Project ref. no.	IST-2000-25426
Project acronym	**VICO**
Project full title	**V**irtual **I**ntelligent **Co**-Driver

Security (distribution level)	Project internal
Contractual date of delivery	June 2001 (M 03)
Actual date of delivery	June 29th 2001
Deliverable number	D3
Deliverable name	Report on Market Situation, Technological Trends and User Expectations
Type	Report
Status & version	Final 1.0
Number of pages	52
WP contributing to the deliverable	WP 2
WP / Task responsible	Robert Bosch GmbH (Bosch)
Other contributors	DaimlerChrysler AG (DCAG), Istituto Trentino di Cultura (ITC-irst) University of Southern Denmark (SDU) Phonetic Topographics (PT)
Author(s)	s. page 2
EC Project Officer	Domenico Perrotta
Keywords	Market Overview, product visions, related projects, technological trends, user expectations and requirements
Abstract (for dissemination)	This report contains a market overview concerning speech-operated in-car systems focusing on the navigation application. The product visions of the consortium, and VICO-related projects are presented. Finally, technological trends as well as user expectations and user requirements are described.

Author(s)	Petra Geutner (Bosch) Frank Steffens (Bosch) Niels Ole Bernsen (SDU) André Berton (DCAG) Marcela Charfuelan (SDU) Fritz Class (DCAG) Paolo Coletti (ITC-irst) Luca Cristoforetti (ITC-irst) Laila Dybkjær (SDU) Marco Matassoni (ITC-irst) Maurizio Omologo (ITC-irst) Luc Peirlinckx (PT) Piergiorgio Svaizer (ITC-irst)

Table of Contents

1 Market Study

1.1 Market Overview

In the following, an overview of commercially available speech-operated in-car systems is given. A general survey of the current market situation concerning driver information systems focusing on the navigation part is presented, and a prospective view into the future is given.

1.1.1 Commercially available Speech-Operated In-Car Systems

Very few systems are currently on the market that allow, in addition to the traditional tactile interface, input by voice. Among them are:

Original Equipment Manufacturers (OEMs)

1. A system by Visteon available in the Jaguar S-type that allows to operate the audio devices radio, cassette and CD-player, as well as the control of the telephone and the climate control by voice. Even though announced several times for the German market, the system still is only available for the English language (see also section 3.2.3.18).
2. A system that can be purchased in addition to the driver information systems available for the BMW 5 and 7 models. The speech control here enables the operation of the telephone by voice, a notebook functionality where short voice messages can be recorded, as well as the speech operation of a very limited set of navigation commands (excluding destination entry of city or streets by voice).
3. A system called “COMAND” (Cockpit Management and Data System) available in the Mercedes S-class. Here, speech control is possible for audio devices as well as the telephone.

After Sales Products

4. A system by Robert Bosch GmbH, VOCS (Voice Control System), that can be purchased in addition to so-called RadioPhone systems that combine audio devices like radio and CD-player or cassette with telephone functionality. VOCS allows speech-enabled operation of the audio functionalities of the RadioPhone as well as the entry of a telephone number.
5. A system called “ECLIPSE Commander 9002” by Fujitsu Ten that is available on the US market. It contains a voice control and navigation unit, and allows speech operation of audio functions, telephone as well as destination entry by spelling. Pauses are required after each two digits respectively letters.

All of the systems introduced above are operated by the usage of a push-to-talk (PTT) button that starts the speech recognition process. Alternatively 5.) can be activated by keyword. All systems operate on a very limited set of keywords or command phrases.

Most of them offer a reasonable good performance when entering telephone numbers by voice. Except 2.) all of the systems also provide the possibility to operate audio devices like radio and cassette or CD-player via speech input. In addition 1.) also offers speech-enabled climate control, whereas 2.) and 5.) provide a very restricted set of commands that can be used within the navigation menu of the system. In addition, 5.) allows destination entry by spelling, where only two letters at a time can be entered without pausing.

Beside the entry of a telephone number all of the systems except 5.) offer the possibility to define name or voice tags (either for entries into a telephone book, or labelling of radio stations or CDs). Both of these functionalities require the usage of small application-dependent dialogues. Concerning dialogue structure only very simple and rigid dialogues are realized.

When looking at the functionality of navigation only two systems offer at least a start into this application by allowing to enter some control commands by speech. Only one system allows simple two-letter spelling of destinations. However, the most important and simultaneously most desirable and difficult parts of input are the entries of destinations such as cities, streets, street numbers and points of interest (POI) in a natural way where large lists of possible destinations may be very confusable by voice.

1.1.2 Speech-Enabled Navigation Systems – from Present to Future

Taking a general view on the market for embedded navigation systems, the distinction between the actual market situation and future products is the following: currently available speech-enabled navigation products for the embedded automotive market are limited to:

1) voice-enabled route guidance using pre-recorded voice or simple low cost and low performance text-to-speech (TTS).
2) destination entry by means of full name recognition is not available in commercial products. Spelling mode (i.e. entering a destination through spelling letter by letter) is available in very few products that can be used as route guidance systems where a static route is calculated. However, in-car navigation or driver information system that provide the possibility of entering a destination by spelling are only announced to appear on the market in 2002.

For these currently commercially available products the following applies:

Time table:	From now on through a couple of years
Cost:	Low cost (no speech engines, no phonetic data is used)
Market:	Staple (mass-) market: high volume but low revenue per item
Speech Content:	Simple and poor
After sales:	Not upgradeable, no after sales market for extra revenue

For the future generation of products in this market segment it is expected that the performance enhancements of both the speech output (TTS systems) and speech input, together with advanced embedded systems and operating systems, will allow more sophisticated operations by speech for future navigation platforms. This includes more natural and intelligible speech output as well as advanced speech input facilities. Concerning speech input a destination entry will be possible by voice of city and street names, or alternatively by POI name. Also an enhanced TTS for route guidance and other services is in pre-development. For these future product developments the following features are characteristic:

Time table:	Estimated implementation period from beginning of 2002
Cost:	higher cost level (speech engines, phonetic

	data)
Market:	Limited edition strategy at the first initial step market size: lower volume but higher revenue per item
Content:	Rich and attractive through high quality speech recognition (SR) & TTS
After sales:	Possibilities for upgrades and extra revenues

Also the upcoming Location Based Services (LBS) should be mentioned. The integration of third generation mobile phones and personal digital assistants (PDAs) will enable user-specific navigation and potentially combined use in the car. As a consequence, a link with the embedded navigation products and markets should be foreseen and anticipated. Speech-enabled LBS are under development, and commercial product launch will depend on the GPRS operators.

Time table:	Estimated implementation period from beginning 2004 or later depending on the whole infrastructure of the operators
Cost:	higher cost level
Market:	Limited edition strategy at the first initial step market size: lower volume, but higher revenue per item
Content:	Rich and attractive through high quality SR & TTS
After sales:	Possibilities for upgrades and extra revenues
Other:	1. Can come into competition with other new to be invented solutions. This is a risk when the launching is subject to a quite long waiting time. 2. As it is open for any other kind of information it is possible that commercial use keep its future alive (publicity, advertising, V-commerce, and the like).

1.2 Product Visions

The expected market growth for speech-enabled applications is enormous and might reach as much as 10,000 million Euro within 3 to 4 years. The Business To Business (B2B) and Business To Consumer (B2C) market will no doubt be significantly affected by industry specific fundamental changes. These changes will most certainly first materialize in Telecommunication, Internet and E-commerce. The breakthrough in these industries is definitely happening now, but with gradually increasing complexity of the dialogue solutions. The full exploitation of speech-enabled geographic databases in car-navigation applications should follow from 2002 onwards.

Research conducted within the VICO project will contribute to a general European technological progress by means of its impact on the navigation and the consumer and GIS related markets. World-wide, results of this project will make it possible to extend the input

towards the ISO standardisation processes, and as such will have an impact of future product developments within this market.

The industrial partners within the VICO consortium consider the development of conversational speech interfaces by means of the concept of an intelligent agent as a crucial success factor for the migration of modern information technologies into the automobile. Therefore, they expect that the results of the VICO project will help to maintain their strong position in the current market. The outcome of the project will provide an excellent position to tackle the future market of next generation information and communication services, especially in the automotive environment represented through driver assistance and information systems. Major innovations developed throughout the project will be exploited directly in products of the industrial partners. This includes the development of products for multimedia services as well as input to the standardization processes for speech-enabled navigational map databases. In the following individual product visions and exploitation plans of the various partners are described.

1.2.1 Bosch

Bosch is Europe's major provider of driver information and entertainment systems featuring a high degree of functionality. Interacting with these systems might interfere with the driver's primary task, namely driving, thus affecting the safety considerably. As a consequence, speech-enabled input is available in currently commercially available Bosch products already. Bosch is not developing speech technology algorithms but concentrates on the application and integration of this technology into the human-machine interface (HMI) of its car multimedia systems with the goal of making them an integral part of the HMI. For this purpose strategic partnerships with leading vendors of speech recognition and speech synthesis have been established.

Commercially available since 1999 is VOCS, an after-sales market product that enables speech input by digits and a small set of command & control words for operating audio devices and telephone. Announced for 2002 is the introduction of the independent platform VASCO (vehicle application specific computer). In addition to traditional audio functionalities, telephone and dynamic navigation VASCO also offers integration of SMS, WAP, e-mail, bluetooth, MP3 and DVD. Voice operation is integrated and input of city and street names out of large lists by spelling will be possible. As a next step destination entry by speaking the name of cities and streets in a natural way is planned around 2003. The possible choices out of city and street lists are restricted by the dialogue, e.g. by selecting the respective region or zip code. For future generations of driver information systems spontaneous speech input is planned.

By means of the natural language technologies developed within the VICO project, Bosch will be enabled to realize user-friendly and safe interfaces for advanced driver information systems. In this way, the market for these systems will be further extended by increased user acceptance. The developed technologies for speech interfaces and dialogue management might furthermore be transferred to the control of other car devices and products from Bosch.

1.2.2 DaimlerChrysler

DaimlerChrysler is active in different fields of speech-input applications. One example is the very important field of traffic systems like cars, trucks and railway systems, where

DaimlerChrysler has operating business subsidiaries. Other areas of importance are civil and military aircrafts like Airbus and European fighter. Furthermore, DaimlerChrysler has activities in telephone speech applications like call center automation.

In all these systems natural speech input will play an important role in upcoming advanced human-machine interaction. Systems get more and more complicated, and system safety will simultaneously play a growing role. Reliability of speech input is a basic requirement for all these applications. Even when considering all the different conditions in the applications mentioned, reliable recognition rates are the critical aspect for a broad acceptance of speech input systems. To this account research related to the project VICO will contribute to further progress in this new technology field as well as future product development.

1.2.3 ITC-irst

In the past years, ITC-irst developed a speech recognition technology for speaker-independent continuous speech recognition. This technology was used for various applications (dictation, data-entry, dialogue over the telephone line, etc.) under different operating systems among which are Linux, Win9x, and WinNT. The technology is also implemented with a set of API, called SPINET, based on a client-server architecture and, from a conceptual point of view, very similar to JavaSpeech API. Both prototype developments and technology transfers are based on the use of this software platform.

Recently, a new release was derived in order to tackle adverse noisy conditions as those characterizing the in-car environment. With this new version of SPINET, a first demonstrator for in-car telephone dialing has already been developed. Main features of this system are: hands-free interaction, no need of push-to-talk functionalities, continuous speech recognition, no need of environmental noise adaptation before using the system. This development was partially supported by a private Italian company interested in this technology.

The VICO project may contribute to further improve this technology for what concerns robustness, and to adapt it to complex dialogue management scenarios, where large vocabularies as well as interaction with dialogue manager and natural language understanding components are required. Also, the degree of portability and flexibility of the given software APIs will be demonstrated by research performed within VICO. Finally, the project will provide a software component set-up for future in-car dialogue system developments based only on ITC-irst technology.

1.2.4 Phonetic Topographics

The VICO project has significant strategic impact as it encourages the use of speech-enabled navigation applications on a European level. Furthermore, it will enhance the availability and quality of speech-enabled data in map databases to be used in various applications.

The qualitative phonetic transcriptions of topographical data enable a high recognition-level, which is a prerequisite for dialogue applications. Particularly the cartographic and topographic markets offer good opportunities for speech-enabling the human-machine interface (HMI). The need of a one-to-one link between the maps and related databases (e.g. points of interest, editorial data, geographic data) is more important than ever. Making this link speech-enabled is the ultimate goal of PT and will influence future product development significantly.

2 Related projects

The following section comprises an overview of projects whose objectives and results are closely related to topics addressed within the VICO project.

2.1.1 CeMVocAS

(**Ce**ntralised **M**anagement of **Vo**cal Interfaces aiming at a Better **A**utomotive **S**afety)

Duration: December 1997, 12 months
Funding: EC ESPRIT 4th framework programme
Partners: AKG (A), Fiat (IT), INRETS (FR), Matra Nortel (FR), MetraVIB (FR), Renault (FR), Universidade Técnica de Lisboa (P)
Website: http://www.inrets.fr/ur/lescot/CeMVocAS/Pagedepa.htm

Today's car drivers have to cope with several tasks in parallel: driving, internal and external communication (with passengers or via the phone), interacting with in-car devices and services, etc. This may lead to a certain degradation in performance particularly in terms of reaction time. CeMVocAS focuses on the problem of divided attention when performing two tasks at the same time. Whereas the usage of speech input already alleviates the usage of in-car systems and services, the attentional load of drivers might still be high when performing specific tasks. As a consequence the overall attention of a driver towards his/her primary task, driving, may still be reduced. Within CeMVocAS a group of human factor specialists, electronic and acoustic engineers as well as a car manufacturer (RENAULT) were working together to develop a centralized vocal information system, also considering the specific situation of an ever growing number of elderly drivers. The main goal of the project was to provide an advanced ergonomic interface relying on robust voice input-output technologies interacting with a driver activity measurement system. This system will continuously give information on the attentional availability of the driver so that external requests are presented to the driver when he is in a situation where his activity allows him to receive such a request. In addition to increased comfort, expected benefits on safety are foreseen for all categories of drivers.

Benefits for VICO:
As VICO aims at providing a conversational interface for a large variety of in-car devices and services, also the problem of the attentional load of a driver has to be addressed. As such, the results of the CeMVocAS project might provide valuable input to the design of the system.

2.1.2 CORETEX

Duration: April 2000, 36 months
Funding: EC IST programme
Partners: Rheinisch-Westfälische Technische Hochschule Aachen - RWTH (DE), University of Cambridge (UK), Istituto Trentino di Cultura – ITC-irst (IT), Centre National de la Recherche Scientifique - CNRS/LIMSI (FR)
Website: http://coretex.itc.it

The aim of the project is to improve core speech recognition technologies. The participants of this project, four world-class speech recognition laboratories, join their efforts to address some of the outstanding research issues:

- Genericity and adaptability
  i.e. the capability of technology to work properly on a wide range of tasks and to dynamically adapt using contemporary data.
- Portability
  i.e. the capability of porting technology to different languages and tasks at reasonable cost.
- Enriched transcription
  i.e. the capability of providing more information than a simple textual transcription which can be used as meta-data for indexing and retrieval purposes.

The objectives of the Coretex project are to develop generic speech recognition technology that works well for a wide range of tasks with essentially no exposure to task-specific data and to develop methods for rapid porting to new domains and languages. A further research area covers enriched transcription, which aims to produce an enriched symbolic speech transcription with extra information for higher level (symbolic) processing.

Benefits for VICO:
Speech recognizers are quite sensitive to the acoustic properties of the data, and in particular to mismatches between the training and the real usage conditions. The adaptability issues addressed by the Coretex consortium may offer to VICO new methods to improve the acoustic modeling or quickly adapt to different conditions.

2.1.3 DARPA Communicator

Duration: 1998, 72 months
Funding: **D**efense **A**dvanced **R**esearch **P**rojects **A**gency (DARPA) of the United States Government
Participants (Communicator sites): Carnegie Mellon University (US), Center for Speech and Language Research at the University of Colorado at Boulder (US), MIT Spoken Language Systems (US), Artificial Intelligence Centre at SRI International (US), MITRE (US), IBM (US), AT&T (US), BBN (US)
Website: http://fofoca.mitre.org

2.1.3.1 Goals and architecture

The overall objective of the DARPA COMMUNICATOR (1998-2003) project is to support rapid, cost-effective development of speech-enabled dialogue systems. There are several enabling goals for the Communicator program. These are [1]:

- To provide a common architecture, so that researchers can furnish subcomponents without having to build an entire system.
- To provide a test bed with shareable components that lower the entry bar to building speech-enabled dialogue systems.
- To provide a shared research environment, including common data and a common evaluation framework, to encourage cross-group comparison a rapid sharing of technological innovations.
- To further innovate research on dialogue management and interface design to support conversational systems.

- To encourage the transfer of this technology to real users, in particular, military users.

The program has chosen MIT's Galaxy II architecture as its common architecture. This architecture consists of a central hub that controls the flow of information among a suite of servers, which may run on the same machine or at remote locations. The hub interaction with the servers is controlled via a scripting language. A hub program includes a list of the active *servers*, specifying the host, port, and set of operations each server supports, as well as a set of one or more *programs*. Each program consists of a set of rules, where each rule specifies an operation, a set of conditions under which that rule should "fire", a list for INPUT and OUTPUT variables for the rule, as well as optional STORE/RETRIEVE variables into/from the discourse history. When a rule fires, the input variables are packaged into a token and sent to the server that handles the operation. The hub expects the server to return a token containing the output variable at a later time. The variables are all recorded in a hub/internal master token. The hubs communicate with the various servers via a standardised frame-based protocol [2].

A number of groups are now building systems using the Galaxy architecture and hub, coupled with in-house developed servers. These systems provide end-to-end functionality in the initial Communicator challenge task, air travel planning.

CU-Move project

One of the groups that are building systems using the Galaxy architecture is the Center for Spoken Language Research (CSLR) at the University of Colorado, Boulder (USA) [3]. Here one of the research projects is CU-Move (DARPA In-Vehicle Automatic Speech Recognition & Navigation System) [4]. The goal of the University of Colorado CU-Move project is to develop algorithms and technology for robust access to information via spoken dialogue systems in mobile, hands-free environments. The novel aspects include the formulation of a new microphone array and multi-channel noise suppression front-end, corpus development for speech and acoustic vehicle conditions, environmental classification for changing in-vehicle noise conditions, and a back-end dialogue navigation information retrieval sub-system connected to the web. The CU-Move system [5] consists of a front-end speech collection/processing task that feeds into the speech recognizer. The speech recognizer is an integral part of the dialogue system (tasks for understanding, discourse, dialogue management, text generation and TTS). The back-end processing consists of the information server, route database, route planner and interface with the navigation database and navigation guidance systems. Focus is in particular on multi-channel noise suppression, automatic environment characterization, and on a prototype navigation dialogue.
A similar application, an in-vehicle navigation system, is being developed and implemented by one of the CU-Move project collaborative partners, HRL Laboratories [6]. A description of this system can be found in [7].

CU-Move results

A prototype dialogue system for data collection in the car environment has been developed within CU-Move [5]. The dialogue system is based on the MIT Galaxy-II Hub architecture with base system components derived from the CU communicator system. Users interacting with the dialogue system can enter their origin and destination address by voice. Currently, 1,107 street names for Boulder, Colorado area are modeled. The dialogue system automatically retrieves the driving instructions from the internet using an online web route direction provider. Once downloaded, the driving directions are queried locally from an SQL database. During interaction, users mark their location on the route by providing spoken odometer readings. Odometer readings are needed since global positioning system (GPS) information has not yet been integrated into the prototype dialogue system. Given the

odometer reading of the vehicle as an estimate of position, route information such as turn descriptions, distances, and summaries can be queried during travel (e.g., "What's my next turn", "How far is it", etc.).

Benefits for VICO:
To decide the optimal system architecture for VICO, different varieties of system design and information exchange are currently being looked at. Among other solutions, like e.g. CORBA (Common Object Request Broker Architecture), the central hub in Communicator should be looked at as a possible solution to system management in VICO. If the existing hub architecture is available for projects outside communicator and well-suited for VICO, a considerable workload might be saved.

2.1.4 DISC

(Spoken Language **Di**alogue **S**ystems and **C**omponents: Best practice in development and evaluation)

Duration: June 1997, 42 months
Funding: EC/HLT
Partners: Natural Interactive Systems Laboratory (NISLab) (DK), Human-Machine Communication Department at the Centre National de la Recherche Scientifique (CNRS-LIMSI) (FR), Institut für Maschinelle Sprachverarbeitung (IMS) Universität Stuttgart (DE), Department of Speech, Music and Hearing Kungliga Tekniska Högskolan (KTH), Vocalis Ltd (UK), DaimlerChrysler AG (DE), ELSNET
Website: http://www.disc2.dk

The DISC project and its successor DISC2 lasted from 1997 to 2000. The aim of the projects was to develop a first Best Practice Guide on how to develop and evaluate spoken language dialogue systems and their components.

The first phase of DISC was dedicated to the development of current practice reviews of the DISC spoken language dialogue system (SLDS) aspects (see below), a detailed best practice dialogue engineering development and evaluation methodology, and a range of design support concepts and software tools. The second phase, i.e. DISC2, focused on testing the validity and usability of the draft Best Practice Guide, the concepts and the tools, and on the integration, packaging and dissemination of the final DISC Best Practice Guide.

The DISC current practice reviews charted current SLDS's development and evaluation practice, producing about 50 in-depth analyses of existing SLDSs and components together with the following approach to dialogue engineering best practice. An SLDS is viewed as having six major *aspects:* speech recognition, speech generation, natural language understanding and generation, dialogue management, human factors, and systems integration. Each aspect can be analyzed in terms of a *'grid'*. A grid contains an aspect-specific description of the state-of-the-art technical problem space facing the developer, including technical properties, interrelationships among properties, and advice on which properties to include in particular applications. Within the grid problem space – or outside it, since new ideas appear all the time - the developer must make the decisions most appropriate for the application to be developed. In DISC, the grid problem space is structured in terms of the *issues* facing the developer, the *options* the developer must choose from per issue, and the *pros and cons* with respect to each option.

Orthogonal to the "static" grid description, each aspect may be analyzed in terms of a development *life-cycle* which decomposes the development process into iterative phases and issues to be addressed in each phase. Integral to the life-cycle is the continuous *evaluation* of progress and results. As DISC progressed, evaluation of SLDS aspects gained prominence due to the many unsolved research issues in the field. In response, DISC has developed a generic *evaluation template* [8] which can be used to characterize each evaluation criterion for use in evaluating aspect-specific properties of SLDSs. Furthermore, the grid analyses have been used to systematically generate a set of evaluation criteria per aspect.
In addition to the above, best practice guidance must incorporate guidance on available platforms, methods and supporting tools per aspect. These have been surveyed in a series of DISC reports. Moreover, DISC itself has produced a series of development support tools and guidelines. These include:

- guidelines and testing protocols for the development of speech recognition components for SLDSs;
- a software tool for evaluating speech synthesis components for SLDSs;
- guidelines for the acquisition of lexical data for SLDSs;
- CODIAL, a software tool in support of cooperative system dialogue design;
- SMALTO, a software tool in support of speech functionality (pertaining to what the speech modality is (not) good for) decisions in early design.

DISC results
The core result of DISC is the web-based DISC Best Practice Guide (www.disc2.dk) which resulted from turning everything mentioned above into a comprehensive website. In addition to the ingredients described above, the DISC Best Practice Guide includes a comprehensive *glossary* of dialogue engineering terminology, *references* to the literature including all DISC publications, and brief *checklists* per aspect.

Benefits for VICO:
The DISC Best Practice Guide developed in this project as well as the developed software tools will be applied during the development and evaluation of VICO.

2.1.5 NESPOLE!

(**NE**gotiating through **SPO**ken **L**anguage in **E**-commerce)

Duration: January 2000, 30 months
Funding: EC IST programme, NFS
Partners: ITC-irst (IT), Carnegie Mellon University (US), Universität Karlsruhe (DE), Université Joseph Fourier (FR), Aethra (IT), Azienda per la Promozione Turistica (APT) del Trentino (IT)
Website: http://nespole.itc.it

Nespole! project is an American European project aiming to improve the technology of speech-to-speech translation. The project realizes two prototypes: the first one shows a touristic translation system through internet (using Microsoft Windows Netmeeting); the second one enlarges the domain of the first one and adds a help desk domain. Both systems are designed to let a non-expert client use the system talking with an expert agent.

Nespole! in its early stage studied touristic dialogues: what do people ask, how do they ask information, how do touristic agents provide information. The main issue is that the user asks for redundant information, which can more easily be found on the APT website, and quite often wants printed or visual material. Other frequently asked topics include the immediate availability of rooms in hotels, and telephone numbers.

Data collection in Nespole! consisted of monolingual dialogues between a tourist operator, located in APT, and a naive tourist who had to follow a predefined scenario, talking in his/her language from Germany, France, US or Italy.

The dialogues' language was very free and, as a result, many out-of-domain sentences were observed. These dialogues were transcribed manually and then used, but only for statistical language modeling purposes as the corresponding audio quality was very poor (due to the distortion introduced by the real internet connection). The dialogues also included exchange of multimedia material and use of gestures for a multimodal interaction.

Benefits for VICO:
Nespole!'s email study will help VICO gain experience with standard topics asked by tourists and their mean way of interacting with an already present information. Nespole!'s data collection provides also a lot of examples about spontaneous monolingual touristic dialogues.

2.1.6 RESPITE

(**RE**cognition of **S**peech by Partial **I**nformation **TE**chniques)

Duration: January 1999, 36 months
Funding: EC ESPRIT programme
Partners: University of Sheffield (UK), DaimlerChrysler AG (DE), FPM (BE), IPC (FR), BaBel Technologies(B), IDIAP (CH), ICSI (US)
Website: http://www.dcs.shef.ac.uk/research/groups/spandh/projects/respite

Background:
Automatic Speech Recognition has made its way from research to products within the last few years. These products, like dictation systems or simple command & control systems, operate reliably only in predictable, quiet situations. Now that speech recognition has gained user acceptance in quiet surroundings, it has to be made robust in noisy, real-world environments. To achieve this, new methods have to be developed which enable the recognizer to handle strongly deteriorated speech signals as well as a human listener.

Objectives:
Research within the RESPITE project focuses on new methods for robust Automatic Speech Recognition. All modules of a state-of-the-art speech recognizer are being investigated and extended using new approaches, such as missing-data theory, multistream classification and hybride decoding. RESPITE will provide a closer link between the theory of human hearing perception and the modeling of speech in recognition systems. Computational Auditory Scene Analysis and long term modulation-filtered spectral features will help to increase the robustness of the recognizers front-end. These new methods will target recognition in noisy environments and will hence be of a great value for in-car and cellphone applications. New algorithms are evaluated using the Aurora-II connected digit database, featuring various noise conditions.

The specific measurable objectives are:

1. to develop techniques for identifying reliable data;
2. to advance the theory of multistream processing;
3. to advance the theory of missing and masked data handling;
4. to obtain new perceptual data on speech recognition to be used in the above;
5. to combine missing data and multistream processing with existing robust automatic speech recognition (ASR) techniques;
6. to evaluate all of the above within a framework of demonstrator ASR applications (cellphone and in-car applications).

Scientific Highlights:

- Identifying Reliable Evidence
  The maximum likelihood decoding process has been extended with a dynamic combination of subband likelihoods using expert weights for each subband. The cues for consonant identification were investigated in a psychoacoustic study, which proved that the point of articulation for consonants is a primary spectral cue and that voicing is a robust consonant feature.
- Dealing with Missing Data
  The gist of the Missing Data theory is to recognize speech only in those spectral-temporal regions which carry reliable speech evidence. It was found that even in noisy speech, some spectral-temporal regions hardly suffer from the signal corruption. Two methods, „marginalization“ and „state-based imputation“ have been proposed which base the recognition only on relatively clean speech regions. The evaluation of these methods on speech under non-stationary factory-noise conditions showed a significant 30% relative improvement in recognition accuracy.
- Multistream Formalism
  A new method, which was dubbed „Tandem approach“, was developed. This approach combines a neural network with Gaussian mixture models that are used in state-of-the-art recognition systems. The neural network estimates the probabilities for the Gaussian mixture weights. This combination leads to another highlight, which is the combination of multiple simultaneous feature representations, which can easily be combined after the neural network stage. Experiments on the Aurora-II database showed in a remarkable 30% – 50% relative improvement in recognition accuracy.

Benefits for VICO:
The new techniques for robust speech recognition, particularly the new multistream front-end offers a great potential for improving the recognition rate of speech in noisy environments such as the VICO in-car application. The Tandem approach should be investigated using different feature representations that cover both short term and long term speech properties. The approach has to be extended to both context-dependent models and also to large vocabularies.

2.1.7 SENECA

(**S**peech control modules for **E**ntertainment, **N**avigation and communication **E**quipment in **CA**rs)

Duration: May 1998, 42 months
Funding: EC ESPRIT 5th framework programme

Partners: Robert Bosch GmbH (DE), DaimlerChrysler AG (DE), TEMIC Telefunken (DE), Motorola Germany (DE), Motorola Semiconductor Israel (IL), Centro Ricerche Fiat CRF (IT), Renault Recherche Innovation (FR)
Website: http://www.seneca-project.de

Within the project SENECA a robust and inexpensive multilingual speech dialogue demonstrator system has been developed in three languages (German, French and Italian). The developed system uses command-word based speech input in order to control navigational, entertainment and communication devices in cars and trucks. As a major objective of the project is to integrate and further develop speech recognition technology for use in the car, the following goals have been pursued throughout the project:

- to extend the functionality of speech recognition from spelling to full-word recognition on destination input (for a vocabulary of approx. 3,000 words)
- to achieve more robust speech recognition
- to reduce the cost of the feature speech recognition for the end user

Special focus has been put on speech enhancement and the development of a two-channel solution for noise reduction. Also higher quality for hands-free telephony is achieved by means of doubletalk echo compensation.

The developed demonstrators in the respective three languages have been evaluated focusing on user acceptance and usability tests. The evaluations have been conducted on driver-level in a moving car.

Benefits for VICO:
For VICO results of SENECA concerning robust speech recognition as well as innovative methods for improved destination entry should be considered. Also, results of the performed user acceptance and usability tests are of high concern for the development of the VICO system.

2.1.8 SIRIDUS

(**S**pecification, **I**nteraction and **R**econfiguration **I**n **D**ialogue **U**nderstanding **S**ystems)

Duration: January 2000, 36 months
Funding: EC IST programme
Partners: SRI International (UK), Universidad de Sevilla (ES), Goeteborg University (SE), Universität des Saarlandes (DE), Telefónica Investigación y Desarrollo (ES)
Website: http://www.cam.sri.com/siridus/index.html

Spoken language dialogue systems, such as automated telephone inquiry systems and hands-free in-car device control, are rapidly evolving towards commercial products. SIRIDUS aims to improve the understanding of what is required to provide reusable, robust and user-friendly spoken dialogue systems.
Particular concerns in SIRIDUS are:

- achieving robustness when user utterances are unpredictable and speech recognition is noisy
- showing that generic strategies for dialogue management can be applied to a wide range of dialogues including "command" dialogues and negotiation dialogues

- providing architectures which allow appropriate sharing of information between modules, in particular enabling dialogue systems being sensitive to different stress of individual words.

The SIRIDUS project is building two main demonstrators. The first is a telephone operator dialogue system in Spanish which already exists in a text-to-text version. A first version of an automated speech-to-speech telephone operator system will be available in 2001. The second demonstrator is an integrated toolset for dialogue researchers. The aim is to provide both a library of modules and a toolkit in which researchers can plug in their particular module to test its effect on a whole system.

The project aims at providing innovative research which is applicable for real systems in the near future. Within the first year of the project innovative research work on how to provide systems for natural command languages and negotiation dialogues has been provided.

A key to the work on robust interpretation is that it provides a uniform way to express rules based on keyword/key phrase spotting or more detailed linguistic descriptions. This makes it easier to ensure that the system always performs at least as well as a keyword/key phrase spotting system, since extra linguistic information is only used if it is both available and likely to be helpful. Innovative work is aimed at integrating repair strategies for robustness, and to show that also more detailed linguistic descriptions can be used for approaches which aim at a full linguistic analysis of user utterances.

Benefits for VICO:
As an important research topic within VICO is the design of an intelligent and user- and situation-adaptive dialogue, the second demonstrator of SIRIDUS should closely be examined. The emerging integrated toolset might be reusable for the VICO project.

2.1.9 SMADA

(**S**peech Driven **M**ultimodal **A**utomatic **D**irectory **A**ssistance)

Duration: January 2000, 36 months
Funding: EC IST prgramme
Partners: Koninklijke KPN (NL), Stichting Katholieke Universiteit (NL), France Télécom - CNET (FR), Université d'Avignon et des Pays de Vaucluse (FR), Alcatel Deutschland (DE), Centro Studi e Laboratori Telecomunicazioni - CSELT (IT), Politecnico di Torino (IT), Swisscom (CH).
Website: http://smada.research.kpn.com

SMADA lasts from 2000 to 2002. The overall objective of the SMADA project is to improve all aspects of the technology needed to automate a large part of the calls to a Directory Assistance service without compromising customer satisfaction. Directory Assistance (DA) is the service that provides information on telephone numbers, once a name and address are provided by the customer. The project has as its starting point the current prototypes of four Telecom Operators - KPN, France Telecom, Telecom Italia and Swisscom - which aim to automate part of their DA service.
SMADA aims at a breakthrough in the robustness and accuracy of automatic speech recognition (ASR) technology by developing confidence measures, noise robust decoding, progressive search and techniques for unsupervised learning from the speech recorded during operation of the service. There are large differences in the number of cities that must be

recognized. The Italian system contains a lexicon of 9,325 city names. The French demonstrator system aims for the recognition of about 36,000 French town names. For the Dutch system a lexicon of about 2,400 is sufficient to cover half of the named localities in the Netherlands. The Swiss system that only covers the German speaking part of the country, contains a lexicon of about 2,600 city names [9].
SMADA will build demonstrators for uni-modal voice access and multimodal access to web-based directory assistance services, which will be used for realistic field trials. The project will carry out human factors research addressing issues related to caller-system interaction, and in the case of partial automation, operator-system interaction. Attention will focus on calls from fixed and cellular networks. Protocols supporting distributed speech recognition will also be investigated.

SMADA results
In the SMADA project [9], three evaluations of the Directory Assistant systems are planned: one took place at the end of 2000, one will be performed in the middle of the project and one at the end of the project (December 2002). The last two evaluations include human factors and technology, the first one was mainly focused on technology.

Human Factors evaluation
The best speech recognition performance will be reached when customers stick to predictable expressions as much as possible. To accomplish this, much attention must be paid to the formulation of prompts. In the Italian system many different prompt formulations have been tried. The best prompts turned out to be those that give explicit examples of what customers should say. In the Dutch demonstrator system, short direct prompts were used. Also these prompts resulted in a relatively large proportion of predictable utterances for city names (85%). The remaining 15% include silence, utterances that are too long and out-of-vocabulary (OOV) utterances. However, a marketing study involving a large number of customers showed that the prompts were considered as rather unfriendly and impolite.

Technology evaluation
First experiments by CSELT have shown that substantial improvements can be obtained for city name recognition by adding a small number of application-specific acoustic models. Also a richer set of acoustic features helped to improve the performance. Some experiments were carried out on a spontaneous speech database consisting of 8,775 tokens of city names recorded in the Italian Directory Assistant service prototype. The use of vocabulary dependent sub-word units for the most common city names significantly improved the performance; further improvement was obtained by the addition of word models for the most frequent city names. The following table summarizes the results obtained with Continuous Density HMMs.

Models	Errors	Error rate %
Vocabulary independent	1536	17.8%
+40 vocabulary dependent sub-words	1122	12.8%
+20 word models	1000	11.4%

Benefits for VICO:
The recognition of city names is one of the difficult problems that has to be faced in VICO. SMADA techniques such as addition of application-specific acoustic models or vocabulary dependent sub-word units for the most common city names could be tested in VICO.

2.1.10 SmartKom

(Human Machine Interaction using coordinated analysis and generation of multiple modalities)

Duration: January 2000, 48 months
Funding: BMBF+industry
Partners: DFKI (DE), DaimlerChrysler AG (DE), Philips GmbH (DE), Siemens AG (DE), Sony International GmbH (DE), European Media Lab(DE), MediaInterface Dresden GmbH (DE), FAU, IMS (DE), LMU (DE), ICSI (US)
Website: http://smartkom.dfki.de

Background:
State-of-the-art user interfaces for Human Machine Interaction offer simple, template dialogues, (the initiative is always with the machine), or command & control systems. One of the great challenges in the information society is to create intelligent multimodal and multimedia user interfaces that allow for natural communication and thus can also be used by people unfamiliar with high-tech equipment. Efficient and natural interfaces will be needed in the near future to access information and to facilitate new applications.

Objectives:
The goal of SmartKom is to create an innovative new user interface which incorporates not only speech dialogues but also mimics, gestures, and graphics. Information services such as cinema and hotel reservations, navigation guidance, tourist and weather information, etc. will be provided via a new interface.
SmartKom will provide the following improvements to current dialogue systems:

- Semantic interpretation and combination of mutually complementing input modalities, e.g. speech, gesture, mimics, touchscreen/graphics, biometrics
- Robust processing of each input channel in spite of signal deterioration
- Context-sensitive interpretation of dialogue interaction using dynamic discourse and task models
- Adaptive generation of coordinated, cohesive multimodal presentations
- Automatic completion of delegated tasks using integrated information services
- Personalized presentation agent and intuitive help function
- Adaptive dialogue control using situation- and user-adapted interaction models

Scenarios:
The following functionalites will be investigated in three application scenarios:

- Multimodal communication booth:
  A public phone booth will be transformed into a multimodal communcation booth, which will offer information services via broadband transmission. This communication booth integrates a telephone (speech dialogue), a display (touchscreen), a document camera (transfer of documents), and an infrared camera (gesture). Connection ports for PDAs and camcorders will allow data exchange with personal gadgets. A personal chip card will enable a data connection to the PC at home or at office. The user-specific style and the last dialogue state will be stored on the chip card if this feature is desired.
- Mobile communication assistant

The mobile communication assistant will constantly accompany the user at his office, in his car, at home, or while walking. It will offer internet and phone connection via GSM and location-based services using an integrated GPS antenna. A microphone, loud-speakers, a display and a camera complete the PDA. This allows for input through speech, mimic and touchscreen.

- Intuitive working environment at home and in the office
  The home computer becomes a control centre for the intelligent house. Home entertainment and house technology (air condition, humidifier) can be controlled by both touchscreen or voice activation. Interaction is based on everyday life habits but nevertheless meets the varying needs of the user and adjusts to the user‘s communication abilities.

Benefits for VICO:
The speech database collected at the University of Munich will help the VICO project to gain experience with noisy, spontaneous speech dialogues. Recognizers will be adapted to noise conditions similar to those in VICO. Statistical class-based language models are investigated by Sony. These models might be advantageous to cover spontaneous speech and to reduce the perplexity of speech input. Based on the language model training procedure provided by Sony, DaimlerChrysler hopes to be able to either improve our language model training or implement a new training procedure from scratch. The decision depends on the results achieved in the SmartKom project. The SmartKom dialogue structure of the mobile scenario will give the VICO partners a first insight into how navigation and hotel information tasks can be implemented using natural speech input.

2.1.11 SpeechDat-Car

(**Speech** **Dat**abases for Voice Driven Teleservices and Control in Automotive Environments)

Duration: March 1998, 30 months
Funding: EC Telematics Applications programme
Partners: Lernout & Hauspie France (FR), Alcatel Mobile Phones (FR), University of Aalborg (DK), Robert Bosch GmbH (DE), BMW (DE), Digital Media Institute (FI), ITC-irst (IT), Lernout & Hauspie (BE), Knowledge (GR), Nokia (FI), Renault Recherche Innovation (FR), Seat (ES), Speech Processing Expertise Centre-SPEX (NL), Universität München (DE), Universitat Politecnica de Catalunya (ES), Vocalis (UK), VW (DE)
Website: http://speechdat.phonetik.uni-muenchen.de/SP-CAR

Within the European project SpeechDat-Car large-scale spoken language resources (speech databases) have been collected for developing in-car voice-driven information services and control systems. In total speech databases in 9 European languages (English, Danish, Finnish, Flemish-Dutch, French, German, Greek, Italian and Spanish) covering 300 different speakers have been collected. Recordings were performed with speakers of different dialects and gender, also varying the environmental noise and driving conditions (windows open/closed, fan on/off, varying car speed, weather and road conditions). Several microphones have been used from close-talk signals up to far microphone signals for hands-free interaction. The data collections consists of:

- Isolated digits
- Digit strings
- Natural numbers
- Monetary amounts

- Dates and times
- Names
- Spelling
- Phonetically rich words and sentences
- Pre-selected command words
- Application keywords and spontaneous sentences

Benefits for VICO:
As ITC-irst has participated in the project and owns the Italian data, this part of the database will be used for training of the Italian speech recognition engine.

2.1.12 SPEECON

(**SPEE**ch Driven Interfaces for **CON**sumer Applications)

Duration: February 2000, 24 months
Funding: EC IST programme+industry
Partners: Siemens AG (DE), Ericsson (DE), IBM (DE), Nokia (FI), Lernout & Hauspie (BE), Matra Nortel Communications (FR), DaimlerChrysler AG (DE), Philips Speech Processing (DE), Sony International (DE)
Website: http://www.speecon.com

Background:
As speech recognition technology finds its way from the research and development labs to the market, consumer devices for many languages and acoustic environments will incorporate this technology as a new mode of interaction. Examples of such devices are mobile telephones, TV-control sets and car navigation kits. In order to transfer the speech-driven interfaces from one to various other languages and different acoustic environments, large language and application-specific speech databases need to be collected for recognizer training.
Speech databases covering 18 languages and typical acoustic environments of the application areas are developed. Due to the fact that these databases cannot cover all environmental conditions sufficiently, new methods are investigated for adapting the speech databases to the specific acoustic environment of the used device.

Objectives:
The SPEECON project focuses on the transfer of speech recognition technology to many languages and acoustic environments. Particularly, the project deals with speech recognition interfaces for consumer devices such as mobile telephones, TV-control and home appliances. The feasibility of the transfer approach is shown by three demonstrators.

Developments :
Before collecting speech data for several languages and acoustic environments, a detailed requirement analysis and specification phase has to be undertaken:

- Market analysis for voice-driven consumer devices
- requirements for the functionality of the recognizers
- specification of the speech databases
- definition of the recording platforms, the acoustic environments for recording and the corpus

- specification of annotation
- specification of speakers to be recorded

Eighteen speech databases will be created by the steps :

- building and testing the recording platform
- recruitment and recording of speakers
- annotation of the recorded items.

An external evaluation centre will assess the quality of the databases

While collecting speech databases, the partners also investigate algorithms that allow for transferring a given database from one acoustic environment to another. Such new methods could significantly decrease the effort to collect databases.

Benefits for VICO:
Collected speech databases in German, English and Italian might be included in the training of acoustic models for VICO to improve speech recognition accuracy. New algorithms to transfer speech databases from one acoustic environment to another could enable the VICO consortium to use speech data from different databases and transfer them to the VICO typical environment. Hence, far more speech material would be available for training the acoustic models.

2.1.13 SPOTLIGHT

(Mass Market eCommerce Services Using Multi-language Natural Spoken Dialogues)

Duration: January 2000, 36 months
Funding: EC IST programme
Partners: University of Edinburgh (UK), Lloyds-TSB Group (UK), British Midland Airways (UK), Periphonics Voice Processing Systems (UK), Advance Bank (DE), Deutsche Bahn Reise & Touristik (DE), Centro Studi E Laboratori Telecomunicazioni - CSELT (IT), Comune di Roma (IT), SARITEL (IT**).**
Website: http://spotlight.ccir.cd.ac.uk

SPOTLIGHT (Mass Market eCommerce Services Using Multi-language Natural Spoken Dialogues) lasts from 2000 to 2002. The project aims to research and develop advanced technology demonstrators for extending the use of spoken natural language technologies for mass market eCommerce services. The demonstrators being developed are in two domains: the financial services sector and the travel industry in Europe. Customizable services are being developed for German, English and Italian. The use of three different grammars in two different industry sectors will form a solid basis for evaluating and validating the SPOTLIGHT concept. The technology demonstrators created during the project will be tested with major European user groups in each of the three countries of the project.

The objective of the project is to define a series of demonstrators which will produce a visible impact on the market positioning of spoken natural language technology and its application in the two sectors. The project aims to produce, as a first step, a series of requirements definition reports detailing the demonstrators. A total of 21 spoken natural language demonstrators are planned by SPOTLIGHT. The project demonstrators will subsequently be made widely

accessible in a web 'showcase' so that they can be experienced by a wide audience of interested industrial and consumer users.

SPOTLIGHT results

To date, five demonstrators have been successfully completed in the first phase of the project. These are [10]:

1. British Midland Airways Flight Status Demonstrator - allows travellers to check on the status of their specific flight, including inquiries about flight delays. It has been designed to incorporate several key dialogue features, including the ability to handle spoken natural language and to recognize and deal with utterances containing multiple semantic slots. Another key feature of the Flight Status demonstrator is the use of conditional confirmation. A usable interaction would involve the use of confirmation only when the service is unsure about what has been said. By making use of the confidence scores passed back from the recognizer this capability has been introduced into the dialogue for the demonstrator. Whole utterances or semantic slots whose confidence scores lie above a carefully defined threshold are accepted without need for confirmation, thereby speeding up the interaction and reducing the need for confirmation. The Flight Status demonstrator also has the ability to deal with ambiguous input, like a city name when the city contains more than one airport and/or a time which is ambiguous e.g. eight o'clock. In the case of an ambiguous airport a dynamic grammar of the possible airport names is created, and the caller is asked to select from the various options. Finally, throughout the dialogue, three levels of error recovery are employed. The error handling is designed to distinguish between silence errors (where no speech is detected) and reject errors (where an out-of-grammar utterance is detected). The advantage of these conditional error messages is that callers can correct errors more effectively if they know more about them.

2. Lloyds TSB Banking Demonstrator - features a wide range of banking services, such as balances, funds transfer and account monitoring. The dialogue design allows the use of spoken language inputs where customers can use speech to input multiple bits of information at a time. The spoken language dialogue is supported by a system of error recovery dialogues. Each error recovery level contains feedback for repair of the type of error made (e.g. silence) and then a re-prompt with additional information which aims at clarifying the required input at any specific point in the dialogue.

3. Comune di Roma Taxation Data Purity Demonstrator - is a 'front-end' to a system for the administration of waste management taxation in the city of Rome. It allows private citizens of the Commune to use a voice-operated automated telephone service to enter their fiscal data into the administration database. Most of the data items are recognized using a grammar-driven approach, where grammar rules of differing levels of complexity have been developed for the recognition of items such as date, fiscal code and occupation code which may be entered by callers in many different ways. In the case of city and province of birth, the demonstrator uses an innovative approach driving the speech recognizer with the data contained in a relational database which is also used to drive the dialogue. This new approach allows the system to successfully retrieve the correct city (and province) from a list of more that 13,600 items.

4. Deutsche Bahn Reise & Touristik National Rail Timetable Demonstrator - is designed to handle the 25% of customer inquiries not involving a transaction (i.e. sale of ticket). The demonstrator is built around a number of dialogue stages, including: request for departure station, request for arrival station, request for time of travel, announcement of the connection, option to request return connection, input of new time or new route. The DB National Train Timetable Demonstrator handles train timetable information on both long-distance and regional train routes in Germany and other countries. It incorporates approximately 7,000 locations serviced by Deutsche Bahn. The demonstrator has the ability to deal with

ambiguous input in cases where callers supply an inadequately defined station name. When this is detected and appropriate statistical information exists, the system provides the caller with a possible station name to confirm. If such statistical information does not exist, the caller is presented with a list of station names to choose from – these are presented in blocks of five to reduce calling time. Taking into account the large number of station names that the service can recognize and the particularities of some German towns (e.g. more than one town with the same name, distinguished only by the region they are in), the input of ambiguous names is a common experience. The above feature ensures that these are correctly identified and dealt with. The application is able to recognize input containing multiple semantic slots in the case of travel time input and three levels of error recovery are used throughout the dialogue.

5. Telethon Demonstrator - a transaction-based service which enables members of the public to give donations to charity as part of a national charity appeal campaign. The service is operated in conjunction with television appeals ("Telethons") and has previously allowed touch-tone input only.

Benefits for VICO:

The use of confidence scores and different levels of error recovery are features that could be applied in the VICO project. SPOTLIGHT also addresses the problem of ambiguous station name recognition. Their approach to deal with ambiguous input names could be interesting for VICO.

3 Technological Trends

3.1 Trends in internet and telematic services

In-car application of telematics concerns the integration and convergence of wireless systems, global positioning, onboard computer and onboard automotive electronics.

The vision of the industry is to enhance driver and passenger safety, productivity, and security through communication, information, and convenience services. Telematics systems provide interaction between the vehicle and some outside service provider who delivers traffic information, news, e-mail and other data useful to people on the move. From the automotive manufacturers' viewpoint, telematics is a mechanism for "transforming the automobile into the next mobile portal" or a "hands-free in-vehicle communications system."

Most common telematics systems on the road today are navigation systems using the global positioning system (GPS). This system is available in many different car models, but is most commonly provided in high-end luxury vehicles. The following companies include telematics systems in their vehicles, either as an available option or as a standard feature: General Motors, Mercedes-Benz, Peugeot, Lincoln, Jaguar, BMW, Infinity, Toyota, Fiat, Honda, and Mazda. Alliances and relationships among service providers, content providers, telecommunication companies, automotive original equipment manufacturers (OEMs), and electronic hardware providers will produce the network coordinating telematic content and services.

Current uses for telematics is primarily for safety and security, and for navigation. Examples include: 24-hour roadside assistance, accident alerts on airbag activation, theft tracking, remote door unlocking, in-vehicle digital mapping, access to call center to get directions, traffic alerts, traffic accident reports and re-routing. Future uses may include in-vehicle Internet access, "infotainment" (information + entertainment), in-vehicle E-commerce (purchase of gas or parking, or download of music or information), location-based commerce and advertising (e.g. special offers from stores near the current location), remote diagnostics (e.g. potential problem of the car can be monitored or even fixed remotely), mobile office facilities, virtual meeting participation, potential collision alerts and augmented reality assistance to driving.

Currently telematics systems are already provided by some vehicle producers: General Motors, Ford, DaimlerChrysler, Toyota, Honda, Renault, Fiat, BMW, VW, Audi. These are mainly based on GPS with hardware by Delphi, Visteon, Motorola, Bosch, Alpine, Matsushita and are connected to service providers as OnStar (General Motors, Toyota, Honda), Wingcast (Ford, Nissan), ATX (Mercedes, Lincoln, Jaguar), Tegaron (Renault, Fiat, Volkswagen, Audi), Mannesmann Passo (BMW).

General Motors Corp.'s OnStar system is one of the first system brought to the market. At its heart it's a cellular phone, but the system adds a variety of driver-friendly features. Drivers can use it to get directions, schedule airline reservations or find a local dealer should they break down while traveling. If keys have been accidentally locked in the car, OnStar's service center can transmit a signal telling the car to unlock its doors. OnStar even sends an automatic alert to authorities in the event of an accident serious enough to trigger a car's airbag.

General Motors is estimating four million OnStar subscribers by the end of 2003. Ford is estimating that telematics services will be in virtually all of its cars and trucks by the end of 2004.

In the following the functionalities of current and envisaged products for in-car telematics are briefly summarized.

3.1.1 Safety and Security

Beside in-car driver assistance systems like adaptive cruise control (ACC) or collision warning systems, the following telematics safety and security systems and services should be mentioned.

3.1.1.1 Automatic Vehicle Tracking

A vehicle with a GPS receiver can be queried for its precise location even from a cell phone.

3.1.1.2 Automatic Emergency Call

In the event of an accident serious enough to set off a vehicle's airbags, the system automatically calls for help over the vehicle's cell phone (e.g OnStar service).

3.1.1.3 Stolen Vehicle Recovery Systems

This technology provides a crime-fighting twist on automatic vehicle tracking (e.g. Lo Jack system).

3.1.2 Infotainment

3.1.2.1 AutoPC

By means of a wireless connection the AutoPC can access the type of information and entertainment services currently delivered by a desktop PC or cable television system.

3.1.2.2 Back Seat Entertainment

A growing list of automakers and aftermarket vendors are providing video and audio systems for back seat passengers. Most systems integrate videotape players, while some add digital gaming systems. DVD players are starting to appear on the market.

3.1.2.3 E-mail

Text-to-speech synthesis is used to read a message. Replies are sent verbally, either in the form of an attached "WAV" or audio file, or translated to text by a voice recognition system.

3.1.2.4 Internet Access

The Internet has become the fastest-growing mass medium in history. As with e-mail, it will now become accessible while driving, thanks to the AutoPC. Again, as with e-mail, it is still a matter of debate how much access to provide. Due to safety and other considerations, most onboard computing systems will likely limit access to specific sources of information or entertainment which can easily be played in audio form, rather than displayed in text or graphic form.

3.1.2.5 Onboard Navigation

This technology is extraordinarily popular in Japan, and increasingly commonplace in Europe. At its most essential, it is a digital mapping system that tracks a vehicle by matching

the data from a GPS receiver to a digitized map. Today's systems typically add turn-by-turn navigation, where directions are provided to a pre-specified destination.
Costs are rapidly falling concerning navigation hardware, and industry leaders believe it could become as commonplace as today's car CD systems once prices reach $500 to $1,000. Unto now, navigation systems have operated autonomously, with all hardware and software located inside the vehicle. However, several vendors of traditional navigation systems are studying the idea of so-called “offboard navigation” to hold down costs by "offloading" the mapping software and load limited bursts of data to the vehicle, as needed.

3.1.2.6 Out-of-Car Navigation

Experts predict many of tomorrow's drivers will use their own hardware, including Internet-ready cell phones and personal digital assistants to access telematic services. These devices will offer many of the same features as in-car systems, including wireless e-mail access, navigation and traffic advice.

3.1.2.7 Real Time Traffic Advice

There are a variety of methods to deliver real-time traffic data. Europeans have long relied on the RDS system, where verbal advice overrides the car's audio system. Several vendors are offering live tracking services on the Internet, which can be checked before leaving home or office. The most sophisticated systems under development send data bursts to a vehicle's onboard navigation system, where it can be displayed on a regional map or used to calculate alternative routes around traffic tie-ups.

3.1.2.8 Satellite Radio

The Federal Communications Commission has licensed two new broadcasters, CD Radio and XM Satellite Radio, to provide satellite-broadcast audio services aimed at American motorists. (Similar systems are under development, such as Japan's Nihon Mobile Broadcasting Service, owned by a consortium including Toyota as a senior partner.) Because the satellites are located almost directly overhead, these services will be accessible virtually everywhere in the United States. Both companies promise 100 channels of news, music and other audio channels, a large number of them commercial-free. Ford Motor Co. has signed on to offer CD Radio, which begins operating in 2002, while GM has inked a deal with XM. Basic service is expected to start at around $10 a month.

3.1.3 Technologies

3.1.3.1 Cellular Telephony

For better or worse, the cellular telephone system has become the wireless lifeline for telematics technology. Yet it poses a wide range of problems. Though costs are coming down, airtime rates are still prohibitive for such uses as streaming (or continuous) data. In the United States, Eastern Europe and most of the rest of the world, there are still wide areas where service is spotty or unavailable. Limited bandwidth is also a problem. A number of improvements are on the horizon. The GSM cell phone standard is nearly ubiquitous in Western Europe and within the coming year, many carriers will begin offering packet switching, a method used on the Internet to speed data flow. By 2003, both European and U.S. communications carriers expect to begin offering universal cell phone service (3G).

3.1.3.2 GPS

GPS technology can now be built into chips small enough for cell phones or PDAs, even as costs for a typical receiver have dipped well below $100.

3.1.3.3 PDAs

The latest Palm Pilot offers a wireless, two-way radio link. Some believe PDAs will supplant installed telematic systems, much the same way the hand-held cell phone has largely replaced the hard-wired car phone. Using Bluetooth technology, a PDA could serve as a vehicle's main "infotainment" computer, then be carried to home or office.

3.1.3.4 Reconfigurable Displays

Today's instrument panel is an assortment of mechanical, electromechanical and electronic displays. In the near future, all that gear could be supplanted by a single, digital display. Like the monitor of a computer, it could be programmed to display a wide range of sizes, shapes and colors. A driver could be given the option of customizing the display according to personal tastes and needs. An older driver might prefer larger text and numbers. And much like the "glass" cockpits found in the latest jets, reconfigurable displays could be programmed to present only must-see information during normal conditions. In the event of a malfunction, additional images would be projected.

3.1.4 Market trends

The potential market for telematic hardware, software and services could be enormous, according to industry analysts. It is foreseen that within a decade, navigation and automotive PC technology will be installed in one third of all cars.

A recent study by Frost and Sullivan [11] looked at the European segment of the telematics industry and valued the market there for systems and services at 1.03 billion Euros in 2000, growing steadily until 2004 when most volume automakers will launch affordable telematics systems across the range, accelerating growth and propelling revenues to a staggering 8.55 billion Euros in 2007.

Hardware systems hold the biggest share of the market at this stage, bringing in around 82% of revenues in 2000. However, Frost and Sullivan predicts this will drop to 42% in 2007 as services take over the dominant position.

3.2 Emergent new software and hardware platforms

3.2.1 Introduction

In-car computer/telecommunications technology market is at its starting point. There will be a fast growth in the next 5-10 years and firms are fighting to get a leadership position. Three kind of firms are joining this competition: automakers, autosuppliers and computer/communications companies. Automakers and autosuppliers are trying to ally with companies already in the Hi-Tech market to have in-car technology as soon as possible.
Customers on their hands seem to prefer car-related products and do not seem to be willing to pay the large amount of money which is needed for communications/computer devices. They are accustomed to desktop computers that crash, but computers in cars must run flawlessly, especially when they are in charge of security or engine systems and they must survive vibrations and large temperature excursions. And there is no guarantee that consumers will be willing to pay more for car gadgets. Price, size and weight, security and robustness seem now the leading problems of in-car technology. According to a study by DBH Consulting presented by University of Virginia, users of in-car technology mostly want:

- 48% navigation systems
- 44% traffic alerts
- 45% car's diagnostic
- 38% automatic call to emergency aid
- 35% connect to home
- 27% yellow pages access
- 25% internet
- 18% television.

On the other hand, carmakers are fairly sure that in few time the market will explode like the computer market did in the 80s and 90s. They declare that millions of drivers will be willing to pay (as much as $30 a month) for the Internet-access gadgets, which are expected to be standard equipment in all new cars by 2004 or 2005.

At the moment GPS and navigation system are a luxury car's standard equipment, while other type of communications devices are at an embryonic state: every company seems to have produced a prototype or a very basic computer/communication system which provides only very generic or pre-personalized information and therefore does not guarantee the flexibility of standard Internet information retrieval. In fact, no product allows a free browsing while e-mail access is usually only promised. On the other hand, every product and prototype is oriented, for security and usability reasons, towards automatic speech recognition and speech synthesis.

As a result, several electronic car gadgets have been presented in conventions (Convergence, SAE, Frankfurt Auto Show, British International Motor Show, Consumer Electronics Show) such as dash boards, navigation computers, GPS locators, parking aids, etc. but now automakers desperate for new sources of income and a competitive edge are pushing to make advanced auto technology more than a curiosity. The auto industry is trying to move faster to avoid having non-car firms taking a leadership position in this quickly growing market.

In this large range of companies, applications and perspectives there is an equally large range of possible system architectures, basic hardware platforms, and software solutions. It is, at the moment, a fluent situation: the following of this section reports on companies presently working on more emerging HW/SW solutions. However, the document is not exhaustive, as the situation changes day by day and most of the reported information was loaded from Internet, with a consequent lower degree of reliability and completeness with respect to other possible sources.

3.2.2 Car Manufacturers' situation

3.2.2.1 BMW

A Sprint PCS digital tri-mode CPT 8000 portable phone, based on the Motorola Timeport, wireless phone with integrated hands-free voice command operation is available as standard equipment on 7-Series and Z8 vehicles and will be available as a BMW center installed accessory on 3-Series, 5-Series and X5 SAV vehicles.
The system will be able to send/receive e-mail, access general information (traffic and weather) or specific ones (news and stock quotes). In the Z8 models, an Andrea Electronics' DA-310 digital microphone array is going to be included to improve the speech recognition robustness to noise.

3.2.2.2 DaimlerChrysler

DaimlerChrysler uses speech technology since autumn 1996 in its C, E and S-class. This system is called *Linguatronic*, and is for handling car phone and audio components (radio, cassette, CD-player) by voice. It is a speaker-independent, continuous speech system for command & control input and a speaker-dependent system for user-defined names (e.g. phonebook) as well. In the next S-class there will be an extended voice system integrated into the *COMAND* system with additional functionalities according to navigation.

3.2.2.3 Ford

On Ford Focus series, and in the next 5 years on every Ford car, there is a button system which provides traffic information (and will be later extended to news, stock quotes and weather) and a GPS, as well as an automatic emergency call system when the airbag is activated. This system presently uses Vodafone network in Germany and United Kingdom.

For what concerns speech recognition, Ford Motor Co. is allied with speech recognition developer Lernout & Hauspie (Ieper, Belgium).

3.2.2.4 General Motors

General Motor produces *OnStar* system. This system is a human operator based system mounted on GM cars; it consists of a three button interface and provides assistance for the driver. Now GM is starting to add automatic features to this system.

With the *OnStar* system produced by Motorola, General Motors was the first to introduce Hi-Tech into a car. OnStar uses speech technology provided by Nuance (Menlo Park, Calif.) and General Magic (Sunnyvale, Calif.). General Motors' OnStar unit also announced a deal with Sun Microsystems to try to make Java technology the computing standard for the automotive industry.

Starting with the 2001 model year, General Motors will offer an upgraded version of the *OnStar* system, including satellite radio and pre-selected news, sports, and financial information. Hands-free cellular phone service will be provided on the 30 models with OnStar, and Internet and e-mail access on Cadillac DeVille and Seville.

In 2002, Motorola's models will roll out *iRadio*, a somewhat voice-activated system that lets users download music and audio books, access voice mail and e-mail and get all sorts of over-the-Net information read to them. Like OnStar, it will be button - and voice - activated, so drivers can keep their eyes on the road.

3.2.2.5 Honda

Honda developed a system called *Internavi* with modem and cellular phone from which the user is able to browse the Internavi web site and access the navigation pages. This system does not work while moving (as safety precaution) and its use is 6 times slower than a desktop home computer. Therefore a memory card is used to travel with pre-downloaded information needed for the journey.

3.2.3 **Hardware-Software platforms**

3.2.3.1 ART

Advanced Recognition Technologies (Israel) is attempting to solve the car noise problem by combining speech and a touchpad on the vehicle console, replacing the conventional touchpad controls with a combination of speech and a touchpad.

3.2.3.2 Bosch

Robert Bosch GmbH, as a major provider of driver information and entertainment systems, is also offering VOCS, an after-sales market product, allowing speech input for their RadioPhone systems. Audio functionalities like radio, cassette and CD-player can be operated by speech commands. Also, dialing of a phone number is enabled by voice. Moreover the definition of voice or name tags is possible for radio stations, CDs and phonebook entries.

Speech input is seen as an essential input mode for future generations of driver information systems. As a further step in its product line, the company has announced speech control of navigation features for the year 2002. The independent platform VASCO (vehicle application specific computer) offers in addition to traditional audio functionalities, telephone and dynamic navigation the integration of SMS, WAP, e-mail, bluetooth, MP3 and DVD. Voice operation and input of city and street names out of large lists by spelling will be possible.

3.2.3.3 Clarity LLC

Clarity LLC (Troy, Mich), established in April 1998, offers a technology known as Clear Voice Capture, which extracts the voice signal of interest. The company says the technology (based on a bio-inspired approach) provides an improvement over standard noise suppression systems, which have difficulty with signals that have components overlapping with voice signals. Specifically, Clarity products provide an effective signal separation in the automobile.

3.2.3.4 Clarion

Clarion, allied with Microsoft, produced the first car semi-PC, called *AutoPC*, a dashboard which presently uses Windows CE on an Hitachi 66MHz computer with 60 MB of flash memory, a CD-ROM drive, and a speech recognition and synthesis system. The AutoPC offers: GPS and navigation system, download of personalized data (news, sport scores and stock quotes), a send/receive e-mail system, car audio (radio, CD), computing functions, wireless communications, hands-free voice interaction, Palm PC information exchange, address book and voice memo, Infrared data port, Universal serial bus.

3.2.3.5 Conversay

Founded in 1994, Conversay provides solutions that enable voice interaction with networked information, including the Internet, when other interfaces are difficult or impossible. Built on an innovative speech engine, Conversay™ technology is speaker-independent, modular, scalable and accommodates unlimited vocabulary, making it ideally suited for embedded applications. It also drives the award-winning line of Conversation™ products including a voice browser, a server, and developer tools for the web, desktop, server, and embedded markets.

Conversay also offers filtration techniques that separate speech signals from noise signals and narrowly focus on the speaker. The system employs two microphones, one on the passenger side and another on the driver side, and is focused more on distributed speech, for which processing power is split between the client and server.

Recently, Conversay announced it is partnering with Munich-based ComROAD AG, worldwide leading provider of telematic networks. As the partnership's initial project, Conversay's Conversation Server™ will be integrated into "Voicecom," ComROAD's portal extension to its GTTS (Global Transport Telematic System) network. The system debuted at CeBIT 2001, March 22-28 in Hannover, Germany, where users sitting in a Chrysler PT Cruiser used their voice to quickly access local weather, financial data, road conditions and sports hosted on the ComROAD portal.

3.2.3.6 Delphi and Palm

Delphi Automotive Systems, the world’s No. 1 auto supplier, and Palm joined in MobileAria agreement to develop by mid-2001 Communiport, a dashboard device to interface a Palm computer with a hands-free speech recognition system. Communiport includes navigation, mobile computing, communications and information systems, and will be able to let the computer download news, weather, finance information and send/receive e-mails. At the moment, drivers can retrieve information from their Palm V or Vx hand-helds and then make calls via certain Ericsson cell phones but it is planned to combine these devices with consoles with DVD players on back seats.
Delphi Automotive Systems recently announced that its Communiport Mobile multimedia technologies, which control various features in automobiles, generated more than $2.5 billion in orders.
Delphi Automotive Systems and Palm said they invested in a new venture that plans to offer Internet service in autos by mid-2001, bringing together the largest makers of auto parts and electronic organizers.

3.2.3.7 I/NET

I/NET (www.inetmi.com) is a software company with over ten years experience in building conversational interfaces.
Their conversational technology is based on a high-level task control architecture and event recognition system for both task control and dialogue management.
I/NET systems have been used by NASA and others for state-of-the-art control and dialogue management. Now, I/NET is working at the porting of their technology to in-car applications.

3.2.3.8 Intel

Intel is pushing towards putting its new generation of processors in cars' technology. It formed a consortium with Microsoft, QNX Software Systems (for operating system and development tools), Wind River Systems (for VxWorks OS, tools and Java technology), IBM, Fonix, and Lernout & Hauspie to build in-car computers. These computers seem to be based on Strongarm or Xscale Intel processors with a speech interface provided by L&H, Fonix and IBM, and a middleware platform by IBM. They are working under Microsoft Windows CE. Otherwise known as telematic systems, these in-car computers promise to perform any number of tasks from facilitating hands-free cell phone calls and finding urgently needed directions to providing backseat entertainment systems and in-car commerce. In-car computers based on Intel's chips could begin shipping as soon as the second half of 2002.

For what concerns speech recognition, IBM, Fonix and Lernout & Hauspie will offer a software that allows drivers to use voice commands to operate their in-car computers. Lernout & Hauspie, for example, is tweaking its noise-resistant speech recognition ASR 200 and ASR 1600 software and its text-to-speech TTS3000 application to run on StrongARM and XScale chips.

Intel is also working with consumer electronics companies such as Sony and Clarion, as well as with automakers to develop a range of in-car computers—from very basic ones for communication and navigation to full-on Internet-based entertainment and commerce systems.

IBM and Intel, too, have announced plans to collaborate on a nonproprietary standard for dashboard "telematics," a term for cellular and Internet services in vehicles such as navigation systems, roadside assistance and entertainment.

3.2.3.9 Linux

Leading automobile manufacturers in Europe are developing plans to test-drive Linux and other open source software in cars, according to industry representatives. In recent weeks, manufacturers DaimlerChrysler, in-car technology manufacturers Delphi Delco Automotive Systems, and Visteon have all revealed plans to use open source software in products.
This comes at a crucial point for the automobile industry, which is digesting the first industry-wide specification for intelligent devices in cars. It is expected to help manufacturers build onboard systems capable of linking mobile phones, navigation and in-car entertainment. It will also drive down the costs of such technology for customers. Visteon has unveiled the Mach MP3 Jukebox, a Linux-fuelled car audio system.
In the opinion of many manufacturers Microsoft's Windows operating systems technology is not stable enough to be implemented in cars.

3.2.3.10 Lucent

Lucent has introduced a hands-free speakerphone powered by voice technology from both Philips Speech Processing and Lernout & Hauspie, and has announced plans to introduce its own speech recognition engine later this year. Several phone makers, among which Lucent, announced plans to incorporate the Bluetooth communications standard in the next generation of devices. When coupled with a Bluetooth-enabled vehicle, the phones would be able to use microphones built into the car to enable hands-free operation.

3.2.3.11 Motorola

The *OnStar* system is a three-button dash system mounted on Cadillacs able to operate a cellular phone, or to connect with an human operator (who receives also information from the car's GPS) in order to ask for information or require assistance. By mid-2001 it will turn into a computer with Windows CE, Sun software and a single word recognition but there will always be operators to be contacted. Moreover the system mounted on Cadillac De Ville (or Seville) will be able to receive e-mails, to retrieve traffic data and personalized information (sport and financial news).

By end-2001 Motorola will develop *iRadio*, which it plans to sell to automakers. *iRadio* allows drivers to listen to Internet radio stations, traffic reports, news and stock information, and access e-mail by using speech recognition (a system with a single word activation) and text-to-speech technology. By mid-2002 it will be able to downlad MP3, audio books, send/receive e-mails, and have GPS navigation information.

Also, Motorola recently signed a cooperative deal with *Phone-Or*, an Israel-based company that makes noise-suppression systems, which will be available with *OnStar* for the 2001 model year. Motorola also inked a pact with IBM to develop new uses for telematics and work on the reliability of car-to-wireless-network connections.

Motorola is then looking beyond *OnStar* and first-generation *iRadio* with new partnerships and equity positions. Lear Corp., a new partner, provides interior dashboard designs and parts to Ford. Command Audio, in which Motorola took an equity position, wirelessly transmits audio programs from print and broadcast.

3.2.3.12 NSC

Natural Speech Communication is a world technology leader in providing cost-competitive, high-density, and noise-robust speech recognition solutions for cellphony, telephony and Internet infrastructure vendors. NSC provides extremely compact DSP-based firmware and hardware products, with up to hundreds of speech recognition channels in a single telephony server, at a highly competitive cost-per-channel. NSC's award-winning products have shown outstanding performance in severe noise conditions for various telephony networks.
NSC's unique SR engine is designed for handling speech in severe noise environments, such as car-phones or speakerphones in the office.

3.2.3.13 Philips Semiconductors

Philips (http://www.semiconductors.philips.com) next generation IC for telematics/navigation applications is a complete telematics processor, developed using the Nexperia CIP concept. By drawing together all the necessary hardware blocks onto a single piece of silicon and dedicated software from Philips Semiconductors, this truly integrated telematics solution offers potential safety and security features to the occupants or owner of a vehicle using a

combination of location information (GPS and Dead Reckoning) and communications via a cellular phone link.

Philips Semiconductors' Nexperia Car Infotainment platforms provide a highly flexible design methodology for fast-moving markets where getting there first is essential. As product complexity increases and consumers demand ever-more features and functionality, Nexperia platforms offer a consistent architecture for IP re-use and sophisticated tools for rapid, versatile and cost-effective design. Central to Nexperia platforms are powerful programmable embedded processor cores: including 32-bit MIPS RISC CPUs and Philips-developed advanced 32-bit VLIW dedicated media processor Trimedia high-level programmability enables functionality to be added or modified late into the design process and guarantees future-proofing, allowing enhancements to be incorporated in the field. Product differentiation is easy, and without the need for hard-wired components, development risk and costs, as well as design time, are kept low.
Due to their multi-processor environment, Nexperia platforms ensure extensive scope for enhancing processing power, and their scalability gives the freedom to combine IP blocks in innovative ways for high, medium and low-end applications.

Software forms an integral part of Nexperia platforms, with a consistent set of APIs (Application Programming Interfaces) providing functionality abstracted from the hardware to facilitate re-use and eliminate the need for low-level programming.
Supporting multiple operating systems, including Java, Windows CE and ISI's pSOS, the Nexperia software architecture incorporates a rich library of streaming software for media components (for video, audio, graphics and speech), and numerous OS-independent device drivers. A range of software development tools including a dedicated TriMedia Software Development Environment, further speeds Nexperia design.

Beside Trimedia, Philips processor cores for audio and video processing includes R.E.A.L. - a dedicated low-power, real-time re-configurable embedded DSP core for audio and speech recognition. Front-ends include GPS and radio, and interfaces to various car bus systems can be supported such as CAN. Peripheral I/O blocks are already available to cover speech recognition and audio/video functions for example, and others will be added. Software includes device libraries, functional libraries (speech recognition, graphics, etc) and support for an operating system.

3.2.3.14 Sensory Inc.

Sensory Inc. is a leader in speech recognition for the consumer and embedded markets, including large and small speaker-independent recognition, speaker-dependent recognition, noise and echo cancellation.
Sensory recently acquired Fluent Speech Technologies, a group spun out of the Oregon Graduate Institute with a large vocabulary speech recognition engine. It is now moving this technology into cars for natural language interaction.

3.2.3.15 SGS Thomson Microelectronics (ST)

ST has developed a specialized 24-bit DSP-based chip called Euterpe that can perform the functions of voice recognition, text-to-speech rendition, noise and echo cancellation, and biometric verification on audio data streams. At present, the ST system is declared to be well-suited for command & control applications.

3.2.3.16 Siemens

Siemens has recently presented WIRE (Web-based Interactive Radio Environment). This system is able to browse web pages using a speech recognizer and synthesis. The system provides an option to save graphically rich pages in order to let the driver look at them later.

WIRE was conceived to act as a car radio. While browsing the WWW, you bookmark a particular website just like selecting a radio channel. Once selected, you listen to the audio rendering of the website page, as if listening to a certain song on the radio. You then have the option to use voice commands for additional features, such as following a link to another Home Page, or "turning the channel" to visit an unrelated website.

Separately, an Israeli joint venture between Altec-Lansing and STMicroelectronics was created to develop speech DSP technology, in particular for speech recognition applications.

3.2.3.17 Sprint PCS

Sprint PCS operates the largest 100 percent digital, 100 percent PCS nationwide wireless network in the United States, already serving the majority of the nation's metropolitan areas including more than 4,000 cities and communities across the country. Sprint PCS has licensed PCS coverage of nearly 270 million people in all 50 states, Puerto Rico and the U.S. Virgin Islands (for more information, visit the Sprint PCS Web site at “http://www.sprintpcs.com”) .

3.2.3.18 Visteon

Visteon Corporation implemented a voice-activated system in specially equipped Jaguar S-Types in 1999, allowing drivers to control their audio systems, climate control and phone by voice. In January, the company announced a contract to produce the devices for the 2002 Infiniti Q45 vehicle.

Visteon Corporation purchased C-REC™ and SDX speech recognition technology from L&H Holdings USA, Inc., a wholly owned L&H subsidiary. C-REC™ is a continuous speaker independent, phonetic-based speech recognition engine that enables a wide range of applications to the automotive telematics/multi-media market and has direct applicability to non-automotive markets as well. SDX is a layer of additional software that contains a programming interface and application software that enables efficient integration of C-REC™ into commercial products.

Visteon will form a subsidiary to do research/development and application on speech interfaces in automotive and non-automotive applications. Concentrating on using voice technology, the new company will develop leading-edge speech solutions that will enable faster-to-market delivery of new and innovative products in the areas of navigation, communication, entertainment, in-vehicle computing and audio systems.

3.2.3.19 Wavemakers

Similarly to Clarity LLC, Wavemakers, a Canadian supplier of speech enhancement systems, develops products that provide signal separation in the automobile.

Recently, it announced its commitment to the Aurora Project. As part of this commitment, the company tested its ClearStream noise extraction system using Aurora's standardized sound files and Microsoft's Entropic speech recognition engine. ClearStream increased speech recognition accuracy on city streets, in cars and inside train stations by as much as 35% in low

signal-to-noise cases (SNRs of 0 dB to 5 dB), representing an error recovery rate of 90%. These results are available at “www.wavemakers.com/eval.html” for review and download.

3.2.4 Special devices

3.2.4.1 AcousticMagic

Acoustic Magic is a privately held company building a family of superior "far talking" microphones for speech recognition, teleconferencing and automotive applications.
Their patented array microphone technology improves sound fidelity and signal-to-noise performance, enabling speech recognition for hands-free command and control of computers and PDAs in noisy environments as automobiles (see www.acousticmagic.com).

3.2.4.2 Andrea Electronics

Andrea Electronics delivers software and hardware audio input solutions for hands-free cellular communications and speech-enabled telematics applications. Patented and proprietary technologies include adaptive beam-forming, beam-steering, blind source separation, noise cancellation and echo cancellation algorithms. Andrea Electronics’ digital microphone solutions offer a flexible array structure which utilizes between 2 and 8 microphones and can be customized to suit a particular application or hardware configuration, such as a rear view mirror, overhead console, headliner or steering column.

The Andrea DA-310 digital microphone array, its second generation far-field microphone array, is optimized specifically for the automotive environment. Andrea Electronics’ DA-310, incorporating patented Digital Super Directional Array (DSDA®) 2.0 and patent-pending PureAudio™ 2.0 technologies, utilizes an adaptive beamforming technique and an effective de-reverberation process to significantly reduce ambient automotive noises such as tire, engine and wind noise, as well as other passenger voices. By eliminating extraneous noise, the DA-310 enables speech-activated functions to perform optimally, while the speaker is at a distance from the microphone, ensuring safe and accurate mobile communications.

3.2.4.3 PhoneOr

Phone-Or’s BNS technology provides clear and noise-free sound input, resulting in the highest speech recognition performance, unattainable with other standard technologies in the market today. As opposed to DSP technology, Phone-Or’s technology, is based on a unique and smart algorithm of real-time optical-acoustic signal processing, which controls and adjusts the optical microphone’s Figure-of-8 polar pattern to changes in the speaker’s environment resulting in real-time cancellation of background, random, cyclic and harmonic noise, as well as providing excellent directionality and echo reduction.

3.3 High-bandwidth wireless communication facilities

The term "wireless" is referred to telecommunication in which electromagnetic waves (rather than a wire) carry the signal over part or the entire communication path.
Radiotelegraphy, with Morse code (in the beginning of 20th century), was the first wireless transmission that went on the air. Later, with the use of modulation, it was possible to transmit voice and music via wireless and the name of transmission changed into "radio". Nowadays the spectrum is used in a broad portion for data communication, television, fax and the term "wireless" has been used again.

Wireless can be divided in four categories [12]:
- Fixed wireless: wireless systems or devices installed in homes and offices
- Mobile wireless: equipment used on moving vehicles, like an automotive cell phone
- Portable wireless: autonomous equipment to be used outside homes and vehicles, like a mobile cell phone
- IR wireless: devices that use an infrared interface to communicate (in a limited range) with other devices, e.g. laptops, mobile cell phones and PDAs

Wireless technology is evolving and its role is becoming primary in the lives of people. Portable wireless in particular is improving very fast and this has led to a wide utilization of mobile phones in the whole world. In few years three generations of mobile phones [13] hit the market, yet still a lot of improvements are possible. Third generation (3G) will lead to a high bandwidth phones able to connect to the Internet, listen to music, and watch videos.
The trend is to increase the number and quality of the services with the increasing demands:
- high capacity
- high data rate

and the constraints:
- portability
  - low power consumption
  - small form factor
- fast time-to-market

A limitation of portable wireless communication is the limited bandwidth. Actually only few kilobits per second can be transmitted with a mobile phone and this is limiting dramatically the applications. The term "high bandwidth" refers to a data communication whose speed is at least 2Mbps and this should be reached when the 3rd generation will be available.

3.3.1 The three generations of mobile phones[1]

3.3.1.1 First Generation(1G): analog

The first generation was born in the early 80s, for voice transfer. AMPS, TACS, etc are included among first generation systems. This technology is based on a modulation of radio phone signals, varying continuously their frequency.
With the recent proliferation of post-analog technology, only a very few analog systems remain in existence.

[1] http://www.nokia.com/3g

3.3.1.2 Second Generation (2G): digital

This generation was born in the early 90s and it is widely used. The major part of mobile phones at the moment uses this digital technology that converts the sounds in streams of bits. The bits are then used to module the wireless signals. Digital networks are perfectly suitable for voice/data/fax transfer and other services. At present, second generation systems are still evolving with ever-increasing data rates via new technologies such as HSCSD and GPRS. Second generation systems include GSM, US-TDMA and PDC.

3.3.1.3 Third Generation (3G): high bandwidth digital

This generation is expected in the next few years. Wireless companies are promising new networks able to handle high bandwidth communications, with data transmission much faster than the actual 10Kbps. 3G is not a standard but a set of different approaches to reach a download speed close to 2.4Mbps. Applying high-speed data transfer and state-of-the-art radio terminal technology, third generations systems will enable multimedia services among traditional services.

3.3.2 Transmission technologies[2]

3.3.2.1 AMPS (Advanced Mobile Phone Service)

Analog cellular communications system developed and used in the US. It is the most used kind of analog nets for mobile phones. AMPS works on a frequency of 800MHz and uses the FDMA technology. This system is more suitable for transmission of voice than data, so is has been replaced by digital networks.

3.3.2.2 TACS (Total Access Communications System)

An analog cellular communications system derived from AMPS. It has been adopted in the UK (ETACS) and operates in the 900MHz band.

3.3.2.3 FDMA (Frequency Division Multiple Access)

Each user is given a different frequency channel that is used for the whole length of the communication. I.e. an AMPS network has 832 channels, spaced by 30kHz. In digital networks FDMA is used in combination with CDMA or TDMA.

3.3.2.4 CDMA (Code Division Multiple Access)

Digital technique that allows multiple users to share the same frequency channel. Each signal is divided in data chips; every chip is labeled with a user code and sent on a different band frequency. On the receiving side, the chips are assembled to recreate the initial signal.

3.3.2.5 TDMA (Time Division Multiple Access)

Digital technique that allows multiple users to share the same frequency channel. Each user is given a time interval (time slot) repeated in the same channel. Since the data of every user are always in the same time slot, the receiver can split the signals and recreate them.

3.3.2.6 US-TDMA (US Time Division Multiple Access)

A second-generation system used in the US. Also referred to as D-AMPS (Digital AMPS). First digital system adopted in the US and covers the entire country.

[2] http://whatis.techtarget.com

3.3.2.7 GSM (Global System for Mobile communications)

It is an European standard for digital networks than ensures the compatibility among devices. It uses a variation of TDMA technology and works on 900/1800MHz frequencies. Data transmission is limited to 9600bps. GSM digitizes and compresses data, then sends it down a channel with two other streams of user data, each in its own time slot.

3.3.2.8 PCS (Personal Communication Services)

It is a digital standard used in North America, on a frequency of 1900MHz. It is a wireless phone service somewhat similar to cellular phone service but emphasizing personal service and extended mobility. It is sometimes referred to as digital cellular. PCS is for mobile users and requires a number of antennas to blanket an area of coverage.

3.3.2.9 HSCSD (High-Speed Circuit-Switched Data)

It is a circuit-switched wireless data transmission for mobile users at data rates up to 38.4Kbps, four times faster than the standard data rates of GSM. HSCSD is comparable to the speed of many computer modems that communicate with today's fixed telephone networks.

3.3.2.10 EDGE (Enhanced Data rates for Global Evolution)

It is an improvement of the GSM and TDMA based networks. EDGE should lead to a data transmission of 473Kbps, enabling value-added Mobile Multimedia services. It is a target of GMS networks in Europe and AT&T wireless in USA.

3.3.2.11 GPRS (General Packet Radio Service)

It is a planned improvement for GSM networks that uses packet commutation for data communication. Instead of using data on dedicated circuits, a packet commutation network subdivides the information in packets and sends them on one of the available channels of the network. GPRS promises data rates from 56Kbps up to 114Kbps and continuous connection to the Internet for mobile phone and computer users.

3.3.2.12 UMTS (Universal Mobile Telecommunication System)

It is the standard for 3G wireless communications made by ETSI (European Telecommunications Standard Institute) within the ITU's (International Telecommunications Union) IMT-2000 framework. UMTS is based on W-CDMA

technology for voice signals and TD/CDMA for data transmission and the speed should be around 2Mbps. It is the planned standard for mobile users around the world by 2002. Once UMTS is fully implemented, computer and phone users can be constantly attached to the Internet as they travel. UMTS will use a packet-switched connection, using the Internet Protocol (IP). This means that a virtual connection is always available to any other end point in the network.

3.3.2.13 W-CDMA (Wideband Code Division Multiple Access)

It is the improvement of CDMA technology, it distributes data chips on a wider frequency band than CDMA. It can support mobile/portable voice, images, data, and video communications at up to 2Mbps (local area access) or 384Kbps (wide area access). The input signals are digitized and transmitted in coded, spread-spectrum mode over a broad range of frequencies.

3.3.3 New wireless transmission technologies[3]

3.3.3.1 UWB (Ultra Wide Band)

This technology was invented in the 60s and it is not yet used for mobile telephony. The working principle is the absence of the carriage signal, present in all the other transmission technologies. UWB signals are "pure signals", they are a sequence of short impulses. A single impulse lasts $1*10^{-9}$ seconds and there can be more than 40 millions impulses in a second. Similar to a ultra fast Morse code, the impulses follow a precise coding scheme. The noise created by UWB should be ignored by other electronic equipment, if at low power. It is still early to think of wireless communications using UWB because the transmission range is too small, but there could be some improvements in the next years.

3.3.3.2 W-OFDM (Wideband Orthogonal Frequency Division Multiplexing)

Multiplexing is used to split the signal in several small signals at lower speed. This has advantages because fast signals are more disturbed by noises and interferences, while slow signals are more suitable to be filtered from noises. I.e. a channel of 10MHz in a 900MHz band can be split in 10 different channels, each one transporting 1Mbps data. This technology allows to use the spectrum in a more efficient way, but the problem is that there are not DSPs so fast at the moment and nowadays networks are not yet suitable.

3.3.3.3 Bluetooth

It has been created by a consortium of mobile phones producers and it is a wireless connection among electronic devices. Bluetooth was invented to connect the earphones to the mobile phone, but now more than 1000 societies want to construct devices that use this standard, from toys to microwave ovens. This technology is a point to point radio connection at a band of 2.4-2.5GHz and can transmit data in the distance range of 10 meters. To avoid interferences with other devices, the frequency is switched 1600 times in a second. Data transmission speed should be around 1Mbps. Bluetooth will improve the connections among devices, but its utilization for mobile phones data transmission seems not so easy.

3.3.3.4 Satellite transmission

An interesting solution for wireless communication comes from the use of LEO (Low Earth Orbit) satellites. GEO (Geostationary Earth Orbit) satellite communications are afflicted by a significant delay in the retransmission of data. This delay is about one half second and is due to the distance between the satellite and the Earth receivers. LEO satellites spin at a distance 25 times closer than GEO satellites (only 1400 Km from the Earth surface) and this reduces sensibly the delay. Being non geostationary, there is the need of several satellites to cover the whole terrestrial surface, and at the moment big zones are not yet covered by this service. One of the major problems of this solution is signal reception. Satellite mobile phones can receive the signal only on open-space places and no signal is received inside buildings. A mixed solution is to use GSM network when the signal is not available by satellite coverage. Existing companies assure anyway a limited speed in data transmission, less than 10Kbps, but the future seems promising. In less than five years there should be companies with more than 250 LEO satellites, promising a global broadband "Internet in the sky" network. There should be available a variety of telecommunication services, interactive multimedia and voice transmission. Million of users should have a fast two-way connection (2000 times a standard modem speed). The costs should be limited and similar to a standard cost of a broadband service. It is not clear anyway if new satellite communications will be available for mobile

[3] Scientific American, "The Wireless Web", October 2000.

phones. The power needed to transmit data to a satellite is still high and not suitable for a hand-held device.

3.3.3.5 Aerial arrays

A solution to reuse existing structures is the utilization of a matrix of aerials. This solution should guarantee 1Mbps for each user and every cell should be used by a maximum of 40 users. The technology uses in a better way the aerial matrixes present in the mobile phones base stations. At the moment aerials are used in a omnidirectional way, but in other telecommunication fields single aerials are used in conjunction to send signals only in certain directions. Each aerial signal is combined with the others; in this way it is possible to amplify it in some zones and inhibit it in some others. This technology is based on DSPs developed by the American Army to intercept foreign radio transmissions. When connected to an aerial array, these DSPs can send signals to a single user with a good precision and can follow him. The system can reuse the same frequencies for multiple users and this leads to a very efficient use of the available spectrum. Aerial arrays are already installed and many base stations are equipped with sufficient powerful DSPs.
One of the big problems anyway is the relatively slowness of the system: although it can follow a walking person, it cannot follow a running vehicle.

3.3.4 A real case: the WAP (Wireless Application Protocol)[4]

WAP is a standard for wireless Internet connection with mobile phones. It is a set of specifications to transmit Web documents to phones and other portable devices. At the moment WAP is present only in Europe but will be introduced soon also in the US. Mobile phones have small displays and their networks are unable to use HTML. This results in the incapacity for WAP to transmit images and sounds, also due to the limited data transmission speed. To avoid this problem it was created HDML (Hand-held Device Markup Language), a language suitable for wireless networks. From HDML, with the help of phone companies, it was developed WML (Wireless Markup Language), that became the core of WAP specifications. The display of a WAP phone acts like a mini browser. The user digits a Web address and the request is transmitted to a WAP gateway. The gateway, connected to Internet, finds the requested page and converts it from HTML to WML. WML code is then sent to the user's phone that displays the text. The conversion from HTML to WML is not painless: every image is stripped away and text formatting is removed. The result can be unreadable or not very usable. This fact led some companies to create new pages written in WML or to optimize old HTML pages to be easily converted in WML. There are more than 5000 WML pages available on the Web and this number is rapidly increasing. With WAP, mobile phone users can know sport results, book flights, find cinema timetables, read books and news. WAP could become obsolete with the progress of technology: with a wider band, WAP will be unused soon. In the meanwhile, the come of GPRS with a data transmission speed 10 times faster than GSM, should increase the use of WAP, limiting costs and time of connections. This uncertain future is limiting the development of WAP pages because no one wants to spend effort creating pages that could be based on a poor standard. WAP has another problem: security. The standard at the moment includes dispositions called WTLS (Wireless Transport Layer Security) specifying data coding during the transmission from the phone to the network. WTLS are requiring less power and memory than SSL (Secure Socket Layer) technology, which is used for credit card numbers on the Internet. The problem lies in the WAP gateway, which decodes WTLS data and converts it into SSL data. For a short period of time, user data is not protected and could be intercepted by hackers. This and other little problems are limiting the use of WAP and the companies are controlling too much what is

[4] Scientific American, "The Wireless Web", October 2000.

happening. WAP should be put free like Internet to go under a massive develop, so at the moment is still to early to foresee a rosy future.

4 User Expectations and Requirements

4.1 User Expectations

The need of navigation systems in a quite populated area as Europe is obvious. Especially within the travelling domain, provision of in-car route guidance systems is desirable. As more and more functionalities for the automotive environment are offered by an ever-growing number of devices and services, there is also the need to integrate all these functionalities into one single product.

According to user inquiries that have been conducted by companies from the automotive industry as well as information from the ADAC (Allgemeiner Deutscher Automobil Club) the most important features in customer priority that should be covered by a driver information system are:

- Dynamic navigation guidance (high importance)
- Traffic information
- Hotel reservation
- SMS
- Parking assistance
- Messages
- Emails
- Mobile office features
- Restaurant reservation
- Event notes
- Appointment calendar
- Info- and entertainment (low importance)

These studies partly coincide with the functionalities that emerged from Wizard-of-Oz experiments within the VICO project. Here the following order of desired features controllable by speech for in-car driver information systems was observed:

- Traffic Information
- Navigation
- Telephone
- Hotel Reservation/Tourist Information
- Radio
- Cassette and CD-player
- News Reading
- Car Manual

Currently commercially available driver information and assistance systems have their focus on audio functionality, traffic applications, and the usage of a phone. For current users the most important features are traffic- and travel-related, as these are functionalities they are familiar with. Features of such a system that provide the user with information in general and serve as entertainment functionality are mostly considered as being much less important

where one of the reasons certainly is that the advantage of some of the applications can not be imagined when not having been experienced by the user.

Concerning the human-machine interface offered by those systems the following features are favoured by potential users:

- Appropriate system design
- No distraction from primary driving task
- speech input is preferred over traditional tactile interfaces
- usage of natural everyday language is preferred over command language
- Simple, easy-to-understand and safe operation of all system functions
- If not self-explaining, simple and easy-to-follow guidance through the system's menu (in case of speech input through the dialogue respectively)
- If possible, usage of a keyword to start the recognition process is preferred over the usage of a push-to-talk (PTT) button
- Useful combination of visual and acoustic feedback (e.g. concerning map display, etc.)
- Sufficient speech recognition accuracy
- Robustness against environmental noise (car noise as well as crosstalk)
- Intelligible and natural good-quality speech output

Looking at the overall system design of driver information systems, guidance throughout the conversation by means of intelligent and understandable dialogue strategies is desired. Users attach high importance to a flexible and, above all, consistent dialogue structure. Appropriate dialogue design is expected to assist a user in fulfilling his/her task. Also, visual feedback is felt as being important to give a fast feedback to the user.

Especially the quality requirements asked by potential users of in-car voice-operated systems are hard to meet. This includes very subjective opinions on what exactly constitutes high-quality speech output. Whereas for some people naturalness is the most important characteristic, others are willing to put this second if the intelligibility is ensured. The same applies to the prerequisites a "good" speech recognition engine should meet. Clearly the optimal solution would be the adequate system reaction to any kind of speech input, ranging from spontaneous sentences to command words, and covering clean speech as well as all kinds of environmental and human noise. When interrogated, people do fortunately not expect a speech system to have 100% recognition performance, but claim to tolerate 5-10% errors [15]. However, most people do not consider that 10% error rate means that every tenth word is misrecognized, sometimes severely obstructing successful speech recognition and the adequate system reactions to user responses.

Concerning erroneously recognized speech, misrecognitions are usually judged by users much worse when a wrong sentence is hypothesized by the system. In contrast rejection of the input and the request of the system to repeat the user input is felt to be rather acceptable. In general users are very sensible to undesired system reactions preferring cooperative systems with comprehensible system responses.

4.2 User Requirements

In general terms, a usable spoken language dialogue system (SLDS) must satisfy user needs which are similar to those which must be satisfied by other interactive systems. Thus, what users need are SLDSs with which they are generally satisfied in the overall context of use and

which they feel are easy to understand and interact with. Interaction should be smooth rather than bumpy and error-prone, and the user should feel in control throughout the dialogue with the system. SLDSs differ in many ways from other classes of interactive systems which is why the above, general guidelines for interactive systems specification and design must be specialised to guide the development of usable SLDSs [18]. Furthermore, the VICO system has a number of properties which make VICO different from many other SLDSs. These properties must be taken into account as well when specifying user requirements to the VICO system. This has been done below.

4.2.1 Interactive system setup

Input: push-to-speech button, speaker independent spontaneous speech.
Output: synthetic speech (TTS), static graphic text on screen (1st prototype), static graphics text and possibly maps and more on screen (2nd prototype).
Setting: car environment, traffic environment, safety-critical, user mostly occupied by safe driving, hands and eyes mostly occupied by activities which are more important than activities to do with manipulating the system and reading the text on the screen.
It follows that most of the communication with the users should be made through speech which leaves hands and eyes free to safely steer the car. The VICO screen text should be static, so that it can be read whenever it is safe to do so. The screen text should be parsimonious and, as a general guideline, duplicated by the speech output from the system. In this way, the user primarily has to resort to the screen when the user has missed the system's spoken output, for instance because of some attention-grabbing event in the traffic, or when the user in a stationary car wants to get an overview of VICO's functionalities and information about these. The push-to-talk button should be placed within easy reach of the user, for instance on the steering wheel. The screen should be similarly placed so that the user can read it with as little effort as possible. This requirement also affects the font sizes and contrast used on the screen.

4.2.2 The users

4.2.2.1 Age and capabilities

The users of VICO will be anybody who has or uses a car with the VICO system installed. This means that the VICO users will be aged from 18 (or in some countries 16) to +80 years and may have any background in terms of education and computer literacy. The users will also differ widely in their mental and intellectual capabilities.

4.2.2.2 Frequency of use

VICO is not strictly speaking a walk-up-and-use system because most users of VICO are expected to use the system on a regular basis. However, to become regular users of VICO, the system should be so easy to learn to use that virtually all intended users can do this with minimal effort. Also, single-time users should have a good chance of quickly familiarizing themselves with the system. On the other hand, once a user has learnt to use VICO as a matter or routine, this user is likely to want a system which offers the necessary shortcuts, by-passing system information which may be useful to first-time users but which is a nuisance to experienced users.

4.2.2.3 User experience

VICO's users cannot be assumed to have prior experience with (a) in-car non-speech input navigation systems or (b) spoken language systems. The reason is that these technologies are so recent on the market that many users will not have had the occasion to become familiar

with them. Still, VICO's users may be expected to have some knowledge of the tasks and domains supported by VICO. However, not all users will be task and domain experts. This must be taken into account when designing VICO's user-system dialogue, for instance through explaining terms which not all users can be expected to understand. For instance, the 18-years old might not be familiar with the star system for hotels (but might be able to afford a hotel).
It follows that VICO must be very easy to understand and use, and that VICO must support different interaction strategies depending on whether the user is a novice to the system or has acquired sufficient experience to use it as a matter of routine. Even then, it may be advisable, or even necessary, to present on request the basics of how to operate VICO on the screen and via speech output. It is doubtful if VICO can be designed in a way which does not require how-to-operate information to be presented to at least some users. How-to-operate information includes not only the standard what-to-do-when but may also have to include information such as that it does not help to speak very slowly to become recognized by VICO. Speaking at a non-standard rate may only make VICO's speech recognition rate worse. Similarly, users may need to be informed of VICO's barge-in capability.

4.2.2.4 Cooperativity

To support ease-of-use and minimize the need for meta-communication, VICO's spoken and written output should be maximally cooperative and conform to the cooperativity principles described in e.g. [19] and supported by the CODIAL tool [20].

4.2.2.5 Naturalness

To support naturalness of interaction, VICO should be capable of mixed-initiative spoken dialogue in which both VICO and the user may take the initiative at any time. Similarly, VICO's conception of the order in which information has to be exchanged with the user to get a particular task done, should conform to the user's intuitive conception of information exchange order, if any.

4.2.2.6 Other user differences

To support different user backgrounds, users changing their minds during dialogue, etc., VICO should preferably have barge-in.
Since VICO is being designed to work with users from different European cultures, relevant cultural differences, if any, should be identified and taken into account when designing VICO's output language as regards, e.g., politeness and other factors of speaking style.

4.2.2.7 Feedback

To support users' need to feel in control during interaction, VICO should provide adequate spoken and visual information feedback on the user's input. This should be kept in mind when designing information feedback in domain communication situations as well as when designing information feedback in meta-communication situations. Similarly, ways should be found to provide suitably informative process feedback when VICO is spending some time processing the user's input. Having barge-in, VICO must be protected from impatient users when VICO has not been speaking for some time. So, impatient users must be given the impression via VICO-produced audio, that VICO is processing their input.

4.2.2.8 Reasoning

Implicitly, users expect VICO to be able to perform bits and pieces of reasoning which humans normally perform without thinking. VICO should be able to meet these user expectations. For instance, if a user asks VICO to support navigation to a McDonald's in Hamburg, VICO should infer that the relevant McDonald's is the nearest one in terms of, probably, distance from where the query was made. An alternative solution is to provide the

nearest McDonald's in the direction in which the car is going. In any case, since all McDonald's are more or less identical in what they offer, there is no reason to start a dialogue about which McDonald's the user wants in a city which may be totally unknown to that user.

4.2.2.9 When should VICO be active?

Given the safety-critical nature of car driving, VICO should work as non-intrusively as possible. For this reason, VICO should only be switched on when the user wants its advice, VICO should stop listening when the user has finished inputting the information needed for a particular task, and VICO should stop outputting information whenever there is indication from some in-car device that the driver's attention is fully needed elsewhere.

4.2.2.10 Special user groups

Some users who are hard-of-hearing may have difficulty using VICO. It is probably no (safe) solution to provide these users with an alternative, complete screen-based dialogue (with or without spoken input).

4.2.3 User model

4.2.3.1 User model information

To facilitate adaptation to individual users, VICO will create and maintain user models of the different users of the system in a particular car. As soon as VICO knows who is driving the car on a particular occasion, VICO will consult the model of this user, if any, and use the information in the model to facilitate the user's tasks. During its use, VICO will maintain and revise its user model in order for the model to conform as well as possible to the needs of that particular user. It is too early to provide a complete list of the parameters which VICO will use in categorizing and supporting its users. However, driving habits concerning the areas which a particular user normally visits will certainly be among those parameters. The user's degree of familiarity with VICO is another obvious parameter. A third parameter could be to keep track of the kinds of difficulties a particular user has using VICO, such as pronunciation difficulties. VICO could use this information to better support those users.

4.2.3.2 Knowing who is driving

An important question concerns how VICO will know who is driving the car on a particular occasion. This knowledge is required for VICO to correctly access the model of that user. There are several possible solutions, such as speaking one's name, using a password, or simply being identified from one's voice. User identification will not be implemented in the first prototype.

4.2.4 Domain and task

4.2.4.1 Explaining what VICO knows

Given the broad range of tasks to be supported by VICO, users will have to be informed of VICO's conceptions of those tasks including any unexpected limitations to those conceptions. Lack of understanding of what VICO can and cannot do it likely to cause user dissatisfaction. One solution is to use the screen as an additional facility which can inform the users what VICO can and cannot do for them. Users can browse this information when the car is parked or in other safe circumstances. By the same argument, VICO might come with a reference card which explains what is can do and how to operate it. Users will expect that the tasks supported by VICO are either (a) intuitively clear or (b) clearly described including their possibly counter-intuitive limitations. For instance, if all VICO knows about hotels is whether

a particular hotel is a, say, three-star or four-star hotel, this limitation should be clearly stated. Otherwise, users may expect that VICO can answer all manner of questions about hotels. When they find out that VICO cannot do what they expected, they are likely to be disappointed with the system.

4.2.4.2 Naturalness

To support ease-of-use, it is important that VICO is able to recognize and understand the widest possible range of expressions which users might find it natural to use when asking for information from VICO.

4.2.5 Miscommunication

Obviously, VICO's speech recognition and understanding capabilities should be as high-quality as possible. Like humans and other spoken language dialogue systems, however, VICO will sometimes have difficulty recognizing or understanding what the user says. Similarly, users may sometimes have difficulty understanding what VICO said or meant. This means that VICO needs state-of-the-art error-handling (meta-communication) mechanisms and strategies: to handle its own recognition errors and clarification needs as well as to handle the user's problems of hearing what VICO just said, for instance because the user's attention was focused on the traffic, and understanding what VICO meant. These strategies are likely to include user spelling of difficult-to-distinguish destination names. VICO's user model might provide help in interacting with users who tend to have miscommunication with the system.

4.2.6 Evaluation

For the purpose of optimizing VICO's user-friendliness, it is important to evaluate the extent to which VICO satisfies the above user requirements throughout the development life-cycle.

5 Bibliography

[1] M. Walker, L. Hirschman, and J. Aberdeen. Evaluation for DARPA Communicator Spoken Dialogue Systems. In Proceedings of *Second International Conference on Language Resources and Evaluation*, Athens, Greece, May 2000.

[2] J. Polifroni and S. Seneff. Galaxy-II as an architecture for Spoken Dialogue Evaluation. In Proceedings of *Second International Conference on Language Resources and Evaluation*, Athens, Greece, May 2000.

[3] http://cslr.colorado.edu/

[4] http://cumove.colorado.edu/

[5] J.H.L. Hansen, J. Plucienkowski, S. Gallant, B.L. Pellom, and W. Ward. CU-Move: Robust Speech Processing for In-Vehicle Speech Systems. In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP'2000),* vol. I, pp. 524-527, Beijing, China, Oct. 2000.

[6] http://www.hrl.com/TECHLABS/isl/index.html

[7] R. Belvin, R. Burns, and C. Hein. Spoken Language Navigation Systems For Drivers. In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP'2000),* vol. II, pp. 591-594, Beijing, China, Oct. 2000.

[8] http://www.disc2.dk/

[9] F. Béchet, E. den Os, L. Boves, J. Sienel. Introduction to the IST-HLT Project Speech-Driven Multimodal Auomatic Directory Assistance (SMADA). In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP'2000),* Beijing, China, Oct. 2000.

[10] http://www.hltcentral.org/usr_docs/project-source/SPOTLIGHT/AR-2000/annual_report_2000.htm

[11] Frost & Sullivan. European Automotive Telematics for Hardware and Services-, Analysis Report, 2001.

[12] http://whatis.techtarget.com

[13] http://www.nokia.com/3g/

[14] Speech Technology Magazine, http://www.speechtechmag.com

[15] P. Geutner, L. Arévalo, and J. Breuninger. VODIS – Voice-Operated Driver Information Systems: A Usability Study on Advanced Speech Technologies for Car Environments. In *Proceedings of the 6th International Conference on Spoken Language Processing (ICSLP'2000),* pp. 378-381, Beijing, China, October 2000.

[16] D. van Compernolle. Speech Recognition in the Car: From Phone Dialing to Car Navigation. In *Proceedings of the 5th European Conference on Speech Communication and Technology (Eurospeech'97),* pp. 2431-2434, Rhodos. Greece, September 1997.

[17] P. Geutner, M. Denecke, U. Meier, M. Westphal, A. Waibel. Conversational Speech Systems for On-Board Navigation and Assistance. In *Proceedings of the 5th International Conference on Spoken Language Processing (ICSLP'98)*, Sydney, Australia, December 1998.

[18] L. Dybkjær and N.O. Bernsen. Usability Issues in Spoken Language Dialogue Systems. In *Natural Language Engineering*, Special Issue on Best Practice in Spoken Language Dialogue System Engineering, 6, 3/4, pp. 243-272, September 2000,

[19] N.O. Bernsen, H. Dybkjær, and L. Dybkjær. Designing Interactive Speech Systems. From First Ideas to User Testing. Springer Verlag 1998.

[20] http://www.disc2.dk/tools/codial/index.html