HANDS-FREE IN-CAR SPEECH INTERACTION IN THE VICO PROJECT

Marco Matassoni, Maurizio Omologo and Piergiorgio Svaizer

ITC-irst, 38050 Povo (TN), Italy

ABSTRACT

This paper presents goals, system architecture and main results of VICO (Virtual Intelligent CO-driver), a project aiming at the development of an intelligent conversational agent enabling natural hands-free interaction between humans and digital devices in the car. The underlying design principles and system framework are given, as well as an overview of the components of the final prototype working in a real environment. Finally the results of the evaluation phase are discussed.

1. INTRODUCTION

In the automotive environment, speech recognition technology is progressively being introduced as an alternative and extension of the traditional tactile interfaces used in driver information systems.

During the last years simple command-and-control applications have been studied in research activities, such as inside the European projects VODIS [1] and SENECA, and led to current commercially available in-car systems. Here the strictly menu-based input structure is guided by a very rigid dialogue, requiring to separately enter each item of a complex user request. Hence, the utilization of natural language as means of operation is a logical consequence to facilitate the usage of in-car services and devices.

This paper presents the final achievements in the VICO project [2], where the overall objective was the creation of a conversational speech interface allowing natural, user-friendly, safe and comfortable communication with a virtual co-driver in spite of the adverse conditions typical of the automotive environment. Section 2 introduces the consortium and the goals of the project. The framework of the developed system and its modules are briefly presented in Section 3, while Section 4 describes the evaluation activity carried out with the final prototype. Some conclusions are drawn in Section 5.

2. THE VICO PROJECT

VICO[1], the Virtual Intelligent Co-Driver, involved Robert Bosch GmbH (Germany) as consortium leader, DaimlerChrysler AG (Germany), ITC-irst (Italy) and Tele Atlas N.V. (Belgium) and aimed at the development of an intelligent conversational agent enabling ubiquitous natural interaction between humans and digital devices and services. Focusing on the automotive environment, the project provided a user-friendly natural language interface for services, such as navigation, on-the-fly route planning, interactive hotel and restaurant reservation and tourist information. The main concerns of the project were robustness of speech technology in adverse environments and design of adaptive mixed-initiative dialogue strategies integrated in a user-friendly and safe-to-use vocal interface.

The virtual intelligent co-driver acts as a travel guide allowing queries to certain destinations (specific address) or points-of-interest (POIs, e.g. parking, airport, or the "next" available gas station), information about tourist attractions (such as history, opening hours, entrance fees etc.) as well as information on restaurants or hotels. It also serves as a booking assistant to the driver for what concerns hotel and restaurant reservation. The applications are demonstrated in the geographic area of Trentino (Italy) and Baden-Württemberg (Germany). The prototype covers three languages (English, German and Italian).

3. SYSTEM DESIGN AND COMPONENTS

When designing the VICO system architecture and framework [3] the following design issues have been taken into account: extensibility to new features, devices and services; easy integration of additional languages; portability to other domains; robustness and fault tolerance; distributed software development at various partner sites.

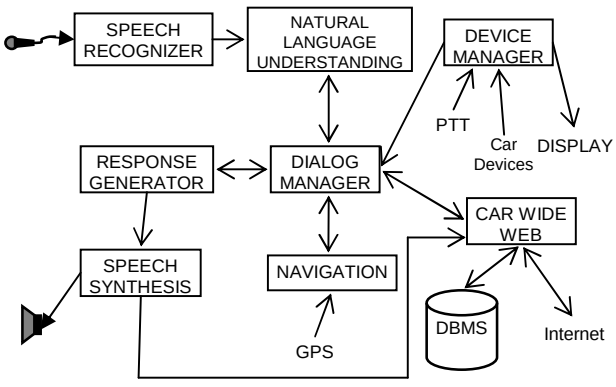

Figure 1: The VICO architecture: the data exchange among modules is provided by the System Manager within a CORBA framework.

[1] The project was partly funded by the European Commission (IST Programme 2000-25426).

The analysis of the planned functionalities led to the development of a set of modules, each covering a certain function within the system as shown in the VICO architecture reported in Figure 1:

- the System Manager component (SM) implements the communication flow inside the VICO system and provides central maintenance with supervising, logging and testing functionalities through a comfortable graphical user interface.
- the Speech Recognizer (SR) has to cope with low-SNR signals acquired in hands-free modality in an adverse environment and it must face spontaneous speech effects as well as human and non-human noises, non-native speaker pronunciations, out-of-vocabulary words (the dictionary size is more than 10000). The driver enables the system pressing a button on the wheel (push-to-talk strategy) and then talking freely as the microphone is mounted near the sun-visor. Robust end-point detection was developed and integrated in order to provide speech-only input to the recognizer. A comprehensive word hypothesis graph with scored results is delivered to the NLU component.

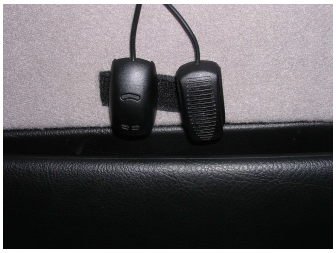

Figure 2: The microphone mounted near the sun-visor.

- the Natural Language Understanding (NLU) component relies on a robust island driven parser with anytime behaviour. The resulting semantic representation contains a set of slots instantiated by the meaningful words of the utterance, depending on the context of the on-going dialogue. The contextual information for interpreting the utterance within the dialogue context is provided by the DM. The semantically analyzed output is then passed on, via an XML interface, to the DM.

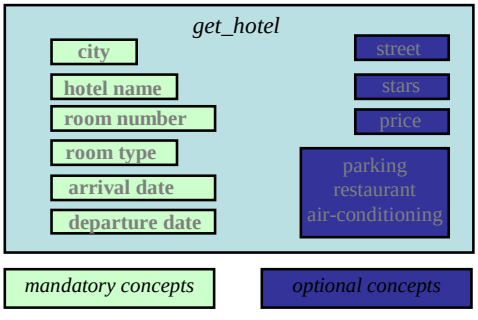

Figure 3: Example of context *get_hotel* and related concepts (the mandatory ones are requested before performing a query to the database).

- the Dialog Manager (DM) has to provide adequate reactions to all kinds of user requests in all circumstances in order to obtain and manage the information needed to fulfill the current goal. The core of the module is derived from a dialogue engine [4] that interprets the description of the specific application (characterized by a set of *contexts*, each one containing some semantically relevant *concepts* – see Figure 3 for an example) and a set of procedures describing the actions to be performed during the dialogue. The goal of the engine is to fill in the concepts of the active context in a consistent way, to trigger the necessary external actions and finally to move to another context. The DM implements different dialogue strategies (e.g. explicit/implicit confirmation) and dynamically selects the best one during the interaction.
- the DM sends a semantic representation of the system response to the Response Generation (RG) component, including a set of concepts with values and dialogue status. The RG builds the sentence using every received concept, a reliable grammar, and a corresponding natural language processing.
- a commercially available product was used in this project for the Speech Synthesis (SS). Integrating transcriptions of geographic names and cross-language transcriptions developed within the VICO project further enhanced the quality of the synthetic speech.
- the CarWideWeb (CWW) module is responsible of retrieval of geographic data for navigation, hotel and restaurant reservation as well as general tourist information. Access to the databases is provided through a specific API, so that retrieval of data is independent of the physical database structure. The CWW provides a seamless integration of different databases, and can be extended to applications such as novel multimedia in-car services.
- the Device Manager (DvM) component realizes the interface to the remote control for the push-to-talk button (PTT), that allows also to interrupt the synthesis if the user wants to speed up the dialogue or change task. In addition, it displays optional textual output generated by the RG on the user screen. An icon shows the actual status of the system to the user (listening, non-listening, system busy).

More details on the system can be found in [5,6].

4. EVALUATION OF THE ITALIAN PROTOTYPE

The final prototype covered the functionalities of navigation, hotel and restaurant reservation, and tourist information. Some videos of real interactions are available on the VICO web site [2] and an example is given in Figure 2.

VICO:	*Good morning. I am VICO, your virtual intelligent co-driver. How may I help you?*
User:	I would like to go to Trento.
<misrecognition: "Trento" is confused with "Centa">	
VICO:	*To which street in Centa do you want to go?*
User:	No, not to Centa, to Trento, please.
VICO:	*To which street in Trento do you want to go?*
User:	To the centre.
VICO:	*Do you want to go to the city centre?*
User:	Yes, right.

VICO:	*The route is being calculated.*
User:	I also need a hotel, please.
VICO:	*I found 32 hotels in Trento with 1 up to 4 stars. How many stars do you prefer?*
User:	4 stars would be great.
VICO:	*I found 4 hotels with 4 stars between 84 and 106 Euro. How much do you want to spend for a room?*
User:	Approximately 100 Euro.
VICO:	*I found 2 hotels in Trento with a price up to 100 Euro for a single room. Do you prefer the hotel "Adige" or "Accademia"?*
User:	Hotel "Accademia" sounds good.
VICO:	*Hotel "Accademia". Do you want to make a reservation?*
User:	Right.
VICO:	*How many single and double rooms do you want to reserve at the hotel "Accademia"?*
User:	One single room for tomorrow night.
VICO:	*Arriving Wednesday, 6. August. How long do you plan to stay?*
User:	7 days.
VICO:	*Do you like to reserve the hotel "Accademia" for 7 days?*
User:	Yes.
VICO:	*Your reservation is being made.*

Table 1: Example of dialogue.

The goal of the evaluation was to investigate both, the usability and the performance of the VICO system in all three applications with different levels of complexity.

The VICO final demonstrator was evaluated by Bosch, DCAG and ITC-irst. Therefore, each of the partners implemented a field experiment to have 20 subjects operate the system in a real driving situation. Every partner provided a VICO system integrated into a demonstrator car, defined a driving course and acquired subjects. While the subjects drove the car they were confronted with tasks, that they had to execute by the means of speech interaction with the VICO system. Before the experiment, the subjects were instructed to use a language that appears natural and that they felt comfortable with. Note that in this paper only the evaluation results of the Italian prototype are presented.

Figure 3: Starting screen of the system.

4.1. Setup

The VICO demonstrator was evaluated in the car in a real driving situation. Twenty Italian subjects were chosen, most of which unfamiliar with the prototype, all with medium to extensive driving experience. Twelve tasks of mixed complexity were defined, from easy "navigate to next gas station" to complex hotel reservation (see Table 2).

4.2. Experiments

The subjects were given a written introduction and a short video presentation of how the system works. The introduction included the use of the PTT button and the display, and it emphasized the fact that they had to use natural language. Next, they accomplished three trial tasks in the parked car, to get accustomed to the system. Then they started driving and the nine tasks (different from those previously shown) were given to them one by one by the supervisor in the passenger seat. All tasks were given in keywords, so as to not bias the subject with a certain wording. The full experimental time per subject was limited to 75 minutes. At the end the subjects were given a questionnaire to put down their subjective impression of the system.

1.	Zambana / Vicolo Basso
2.	Castel Beseno / description
3.	restaurant reservation / Storo / chinese food
4.	petrol station / closest
5.	Faedo / center
6.	San Lorenzo in Banale / pharmacy / opening times
7.	restaurant reservation / Andalo / restaurant La Romantica / 14:00 / 4 people
8.	hotel reservation / Riva del Garda / hotel Bellariva / 2 single rooms / tomorrow
9.	hotel reservation / Levico / 2 stars / swimming pool
10.	Terlago / restaurant / fish
11.	hotel reservation / Moena / 2 weeks / december / cheapest
12.	hotel reservation / Andalo / from 28th May / to 3rd June / 2 double rooms / for less 50 euro

Table 2: Tasks presented to the users in terms of keywords.

4.3. Results

In the experiments all subjects' utterances were recorded as well as all system steps logged. The utterances were then transcribed and merged with the dialog logging. From this data the word error rate (WER), the task completion rate (TCR) and the turns/duration per task were calculated (see Table 2). The average WER was 23% and the average TCR 82%. The dialogs took on average 10 turns or 160 seconds. To get a more comparable number across the various complexities of the tasks, times and turns were normalized to the number of Semantic Items (SI), i.e. pieces of information involved in a task (e.g. cities, streets, dates, times). Then we get 2.4 turns or 37 seconds per SI.

Task result	cases	percent
Success	201	82%
User changes some items in task	15	6%
User gives up	24	10%
End of total time (1 hour)	2	1%
Hardware problem	3	1%

Task duration	Average	Std. dev.
Time (s)	162.2	140.9
Time (s) / items	37.6	29.9
Turns	9.50	6.84
Turns / items	2.38	1.63

Table 3: Main results of the evaluation: task completion rate and duration.

Finally, on the basis of the post-questionnaires (Table 4), where a scale from 1 (worst) to 10 (best) was adopted, it turned out that the impression of usability was good. The subjective impression of the performance of the system components was satisfactory, while word recognition mistakes and unwanted task changes were named as annoying problems. The subjective judgment of distraction was moderate while safety was rated very good. The system implementation (handling, learnability, IO-scheme) was rated good while the general concept was rated very good.

Question (answer scale 0-10)	Average	Std. dev.
General impression	7.0	1.2
Utility (navigation, POI, booking tasks)	8.1	1.8
User is not distracted	3.6	2.3
VICO does not understand	4.3	2.3
Automatic task switching	3.8	2.6
Easy to learn	9.2	1.6
Usability	9.3	1.0
Push-To-Talk button	8.6	1.5
Audio quality	9.8	0.1
Sentence formulation	8.1	1.2
Monitor utility	5.9	3.3

Table 3: Results of the post-questionnaire.

5. CONCLUSIONS

This paper has briefly described system design and architecture of the VICO system as well as on-the-field evaluation results. The integrated system modules implement the required algorithms and techniques to guarantee robust speech recognition and natural language processing. Also, effective and flexible dialogue strategies were designed to allow natural interaction between VICO and a potential user.

The objective measurements showed that a speech-only input based on natural language is possible, feasible and promising. These results were confirmed by the judgment of the subjects. All parts of the conceptional setup, including the speech input and output, the push-to-talk button and the supporting display, were awarded good to outstanding grades. The VICO final demonstrator proves that a dialogue interaction in natural language for the operation of in-car devices by the driver is usable, comfortable and safe. The weaknesses in the system implementation that still challenged the users at some points will be subject to further research and fine-tuning.

Acknowledgements: This work was possible thanks to the effort of many people. In particular, we would like to thank Petra Geutner, Frank Steffens, and all the other researchers of Bosch, Daimler-Chrysler and TeleAtlas, who contributed to the success of the VICO project.
We also thank all of our colleagues of ITC-irst involved in the project: Alessio Brutti, Paolo Coletti, Luca Cristoforetti, Alessandro Giacomini, Roberto Gretter, and Mirko Maistrello.

7. REFERENCES

[1] P. Geutner, L. Arévalo and J. Breuninger, “VODIS – Voice-Operated Driver Information System: A Usability Study on Advanced Speech Technologies for Driver Information Systems”, *Proceedings of the International Conference on Spoken Language Processing (ICSLP 2000)*, Beijing, China, October 2000.

[2] *www.vico-project.org*

[3] P. Geutner, F. Steffens and D. Manstetten, “Design of the VICO Spoken Dialogue System: Evaluation of User Expectations by Wizard-of-Oz Experiments”, *Proceedings of the Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas, Spain, May 2002.

[4] D. Falavigna, R. Gretter, "Flexible Mixed Initiative Dialogue over the Telephone Network", in *Proceedings of ASRU 99*, Keystone, Colorado, 12-15 December 1999.

[5] P. Coletti, L. Cristoforetti, M. Matassoni, M. Omologo, P. Svaizer, P. Geutner, F. Steffens, “A speech-driven in-car assistance system”, in *Proceedings of Intelligent Vehicles Symposium*, Columbus (OH), 2003.

[6] A. Brutti, P. Coletti, L. Cristoforetti, P. Geutner, A. Giacomini, M. Maistrello, M. Matassoni, M. Omologo, F. Steffens, P. Svaizer, “Use of Multiple Speech Recognition Units in a In-car Assistance System”, chapter 6 of the book: Abut, Hansen, Takeda, Kazuya (eds.), *DSP for In-Vehicle and Mobile Systems*, Springer 2005.