A Multimodal Virtual Co-driver's Problems with the Driver

Niels Ole Bernsen and Laila Dybkjær
NISLab, University of Southern Denmark, Odense
nob@nis.sdu.dk, laila@nis.sdu.dk

Abstract: The paper discusses a series of four user-oriented design analysis problems in a research prototype multimodal spoken language dialogue system for supporting drivers whilst driving. The problems are: (a) when should the system (not) listen to the speech and non-speech acoustics in the car; (b) how to use the in-car display in conjunction with spoken driver-system dialogue; (c) how to identify the present driver as a basis for building a user model of the driver; and (d) how to create useful adaptive user modelling of the driver. The system discussed is under development and is called VICO or Virtual Intelligent CO-driver.

Key words: in-car spoken dialogue systems, in-car multimodal interaction, driver identification, in-car adaptive user modelling

1. INTRODUCTION

Spoken language dialogue systems (SLDSs) are now firmly positioned in the market and appear set to become available in an increasing number of languages and for a rapidly increasing number of tasks. Current commercial SLDSs help people solve a single task or sometimes several independent tasks through spoken dialogue. The dialogue is still mostly being conducted over the phone but open microphone applications are beginning to proliferate as well. The tasks solved are mainly information retrieval and/or information entry tasks but also in this respect the field is rapidly diversifying into reservation tasks, tasks involving important elements of negotiation between user and system, etc.

A typical commercial SLDS has speaker independent speech recognition; up to several thousand words in its vocabulary; modest natural language processing of the recogniser's output; increasingly modular dialogue management which often interacts with a domain database; no natural language generation; and spoken output production by means of concatenated speech, i.e. through on-line combination of recorded sentences, phrases, and words. However, due to its superior flexibility and reflecting recent increases in perceived quality and immediate intelligibility, speech synthesis is now also starting to be used for certain languages. In order to maintain control of users' spoken input behaviour, the dialogue is mostly system-directed, especially in systems intended for irregular and infrequent use. The system thus "takes the user through the task" to its completion in a structured, more or less flexible fashion [Bernsen et al. 1998].

Meanwhile, next-generation systems are gathering in the pipeline based on progress in research. These SLDSs will be able to solve a series of next-step technical challenges, including robust speech recognition in noisy conditions, large-vocabulary speaker-independent speech recognition and understanding, more efficient dialogue manager use of speech recogniser and natural language processing confidence scores; natural language processing of fully spontaneous spoken input, so that the SLDS no longer has to conduct system-directed dialogue when mixed-initiative dialogue or conversational dialogue is more appropriate; dialogue management of mutually dependent tasks; integration of adaptive user models built on-line from observations of users' behaviour; situation-aware system dialogue; and concept-based natural language generation. Moreover, next-generation SLDSs will probably no longer be mainly speech-only, or unimodal, systems but will increasingly combine speech with other modalities for information representation and exchange, enabling multimodal dialogue. Also, systems will increasingly migrate to mobile environments and devices.

This paper discusses a cluster of user-oriented design analysis problems in a research prototype multimodal SLDS which addresses the next-step challenges mentioned above in the context of supporting car drivers whilst driving. The system is called VICO (Virtual Intelligent CO-driver) and is being developed in the European HLT VICO project which began in March 2001 and has a duration of three years. The project partners are Robert Bosch GmbH, DaimlerChrysler AG, Istituto Trentino di Cultura, Phonetic Topographics N. V., and NISLab. NISLab is developing VICO's natural language understanding, dialogue management and response generation components for three languages: English, German and Italian. In the following, Section 2 provides a general description of VICO functionality and architecture. Sections 3 through 6 discuss the following problems: when should VICO listen? (Section 3), why use multimodal speech-graphics output in the car? (Section 4), how to identify the driver? (Section 5), and which aspects of the driver's behaviour should VICO model? (Section 6). Section 7 concludes the paper by discussing some of the issues for which additional research is clearly needed.

2. THE VICO SYSTEM

The car driver's environment is both a challenge and an opportunity for next-generation SLDSs developers. Important challenges include noise, from the car itself, rain, passengers, and in-car entertainment systems, large-vocabulary recognition, such as of 80.000 names of German regions, cities, streets etc., traffic safety, and ease of use by large and heterogeneous user populations. The opportunities are equally important. Car driving is a safety-critical, heads-up, hands-occupied activity in which the driver is mostly free to speak to fellow passengers and equipment but can only to a very limited extent expend valuable attention resources on GUI (graphical user interface) devices, such as screens, hand-held remote controllers, or keyboards. The car industry and user need studies concur that navigation is the "killer application" task for in-car SLDSs but that spoken interaction might be useful for many other tasks as well [Manstetten et al. 2002]. Moreover, there are strong indications that spoken car navigation and use of speech in the car more generally, cannot useably be realised by command-based SLDSs. The reason is that drivers are not able to remember the required, increasingly large number of spoken commands needed to operate in-car SLDSs. For reasons such as the above, the development of a usable and versatile in-car SLDS is an obvious "technology push" challenge whose "user pull" can be taken for granted.

To address this challenge, we are building the first of two planned prototypes of a natural interactive and multimodal in-car spoken dialogue system. The first prototype will enable navigation assistance in three languages, including streets and street numbers, parts of cities, cities, and, when relevant, parts of country for Germany, Greater London and the Trentino Province in Italy; navigation to various points of interests, such as cinemas, petrol stations, and airports, and hotel reservation over the web based on a number of driver-defined hotel selection constraints. The first prototype will also include V.0.1 of a user modelling module which will enable VICO to adapt its behaviour to the current driver. The second prototype will add full user modelling based on on-line gathered data on particular drivers as a basis for adaptive system behaviour; scenic route planning including web-based information on touristic points of interest, such as castles and churches, which will be accessed using GPS-based location awareness; car manual information; news reading; and spoken operation of in-car devices. Throughout its interaction with the driver, VICO will maintain an amount of situation awareness with respect to the car, avoiding intrusion on the driver in dangerous traffic situations. The driver-VICO dialogue will be spontaneous natural interactive dialogue, allowing the driver to address any task and sub-task in any order and using any appropriate linguistic form of expression. Finally, taking into account the in-car and out-of-car environment, VICO will incorporate aspects of multimodal communication. Thus, VICO will be activated by pushing a button and the system will provide both spoken output and graphics display output.

In the following sections, we describe our approach to some of the challenges facing VICO interaction design and development.

3. VICO HAPTICS: HOW AND WHEN TO MAKE VICO LISTEN?

An in-car spoken dialogue system faces the problem of figuring out when the registered acoustics in the cabin is actually input meant for the system or just background noise. The system also has to cope with different speaker profiles. Some of the latter factors, such as slow or fast speaking style or high or low voice, can be normalised by using adaptation methods. Noise factors such as radio or CD playing can be eliminated by echo compensation. Noises from, e.g., rain and screen sweepers can also be coped with by the recogniser and their influence on the system eliminated. Problems caused by passenger background speech can be reduced by using a microphone array with beamforming capabilities. The really hard problems come from strong accents or dialects which will not be discussed further here, and from cross-talk between the driver and the passengers while the system is listening.

In order to reduce the amount of recognition problems and non-sense dialogues which may arise from driver-passenger cross-talk, one may limit the periods during which the recogniser is listening. In the VICO project it has been decided to introduce a push-to-talk button for this purpose. To make VICO listen, the user must push the button.

3.1 Button design and interaction

The design of the push-to-talk (PTT) button has not been finally decided yet. However, it seems likely that the button will be positioned on the steering wheel. The button will be red when the recogniser is inactive and green when the recogniser is active. If the button is red and the user pushes it, it will turn green as soon as the recogniser is ready.

In addition, we will experiment with acoustic awareness so that the user does not have to look at the button to see whether it is actually red or green. Acoustic awareness may be stimulated through a non-speech sound or through spoken words or phrases, such as "hello", or "good morning". We expect that a non-speech sound will be felt less intrusive during daily use compared to using words or phrases to indicate that the system is ready. When the recogniser goes inactive after a period of input inactivity (see below), this may be indicated through a non-speech sound as well in addition to the button turning red. Using speech for this purpose, such as saying "bye", would seem less appropriate since the system may still be talking to the driver about the task.

The need for some kind of acoustic feedback on when the system is listening is supported by a set of Wizard-of-Oz experiments (see also Section 4). In those experiments, we only used a "button" on a display. The user was not supposed to push anything. The "button" would become green when the system was ready to listen. However, users were not always aware of the state of the button because they were occupied driving the car and thus were not sure when they were supposed to start speaking. Although there is a difference between just passively waiting for the button to turn green and actively pushing a button which is then expected to become green soon thereafter, it still seems likely that acoustic feedback will be appreciated since it relieves the driver from having to keep an eye on the red/green colour of the button before speaking. The acoustics is sufficient to tell the driver when the recogniser is open and when it has closed.

3.2 When to turn off the recogniser

Since we have decided that the recogniser will not just remain open once the PTT button has been pushed, we also have to find out when it is appropriate to turn off the recogniser. It would clearly not be acceptable that the driver has to push the button each time s/he wants to say something during an ongoing dialogue with VICO. On the other hand, the longer the recogniser remains open, the larger is the risk that it attempts to recognise speech not meant for the system, such as driver-passenger cross-talk. We have identified the following cases in which it seems appropriate to turn off the recogniser:

– a task has been completed and the driver does not initiate a new one within the following, say, 8-20 seconds;
– a driver stops interaction in the middle of a task but does not provide input for 10-20 seconds.

A task is considered “completed” when the negotiation with the user is finished. A user may, e.g., have asked for route guidance to a particular location. Once the system and the user have agreed where to go, the task is “completed” although the system may continue to provide route guidance output for the next 100 kilometres or more.

The system stacks a non-completed task in case the user wants to return to the task in order to complete it. For instance, a traffic situation may occupy the driver’s attention for more than 10-20 seconds, which means that the recogniser closes down. The system must be able to easily restore the dialogue state when the user pushes the button again in order to continue the unfinished dialogue.

Clearly, the solution just proposed does not completely remove the background noise problem caused by driver-passenger cross-talk. The driver may still be talking to passengers while at the same time trying to have a dialogue with VICO. We do not have data that tells us how often this will be a problem. The system may try to reduce the problem, when it occurs, through out-of-vocabulary word modelling and confidence score analysis. Thus, measures to identify input which was not meant for VICO may have to be taken by system modules other than the recogniser.

4. VICO GRAPHICS: WHEN MIGHT THE DRIVER LOOK?

Existing car navigation systems include a display on which output to the driver is shown throughout interaction. The display may be small and without map information, using an arrow to show in which direction to turn next, or it may be somewhat larger and display a map showing the present location and direction of the car in addition to the textual and iconic information which is available on the small display. Navigation information on the screen is accompanied by spoken instructions on when and where to turn. This output combination generally seems to work quite well. Even if the driver does not have much time for studying the display, many drivers still seem to appreciate the availability of display output. The advantage is that the text and graphics on the screen remains there long enough for the driver to inspect them a second time, which is not the case with speech.

What is new in VICO as regards navigation is the spoken negotiation of where to go. For input, today’s navigation systems require a remote control which the driver uses to specify the destination through prolonged interaction with the display, doing spelling, on-screen navigation, between-screens navigation, etc., cf. Figure 1. This is definitely not very traffic-safe to do. Spoken interaction will change that, of course, but, very likely, the spoken output during destination negotiation could benefit from being supported by output on the display as long as interaction mainly takes place through speech.

Figure 1. Inputting a destination with one of today's car navigation systems.

To investigate if and when the driver might want to look at the display during a destination negotiation dialogue with VICO, and to investigate the kind of information users might want on the display, we made a series of Wizard-of-Oz (WOZ) experiments with spoken input and spoken and textual output in December 2001 (3 subjects) and in January 2002 (10 subjects), see [Bernsen and Dybkjær 2001] for details on the December WOZ experiments. To simulate driving the car in traffic we used a PC car game. Subjects were seated in front of a 42" flat screen display showing the road ahead in wind-screen view, the rear-mirror view, and whatever traffic there might be, cf. Figure 2. To control the car, subjects had a force-feedback steering wheel and pedals (accelerator and brakes). Next to the large screen was a small portable computer simulating the car display and showing textual system output. For the spoken output, the Festival synthesiser was used. Each user was asked to carry out three scenarios. Subjects were interviewed after their interaction with the system.

Figure 2. Driving the simulated car.

The dialogue with VICO was in English. The scenarios all concerned route planning for Danish destinations. We experimented with three different text versions on the car display. All three versions were displayed to each user (one version per scenario) but in differing order. One version was a full textual repetition of the spoken output, a second version only included the keywords of the spoken output, and the third version only provided the agreed-upon goal as text output at the end of the dialogue.

The experiments aimed to provide data on what users would like to see on the display and in which situations they would look at the display. In the following we describe our findings.

Many subjects found it less stressful to use the car game than to drive a real car. As a major reason they indicated the fact that they knew that nothing would happen even if they crashed the car. However, some subjects found that it required much more concentration to play the car game than to drive a real car. Typically, these subjects also found it unsafe to look at the car display whilst driving.

Although most subjects found it less stressful to play the car game than to drive a real car, only a couple of them stated that they used the display quite frequently. Most subjects did not use it much, either because they found it unsafe or because they did not feel a need for it. When the display was used it was typically to check the output. The quality of the output speech was fairly low. Danish location names pronounced by an English synthesiser are often rather difficult to understand. Moreover, drivers were from time to time "disturbed" by a passenger talking to them, which meant that they missed what was being said by the system. Better synthesis is certainly available which will reduce the first problem while the second problem is not likely to go away in real driving situations.

When subjects did not hear what the system said or were not sure that they got it right, they would typically either ask for repetition or look at the display. Since most subjects only used the display infrequently, they were also not aware of the changing amount of feedback. A couple of users complained that there was nothing on the display when they looked. They probably checked the display when the system provided the shortest version of its text output, i.e. only the finally agreed destination would be displayed but nothing would be displayed during negotiation. Although, therefore, we did not collect that much data on the length and contents of the text displayed, it appears from the subsequent discussions with subjects that the medium-length text version was the most appropriate. The short version does not provide sufficient support since there will not always be information available on the display, while the long version contains too much superfluous information and may be hard to catch at a glance. The long version is therefore less safe and less to-the-point than the medium version which contains the key information expressed as briefly as possible. The main functions of the text output are to allow the driver to check correctness and the present state of progress of the dialogue. Even if oral dialogue requires less effort from the user while driving than reading text on a display, situations may occur in which the driver stops listening in order to handle a difficult traffic situation. Returning to the dialogue, the driver may either catch up by asking the system where they were at or by checking the display. We expect that there will be individual differences as to which of the two options people prefer in the same way as our subjects behaved differently with respect to their use of the display versus spoken dialogue. People have different driving experience and different habits which are likely to influence their preferences, so both options should probably be enabled.

5. WHO IS DRIVING THIS TIME?

VICO will incrementally build and use a user model for each driver of the particular car in which the system is installed (see Section 6). In order for VICO user modelling to be of any use, VICO must be able to determine which of the car's drivers is currently driving, given the fact that cars often have several different drivers. Furthermore, it would seem that driver identification has to be made with *near-certainty*. If it is too uncertain that VICO has correctly identified the current driver, driver misidentification will happen too often. The misidentified drivers will be "mistreated" by VICO because VICO adapts to them based on the wrong user model, and their behaviour will fudge up the misallocated user models with irrelevant and misleading information. In addition, the driver must be *identified*, or recognised as a new driver, *up front*, i.e. before or, at the latest, as soon as that driver starts the dialogue with VICO. If the driver is identified at any later time, the driver will be helped less by the user model and the user model may be missing important information on the driver's behaviour.

So, how to identify the driver? There are several possibilities, at least:

- *voice identification.* Even though today's voice identification technology is not perfect, it might be possible to get near-certain identification in the VICO case, simply because most cars have rather few drivers. Voice identification is also an elegant solution because the driver does not have to do anything other than speak to VICO about some task. It is not necessary for the driver to even explicitly address the identity issue. Finally, voice identification happens up front as soon as the driver speaks to VICO;
- *driver's code.* "Password" may be a bit misleading here since the purpose is not to create in-context "unbreakable" passwords for security but that of making the minimal distinction needed among the car's different drivers, so that, for instance, one driver is 1, another is 2, and so on. Contrary to the case of voice identification, the driver's code must be input to VICO explicitly. The simplest way to do so is to speak the code to VICO up front. VICO might even start by asking for it. Alternatively, the code could be entered through some other modality, such as haptically through one or several keys. This raises a new issue of adding key(s) to the VICO system, which has not been foreseen and which may complicate the VICO system unnecessarily. Simplicity as well as traffic safety, therefore, speaks for an acoustic code per driver. An advantage of the code solution is that correct driver identification is guaranteed, assuming that the driver remembers the code correctly and VICO provides feedback on the code provided. One drawback is that the driver has to remember yet another code/password. Another possible drawback, which the driver's code shares with voice identification, is that VICO will not be able to respond by saying, as people would naturally do, e.g. "Hello Ole, what would you like to do today?". That would require VICO to be able to match the code, say, "2", to Ole, which again would take keys or something. It should be added that an

unknown percentage of drivers might not appreciate this kind of feedback anyway.

- *driver's spelled first name.* This solution would enable VICO to give verifiable non-coded feedback to the driver that the driver has been correctly identified, simply by composing the word spelled and sending it through the synthesiser. However, spelling one's full first name each time one has to interact with VICO is an awkward thing to do, the more so the longer the name is (e.g. Elisabeth). Maybe the driver would only have to spell, say, the first two letters in order to be called by VICO "El" forever after? Even assuming that most drivers could live with that, it would not work in a family in which the woman was called Petra and the man Peter. And sometimes father and son have identical first names. Such drivers might have to return to personal codes/passwords;
- *combinations of the above*, such as voice identification plus spelling of their first names by first-time users. This couples the elegance of voice identification with the non-coded feedback from VICO on which driver VICO has identified.

In conclusion, voice identification is the simpler and most elegant of the solutions possible. It is not clear at this point if this solution will be able to achieve the required virtual certainty of identification. If not, we will have to look for other solutions. However, also voice identification suffers the drawback that VICO will have no match between the identified voice and, e.g., the driver's name, which could be used to confirm to the driver that driver identification was successful. It is not clear what to do about that in any simple way since the first name-spelling option has problems of its own.

Another issue which has not been mentioned above, is that *passengers* might want to talk to VICO as well. Normally, it seems, a car has fewer different drivers than different, front seat or other, passengers. If these also talk to VICO, user modelling becomes significantly more diverse. VICO might come to have dozens of user models for a particular car, most of which are not being used at all since they were simply created when some passenger wanted to try to speak to VICO once and for all. We might assume that all or most passengers will only talk to VICO once or a few times. This should not fudge up too much the user model statistics on a particular driver. In voice identification mode, no fudge-up will happen because VICO will simply create a new user model for that passenger. Still, a driver's code or a driver's spelled first name might come out more elegant by comparison because, in these cases, VICO does not necessarily have to establish a user model for unidentified speakers in the first place. Possibly, VICO could talk with anyone but would only establish and maintain user models for those who provide their code/spelled first name up front.

6. MODELLING THE DRIVER

Once the driver starts speaking to VICO, the system must try to identify that driver and retrieve the appropriate user model, if any. If identification fails, VICO assumes that the driver is new to VICO and creates a new user model (UM). In both cases, VICO will collect relevant information on the driver's behaviour during the spoken interaction and use that information to update its model of the driver. Expressed in a slightly more systematic fashion, VICO's UM-related tasks are:

1. identify the present driver (cf. Section 5);
2. retrieve the present driver's user model;
3. create a new user model UM(Dx) for a new driver, Dx;
4. make appropriate use of the present driver's user model during the driver's dialogue with VICO;
5. collect new information on the present driver during the driver's dialogue with VICO;
6. update the present driver's user model with the new information gathered;
7. store the user model whenever it has been updated with new information.

From a design viewpoint, the hardest problem in the list above probably is Point 4 followed by points 6 and 5. However, before addressing those problems, decision must be made on *which type(s) of information* on the driver the system should collect, store, and use. This problem appears to be the hardest of all. The reason is a rather general one. When embarking on adaptive user modelling in VICO, we enter a technical area fraud with difficulty and failure. Adaptation is among the most difficult things to do in developing interactive computer systems, independently of whether those systems use speech or other modalities. In fact, user adaptation has proved so difficult to do that it seems fair to say that, by and large, and despite numerous attempts in the past 15-20 years, research and industry have failed in developing useful adaptive functionality in the huge numbers of interactive systems which already exist. There are successful exceptions, of course, but these are few and functionally simple. The conclusion we should draw from that fact is that we must be extremely careful in selecting what we want to do. It is better to succeed with one, or a few, adaptive functionalities in VICO than to fail through ignorance of the difficulties involved by trying to develop an unrealistic number of poor adaptive functions.

As it turns out, it appears that we may distinguish between several different types of information which VICO could collect and use adaptively. A possible typology of information which VICO might use is the following. At least three different generic kinds of information about particular drivers may be distinguished:

1. information on the driver's task objectives;

2. information on the driver's communication with VICO;
3. information on the driver's experience of various kinds.

In the VICO context, each of these generic kinds of information subsume several more specific information types, such as the driver's hotel preferences (1), the driver's difficulties in being understood by VICO due to strong accent or dialect (2), or the driver's experience in using VICO itself (3). In other words, the information typology helps generate a structured space of candidates for adaptive user modelling.

To further constrain the user modelling capabilities of VICO, we have identified the following criteria which should be met by a particular kind of driver information in order for that information to be collected and used by VICO:

1. include at least one user modelling functionality belonging to each type in the typology of generic information about the driver described above;
2. the chosen user modelling functionalities should be top quality in terms of their usefulness to all or most drivers;
3. the user modelling functionalities should provide genuine driver adaptivity without significant drawbacks;
4. the user modelling functionalities should be possible to implement without extreme effort (since we do not have the time for putting extreme effort into them);
5. the user modelling functionalities must be based on clearly verifiable information about the driver.

Space does not allow to present the pros and cons with respect to each candidate on the long list of potential information sub-types which we have analysed. An example sub-type of Type (1) in the typology above, i.e. *information on the driver's task objectives*, is: store the driver's past hotel preferences, such as number of stars and possibly other selection constraints as well. Even if not told about them by the driver, VICO could offer to use those constraints as selection criteria when finding a hotel. It is important, of course, that the driver is able to override those constraints and provide new ones. If not, all we will be doing is to produce yet another failed attempt at creating useful system adaptivity. However, in this case it is easy for the driver to override VICO's suggestions because the driver will be asked before VICO proposes any hotels meeting those selection constraints. This user modelling functionality can be implemented without extreme effort, cf. Criterion (4) above. The functionality is based on clearly verifiable information about the driver (Criterion 5 above), i.e. it is possible to write an update algorithm which only produces VICO hotel property offerings when a clear pattern can be discerned in the driver's hotel preferences. Also, the functionality under consideration does not appear to have any significant drawbacks (Criterion 3 above). So, the final question is whether information on the driver's observed hotel preferences is top quality in terms of its usefulness to all or most drivers (Criterion 2 above). This question is a difficult one, because the answer to the question depends on (i) how many users of VICO will actually need hotels, and (ii) how many of those users have systematic hotel preferences. We do not know the answer to that question at this point but clearly need to find out the best we can in order to be able to rank the user modelling option just described among its competitors.

Compared to the problem of identifying a user modelling candidate for information about the driver's task preferences, it would seem considerably harder to build adaptive user modelling with respect to Type (2) *information on the driver's communication with VICO.* An example is a system which adapts its dialogue behaviour to drivers whose strong dialect or accent makes their dialogue contributions difficult to recognise. One issue, of course, is that we might need two significantly different dialogue structures to accommodate both standard drivers and drivers with strong dialect or accent, making a solution relatively costly to implement (Criterion 4 above). A second problem is that any solution may be at risk as long as we do not have efficient ways of determining the causes of recognition problems. Recognition confidence scores, for instance, cannot tell VICO whether the cause of recognition problems are a strong dialect or accent or something entirely different, such as a driver who regularly talks to passengers whilst having a dialogue with VICO. Similarly, the measurable fact that the driver makes unusually many error corrections may be due to many different causes (cf. Criterion 5).

Type (3) *information on the driver's experience* includes an obvious candidate for adaptive user modelling, i.e. the driver's experience with VICO itself. The idea is to offer information on VICO to all new drivers independently of whether a new driver asks for it or not. Provision of this information would seem to be top quality in terms of its usefulness to all or most drivers (Criterion 2) as well as providing genuine adaptivity without any significant drawbacks (Criterion 3). This assumes, of course, that drivers are identified with near-certainty, as discussed above. Implementation is simple (Criterion 4), and, as argued in Section 5, it is clearly verifiable if the driver is new to VICO or not (Criterion 5).

7. CONCLUSION AND FUTURE WORK

In this paper we have discussed four issues of importance to future in-car information systems development. At the time of writing, only one of these issues have been resolved to our satisfaction, i.e. the issue of which driver-system dialogue-relevant information to present on the in-car display. And even that issue still leaves several aspects open, such as whether to continue to graphically display, in addition to the dialogue-relevant (text) information, diagrammatic and iconographic information, such as maps, arrows, and the like, including the offering of less traffic-safe functionality, such as on-screen navigation, between-screens navigation, display information customisation, etc. The car display can be used to display a wealth of information in addition to the spoken dialogue-supporting text proposed in this paper. The problem is to find a balanced solution which will not re-introduce the remote control.

As for the three other issues discusses above, we are still investigating the pros and cons of different solutions. Thus, the duration of the time window in which the system should be listening to the driver will form the topic of future experimentation. Similarly, the problem of driver-passenger conversation while the system is listening continues to demand a more efficient solution than any we have investigated so far. As regards driver identification, we are still investigating what the most elegant, useful and usable solution might be. The same applies, in part, at least, to the issue of adaptive driver modelling. Potentially, adaptive driver modelling could be extremely useful to drivers, yet the complexity of the options, trade-offs and technical issues involved would seem to make adaptive driver modelling a highly interesting research challenge which is likely to occupy researchers for some time until the terrain has been appropriately charted and useful solutions identified.

8. ACKNOWLEDGEMENTS

VICO is supported by the EU HLT (Human Language Technologies) Programme under Contract IST-2000-25426. We gratefully acknowledge the support.

9. REFERENCES

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

Bernsen, N. O. and Dybkjær, L.: Exploring Natural Interaction in the Car. Proceedings of the International Workshop on Information Presentation and Natural Multimodal Dialogue, Verona, December 2001, 75-79.

Manstetten, D., Berton, A., Krautter, W., Grothkopp, B., Steffens, F. and Geutner, P.: Evaluation report from simulated environment experiments. VICO report D7, DaimlerChrysler, January 2002.