Multimodal Menu-based Dialogue with Speech Cursor in DICO II+

Staffan Larsson
University of Gothenburg
Sweden
sl@ling.gu.se

Alexander Berman
Talkamatic AB
Sweden
alex@talkamatic.se

Jessica Villing
University of Gothenburg
Sweden
jessica@ling.gu.se

Abstract

This paper describes Dico II+, an in-vehicle dialogue system demonstrating a novel combination of flexible multimodal menu-based dialogueand a "speech cursor" which enables menu navigation as well as browsing long list using haptic input and spoken output.

1 Introduction

Dico is a multimodal in-car dialogue system application, originally developed in the DICO (with capital letters) project by Volvo Technology AB and the University of Gothenburg. Dico is built on top of the GoDiS dialogue system platform (Larsson, 2002), which in turn is implemented using TrindiKit (Traum and Larsson, 2003).

The main goal of the original Dico application (Olsson and Villing, 2005), (Villing and Larsson, 2006) was to develop an interface that is less distracting for the driver, and thus both safer and easier to use than existing interfaces. (Larsson and Villing, 2009) described the Dico II system resulting from work in the DICO project. Since then, the Dico demonstrator has been further developed.

In this paper, we describe the Dico II+ demonstrator which introduces a novel combination of flexible Multimodal Menu-Based Dialogue and a Speech Cursor which together enable flexible multimodal interaction without the need for looking at the screen. In the following, we will first argue for the usefulness of in-vehicle dialogue systems. We will then briefly present the GoDiS platform which Dico II+ is based on, as well as some aspects of flexible dialogue enabled by the GoDiS dialogue manager.

2 In-vehicle dialogue systems

Voice interaction is a very natural means of communication for humans, and enabling spoken interaction with technologies may thus make it easier and less cognitively demanding for people to interact with machines. However, this requires that the spoken interaction is similar to ordinary spoken human-human dialogue. A problem with available in-vehicle voice control technologies is that they are not flexible enough in terms of the interaction strategies and modalities offered to the user.

3 GoDiS features in Dico

GoDiS (Larsson, 2002) is an experimental dialogue system implementing a theory of Issue-Based Dialogue Management based on Ginzburg's concept of Questions Under Discussion (QUD). GoDiS is implemented using the TrindiKit, a toolkit for implementing dialogue move engines and dialogue systems based on the Information State approach (Traum and Larsson, 2003). GoDiS has been adapted to several different dialogue types, domains, and languages, including menu-based multimodal dialogue when acting as an interface to an mp3 player (Hjelm et al., 2005).

The GoDiS dialogue manager allows the user to interact more flexibly and naturally with menu-based interfaces to devices. General dialogue management issues such as accommodation, task switching and grounding are handled by the application-independent dialogue manager. Re-using these technologies in new applications enables rapid prototyping of advanced dialogue applications.

3.1 Accommodation

The applications in Dico II+ are based on existing menu interfaces, using a method for converting menus into dialogue plans (see below). While it is possible for the novice user to let the system take initiative and guide the user through the menus step-by-step, expert users might prefer to use *accommodation* strategies enabling the user to skip through the menus and getting right to the point. Two examples are shown below.

Ex. 1:

S: "What do you want to do? Go to the phonebook, manage settings, or..."

U: "Call Lisa's home number"

S: "OK. Calling Lisa's home number."

Ex. 2:

U: "Lisa's home number"

S: "OK. Do you want to change the number or make a call?"

U: "Make a call."

S: "OK. Calling Lisa's home number."

3.2 Multiple simultaneous tasks and task switching

GoDiS enables arbitrarily nested subdialogues. It is possible to start one dialogue to perform a task, and then start a subdialogue before the first task is completed. When the second task is completed the system automatically returns to the first task, and explicitly signals this. This gives the user freedom to switch task at any time:

Ex. 3:

U: "Change Lisa's home number."

S: "Okay. Let's see. What phone number do you want instead?"

U: "Check my messages."

S: "You have got this message. Hi! I have got a new home number, it is (031)234567. Best regards, Lisa."

S: "Returning to change an entry. What phone number do you want instead?"

U: "oh three one twentythree fourtyfive sixtyseven."

U: "Okay. Changing Lisa's home number to oh three one two three four five six seven."

3.3 Feedback and grounding

The GoDiS dialogue manager provides general feedback strategies to make sure that the dialogue partners have contact, that the system can can hear what the user says, understands the words that are spoken (semantic understanding), understands the meaning of the utterance (pragmatic understanding) and accepts the dialogue moves performed in utterances.

As an example, the single user utterance "Lisa" may result in positive feedback on the semantic level but negative on the pragmatic, resulting in a system utterance consisting of two feedback moves and a clarification question: "Lisa. I don't quite understand. Do you want to make a call, change an entry in the phonebook, or delete an entry from the phonebook?"

4 Multimodal menu-based dialogue

Dico II+ implemented a concept of Multimodal Menu-based Dialogue (MMD). Technologies for MMD in menu-based applications have already been developed for other GoDiS applications (Hjelm et al., 2005) and the ideas behind these solutions were re-implemented and significantly improved in Dico.

A common argument for using spoken interaction in a car context is that the driver should be able to use a system without looking at a screen. However, there are many situations where current technology requires the user to look at a screen at some point in the interaction. The idea behind MMD is that the user should be able to switch between and combine modalities freely across and within utterances. This makes it possible to use the system using speech only, using traditional GUI interaction only, or using a combination of the two.

MMD enables *integrated multimodality* for user input, meaning that a single contribution can use several input modalities, e.g. *"Call this contact [click]"* where the [click] symbolises haptic input (e.g. a mouse click) which in this case selects a specific contact. For output, MMD uses *parallel mul-*

timodality, i.e., output is generally rendered both as speech and as GUI output. To use speech only, the user can merely ignore the graphical output and not use the haptic input device. To enable interaction using GUI only, speech input and output can be turned on or off using a button which toggles between "speech on" and "speech off" mode.

The GUI used in Dico II+ is a generic graphical interface for the GoDiS system, developed by Talkamatic AB with graphical adaptations for Dico. It represents GoDiS dialogue moves graphically as menus using a refined version of the conversion schema presented in (Larsson et al., 2001) . For example, alternative questions are represented as multiple choice menus, and wh-questions are represented as scrollable lists. Conversely, haptic user input from the GUI is interpreted as dialogue moves. Selecting an action in a multiple-choice menu corresponds to making a *request* move, and selecting an item in a scrollable list corresponds to an *answer* move.

5 Speech Cursor

This section describes an important addition to the GoDiS dialogue manager and Dico demonstrator, which enables the user to use spoken interaction in combination with haptic input to access all functionality (including browsing long lists) without ever having to look at the screen. In combination with the flexible dialogue capabilities of the GoDiS dialogue manager, and the concept of MMD, we believe that a Speech Cursor provides a powerful and user-friendly way of interacting with menu-based interfaces in cognitively demanding environments such as the in-vehicle environment.

5.1 The problem

A common argument for using spoken interaction in a car context is that the driver should be able to use a system without looking at a screen. However, there are many situations where current technology requires the user to look at a screen at some point in the interaction. This was true also for Dico II in the case of browsing lists; for example, to find out which contacts were listed in the phonebook, the user would at some point have to look at the screen.

Imagine that the user wants to select a song from a song database, and that the user has made restrictions filtering out 30 songs from the database. The dialogue system asks the user which of the songs she wants to hear displaying them in a list on the screen.

The user must now either look at the screen and use a scroll-wheel or similar to select a song, or look at the screen to see which songs are available, and then speak the proper song title. This means that part of the point of using spoken interaction in the car is lost. The example discusses car use, but is applicable any time when the user cannot or does not want to look at a screen, for instance when using a cellphone walking in a city, or when using a web application on a portable device.

An existing interaction strategy for addressing the problems of browsing lists is to allow a kind of meta-dialogue, where the system verbally presents a number of items (for instance 5) from the list, then asking the user if she or he would like to hear the subsequent 5 items, until the list has been read in its entirety or until the users responds negatively. While this strategy in principle solves the problem, it is rather time-consuming compared to browsing the list using a screen and a haptic input device (such as a scroll-wheel); this may decrease the perceived usability of the voice interface in comparison with traditional GUI-based interaction.

Some existing voice interaction systems use a technology to establish understanding which consists of displaying the top N best recognition hypotheses to the user, each one associated with a number, together with a verbal request to the user to say the number corresponding to the desired result. This situation, however, requires the user to look at a screen, and is arguably quite unnatural.

5.2 The solution: Speech Cursor

Dico II+ requires a haptic menu navigation device, such as a mouse (trackball, touch pad, TrackPoint^{TM}) with buttons, pointers and drivers, keyboard with arrow keys, or jog dial/shuttle wheel. A typical in-vehicle menu navigation device consists of three or four buttons (UP, DOWN, OK and possible BACK).

Every time a new item gets focus, the system reads out a voice icon - a spoken representation of the item. This representation can be textual, intended to be realised using a TTS, or in the form

of audio data, to be played directly. Every time a new element gets focus, all any ongoing voice output is interrupted by the voice icon for the element in focus.

6 Combining flexible dialogue and Speech Cursor

Combining flexible dialogue processing with the Speech Cursor concept[1] enables interaction where the user may freely choose between using domain-level spoken utterances (requests, confirmations, questions, answers etc.) and using the Speech Cursor. This is an improvement over existing technology in that it offers a greater variety of interaction styles which can be freely chosen and combined by the user, as well as offering flexible dialogue interaction. Specifically, it provides flexible spoken dialogue interaction in combination with user-friendly browsing of long lists without the need for looking at the screen.

Here is a walk-through of a sample interaction:

Ex. 4:

U: "I want to listen to Madonna"*(This utterance uses accommodation to allow the user to supply unrequested information)*

S: "There are 3 songs by Madonna. What song do you want?" *(Showing list of all songs by Madonna: 'Like a Prayer', 'La Isla Bonita', 'Music', ...)*

U: [DOWN]

S: "Like a Prayer from the album Like a Prayer" *("Like a Prayer" is highlighted)*

U: [DOWN]

S: "La Isla Bonita from the album True Blue" *("La Isla Bonita" is highlighted)*

U: [DOWN]

S: "Music from the alb+" *("Music" is highlighted)*

U: [UP]

S: "Like a Prayer"

U: [OK]

S: "OK, playing Like a Prayer."

Here is a further example:

Ex. 5:

U: "I want to add an ABBA song"

S: "What album?" *(Shows "Waterloo" and "Arrival)*

U: [DOWN]

S: "Wat..."

U: [DOWN]

S: "Arrival"

U: [OK]

S: "what song?" *(Shows "Mamma Mia" and "Money Money Money")*

U: "Mamma Mia"

A promotional demo film showing interactions similar to the above is available[2], courtesy of Talkamatic AB.

Acknowledgments

The work reported here was funded DICO, Vinnova project P28536-1. Additional funding was provided by Talkamatic AB, and by CLT, the Centre of Language Technology at the University of Gothenburg. Dico II+ was implemented by the authors, Fredrik Kronlid, Peter Ljunglöf and Andreas Wiberg. The authors gratefully acknowledge the assistance of Volvo Technology AB and the DICO project group. The GoDiS system is the property of Talkamatic AB.

References

David Hjelm, Ann-Charlotte Forslund, Staffan Larsson, and Andreas Wallentin. 2005. DJ GoDiS: Multimodal menu-based dialogue in an asychronous isu system. In Claire Gardent and Bertrand Gaiffe, editors, *Proceedings of the ninth workshop on the semantics and pragmatics of dialogue*.

[1] The combination of a (simple or flexible) spoken dialogue system and the voice cursor technology is patent pending.

[2] `www.youtube.com/watch?v=yvLcQOeBAJE`

Staffan Larsson and Jessica Villing. 2009. Multimodal menu-based dialogue in dico ii. In Jens Edlund, Joakim Gustafson, Anna Hjalmarsson, and Gabriel Skantze, editors, *Proceedings of DiaHolmia, 2009 Workshop on the Semantics and Pragmatics of Dialogue*.

Staffan Larsson, Robin Cooper, and Stina Ericsson. 2001. menu2dialog. In *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, pages 41–45.

Staffan Larsson. 2002. *Issue-based Dialogue Management*. Ph.D. thesis, Göteborg University.

Anna Olsson and Jessica Villing. 2005. Dico - a dialogue system for a cell phone. Master's thesis, Department of Linguistics, Goteborg University.

David Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In Ronnie Smith and Jan Kuppevelt, editors, *Current and New Directions in Discourse & Dialogue*. Kluwer Academic Publishers.

Jessica Villing and Staffan Larsson. 2006. Dico - a multimodal in-vehicle dialogue system. In D. Schlangen and R. Fernandez, editors, *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue*.