Adding a Speech Cursor to a Multimodal Dialogue System

Staffan Larsson 1, Alexander Berman2, Jessica Villing1

1Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden
2Talkamatic AB, Gothenburg, Sweden

sl@ling.gu.se, alex@talkamatic.se, jessica@ling.gu.se

Abstract

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

Index Terms: dialogue systems, multimodality

1. Introduction

Dico is a multimodal in-car dialogue system application, originally developed in the DICO (with capital letters) project by (self-reference deleted). Dico is built on top of the GoDiS dialogue system platform [1], which in turn is implemented using TrindiKit [2].

The main goal of the original Dico application [3], [4] was to develop an interface that is less distracting for the driver, and thus both safer and easier to use than existing interfaces. [5] described the Dico II system resulting from work in the DICO project. Since then, the Dico demonstrator has been further developed. An important feature of the most recent version, Dico II+, is that it enables 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 [1] 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 [2]. 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.

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.

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 [6] 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 multimodality, 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.

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 "metadialogue", 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 possibly also 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:

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 album +" ("Music" is highlighted)
U: [UP]
S: "Like a Prayer"
U: [OK]
S: "OK, playing Like a Prayer."

A promotional demo film showing interactions similar to the above is available at http://www.youtube.com/watch?v=yvLcQ0eBAJE, courtesy of Talkamatic AB.

7. Acknowledgments

The work reported here was funded DICO, Vinnova project P2836-1. 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.

8. References


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