Comparing system-driven and free dialogue in in-vehicle interaction

Fredrik Kronlid[1], *Jessica Villing*[2], *Alexander Berman*[1], *Staffan Larsson*[2]

[1]Talkamatic AB, Gothenburg, Sweden
[2]Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, Sweden
fredrik@talkamatic.se, jessica@ling.gu.se, alex@talkamatic.se, sl@ling.gu.se

Abstract

It is widely held that a free, natural dialogue model is more efficient and less distracting than system-initiative, state based dialogue. This paper describes an evaluation of two systems - one using system-directed dialogue and one using a more "free" dialogue - focusing on distraction and efficiency. The level of distraction is measured using an automotive industry standard test (LCT), and the efficiency is measured by counting the number of completed tasks. The efficiency is increased by 42 % using the free, natural dialogue model while the LCT results are unclear. Using a free dialogue model increases the efficiency and reduces the distraction in some cases.

Index Terms: dialogue systems, system initiative, mixed initiative, lane change task

1. Introduction

The purpose of this evaluation was to measure how efficiency and distraction differs between a dialogue system using the dialogue model imposed by the GoDiS dialogue manager [1] and a commercially availablevoice enabled navigator, also offering telephone voice control (Navigon 8410), henceforth called "the reference system". The reference system was selected because of favourable comments in tests regarding the voice interaction functionality [2, 3].

2. Background

2.1. The reference system

According to the reference system product website[1], the reference system features "Voice Interaction Pro" which "understands you even in colloquial language". Voice interaction is system-directed, which means that the system asks a questions, and the user provides answers system questions. The user must give a direct answer to the question asked, and cannot supply more information or other information than requested. Here is a sample interaction with the reference system:

> System: Address, POI nearby, or My Destinations?
> User: Address
> S: What is the name of the town?
> U: Gothenburg
> S: What is the name of the street?
> U: Avenyn
> ...

[1]http://www.navigon.com/portal/int/produkte/navigationssysteme/navigon-premium/navigon_8410.html

2.2. The VCC/Talkamatic system

Talkamatic[2] has developed an advanced dialogue system to complement a graphical user interface for a concept audio system developed at VCC (Volvo Cars). The dialogue system is based on the dialogue manager GoDiS [1], developed at the University of Gothenburg, and currently owned and maintained by Talkamatic. We will refer to the the dialogue system applications (telephone and navigation) based on the VCC audio system and implemented on top of the Talkamatic dialogue system (which in itself is application independent) as as the "VCC/Talkamatic system" below.

The VCC/Talkamatic system represents a novel dialogue model, compared to the dialogue systems usually offered in a car setting. The dialogue system allows the user to give little or much information in one utterance, offers integration with a graphic/haptic interface and contains many linguistically motivated features, such as a rich model for establishing shared understanding (a.k.a. grounding). A similar system, built on an earlier version of the same platform, is described in [4].

GoDiS [1] is a dialogue manager implementing a theory of Issue-Based Dialogue Management based on Ginzburg's concept of Questions Under Discussion (QUD). GoDiS is implemented using TrindiKit, a toolkit for implementing dialogue move engines and dialogue systems based on the Information State approach [5]. 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 [6].

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.

2.3. Accommodation

The applications in GoDiS 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. An example is shown below.

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

[2]Talkamatic AB is a spin-off company from the Department of Philosophy, Linguistics and Theory of Science, University of Gothenburg, aiming to commercialise the GoDiS dialogue manger.

U: Call Lisa's home number
S: OK. Calling Lisa's home number.

2.4. 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:

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.

2.5. 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?"

2.6. Multimodality

The Talkamatic dialogue system based on GoDiS also 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 have been re-implemented and significantly improved.

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.

The GUI used in the VCC/Talkamatic system is a generic graphical interface for the GoDiS system, developed by Talkamatic AB. It represents GoDiS dialogue moves graphically as menus using a refined version of the conversion schema presented in [7] . 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.

3. Method

We selected LCT (Lane Change Task) [8] for determining the level of user distraction. LCT measures the level of driver distraction by recording deviations from an ideal track on a simulated roadway when performing a secondary task such as using a Kleenex, tuning to a certain radio station etc.

In order to reduce effects from graphical/haptic interaction differences (as the purpose of the study was to evaluate dialogue models) the subjects were only able to give input to the system using their voices. The subjects were, however, allowed to look at the screens, as some vital information was displayed graphically only. The screens for both systems had the same visible size and were placed in the same direction relative to the driver.

The efficiency was estimated by measuring how many times the users were able to complete the secondary tasks during the length of one road segment (3 minutes).

Complementary information was acquired using a questionnaire. The complementary information consists of user estimated distraction level, preference of voice interaction model and personal information such as age, etc.

The tasks given to the users (one telephone task and one navigation task) were possible to perform in both models. The telephone task consisted of calling contacts in the telephone book and alternating between four contacts[3], some of which had a single phone number, some of which had several phone numbers. The navigation task consisted of entering destinations (well-known addresses in Gothenburg and in Stockholm).

The 13 subjects[4] were introduced to the LCT setup, and were encouraged to practise driving in the simulator. Before driving with a certain secondary task, the subject was encouraged to practise executing the task.

The subjects followed the following schedule:

1. Baseline driving, no secondary task
2. System A Telephone Task
3. System A Navigation Task
4. System B Telephone Task
5. System B Navigation Task
6. Baseline driving, no secondary task

In order to compensate for priming effects resulting from subjects using one voice interaction model before another, the subjects were divided into two groups. The first group (7 subjects)

[3] To make it easier for the subjects to remember the names of the contacts, their names were identical to those of some well-known characters from Swedish author Astrid Lindgren.

[4] Originally 14. One subject could not complete the tests due to motion sickness. One of the subjects did not finish the navigation tasks due to scheduling problems.

were assigned VCC/Talkamatic as system A and the reference system as system B, while the second group (6 subjects) were assigned the opposite order.

After each of the tasks the subjects were asked to estimate their driving ability, using a number between 0 and 10, where 10 represented "on par with driving with no secondary task" and 0 represented "unable to drive".

4. Results

4.1. LCT

When comparing the LCT results for the telephone task, VCC/Talkamatic (9.76, SD 0.36, range 0.15-0.98) is indicated as less distracting than the reference system (11.27, SD 0.53, range 0.39-0.93). For the navigation task, the results are the opposite: 10.75, SD 0.55 and range 0.27-2.39 for VCC/Talkamatic while the reference system has a mean deviation from the normal driving track of 9.38 (SD 0.26, range 0.58-1.01).

These results are reflected also in the individual LCT scores. VCC/Talkamatic has a better LCT score in 8 cases out of 13 (62 %) in the telephony task, while the reference system has the better score in 7 cases out of 12 (58 %) in the navigation task.

4.2. Estimated driving ability

The user estimations of driving ability mostly favour the VCC/Talkamatic system. In the telephone task 9 out of 13 estimated their driving ability as higher with VCC/Talkamatic than with the reference system, 2 out of 13 being equal and 2 out of 13 lower than with the reference system. In the navigation task the corresponding numbers were 6 out of 12 higher, 4 out of 12 equal and 2 out of 12 lower than the reference system. The average estimations are shown in Table 1.

System	VCC/Talkamatic		Reference system	
Task	Tele	Navi	Tele	Navi
Mean	7.15	7.33	6.38	6.67
SD	1.46	1.72	1.26	1.72

Table 1: Average estimates of driving ability

4.3. Efficiency

The subjects were together able to perform a total of 126 tasks with the VCC/Talkamatic system, compared to 89 completed tasks using the the reference system. The VCC/Talkamatic system was thus on average 42 % more efficient than the the reference system. The total of completed telephone tasks using VCC/Talkamatic was 84, to be compared to 64 the reference system tasks (31 % more efficient). The total of navigation tasks (42 and 25) shows a difference of 68 % in favour of VCC/Talkamatic.

	Tele	Navi	Total
VCC/Talkamatic	84	42	126
Reference system	64	25	89

Table 2: Total number of completed tasks

Looking at the individual subjects, 11 subjects out of 13 (85 %) were able to perform more telephone tasks using the VCC/Talkamatic system than when using the the reference system. The corresponding number for the navigation task was 8 out of 12, or 67 %.

System	VCC/Talkamatic		Reference system	
Task	Tele	Navi	Tele	Navi
Mean	6.46	3.50	4.92	2.08
SD	1.98	2.02	0.76	1.08

Table 3: Average number of completed tasks

The mean time for task completion are shown in Table 4. The subject that completed the highest number of telephone tasks using VCC/Talkamatic used on average 18 seconds per task. The corresponding number for the reference system is 30 seconds per task. For the navigation tasks, the result is 23 seconds for VCC/Talkamatic and 60 seconds for the reference system. It can be noted that the subject that completed the most tasks using VCC/Talkamatic was also among the fastest the reference system users – this is true for both the telephone and navigation tasks.

	Tele	Navi	Total
VCC/Talkamatic	28 s.	51 s.	36 s.
Reference system	37 s	86 s.	51 s.

Table 4: Mean time in seconds for task completion

4.4. Comments and Observations

11 out of 13 subjects said that they would prefer a system like VCC/Talkamatic to be installed in their cars. One preferred the reference system, and one had no preference. It can be noted that the persons that did not prefer VCC/Talkamatic also had the highest age. Some users, including the ones that preferred the reference system, argued that it was easier to understand what you were supposed to say at every given moment using the reference system.

We saw some unexpected usage of VCC/Talkamatic. One subject consistently used the following interaction pattern:

S: What would you like to do?
U: (streetname) (housenumber) in (city)
S: Okay. Enter an address.
U: (streetname) (housenumber) in (city)
S: Would you like to get more information on the destinaton or add the destination to the itinerary?
U: Add.

With the third utterance ("Okej. Ange address" in Swedish) the system wants to verify that the recognized user intention of entering an address is correct. The expected user behaviour is to answer "yes" or "no" or to accept the intention recognition with silence.

However, this subject – and probably others – instead interpreted the intention recognition as a request to the user to enter an address. The non-optimal behaviour of repeating the entered address worked as an alternative strategy, and this behaviour was repeated in this user's subsequent interactions. To increase optimality, the realization of the system's intention recognition has been changed after the evaluation to "Okej, du vill ange en adress." ("Okay, you want to enter an address.").

When asked to motivate which system they preferred, nine of the subjects that preferred VCC/Talkamatic mentioned that it was less tedious since it was possible to speak more freely. They thought it was convenient to be able to give all information in one utterance instead of answering one question at a time. They also found it less tedious to use the VCC/Talkamatic system since they could interrupt it if they already knew what to say, instead of waiting for the system to finish the prompt as needed when using the reference system. This was also the main reason why the subjects found the VCC/Talkamatic system more efficient than the reference system.

The reference system was considered most distracting, mainly because the user has to wait for a beep signal that indicates that it is the users turn to speak. There was a time difference between the end of the system prompt and the beep signal, which made it easy to forget to wait for the signal. The main complaints on the VCC/Talkamatic system when it comes to distraction was that it sometimes misheard, and one subject thought that the VCC/Talkamatic system had an unnatural voice.

Results in terms of user preference, estimated driving ability and efficiency show a strong preference for the VCC/Talkamatic system. However, the LCT results are partly contradictory. The telephone task is less distracting with the VCC/Talkamatic system, whereas the reference system is somewhat less distracting for the navigation task.

5. Conclusions

First of all it should be noted that the number of subjects in this evaluation is very low – 13 for one experiment and 12 for the other. This means that the results from this evaluation should be interpreted as interesting tendencies and as a starting point for further investigations.

There are several possible explanations to the differences in the LCT results. One explanation could be that the navigation dialogue model of the reference system is less distracting than the VCC/Talkamatic one, but since the dialogue for entering destinations is very similar to the telephone dialogue system application this seems unlikely. We believe that for most people the VCC/Talkamatic dialogue model is the easiest to use, and that the data for the telephone application supports this. The reason for VCC/Talkamatic's higher LCT values for the navigation application could be that this application is not as well engineered as the telephone application (for instance the feedback issue mentioned above).

Our investigation clearly show better results for the Talkamatic dialogue model than for the reference system with the exceptions mentioned above. One could also argue that the distraction level is of less importance since the efficiency of VCC/Talkamatic is significantly higher. A somewhat elevated distraction level is compensated by the fact that the interaction takes place during a very short period of time.

As different users prefer different dialogue models it would be preferable to be able to offer to the user the possibility to customize the dialogue interaction model.

As some users commented that it was hard to understand what you were supposed to say at a given time using VCC/Talkamatic, it would be interesting to perform user studies of long term use of the system. One can expect that the users' understanding of VCC/Talkamatic's system would increase with time and use, leading to increased efficiency and lower distraction levels.

6. Acknowledgements

The evaluation is funded by Innovationsbron Väst (Fokus Verifiering). Implementation of the dialogue system applications were funded by Volvo Car Corporation (VCC). We would like to thank Peter Ljunglöf (Talkamatic and University of Gothenburg) for invaluable help during implementation and testing.

7. References

[1] S. Larsson, "Issue-based dialogue management," Ph.D. dissertation, Göteborg University, 2002.

[2] M. Lindkvist, "Navigon 7310: LyhÃ¶rd toppmodell." 2009, http://m3.idg.se/2.1022/1.232690/navigon-7310-lyhord-toppmodell.

[3] I. Walke-Chomjakov, "Navigon 8410 im test," 2009, http://www.pcwelt.de/start/mobility_handy_pda/navigation/tests/2103472/navigon_8410_im_test/.

[4] J. Villing and S. Larsson, "Dico - a multimodal in-vehicle dialogue system," in *Proceedings of the 10th workshop on the semantics and pragmatics of dialogue*, D. Schlangen and R. Fernandez, Eds., 2006.

[5] D. Traum and S. Larsson, "The information state approach to dialogue management," in *Current and New Directions in Discourse & Dialogue*, R. Smith and J. Kuppevelt, Eds. Kluwer Academic Publishers, 2003.

[6] D. Hjelm, A.-C. Forslund, S. Larsson, and A. Wallentin, "DJ GoDiS: Multimodal menu-based dialogue in an asychronous isu system," in *Proceedings of the ninth workshop on the semantics and pragmatics of dialogue*, C. Gardent and B. Gaiffe, Eds., 2005.

[7] S. Larsson, R. Cooper, and S. Ericsson, "menu2dialog," in *Proceedings of the 2nd IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, 2001, pp. 41–45.

[8] S. Mattes, "The Lane-Change-Task as a Tool for Driver Distraction Evaluation," in *Proceedings of the IGfA*, 2003.