Topic Switching Strategies for Spoken Dialogue Systems

Tobias Heinroth, Savina Koleva, and Wolfgang Minker

Institute of Information Technology, Ulm University, 89081 Ulm, Germany

prename.surname@uni-ulm.de

Abstract

One of the most important challenges researchers are facing within the field of Spoken Dialogue Systems is that life is neither domain dependent nor driven by a single task. In the recent past various methods to handle multiple tasks or topics in parallel within spoken human-human and human-computer dialogues have been investigated. In this paper we compare several task switching approaches such as discourse markers and task recovery methods. The aim of our study is to reveal which strategies users prefer regarding metrics such as efficiency, friendliness, and reliability. Furthermore we investigate how the different strategies influence the cognitive capacity of the subjects. The dialogues used for the study have been implemented utilising the OwlSpeak Spoken Dialogue Manager, which applies ontologies as dialogue models that can be dynamically combined during runtime.

Index Terms: spoken dialogue strategies, multitasking, HCI

10.21437/Interspeech.2011-545

1. Introduction

Compared to developers of telephony-based or in-car Spoken Dialogue Systems (SDS) the developers of SDS for real-life are confronted with various sources of complexity. One of the main sources of complexity is the multi-domain and thus multi-topic nature of real-life processes. If the dialogue system domain is not clearly defined it is nearly impossible to collect a corpus or to define valid rules to control the dialogue flow. Within the framework of the EU-funded project ATRACO [1] we have developed a model-based Spoken Dialogue Manager (SDM) called OwlSpeak. It provides a spoken interface to an existing Intelligent Environment (IE) in real-life situations [2]. The evaluation of OwlSpeak revealed several questions that arise from the specific domain-independent situation:

- How much contextual information must be taken into account in order to allow for a meaningful dialogue between the user and the system?
- How can the information be modelled in order to be computer readable?
- How can new information be easily assimilable into a probably ongoing dialogue?
- How can users switch to a different topic or task when using a spoken interface?
- How can the SDS switch to a different topic while the user is probably already involved in a spoken dialogue?

Answering all these questions is a challenging task and we have already tried to take some of them into account during the development phase of OwlSpeak. By utilising ontologies as dialogue models we have found a way to easily add new and remove old dialogue models from the knowledgebase and thus always provide *valid* spoken dialogues. We refer to *valid* with respect to the current situation of the user within a specific (intelligent) environment and a specific context. Furthermore we have implemented a task selection algorithm based on priorities that decides which task to be activated or deactivated [3]. The SDM we have developed technically solves most of the questions listed above. However, to deliver a technical solution to a specific problem and to behave in a human-like manner are two different challenges. Especially when it comes to switching between independent tasks or topics during an ongoing conversation, no adequate solutions seem to be available so far. In this paper we focus on the task or topic switching issue for SDS and compare the available approaches. In Section 2 we present previous work that has been conducted on the task switching topic so far. In Section 3, we describe the motivation of our work and provide a detailed look on the advantages and disadvantages of the established task switching strategies. Section 4 presents our experimental setup followed by the results in Section 5 and corresponding conclusions in Section 6.

2. Related Work

Several approaches to task or topic switching in spoken human-human and human-computer dialogues have been investigated. In this section we present the most common strategies that guided the implementation of the dialogues that have been used for our evaluation. Flanke et al. propose mechanisms how to successfully handle interruptions during spoken dialogues in order to reduce their disruptive effects [4]. Furthermore these mechanisms should improve the human performance during a multitasking situation. These interruption techniques have been applied to an SDS deployed for managing military logistics tasks. The standard manner to switch between various tasks is to interrupt an ongoing task and to insert new tasks that might be more urgent. Such system behaviour may negatively affect the human information management capacity. To avoid those distracting effects of dialogue interruption the authors set the main goal for their approach to furthering efficient task resuming.

In their approach Flanke et al. divide the user interface design for handling interruptions into three phases: the pre-interruption, the mid-interruption and the post-interruption phase. The former phase disposes the user for switching from the main task to the interrupting task while providing warnings. These cues can support the user to distinguish the main task from the interrupting one. In their case study for this purpose the authors utilised a different voice for the interrupting task. The mid-interruption phase is focused on the user's transition to the interrupting task and the user's ability to maintain situational awareness of background tasks. To implement the mid-interruption the authors rely on McFarlane's taxonomy of human interruption [5]. McFarlane defines four primary methods for coordinating human interruption: (1) immediate, (2) negotiated, (3) mediate and (4) scheduled interruption. Which of these interruption solutions should be applied to the application depends on the relative importance of the alerting task related

Copyright © 2011 ISCA

to the current task. If the interrupting task is critically important compared to the ongoing task the interruption occurs immediately. If it is less important it occurs case-scheduled. In all other situations the interruption is negotiated on a per-case basis as this was measured to be the most promising strategy. The post-interruption phase provides task recovery to the original task that was interrupted. In their test application a user request provided information about the previous task. To test their approach to efficient interruption support the described mechanisms were implemented into the Listen, Communicate, Show (LCS) system [6] in which mixed-initiative spoken dialogue interaction has been integrated.

Human-human Multitasking Dialogues (MTD) characterised by speakers who switch from a current task to an alerting task have been studied in [7]. In this work Yang et al. investigated guidelines for building human-computer spoken dialogue interfaces that support multitasking. They concentrated on finding definitions when and how to switch between tasks and found out that the dialogue partners usually use so-called "discourse markers" and prosody cues to signal task switching. Furthermore the subjects tried to switch at less disruptive places. To foster the understanding of human-human MTDs, the authors collected the Heeman's MTD corpus [8] in which dialogue partner pairs perform overlapping verbal tasks. To collect this corpus the following dialogue situation has been defined: A poker game forms the main task. The users are playing together and communicate over a speech interface. The game interface is displayed on a screen. The (probably interrupting) real-time task is formed by a second game where the players should find out whether the opponent has a certain picture on the bottom of his display. Two flashing bars have been used to alert the players. The authors analysed the collected data in order to figure out how the subjects switch between tasks. They found out that the players strive to switch to an alerting task at a less disruptive point during the ongoing task namely at the end of a round or at the end of the game.

Furthermore the authors examined how the users signalised the task switching. The players often used discourse markers such as "oh" and "wait" while switching from the ongoing task to the interrupting, real-time task. The prosody of the phrase "Do you have" during the current task and by task switching was studied as well. For this purpose the pitch, the duration and the energy was measured. Relying on the test results Yang et al. established that higher pitch is related to task switching but the energy and the duration are merely irrelevant. In addition it was shown that the pitch correlates to the situation of switching: the higher the pitch the more disruptive the interruption is. Finally machine learning experiments have been conducted to determine if the described cues are reliable for discriminating task switching. For this purpose a decision tree classifier was used. The results of the experiment showed that discourse context and normalized pitch are useful features for task switching recognition. In the next section we show how the presented related work guided the design of the dialogues that have been used for our evaluation.

3. Motivation

The way interruptions are handled influences the cognitive resources and the information management capacity of humans. In order to improve the user performance during a multitasking situation we focus on studying and comparing different methods of interruptions. These methods can be automatically applied to the SDS by the SDM depending on the actual dialogue situation, the urgency of the dialogue turns, or the user characteristics. On the one hand we have compared the immediate form of coordinating task breaks and switches compared to a negotiated approach in relation to the human acceptance and cognition. On the other hand we have examined different attention cues that could proof usefulness for the dialogue participants and compared the results with a baseline setup.

Speaker	Utterance
System:	In which city in America do you want to stay?
User:	Chicago.
System:	*Sorry, may I interrupt you for a moment?*
User:	Yes.
System:	The washing machine is done.
User:	OK.
System:	How long do you want to stay?

Table 1: An exemplary dialogue applying negotiated task control to the short reminder.

For this purpose we have conceptualized four different strategies for the switching of tasks based on the previous work presented in Section 2 and on our own hypotheses. We have applied them to an exemplary dialogue scenario where three interrupting tasks cross over a main task one after another. During the main task the user is supposed to book a hotel and a flight for a fictive travel. In between a short notification (see Table 1), a long reminder with detailed additional information (see Table 2), and furthermore a complex subtask (see Table 3) interrupt the main dialogue. The objective of this conversional situation is to investigate the correlation between the user's ability of maintaining situational awareness and a preferred form of task interruption.

Speaker	Utterance
System:	Which country do you want to travel to?
User:	America.
System:	*Ah!* Your friend Mario has his birthday on Thursday. You should buy him a present.
System:	*So,* in which American city do you want to stay?

Table 2: An exemplary dialogue applying discourse markers (in italics) to a detailed long reminder.

Three out of four strategies integrated the interruption immediately into the main dialogue. Our approach here is focused on mechanisms providing efficient task switches such as attention markers and task resuming after the interruptions have occurred. As a baseline we have implemented the most common way of task switching: an unassisted immediate topic shift. The system immediately initiates topic shifts regardless of any cognitive limitations of the human dialogue partner. Neither warning signals for switching between tasks nor context recovery after interruption has been provided. In the following we refer to this strategy as *baseline*.

To introduce a more sophisticated strategy, we use discourse markers combined with prosody cues for signalling task switching. With conversation cues such as "wait", "oh" and "well" we intent to catch the user's attention to the upcoming interruption. The higher pitch and volume at the beginning of the interrupting task are ought to support the situational awareness to more than one task. The discourse markers "so" and "further" are used to help the user to resume the main task. An exemplary dialogue applying this strategy, called *discourse* is shown in Table 2. A further strategy we propose is to employ full sentences initialising topic shifts instead of short discourse

cues in order to produce a more natural dialogue flow and to increase the timespan the user has for task switching detection and realisation. Complex task recovery to the primary task that has been interrupted is provided by the system as well. The dialogue snippet in Table 3 demonstrates the implementation of this more elaborated option to switch between tasks that is further referred to as *explanation* strategy.

Speaker	Utterance
System:	On which date do you want to return?
User:	On the 24th of March.
System:	*Just for a moment something else.* You are supposed to meet Sandra for dinner on the 10th of March. Do you want to make a reservation now?
User:	Yes.
System:	You can choose between "San Marco", "Enchilada" and "Chinese World".
User:	I prefer "Enchilada".
...	
System:	*Well, let's continue with the flight booking now.* You want to flight back from New York on the 24th of March, right?
User:	Yes. That's correct.

Table 3: An exemplary dialogue applying the *explanation* strategy to the subtask.

The last strategy we have investigated is called *negotiation*. It supports negotiated task interruption in the form of queries such as "May I interrupt you?". Thus the system allows the user to have full control of the task delivery. In case of a positive confirmation the system switches to the interrupting task. If the user denies, the main dialogue is continued. Table 1 shows the mechanism of the negotiated task switching control. The three described strategies give a competitive edge compared to the *baseline* in relation to the human cognitive resources and situational awareness. However, they have some drawbacks as well. The precise explanation of each topic shift may be boring and even frustrating to the user. The ambiguities of the discourse markers "so" and "than" could lead to confusion. By negotiation some critically important tasks may be rejected. In order to prove the efficiency and the reliability of the presented strategies we have conducted a study to test the approaches during a real user-system interaction with an OwlSpeak-based SDS. In the following section we show our experimental setup before we present the results of the evaluation in Section 5.

4. Experimental Setup

In order to provide a realistic evaluation setup we refrained from using a Wizard-of-Oz layout but used an up-to-date German language SDS instead. Since our main aim was to investigate the drawbacks and benefits of the presented task switching strategies we wanted to deploy a system with a performance (regarding recognition rate etc.) that is comparable to established systems. Figure 1 shows an architectural overview on the SDS that utilises OwlSpeak as core SDM component. The subjects used a SIP phone running on a laptop providing microphone and speakers to connect to a SIP gateway that grants access to a VoiceXML browser. For the experimental setup we used Voxeo Prophecy as the SIP gateway and a VoiceXML browser. The Nuance Recognizer 9.1 and Nunace Vocalizer 5.1 are accessible via MRCP [9]. They perform the automatic speech recognition and the text-to-speech tasks. All dialogue descriptions (i.e., the spoken dialogue ontologies as described in [2]) are in German language since all subjects were fluent or native speakers. All in all 80 persons took part of the experiment. A majority of the subjects aged between 19 and 47 (25.5 on average) were not experienced with spoken dialogue systems but were used to computers. The 26 women and 54 men were separated into 4 groups of 20 people in order to equally cover each of the interruption strategies. The subjects' default task was to make a reservation for a hotel in New York and an according flight to the USA.

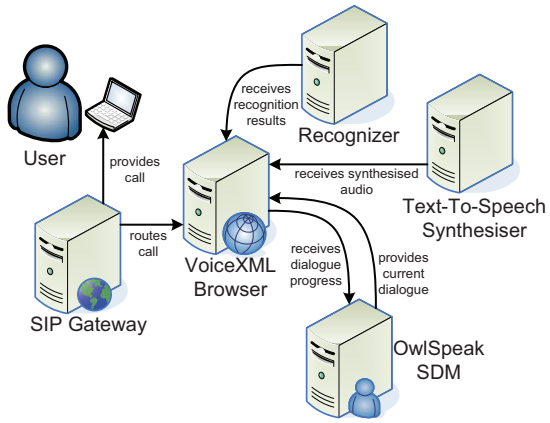

Figure 1: Architectural setup of the OwlSpeak-based SDS used for the evaluation.

No information about the interrupting tasks was given in advance in order to examine the user reaction and situational awareness similarly to real-life processes. After having successfully finished the dialogue the test persons had to fill in a questionnaire related to the main task of the dialogue as well as to the interrupting tasks. We adopted SASSI [10], a standardized method to assess whole dialogues. We have tailored the comprehensive questionnaire to 10 questions related to the following quality factors:

- EFF – Efficiency shows how efficient the system is, i.e., how well the dialogue flow can be followed by the user.
- FRI – Friendliness describes how user-friendly the system is and how pleasant it is for the subject to interact with the system.
- REL – Reliability shows how reliable the system is regarding to mistakes and understanding problems.
- IRR – Irritating presents how irritating the task switches are for the user.
- CLS – Clearly separated describes the subject's ability to separate the interrupting tasks from the main dialogue.

All user ratings are in form of checkboxes with values corresponding to an evaluation scale from 1 (very bad) to 5 (very good). In order to derive a valid estimation about the topic switching strategies the subjects furthermore had to answer several comprehension questions regarding the content of the three reminders. The results and their interpretation are detailed in the following section.

5. Results

Figure 2 shows the subjective results of the questionnaire. During all dialogues the main tasks of booking a hotel and a flight have been accomplished successfully. For each quality factor a comparison of the four task switching strategies is presented. The CLS coefficient has not been collected for the *negotiation* strategy since here the task switches occurred explicitly

and user-defined. The *discourse* and *explanation* strategies performed best regarding the EFF and FRI categories. The *explanation* and *baseline* strategies are assessed as being the most reliable ones. Furthermore, the *baseline* strategy shows the distinction of being least irritating. In contrast, *negotiation* appears to be the most irritating strategy. The *discourse* method exhibits the best results with regard to the CLS factor followed by the *explanation* strategy.

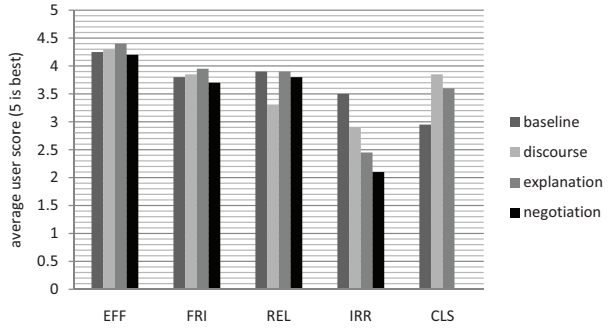

Figure 2: The five combined quality factor scores.

The main goal of the comprehension questions was to establish the ability of the subjects to memorise the details of the reminders. We have expected that depending on the attached strategy the users kept different amount of information in mind. The results are shown in Figure 3. Again the four interrupting strategies have been compared. The y-axis shows the percentage of correct user answers to the detailed questions regarding the short reminder, the long reminder and the subtask. The *explanation* and the *negotiation* strategies showed the highest values of correctly remembered details related to the short and long reminders. However, in relation to the percentage of memorised details in the subtask, the *explanation* outperforms all other strategies.

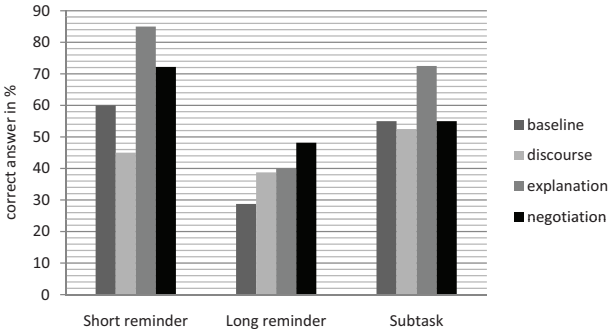

Figure 3: The average amount of information the subjects kept in mind.

Furthermore, we have examined that some subjects completely suppressed and ignored the short reminder depending on the applied strategy. On average only 2.1 out of 3 interrupting tasks were recognized when the *baseline* has been utilised and 2.4 of 3 by applying the *discourse*. Here the score of the *explanation* strategy is approximately 2.65. The maximum of 3 (i.e., all) perceived interruptions is reached by the *negotiation* strategy. The fact that the subjects who tested the *negotiation* strategy rejected at least one of the interruptions is also of interest in this context. We have observed similar user reactions regarding all strategies in case the user has been asked if the main task could be interrupted: one third of the subjects denied. The most common explanation was the desire to focus on the main task.

6. Conclusion

The main goal of our work was to find out an efficient way of switching tasks in order to improve the user acceptance and the SDS performance within real-life situations. For this purpose we have examined four different methods for handling interruptions. According the represented evaluation results all of them demonstrate strengths as well as weaknesses. The *explanation* strategy showed high scores regarding general characteristics such as efficiency and user-friendless. It seems that this strategy furthermore supports the user to memorise the reminders. Since the other strategies showed advantages such as being less irritating and providing a clearer separation form the main task a combination of the *explanation* and the *discourse* strategies seems to be interesting for future work. Depending on the importance of the interrupting task an appropriate strategy could be applied here. Furthermore the influence of the objective data that has been collected has to be investigated. For this purpose the number of interruptions the user has recognized and the influence of the total number of non-understandings could be taken into account.

7. Acknowledgement

The research leading to these results has received funding from the European Community's 7th Framework Programme (FP7/2007-2013) under grant agreement n° 216837.

8. References

[1] A. Kameas, C. Goumopoulos, H. Hagras, M. Gardner, T. Heinroth, W. Minker, A. Meliones, D. Economou, Y. Bellik, and G. Pruvost, "A Pervasive System Architecture that supports Adaptation using Agents and Ontologies," in *I-SPAN*. IEEE, Dec. 2009, pp. 148–153.

[2] T. Heinroth, D. Denich, A. Schmitt, and W. Minker, "Efficient spoken dialogue domain representation and interpretation," in *LREC'10*. ELRA, 2010.

[3] T. Heinroth and D. Denich, "Spoken Interaction within the Computed World: Evaluation of a Multitasking Adaptive Spoken Dialogue System," in *COMPSAC*. IEEE, 2011.

[4] J. Franke, J. Daniels, and D. McFarlane, "Recovering context after interruption," in *CogSci'02*, 2002, pp. 310–315.

[5] D. McFarlane, "Comparison of four primary methods for coordinating the interruption of people in human-computer interaction," *Hum.-Comput. Interact.*, vol. 17, pp. 63–139, 2002.

[6] J. Daniels, "Integrating a spoken language system with agents for operational information access," in *AAAI*, 2000, pp. 1002–1007.

[7] F. Yang, P. Heeman, and A. Kun, "Switching to real-time tasks in multi-tasking dialogue," in *COLING'08*. ACL, 2008, pp. 1025–1032.

[8] P. Heeman, F. Yang, A. Kun, and A. Shyrokov, "Conventions in human-human multi-threaded dialogues: a preliminary study," in *IUI'05*. ACM, 2005, pp. 293–295.

[9] S. Shanmugham, P. Monaco, and B. Eberman, "A media resource control protocol (mrcp) developed by cisco, nuance, and speechworks," The Internet Society, Tech. Rep., 2006.

[10] K. S. Hone and R. Graham, "Towards a tool for the subjective assessment of speech system interfaces (sassi)," *Nat. Lang. Eng.*, vol. 6, pp. 287–303, 2000.