A STRATEGY FOR MIXED-INITIATIVE DIALOGUE CONTROL

Lars Bo Larsen
Center for PersonKommunikation
Institute of Electronic Systems
Aalborg University
DK-9220 Aalborg Denmark
Email: lbl@cpk.auc.dk

ABSTRACT

This paper presents and discusses a strategy for mixed-initiative dialogue management within a home banking application. The strategy tries to utilise the guidance of system-directed dialogues, while accommodating user initiated focus shifts by the inclusion of *short-cuts* in the dialogue.

The paper reports on two experiments, one with a simulated speech recogniser (WOZ), and the second with a fully automated system. Both experiments shows that users use the possibility for *short-cuts*, even when not instructed of their existence. A tendency towards user habituation is also demonstrated.

1. INTRODUCTION

This paper describes a strategy for mixed-initiative spoken dialogue management. The strategy is outlined, and two experiments are carried out to investigate the methodology.

The experiments are carried out within the Esprit OVID[1]-project. The OVID project is concerned with the development of trial applications within automated banking services [1]. Among other things, the OVID user specifications [2] states that the customer must be in control of the interaction. However, this may not always lead to the most natural or efficient mode of communication, as humans often expect others to hold or take the initiative in conversations. Therefore, a mixed-initiative strategy is proposed.

The overall goal of the OVID project is to measure user acceptance of voice controlled home banking systems. Apart from this, the purposes of the dialogue experiments reported here are twofold:

- To test the implemented dialogue management strategy.
- To identify and delimit the application vocabulary

[1]The OVID Esprit 20717 Project consortium comprises The Royal Bank of Scotland and Barclays Bank in the U.K., Lån & Spar Bank in Denmark, CCIR Edinburgh University, U.K, CPK, Aalborg University, Denmark, Brite Voice Technology U.K., and AGORA Consult, France as coordinating partner. The work presented here is partly funded by the ESPRIT programme, and partly by CPK

The customers use unconstrained natural speech, and the speech recognition technology chosen for the task is a combination of digit string recognition and spotting of keywords and -phrases. The experiments must therefore include identification of the application vocabulary. Consequently, the experiment is carried out in two phases. First with a simulated speech recogniser (Wizard of Oz.), denoted Trial 1 or "WOZ-trial" and the second with a fully automated system, denoted Trial 2.

This paper focuses on the dialogue management issues, and reports on the experiments carried out in Trial 1 and preliminary results from Trial 2.

2. DIALOGUE SPECIFICATIONS

The overall functionality of the automated home banking application is:

- The service must first enquire the customer for his/her identification (Id) number, and subsequently a PIN code. The formats are identical to those used by Danish banks.
- The service provides the customer with a balance and an overview of the most recent transactions on his/her accounts (denoted a Mini Statement). Each customer has three accounts.
- DTMF interpretation of at least Id- and PIN codes must be available to ensure privacy.

On the basis of the overall specification a simple dialogue structure with five tasks is implemented. These are the Main task, Id- and PIN sub tasks and Balance and MiniStat subtasks.

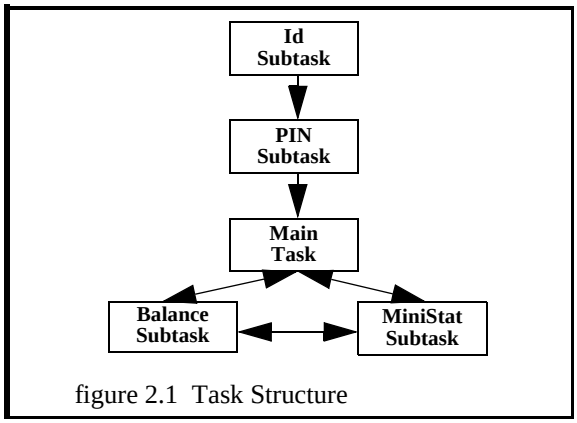

figure 2.1 Task Structure

The task structure is shown in figure 2.1. Instead of building specific DTMF subtasks, all tasks accept spoken and DTMF key pad input in parallel. This is achieved by including two sets of prompts in the dialogue, and switching between them depending on which modality the user chooses.

3. DIALOGUE INITIATIVE

The question of system directed vs. user-driven (or mixed-initiative) dialogue control strategies has been the focus of discussion for a number of years. In general, user controlled dialogues is considered preferable, as this allows the user to gain the control over the interaction, and hence achieve his goals more directly. In contrast, system-directed dialogues tend to be more rigid and menu-like.

However, this might not always be the case. A problem that might arise in user-driven dialogues, is that the user is left without a clear understanding of his options at a given point in the dialogue. This can cause frustrations or even breakdown of the communication. In [3] it is demonstrated that for a train information task, users actually preferred the system directed mode. On these grounds, and in the case of inexperienced users, the system directed mode might be preferable, while experienced users will choose to gain the initiative. Consequently, a combined system directed and user-driven dialogue (mixed-initiative) management strategy is employed in the present case. By default, the system has the initiative, and the user responds to system prompts. This works well for inexperienced users, who will be guided throughout the dialogue. However, for experienced (or impatient) users this strategy is too rigid. There clearly exists a need for the user to be able to take the initiative and directly request the desired information from the service. This is achieved by including a number of *short-cuts* in the rigid system directed dialogue structure.

By performing a short-cut, the user overrules the dialogue task structure, and forces the system to switch from one subtask to another. The shortcuts are shown as dashed (red) arcs in the simplified diagram of the overall dialogue flow structure shown in figure 3.1. The text in the boxes denotes system utterances, and the semantics of the user resonses are shown on the connecting arcs.

Two types of arcs are shown. The fully drawn lines show the system initiated transitions, and the dashed arcs depicts the user initiated transitions (*short-cuts*). This means that if the user answers all system prompts faithfully, the possible dialogue state transitions will reduce to the fully drawn lines. Incidentally, this corresponds to the dialogue that would be valid if only DTMF input was available.

Note that the user can generate a transition to “Dialogue End” from any point in the dialogue simply by hanging up.

.

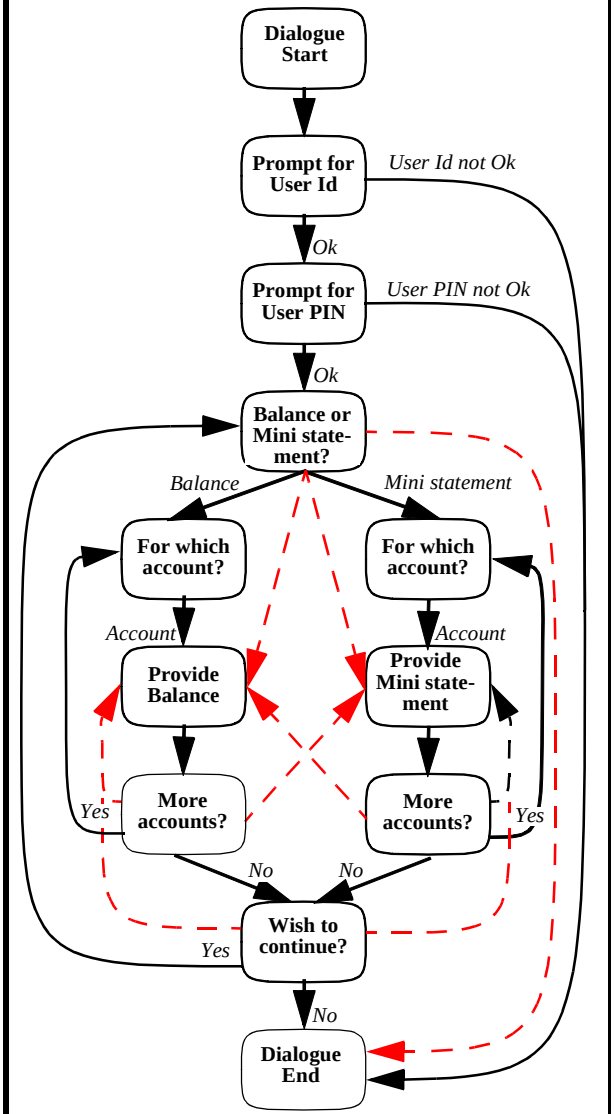

figure 3.1 Dialogue Flow

4. DIALOGUE MODEL

Further to the task structure and flow control, the dialogue model comprises a number of elements. Among the most important are:

- User Profile.
- Dialogue History.

The system builds and maintains a profile of the user’s behaviour. This includes the information already given to the user, and whether the user has been given specific instructions about the use of the system. The user profile and dialogue history is used to determine the way the system will respond to specific user input. E.g. in the case of a rejected user utterance, the system response will be dependent on previous instructions given to the user.

5. EXPERIMENTS

In the experiments, all subjects received a letter describing

the application, and defining two scenarios. Furthermore, they received a usability questionnaire to be filled out and returned.

Delibately, the subjects were not informed of the short-cuts in the dialogue. By this, it is possible to investigate to what extent users will naturally take the initiative, and also how quickly users can be termed "experienced".

In both trials each user completed two scenarios; A and B.

- Scenario A: Obtain the balance for all three accounts
- Scenario B: Obtain the balance and a mini statement for the budget account.

6. RESULTS

6.1. Results of Trial 1 (WOZ-Trial)

Trial 1 was carried out with a limited number (20) of participants. All the participants had either some connection with the university or with Tele Denmark. They had no prior knowledge of the application, although a number of them had experience with speech technology. They were not told that they participated in a simulated trial. The results of the experiments in Trial 1 is shown below in table 1 in terms of number of turns and dialogue completion

Scenario:	A	B
Nominal number of turns	7	9
Minimal[1] number of turns	5	4
Average number of turns	8.1[2]	7.8
95% confidence interval (turns)	0.5	1.0
Average duration of dialogues[3] (seconds)	105	112
95% confidence interval (duration)	6.1	13.9

table 1 Turns and dialogue completion times

[1] This includes the user hanging up immediately after the desired information has been obtained.

[2] Some users requested the information twice, or asked for repetition. Therefore the average number of turns is larger than the nominal.

[3] It should be taken into account that the Mini Statement includes full description of three postings, and hence the average completion time for scenario B is influenced by this.

times. The figures are based on 40 dialogues. The nominal number of turns is the number of turns a user would have to go through if he/she answered all system prompts without gaining the initiative at any point.

If the user gains the initiative e.g by supplying additional information, or by answering a yes/no question with a new request, he can short-cut the rigid system controlled dialogue structure. Full utilisation of this yields the minimal number of turns.

The scenarios are designed in such a way that the nominal number of turns are almost equal for both, but scenario B can be completed with less that half the nominal number.

Inspecting the average number of turns does not directly give an indication of how users perform, but a closer investigation of the transcribed dialogues shows that the subjects now are separated in two groups. This can also be observed in the confidence intervals, which have doubled for scenario B as compared to scenario A.

One group follows the system directed dialogue, whereas the other group have started utilising the short-cuts. This tendency is even more pronounced when taking the user identity and verification procedure into account. This "costs" two turns in all cases. This tendency is illustrated in figure 6.1

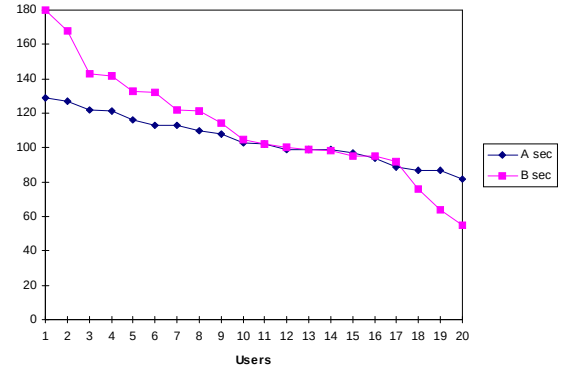

figure 6.1 Duration of scenario A and B in Trial 1

Another objective of Trial 1 was to identify the vocabulary for the word spotting speech recogniser to be used in Trial 2. A total of 20 words were found to be sufficient for the task.

6.2. Results of Trial 2

Trial 2 was carried out with 350 customers from the Danish Lån & Spar Bank. As Lån & Spar is a "Direct Bank" depending heavily on automatic services all users were experienced users of DTMF systems, but had not used a speech controlled system before. The users were selected evenly from geographic regions and age groups.

The results reported for Trial 2 are preliminary and are based on data from 80 users and a total of 176 dialogues. The users were given the same scenarios as in Trial 1, but 50% of the users were instructed to perform scenario B first. As in Trial 1, the CPK Generic Dialogue System (GDS) platform [4],[5],[6] was used to implement the dialogue. Trial 2 was carried out using the CPK SUNCAR real-time speech recogniser [7]. The Danish SpeechDat M 1000 speaker corpus [8] was used for training of the acoustic models. The dialogue model was identical to that of Trial 1.

It was not possible to identify a similar tendency for Trial 2 as shown in table 1 and figure 6.1 concerning the overall duration of the dialogues. However, the users did take the initiative throughout the dialogues, as indicated in figure 6.2

The numbers shows whether it is the users' first or second

dialogue. The figure shows clearly that more experienced

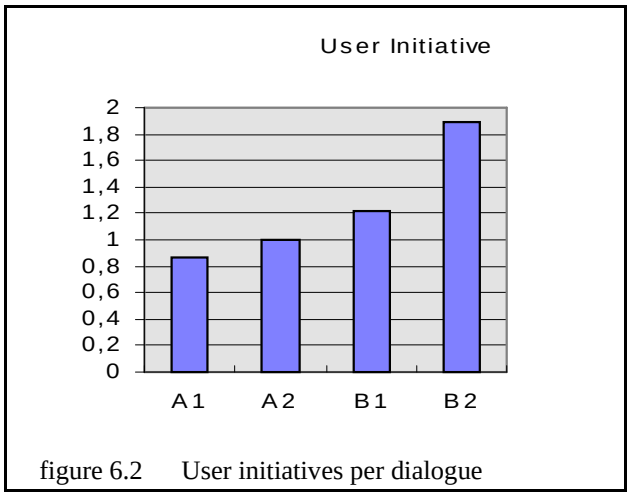

figure 6.2 User initiatives per dialogue

users tend to take the initiative more often (e.g. compare A1 to A2).

Turns	Total		Id.Num		PIN		Main		Balance		Mini St.	
Scenario	A	B	A	B	A	B	A	B	A	B	A	B
Nominal	7	9	1	1	1	1	2	3	3	2	0	2
Actual	8.0	7.2	1.3	1.3	1.3	1.2	2.0	2.0	3.2	1.7	0.2	0.9
Minimal	5	4	1	1	1	1	1	1	3	1	0	1

table 2 Nominal, Actual and Minimal number of turns per task

In table 2 and figure 6.3 it is shown how many turns the us-

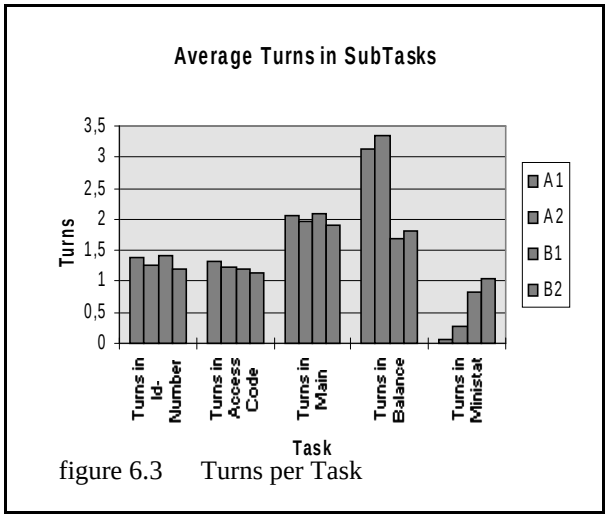

figure 6.3 Turns per Task

ers spend on average in each subtask. Note that the average number of turns in the Id- and PIN code task are very close to one. The nominal number of turns in the Balance task is three for scenario A and two for scenario B. Again, the actual figures come very close, and is even a little below the nominal number. The nominal number of turns in Ministat is zero for scenario A and one for B. The average number of turns in scenario B actually drops below one, which indicates that not all users succeed in getting the mini statement required in the scenario.

7. CONCLUSIONS

The emphasis has been put on naturalness and flexibility of the spoken input/output and the dialogue structure. The user tests indicate that this goal has been accomplished, as the users were able to start gaining the initiative and short-cut the system controlled dialogue structure without prior instructions or informations about this opportunity.

Trial 1 indicates that after only one exposure to the dialogue, some of the users have become acquainted with the dialogue structure. This indication could also be found in Trial 2. It furthermore showed that users immediately started to go beyond the limits of the system directed dialogue structure and utilising the built-in short-cuts. An even more pronounced effect can be expected when users are exposed to a larger number of dialogues.

Speech recognition error rates for digit strings and word spotting were found to be approximately 10%. This seemed sufficient to ensure user acceptance.

The preceding sections have shown that the strategy of maintaining a system directed dialogue on the surface and then provide *shortcuts* for more experienced users has proved successful.

Trial 2 was, in fact, a usability trial, with the aim to investigate to what extent customers are prepared to accept voice controlled access to their bank accounts. The preliminary results strongly indicates that this is the case.The tested dialogue was very small, containing only a few sub tasks, so the next step will evidently be to expand the dialogue to cover a larger number of tasks, and a more complex task structure.

8. ACKNOWLEDGEMENTS

The author wishes to thank Børge Lindberg, Bo Bai and Jesper Olesen from CPK for their help in getting this experiment done.

9. REFERENCES

[1] Lars Bo Larsen, “Voice controlled home banking - objectives and experiences of the Esprit OVID project” in Proc IVTTA-96, New Jersey September 1996 (ieee 96TH8178).

[2] ESPRIT 20171 Project OVID - Trial application of Voice Processing in Automated Telephone Banking Services: “User Requirements (Deliverable D1)” CCIR Edinburgh March 1996.

[3] Louis Boves et al. “Localization and Field Test of a Dutch Train Time Table Information System” oral presentation and in Proc. IVTTA-96, New Jersey September 1996 (ieee 96TH8178).

[4] Anders Bækgaard, “The GDS PLatform” Report 10 from the Danish Dialogue Project, CPK, Aalborg University 1996.“

[5] L.B. Larsen A. Baekgaard, “Rapid Prototyping of a Dialogue System using a Generic Dialogue Development Platform” in Proc. ICSLP-94, Yokohama 1994.

[6] L.B. Larsen, “Development and evaluation of a spoken dialogue for a telephone based transaction system”, in proc. EUROSPEECH-95, Madrid 1995.

[7] Børge Lindberg, Jan Kristiansen, “Realtime Speech Recognition within Spoken Dialogue Systems”, Report 8 from the Danish Dialogue Project, CPK, Aalborg University 1996.

[8] H. Christensen, B. Lindberg and P. Steingrimson, “Documentation of the Danish SpeechDat (M) Database”, CPK, Aalborg University, Aalborg 1996.