Annals of DAAAM for 2011 & Proceedings of the 22nd International DAAAM Symposium, Volume 22, No. 1, ISSN 1726-9679
ISBN 978-3-901509-83-4, Editor B. Katalinic, Published by DAAAM International, Vienna, Austria, EU, 2011
Make Harmony between Technology and Nature, and Your Mind will Fly Free as a Bird
Annals & Proceedings of DAAAM International 2011

DESIGNING A PERSONAL ASSISTANCE APPLICATION USING WIZARD OF OZ METHODOLOGY

NESTOROVIC, T[omas] & STRUHAR, V[aclav]

Abstract: *This paper concerns a Wizard of Oz (WoZ) method applied to a PDA applications domain speech interface design. Below, we define rules for selecting common PDA relevant functions, design them using SUEDE tool, and conduct WoZ experiments. Preliminary results show approximately 80 percent of user satisfaction in average.*

Key words: *wizard of oz, user interface, artificial intelligences*

1. INTRODUCTION

Today's graphical user interfaces (GUI) are the most developed and preferred means for exchanging information between a human and a machine (Oviatt, 2002). They can give a comprehensive non-linear overview of application's current state and can provide sufficient interactive environment. However, especially for new users they tend to be slow to deal with, incomprehensible, and unnatural. Hence during the past years, the attention focused on developing alternative *speech user interfaces* (SUI). The benefit of them is naturalness (so that users can express demands using speech instead of traversing through a system of menus), sometimes having also an indirect impact on the hosting environment (e.g. supporting a traffic safety in the case of a spoken car navigation system).

2. DESIGNING A PDA SPEECH INTERFACE

In this section, we describe our PDA domain with respect to its decomposition and decisions made while designing it. We employed SUEDE (Klemmer *et al*., 2000) as a prototyping tool. To get involved with SUIs, we need to take into account two important aspects – the speech recognition component performance and the structure of the speech interface itself.

Current *speech recognizers* can deal with just a fraction of language needed to cover a specific domain (Oviatt, 2002). An usual approach to this issue is to manage a dialogue in a system- or mixed-initiative way (Gustafson, 2002). However, if we want to allow more flexible interaction and yet avoid eventual state explosion, we need to constrain the set of functions available to the user through a SUI (let then each of them be managed in a mixed-initiative manner). PDAs can accomplish a variety of different tasks (e.g. messages management, contact list management and phoning, or Internet browsing). However, not each of them is necessary to have a spoken alternative, more particularly, we consider only those functions which hold at least one of the following conditions:

- The user uses it often during her/his PDA interaction (e.g. telephone functions or appointments management).
- It facilitates the manipulation with the PDA (e.g. a counter example may be speech controlled Internet browsing – although (Christian *et al*., 2000) shows a limited feasibility, we do not think this interaction makes an assistant more user-friendly).
- Its results are easy to present using speech (otherwise speech interaction is useless; e.g. all system utterances must follow the “3+2 rule” on the number of results an average user is capable to memorize (Yankelovich, 1996)).

Although more PDA function meet these three criteria, we constrained our design to messaging and appointment services only as we thing these two are of sufficient complexity to show a preliminary SUI approach.

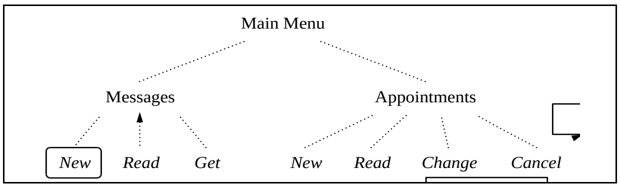


Fig. 1. PDA applications speech interface structure

Let us now focus on the *structure* of the SUI itself. We follow a hierarchical system of menus in which nodes represent particular functions we picked up above (Fig.1). More specifically, this hierarchy matches to an extend the original graphical structure menus that users are already familiar with. With this decision, we take into account two observations:

- Users learn an unknown environment quicker if they can find something familiar in it. (In the past, there was a discussion on whether the design of SUIs should or should not adopt GUI design habits. Oviatt (2002) argues that for a speech interface to be as effective as human language it should avoid any GUI guidelines. However, this is in contrast with the following observation.)
- People currently do not expect complex behaviour from a system and like consistency (Gustafson, 2002).

Thus in our design, a user may traverse through the menus (following the GUI-like decomposition) or say directly the name of the desired function (once implemented as a real system, the best interaction style will be determined by the underlying dialogue manager (Nestorovič, 2010) as a results of evaluating the previous interaction with the user).

Finally, let us make a note on reducing system's initiative. To facilitate orientation in the menus, the user is always told where s/he currently is (e.g. W_3 or W_6 in Table 2). Furthermore, we suggest prompting this information in the Main Menu *only when entering it* (i.e. being idle in the Main Menu does not cause the system to push the user to say more as this can be annoying, e.g. while driving a car).

3. WoZ EXPERIMENT

During our experiments with the PDA domain, there were N = 8 users interacting with a wizard. To meet the minimum requirement of the WoZ methodology,[1] users and the wizard were located in different rooms, interacting with each other through microphone (user only) and headsets (user and wizard). Each session has been logged and voice recorded to allow further processing (i.e. ASR and Parser modules training).

[1] The minimum requirement is that the user must not reveal that s/he is faked by a wizard. Additional requirements are user's willingness to participate in the experiments (e.g. as Klemmer et al. (2000) point out “The user should not buy the first washing machine offered by the system but attempt to negotiate its parameters.”) and a wizard convincingly pretending a system.

Each user was given a four task scenario to accomplish (Fig.2). At the end of each tasks, users were obliged to write down information they obtained from the wizard. When finished with their tasks, users were given a short survey focused on telling us what they reckon about the interaction. We wanted to know their overall feeling (Question 1) broken down into narrowed questions on the way of interaction quality (Questions 2 and 3), speed of the interaction (Question 4), opinion on system prompts clarity (Question 5), and attitude to eventual future usage (Question 6). Responses ranged over five predefined values (representing that the user does not agree, rather does not agree, does not know, rather agrees, and agrees) mapped to integers 1…5 (with 5 representing full agreement). Their responses can be found in Table 1.

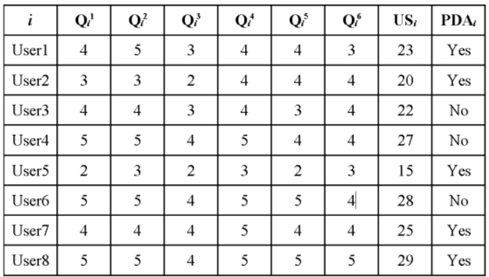
Task 1. You sent a message to **Jane**. Check that she already answered, and **reply** to her eventually. (Note: You will not write any message, just say Jane's name once requested.)

Task 2. **Peter** asked you to **reschedule** an **appointment** with him. **Cancel** all appointments that collide with it.

Task 3. Plan an **appointment** with **Charles**. The appointment should last **half an hour** and should take place any time between **17:00 and 19:00** at **your flat**. **Postpone** any appointment that collides with it.

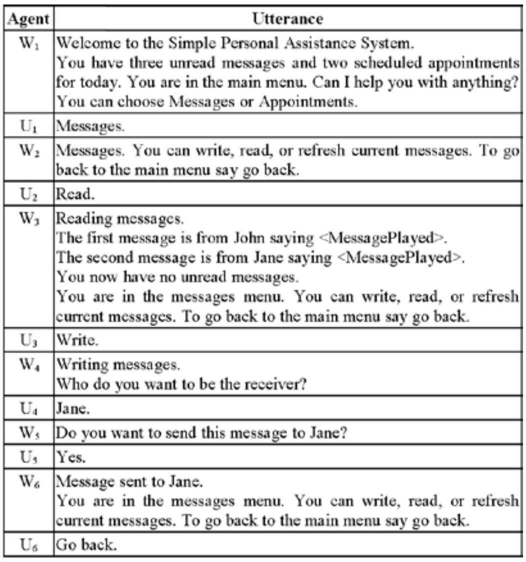
Task 4. You are going to pick up a car from the repair shop between **11:00 and 14:00**. Plan it as an **appointment**. If there are any collisions, solve them your way.

Fig. 2. Task scenario



i	Q_i^1	Q_i^2	Q_i^3	Q_i^4	Q_i^5	Q_i^6	US_i	PDA_i
User1	4	5	3	4	4	3	23	Yes
User2	3	3	2	4	4	4	20	Yes
User3	4	4	3	4	3	4	22	No
User4	5	5	4	5	4	4	27	No
User5	2	3	2	3	2	3	15	Yes
User6	5	5	4	5	5	4	28	No
User7	4	4	4	5	4	4	25	Yes
User8	5	5	4	5	5	5	29	Yes

Tab. 1. Questionnaire results. $Q_i^{\{1…6\}}$ = *i*-th user's answer to Question {1…6}, US_i = *i*-th user's satisfaction = $\sum_{(j)} Q_i^j$, PDA = *i*-th user can use a PDA.



Agent	Utterance
W_1	Welcome to the Simple Personal Assistance System. You have three unread messages and two scheduled appointments for today. You are in the main menu. Can I help you with anything? You can choose Messages or Appointments.
U_1	Messages.
W_2	Messages. You can write, read, or refresh current messages. To go back to the main menu say go back.
U_2	Read.
W_3	Reading messages. The first message is from John saying <MessagePlayed>. The second message is from Jane saying <MessagePlayed>. You now have no unread messages. You are in the messages menu. You can write, read, or refresh current messages. To go back to the main menu say go back.
U_3	Write.
W_4	Writing messages. Who do you want to be the receiver?
U_4	Jane.
W_5	Do you want to send this message to Jane?
U_5	Yes.
W_6	Message sent to Jane. You are in the messages menu. You can write, read, or refresh current messages. To go back to the main menu say go back.
U_6	Go back.

Tab. 2. Task 1 dialogue between a user (U) and wizard (W)

Each particular session took about 15 minutes in average, depending on user's previous familiarity with PDAs. As it can be seen from Table 2, most users found the system's structure (with respect to menus and function flow) rather user-friendly (Questions 2 and 3) and were able to accomplish all given tasks. During the interaction, the wizard generated random errors (with probability of 15 %) to emulate the imperfectness of an ASR module. Furthermore, users rated the average pace of information provided by the system as “neutral” (in terms of the scale from 1 to 5). However, the pace of information is hard to model in state-based SUEDE due to the danger of state explosion and eventual inconsistencies in the resulting dialogue flow model. For example in our current experimental proposal, the system reads all messages at once (utterances W_3 in Table 2), whereas more preferred way would be first offering a summary of new messages followed by letting the user choose which one to read completely (i.e. an approach that follows Yankelovich's (1996) incremental method for exchanging information in SUIs). However, we did not included this feature at this experimental level due to the state explosion mentioned above and we let this issue for the implementation when filling an empty multimodal dialogue manager with domain specific information (Nestorovič, 2010).

Finally, we can compute an overall user satisfaction by summing all users' relative satisfactions (Table 1), i.e. US = ⅛ $\sum_{(i)} US_i$ / 30, where 30 is the maximal satisfaction. This yields an overall user satisfaction of 80 % which can be interpreted as the ratio of meeting user's requirements. Thus, there is a space for improvements (e.g. the subsequent detailing of received messages mentioned above, however, as we already pointed out, this will be sorted out at the implementation level as a dialogue manager behavioural specification).

4. CONCLUSIONS

This paper conceived a partial design of a PDA domain SUI by a WoZ method. However, because recognizing a wide portion of a language yields low recognition scores and errors, we needed to involve heuristic rules to pick up functions worth a speech alternative. We designed a subset of these functions using SUEDE and conducted experiments to verify the applicability of the resulting SUI. Our results show that the overall user rating is approximately 80 %. However, as SUEDE can simulate just a rigid way of dialogue management, additional changes will be required when implementing the design (e.g. subsequent detailing of received messages or more sophisticated logic in appointment rescheduling, both of which would result in state explosion if involved in SUEDE).

6. ACKNOWLEDGEMENTS

This paper has been supported from grant UWB Grant SGS-2010-028 Advanced Computer and Information Systems.

5. REFERENCES

Christian, K., Kules, B., Shneiderman, B., and Youssef, A. (2000). A Comparison of Voice Controlled and Mouse Controlled Web Browsing, In: *Conference on Assistive Technologies*, pp. 72-79, ACM Press

Gustafson, J. (2002). *Developing Multimodal Spoken Dialogue Systems – Empirical Studies of Spoken Human-Computer Interaction*, Universitetsservice US AB, Stockholm

Klemmer, S., Sinha, A., Chen, J., Landay, J., Aboobaker, N., and Wang, A. (2000). SUEDE: A Wizard of Oz Prototyping Tool for Speech User Interfaces, In: *User Interface Software and Technology*, pp. 1-10, ACM Press

Nestorovič, T. (2010). Frame-based Dialogue Management Automated Error Recovery Approach, In: *Australasian Joint Conference on Artificial Intelligence*, pp. 32–41

Oviatt, S. (2002). Multimodal Interfaces, In: *The human-computer interaction handbook*, pp. 286-304, Lawrence Earlbaum, New Jersey

Yankelovich, N. (1996). How Do Users Know What to Say? *Interactions*, vol. 3, pp. 32–43