

Conducting a Wizard of Oz Experiment on a Ubiquitous Computing System Doorman

Kaj Mäkelä, Esa-Pekka Salonen, Markku Turunen, Jaakko Hakulinen and Roope Raisamo

Computer-Human Interaction Unit, Department of Computer and Information Sciences

FIN-33014 University of Tampere, Finland

+39 555 321 7654

{kaj, eps, mturunen, jh, rr}@cs.uta.fi

1 Abstract

A problem in developing and testing ubiquitous computing systems is the fact that they are environments and cannot be tested in conventional laboratory settings. Here we have addressed the problem of testing such an environment by applying the Wizard of Oz method to a ubiquitous computing system called Doorman. Doorman uses spoken language input and multimodal speech output (speech synthesis combined with pointing gestures) to control the access of incoming visitors and staff members to our office premises and to guide the visitors to find the people or the room they are seeking. The experiment was conducted by simulating speech recognition with a human wizard operating the otherwise fully working system. The user-initiated dialogue strategy was mostly successful, but did not meet the requirements in some cases. We also found that guiding visitors using multimodal spoken output was not successful and should be redesigned.

1.1 Keywords

Wizard of Oz, ubiquitous computing, spoken language dialogue, speech user interfaces, evaluation

2 Introduction

The speech-based ubiquitous computing system called Doorman [4] is located and being developed in the premises of TAUCHI, the Computer-Human Interaction Unit in the University of Tampere. Doorman opens the door to the visitors and the staff members, and guides the visitors in the premises of TAUCHI.

This paper describes a Wizard of Oz (WOZ) experiment conducted during the iterative development of the Doorman system. The Wizard of Oz testing is an experimental user interface evaluation method in which the user of the system is made to believe that he or she is interacting with a fully implemented system though the whole or a part of the interaction of the system is controlled by a human, “a wizard”, or several of them. The interaction is logged and/or recorded for further analysis. Wizard of Oz tests are useful in supporting the design process and evaluating the interface [2]. The method has been used to test natural language dialogue systems (for example, [6]) and multimodal systems (for example, [5]). Here we apply this

method to speech-based, multimodal ubiquitous computing applications.

Human-computer communication has been found to differ from human-human communication [1]. Therefore, to gather reliable information about human-computer communication it is important to observe the human behaviour in a situation in which humans believe to be interacting with a real computer system. It is important that the user thinks he or she is communicating with the system, not a human, as noted for example by Dahlbäck *et al.* [3].

3 Description of the system

Doorman is used to help the members of TAUCHI and their visitors in their everyday life. The Doorman system identifies the staff members and recognises the visitor’s target of the visit. The door is opened automatically to identified staff members and to visitors if the target of their visit is recognised. Doorman guides visitors in TAUCHI premises to the person or the room they are seeking and gives the staff members information about organisational messages, e-mails, instant messages, phone calls and about visitors who have been seeking them.

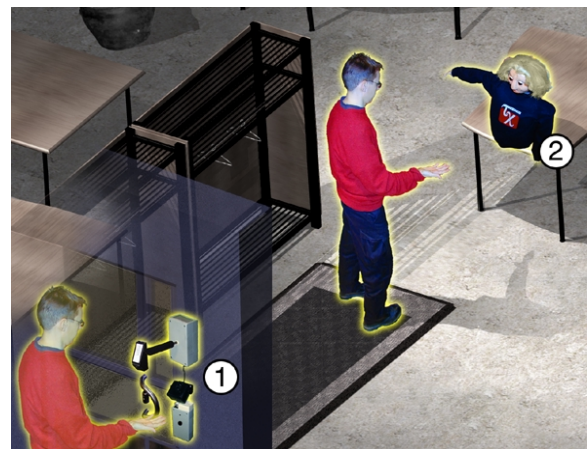


Figure 1. The basic setup of the Doorman system. Outside the door is a microphone, a loudspeaker, a doorbell button and a led light (1). In the lobby there is a guide robot containing a loudspeaker and a microphone (2).

Doorman uses spoken language to communicate with the users. Speech recognition is used for input and multimodal speech synthesis as the output method. In addition to speech, output contains pointing hand gestures done by the guiding robot. The target of the visitor's visit or the identity of the staff member is recognised from the user's initial phrase by using speech recognition. In addition, speaker recognition will be used. The Doorman system has some resemblance to Office Monitor [9].

The system gathers information about the situation at the door with a microphone, a doorbell switch and a door micro-switch. The output of the system is presented to the user with synthesised speech via speakers installed at the door and lobby. The spoken guidance is given by an anthropomorphic robot pointing to the direction the user should go to find the target. Guidance is formed dynamically and spoken to the user using speech synthesis. The system is illustrated in Figure 1.

The Doorman system is based on a distributed software architecture called Jaspis [7]. Jaspis is a Java-based adaptive speech user interface architecture that has been developed by Speech-based and Pervasive Interaction Group at the University of Tampere. It was originally designed for spoken dialogue applications and has been expanded to include features that support ubiquitous computing applications and Wizard of Oz studies.

The dialogue control model of the system is implemented as a finite state machine. Each system state (usually one turn in a conversation) is implemented as an independent dialogue agent. No modifications were needed in the dialogue components in the WOZ experiment. Therefore, the system was fully functional except for the fact that a wizard simulated the speech inputs using the WOZ user interface. The wizard was not able to control the behaviour of the system in other ways.

The functions of the current system can be divided into the following stages: (1a) recognition of the staff member or (1b) recognition of the target of the visit, (2) opening the door and (3a) greeting the staff member or (3b) guiding the visitor to the target of the visit; the target can be a person or a room.

4 Example dialogues

In the current dialogue model the system prompts are formed to guide the user to answer briefly. The visitors are assumed to push the doorbell, after which they are asked and expected to tell the target of their visit. It is assumed that the staff members will not push the doorbell but say a greeting and their name straight away on the door. The following examples are translated from Finnish.

4.1 Example dialogue of a visitor

[Visitor] Pushes the doorbell button.

[Doorman] *"What is the name of the person or the room you are looking for?"* // target request

[Visitor] *"John Doe"* or *"Usability lab"* // target name

[Doorman] *"Good morning. I will open the door for you"* // the door is opened.

[Doorman] (continues inside the premises) *"Good morning. The person you are looking for is in room 444. To find there turn left, go seven meters forward, turn right, go five meters forward, turn left. The room of John Doe that you were looking for is to the right seven meters from you."*

4.3 Example dialogue of a staff member

[Staff member] *"John Doe here, hello. Could you open the door?"* // name and greeting

[Doorman] *"Good morning, John Doe. I will open the door for you."* // the door is opened

[Doorman] (continues inside the premises) *"Good morning, John Doe."*

4.4 Error handling

[Visitor] Pushes the doorbell button.

[Doorman] *"What is the name of the person or the room you are looking for?"* // target request

[Visitor] speaks a non-recognisable name or a non-recognisable room.

[Doorman] *"I am sorry, I did not understand. Say the name of the person or the room you are looking for."*

[Visitor] speaks a non-recognisable name or a non-recognisable room.

[Doorman] *"Say the name of the person or the room you are looking for."* // repeated three times.

[Doorman] *"I am sorry, I cannot open the door. Use the key or push the doorbell button within 15 seconds to ring the doorbell."*

5 Description of the experiment

The aim of the study was to test and analyse the spoken language and multimodal dialogue model designed in the system before constructing the actual speech recognisers. In addition, we wanted to recognise the actual behaviour of the user and the problems occurring in the following situations: the user understanding the question asked by the system using synthesised speech, the visitor responding to the question and stating the target of the visit, the staff member declaring his or her identity, the behaviour of the user while entering the premises and the visitor understanding and responding to the guidance given by the system.

The collected data was to be used to evaluate how well the current dialogue model works, and to give insight on how to improve it when the system is further developed.

The test was conducted in five days, one of which was used for training and pilot testing the setup. The test was run for approximately 4 hours per day, on a quite varying basis. The test sessions lasted from 45 minutes to 1.5 hours each time. Two persons were required: one was acting as a wizard and one was gathering permissions from visitors for recording. The wizard group members changed roles many

times during the test, because the wizard task was quite demanding and their alertness would go down in a long-lasting session.

The visitors were informed of the system with posters outside the door. The poster explained that the system uses speech recognition and that it can be bypassed by pushing the doorbell button three times in a sequence.

The staff members were informed of the system via e-mail before the experiment was started. However, the nature of the system was not revealed. They were asked a permission to gather voice samples and log information to create personal profiles. These voice samples will later be used to improve speech recognition accuracy and to recognise the users from their voice. They were not given any specific instructions on using the system, only to greet and introduce themselves to the system. This was done to see which way they would behave without detailed instructions.

The door that the Doorman opens to the users can also be opened with traditional keys or with electronic key cards. Key owners were asked but not required to use the Doorman system. This request may have had an effect on the naturalness of the test setting. However, this was necessary to ensure gaining data on how the staff members used the system.

6 The wizard's tool

The Wizard of Oz experiment was conducted by substituting the speech recognition module of the Doorman system with a control application used manually by the wizard. The user interface of the control application is shown in Figure 2. The speech of the users was recorded and all the system tasks and sensor inputs were logged to be thoroughly analysed later.

The wizard listened constantly to the voice input gathered with the microphone installed at the door. When a user spoke the wizard interpreted the input to the system.

The system was otherwise fully functional and the wizard was unable to alter the behaviour of the system in any other means except giving simulated speech inputs. Furthermore, the simulated speech inputs were always either legal inputs or indicated recognition errors. Only one kinds of recognition errors (not recognized) were simulated mainly because we wanted to simplify the WOZ experiment and the work of wizards.

Technically the WOZ tool is a Java applet (WOZlet), which is connected to the underlying Jaspis system using socket connections. Jaspis easily enables the Wizard of Oz testing because of its modular manager, agent and evaluator based structure [8]. Each of the components of the Jaspis architecture can be replaced with a wizard. Therefore, implementing the Wizard of Oz version of the system did not require extensive effort. We also added several new features in the Jaspis architecture to support WOZ experiments, such as data logging tools.

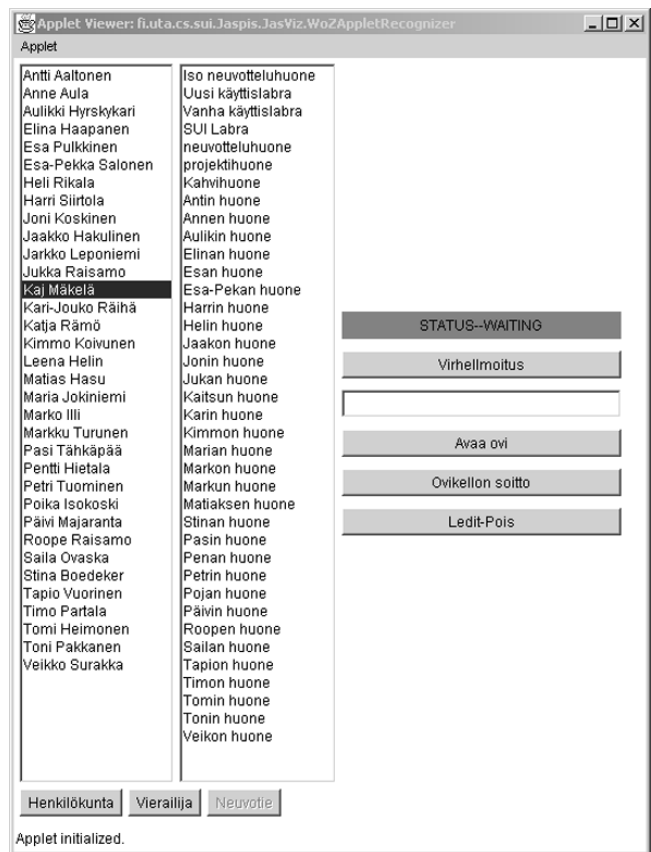


Figure 2. The user interface of the control application. The two lists contain all the staff members (on the left) and the rooms (on the right) in TAUCHI. There are buttons used to inform the system that the object chosen from the list is the identity of the staff member (“Henkilökunta”) or that it is the target of the visit (“Vierailija” or “Neuvo tie”). There is also a button for error messages (“Virheilmoitus”). For exceptional cases there is a possibility to open the door manually (button “Avaa ovi”) or call for help by ringing the doorbell (button “Ovikellon soitto”).

7 Findings

During the experiment, the system was used in 74 situations, of which 22 were visitors, and 52 were staff members. Fifteen visitors (68 %) used the system so that they responded to the first prompt in the way that was expected. One user succeeded in the second try and one in the third try. Three users bypassed the system by pushing the doorbell three times. Thus, in most of the visitor cases (77 %, 17 persons) the system successively served the visitors.

7.1 Visitors

During the tests the system failed to serve the user in two cases. Although this is a small amount, it should be seriously considered. In one case, a user did not speak to the system at all and did not use the possibility to bypass the system. In the other case, the person the visitor was searching for was not a staff member. In this case, the user got frustrated when his speech was not recognized, and

used his cellular phone to contact the person he was supposed to meet. In this case, the user tried three times and stopped after this. The system was implemented so that after the fourth try it will go to a state where the user can push the doorbell button to actually ring the bell inside. Based on the findings the system should go to the manual mode after three or already after two failed recognitions and ring the bell in a normal way. Furthermore, we should have other ways to handle situations of this kind.

The other problem in this case was that the system is able to guide the visitor only to the members of the TAUCHI staff listed in the system. However, also students and members of other organisations use the usability laboratory that is situated in the premises. A visitor arriving to the test is usually searching for the person who is conducting the usability test. Thus, the system will be unable to recognise the person if she or he is not a staff member. It is also possible that the visitor arriving to the test does not even know the name of the person conducting the test or the name of the room the test is held in. It is very difficult to detect automatically when the users speak names not known to the system. This out-of-vocabulary detection is one of the weak points of the current speech recognition systems. It is also out of the question to list every possible option (staff members, rooms) to the user using spoken prompts.

During the experiment, the visitors never used the names of the rooms when stating the target of their visit. Instead, they always asked for persons. It is noteworthy that the visitors were asked to give a name of either a person or a place.

7.2 Staff members

Some of the staff members started the dialogue by greeting the system and waited for the system to respond before stating their name. This may have been because they wanted to make sure that they are heard and to make sure that the connection with the system is established. In these cases, the users behaved much like in human-human interaction and expected the system to behave similarly. They maybe expected the system to be able to work on a more sophisticated level of conversation than it really was. This is one of the learned communication patterns which people use in their daily communication. Even if people know the limits of the system they use their learnt skills. One reason may also be that in the introductory e-mail message we gave instructions for staff members to greet the system.

The staff members who used the system on a more regular basis changed their way of speaking to the system based on their former experiences. They learned from their mistakes and adapted their interaction to the system. However, since some people stopped the use of the system after few attempts, it is possible that they did not want to adapt to the system.

7.3 System responses

The delays in the system were found irritating and the users having a key or a key card often chose to use one instead of waiting for the system to react. This was partly because the

user was not sure when the system was processing the input due to lack of any indicator or feedback showing the current state of the process to the user. The human wizard, detailed event logging and limited hardware resources caused some delays. Also the length of the sentences spoken by the system annoyed the users. Especially the users using the system on a regular basis were irritated, because the speech delayed their entrance and the prompts remained unchanged.

7.4 Guidance

The guide robot was often passed without listening to the instructions. The reason was mostly that the visitor already knew where he or she was going. The other reason was that the robot and the guidance were not given attention. The robot had also been on location long before the system was functioning and the users may have not expected it to act. Sometimes, the timing caused that the robot started guiding too late and the visitor had already passed it.

The guidance given by the guide robot was also found too long, slow and unclear. The timing, the length of the prompt, the speech rate and volume altogether caused that most of the visitors ignored the guidance. The length of the guidance also made it hard to remember the guided route.

7.5 Miscellaneous findings

In informal interviews the users told that the speech synthesis was unclear and therefore sometimes hard to understand. However, the synthesizer used was quite intelligible, although not very pleasant. Due to the nature of the system it is not possible to use recorded system prompts in all situations. One solution may be mixing natural speech and synthetic speech, but the most important information is given in dynamic system outputs.

On the basis of this experiment it seems that the users will choose the easiest and quickest way to handle the task and if the system is not able to serve them properly they will choose an alternative method. However, in the current stage the system did not offer any added value to most of the users. In the future versions the system will contain other services, such as information about e-mails, general information and other services. This may change usage patterns dramatically.

8 Lessons learnt

As a part of the iterative design process, improvements are going to be made in the Doorman system on the basis of the test. To make the interaction with the system quicker, and to have more people use the system, the level of the system initiative is going to be increased by implementing sensors to recognise the presence of the user on the door.

The results show that some additional consideration should be given in the form of the speech output and the delays of the system. Long and static system utterances are going to be shortened, varied and made more informal and natural, for example, by utilising the user profiles of the staff members and by varying greetings. The speech rate will be increased to make the speech more pleasant and fluent.

The order of speech and other system actions is going to be rearranged to shorten the delays, and the flexibility of dialogue is going to be increased by giving the user a possibility to interrupt the synthetic speech. The user should be given feedback to confirm that the system is reacting. This is going to be done by using short utterances like, for example, 'let's see' and non-speech sounds.

In some problematic situations, when the user did not use assumed words, the dialogue model failed. This happened, for example, when the user said something beyond the vocabulary of the system. The error loop after a failed speech recognition attempt was found to be too long and it is going to be shortened. Each turn of the loop should also adapt to the situation and give more detailed instructions to the user. In the dead-end situation, the system should be able to offer alternative solutions such as to contact human operators.

We should alter the dialogue model to better support system-initiative dialogues. For example, many staff members did not take initiative by speaking, but instead acted like visitors and pushed the doorbell button.

The guide robot and the guidance messages were unsuccessful and we should really focus on these issues. Especially we need to change the guidance messages to use landmarks instead of direct walking instructions to make the guidance shorter and more understandable. The appearance and position of the guide robot should be altered to make it more noticeable. The association between the speech outside the door and the guide robot needs to be evident. We will also make several robots to work co-operatively in order to divide the dialogue into more compact parts.

The test also brought up a need for changes in the architecture level: at the moment, the Jaspis architecture does not support simultaneous dialogues. The changes that allow this feature are going to be implemented in the near future.

9 Conclusions

This paper described a Wizard of Oz experiment that was used to find out the way the users interact with the Doorman system. Although informal, the results are useful in giving guidelines to the following iterations of the design process.

The most important findings that help in the further development of the system were related to the structure of the dialogue, the need for system-initiated dialogue and better error handling, and the way the guidance is arranged. The guide robot needs to speak in a more natural way and to be easily recognisable.

The experiment gave us valuable information on how to improve the system. It also showed that setting up a Wizard of Oz experiment is straightforward by using the tools provided by the Jaspis framework. Based on our experience we recommend using the Wizard of Oz method during the iterative development of ubiquitous computing systems.

11 Acknowledgments

This work was funded by the Academy of Finland (project 163356).

REFERENCES

1. Baber, C. and Stammers, R.B., Is it natural to talk to computers: an experiment using the Wizard of Oz technique. In E.D. Megaw, (Ed.), *Contemporary Ergonomics 1989*. Taylor & Francis, 1989.
2. Bernsen, N.O., Dybkjær, H., and Dybkjær, L., Designing Interactive Speech Systems. *Springer-Verlag, London*. 1999, 127-160.
3. Dahlbäck, N., Jönsson, A., and Ahrenberg, L., Wizard of Oz Studies – Why and How. *Proceedings of the 1993 International Workshop on Intelligent User Interfaces (IUI'93)*, ACM Press, 1993, 193-200.
4. The Doorman system
http://www.cs.uta.fi/hci/spi/Ovimies/index_en.html
5. Gustafson, J., Bell, L., Beskow, J., Boye, J., Carlson, R., Edlund, J., Granstrom, B., House, D., and Wiren, M., AdApt-A Multimodal Conversational Dialogue System in an Apartment Domain. *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.
6. Johnsen, M., Svendsen, T., Amble, T., Holter, T., and Harborg, E., TABOR – A Norwegian Spoken Dialogue System for Bus Travel Information. *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.
7. Turunen, M., and Hakulinen, J., Jaspis - A Framework for Multilingual Adaptive Speech Applications. *Proceedings of 6th International Conference of Spoken Language Processing (ICSLP 2000)*, Beijing, China, 2000.
8. Turunen, M. and Hakulinen, J., Agent-based Adaptive Interaction and Dialogue Management Architecture for Speech Applications. In Text, Speech and Dialogue. *Proceedings of the Fourth International Conference TSD 2001*, 357-364
9. Yankelovich, N., and McLain, C.D., Office Monitor. *CHI 1996 Conference Companion*, ACM Press, 1996 173-174.