IMPROVING PROTOTYPE CONSISTENCE FOR WIZARD-OF-OZ SIMULATIONS AND EVALUATIONS

Andol X Li and John V H Bonner
Department of Informatics, School of Computing & Engineering
University of Huddersfield, Queensgate, Huddersfield HD1 3DH, UK

ABSTRACT

In this paper we describe a study of the technique of Wizard of Oz (WoZ) in perspectives of maintaining the simulation consistence for domestic communication evaluations. This study employs evaluation methodologies such like heuristic evaluations and video analysis to compare and review the consistence during WoZ experiments. In specific, we address this consistence study through rapid prototype designs and iterative evaluations. Series of prototypes were developed based on our domestic communication platform and these prototype features were compared across iterative evaluations. The consequences of our prototype development and experimental evaluations show that the role of the wizard is vital to the effectiveness of WoZ studies, and variables which relate to the wizard's operations such like personal preferences and interaction paces need to be carefully addressed to improve the consistence of system simulations. The conclusions of this study are based on comparisons of different control panel designs for the wizard, which indicate that simulation consistence can be augmented through interfaces with specific flexibility and customisations.

Keywords wizard of oz, simulation consistence, domestic communication, interface designs

1 THE INTRODUCTION

Emerging technologies have been heavily weighted since last decades, towards adoptions of these technologies into diverse scenarios such like the domestic environments. To date these technologies have permeated into every aspect of our life, however identifying proper ways of integrating technologies are still challengeable to majority of researchers [Taylor et al. 2007]. One of reasons we propose is the lack of efficient methods which can have our insights into futuristic technologies and reality integrations. To explore the patterns of technology adoptions we shift our attentions to a method of Wizard of Oz. The method of WoZ can extend our sights beyond constrains of underlying technology levels, and our prototype capability can be thoroughly improved to build up wide range of systems. A typical wizard of oz experiment is like this [Dahlback et al. 1993; Dow et al. 2005; Molin 2004]. A wizard who is usually played by the experimenter sits apart invisibly and intercepts the communications between the system and subjects. Therefore the system can export various simulations which can be used as components of system for prototype developments and technology evaluations. However at this stage the WoZ is insufficient in perspectives of exporting stable and consistent simulations by the wizard. Since the role of the wizard is dynamic and dependent on the experimenter's performances, this uncertainty needs to be addressed before we adopt this method to gain creditable understandings towards futuristic technologies and interactions in the home.

This paper starts with an analysis of current state of WoZ, which leads to several requirements of simulation consistence. Following that the paper exemplifies an example of trial experiment in our laboratory to gain preliminary understandings of system consistence. After that the paper indicates our further studies in details, and this section is completed with several relevant experimental cases. Finally the paper is summarised with a discussion of experiment faults and some inspirations for future developments, meanwhile based on the evidences we have gained from the experiments the paper proposes a conclusion that providing the wizard a flexible and organisable interfaces can hugely improve the wizard's performance consistence as well as simulation effectiveness.

2 THE STUDY BACKGROUNDS

The range of new interactive technologies is like augmented reality technologies [Haller et al. 2007], ubiquitous technologies [Weiser 1991], smart technologies [Chan et al. 2009] and so forth. Identifying proper forms to integrate various types of technology is still challengeable due to conventional methods of prototype developments and evaluations constrain our understandings towards testing and applying technologies in practical scenarios. Unlike the historical evolutions of buildings, we have gained implicit understandings of what people expect and how these technologies can be used to satisfy the requirements [Rodden and Benford 2003]. To date researchers have conducted many studies to investigate the patterns of pushing technologies into real scenarios. In 1996 Venkatesh examined the technologies used in the home such like the cordless telephones and answering machines, and his considerations lead to the gaps between what technology can do and what people want to do with the technology [Venkatesh 1996]. In 2002 Blythe adopted ethnography methodology and found some gendered technology uses [Blythe and Monk 2002]. And Hutchinson adopted a new method of technology probes to collect experimental data about the use of the technology in a real-world setting [Hutchinson et al. 2003]. A common thing among these studies is that they were conducted in practical scenarios which were processed with prototype designs and evaluations. These examples show the importance of how iterative studies with practical prototypes and evaluations can benefit researchers to understand the technology adoptions.

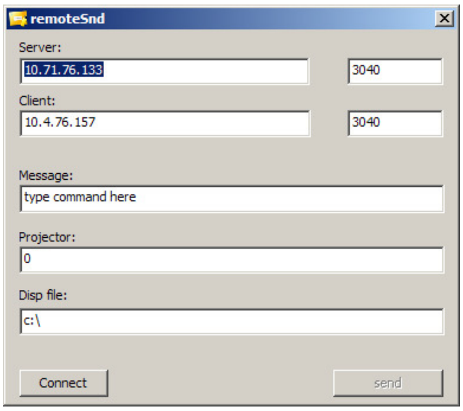
However there are two factors which constrain us to execute such iterative studies. The first is the constrain of current technology levels, which hinder our imaginary ideas evolving into practical products such like a robot with human-like intelligence. The other constrain is the low effectiveness of building experimental prototypes. For instance the speech recognition technology has been developed for decades and now it has been preliminary for commercial use, however researchers may still pay huge costs to design a system which may not fully understand human speeches. These two constrains bring us to consider WoZ to avoid getting locked into particular limitations. WoZ is conceived to build prototypes beyond underlying technologies and it can also evolve with the systems iteratively [Dow, MacIntyre, Lee, Oezbek, Bolter and Gandy 2005]. Through the wizard's simulations designers can gain vast space to provide subjects possibilities for interactions within different scenarios. Typically WoZ-based interactions involve a wizard operator who plays some roles such like controller and overseer, and the simulations the wizard is asked to provide are like sensors and intelligence. Combining these roles harmonically and exporting stable and consistent simulations are much challengeable to the wizard particularly with dominating interfaces which contains dynamic manipulation elements such like text boxes and scroll bars. Operations relevant to these dynamic elements may generate ambiguous information to subjects, and this reduces the creditability of WoZ simulations since the subjects may be aware the system not acting as ways of intelligent computers with consistent and predictable exports. Many studies from other researchers have adopted the WoZ to provide simulations but only limited in playing simple components (e.g. the body gesture recognising) rather than synthetic systems such like natural language-based speech interactions [Bradley et al. 2009; Hoysniemi et al. 2004; Minkyung and Mark 2008]. In some speech dialogue research we have noticed that some wizard relevant variables have been identified such like the tones of speech and the pauses of phrases [Fraser and Gilbert 1991], yet the solutions to address these variables are still underground. To step further our study emphasises on how designs should be achieved to assist the wizard export stable and consistent simulations.

The site we are about to locate our studies is the domestic scenario. Since we have observed that some technologies have successfully integrated into the home such like the telephones and digital televisions, we are convinced to push more complex technologies into this environment to evaluate the interactions. Meanwhile daily routines of householders have been exposed to more and more computations as part of our life experiences, which provides huge space for our application designs. And the convergent social relationships in the home such like friendships and intimacies also enable us to utilise the technologies at most in perspectives of social influences.

3 PRELIMINARY EXPERIMENTS

Our study began from a preliminary experiment which tried to clarify some intuitive design faults. Through a walkthrough of the prototype, we generally discovered some drawbacks of the prototype which might affect future WoZ experiments. The prototype interface for the wizard was sketched as below (Figure 1). Through this panel the wizard was enabled to connect the remote client with IP address in local networks and sent direct messages to perform text-based natural dialogues as well as to specify different content displays in the laboratory. The panel gave the wizard vast flexibilities to provide simulations, from a speech recognised multimedia control system to a natural language-based human-computer dialogue with instant replies. Specifically the message sent by the wizard can have many forms to respond the subjects' activities such like full sentences and a single exclamatory mark. In this prototype every message was sent after clicking the send button and the messages were received immediately by the subjects as long as the connection was on.

The walkthrough involved an expert instructor who was asked to experience a full list of prototype functions and give comments. Discussions were held after the walkthrough and critical experiment data was collected through this procedure. Since the instructor was around the laboratory and reviewed all system components, he had straightforward feelings to immerge himself in the simulated interactive environment. Hence the comments we gained from this phrase were relatively subjective, yet in perspectives of presenting subjects comfortable interactions this was still effective. The main procedures of the walkthrough were like this. The instructor was in the laboratory and he used speeches to control content displays around the space such like the floor and the coffee table (Figure 2). These speeches were based on natural languages but combined with few computer-understandable phrases such like 'system start', 'new user' and so forth. In this walkthrough the system accepted speeches to display videos, audios, PDF files and web pages on various domestic surfaces, besides there laid a cube on the table to specify content displays through projectors. The instructor repeated these operations iteratively and compared how smooth the system interactions were and where was able to be improved potentially.



Server: 10.71.76.133 3040

Client: 10.4.76.157 3040

Message: type command here

Projector: 0

Disp file: c:\

Figure 1. The panel for the wizard

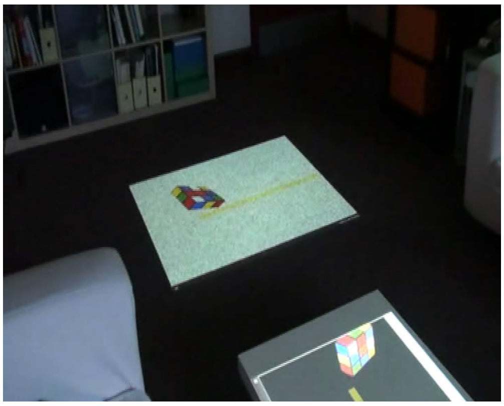


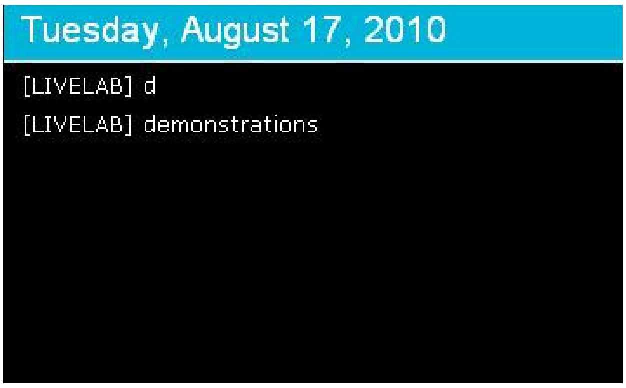
Figure 2. Displays in the laboratory

The walkthrough of the trial run went informally due to our goal at this stage aimed to reduce noticeable prototype flaws, and through this we achieved the goal. The feedbacks from the instructor included those of the massive components of applications, the challenges of handling simultaneous displays and the lack of speech interfaces. In this prototype we integrated three main modules for the cube manipulations, the drag and drop manipulations and speech manipulations. Multiple applications coexisting in one system did not support the instructor multimodal interactions, rather the instructor could not remember all these manipulations and used to focus on one component of the speech manipulations. Similarly the multiple displays around the room provided few clues for the instructor to deal with these simultaneously. Flashing contents across the environment with dynamic information made the instructor distractive. And the last feedback of the lack of speech interface indicated us the necessity of dialogues which take the roles of avatar to communicate with people.

4 CASES OF STUDIES

Based on the implications we have drawn from the trial run of prototype, we improved the system as a specific application which integrated a calendar for domestic communications. Meanwhile we concentrated on one surface of the coffee table which was frequently used in daily lives. And the new prototype was also designed with speech dialogue interfaces which indicated the subjects that they were interacting with a computer system. Due to the requirements of calendar simulations and manipulations, the wizard's simulation panel was added new features relevant to date selections and appointment managements (Figure 5). Through the projection on the table we also integrated the calendar seamlessly into the table which seemed as one of the table's functions (Figure 4). The calendar allowed the subjects to check appointments of a day and add new appointments as well as to delete appointments. All these operations of adding and deleting were achieved through speech commands such like 'add event' and 'delete event'. Since the main representative interface of the system was projected on the coffee table, the subjects could sit in the sofa and give commands in means of talking to the table.

At this stage we adopted complex methods to observe and record experiment activities such like the video recording and informal interviews. The main method of this experiment was the heuristic evaluation which employed one or more reviewers, preferably experts, to comment system performances according to criteria [Jeffries and Desurvire 1992; Nielsen and Molich 1990]. One advantage of using this method in our experiments was that this could be conducted based on WoZ simulations. Hence the experiments involved two dimensions of evaluations in terms of the wizard's simulations and the subjects' (the expert reviewers) activities. Corresponding to the subjects' comments we were able to analyse which types of operations caused the inconsistence throughout the wizard's simulations. Comparing with direct evaluations of the wizard's operation consistence, the analysis of the subjects' interactions showed us more clues to locate inconsistent operations which probably confused subjects. The collections and analysis of subjects' activity data had many sophisticated methods which revealed intuitive understandings. In this study we took the experiment video into manual analysis frame by frame to capture subjects' tiny emotional expressions. Emotion observations gave us sufficient clues of how the subjects' attitudes changed from one scene to another such like hesitations, confusions and satisfactions.



Tuesday, August 17, 2010

[LIVELAB] d

[LIVELAB] demonstrations

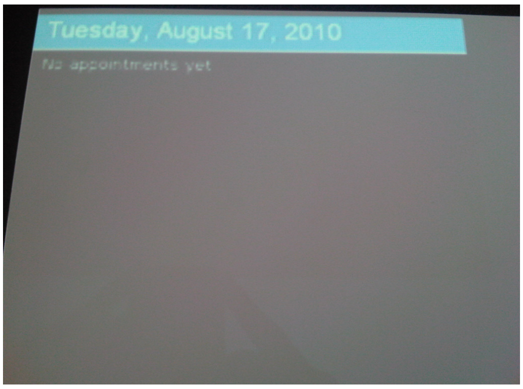


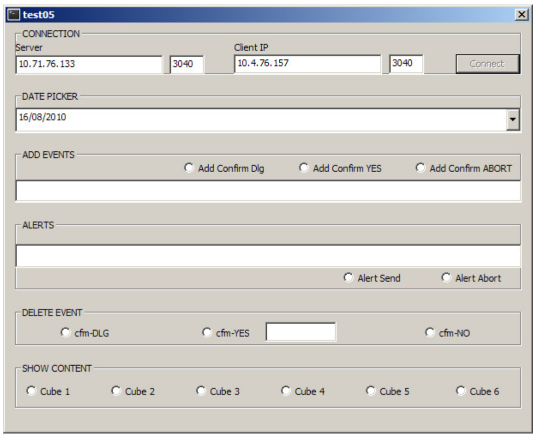
Figure 3. The interface of domestic calendar

Figure 4. The calendar on the table

Iterative system designs were conducted during the experiments, after each experiment analysis the design was correspondingly changed according to the implications. The first prototype related to the domestic calendar with interfaces of the event list and the wizard's control panel (Figure 3 and Figure 5). One subject was invited to our experiment to use this new developed intelligent system. The subject had sound backgrounds in interactions but he knew nothing about this system until the experiment was introduced. Tasks the subject was asked to perform were 1) to pick a day and check events of the day, 2) to add an event relevant to the subject itself and 3) to delete any event of a day. All task conductions required speech commands and these commands had a specific format. For instance, to show the events of a day the speech format was like 'today', to add events it was like 'today, add event, have a meeting with John, confirm' and the format for event deleting was like 'today, delete event, go swimming, confirm'. After the tasks we also encouraged the subject to manipulate the system by natural language instead of computer-understandable commands. Through this we observed which communication form was more natural to have speech dialogue interactions with people and whether the wizard can afford this type of requirements. Each round of experiment took about five minutes, exclusive of interviews and discussions.

After the first run of experiment we noticed the slow responses of the wizard which caused the subject's anxieties for system progresses, while some rapid responses also existed occasionally. This phenomenon reveals risks of exposing the wizard due to the inconsistent responses with unpredictable speeds. Our video analysis of the wizard's operations shows that the fast responses came from easy accessibility. For instance when the wizard's mouse was right over the data picker then the date can be selected immediately, when mouse was far away then the movements wasted unnecessary time. These hypothesises towards causes of inconsistence lead us to one direction, which is to reconstruct wizard's control panel. To address this issue we broke the panel into several respective modules and made these modules flexible for rearrangements (Figure 6). This improvement enabled the wizard to shape a layout with most effectiveness.

The procedures of new experiments were the same as the previous. The subject was invited and asked to accomplish tasks and followed by help-yourself manipulations. The whole procedures were recorded and the video analysis followed with comparisons of previous experiments. In this experiment, we noticed that the wizard exported more stable simulations to subjects and the whole interaction went with less unpredictable clashes which convinced the subject to believe the intelligent computer system. Meanwhile through the wizard (the experimenter) we also found that the task loads were relatively reduced since the wizard can organise these small units to simulate a complexity component. The wizard can change the layout of these units and make them less time consuming, furthermore the wizard can also learn these small units rapidly due to each piece of panel contained an intuitive function with high acceptability.



test05

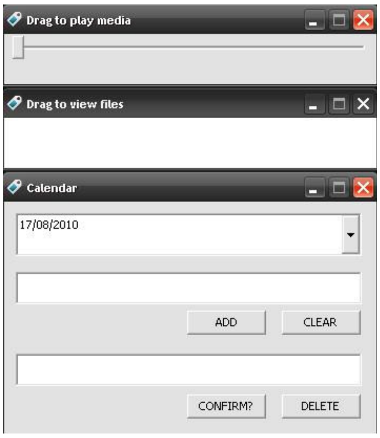
CONNECTION

Server: 10.71.76.133 3040

Client IP: 10.4.76.157 3040

DATE PICKER: 16/08/2010

Figure 5. The conventional control panel



Calendar: 17/08/2010

Figure 6. The new control panel

5 DISCUSSIONS AND CONCLUSIONS

So far we have identified the influences of the control panel which can support the wizard with efficient operations, and also we have gained some understandings towards interface designs for Wizard. Meanwhile the requirements of integrating an interactive and ambient calendar for domestic communications have been clarified. However this is still far from goals of our WoZ studies of wizard's consistent performances. In this study we have addressed designs for WoZ, yet we have considered few behaviours of wizard such like the paces of typing the messages and the degree of exporting humanities. To address these aspects, we need to develop much complicated systems with proper simulations, and we need more iterative experiments to reveal rules of exporting consistent simulations during WoZ evaluations.

Fortunately this study has provided us some clues of interface designs and how these interfaces may influence the WoZ simulations. Based on evidences we have gained, in this paper, we propose our conclusions as below. Firstly, the interfaces for the wizard require much more flexibilities than interfaces for normal applications. Usual interfaces are designed for heavy inputs into systems, in contrast interfaces designed for the wizard tend to have more outputs than inputs. The outputs require efficient organisations of manipulative elements, therefore enhancing interface flexibilities helps the wizard shape a customised environment. Secondly, the consistent performances are inadequate without proper configurations of other WoZ variables. These variables are various and dependent on systems, such like in speech systems a variable can be the speech tones while in robot systems a variable can be the velocity of movement. Taken together, the consistence of wizard's performances heavily lies on how researchers deal with the collection of variables such like the interface designs.

REFERENCES

BLYTHE, M. AND MONK, A. 2002. Notes towards an ethnography of domestic technology. In *Proceedings of the Proceedings of the 4th conference on Designing interactive systems: processes, practices, methods, and techniques*, London, England2002 ACM.

BRADLEY, J., MIVAL, O. AND BENYON, D. 2009. Wizard of Oz experiments for companions. In *Proceedings of the Proceedings of the 2009 British Computer Society Conference on Human-Computer Interaction*, Cambridge, United Kingdom2009 British Computer Society.

CHAN, M., CAMPO, E., ESTEVE, D. AND FOURNIOLS, J.-Y. 2009. Smart homes - Current features and future perspectives. *Maturitas 64*, 90-97.

DAHLBACK, N., JONSSON, A. AND AHRENBERG, L. 1993. Wizard of Oz studies: why and how. In *Proceedings of the Proceedings of the 1st international conference on Intelligent user interfaces*, Orlando, Florida, United States1993 ACM.

DOW, S., MACINTYRE, B., LEE, J., OEZBEK, C., BOLTER, J.D. AND GANDY, M. 2005. Wizard of Oz Support throughout an Iterative Design Process IEEE Educational Activities Department, 18-26.

FRASER, N.M. AND GILBERT, G.N. 1991. Simulating speech systems. *Computer Speech and Language 5*, 81-99.

HALLER, M., BILLINGHUSR, M. AND THOMAS, B. 2007. *Emerging Technologies of Augmented Reality Interfaces and Design*. IDEA GROUP PUBLISHING.

HOYSNIEMI, J., HAMALAINEN, P. AND TURKKI, L. 2004. Wizard of Oz prototyping of computer vision based action games for children. In *Proceedings of the Proceedings of the 2004 conference on Interaction design and children: building a community*, Maryland2004 ACM.

HUTCHINSON, H., MACKAY, W., WESTERLUND, B., B.BEDERSON, B., DRUIN, A., PLAISANT, C., BEAUDOUIN-LAFON, M., CONVERSY, S., EVANS, H., HANSEN, H., ROUSSEL, N., EIDERBACK, B., LINDQUIST, S. AND SUNDBLAD, Y. 2003. Techonology probes: Inspiring design for and with families. *CHI*, 17-24.

JEFFRIES, R. AND DESURVIRE, H. 1992. Usability testing vs. heuristic evaluation: was there a contest? ACM, 39-41.

MINKYUNG, L. AND MARK, B. 2008. A Wizard of Oz study for an AR multimodal interface. In *Proceedings of the Proceedings of the 10th international conference on Multimodal interfaces*, Chania, Crete, Greece2008 ACM.

MOLIN, L. 2004. Wizard-of-Oz prototyping for co-operative interaction design of graphical user interfaces. In *Proceedings of the Proceedings of the third Nordic conference on Human-computer interaction*, Tampere, Finland2004 ACM.

NIELSEN, J. AND MOLICH, R. 1990. Heuristic evaluation of user interfaces. In *Proceedings of the Proceedings of the SIGCHI conference on Human factors in computing systems: Empowering people*, Seattle, Washington, United States1990 ACM.

RODDEN, T. AND BENFORD, S. 2003. The evolution of buildings and implications for the design of ubiquitous domestic environments. In *Proceedings of the Proceedings of the SIGCHI conference on Human factors in computing systems*, Ft. Lauderdale, Florida, USA2003 ACM.

TAYLOR, A.S., HARPER, R., SWAN, L., IZADI, S., SELLEN, A. AND PERRY, M. 2007. Homes that make us smart. *Pers Ubiquit Comput.*

VENKATESH, A. 1996. Computers and other interactive technologies for the home. *Communication of the ACM 39*, 8.

WEISER, M. 1991. The computer for the 21st century. *Scientific American 265*, 94-104.