

A WIZARD OF OZ PLATFORM FOR THE STUDY OF MULTIMODAL SYSTEMS

Daniel Salber, Joëlle Coutaz

Laboratoire de Génie Informatique, IMAG
B.P. 53 X, 38041 Grenoble Cedex, France
Phone: +33 76 51 44 40, E-mail: salber@imag.fr, joelle@imag.fr

ABSTRACT

The Wizard of Oz (WOz) technique is an experimental evaluation mechanism. It allows the observation of a user operating an apparently fully functioning system whose missing services are supplemented by a hidden wizard. In the absence of generalizable theories and models for the design and evaluation of multimodal systems, the WOz technique is an appropriate approach to the identification of sound design solutions. We show how the WOz technique can be extended to the study of multimodal interfaces and we introduce the Neimo platform as an illustration of our early experience in the development of such platforms.

KEYWORDS: Multimodal interaction, Wizard of Oz, Evaluation techniques.

INTRODUCTION

Evaluation techniques and user observation provide ways to improve the design of interactive systems. Predictive evaluation techniques such as GOMS [2] or KRI [6] are based on predictions from theory, and so, they may be limited in scope when the case study (e.g., multimodal interaction) is not supported by the underlying theory. In addition, the setting and interpretation of such models are sometimes complex and as time consuming as an effective implementation. At the opposite, experimental techniques deal with real data observed from real users accomplishing real tasks, and operating a physical artefact. Artefacts may be paper scenarios, mockups, computer system prototypes, or Wizard of Oz systems. In this short paper, we investigate how the Wizard of Oz (WOz) experimental evaluation technique can be extended to the evaluation of multimodal interfaces and we introduce the Neimo platform as an illustration of our early experience in the development of WOz platforms for the study of multimodal systems.

WIZARD OF OZ SYSTEMS

A WOz system allows the observation of a user operating an apparently fully functioning system whose missing services are supplemented by a hidden wizard. The user is not aware of the presence of the wizard and is led to believe

that the computer system is fully operational. The wizard observes the user through a dedicated computer system connected to the observed system over a network. When the user invokes a function that is not available in the observed system, the wizard simulates the effect of the function. Through the observation of users' behavior, designers can identify users' needs when accomplishing a particular set of relevant tasks and evaluate the particular interface used to accomplish the tasks. During a WOz experiment, the data exchanged between the observed user and the wizard is recorded for further analysis by interface evaluation experts.

Most of the existing WOz systems have been developed to study the usage of natural language for retrieval information systems such as telephone information services or databases [5, 7] as well as dialogues with expert systems [3]. These WOz systems involve the simulation of only one input medium (spoken or typed natural language, for example). Apart from the results of the WOz experiments themselves, these experiments led to interesting results for the design of WOz systems. The main point of interest is that wizards' tasks, although apparently simple, are cognitively expensive. The realism of the simulation requires wizards' actions to be consistent in content, style, and pace. To achieve an acceptable consistent behavior, wizards must be thoroughly trained at well defined tasks.

In order to alleviate cognitive overload, recent studies suggest a two-wizard configuration where one wizard is specialized in I/O whereas the second one performs task level processing [1]. The I/O wizard acquires the observed user's requests and transmits simulated answers. The task wizard interprets the requests translated by the I/O wizard and generates the answers to be formulated by the I/O wizard. This collaborative task sharing is more likely to guarantee consistency. It doesn't add noticeably to the response time provided that the wizards are appropriately trained. Another experiment using a two-wizard configuration has proven to be successful [4].

PARTICULARITIES OF A MULTIMODAL WOZ

From the wizard's perspective, multimodality (i.e., the user can use more than one input medium at the same time) increases the complexity of the task as well as information bandwidth. Thus, a single wizard might not be able to achieve a realistic and consistent simulation. Hence, we propose to have many wizards, each being specialized in the

simulation of a specific part of the interaction. A multiwizard configuration is an answer to the problems of task complexity and high information bandwidth in the case of multimodal interaction. However, since many wizards are involved in the simulation process, consistency might be more difficult to achieve. Thus, we feel the need for a specialized wizard who would act as a supervisor. This superwizard would not accomplish any regular wizard task but would regulate the wizards' behavior and monitor the WOz session.

Task allocation among wizards is a major issue. It strongly depends on the peculiarities of the experiment, but general rules could be applied. For example, we have found useful to consider having one wizard specialized in I/O and other wizards performing task level processing as suggested in [1]. Another approach is to assign to each wizard the simulation of a single modality, but in this case the combination of modalities may require an extra wizard specialized in the fusion of modalities.

THE NEIMO PROJECT: A MULTIMODAL WIZARD OF OZ PLATFORM

The Neimo platform is a prototype of a Wizard of Oz system for the study of multimodal interaction. It has been designed to allow experiments involving many wizards for the simulation. Moreover, the Neimo platform provides some original features:

- **Flexibility:** since the allocation of tasks among wizards is still a field of study, the Neimo platform allows for any number of wizards; wizards' tasks are dynamically redefinable, even during an experiment, and a wizard can play many roles, at once or successively during the experiment. This feature will allow us to try and study various wizards configurations.
- **Reusability:** any software can be easily adapted to the Neimo platform to be studied within a WOz experiment. If the software requires the simulation of special functionalities, the necessary simulation tools can be added without modifying the Neimo platform itself.
- **Genericity:** the studied software can request the simulation of functionalities of any level. This allows the designers to use the WOz experiments at any stage of the development process. For example, for designing software involving a handwriting recognizer, the WOz platform could first be used for the study of the integration of handwriting in the interface; at this stage, the recognizer would be entirely simulated by the wizards. In a further step, the designers could setup another experiment, with a functional handwriting recognizer; the WOz platform would then be used to confirm the results of the first experiment, and could also be used to check the accuracy of the recognizer.
- **Extensibility:** the Neimo platform has been designed to allow the simulation of any functionality and any modality by the wizards. To meet this requirement, the Neimo system consists in two main parts. On one hand a communication framework handles the circulation of requests from the observed software to the wizards, and of simulation actions from the wizards towards the observed

software. This communication framework can convey any information, regardless of its contents. On the other hand, plug-in modules are in charge of all aspects that are specific to the studied software requests or to the wizards' tasks. Thus, the addition of a new functionality (e.g., the simulation of a new modality) only requires the development of a new plug-in module on the wizards side.

During a WOz experiment with the Neimo platform, all the interactions between the observed user and the wizards are collected in an history. This material can then be used by interface evaluation experts to replay and analyze the session. Since a multimodal WOz experiment can represent a huge amount of data, we plan to provide analysis tools to help the evaluation experts' task, and possibly services to automate parts of the analysis process.

The Neimo system is developed on Apple Macintosh Quadra workstations using MacApp, an object-oriented application framework.

CONCLUSION AND PERSPECTIVES

We exposed how the WOz technique can be extended to the study of multimodal interaction, and we presented Neimo, a multimodal WOz platform. Two major areas still need further research. First, the organization of the wizards' work requires a lot of testing and experiments; we intend to define a set of rules that would help configurate the wizards' operations. Second, we plan to work on the partial automation of the analysis of the data collected during a WOz experiment.

REFERENCES

1. R. Amalberti and C. Valot: "Le Magicien d'Oz", internal presentation, CERMA, 1992.
2. S.K. Card, T.P. Moran and A. Newell: "The Psychology of Human-Computer Interaction", Lawrence Erlbaum Associates, 1983.
3. D. Diaper: "The Wizard's Apprentice: A Program to Help Analyse Natural Language Dialogues", People and Computers V, proceedings of the 5th Conference of the British Computer Society, 1989.
4. J.-M. Francony, E. Kuijpers and Y. Polity: "Towards a methodology for Wizard of Oz experiments", Third Conference on Applied Natural Language Processing, Trento, Italy, 31 march - 3 April 1992.
5. A. Jönsson and N. Dahlbäck: "Talking to A Computer is Not Like Talking To Your Best Friend", Proceedings of the Scandinavian Conference on Artificial Intelligence '88, pp. 53-68.
6. J. Löwgren and T. Nordqvist: "A Knowledge-Based Tool for User Interface Evaluation and its Integration in a UIMS", Interact '90, pp. 395-400.
7. M. Richards & K. Underwood: "How Should People and Computers Speak to Each Other", Proceedings of Interact '84, pp. 268-273.