WebWOZ: A Platform for Designing and Conducting Web-based Wizard of Oz Experiments

Stephan Schlögl
Institut Mines-Télécom
Télécom ParisTech, CNRS LTCI
Paris, France
schlogl@enst.fr

Saturnino Luz, Gavin Doherty
Trinity College
University of Dublin
Dublin, Ireland
{firstname.lastname}@scss.tcd.ie

Abstract

The Wizard of Oz (WOZ) method has been used for a variety of purposes in early-stage development of dialogue systems and language technology applications, from data collection, to experimentation, prototyping and evaluation. However, software to support WOZ experimentation is often developed ad hoc for specific application scenarios. In this demo we present WebWOZ, a web-based WOZ prototyping platform that aims at supporting a variety of experimental settings and combinations of different language technology components. We argue that a generic and distributed platform such as WebWOZ can increase the usefulness of the WOZ method.

1 Introduction

The use of language technologies such as Automatic Speech Recognition (ASR), Machine Translation (MT) and Text-to-Speech Synthesis (TTS) has significantly increased in recent years. Drivers of adoption have been enhanced quality and increasingly ubiquitous access to products and services. However, the technology is still far from perfect and typically substantial engineering effort is needed before prototypes can deliver a user experience robust enough to allow potential applications to be evaluated with real users. For graphical interfaces, well-known prototyping methods like sketching and wire-framing support the designer in obtaining early impressions and initial user feedback. These low-fidelity prototyping techniques do, however, not map well onto systems based around speech and natural language. Wizard of Oz (WOZ) tries to fill this gap by using a human 'wizard' to mimic some of the functionality of a system, which allows for evaluating potential user experiences and interaction strategies without the need for building a fully functional product first (Gould et al., 1983).

2 The WebWOZ Platform

WebWOZ is an entirely web-based, open-source Wizard of Oz prototyping platform[1]. It allows for testing interaction scenarios that employ one or more Language Technology Components (LTC). The integration of these LTCs is done via web services. Currently we have integrated ASR from Google using HTML-based Speech Input[2], on-the-fly MT from Microsoft[3] and TTS provided by the Muse Speech Technology Research Platform[4]. In addition we support pre-recorded audio and video files that are accessible through a web server. Table 1 shows the different components currently integrated into WebWOZ. Depending on the application scenario those components can be turned on and off as well as be used in combination (Schlögl et al., 2010; Schlögl et al., 2011).

2.1 Software Requirements

WebWOZ is written in Java and therefore can be hosted on a typical application server (e.g. Apache Tomcat). In addition a relational database (e.g. MySQL) is needed. In order to run experiments we further recommend the use of an up-to-date web browser that is able to adequately interpret recent HTML5 commands. For the moment, the Chrome browser is probably the best choice, since it supports speech input without the need for installing an additional plug-in. However, we are convinced that soon most web browsers will support the majority of HTML5 features required by WebWOZ.

---

[1] https://github.com/stephanschloegl/WebWOZ/
[2] http://lists.w3.org/Archives/Public/public-xg-htmlspeech/2011Feb/att-0020/api-draft.html
[3] http://msdn.microsoft.com/en-us/library/ff512419.aspx
[4] http://muster.ucd.ie/content/muse-speech-technology-research-platform

Table 1: WebWOZ Component List

ASR	HTML Speech Input
MT	Microsoft Translate
TTS	Muse Speech Technology
	Pre-recorded Audio Files

2.2 Supported Scenarios

One of the main features of WebWOZ is its integrated CMS-like editing functionality. This permits researchers/designers to create their own WOZ experiments without requiring from them any programming skills. They can add, edit, and delete utterances and organize them in different tabs (dialogue stages) using the wizard interface (cf. demo video[5]). Corresponding client (i.e. non-wizard) user/password combinations can be added and distinct interaction modes for the experiment can be set (e.g. ASR on/off, TTS on/off, MT on/off, etc.). The client interface itself runs in a separate browser window, which allows for an easy integration into already existing web applications.

Following this architecture WebWOZ supports the design of a variety of experimental settings. Different scenarios from classic monolingual text-to-text to multi-lingual speech-to-speech interactions are possible. From a wizard's perspective, tasks can reach from pure dialogue management to augmenting LTC output. That is, in WebWOZ a wizard can act as the substitute for a working dialogue manager, linking a test persons' input with an appropriate response by choosing from a set of pre-defined answer possibilities. Alternatively, however, one could be focusing on enhancing the quality of a single LTC by augmenting its output. Examples might include choosing from an n-best list of recognition results or the post-editing of output produced by an MT service.

3 Why a Web-based Solution?

The WOZ technique is usually used for four main purposes related to the design and implementation of dialogue systems: (1) it is used for dialogue data collection, (2) for controlled experimentation (including system evaluation), (3) for exploration of design alternatives and (4) for teaching of system design. Given this context, why should one build a web-based WOZ platform? What are the

benefits of such a solution? As it turns out, one can identify benefits to each of the above mentioned main uses of the WOZ method.

In terms of data collection, the gathering of multimodal dialogue corpora is often a complex and time consuming enterprise. It requires standardization and uniformity with respect to data format, timing and encoding, as well as collection settings and procedures. WOZ techniques have been increasingly used for this purpose, particularly in the gathering of data for studying multimodal information presentation and interaction e.g. (Rieser et al., 2011). A Web-based platform such as WebWOZ can facilitate data collection by geographically distributed groups while guaranteeing adherence to the requisite standards.

As regards experiments, a crucial requirement from the perspective of scientific methodology is reproducibility. Different research groups need to be able to replicate experiments according to precisely prescribed procedures and settings. Wizard of OZ experiments, however, are usually conducted using purpose built, ad hoc tools and software. This makes replication difficult, if not impossible. WebWOZ provides a widely available, standardized environment in which experimental protocols can be precisely specified and shared with interested research groups, thus supporting reproducibility. These features are similarly important for extrinsic system components evaluation e.g. (Schneider and Luz, 2011) where the overall system functionality should be kept constant while a specific component to be tested (say, an MT module) is varied.

WOZ techniques are also employed for exploration (through prototyping) of design ideas and alternatives, particularly at the early design stages of interactive systems that involve diverse language technology components. In this case, reproducibility and controlled conditions are less important. However, as distributed system development becomes a common practice WebWOZ can be used in such scenarios as a shared design artifact to support the activities of geographically distributed design teams as well as the communication among them.

Finally, WebWOZ can be (and has been) used in support of teaching the development of dialogue systems. While students are usually introduced to WOZ (i.e. written on a lecture slide) only a small portion of them receives actual hands-on experi-

---

ence. One reason for this lack of practical usage might be that in order to be applicable in a teaching context, any approach would have to have a low logistical and technical overhead to enable students to quickly design and carry out evaluations. Our experience with WebWOZ has shown that the web-based approach significantly lowers this barrier. To date more than 50 students were able to design experiments and hence improve their understanding of the complexity of dialogue systems.

4 Uses of WebWOZ in Research

WebWOZ has already been employed in two different research studies. The first study explored the effects of MT when it is used in combination with TTS (Schneider et al., 2010). The second study aimed at building and evaluating a corpus of feedback utterances sent to language learners who try to improve their pronunciation (Cabral et al., 2012).

The experimental set-up of these two studies differed greatly, highlighting the flexibility of WebWOZ. The first study tested the scenario of an intelligent computer system recommending appropriate Internet connection bundles to German speaking customers. To support this scenario a set of pre-defined dialogue utterances as well as the relevant domain utterances (i.e. examples of Internet connection bundles) were collected, automatically translated and then added to WebWOZ. On-the-fly translation was not used as the experimenters wanted to control for any possible inconsistencies. The TTS part of the experiment did not utilize a synthesis directly, but rather used the possibility of WebWOZ handling pre-synthesized audio files. ASR was simulated by the wizard. Voice-over-IP was used to transmit the participant's voice to the wizard, who then selected an appropriate response.

The second study was less restrictive. Here the researcher's goal was to built up and evaluate a corpus of feedback utterances, for which the wizard could be more open in terms of responses. Similarly to the first study a set of pre-defined responses was added to WebWOZ. However, in cases were those utterances were not sufficient, the wizard could use a free-text field to reply. Again Voice-over-IP was used to transfer speech input from a test user to the wizard and TTS was turned off, as the experiment design used textual feedback only.

5 Conclusion and Future Work

We presented WebWOZ a Wizard of Oz prototyping platform that is developed in our research group. WebWOZ differs from existing WOZ tools by being entirely web-based and through its goal of supporting various types of application scenarios. The different features of WebWOZ were highlighted and it was described how two independent studies already made use of them. Future work aims to optimize WebWOZ, to generalise it to further experimental settings and to extend it by integrating additional modalities. To do so the system has been installed in our partner institutions where it has currently been adapted to support additional settings in at least two other research projects. Although we are aware of the fact that the great difference between the interests of individual researchers pose challenges to the design of a truly generic WOZ tool, we believe that our platform can be a helpful starting point for a variety of researchers and designers who may wish to use the WOZ method.

References

J. P. Cabral, M. Kane, Z. Ahmed, M. Abou-Zleikha, É. Székely, A. Zahra, K. U. Ogbureke, P. Cahill, J. Carson-Berndsen, and S. Schlögl. 2012. Rapidly Testing the Interaction Model of a Pronunciation Training System via Wizard-of-Oz. In *Proceedings of LREC*.

J. D. Gould, J. Conti, and T. Hovanyecz. 1983. Composing letters with a simulated listening typewriter. *Communications of the ACM*, 26:295–308.

V. Rieser, S. Keizer, X. Liu, and O. Lemon. 2011. Adaptive Information Presentation for Spoken Dialogue Systems: Evaluation with human subjects. In *Proceedings of ENLG*, pages 102–109.

S. Schlögl, G. Doherty, N. Karamanis, A. H. Schneider, and S. Luz. 2010. Observing the wizard: In search of a generic interface for wizard of oz studies. In *Proceedings of Irish HCI*, pages 43–50.

S. Schlögl, A. H. Schneider, S. Luz, and G. Doherty. 2011. Supporting the wizard: Interface improvements in wizard of oz studies. In *Proceedings of BCS HCI*.

A. H. Schneider and S. Luz. 2011. Speaker alignment in synthesised, machine translated communication. In *Proceedings of IWSLT*, pages 254–260.

A. H. Schneider, I. Van der Sluis, and S. Luz. 2010. Comparing intrinsic and extrinsic evaluation of mt output in a dialogue system. In *Proceedings of the IWSLT*, pages 329–336.