# QuickWoZ: A Multi-purpose Wizard-of-Oz Framework for Experiments with Embodied Conversational Agents

*Jan Smeddinck, Kamila Wajda, Adeel Naveed, Leen Touma, Yuting Chen, Muhammad Abu Hasan, Muhammad Waqas Latif, Robert Porzel*
Digital Media, FB 3, University of Bremen
confetti | porzel@informatik.uni-bremen.de

**ABSTRACT**
Herein we describe the QuickWoZ system, a Wizard-of-Oz (WoZ) tool that allows for the remote control of the behavior of animated characters in a 3D environment. The complete scene, character, behaviors and sounds can be defined in simple XML documents, which are parsed at runtime, so that setting up an experiment can be done without programming expertise. Quick selection lists and buttons enable the *wizard* to easily control the agents' behavior and allow for fast reactions to the subjects' input. The system is tailored for experiments with embodied conversational agents (ECAs) featuring multimodal interaction and was designed as a rapid prototyping system for evaluating the impact of an agent's behavior on the user.

**Author Keywords**
HCI, embodiment, conversational agents, evaluation

**ACM Classification Keywords**
H5.2. Information interfaces and presentation: Graphical user interfaces (GUI)

**General Terms**
Experimentation, Measurement

**INTRODUCTION**
ECAs featuring natural language and other natural modalities processing capabilities promise interactional benefits as an HCI metaphor, but their development as a suitable intelligent user interface remains a challenging enterprise. In order to explore various aspects and implications of HCI systems – without the need to develop the system first – WoZ experiments have been employed as a suitable tool for testing and evaluating the interaction with the system [5]. Such experiments, therefore, allow the compilation of user requirements concerning ECAs [4]. Typical WoZ-tools are often difficult to use and designed for a single purpose. Additionally, we noted a complete lack of systems offering means to include natural 3D motion and animation of the character. This motivated the creation of this a more generally applicable system – based on popular formats and open definition files – that enables the designer to record and import libraries of movements, which can be combined with each other and more modalities such as text-to-speech audio. In the following we describe the architecture of the QuickWoZ system and discuss some of the implications concerning future improvements that are motivated by its experimental employment for testing the impact of agent character traits and gestural behavior on user expectations of ECA systems.

**RELATED WORK**
As noted above a whole range of tools for controlling 2D virtual characters and voice-based systems exists [4], but none that allow for integrating fully embodied 3D animation of ECAs. Additionally, a host of studies exist that examine the potential effects of an agent's design and personality on the interaction [3,6]. The work presented herein can be thought of as an enabling tool to further this type of study by the creation of a more generally usable WoZ system that allows the wizard to control 3D embodied agents within their environments.

**FRAMEWORK AND PIPELINE**
The framework and pipeline were designed for specific types of experimental settings, where the goal is to test not only the effects of the verbal but also the non-verbal behavior of an agent. Given its particular construction, however, the system can also be employed for different purposes as it is intuitively adaptable. For this reason, QuickWOZ can be seen as a multi-purpose experimental framework that could be swiftly applied (without programming) to a large number of possible experiments.

QuickWoZ is programmed in C# and uses Ogre3D with the C# wrapper library MOGRE for .NET integration as the visual engine, supporting both DirectX and OpenGL rendering [7]. The sound streaming is managed with the BASS.NET library [1]. With these basic technologies QuickWoZ supports the creation of scenes, characters and animations with all industry standard 3D modeling and animation tools and the most common sound file formats.


*IUI'10,* February 7–10, 2010, Hong Kong, China.
ACM 978-1-60558-515-4/10/02.

Funded by the DFG within the SFB 637.

An arbitrary amount of QuickWoZ instances can be synchronized through a relay-server. All engine-controlling functions that are influencing a given scene are then automatically synchronized, while user interface actions run in a different thread and are not synchronized. This allows for one client to be put into full-screen mode, while the instance on the wizard's computer runs the engine in window mode. The experiment operators can thereby easily observe the exact behavior of the agent as it is presented to experiment participants.

In order to obtain natural human-like animations in comparably short time we decided to capture animations with a motion capture suit. Inverse-kinematics, physically, or behaviorally driven animations can be used alternatively. These animations were mapped to a biped skeleton in a 3D character animation software and skinned with a simple humanoid body. Ogre exporter tools exist for all common 3D-modeling and -animation suites and can be used to generate the necessary meshes and skeleton binaries. The animations in the skeleton file can be linked to the according sound files via XML documents. Next to such a library of movements and gestures, natural language responses of the agent can be prerecorded and distorted with audio filters to mimic the sound of a computer-synthesized voice or spoken directly by the wizard using online distortion filters.

As other WoZ tools, the QuickWoZ system will typically be running on two or more computers in different rooms. The connection can be set up in the 'Options' panel of the menu bar. To be able to control the prerecorded libraries, the operator first has to start Ogre and synch the two machines using the accordant 'Synch'-button. In the 'File' panel, the wizard can load an XML expression file whose content appears as alphabetically sorted list on the right side of the window. In order to play an animation the wizard can double-click an expression from the list after running Ogre. To find an expression from the list in short time, the operator can also use the search bar on the right top of the application. As depicted in Figure 1, individual buttons are offered for the most frequently used animations which are automatically generated from the expression list. In case the user asks a question which was not prerecorded, the wizard can use the 'push to talk' button which activates a prerecorded animation and the answering is done live by a speaker using sound distortion filters.

The QuickWoZ system has been employed in an experiment testing the impact of the character traits and gestures of an agent on user-expectations and trust. In this case, two wizards controlled the agent together: One selected the behaviors as replies to user requests and the other one acted as a "live speaker", when no prepared answer was available as a matching response to user requests. For this test, live speech was achieved using an external sound-distortion tool. For the next version of the QuickWoZ system we plan to integrate this functionality

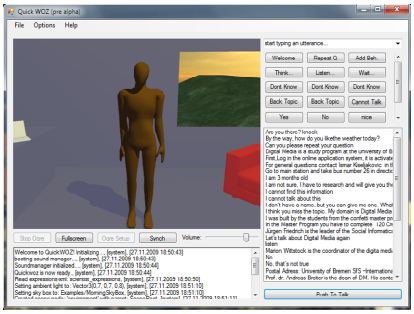

**Figure 1. Screenshot of a QuickWoZ sample scene with quick-play buttons on the top right.**

into the core system. The deployment of the system delivered valuable feedback both in terms of the experimental design as well as for the architecture of the system. The tool ran stable and allowed for variable length synchronized sessions. The illusion of the ECA was such that none of the participants noticed that they were interacting with a wizard-system and not a real one.

## CONCLUSION AND FUTURE WORK

The outcome of the deployment showed that the QuickWoZ system provided an appropriate illusion of a real naturally behaving artificial intelligent. Participants described the system as an engaging interaction experience. The next version of QuickWoZ will include real-time filters for microphone sound-distortion so that the number of tools required by a wizard drops to one and future experiments can be handled by a single wizard in a more convenient manner. Also, we will add speech synthesis technology to enable faster extensions with more expressive capabilities for different characters.

## REFERENCES


1. BASS audio library (http://www.un4seen.com/).
2. Cassell, J. and Ryokai, K. 2001. Making Space for Voice: Technologies to Support Children's Fantasy and Storytelling. *Personal Ubiquitous Comput.* 5, 3 2001.
3. Cassell, J. and T. Bickmore. External Manifestations of Trustworthiness in the Interface. Communications of the ACM, 43(12):50-56, 2000.
4. Dahlback, N., Jonsson, A., Ahrenberg, L. Wizard of OZ studies-Why and How, In Readings in intelligent user interfaces (1993), 610-619.
5. Francony, J.-M., E. Kuijpers, and Y. Polity. Towards a methodology for Wizard of Oz experiments. In 3rd Conference on Applied NLP, 1992.
6. Kraemer, N., N. Simons, and S. Kopp. The effects of an embodied conversational agent's nonverbal behavior on user's evaluation and behavioral mimicry. In Proc. Of 7th IVA, LNAI 4722, Springer-Verlag, 2007.
7. MOGRE .NET wrapper for Ogre3D. http://www.ogre3d.org/wiki/index.php/MOGRE