

# ChatWoz: chatting through a Wizard of Oz

Pedro Fialho  
INESC-ID  
Universidade de Évora  
pedro.fialho@l2f.inesc-id.pt

Luísa Coheur  
INESC-ID  
Instituto Superior Técnico  
Universidade de Lisboa  
luisa.coheur@l2f.inesc-id.pt

## ABSTRACT

Several cases of autistic children successfully interacting with virtual assistants such as Siri or Cortana have been recently reported. In this demo we describe CHATWOZ, an application that can be used as a Wizard of Oz, to collect real data for dialogue systems, but also to allow children to interact with their caregivers through it, as it is based on a virtual agent. CHATWOZ is composed of an interface controlled by the caregiver, which establishes what the agent will utter, in a synthesised voice. Several elements of the interface can be controlled, such as the agent's face emotions. In this paper we focus on the scenario of child-caregiver interaction and detail the features implemented in order to couple with it.

## Categories and Subject Descriptors

K.3.1 [Computer Uses in Education]: Computer-assisted instruction (CAI); K.3.1 [Computer Uses in Education]: Computer-managed instruction (CMI); K.4.2 [Social Issues]: Assistive technologies for persons with disabilities

## Keywords

Wizard of Oz, autistic children, interaction, emotion, speech

## 1. INTRODUCTION

A Wizard of Oz (WOZ) is the name given to the family of experiments in which a user interacts with a computer that is being controlled by an unseen human, being the user not aware of this fact. WOZ are important as they allow to understand what the specific (linguistic) needs of the application are [1]. In the literature we can find several studies regarding the building of WOZ applications (*e.g.* [2]), but also that show how WOZs have been used to collect data/improve different systems. For instance, the work from [4] focus on the type of data gathered by an WOZ (audio, video, etc.), while the study described in [3] shows how data collected by a WOZ can be used to learn dialogue strategies.

In the context of a project where we try to boost several

capacities of autistic children, and being aware of recent reports of children enthusiastically interacting with virtual assistants<sup>1</sup>, as well as studies, such as the one described in [5], we felt the need of developing a WOZ that would allow the child to interact with a caregiver through an avatar. In this paper we describe that system, the CHATWOZ.

## 2. GENERAL OVERVIEW

The design and deployment of CHATWOZ's front end is performed in Unity<sup>2</sup>. Several modules communicate by means of specific protocols, thus leveraging the capabilities of independent modules such as Text-to-Speech (TTS) (remote) or 3D rendering/animation. The main interface (Figure 1 – in Portuguese) is handled by one of the users (from now on, the caregiver), who will control what the agent will say.

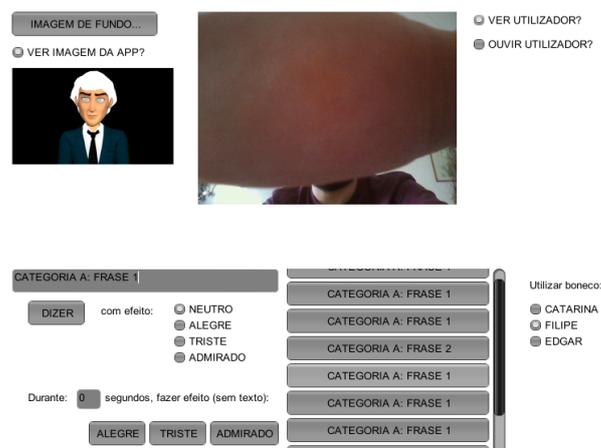


Figure 1: The main interface.

After choosing between predefined utterances or by entering a new sentence, the chosen text will be synthesised. The audio is played while the corresponding phonemes are mapped into visemes, represented as morph animations, being synchronized according to phoneme durations, provided by the speech synthesizer. An emotion can also be chosen to animate the agent's face, either when it is uttering the sentence, or when it is simply "listening". The second user (for now

<sup>1</sup>See, for instance, <http://www.nytimes.com/2014/10/19/fashion/how-apples-siri-became-one-autistic-boys-bff.html>

<sup>2</sup><https://unity3d.com>

on, the child) will hear the agent saying the utterances provided by the caregiver, with the chosen emotion. Anything said by the child will reach the caregiver.

### 3. CHATWOZ MAIN FEATURES

The agent utterances and emotions are controlled by the caregiver. Concerning the agent utterances, the caregiver writes the sentences the agent will utter. However, if it is possible to know in advance which utterances the agent will be saying, it is possible to pre-record them (either by a human or by using a synthesiser). These utterances can also be clustered in categories. In the main interface, a menu with these utterances appears (keeping the categories organisation), which allows to ease/speed the process of choosing what the agent will say. Instead of writing, the caregiver only has to click on the appropriate utterance, which was already previously synthesized (or recorded by an human), which also speeds the process. Regarding the agent's emotions, the caregiver can choose between five different emotions regarding the agents's face when uttering a sentence. These are "neutro" ("neutral"), "alegre" ("joyful"), "triste" ("sad") and "admirado" ("surprised") (Figure 2). In what concerns the emotions expressed by the agent when it is listening to the child, these can be "alegre" ("joyful"), "triste" ("sad") and "admirado" ("surprised"). After expressing that emotion, the agent will go back to its neutral state.



Figure 2: The agent emotions

Also, several elements of CHATWOZ can be customised, namely:

a) The scenario, which can be changed, by simply uploading a new image; b) The audio and video regarding the child. The caregiver can decide, during the session, if he/she wants to hear/see the child or not. Obviously, activating these features can slow the interface (specially the video), which can be a problem with high latency connections; c) The child's interface, as the caregiver can also see it as displayed in the child's device screen; d) The agent. For the moment we can opt by Filipe, Catarina or Edgar (Figure 3). Agents can be changed during a session. One leaves and the other walks in.



Figure 3: Currently available agents (from left to right, Catarina, Filipe and Edgar)

### 4. DEMONSTRATING CHATWOZ

A person will act as the caregiver and a second person will act as the child. The first will have a laptop and will see

CHATWOZ's main interface; the latter will see the agent. They will "talk" through the avatar. The "caregiver" can choose what the agent will say, between a set of predefined sentences; he/she can also enter a new sentence. The "child" will simply hear and talk with the agent.

Besides deciding what the agent will say and which emotion it should have in its face, the caregiver can, in runtime: a) Change the agent ("Utilizar boneco" ("use agent") in the bottom right of the main interface); b) Change the back figure of the agent ("Imagem de fundo" ("back image"), in the top left corner of the main interface); c) Decide to have (or not) audio/video regarding the child actions ("Ver/Ouvir Utilizador" ("see/hear user) in the top right of the main interface).

We have made some preliminary tests and the only problem is to have the interface working with the video (and also with audios, although in a smaller scale), which can slow the interaction. We concluded that Unity is not suitable for video transmission.

### 5. CONCLUSIONS AND FUTURE WORK

We have described our WOZ, that we intend to use to make children interact with his/her caregiver through an avatar. We are currently trying to add some emotions to the TTS, by voice transfer, and also to fix latency issues related with audio/video transmission.

### 6. ACKNOWLEDGMENTS

This work was supported by Fundação para a Ciência e Tecnologia (FCT) with reference UID/CEC/50021/2013, and by project CMUP-ERI/HCI/0051/2013.

### 7. REFERENCES

- [1] N. Dahlbäck, A. Jönsson, and L. Ahrenberg. Wizard of oz studies: Why and how. In *Proceedings of the 1st International Conference on Intelligent User Interfaces*, IUI '93, New York, NY, USA, 1993.
- [2] S. R. Klemmer, A. K. Sinha, J. Chen, J. A. Landay, N. Aboobaker, and A. Wang. Suede: A wizard of oz prototyping tool for speech user interfaces. In *Proceedings of the 13th Annual ACM Symposium on User Interface Software and Technology*, pages 1–10, New York, NY, USA, 2000. ACM.
- [3] V. Rieser and O. Lemon. Learning effective multimodal dialogue strategies from wizard-of-oz data : bootstrapping and evaluation. In *ACL-08: HLT : 46th annual meeting of the Association for Computational Linguistics: human language technologies*, 2008.
- [4] P. Strauss, H. Hoffmann, W. Minker, H. Neumann, G. Palm, and S. S. et al. Wizard-of-oz data collection for perception and interaction in multi-user environments. In *International Conference on Language Resources and Evaluation (LREC)*, 2006.
- [5] A. Tartaro and J. Cassell. Authorable Virtual Peers for Autism Spectrum Disorders. In *Proceedings of the Workshop on Language-Enabled Educational Technology at the 17th European Conference on Artificial Intelligence (ECAI06)*, Italy, 2006.