This document is published in:

Andre Ponce de Leon F. de Carvalho, et al. (eds.) (2010). *Distributed Computing and Artificial Intelligence: 7th International Symposium*. (Advances in Intelligent and Soft Computing, 79) Springer, 283-290.
DOI: http://dx.doi.org/10.1007/978-3-642-14883-5_37

# A Conversational Academic Assistant for the Interaction in Virtual Worlds

D. Griol, E. Rojo, Á. Arroyo*, M.A. Patricio, and J.M. Molina

Computer Science Department. Carlos III University of Madrid
e-mail: {david.griol,eduardo.rojo,miguelangel.patricio}@uc3m.es, josemanuel.molina@uc3m.es

∗Applied Intelligent Systems Department. Technical University of Madrid e-mail: aarroyo@eui.upm.es

**Abstract.** The current interest and extension of social networking are rapidly introducing a large number of applications that originate new communication and interaction forms among their users. Social networks and virtual worlds, thus represent a perfect environment for interacting with applications that use multimodal information and are able to adapt to the specific characteristics and preferences of each user. As an example of this application, in this paper we present an example of the integration of conversational agents in social networks, describing the development of a conversational avatar that provides academic information in the virtual world of Second Life. For its implementation techniques from Speech Technologies and Natural Language Processing have been used to allow a more natural interaction with the system using voice.

## 1 Introduction

Social Networking has been a global consumer phenomenon during the last few years [10]. The staggering increase in the amount of time people are spending on these sites is changing the way people spend their time online and influence on how people behave, share and interact within their normal daily lives. The development of so-called Web 2.0 has also made possible the introduction of a number of applications into many users' lives, which are profoundly changing the roots of society by creating new ways of communication and cooperation.

The advance of social networking has entailed a considerable progress in the development of virtual worlds [8, 1]. Virtual robots (*metabots*), with the same appearance and capabilities that the avatars for human users, thus intensify the perception of the virtual world, providing gestures, glances, facial expressions and movements

necessary for the communication process. Therefore, these virtual environments are very useful to enhance human-machine interaction.

This way, virtual worlds have become real social networks useful for the interaction between people from different places who can socialize, learn, be entertained, etc. Thanks to the social potential of virtual worlds, they have also become an attraction for institutions, companies and researchers with the purpose of developing virtual robots with the same look and capabilities of avatars for human users. However, social interaction in virtual worlds are usually carried out using only text communication by means of chat-type services. In order to enhance communication in these environments, we propose the integration of conversational agents to develop intelligent metabots with the ability of oral communication and, at the same time, which benefit from the visual modalities provided by these virtual worlds.

Conversational agents [9, 7, 6] can be defined as automatic systems that are able of emulating a human being in a dialog with another person, in order to complete a specific task (usually providing information or perform a particular task.) Two main objectives are fulfilled thanks to its use. The first objective is to facilitate a more natural human-machine interaction using the voice. The second one allows the accessibility for users with motor disabilities, so that the interface avoids the use of traditional interfaces, such as keyboard and mouse.

Our work focuses on three key points. Firstly, since it is very difficult to find studies in the literature that describe the integration of Speech Technologies and Natural Language Processing in virtual worlds, to show that this integration is possible. Secondly, to show a practical application of this integration through the development of a conversational metabot that provides academic information in the virtual world Second Life. Finally, we promote the use of open source applications and tools for the creation and interaction in virtual worlds, such as OpenSim[1] and OsGrid[2].

## 2 Conversational Agents

As stated in the introduction, a conversational agent is a software that accepts natural language as input and generates natural language as output, engaging in a conversation with the user. To successfully manage the interaction with the users, conversational agents usually carry out five main tasks: automatic speech recognition (ASR), natural language understanding (NLU), dialog management (DM), natural language generation (NLG) and text-to-speech synthesis (TTS). These tasks are usually implemented in different modules.

Speech recognition is the process of obtaining the text string corresponding to an acoustic input. It is a very complex task as there is much variability in the input characteristics, which can differ depending on the linguistics of the utterance, the speaker, the interaction context and the transmission channel. Different applications demand different complexity of the speech recognizer. [4] identify eight parameters that allow an optimal tailoring of the speech recognizer: speech mode, speech style,

[1] http://opensimulator.org
[2] http://www.osgrid.org

dependency, vocabulary, language model, perplexity, SNR and transductor. Regarding the speech mode, speech recognizers can be classified into isolated-word or continuous-speech recognizers. Regarding the speech style, discourse can be read or spontaneous, the latter has peculiarities such as hesitations and repetitions that make it more complex to recognize.

Once the conversational agent has recognized what the user uttered, it is necessary to understand what he said. Natural language processing is the process of obtaining the semantic of a text string. It generally involves morphological, lexical, syntactical, semantic, discourse and pragmatical knowledge. In a first stage lexical and morphological knowledge allow dividing the words in their constituents distinguishing lexemes and morphemes. Syntactic analysis yields a hierarchical structure of the sentences, however in spoken language frequently phrases are affected by the difficulties that are associated to the so-called disfluency phenomena: filled pauses, repetitions, syntactic incompleteness and repairs [5].

Semantic analysis extracts the meaning of a complex syntactic structure from the meaning of its constituents. In the pragmatic and discourse processing stage, the sentences are interpreted in the context of the whole dialog, the main complexity of the stage is the resolution of anaphora, and ambiguities derived from phenomena such as irony, sarcasm or double entendre.

There is not a universally agreed upon definition of the tasks that the dialog management module has to carry. [11] state that dialog managing involves four main tasks: i) updating the dialog context, ii) providing a context for interpretations, iii) coordinating other modules and iv) deciding the information to convey and when to do it. Thus, the dialog manager has to deal with different sources of information such as the NLU results, database queries results, application domain knowledge, knowledge about the users and the previous dialog history. Its complexity depends on the task and the dialog flexibility and initiative. When it is necessary to execute and monitor operations in a dynamically changing application domain, an agent-based approach can be employed to develop the dialog management module. The modular agent-based approach to dialog management makes it possible to combine the benefits of different dialog control models, such as finite-state based dialog control and frame-based dialog managing Similarly, it can benefit from alternative dialog management strategies, such as the system-initiative approach and the mixed-initiative approach Furthermore, it makes it possible to combine rule-based and machine learning approaches.

Natural language generation is the process of obtaining texts in natural language from a non-linguistic representation. It is usually carried out in five steps: content organization, content distribution in sentences, lexicalization, generation of referential expressions and linguistic realization. It is important to obtain legible messages, optimizing the text using referring expressions and linking words and adapting the vocabulary and the complexity of the syntactic structures to the user's linguistic expertise. The simplest approach consists in using predefined text messages (e.g. error messages and warnings). Although intuitive, this approach completely lacks from any flexibility. The next level of sophistication is template-based generation, in which the same message structure is produced with slight alterations. Using this

approach, it is possible to provide adapted system prompts that take into account context information.

Finally, text-to-speech synthesizers transform a text into an acoustic signal. A text-to-speech system is composed of two parts: a front-end and a back-end. The front-end carries out two major tasks. Firstly, it converts raw text containing symbols such as numbers and abbreviations into their equivalent words. Secondly, it assigns a phonetic transcriptions to each word, and divides and marks the text into prosodic units, i.e. phrases, clauses, and sentences. The back-end (often referred to as the synthesizer) converts the symbolic linguistic representation into sound. On the one hand, speech synthesis can be based in human speech production, this is the case of parametric synthesis which simulates the physiological parameters of the vocal tract, and formant-based synthesis which models the vibration of vocal chords.

## 3 SecondLife and OpenSim

Second Life[3] (SL) is a three dimensional virtual world developed by Linden Lab in 2003 and accessible via the Internet. A free client program called the Second Life Viewer enables its users, called *Residents*, to interact with each other through avatars. Residents can explore, meet other residents, socialize, participate in individual and group activities, and create and trade virtual property and services with one another, or travel throughout the world, which residents refer to as the grid. Resident population is nowadays of millions of real people from around the world. Each person is represented by an avatar that represents their chosen digital persona. Second Life is currently used as a platform for education by many institutions, such as colleges, universities, libraries and government entities (e.g. Ohio University, Royal Opera House, Universidad Pública de Navarra, Instituto Cervantes, Universidad Carlos III de Madrid, etc.). We own an island in Second Life called TESIS, in which we built its Virtual facilities in which numerous educational activities are performed. Figure 1 shows an image of the TESIS island.

We decided to use SL as an experimental laboratory of our research for several reasons. Firstly, because it is one of the most popular social virtual worlds, and its population is now of millions of residents worldwide. Secondly, because it uses very advanced technologies for the development of realistic simulations, making avatars and the environments more credible and similar to real world users. Thirdly, because the possibility of customizing SL is extensive and supports innovation and user participation, which increases the naturalness of the interactions in the virtual world.

OpenSim is an open source simulator that uses the same standard as Second Life to communicate with their users, and emulates virtual environments independently from the world of Second Life, using its own infrastructure. OsGrid is a network that allows linking free virtual worlds developed using simulators such as OpenSim.

[3] http://secondlife.com/

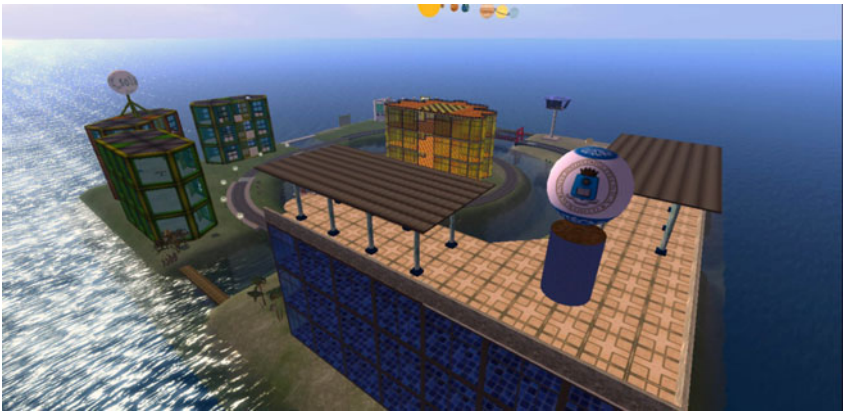

**Fig. 1** An image of the TESIS island in Second Life

## 4 Practical Implementation: Metabot That Provides Academical Information

We have developed a conversational metabot that facilitates academic information (courses, professors, doctoral studies and enrollment) based on the functionalities provided by a previously developed dialog system [2, 3]. The information provided by the metabot can be classified into four categories: subjects, teachers, doctoral studies, and registration. The system has been developed using the typical architecture of current spoken conversational agents, including a module for automatic speech recognition, a dialog manager, a module for access to databases, data storage, and the generation of an oral response through a language generator and a text to speech synthesizer, as described in the previous section.

Figure 2 shows the architecture developed for the integration of conversational metabot both in the Second Life and OpenSim virtual worlds. The conversational agent that governs the metabot is outside the virtual world, using external servers that provide both data access and speech recognition and synthesis functionalities. The speech signal provided by the text to speech synthesizer is captured and transmitted to the voice server module in Second Life (SLVoice) using code developed in Visual C #. NET and the SpeechLib library. This module is external to the client program used to display the virtual world and is based on the Vivox technology, which uses the RTP, SIP, OpenAL, TinyXPath, OpenSSL and libcurl protocols to transmit voice data. We also use the utility provided by Second Life lipsynch to synchronize the voice signal with the lip movements of the avatar.

In addition, we have integrated a keyboard emulator that allows the transmission of the text transcription generated by the conversational avatar directly to the chat in Second Life. The system connection with the virtual world is carried out by using the libOpenMetaverse library. This .Net library, based on the Client /Server

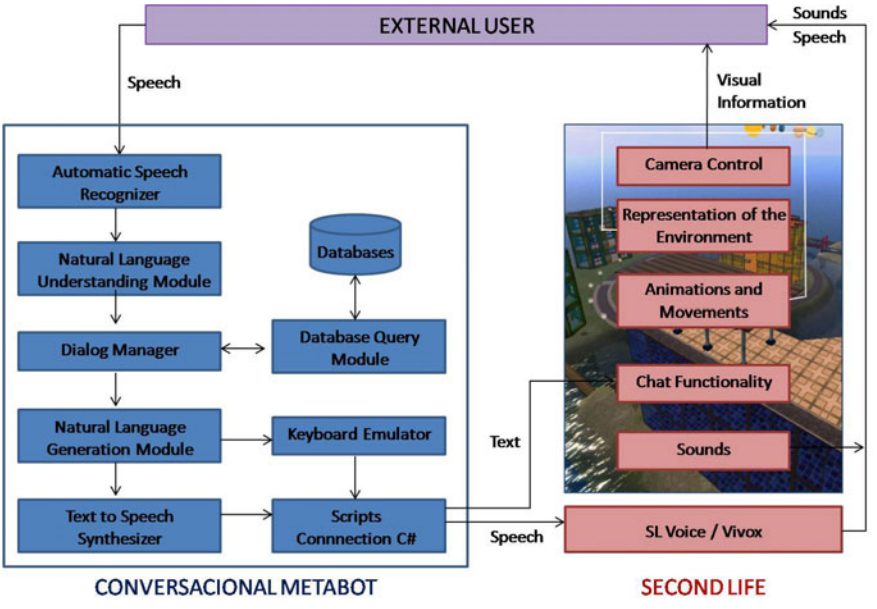


**Fig. 2** Architecture defined for the development of the conversational metabot

paradigm, allows to access and create three-dimensional virtual worlds, and it is used to communicate with servers that control the virtual world of Second Life.

Speech recognition and synthesis are performed using the Microsoft Speech Application Programming Interface (SAPI), integrated into the Windows Vista operating system. To enable the interaction with the conversational in Spanish using the chat in Second Life, we have integrated synthetic voices developed by Loquendo.

Using this architecture user's utterances can be easily recognized, the transcription of these utterances can be transcribed in the chat in Second Life, and the result of the user's query can be communicated using both text and speech modalities. To do this, we have integrated modules for the semantic understanding and dialog management implemented for the original dialog system, which are based on grammars and VXML files. Using OsGrid we have also developed our own free virtual world and integrated our conversational metabot with OpenSim.

Through the participation of students and professors of our university we have acquired a set of dialogs using the same scenarios defined for a previous acquisition using the conversational agent that governs the avatar with real users outside the virtual world of Second Life. Table 1 shows the statistics of the acquisition of 50 dialogs. The main conclusion that can be extracted from this preliminary study is the absence of differences in these statistics among the dialogs acquired using only the conversational agent and the dialogs acquired by means of the interaction with the conversational metabot in Second Life. Figure 3 shows the developed metabot providing information about tutoring hours of a specific professor.

**Table 1** Statistics of the acquired dialogs

| | |
|---|---|
| Average number of turns per dialog | 4.99 |
| Percentage of confirmations from the metabot | 13.51% |
| Questions from the metabot to request information | 18.44% |
| Prompts generated by the metabot after a database query | 68.05% |


**Fig. 3** Conversational metabot developed to interact in virtual worlds

## 5 Conclusions

The development of social networks and virtual worlds brings a wide set of opportunities and new communication channels that can be incorporated to traditional interfaces like conversational agents. In this work, we propose a methodology for creating conversational metabots which are able to interact in virtual worlds. Using our the proposal we have implemented a conversational metabot that provides academic information in Second Life. This virtual world offers a number of possibilities for the development of educational applications, given the possibility for users to socialize, explore and access a large number of educational and cultural resources. As future work we want to evaluate new features to be included in the conversational metabot to improve the communication process, and carry out a detailed analysis of the integration of different modalities for the presentation of information provided by SL in addition to the use of voice.

**Acknowledgements.** Funded by projects CICYT TIN2008-06742-C02-02/TSI, CICYT TEC2008-06732-C02-02/TEC, SINPROB, CAM MADRINET S-0505/TIC/0255, and DPS2008-07029-C02-02.

## References


1. Arroyo, A., Serradilla, F., Calvo, O.: Multimodal agents in second life and the new agents of virtual 3d environments. In: Mira, J., Ferrández, J.M., Álvarez, J.R., de la Paz, F., Toledo, F.J. (eds.) IWINAC 2009. LNCS, vol. 5601, pp. 506–516. Springer, Heidelberg (2009)
2. Callejas, Z., López-Cózar, R.: Implementing modular dialogue systems: a case study. In: Proc. of Applied Spoken Language Interaction in Distributed Environments (ASIDE 2005), Aalborg, Denmark (2005)
3. Callejas, Z., López-Cózar, R.: Relations between de-facto criteria in the evaluation of a spoken dialogue system. Speech Communication 50(8-9), 646–665 (2008)
4. Cole, R., Zue, V.: Survey of the State of the Art in Human Language Technology, pp. 1–49. Cambridge University Press, Cambridge (1997)
5. Gibbon, D., Mertins, I., Moore, R.: Resources, Terminology and Product Evaluation. In: Handbook of Multimodal and Spoken Dialogue Systems. Kluwer International Series in Engineering and Computer Science, vol. 565. Kluwer Academic Publishers, Dordrecht (2000)
6. Griol, D., Hurtado, L., Segarra, E., Sanchis, E.: A Statistical Approach to Spoken Dialog Systems Design and Evaluation. Speech Communication 50(8-9), 666–682 (2008)
7. López-Cózar, R., Araki, M.: Spoken, Multilingual and Multimodal Dialogue Systems. In: Development and Assessment. John Wiley & Sons, Chichester (2005)
8. Lucia, A.D., Francese, R., Passero, I., Tortora, G.: Development and evaluation of a virtual campus on second life: The case of seconddmi. Computers & Education 52(1), 220–233 (2009)
9. McTear, M.F.: Spoken Dialogue Technology: Towards the Conversational User Interface. Springer, Heidelberg (2004)
10. Nielsen: Global Faces and Networked Places: A Nielsen Report on Social Networking's New Global Footprint. Nielsen Online (2009)
11. Traum, D., Larsson, S.: The Information State Approach to Dialogue Management. In: Current and New Directions in Discourse and Dialogue, pp. 325–354. Kluwer Academic Publishers, Dordrecht (2003)