Griol D., Baena I., Molina J.M., de Miguel A.S. (2014) A Multimodal Conversational Agent for Personalized Language Learning. In: Ramos C., Novais P., Nihan C., Corchado Rodríguez J. (eds) Ambient Intelligence - Software and Applications. Advances in Intelligent Systems and Computing, vol 291.pp 13-21 Springer, Cham

DOI: 10.1007/978-3-319-07596-9_2

A Multimodal Conversational Agent for Personalized Language Learning

David Griol, Ismael Baena, José Manuel Molina, and Araceli Sanchis de Miguel

Computer Science Department
Carlos III University of Madrid
Avda de la Universidad, 30, 28911 - Leganés, Spain
{david.griol,josemanuel.molina,araceli.sanchis}@uc3m.es,
100065131@alumnos.uc3m.es

Abstract. Conversational agents have became a strong alternative to enhance educational systems with intelligent communicative capabilities. In this paper, we describe a multimodal conversational agent that facilitates an independent and user-adapted second language learning. The different modules of the system cooperate to interact with students using spoken natural language and visual modalities, and adapt their functionalities taking into account their evolution and specific preferences. The results of a preliminary evaluation show that users' satisfaction with the system was high, as well as the perceived didactic potential and adaptive functionalities.

Keywords: Multimodal conversational agents, e-learning, educative technology, natural language processing.

1 Introduction

Ambient Intelligence is characterized by intelligent, pervasive, and seamless computer systems embedded into everyday devices, tailored to the individual's context-aware needs and providing a natural and intelligent interaction. This way, multimodal conversational agents [1] have became a strong alternative to enhance multi-agent systems with these intelligent communicative capabilities [2].

With the growing maturity of conversational technologies, the possibilities for integrating conversation and discourse in e-learning are receiving greater attention. Using natural language in educational software allows students to spend their cognitive resources on the learning task, and also develop more social-based agents [3].

Current possibilities to employ conversational agents for educative purposes include tutoring applications [4], question-answering [5], conversation practice for language learners [6], pedagogical agents and learning companions [7], and dialogs to promote reflection and metacognitive skills [8]. These agents may also be used as role-playing actors in immersive learning environments [9].

Systems developed to provide these functionalities typically rely on a variety of components, such as speech recognition and synthesis engines, natural language processing components, dialog management, databases management, and graphical user interfaces. Laboratory systems usually include specific modules of the research teams that build them, which make portability difficult. Thus, it is a challenge to package up these components so that they can be easily installed by novice users with limited engineering resources. In addition, due to this variability and the huge amount of factors that must be taken into account, these systems are difficult to develop and typically are developed ad-hoc, which usually implies a lack from scalability. Our work represents a step in this direction.

In this paper we describe a multimodal conversational agent for adaptive second language learning. The system has been developed by means of a modular approach that allows to easily developing multimodal conversational agents for pedagogical applications. This approach facilitates a rapid and cost-effective development. This way, different alternatives can be considered for each module, and the pedagogic knowledge is separated from the technical details, so that teachers and parents can add new contents without having a technical background at the same time as the software includes these new data for the interaction with the students.

2 The *Test Your English* Pedagogical System

The *Test Your English* pedagogical system has been designed with the main aim of facilitating an independent and personalized second language learning. Figure 1 shows the initial screen of the system. As it can be observed, users must register in the application. This way, their previous interactions can be taken into account to provide user adaptation functionalities.

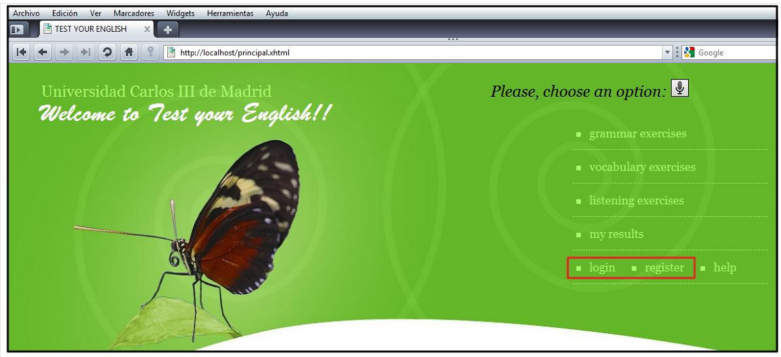

Fig. 1. Main screen of the *Test Your English* system

The current version of the system is web-based. The application consists of the set of core components for multimodal dialog systems (speech recognition, spoken language understanding, dialog management, language generation, and speech synthesis) and split across a server and a client device running a standard web browser. Core components on the server provide speech recognition and speech synthesis capabilities, access to the databases, and a logger component which records user interactions.

The system is accessible to any user with a standard web browser and a network connection. This way, the application can be easily accessed not only from desktop, laptop, and tablet computers, but from a variety of mobile devices as well. In addition, an Audio Controller component runs on the client to capture a user's speech and stream it to the speech recognizer, as well as to play synthesized speech generated on the server and streamed to the client.

Natural language understanding is performed by means of grammars which include the different options that are also visually provided to the student. To do this, we follow the Java Speech Grammar Format (JSGF, `www.w3.org/TR/jsgf/`), which allows specifying these sentences in a compact way, easily adapted and also embedding semantics into the grammar.

Speech recognition hypotheses are passed to the dialog manager architecture for processing. Regarding dialog management, all the events in the application are controlled using JavaScript. Given the requisites of the task, we decided to use a dialog model based on finite states in which at each moment a question is selected and shown in the screen along with the alternative answers and the associated multimedia files.

The main functions of the system can be classified into three main modules: Practice, Assessment, and Contents management. The Practice module includes three kinds of exercises: grammar, vocabulary, and listening exercises. Students can access this module by means of a form in which they can select the level and category of the exercises (e.g., verbal tenses for the grammar exercises, topic for the vocabulary exercises, or title of the text for the listening exercises).

Grammar exercises (see Figure 2 top-left) consist of filling gaps in sentences with a verb, adjective, adverb or other grammar elements from the topic and difficulty level initially selected. Vocabulary exercises (Figure 2 top-right) allow users to dictate or write words that are described by means of images displayed to the user. Listening exercises consist of two main parts (Figure 2 bottom). Firstly, the multimodal system reads the text selected by the user. Then, a set of related questions are presented to the user to evaluate their reading comprehension.

The Assessment module (see Figure 3) allows users to review the answers that they provided to the previously selected exercises, visualize the correct solution in case of errors, and know detailed statistics about the student's specific evolution using the system.

The personalization of the system is carried out by means of the "choose for me" functionality of the Practice module. This functionality takes into account the number of exercises in each category and level correctly solved by each user, so that the system can provide personalized suggestions and select the following

Fig. 2. Example of the different kinds of exercises offered by the system

exercises according to the errors found in previous interactions. To do this, the system manages the database containing the exercises completed by the user for the current category and difficulty level. The system also analyzes whether the number of mistakes made for the current difficulty level is highest than a specific threshold (initially predefined to the half of them). The objective is to personalize users' recommendations taking into their specific evolution with the different categories of exercises and difficulty levels.

Figure 4 shows a set of use cases for the personalization functionality of the system. In the first case, the system recommends trying additional exercises from the same difficulty level that the previously selected given that the student has not already successfully completed at least the half of them. In the second case, the system provides the same recommendation given that the number of mistakes is higher than a specific threshold. In the third case, the system suggests selecting the next difficulty level given that the results for the current level are satisfactory.

Fig. 3. Assessment module of the *Test Your English* system

Finally, the Contents management module allows administrator users to modify and insert new contents in the system. This module is based on different functionalities provided by the phpMyAdmin tool (`www.phpmyadmin.net`) to facilitate creating new exercises and editing or removing existing ones. The system comprises three main databases that contain the learning contents, multimodal elements and the history of the interaction respectively. The first database stores the questions and answers categorized in different topics. For each question, there is a text, optional multimedia contents (audio and video) and several answers. For each answer, there is also text and/or multimedia, as well as the positive and negative feedbacks and hints to be provided to the student in the case he/she selects the answer. For each question only one answer is assumed to be correct. The second database contains the visual rendering of the interface, and the third database stores the information about the previous interactions of the user with the system.

The objective is to facilitate including new questions and editing the existing ones. This way, different people can help in the development of the system without requiring an expert knowledge in conversational agents. For example, teachers and parents can include new questions in the database, and graphic designers can create attractive multimodal contents for the system and include them into the corresponding databases.

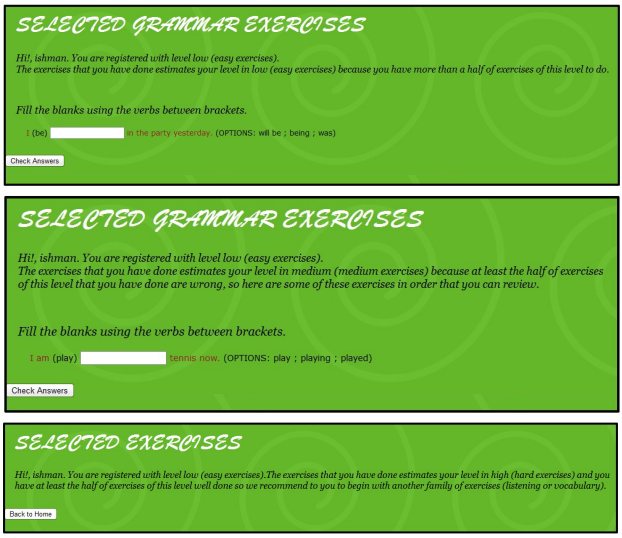

Fig. 4. Use cases describing the operation of the *choose for me* functionality

3 Preliminary Evaluation

We have already completed a preliminary evaluation of the developed pedagogic system. A total of recruited 25 users participated in the evaluation, aged 21 to 57 (mean 27.2), 17 male and 8 female. Prior to the evaluation, an assistant explained to the users what the system is about and how it is used. Then, they were given 30 minutes to get accustomed to the system, while the doubts were solved by a teacher and the assistant. To do this, each user had at his/her disposal a computer and had to wear a microphone headset. They were allowed to break up the interaction at any time for any reason. Then, each user freely interacted with the application during 10 sessions of at list 15 minutes.

At the end of the last session each user was required to complete the questionnaire shown in Table 1 was defined for the evaluation. The responses to the questionnaire were measured on a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree). The users were also asked to rate the system from 0 (minimum) to 10 (maximum) and there was an additional open question to write comments or remarks.

Also, from the interactions of the experts with the system we completed an objective evaluation of the application considering the following interaction parameters: i) Question success rate (SR). This is the percentage of successfully completed questions: system asks - user answers - system provides appropriate feedback about the answer; ii) Confirmation rate (CR). It was computed as the ratio between the number of explicit confirmations turns and the total of turns; iii) Error correction rate (ECR). The percentage of corrected errors.

Table 1. Questionnaire employed for the subjective assessment

Interaction experience and technical quality
IT01. The system is easy to use
IT02. The system provides adequate feedback
IT03. The system is helpful
IT04. The system offers enough interactivity
IT05. It is easy to know what to do at each moment
IT06. The amount of information that is displayed on the screen is adequate
IT07. The system is adapted to my learning degree
IT08. I would use the system again
Learning contents and didactic potential
LD01. The questions were easy to understand
LD02. The questions were easy to answer
LD03. The system help me to learn new things
LD04. The activities support significant learning
LD05. The feedback provided by the agent improves learning
LD06. The system encourages continuing learning after errors
LD07. The system made me appreciate my skills for learning English

Table 2. Results of the evaluation of the system

	Min / max	Average	Std. deviation
IT01	3/4	3.87	0.26
IT02	4/5	4.73	0.40
IT03	4/5	4.85	0.34
IT04	3/5	4.06	0.73
IT05	4/5	4.86	0.31
IT06	4/5	4.93	0.15
IT07	3/5	4.52	0.41
IT08	5/5	5.00	0.00
LD01	3/5	4.26	0.55
LD02	3/5	4.33	0.39
LD03	4/5	4.85	0.22
LD04	5/5	5.00	0.00
LD05	4/5	4.77	0.59
LD06	4/5	4.85	0.38
LD07	3/5	4.32	0.41

SR	CR	ECR
93.05%	17.25%	91.92%

The results of the questionnaire are summarized in Table 2. As can be observed, the system was rated fairly well and most of the users learned new contents and most of them would like to use again the system. The satisfaction with technical aspects was high, as well as the perceived didactic potential. The system was considered attractive and adequate and the users felt that the system is appropriate and the activities relevant. The global rate for the system was 8.7 (in the scale from 0 to 10).

Although the results were very positive, in the open question the users also pointed out desirable improvements. One of them was to make the system listen constantly instead of using the push-to-talk interface. In fact, an analysis of the main problems detected showed that most of these errors were due to the users not holding the push-to-talk button correctly and thus the input was cut, or because they used longer phrases or fillers which were not correctly processed by the system. However, in most cases, these problems could be overcome by

confirming or asking again for the data, as shown by the question success rate of 93.05%. Additionally, the approaches for error correction by means of confirming or re-asking for data were successful in 91.92% of the times when the speech recognizer did not provide the correct answer.

4 Conclusions

According to previous works, multimodal conversational agents can accelerate the learning process, facilitate access to education, personalize the learning process, and supply a richer learning environment. These important points are usually addressed by establishing a more engaging and adaptive relationship between the students and the system. In this paper, we have described the *Test Your English* multimodal conversational agent, which has been developed to provide this enhanced educative environment for second language learning. The system is comprised of different modules that cooperate to interact with students using speech and visual modalities, and adapt its functionalities taking into account their evolution and specific preferences.

Although there are currently many systems for students to learn a second language, most of them are designed to follow the same behavior for every student, not taking into account their specific evolution during the learning process. The experimental results show that the adaptation provided by our system and the natural communication that it provides have a very positive impact on the learning outcomes and satisfaction of the students. For future work, we plan to replicate the experiments with more students to validate these preliminary results and incorporate their suggestions.

Acknowledgements. This work was supported in part by Projects MINECO TEC2012-37832-C02-01, CICYT TEC2011-28626-C02-02, CAM CONTEXTS (S2009/TIC-1485).

References

1. Pieraccini, R.: The Voice in the Machine: Building Computers that Understand Speech. The MIT Press (2012)
2. Corchado, J., Tapia, D., Bajo, J.: A multi-agent architecture for distributed services and applications. Computational Intelligence 24(2), 77–107 (2008)
3. Rodríguez, S., de Paz, Y., Bajo, J., Corchado, J.M.: Social-based Planning Model for Multi-agent Systems. Expert Systems with Applications 38(10), 13005–13023 (2011)
4. Pon-Barry, H., Schultz, K., Bratt, E.O., Clark, B., Peters, S.: Responding to student uncertainty in spoken tutorial dialog systems. Int. Journal of Artificial Intelligence in Education 16, 171–194 (2006)
5. Wang, Y., Wang, W., Huang, C.: Enhanced Semantic Question Answering System for e-Learning Environment. In: Proc. AINAW, pp. 1023–1028 (2007)
6. Fryer, L., Carpenter, R.: Bots as Language Learning Tools. Language Learning and Technology 10(3), 8–14 (2006)

7. Cavazza, M., de la Camara, R.S., Turunen, M.: How Was Your Day? a Companion ECA. In: Proc. AAMAS 2010, pp. 1629–1630 (2010)
8. Kerly, A., Ellis, R., Bull, S.: CALMsystem: A Dialog system for Learner Modelling. Knowledge Based Systems 21(3), 238–246 (2008)
9. Griol, D., Molina, J., Sanchis, A., Callejas, Z.: A Proposal to Create Learning Environments in Virtual Worlds Integrating Advanced Educative Resources. Journal of Universal Computer Science 18(18), 2516–2541 (2012)