Developing multimodal conversational agents for an enhanced e-learning experience

David Griol, José Manuel Molina, 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

KEYWORD

Conversational agents
Multimodal interaction
Chatbots
Speech
E-learning

ABSTRACT

Conversational agents have become a strong alternative to enhance educational systems with intelligent communicative capabilities, provide motivation and engagement, and increment significant learning and helping in the acquisition of meta-cognitive skills. In this paper, we present Geranium, a multimodal conversational agent that helps children to appreciate and protect their environment. The system, which integrates an interactive chatbot, has been developed by means of a modular and scalable framework that eases building pedagogic conversational agents that can interact with the students using speech and natural language.

1 Introduction

Conversational agents [MCTEAR, M-F., et al. 2013], [PIERACCINI, R., 2012], [LÓPEZ-CÓZAR, R., et al. 2005] [MCTEAR, M-F., 2004] have became a strong alternative to enhance multi-agent systems with intelligent communicative capabilities and provide a more natural access to multiagent systems [PINZÓN, C.. et al. 2011].

According to Roda et al. [RODA, C., et al. 2001] and Kerly et al. [KERLY, A., et al. 2008a], the application of these technologies to the educative process should i) accelerate the learning process, ii) facilitate access, iii) personalize the learning process, and iv) supply a richer learning environment. These aspects can be addressed by means of multimodal conversational agents by establishing a more engaging and human-like relationship between the students and the system [CHOU, C-Y., et al. 2003].

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 [RODRÍGUEZ, S., et al. 2011].

For this reason, this kind of agents have been employed to develop a number of educational systems in very different domains, including tutoring [PON-BARRY, H., et al. 2006], conversation practice for language learners [FRYER, L., et al. 2006], pedagogical agents and learning companions [CAVAZZA, M., et al. 2010], dialogs to promote reflection and metacognitive skills [KERLY, A., et al. 2008b], or role-playing actors in simulated experiential learning environments [GRIOL, D., et al. 2012a], etc.

To successfully manage the interaction with users, these agents are usually developed following a modular architecture, which generally includes the following tasks: automatic speech recognition (ASR), spoken language understanding (SLU), dialog management (DM), database management (DB), natural language generation (NLG), and text-to-speech synthesis (TTS).

Due to this variability and the huge amount of factors that must be taken into account, these systems are difficult to implement and typically are developed ad-hoc, which usually implies a lack from scalability. In this paper we describe the *Geranium* system, a web-based interactive software with a friendly chatbot that can be used as a learning resource for children to study about the urban environment. The proposals for the development of the different modules of the system eases the construction of educative conversational by isolating pedagogic from the technical detail, 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.

The developed system, which is accessible using desktop and mobile devices, provides multimodal interaction instead of usually mediated simple text-based forms interaction, including spoken access and a visual representation through an animated bot with gestures and emotional facial displays. Also, the system infers a knowledge level for the students based on their answers, and encourages learners to engage in a dialog to reflect on their self-assessment and any differences between their belief and the expressed by the system.

The remainder of the paper is organized as follows. Section 2 describes related research in the development of educative systems that integrates conversational functionalities. Section 3 describes the main characteristics of the proposed architecture to develop educative conversational agents. Section 4 describes the Geranium educative system. Section 5 presents an evaluation of the system with teachers and children. Finally, conclusions and future work are presented in Section 6.

2 State of the art

The design of conversational agents has reached a maturity based on standards that pervade technology to provide high interoperability, which makes it possible to divide the market in a vertical structure of technology vendors, platform integrators, application developers, and hosting companies [PIERACCINI, R., et al. 2009], [GRIOL, D., et al. 2012b] .

The implementation and strategies of conversational agents employed in e-learning applications vary widely, reflecting the diverse nature of the evolving speech technologies. The conversations are generally mediated through simple text based forms [HEFFERNAN, N-T., et al. 2003], with users typing responses and questions with a keyboard.

Some systems use embodied spoken conversational agents [GRAESSER, A-C., et al. 2001] capable of displaying emotion and gesture, whereas others employ a simpler avatar [KERLY, A., et al. 2008c]. Speech output, using text to speech synthesis is used in some systems [GRAESSER, A-C., et al. 2001], and speech input systems are increasingly available [LITMAN, D-J., et al. 2004], [BOS, J., et al. 1999].

The most popular application of conversational agents to education are tutoring systems. Kumar et al. [KUMAR, R., *et al.* 2011] have shown that agents playing the role of a tutor in a collaborative learning environment can lead to over one grade improvement. Also some studies [ROSÉ, C-P., et al. 2001], [WANG, N., *et al.* 2008], [GRAESSER, A-C., et al. 2001] have evaluated the effect of task-related conversational behaviour in tutorial dialog scenarios; whereas the work in the area of affective computing and its application to tutorial dialog has focused on identification of student's emotional states [D'MELLO, S-K., et al. 2005] and using those to improve choice of task related behavior by tutors.

For example, the AutoTutor project [GRAESSER, A-C., et al. 2005] provides tutorial dialogs on subjects including university level computer literacy and physics. Another tutoring system employing dialog is Ms. Lindquist [HEFFERNAN, N-T., et al. 2003], which offers coached practice to high school students in algebra by scaffolding learning by doing rather than offering explicit instruction.

Also CycleTalk [FORBUS, K-D., et al. 2005] is an intelligent tutoring system that helps university students to learn principles of thermodynamic cycles in the context of a power plant design task. Similarly, ITSPOKE [LITMAN, D-J., et al. 2004] engages students in a spoken dialog to provide feedback and correct misconceptions for tutoring conceptual physics.

Other examples of natural language tutoring are the Geometry Explanation Tutor [ALEVEN, V. et al. 2004], where students explain their answers to geometry problems in their own words, and the Oscar conversational intelligent tutoring system [LATHAM, A., et al. 2012], which uses natural language to provide communication about specific topics with its users and dynamically predicts and adapts to a student's learning style.

Conversational agents as personal coaches integrate information about the domain of the application. For example, Grigoriadou et al. [GRIGORIADOU, M., et al. 2005] describe a system where the learner reads a text about a historical event before stating their position about the significance of an issue and their justification of this opinion. Similarly, in the CALM system [KERLY, A., et al. 2008b] the users answer questions on the domain, and state their confidence in their ability to answer correctly.

Systems of this kind are characterized by the possibility to represent and continuously update information that represents the cognitive and social users' state [WANG, Y., et al. 2007]. The main objective is to guide and monitor users in the learning process, providing suggestions and other interaction functionalities not only with the developed application but also with the rest of students. In order to achieve this goal, these applications usually integrate realistic and interactive interfaces.

Other systems provide a visual representation through an animated bot with gestures and emotional facial displays. These bots have shown to be a good interaction metaphor when acting in the role of counselors [MARSELLA, S-C., et al. 2003], [GRATCH, J., et al. 2005], personal trainers [BICKMORE, T-W. 2003], or healthy living advisors [DE ROSIS, F., et al. 2005]; and have the potential to involve users in a human-like conversation using verbal and non-verbal signals [CASSELL, J., et al. 2001].

Multimodal conversational agents are also a natural choice for many human-robot applications [SIDNER, C-L., et al. 2004], and are important tools to develop social robots for education and entertainment applications [DOWDING, J., et al. 2005], [EDLUND, J., et al. 2005]. A mobile robot platform that includes a spoken dialog system is presented in [GOROSTIZA, J-F., et al. 2005], which is implemented as a collection of agents of the Open Agent Architecture.

An interaction framework to handle multimodal input and output designed for face-to-face human interaction with robot companions is described in [LI, S., et al. 2007]. Authors emphasize the crucial role of the verbal behavior in human-robot interaction. A spoken dialog interface developed for the Jijo-2 mobile office robot is also described in [MATSUI, T., et al. 2003].

An intelligence model for conversational service robots is presented in [NAKANO, M. et al. 2011]. The model includes expert modules to understand human utterances and decide robot utterances or actions. It also enables switching and canceling tasks based on recognized human intentions, as well as parallel execution of several tasks.

Chatbots trained on a corpus have been proposed to allow conversation practice on specific domains [ABU-SHAWAR, B. et al. 2007]. This may be restrictive as the system may only "talk" on the domain of the training corpus, but the method may be useful as a tool for languages that are unknown to developers or where there is a shortage of existing tools in the corpus language.

These agents have also been proposed to improve phonetic and linguistic skills. For example, Vocaliza is a dialog application for computer-aided speech therapy in the Spanish language, which helps in the daily work of speech therapists that teach linguistic skills to Spanish speakers with different language pathologies [VAQUERO, C., et al. 2006]. In addition, the Listen system (Literacy Innovation that Speech Technology Enables) is an automated Reading Tutor that displays stories on a computer screen, and listens to children read aloud [MOSTOW, J., 2012].

Conversational agents may also be used as role-playing actors in simulated experiential learning environments. In these settings, the agent carries out a specific function in a very realistic way inside a simulated environment that emulates the real learning environment [GRIOL, D., et al. 2012a].

However, it is difficult to communicate with these agents whenever needed (i.e. when the user is not in front of a computer but he/she has the need to get suggestions and advices). Therefore, if we want to support student in a continuous way the personal advisor should be available and accessible also on a mobile device. Our work represents a step in this direction.

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 [CORCHADO, J., et al. 2008]. In the next section we describe the proposed architecture, which allows to easily developing a multimodal chatbot for pedagogical applications using a modular approach.

3 Proposed architecture to develop educative conversational agents

To ease the development of educative conversational agents, we contribute the architecture shown in Figure 1, which allows the development of multimodal applications with a chatbot that can interact with the student orally or through the graphical interface. In our architecture, each of the modules might be considered a “black box”, so that third-party software can be used to ease development. This way, commercial ASR and NLU systems can be employed easily within our architecture.

The systems implemented using the architecture generate personalized questionnaires including selected questions, perform the interaction by means of a conversational animated agent, carry out the corresponding analysis of the students’ answers, and provide them with the appropriate feedback.

The student employs the oral interface (which he/she may combine with the GUI) to provide responses to the questions posed by the system. Such oral responses are processed by the ASR, which outputs the most probable word sequence. Such sequence is then interpreted by the NLU module which generates a semantic representation.

Then, the User Answer Analyzer obtains the meaning of the user input from the NLU module and checks whether it corresponds to the correct answer defined in the database. Then, it calculates the percentage of success and the set of recommendations to be made to the student. This can be done by means of grammars in which the student’s answer is compared with the reference answer, assigning a specific score and answer each time a coincidence is detected.

With this information, the dialog manager decides the system response also considering the confidence measures provided by the ASR and NLU modules. Given that speech recognition is not perfect, one of the most critical operations of the design of the dialog manager is related to error handling. One common way to alleviate errors is to use techniques aimed at establishing a confidence level for the speech recognition result, and to use that for deciding when to ask the user for confirmation, or reject the hypothesis completely and re-prompt the user.

This way, the Dialog Manager might decide to confirm the input, ask again for the information or consider it as valid then acting accordingly to its correctness. The simplest dialog model is implemented as a finite-state machine, in which machine states represent dialog states and the transitions between states are determined by the user’s actions. Currently, the application of machine learning approaches to model dialog strategies is a very active research area [WILLIAMS, J., et al. 2007].

The system response decided by the Dialog Manager is presented to the student by means of the result generated by the Natural Language Generation and Text-to-Speech Synthesizer and the Multimodal Answer Generation modules. Natural language generation is the process of obtaining texts in natural language from a non-linguistic representation using predefined text messages. Then, the text-to-speech synthesizer transforms the text strings into acoustic signals. There is also the possibility of reproducing pre-recorded prompts stored in the database with generic messages.

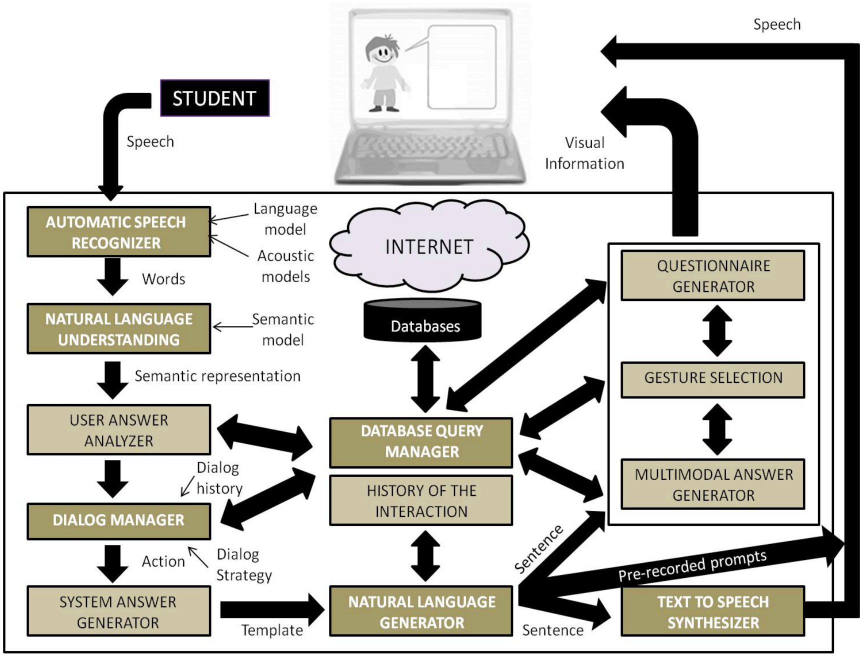


Fig. 1. Proposed architecture to develop educative conversational agents

The Multimodal Answer Generator is in charge of rendering the multimodal output of the chatbot by coupling the speech output with the visual feedback. The Gesture Selection module controls the incorporation of the animated character's expressions. In order to do so, it selects the appropriate gestures from the database and plays them to the user.

Finally, the Questionnaire Generator module manages the current interaction to dynamically modify the selected questions according to the information provided by the different modules in the application.

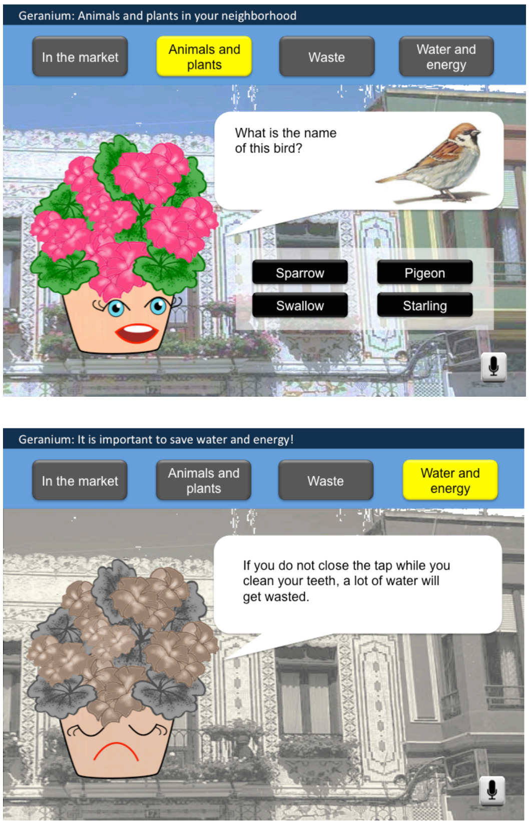
The architecture comprises three databases that contain the learning contents, multimodal expressions of the chatbot and the history of the interaction respectively. The first database stores the questions and answers categorized in different topics. For each question, there is a text, optional multimedia 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 chatbot's gestures and facial expressions, 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, indicating the corresponding responses of the chatbot and making possible the adaptation of the system to new domains. This way, different people can help in the development of the system without requiring an expert knowledge in dialogue systems. For example, teachers and parents can include new questions in the database, and graphic designers can create attractive animations for the chatbot and include them into the corresponding database.

4 The *Geranium* pedagogical system

The *Geranium* system has been developed with the main aim of making children aware of the diversity of the urban ecosystem in which they live, the need to take care of it, and how they can have a positive impact on it. The system has a chatbot named *Gera*, a cartoon that resembles a geranium, a very common plant in the Spanish homes.

Figure 2 shows two snapshots of the system. As can be observed, it has a very simple interface in which the chatbot is placed in a neighborhood. There are several buttons to select the type of questions, the chatbot has a balloon that shows the text, image or videos corresponding to the questions and possible answers, and there is a "push-to-talk" button that enables the oral input. As the chatbot changes its expressions, the background also changes. For example, Figure 2 shows the response of the system to an incorrect response, as can be observed, *Gera* is "sad" and the background has a grey color.


Geranium: Animals and plants in your neighborhood

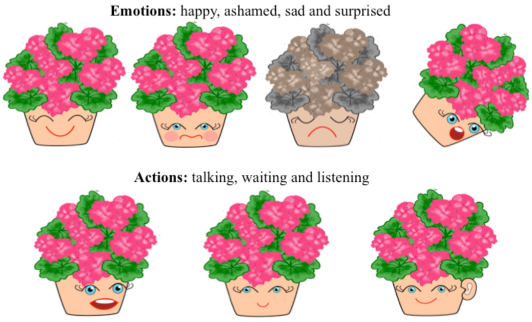
What is the name of this bird?

Geranium: It is important to save water and energy!

If you do not close the tap while you clean your teeth, a lot of water will get wasted.

Fig. 2. Snapshots of the response of the Geranium system to a correct and an incorrect answer

The chatbot poses questions to the children that they must answer either orally or using the graphical interface. Once an answer is selected, the system checks if it is correct. In case it is, the user receives a positive feedback and *Gera* shows a "happy" (usual case) or a "surprised" (in case of many correct questions in a row) face. If the answer selected is not correct, *Gera* shows a "sad" expression and provides a hint to the user, who can make another guess before getting the correct response. *Gera* has 7 expressions: happy, ashamed, sad, surprised, talking, waiting and listening, shown in Figure 3, which can also be extended by adding new resources to the chatbot expressions database.


Emotions: happy, ashamed, sad and surprised

Actions: talking, waiting and listening

Fig. 3. Facial expressions of the *Gera* chatbot

The speech input is activated with a push-to-talk button and the recognition hypotheses generated by the recognizer are passed to the SLU module for processing. Once the input has been processed, the dialog manager chooses the next system action for which a system answer is generated and synthesized. During oral communication, along with the speech and textual response, the chatbot provides visual feedback with the described facial expressions.

The natural language understanding and dialog management modules have been developed according to the Voice Extensible Markup Language (VoiceXML) [DOMINGUEZ, K., 2014], defined by the W3C as the standard for implementing interactive voice dialogs for human-computer interfaces. VoiceXML applications are usually based on the definition of grammars for the SLU module. In our system, grammars are encoded following the Java Speech Grammar Format (JSGF, www.w3.org/TR/jsgf/), supported by any VoiceXML platform.

In the *Geranium* system, for each question type there is a grammar template with the usual structure of the responses, and a new grammar is dynamically generated that makes use of the template and contains the exact response options for the actual question. Each of the options has an assigned code which is used also in the GUI and makes it possible to easily control the synchronization between the different modalities employed. Also the template makes it possible to maintain the same structure for the responses to similar questions. This facilitates the system usage, as it is easier for the users to know what the system expects.

The inclusion of static and dynamic grammars makes it possible to implement flexible dialogs with a wide range of possibilities from system-directed initiative to mixed initiative. Static grammars deal with information that does not vary over time, including the templates for the different question types and the grammars to control the exercises flow (e.g. to repeat a question, ask for help or select an option in the menu). Dynamic grammars include information that varies with time and make it possible to easily update and increment the learning contents.

This way, the SLU and dialog manager modules are simplified by generating a VoiceXML file for each specific question in the system, including the corresponding system prompt and the grammar that defines the valid user's inputs for the prompt. Regarding dialog management, all the events in the application are controlled using JavaScript. The dialog manager selects the next system prompt (i.e. VoiceXML file) by following a JavaScript program that determines the order for the set of questions, which is based on VoiceXML finite states.

The activities proposed by the agent are grouped in four topics: "in the market", "animals and plants", "waste", and "water and energy". In the first topic, the children are asked about fruits and vegetables, the plants where they come from and the seasons when they are collected. The second topic comprises questions about animals and plants that live in the city, showing photographs, drawings and videos of birds, flowers, trees and leaves and how they change or migrate during the year.

In the third topic, the children are asked about recycling, differentiating between the different wastes and the suitable containers. Finally, the fourth topic deals with good practices to save water and energy at home. Currently, there are 20 questions per category (a total of 80 questions), although the system can be extended adding new questions with their respective answers to the database.

Figure 4 shows an example to generate a VoiceXML file and grammars corresponding to the snapshot shown in Figure 2, in which the student is asked to tell the name of a bird. As it can be observed, the VoiceXML file corresponding to each one of the questions can include more than one system prompt. To do this, a prompt counter is defined to track the number of times the prompt has been used since the form was entered.

The values for the properties are computed dynamically taking into account the dialog history. The question template is provided in the *what_is_it.jsgf* file, whereas the exact options for the response are in the question *ex1b3.jsgf* grammar, which is generated at run time. Thus, the student utterances can be short, but also more elaborated, as for example: "Starling", "A sparrow", "It looks like a pigeon", "I think it is a sparrow", or "I am not sure but it can be a swallow".

VoiceXML file

```
<?xml  version=''1.0''  encoding=``UTF-
8''?>
<vxml xmlns=''www.w3.org/2001/vxml''
xmlns:xsi=''www.w3.org/2001/
XMLSchema-instance''
xsi:schemaLocation=''www.w3.org/2001/vxml
www.w3.org/TR/voicexml20/vxml.xsd''
version=''2.0''
application=''Gera.vxml''>
<form id=''ex1b3_form">
<grammar type=``application/x-jsgf''
src=''/grammars/mainGera.jsgf''/>
<field name=``question_ex1b3''>
<grammar type=``application/x-jsgf''
src=''/grammars/question_ex1b3.jsgf''/>
<prompt> What is the name of this bird?
</prompt>
<prompt count=''1''> Tell me the name of
the bird.
</prompt>
<prompt count=''2''> If you think that
the bird in the picture is an eagle,
just say eagle.</prompt>
<help> It is a little brown bird that
lives in your neighborhood and eats
seeds and insects. </help>
<noinput> Have a go! You are very
familiar to this little bird.</noinput>
<filled>
<nomatch> Ohh ohh! I did not get that.
Remember the options are: sparrow,
pigeon, swallow or starling. Please try
again!
```

```
</nomatch>
<return namelist=``follow_question.js''/>
</filled></field></form>
</vxml>
```

Grammar files

```
File question_ex1b3.jsgf
#JSGF V1.0;
grammar question_ex1b3;
public <question_ex1b3> =
<.../grammar_templates/what_is_it.jsgf#
what_is_it> {out.opt = rules.options;};
<options> = sparrow {this.out=''0'';} |
pigeon  {this.out=''1'';}  |  swallow
{this.out=''2'';}       |      starling
{this.out=''3'';};

File what_is_it.jsgf
#JSGF V1.0;
grammar what_is_it;
public  <what_is_it>  =  [<pre_answer>]
<options>;
<pre_answer> = [<certainty>] [<belief>]
[<phrase>];
<certainty>  =  ''Of  course''  |  ''For
sure'' | ''I am sure'' | ''I know'' |
(''I am not sure'' [but]) | (''I do not
know'' [but])
<belief> = I (think | believe);
<phrase>  =  [it  (is  |  ''can  be''  |
''might be'' | ''could be'' | ``may be''
| ''looks like'' | ''seems [to be]'')]
(a | an)
```

Fig. 4. Example VoiceXML file and corresponding grammars for the document

In addition, we have considered different functionalities that allow the adaptation of the system taking into account the current state of the dialog as well as the characteristics of each user. We have captured the main VoiceXML events: *noinput* (the user does not answer in a certain time interval or it was not sensed by the recognizer), *nomatch* (the input did not match the recognition grammar or was misrecognized), and *help* (the user explicitly asks for help).
Regarding the graphical user interface, the system answer generator produces the HTML output for the GUI and the template to be used by the natural language generator to obtain the lexical form of the next system prompt, which is then synthesized. With respect to the input, the visual and oral modalities are synchronized by means of the codes assigned to the answers for each question, both in the HTML form and in the VoiceXML grammars.

5 Evaluation

A preliminary evaluation of the *Geranium* system has been already completed with the participation of 6 primary school teachers of the levels for 8, 9 and 10 years old children, who rated the naturalness and pedagogical potential of the system. Teachers were told to bear in mind that the system was aimed at children of the same age as their students. 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 experts 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:

- Question success rate (*SR*). This is the percentage of successfully completed questions: system asks - user answers - system provides appropriate feedback about the answer;
- Confirmation rate (*CR*). It was computed as the ratio between the number of explicit confirmations turns and the total of turns;
- Error correction rate (*ECR*). The percentage of corrected errors.

Technical quality	
TQ01	The system offers enough interactivity
TQ02	The system is easy to use
TQ03	It is easy to know what to do at each moment.
TQ04	The amount of information that is displayed on the screen is adequate
TQ05	The arrangement of information on the screen is logical
TQ06	The chatbot is helpful
TQ07	The chatbot is attractive
TQ08	The chatbot reacts in a consistent way
TQ09	The chatbot complements the

	activities without distracting or interfering with them
TQ10	The chatbot provides adequate verbal feedback
TQ11	The chatbot provides adequate non-verbal feedback (gestures)
	Didactic potential
DP01	The system fulfils the objective of making children appreciate their environment
DP02	The contents worked in the activities are relevant for this objective
DP03	The design of the activities was adequate for children of this age
DP04	The activities support significant learning
DP05	The feedback provided by the agent improves learning
DP06	The system encourages continuing learning after errors

Table. 1. Questionnaire employed for the subjective assessment of the educative conversational agent

The results of the questionnaire are summarized in Table 2 As can be observed from the responses to the questionnaire, the satisfaction with technical aspects was high, as well as the perceived didactic potential. The chatbot was considered attractive and adequate and the teachers felt that the system is appropriate and the activities relevant. The teachers also considered that the system succeeds in making children appreciate their environment. The global rate for the system was 8.5 (in the scale from 0 to 10).

Although the results were very positive, in the open question the teachers also pointed out desirable improvements. One of them was to make the system listen constantly instead of using the push-to-talk interface. However, we believe that this would cause many recognition problems, taking into account the unpredictability of children behavior. Also, although they considered the chatbot attractive and its feedback adequate, they suggested creating new gestures for the chatbot to make transitions smoother.

The results of the objective evaluation for the described interactions show that the developed system could interact correctly with the users in most cases, achieving a question success rate of 96.56%. The fact that the possible answers to the questions are restricted made it possible to have a very high success in speech recognition. Additionally, the approaches for error correction by means of confirming or re-asking for data were successful in 93.02% of the times when the speech recognizer did not provide the correct answer.

Question	Min/max	Avg.	Std. dev.
TQ01	3/5	4.17	0.69
TQ02	3/4	3.67	0.47
TQ03	4/5	4.83	0.37
TQ04	5/5	5.00	0.00
TQ05	4/5	4.67	0.47
TQ06	4/5	4.83	0.37
TQ07	4/5	4.83	0.37
TQ08	4/5	4.50	0.50
TQ09	4/5	4.83	0.37
TQ10	4/5	4.67	0.47
TQ11	3/5	4.50	0.76
DP01	5/5	5.00	0.00
DP02	4/5	4.67	0.47
DP03	4/5	4.83	0.37
DP04	5/5	5.00	0.00
DP05	4/5	4.67	0.47
DP06	4/5	4.83	0.37

SR	CR	ECR
93.02%	17.25%	91.92%

Table. 2. Results of the subjective and objective assessment of the educative conversational agent

6 Conclusions and future work

In this paper we have described the *Geranium* conversational agent, a web-based interactive software with a friendly chatbot that can be used as a learning resource for children to study about the urban environment. The system has been developed using an architecture to cost-effectively develop pedagogic chatbots. This architecture is comprised of different modules that cooperate to interact with students using speech and visual modalities, and adapt their functionalities taking into account their evolution and specific preferences.

We have carried out an evaluation of the *Geranium* system with primary school teachers to assess its ease of use and its pedagogical potential. The study showed a high degree of satisfaction in the system appearance and interface, and the results were very positive with respect to its pedagogical potential.

For future work, we plan to replicate the experiments with children to validate these preliminary results, incorporate the suggestions provided by the teachers, and also compare the developed system with other pedagogical tools.

7 Acknowledgment

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

8 References

[ABU-SHAWAR, B. *et al.* 2007] ABU-SHAWAR, B., Atwell, E. Fostering language learner autonomy via adaptive conversation tutors. Proc. of Corpus Linguistics, 2007, pp. 1-8.

[ALEVEN, V. *et al.* 2004] ALEVEN, V., Ogan, A., Popescu, O., Torrey, C., Koedinger, K. Evaluating the Effectiveness of a Tutorial Dialog System for Self-Explanation. Proc. of 7th Int. Conference on Intelligent Tutoring Systems (ITS'04), 2004, pp. 443-454.

[BICKMORE, T-W. 2003] BICKMORE, T-W. Relational Agents: Effecting Change through Human-Computer Relationships. PhD Thesis, Media Arts & Sciences, Massachusetts Institute of Technology, 2003.

[BOS, J., *et al.* 1999] BOS, J., Klein, E., Lemon, O., Oka, T. The Verbmobil prototype system - a software engineering perspective. Journal of Natural Language Engineering. 5(1), 1999, 95-112.

[CASSELL, J., *et al.* 2001] CASSELL, J., Sullivan, J., Prevost, S., Churchill, E-F. Embodied Dialog systems, The MIT Press, 2001.

[CAVAZZA, M., *et al.* 2010] CAVAZZA, M., de la Camara, R-S., Turunen, M. How Was Your Day? a Companion ECA. Proc. of AAMAS'10 Conference, 2010, pp. 1629-1630.

[CHOU, C-Y., *et al.* 2003] CHOU, C-Y., Chan, T-W., Lin, C-J. Redefining the Learning Companion: the Past, Present and Future of Educational Agents. Computers & Education, 40, 2003, 255-269.

[CORCHADO, J., *et al.* 2008] CORCHADO, J., Tapia, D., Bajo, J.: A multi-agent architecture for distributed services and applications. Computational Intelligence 24(2), 2008, 77–107.

[DE ROSIS, F., *et al.* 2005] DE ROSIS, F., Cavalluzzi, A., Mazzotta, I., Noviell,i N. Can embodied dialog systems induce empathy in users? Proc. of AISB'05 Virtual Social Characters Symposium, 2005, pp. 1-8.

[DOMINGUEZ, K., 2014] DOMINGUEZ, K. VoiceXML 31 Success Secrets - 31 Most Asked Questions On VoiceXML - What You Need To Know. Emereo Publishing, 2014

[DOWDING, J., *et al.* 2005] DOWDING, J., Clancey, W-J., Graham, J. Are You Talking to Me? Dialogue Systems Supporting Mixed Teams of Humans and Robots. Proc. of AIAA Fall Symposium "Annually Informed Performance: Integrating Machine Listing and Auditory Presentation in Robotic Systems, 2006, pp. 22-27.

[EDLUND, J., *et al.* 2005] EDLUND, J., Gustafson, J., Heldner, M., Hjalmarsson, A. Towards human-like spoken dialog systems. Speech Communication, 50(8-9), 2008, 630-645.

[FORBUS, K-D., *et al.* 2005] FORBUS, K-D., Whalley, P-B., Evrett, J-O., Ureel, L., Brokowski, M., Baher, J., Kuehne, S-E. CyclePad: An articulate virtual laboratory for engineering thermodynamics. Artificial Intelligence, 114(1-2), 1999, 297-347.

[FRYER, L., *et al*. 2006] FRYER, L., Carpenter, R. Bots as Language Learning Tools. Language Learning and Technology, 10(3), 2006, 8-14.

[GOROSTIZA, J-F., *et al*. 2005] GOROSTIZA, J-F., Salichs, M-A. End-user programming of a social robot by dialog. Robotics and Autonomous Systems, 59(12), 2011, 1102–1114.

[GRAESSER, A-C., *et al*. 2005] GRAESSER, A-C., Chipman, P., Haynes, B-C., Olney, A. AutoTutor: An Intelligent Tutoring System with Mixed-initiative Dialog. IEEE Trans. in Education, 48, 2005, 612-618.

[GRAESSER, A-C., *et al*. 2001] GRAESSER, A-C., Person, N-K., Harter, D. Teaching Tactics and Dialog in AutoTutor. International Journal of Artificial Intelligence in Education. 12, 2001, 23-39.

[GRATCH, J., *et al*. 2005] GRATCH, J., Rickel, J., Andre, J., Badler, N., Cassell, J., Petajan, E. Creating Interactive Virtual Humans: Some Assembly Required. Proc. of IEEE Conference on Intelligent Systems, 2002, pp. 54-63.

[GRIGORIADOU, M., *et al*. 2005] GRIGORIADOU, M., Tsaganou, G., Cavoura, T. Dialog-Based Reflective System for Historical Text Comprehension. Proc. of Workshop on Learner Modelling for Reflection at Artificial Intelligence in Education, 2003, pp. 238-247.

[GRIOL, D., *et al*. 2012a] GRIOL, D., Molina, J., Sanchis, A., Callejas, Z. A Proposal to Create Learning Environments in Virtual Worlds Integrating Advanced Educative Resources. JUCS Journal 18(18), 2012, 2516–2541.

[GRIOL, D., *et al*. 2012b] GRIOL, D., Molina, J., Sanchis, A., Callejas, Z. Bringing together commercial and academic perspectives for the development of intelligent AmI interfaces. Journal of Ambient Intelligence and Smart Environments, 4(3), 2012, 183-207.

[HEFFERNAN, N-T., *et al*. 2003] HEFFERNAN, N-T. Web-Based Evaluations Showing both Cognitive and Motivational Benefits of the Ms. Lindquist Tutor. Proc. of Int. Conference on Artificial Intelligence in Education, 2003, pp. 115-122.

[KERLY, A., *et al*. 2008a] KERLY, A., Ellis, R., Bull, S. Dialog systems in E-Learning. Proc. of AI'08, 2008, pp. 169-182.

[KERLY, A., *et al*. 2008b] KERLY, A., Ellis, R., Bull, S. CALMsystem: A Dialog system for Learner Modelling. Knowledge Based Systems, 21(3), 2008, 238-246.

[KERLY, A., *et al*. 2008c] KERLY, A., Bull, S. Children's Interactions with Inspectable and Negotiated Learner Models. Proc. of Int. Conf. on Intelligent Tutoring Systems, 2008, pp. 132-141

[KUMAR, R., *et al*. 2011] KUMAR, R., Rose, C-P. Architecture for Building Dialog systems that Support Collaborative Learning. IEEE Trans. Learn. Technol. 4(1), 2011, 21-34.

[LATHAM, A., *et al*. 2012] LATHAM, A., Crockett, K-A., McLean, D., Edmonds, B. A conversational intelligent tutoring system to automatically predict learning styles. Computers & Education. 59(1), 2012, 95-109.

[LI, S., *et al*. 2007] LI, S., Wrede, B. Why and how to model multi-modal interaction for a mobile robot companion. Proc. of AAAI Spring Symposium on Interaction Challenges for Intelligent Assistant, 2007, pp. 71-79.

[LITMAN, D-J., *et al*. 2004] LITMAN, D-J., Silliman, S. ITSPOKE: An Intelligent Tutoring Spoken Dialog System. Proc. of Human Language Technology Conference: North American Chapter of the Association for Computational Linguistics, 2004, pp. 5-8.

[LÓPEZ-CÓZAR, R., *et al*. 2005] LÓPEZ-CÓZAR, R., Araki, M. Spoken, Multilingual and Multimodal Dialog Systems. Development and Assessment. John Wiley and Sons, 2005.

[MARSELLA, S-C., *et al*. 2003] MARSELLA, S-C., Johnson, W-L., Labore, C-M. Interactive pedagogical drama for health interventions. Artificial Intelligence in Education: Shaping the Future of Learning through Intelligent Technologies, 2003, pp. 341-348.

[MATSUI, T., *et al*. 2003] MATSUI, T., Asoh, H., Asano, F., Fry, J., Hara, I., Motomura, Y., Itoh, K. Spoken Language Interface of the Jijo-2 Office Robot. Springer Tracts in Advanced Robotics, 6/2003, 2003, 307-317.

[MCTEAR, M-F., 2004] MCTEAR, M-F. Spoken dialog technology. Springer, 2004.

[MCTEAR, M-F., *et al*. 2013] MCTEAR, M.F., Callejas, Z. Voice Application Development for Android. Packt Publishing, 2013.

[MOSTOW, J., 2012] MOSTOW, J. Why and How Our Automated Reading Tutor Listens. Proc. of Int. Symposium on Automatic Detection of Errors in Pronunciation Training (ISADEPT), 2012, pp. 43-52.

[NAKANO, M. *et al*. 2011] NAKANO, M., Hasegawa, Y., Funakoshi, K., Takeuchi, J., Torii, T., Nakadai, K., Kanda, N., Komatani, K., Okuno, H-G., Tsujino, H. A multi-expert model for dialogue and behavior control of conversational robots and agents. Knowledge-Based Systems. 24(2), 2011, 248-256.

[PIERACCINI, R., 2012] PIERACCINI, R. The Voice in the Machine: Building Computers that Understand Speech. The MIT Press, 2012.

[PIERACCINI, R., *et al*. 2009] PIERACCINI, R., Suendermann, D., Dayanidhi, K., Liscombe, J. Are we there yet? Research in Commercial Spoken Dialog Systems, Lecture Notes in Computer Science, 5729, 2009, 3-13.

[PINZÓN, C.. *et al*. 2011] PINZÓN, C., Bajo, J., de Paz, J., Corchado, J. S-MAS: An adaptive hierarchical distributed multi-agent architecture for blocking malicious SOAP messages within Web Services environments. Expert Systems with Applications 38(5), 2011, 5486-5499.

[PON-BARRY, H., *et al*. 2006] PON-BARRY, H., Schultz, K., Bratt, E-O., Clark, B., Peters, S. Responding to student uncertainty in spoken tutorial dialog systems. International Journal of Artificial Intelligence in Education. 16, 2006, 171-194.

[RODA, C., *et al*. 2001] RODA, C., Angehrn, A., Nabeth, T. Dialog systems for Advanced Learning: Applications and Research. In: Proc. of BotShow'01 Conference, 2001, pp. 1-7.

[RODRÍGUEZ, S., *et al*. 2011] RODRÍGUEZ, S., de Paz, Y., Bajo, J., Corchado, J.M. Social-based Planning Model for Multi-agent Systems. Expert Systems with Applications 38(10) , 2011, 13005–13023

[ROSÉ, C-P., *et al*. 2001] ROSÉ, C-P., Moore, J-D., VanLehn, K., Allbritton, A. A Comparative Evaluation of Socratic versus Didactic Tutoring. Proc of Cognitive Sciences Society, 2001, pp. 869-874.

[SIDNER, C-L., *et al*. 2004] SIDNER, C-L., Kidd, C-D., Lee, C., Lesh, N. Where to look: a study of human-robot engagement. Proc. of 9th Int. Conference on Intelligent user interfaces (IUI'04), 2004, pp. 78-84.

[VAQUERO, C., *et al*. 2006] VAQUERO, C., Saz, O., Lleida, E., Marcos, J., Canalís, C. Vocaliza: An application for computer-aided speech therapy in Spanish language. Proc. of IV Jornadas en Tecnología del Habla, 2006, pp. 321–326.

[WANG, N., *et al*. 2008] WANG, N., Johnson, L-W. The Politeness Effect in an intelligent foreign language tutoring system. Proc. of Intelligent Tutoring Systems (ITS'08), 2008, pp. 270 - 280.

[WANG, Y., *et al*. 2007] WANG, Y., Wang, W., Huang, C. Enhanced Semantic Question Answering System for e-Learning Environment. Proc of AINAW'07 Conference, 2007, pp. 1023-1028.

[WILLIAMS, J., *et al*. 2007] WILLIAMS, J., Young, S. Partially Observable Markov Decision Processes for Spoken Dialog Systems. Computer Speech and Language, 21(2), 2007, 393–422.