This is a version of the following article:

Griol D., Molina J.M., de Miguel A.S. (2014) The Geranium System: Multimodal Conversational Agents for E-learning. In: Omatu S., Bersini H., Corchado J., Rodríguez S., Pawlewski P., Bucciarelli E. (eds) Distributed Computing and Artificial Intelligence, 11th International Conference. Advances in Intelligent Systems and Computing, vol 290. pp219-226 Springer, Cham

DOI: 10.1007/978-3-319-07593-8_26

The *Geranium* System: Multimodal Conversational Agents for E-learning

David Griol, 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

Abstract. Many e-learning applications use conversational agents as means to obtain enhanced pedagogical results such as fostering motivation and engagement, incrementing 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, provides a modular and scalable framework that eases building pedagogic conversational agents that can interact with the students using speech and natural language.

Keywords: Conversational agents, multimodal interaction, chatbots, speech, e-learning.

1 Introduction

Conversational agents [1] have became a strong alternative to enhance multiagent systems with intelligent communicative capabilities and provide a more natural access to multiagent systems [2]. 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).

According to Roda et al. [3], educative technologies 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. For this reason, this kind of agents have been employed to develop a number of educational systems in very different domains, including tutoring [4], conversation practice for language learners [5], pedagogical agents and learning companions [6], dialogs to promote reflection and metacognitive skills [7], or role-playing actors in simulated experiential learning environments [8], etc.

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.

2 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 1 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 1 shows the response of the system to an incorrect response, as can be observed, *Gera* is "sad" and the background has a grey color.

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 2, which can also be extended by adding new resources to the chatbot expressions database.

The activities 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

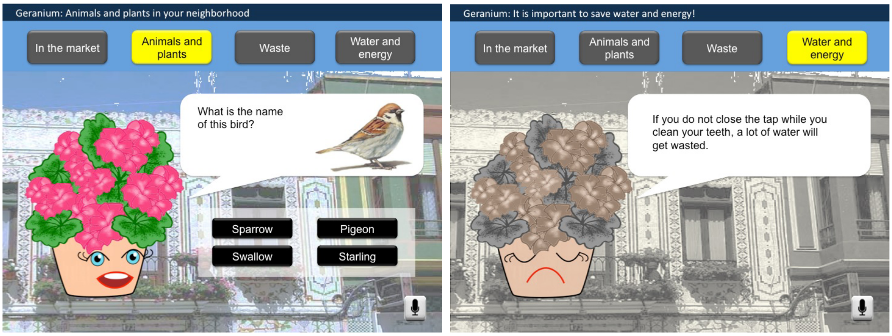

Fig. 1. Snapshots of the response of the *Geranium* system to a correct and an incorrect answer

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.

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 feedback with its expressions. When it is listening to the user, it shows the "listening" face, and if the answer selected was not understood, it shows an "ashamed" expression (see Figure 2). There are two additional expressions: "talking" and "waiting", to resemble the chatbot talking and waiting for the user to provide a response to a question, respectively.

The natural language understanding and dialog management modules have been developed according to the Voice Extensible Markup Language (VoiceXML, `www.w3.org/TR/voicexml20`), 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

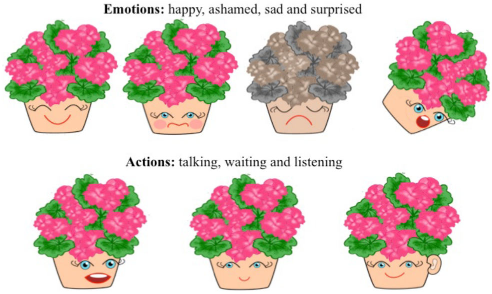

Fig. 2. Facial expressions of the *Gera* chatbot

between the different modalities employed. Also the template makes it possible to maintain the same structure for the responses to similar questions (e.g. to ask for the name of a bird, a plant or a fruit in a photograph) even when they belong to different categories (e.g. birds are in the category "animals and plants" and fruits are in the category "in the market"). 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 the JavaScript program that determines the order for the set of questions, which is based on the transitions summarized in Figure 3 and is based on VoiceXML finite states.

Figure 4 shows an example to generate a VoiceXML file and grammars corresponding to the snapshot shown in Figure 1, 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 *what_is_it.jsgf*, 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",

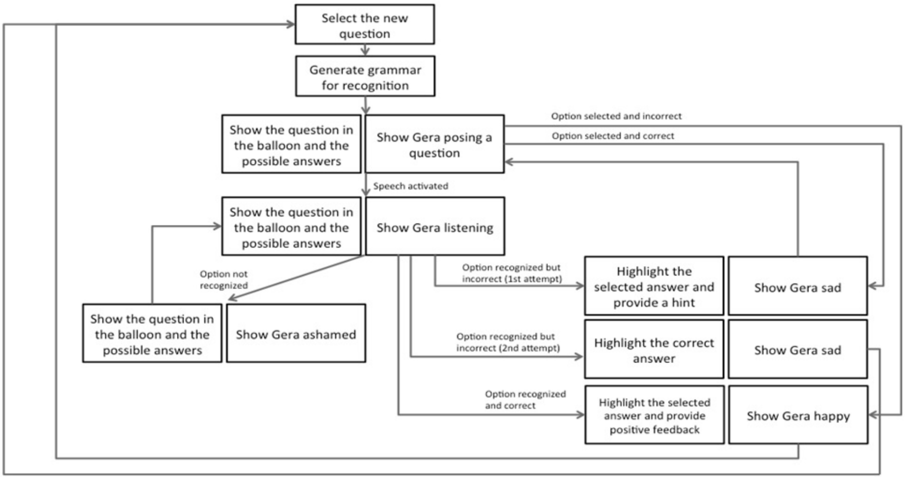

Fig. 3. Transitions to render the chatbot responses in *Geranium*

"A sparrow", "It looks like a pigeon", "I think it is a sparrow", or "I am not sure but it can be a swallow".

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.

3 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.

```
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)
```

```
<?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>
```

Fig. 4. Example VoiceXML document and corresponding grammars

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.

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).

Table 1. Percentage of different dialogs obtained

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
TQ010. The chatbot provides adequate verbal feedback
TQ011. The chatbot provides adequate non-verbal feedback (gestures)
Didactic potential
DP01. The system fulfills 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 2. Results of the evaluation of the system by experts

	Min / max	Average	Std. deviation
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
96.56%	13.00%	93.02%

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.

4 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.

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. 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), 5486–5499 (2011)
3. Roda, C., Angehrn, A., Nabeth, T.: Dialog systems for Advanced Learning: Applications and Research. In: Proc. of BotShow 2001, pp. 1–7 (2001)
4. Pon-Barry, H., Schultz, K., Bratt, E.O., Clark, B., Peters, S.: Responding to student uncertainty in spoken tutorial dialog systems. IJAIED Journal 16, 171–194 (2006)
5. Fryer, L., Carpenter, R.: Bots as Language Learning Tools. Language Learning and Technology 10(3), 8–14 (2006)
6. Cavazza, M., de la Camara, R.S., Turunen, M.: How Was Your Day? a Companion ECA. In: Proc. AAMAS 2010, pp. 1629–1630 (2010)
7. Kerly, A., Ellis, R., Bull, S.: CALMsystem: A Dialog system for Learner Modelling. Knowledge Based Systems 21(3), 238–246 (2008)
8. 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), 2516–2541 (2012)