Duarte Digital - An evaluation of educational and entertaining effectiveness

Ana Mendes, Filipe Martins, Luísa Coheur, João Paulo Neto, Antnio Serralheiro
L²F/INESC-ID Lisboa
Rua Alves Redol, 9, 1000-029 Lisboa, Portugal

Abstract

Duarte Digital is a conversational agent who answers questions about a famous work of Portuguese jewelery, Custódia de Belém. Duarte Digital has two primary goals: teach students and keep them motivated and interested in Custódia de Belém’s details and history. Being so, the main question addressed in this paper is: “Having a conversational agent, playing both educational and entertaining roles, how should the dialogue with the user be conducted”?

With the purpose of answering the previous inquiry, four dialogue strategies were developed and tested with several users. These dialogue strategies range from a situation where the user is given explicitly what (and how) questions should be posed to the agent, to another where no evidence about what the agent is expecting as input is provided.

Results suggest that clues should be provided to the user during the interaction for him/her to have a more effective educational and entertaining experience. Moreover, the analysis of the collected dialogues show that users treat agents as if they were human beings even if they communicate only through text messages: users do not need to see an agent making movements in order to do so.

1 Introduction

Recent years have witnessed an emerging interest in finding new ways of teaching. As important as the effective knowledge transfer between teacher and students, the amusement of the laters while learning is also taken very much under consideration. Edutainment, described in (Adams et al., 1996) as a blend of education and entertainment, is mainly associated with computer games, as well as with more powerful, flexible and intuitive user interfaces, which had their boost not long past. In this context, embodied conversational agents (ECAs) are determining roleplayers, since they allow humans to interact face-to-face with computer systems as if these were another human being.

Duarte Digital is a conversational agent, personified as a 17 years old boy, who answers questions about a famous work of Portuguese jewelery, named Custódia de Belém. Duarte Digital goals are to teach students about this piece of art, while keeping them entertained and bringing up their interest about the Portuguese history and culture.

Knowing that Duarte Digital would stay in the Portuguese National Museum of Ancient Art (MNAA) ¹, the question about the way users should interact with the agent, getting as much knowledge and information as possible, as well as keeping high their amusement levels, arose. We wondered how we should introduce Duarte to users and to keep them interested. Thus, we developed a set of four different dialogue strategies in order to evaluate the system’s impact in users. Nevertheless, instead of evaluating the entire system as a whole (paying attention to the agent’s speech recognition and synthesis modules and gestures), our aim is to know how the dialogue itself ² should be led so the agent's goals are achieved.

Being so, the question we want to address in this paper is: “Having a conversational agent, playing both educational and entertaining roles, how should the dialogue with the user be conducted”?

Four scenarios were implemented and each one

¹http://www.mnarteantiga-ipmuseus.pt/
²For the sake of simplicity and clarity, in this paper we use the word dialogue to express the notion of interaction flow, without speech; when the interaction assumes the existence of speech, through speech recognition and synthesis modules, the expression spoken dialogue is used.
was evaluated regarding its **edutainment** capabilities. Although all the scenarios allowed open dialogue, they all represented different approaches to dialogue. In the first scenario, a list of questions concerning what should be learned was presented to the user, explicitly guiding him/her through the dialogue. In the second scenario, user was confronted with a set of topics indicating the information he/she should learn about. In the third scenario no clue was given to the user regarding the informations he/she should obtain, but a hint driven dialogue was conducted. Finally, the fourth scenario was pure open dialogue, with no hints.

In order to evaluate the educational role of each scenario, a small questionnaire was made at the end of the interaction, with the goal of evaluating what the user really learned about Custdia de Belm. On the other hand, in order to evaluate the entertainment role of the application, user’s expectations/opinions were measured before after the interaction.

This paper continues as follows: Section 2 refers to related work done on the field of conversational agents and their evaluation; Section 3 presents the agent Duarte Digital, and its architecture; Section 4 introduces the different dialogue strategies we evaluated. In Section 5 the evaluation setup is described and in section 6 the results are presented. The paper finishes with some final remarks and future work.

## 2 Related Work

Research on conversational agents have lead experiments towards an “out of the laboratory” paradigm, which aim at investigating how real users interact with computational systems, in realistic scenarios.

Information kiosks can be seen as examples of this practice. Here agents are supposed to provide users with useful information concerning a specific topic, like SGT Blackwell (Leuski et al., 2006), which was originally presented in a real Conference scenario and aimed at giving answers to users’ questions about the Army, the Institute for Creative Technologies and the Conference itself.

The fairytale author Hans Christian Andersen (HCA) (Bernsen and Dybkjær, 2004) represents another example. This multimodal 3D animated swedish character, capable of engaging on spoken dialogues with the user, as well as understanding 2D gesture inputs, was developed to entertain and teach children about the life and work of HCA.

Efforts have been made in order to find how users interact with conversational systems. Previous experiments (de Almeida and Shigeki, 2002) with an interactive character as a virtual guide in a museum show that, when confronted with the option of interacting or not with the agent, users tend to chose the virtual character, nevertheless its limitations. It also revealed that the agent can play an important role in motivating users to explore the museum and its exhibitions. In addiction, the study conducted in (Kopp et al., 2005), focusing on Max, also a museum virtual guide, suggested that users attribute sociality to the agents, as they tend to see characters as human beings.

User testing and evaluation play a crucial part in every conversational system. Being so, various evaluation metrics, as well as tools and frameworks have been developed in order to assess users’ satisfaction and systems’ performance. The PARADISE framework (Walker et al., 1997), for instance, where the performance measure is dependent on both task success and dialogue costs.

As pointed out by (Paek, 2001), when evaluating system’s performance, there is no lack of dialogue evaluation metrics. Nevertheless, from all metrics we had the option to choose from and as far as we know, very few evaluations on conversational agents have been made only to derive how the dialogue should be conducted, disregarding the influence of the speech recognition and synthesis tools on the interaction.

In this paper we describe Duarte Digital and the ideas we followed in order to build the first prototype of an embodied conversational agent on a already functioning framework. Moreover, we evaluate different dialogue strategies and try to get to the conclusion of which one is the more appropriate for an agent that has both educational and entertaining goals.

## 3 Who is Duarte Digital?

Duarte Digital is a 17 years old student, who has done a research work for a school project, which focused on Custódia de Belém. Amazed by the story behind this famous work of art, and armed with his knowledge and great enthusiasm, he is now willing to teach everyone who wants to discover this piece of art and to know the many suprising stories of its origin and life route. Duarte Digi-
tal can answer questions about Custódia de Belém, and he is always trying to keep his audience interested on the details about this ancient jewel.

The previous description could easily be applied to a human being. However, Duarte Digital is not a flesh-and-bone person. He is a conversational agent, with anthropomorphic attributes, which aims at giving informations and teaching people about Custódia de Belém. Duarte Digital was, in fact, born this year, as the result of a need shown by the MNAA to our laboratory: Custódia de Belém is currently being repaired and not available in public exposition. Being so, and due to its “great value, for the country and Portuguese people”, something or someone should talk and present to visitors the work which some consider to be “the finest piece of jewelery ever made in Portugal”. Duarte Digital is the conversational agent developed with this purpose.

Duarte Digital is based on DIGA framework (Martins et al., 2008), originally created to support the deployment of task-driven spoken dialogue systems. DIaloG Assistant (DIGA) is domain-independent and composed by three main modules: an Input-Output Manager, that controls an Automatic Speech Recognition module (Meinedo, 2008), a Text-To-Speech module (Paulo et al., 2008) and a Virtual Agent Face (Viveiros, 2004); a Dialogue Manager (STAR), that interprets the user intentions and generates output messages (Mourão, 2005); and, a Service Manager, that provides a dialogue manager interface to execute the requested services, and an external application interface (Cassaca and Maia, 2002). The dialogue manager’s architecture (Mourão et al., 2004) is frame-based: every domain is described by a frame, composed by domain slots that are filled with user requests until a service can be executed.

Duarte Digital has a 3D face of a boy. The module responsible for the face converts the synthesized sentence phonemes to visemes³; by doing this, the information for lips movement is synchronized with the synthesized audio. Also, it allows us to express facial expressions, like a surprised or angry face (see Figure 1), which improves Duarte’s realism and anthropomorphic characteristics. The module receives as input VHML ⁴ (Beard, 2002) files with the sequence of actions to be performed by the agent’s face. The VHML files are built dynamically, as they aim at expressing the emotional state of the agent during the entire interaction. The emotional states are based on the number of correct/incorrect answers. Also, some yes/no movements ⁵ are provided. Altogether, these movements aim at making the interaction a fun and entertaining visual experience.

Although Duarte is implemented on DIGA, our agent does not follow the task-driven approach. Duarte Digital represents, instead, a first attempt to integrate dialogue and restricted domain question-answering (RDQA) in the same system. Being so, in this prototype, it was decided to model the RDQA as a service. This service is a front-end module to Duarte’s knowledge base, which contains approximately 500 questions and their related answers ⁶. Both questions and answers are focused on Custódia de Belém (its history, details, as well as lateral topics ⁷) and on

---

³A viseme is the visual representation of a phoneme and is usually associated with muscles positioned near the region of the mouth (Neto et al., 2006).

⁴Virtual Human Markup Language: http://www.vhml.org/

⁵By moving the head up-and-down or side-to-side

⁶The knowledge base was first built with the help provided by the MNAA and Faculty of Letters. They gathered a set of questions about Custódia de Belém and the text collections containing the answers. Also, they verified the correctness of every answer in the knowledge base.

⁷We considered lateral topics those indirectly related with Custódia de Belém. For instance, knowing that the gold in which Custódia de Belém was made came from Quiloa, a
Duarte’s life. When receiving a question from the user, the service searches for the most similar question in the knowledge base (based on the Levenshtein distance algorithm, applied to words), and returns one of its associated answers back to the user. If there is no similar question in the knowledge base, the agent replies an answer such as “I don’t know the answer to that question yet! But I will study about that topic!”.

Duarte Digital tries to catch user’s attention by using utterances, like “Did you know that Custódia de Belém has more than 6 kilograms of gold”?. After the user is aware of Duarte’s presence, their subsequent dialogue can be either of user-initiative or mixed-initiative.

With this in mind, Duarte Digital can engage, for instance, in a dialogue like the one presented in Figure 2. This Figure shows a real dialogue situation between Duarte and a user, that belongs to the corpus of interactions collected via a web-based interface. This web-based interface, fairly similar to an instant messaging client, allow us to enrich our interaction corpus (by gathering new dialogues and new vocabulary), and Duarte’s knowledge base (by collecting new questions). Also, this interface is used to perform Wizard-of-Oz experiments. The strategy #4 (Open Dialogue), fully described in section 4, is used in these interactions.

4 Four Dialogue Strategies

This section describes the dialogue strategies we explored. Moreover, it presents exactly what we aim at discovering after collecting the users’ interactions with Duarte Digital, using each one of the strategies.

#1 Questions List: In this first strategy, users are given a set of dissimulated questions about what they can ask the agent about the jewelery piece, e.g.,

If you want to know
what is Custódia de Belém,
who was Custódia de Belém’s author, or
when was Custódia de Belém made,
just ask Duarte Digital!

A set of examples are explicitly provided to the user, with what he/she can ask about Custódia de Belém.

By using this strategy, we aim at knowing if the user simply copies-and-pastes the examples into the interface or uses another vocabulary and question formulation, and if he/she merely asks the given questions or poses another questions to the system.

#2 Questions Topics: In this strategy, users are given a set of topics about which they can ask the agent about the jewelery piece, e.g.,

Definition: 
Author: 
Construction date: 

No other vocabulary is provided to the user, besides the keywords.

With this strategy, we want to know if the user asks questions about the given topics and how does he/she do that, if he asks Duarte about any other topics (which are not given).

#3 Hint Driven Dialogue: In this strategy, users are unconsciously guided in their dialogue with Duarte. Whenever an answer is given to the user, an hint for the user’s next question is also provided. The following example represents 4 different dialogues with Duarte, showing an application of this strategy:

User: Who ordered Custódia’s construction?
Duarte: King Manuel I ordered Custódia de Belém!
User: When was it?
Duarte: I don’t think I understood your question. Can you repeat that please?
User: In which year was she constructed?
Duarte: Custódia de Belém was built between 1503 and 1506. It took 3 years for this work of art to be ready.
User: What is she made of?
Duarte: Custódia de Belém is made of glass, enamel and gold.

Figure 2: Dialogue between Duarte and a user
Hello, I'm Duarte! I am making a project here at the Museum! By the way, do you know what is Custódia de Belém?

Well, Custódia de Belém is a monstrance, which means a receptacle in which the consecrated Host is exposed for adoration.

Custódia de Belém was made in the year of 1503 and its author only finished it in 1506.

The story about Custódia de Belém’s origins is written on the piece of art. Can you guess where?

Yes/no/it’s there in latin

So, lets see... The story is written in a latin inscription on the basis of Custódia! It means...

Custódia de Belém is made of glass, enamel and lots of gold. This last material came from very far away!

Who made Custódia de Belém?

Custódia de Belém was made by Gil Vicente. What else do you want to know?

Which were the materials used in Custódia de Belém’s construction?

Custódia de Belém is made of glass, enamel and gold.

With this strategy we want to know what is the important information about Custódia de Belém that the user wants to know about.

5 Evaluation Setup

The evaluation was adapted from the one conducted in (Laan et al., 1997). In our evaluation, users were asked to fill a form consisting of eight 5-point rating-scale items. This form, shown in Figure 3, is posed to the user immediately before and after the interaction with the agent: the first, aims at measuring users expectations; the late, at measuring users opinions concerning their previous experience with the agent.

Hints can be divided into two different categories: explicit, like in Duarte1 and Duarte3, and implicit, like in Duarte2 and Duarte4. Users can either follow the hint, or not. Nevertheless, the agent is prepared for both cases. In explicit hints, Duarte has a set of prompts independent from the user’s answer. Dealing with implicit hints is exactly the same as the default agent’s behavior.

With this strategy, we want to know if the user is really guided by the Duarte’s hints and which information does the user find important to ask about Custódia de Belém.

#4 Open Dialogue: Finally, in this strategy, the user-agent interaction is completely ruled by the user. The agent has no control on the dialogue and it merely answers what he is asked. The following represents an example of an user-agent dialogue, using this strategy:

<table>
<thead>
<tr>
<th>User</th>
<th>Who made Custódia de Belém?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duarte</td>
<td>Custódia de Belém was made by Gil Vicente. What else do you want to know?</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>User</th>
<th>Which were the materials used in Custódia de Belém’s construction?</th>
</tr>
</thead>
<tbody>
<tr>
<td>Duarte</td>
<td>Custódia de Belém is made of glass, enamel and gold.</td>
</tr>
</tbody>
</table>

Figure 3: User expectations/opinions form.

The evaluation goes as follows:

**Filling the expectations form:** In order to perceive the users expectations concerning the agent and the future dialogue, every item is given an “enthusiasm” pontuation, from 1 to 5. This scale is reverse on items 4 (interesting/not interesting) and 8 (pleasant/unpleasant).

The discretization of the users’ expectations allows us to group them into different categories: [very enthusiastic, enthusiastic, little enthusiastic, not enthusiastic]. By doing this, we can assure that every strategy is tested by an evenly distributed group of users, concerning their previous expectations about the dialogue.
Duarte-User interaction: 9 different facts about Custódia de Belém were previously collected. They represent the basic information people should know about the work of art and include, for instance, Custódia de Belém’s definition, author, weight, construction date and materials.

These facts are used on strategies #1 (Questions List) and #2 (Questions Topics) as dissimulated questions and topics, respectively. It should be noticed that, on strategy #3 (Hint Driven Dialogue), the dialogue was conducted so the user would be taught about every of the facts, and on strategy #4 (Open Dialogue) no hint was given about these facts.

Filling the opinions form: By comparing the results of both expectations and opinions forms (which are, in fact, equal, just differing on the time they are presented to the user) we want to find if the user has enjoyed the experience and whether the agent met the user’s initial expectations. Such comparison aims at evaluating how well our agent’s entertaining goals are met.

Answering the questionnaire: Users answer a questionnaire, similar to the one shown in Figure 4, with 9 questions, directly related with the 9 collected facts about Custódia de Belém. This questionnaire aims at evaluating how well our agent’s educational goals are achieved.

Table 1: Number of users in each of the expectations groups, before and after the interaction with Duarte.

<table>
<thead>
<tr>
<th></th>
<th>before</th>
<th>after</th>
</tr>
</thead>
<tbody>
<tr>
<td>not enthusiastic</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>little enthusiastic</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>enthusiastic</td>
<td>15</td>
<td>13</td>
</tr>
<tr>
<td>very enthusiastic</td>
<td>5</td>
<td>4</td>
</tr>
</tbody>
</table>

As previously mentioned, users tend to interact with the agent as if it would be another human being. This experiment shows that this phenomenon does not just happen when users are face-to-face with an agent with visual appearance; it also happens when the communication is done only through “text messages”. Moreover, and besides treating Duarte as an equal, by asking his opinions and feelings, some users also focused on real-world facts, which Duarte is not capable of understanding. These out-of-domain user sentences seriously compromised the dialogue during some interactions and affected users’ opinions. Like, for instance:

"When was she built?"
"Custódia de Belém was built between 1503 and 1506. It took 3 years for this work of art to be ready."
"Oh well, it was quicker than the metro line between Alameda and São Sebastião!"

Also, users in general were not motivated to ask questions about Custódia de Belém. It will not be erroneous to say that the particular domain in which an agent is proficient highly affects the user reactions to it.

Some users filled the expectations form with the same value for every item. However, no one did this in the opinions form. This indicates that, when confronted with the evaluation approach we followed, users are much more opinionated after the
interaction, since they know clearer what is it being ask and they have a stronger and well formulated opinion about the system.

Concerning each of the strategies, some conclusions can also be derived from the experiments.

It can be seen from Figure 5 that all strategies got higher results on the educational level after the interaction, except the Open Dialogue. In fact, a careful analysis of the dialogues in which this strategy was followed, suggests that users are not aware about what to ask to the agent, even if the image of Custódia de Belém is presented in the interface. Also, their dialogue is not driven by a strong logical path and they seem to digress.

Figure 5: Average points given by users to the educational item, before and after the interaction with Duarte.

Like in the Open Dialogue, users gave lower punctuation to the usefulness item in the Hint Driven Dialogue after the interaction. We believe that the number of hints is very small and Duarte’s capability to continue the dialogue after the user utterance/sentence is still very limited.

As pointed out by (Coheur et al., 2008), users are guided by explicit clues in the dialogue. In fact, in both strategies Questions List and Questions Topics, the users’ questions focused, in average, on more than 4.5 out of the 9 provided topics about Custódia. Because of this happening, users answered correctly more questions than those who followed the Open Dialogue and Hint Driven Dialogue strategies.

Figure 6 also shows the impact of giving users clues to guide them on the dialogue: the average time each user spent interacting with Duarte was considerably higher in Questions List and Questions Topics strategies.

Finally, we should point out that, although Duarte had knowledge to deal with slang and informal expressions, there were some words the agent was not expecting. Due to the fact that the evaluation was conducted via a web-based text interface, many users typed in words used typically in online chats and messenger conversations (like axas instead of the correct achas"). These words misled the agent and affected the dialogue. However, this situation is of little importance, since the agent is intended to make use of an automatic speech recognition module in real situations. In such situations, the input vocabulary to the agent will be constituted only by correctly spelled words.

7 Conclusions

This paper introduces Duarte Digital, a conversational agent to appear in the National Museum of Ancient Art. Its goal is to teach students about a famous work of Portuguese jewelery: Custódia de Belém. The paper presents and describes four different dialogue strategies which were developed in order to assess how an agent, with both educational and entertaining goals, should conduct the interaction with users. The evaluation of the four strategies suggests that, to motivate users for discovering a topic in which they do not seem to be interested about and about which they do not know what to ask, an approach that gives clues during the interaction should be followed.

As future work, we would like to continue performing more user tests on dialogue approaches, having in mind the results we got in this evaluation. In particular, we want to further explore the Hint Driven strategy, as we still believe that this can be a good approach: guiding users in their dialogue with the agent.

We intend to take Duarte to junior and secondary schools, where we can evaluate the agent.
interaction with younger users. According to the results we got in this evaluation, we believe that an interaction where children can get involved with the character and the object of interaction (by means, for instance, of a written short story or a mini interactive game), would be the best approach. It is our objective to test this hypothesis in a near future, and compare it with this evaluation.

Also, we want to increase Duarte’s visual feedback, providing to the user information about the agent’s inner state: for instance, if Duarte has built various different interpretations for one user utterance, he should look confused. We also want to annotate user corpus with emotional information, to be combined with the agent’s own emotional state.

Finally, and since Duarte Digital is available at a museum, often visited by foreign visitors, it would be interesting if the agent could “understand” and “speak” other widely spoken languages, besides Portuguese. Being so, our future plans include the agent adaptation to both Spanish and English languages.

References


