User Preferences for Effective Task Completion while Interacting with a Robotic Conversational Animated Face

Yandí Ricaño
ITESM Cuernavaca
Reforma #182-A
Temixco Morelos 62589
Mexico
A00378827@itesm.mx

Jaqueline Hernandez
ITESM Cuernavaca
Reforma #182-A
Temixco Morelos 62589
Mexico
A00964015@itesm.mx

Luis Enrique Sucar
ITESM Cuernavaca
Reforma #182-A
Temixco Morelos 62589
Mexico
esucar@itesm.mx

Oscar Mayora
CREATE-NET
Via Solteri 38
38100, Trento
Italia
omayora@create-net.it

***Abstract* – We present a study about the preferences that users may have while interacting with a robot by means of a conversational animated interface, and the impact in task completion. An animated face was developed to simulate Spanish visems, and integrated with a conversational system. This system is designed for a robot to give information about a site using speech in an interactive dialogue. We performed a usability study to compare the efficiency, effectiveness and interaction experience of the conversational system, with and without the animated face. The results show that the interaction with an animated face seems more inviting to the user, it inspires confidence and makes easier the communication. However, it may distract the user from the main objective of the system, and could make the interaction slower due to the high computational resources required to animate a realistic face.**

Index Terms – facial animation, usability analysis

I. Introduction

Research in human-computer interaction (HCI) tries to make more efficient the communication between humans and machines. The study of efficient machine interfacing is essential to have a natural and intuitive human-machine communication.

Humanizing computer interfaces has long been a major goal of both computer users and HCI practice [1]. The human face is an extremely visible aspect of a person which, in social interaction, serves the two major functions of identification and communication. Besides, it is also of great importance for communication in terms of both speech and facial expression [2]. For this reason, the human face has attracted the attention of different areas as varied as psychology, criminology and computer graphics.

Computer simulation of human faces capable of reflecting mouth movements has been a flourishing research area for a long time [3]. Three dimensional modeling and animation of the human visage has been a major research field in human animation even though the modeling and animation of the human face is one of the most difficult tasks in computer graphics.

Even if computer interfaces using facial animation are being used more and more every day, very few researchers have studied the effect and the convenience of the use of an animated face in HCI, and in particular, in human-robot interaction [4]. The main motivation of this work is to study the efficiency, effectiveness and interaction experience that users may perceive while communicating with an animated face as a graphic conversational interface.

In this work we present a study about the preferences that users may have while interacting with a robot by means of a conversational animated interface and the impact in task completion, in comparison with the interaction without the animated face.

First the animated face and the voice interaction module will be described. Next we explain the integration of the animated face with the conversational system. Then we present the results obtained from the usability study performed on the interaction with and without the animated face. We conclude with some directions for future work.

II. Jessica: an animated Face

Jessica [5] is an animated face that is based on a conventional pseudo-muscle model, adapted for direct use with MPEG-4 Facial Points (FPs). This face is built up by including an underlying solid structure that cannot be penetrated. This structure is particularly important for the realistic synthesis of facial dynamics outside the FPs, and provides the model with a sense of volume that is absent in other low cost approaches. Jessica contains a mesh of 876 triangles and 28 muscles to allow facial expressions and movement. Fig. 1 shows *Jessica* as displayed in the screen.

The facial characteristics of *Jessica* (eyes, mouth and nose) were done measuring the distances of points based on MPEG-4. Each one of 28 muscles of the face is associated to the corresponding FPs and a predetermined area of influence. The area of influence is a list of the vertices that are affected by each movement of the muscle. The list of vertices associated to each muscle was determined once, but it can be modified to create new models and face movements.

Fig. 1. *Jessica*, the animated face, in neutral state.

A. Jessica adjustments for Spanish dialogues

To make *Jessica* speak Spanish, the files that contain Jessica's Facial Animation's Parameters (FAPs) were modify increasing the number of frames for each file. In order to obtain the desired movement sequence for animation, the values of the mesh coordinates were modified by changing the *X*, *Y* and *Z* position numbers.

The mechanism for integrating a naturally-appealing speech interface in the animated face is based on the animation of the vowels, which constitute the most significant *visems* in Spanish. So the movements that simulate the modulation of the five Spanish vowels: *A, E, I, O, U,* were programmed for facial animation. Fig. 2 shows *Jessica* simulating the vowel *U*.

After the vowels animations are programmed, sequences of vowels are concatenated to simulate words. The same process was done with the sequences of words to simulate complete sentences. These sentences are used in the dialogues of the conversational system.

Fig. 2. *Jessica* simulating the Spanish vowel U.

III. Conversational System for Information Display

CSID is a *Conversational System fo Information Display* designed to be used in a robot which will serve as a guide in a certain location, initially at our campus (ITESM Campus Cuernavaca). The system's purpose is to provide information for visitors about the different areas of the Campus at their verbal request.

The system was developed under the *Voice eXtensible Markup Language* (VXML) framework [6], a markup language for development of voice applications under web environments. The dialog is constructed in a mixed initiative approach, where CSID first gives a welcome message and invites the visitor to join in for a guided tour of the campus. It encourages the visitor to ask for information on a specific section he wishes to know from a list of general areas. Once the choice is taken, the visitor receives a description of the facilities and staff in the Campus and may choose to see videos or images about them. The video or image chosen is then displayed using ASP technology on a screen by a projector.

The complete dialog structure is done by means of hierarchical menus, where the user encounters first a menu of general areas, and then each option can spread into other subareas. As the user interacts with the conversational system based on this hierarchical structure, he can access the specific areas of interest and visualize its corresponding detailed information.

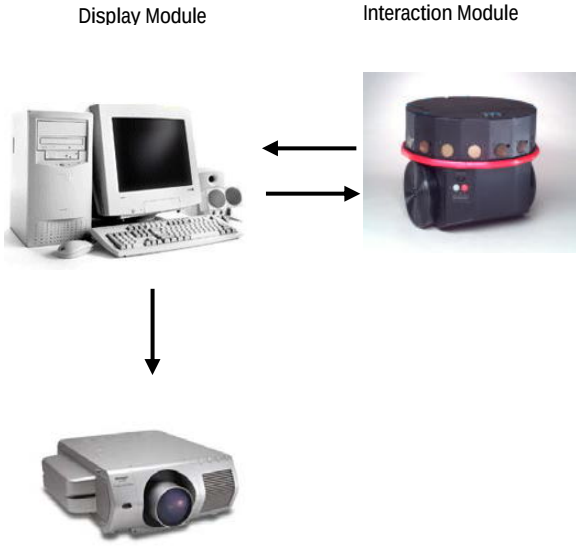

Fig. 3. CSID's functionality. The interaction module (which is on board the robot) receives the verbal requests from the user, which it sends to the display module that shows the information to the user.

The module in charge of the interaction with the visitor is located onboard of the robot, and the one that handles the display of images and videos is located on an external server. Both of these modules communicate using Active Server Pages (ASPs) and Microsoft's Internet Information Service (IIS) through a local wireless network. The

interaction module registers the user's request and uses a set of rules to send the appropriate information to the display module, which will respond with a confirmation message back to the other module in order to proceed with the interaction. This translates into another request from the user, where it may be asked if he wishes to see another video or image from the same area that was selected, visit a different one or exit the tour. Fig. 3 shows CSID's functionality.

IV. Communicating Jessica With CSID

The main goal of integrating the animated face with the conversational system, is to achieve a more complete and friendly interface for the visitor to interact with. The fact of having a face for the robot allows the visitor to improve the engagement while interacting with the robot. According to [7] this makes the user more comfortable as it provides an interface component.

It was decided that the animated face should be more like a cartoon character since a real human voice was recorded in order to make the utterances, as it will be explained later in this section. Boyce states [8] that having a system which sounds too much like a human can create higher expectations from the user which can then lead to errors. The main argument used by human-computer interaction researchers against anthropomorphic systems is that they make users have unrealistic expectations, which means they can expect the system to understand like a human. On this basis we chose to use a cartoon like animated face.

Regarding the dialog-face synchronization, different approaches were considered for the development of this application, including sockets and ASPs.

In this particular case, the authors decided that the most appropriate approach was the use of ASPs, as in the communication among the CSID's modules. This way the same line of work could be continued, since the ASPs developed for CSID provided a strong platform which is used to incorporate the animated face. The first step was to synchronize the animation of each sentence used in CSID together with the recorded voice (WAV) files. This process required the synchronization of each word with its corresponding animated movement. In order to achieve this, several video files were created in AVI format corresponding to each of the system sentences.

Since these files were quite voluminous and the needs of the project dictated an application that runs in real time, an inconvenience regarding time delays during the execution of these videos was encountered. The solution was to turn the videos into executable files, which made them lighter and thus more manageable.

By the inclusion of this avatar, the visitor's interaction is done directly with the animated face. This puts CSID's interaction module in a listening mode; once the animated face enounces its proper line, CSID processes the user's response as described earlier.

CSID's interaction module and the animated face module are both set to be onboard the robot, while the display module resides on an external computer connected to a projector. Fig. 4 shows a representation of how the modules function altogether.

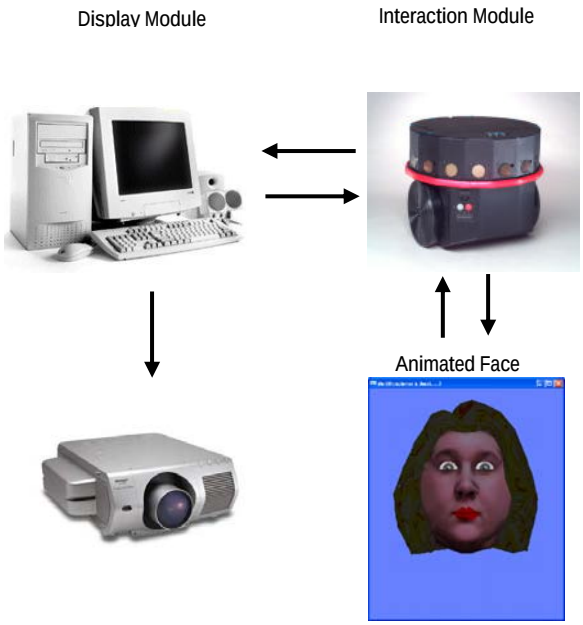

Fig. 4. The conversational system and the animated face working together. The user interaction is through the animated face that communicates with the interaction module, and this with the display module as described in Fig. 3.

V. Experiments and Results

To evaluate user interfaces in human robot interaction three parameters must be determined: efficiency, effectiveness and user's satisfaction [9]. Due to this, in this investigation we formulate a questionnaire to measure the parameters mentioned above for the analysis and evaluation of *Jessica* as a graphical interface within the conversational system.

The usability test consisted of the following stages:

1. **System introduction**: the system is explained to the user (objective and main tasks).
2. **1st Interaction**: the user interacts with the system without the animated face.
3. **2nd Interaction**: the user interacts with the system via the animated face, *Jessica*.
4. **Performance evaluation**: to measure the users' satisfaction with *Jessica*, a questionnaire is applied to the user, which considers the important aspects of the interaction.

In our experiments, all the users first interacted with the conversational system without the animated face, and later with *Jessica*. Although, in general, it is recommended that some interact first without the face and some with the face, in our case the objective is not to compare both options. We wanted the users first to be familiar with the system, and then evaluate the impact of the animated face in the interaction.

B. The Sample

For the purposes of this study, 26 persons were selected for the usability evaluation. All of them are master students at ITESM, Campus Cuernavaca, Mexico. There were 16 male and 10 female, and their age ranges from 22 to 36 years old.

C. Results

We know present the results of the questionnaire, for each aspect: efficiency, effectiveness and user's satisfaction. For the multiple choice questions, we show the number of responses per choice; and for the open questions, we present a representative sample of the most common responses.

Effectiveness

1. Did you accomplish the system's objective (obtain information about the Campus)?
    a. *Yes (26)*
    b. *No (0)*
2. Does the animated face help you obtain a better comprehension of the system?
    a. *Yes (20)*
    b. *No (6)*

Why yes?

- *When Jessica is being used, I have a better interaction with the system.*
- *The interaction is friendlier.*
- *Jessica produces a natural way of communication.*
- *Without Jessica the system's interaction looks impersonal.*

Why no?

- *They focus their attention in the information not in the animated face.*
- *The way that the information is displayed distracts the attention of the users.*
- *The animated face receives all users' attention.*

Efficiency

1. How much time did you spend along the interaction?
    a. *Less than expected (0)*
    b. *As expected (14)*
    c. *More than expected (12)*
2. What do you think is better and why?
    a. *Interact faster with the system without all interaction's details. (20)*

Why?

- Don't loose interest.

- It is better to have a faster interaction than a slower one.

-A slow interaction produces an unexciting interaction with the system.

- In a natural conversation (human-human) answers must be given quickly.

    b. *Interact slower with the system considering all interaction's details. (5)*

Why?

- If some details of the interaction are missing, this may produce confusion.

- Details are very important in a system that presents some kind of information.

- In a system that presents a virtual tour, it is better to have all interaction details.

One person answered that he would like to see the two versions to have an opinion.

User Satisfaction

1. Are you satisfied with the system?
    a. *Highly (8)*
    b. *Average (18)*
    c. *Low (0)*
2. How do you prefer the system to look like (with or without an animated face)?
    a. *With the animated face (22).*
    b. *Without the animated face (2).*

Two persons answered that it doesn't matter.

3. Do you think that the animated face helps you have a better interaction with the system?
    a. *Yes (24).*
    b. *No (2).*
4. What do you think about the animated face?
    a. *It looks friendly but it needs more conversational movements.*
    b. *It avoids a tedious interaction.*
    c. *It is a good support tool.*
    d. *It helps to personalize the system but it may distract the user from the main objective.*
5. Mention some advantages and disadvantages of the animated face.
    - Advantages:
        It attracts the user to interact with the system.
        It produces the impression of being an attendant.
        The communication seems more personalized.
        It makes the systems realistic.
    - Disadvantages:
        It makes the system and the interaction slower.
        The human face is a distraction factor.
        An animated face is very difficult to characterize.

D. Analysis

The results reveal that according to user preferences, the animated face improves the understanding and the interaction of the system making it more realistic and friendly.

The speed appreciation in the interaction with the animated face was in general considered *normal*, but some people consider it *slow* due the time that *Jessica* takes to display. In consequence, the sample majority prefers a quick interaction with the system even if they lose some details, because in this way they do not lose interest.

More than 80% of the sample mentioned that they prefer the interaction with the animated face, because it improves the interaction. In general, the satisfaction produced by the interaction with **Jessica** was *average*, and not *high* as it was expected.

The sample opinion of the animated face was positive. Most mention that is friendly, that the coordination of the lip movements and the sound was very good, that makes people pay more attention, and that is perceived as a good support tool. But there were negative opinions too, like that the animated face was not good looking enough and caused user distractions.

Other advantages of using an animated face as a graphic interface are that the system seems more inviting to the user for interacting, it inspires confidence and makes easier the communication. The major disadvantage perceived by users is that *Jessica* distracts the user from the main objective of the system that is to give information about the Campus. Other disadvantage is that the animated face makes the system slower because it needs high computational resources for its execution.

VI. Conclusions and Future Work

We have presented a study about the preferences that users may have when interacting with a robot by means of an animated face. An animated face was developed to simulate Spanish *visems*, and integrated with a conversational system. This system is designed for a robot to give information about a site using speech in an interactive dialogue. We performed a usability study to compare the efficiency, effectiveness and interaction experience of the conversational system, with and without the animated face. The results reveal that according to user preferences, the animated face improves the understanding and the interaction of the system making it more realistic and friendly. However, the animated face requires high computational resources for a real-time simulation, and may distract the user form the main purpose of the interaction.

There are several possibilities for future work. Regarding the CSID, the use of Markov Decision Processes (MDPs) to optimize the dialog management is work in progress. The idea is to model the dialogue as an MDP, so that the states represent the possible users' requests, and the actions the robot responses. By solving the MDP we will obtain the optimal response given the user request, which we hope will make the dialogue more flexible and efficient. With respect to the animated face, *Jessica*, we want to extend the simulation to include all the *visems* in Spanish, so the interface could be more natural. A more formal and general methodology for usability analysis for human-robot interaction is an important task for future research.

References

[1] J. H. Walker, L. Sproull, R. Subramani. "Using a human face in an interface" Cambridge Research Lab, Digital Equipment Corporation One Kendall Square, Bldg 700 Cambridge, MA 02140 USA, 1994

[2] M Patel. "Colouration issues computer generated facial animation", Computer Group, School of Mathematical Sciences, University of Bath, Bath, Avon, U.K, 1995.

[3] P. Kakumanu, R. Gutierrez-Osuna, A. Esposito, R. Bryll, A. Goshtasby, O. N. García. "Speech driven facial animation". Departament of Computer Science and Engineering, Wright State University, Dayton OH. November, 2001.

[4] S. Palmiter. "The effectiveness of animated demonstrations for computer-based tasks: a summary, model and future research". Journal of Visual Languages and Computing. Vol. 4, 1993.

[5] R. Gutiérrez-Osuna, P. Kakumanu, A. Esposito, O.N. García, A. Borjorquez, J.L. Castillo and I. Rudomin. "Speech-driven facial animation with realistic dynamics", January, 2005.

[6] Voice eXtensible Markup Language VoiceXML Manual. Version: 1.00, March 2000 http://www.voicexml.org/specs/VoiceXML-100.pdf

[7] A. Green and K. Severinson. "Task-oriented dialogue for CERO: a user-centered approach". *Proceedings of Ro-Man'01 (10th IEEE International Workshop on Robot and Human Communication)*, Bordeaux-Paris, pp. 146-151, September 2001.

[8] S. Boyce. "Natural spoken dialogue systems for telephony applications". *Communications of the ACM* vol. 43, no. 9, pp. 29-34, September 2000.

[9] J. Scholtz, S. Bahrami, "Human-robot interaction: development of an evaluation methodology for the bystander role of interaction". National Institute of Standards and Technology, 2004.