Effectiveness and Usability of an Online Help Agent Embodied as a Talking Head

Jérôme Simonin
LORIA, CNRS, INRIA, UHP Nancy 1
BP 239, F54506 Vandœuvre-lès-Nancy, France, +33383593083
Jerome.Simonin@loria.fr

Noëlle Carbonell
LORIA, CNRS, INRIA, UHP Nancy 1
BP 239, F54506 Vandœuvre-lès-Nancy, France, +33383593083
Noelle.Carbonell@loria.fr

Danielle Pelé
France Télécom R&D
BP 91226, F35512 Cesson-Sévigné, France, +33299124111
danielle.pele@orange-ftgroup.com

ABSTRACT

An empirical study is presented which aims at assessing the possible effects of embodiment on online help effectiveness and attraction. 22 undergraduate students who were unfamiliar with animation creation software created two simple animations with Flash, using two multimodal online help agents, EH and UH, one per animation. Both help agents used the same database of speech and graphics messages; EH was personified using a talking head while UH was not embodied. EH and UH presentation order was counter-balanced between participants.

Subjective judgments elicited through verbal and nonverbal questionnaires indicate that the presence of the ECA was well accepted by participants and its influence on help effectiveness perceived as positive. Analysis of eye tracking data indicates that the ECA actually attracted their visual attention and interest, since they glanced at it from the beginning to the end of the animation creation (75 fixations during 40 min.). Contrastingly, post-tests marks and interaction traces suggest that the ECA's presence had no perceivable effect on concept or skill learning and task execution. It only encouraged help consultation.

Categories and Subject Descriptors

H.1.2 [User/Machine Systems]: *human factors, software psychology* – H.3.4 [Systems and Software]: p*erformance evaluation (efficiency and effectiveness)* – H.5.1 [Multimedia Information Systems]: *animations, audio input/output, evaluation/methodology.*

General Terms: Design, Experimentation, Human Factors.

Keywords: Embodied Conversational Agents, Talking heads, Empirical study, Ergonomic evaluation, Eye tracking, Online help.

1. INTRODUCTION

A growing number of commercial websites, online help systems for the general public and learning environments (especially for children) are implementing Embodied Conversational Agents (ECAs) with a view to enhancing user-system communication or improving user support. However, some of these initiatives fail to meet with user acceptance. Thus, the Microsoft Companion which embodies online help to the use of Pack Office application suite irritates a majority of users who deactivate it despite its wide range of embodiments.

ECAs with humanlike appearance and advanced verbal and nonverbal communication capabilities have been designed and implemented in many research laboratories. Main current research efforts focus on creating ECAs which emulate, as best as possible, human appearance, facial expressions, gestures and movements, emotions and intelligent behaviors. Modeling emotion expression [13], human conversational skills [12] and gaze activity during face-to-face exchanges [7] are most active research themes.

Contrastingly, utility, usability and user acceptance of humanlike ECAs have motivated only a few ergonomic studies[1] which address specific issues like: the influence of various multimodal pointing strategies (involving the ECA's speech and hand gestures) on technical information memorization [3]; the influence of the ECA's voice quality (extrovert vs introvert) on children performances [6]; the affective dimension of human-computer interaction resulting from the presence of a human-like ECA (7 of the 9 studies mentioned in [14]); effects of the ECA's social behavior (politeness) on students' motivation and learning progress [15]; effects of ECAs' nonverbal behaviors on users' subjective judgments [8]. Results of these studies cover only a small part of the design space for humanlike ECAs; further research is needed in order to provide designers with meaningful guidance. In addition, most evaluation studies assess users' responses to ECAs in unrealistic Human-Computer Interaction (HCI) situations, and most analyses of empirical data collected are shallow.

The empirical study presented here contributes to increasing knowledge of the design space for ECAs. It aims at assessing a few hypotheses on the design of humanlike characters meant to support users in simple realistic learning tasks. Online help to the discovery of unfamiliar software intended for the general public has been chosen because it is a challenging user support situation. It is common knowledge that online help facilities are ignored by most "lay-users". Will the presence of an ECA contribute to overcome the "motivational paradox" described in [5]? Flash has been selected since basic conceptual (semantic) knowledge is required to create a graphical animation with Flash, in addition to procedural knowledge. The main objectives of this study are to collect and analyze users' responses and subjective judgments towards an ECA in this realistic challenging context, and to assess the actual influence of the ECA's presence on their behaviors and performances. Results will contribute to determine whether the presence of an ECA can increase the effectiveness and user-friendliness of online assistance in

ICMI'08, October 20–22, 2008, Chania, Crete, Greece.

[1] e.g., at IVA'07, only 1 session (out of 8) was devoted to evaluation.

conceptual and procedural knowledge acquisition by increasing novice users' self-confidence and motivation to consult help.

The methodology is presented in the next section together with hypotheses on the design of the ECA's animation. Results are described and discussed in section 3.

2. METHODOLOGY

2.1 Overview

22 undergraduate students who had never used Flash or any other animation creation tool previously, created two animations using two different online help systems successively. Animations had been predefined in scenarios and participants could play them at will in a dedicated window. The same database of multimodal help messages (over 300 speech and graphics messages) was used for both systems. One system, EH, included an ECA, a talking head which "spoke" oral help information, while the other system, UH, was not embodied. The presence/absence of the ECA was the only difference between the two help systems. Participants used EH (or UH) during the first scenario and UH (or EH) during the second one, order being counterbalanced between participants.

Once participants had filled in a background information questionnaire (10 min.), they got acquainted with Flash basic concepts (e.g., scenario, interpolation, etc.) using a short multimedia tutorial they could browse through as long as they wanted to (15-20 min.). Then, they created the two predefined animations (1 hour or so); their gaze movements were recorded throughout their interactions with Flash and the help systems. When they were finished, they filled in two questionnaires, a verbal and a nonverbal one [1]; both questionnaires were meant to elicit their subjective judgments on the ECA. Next, their understanding and memorization of Flash basic concepts and procedures were assessed using a written post-test. Finally, they participated in a debriefing interview. All-in, individual sessions lasted about 2 hours and a half.

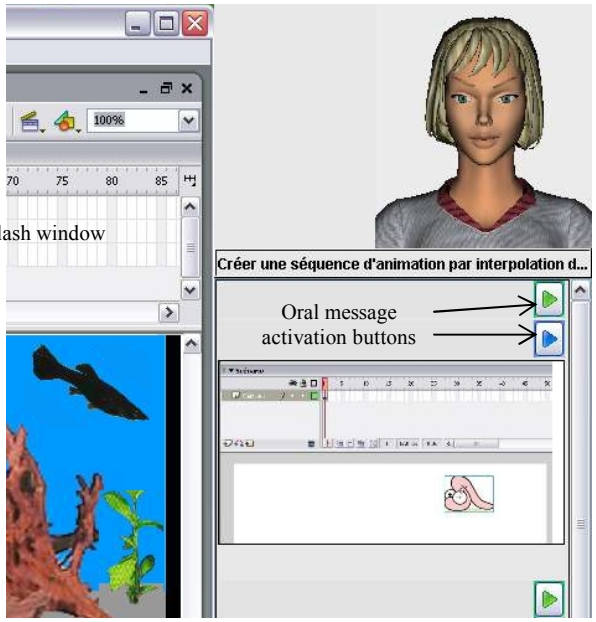

Figure 1. Top half of the EH help window

2.2 Implementation of Multimodal Help

The display includes two permanent windows (see figure 1): a large Flash window and a smaller help window (on the right of the screen) so as to reduce interference between help consulting and animation design activities. Based on earlier empirical work [4], four different types of help information can be asked using four dedicated buttons at the bottom of the help window: procedural know-how (How?), semantic knowledge (What?), explanations of the application current state (Why?) and confirmation or invalidation of the utility of recent actions on Flash (Confirm?). Definitions of technical terms (e.g., "keyframe") can be looked up in a lexicon represented as a pop-up menu above the four request buttons. Oral help messages are activated using colored buttons placed above the screen copies of Flash displays illustrating their information content. Colors indicate the type of information conveyed by messages: warning (e.g., pre-condition), basic information (concept definition or procedure description), additional information (explanations, advice, etc.). The talking head is placed at the top of the EH window.

2.3 Design of the ECA's "Personality"

To improve the effectiveness of user support, the ECA should be perceived as believable, trustable, cooperative, reassuring and engaging, according to the taxonomy presented in [14]. In addition, as it embodies an online help system, its presence and behavior should be unobtrusive, according to the minimal manual [5]; otherwise it might distract users from, and interfere with, their main goal (i.e., completion of the current animation) and activity (i.e., interaction with Flash), which might irritate them and reduce their overall efficiency.

In the absence of design guidelines, the ECA's appearance and personality have been defined using two principles:

P1: The appearance and behavior of the ECA must be in harmony with its role in the interaction.

P2: The ECA will be perceived as having presence and a believable "personality" if its appearance, gestures, facial expressions, voice quality and prosody contribute all to simultaneously express the same stereotypic mood, attitude or feeling, in harmony with the interaction context.

These principles and specific requirements proposed in [5] to overcome the so-called motivational paradox guided design choices. Among the ECAs developed by FT R&D, a 3D female talking head, Eloona, was preferred to a male or cartoon-like face or a fully-embodied character, in compliance with P1. Eloona's facial expressions were defined manually using Facelab graphical editor [2]; an experienced speaker's voice was preferred to speech synthesis so as to obtain more expressive and varied prosody. Oral messages were classified according to their length and information content (using standard speech act categories). For each of the 18 resulting categories, a specific head animation scenario (including facial expressions, gaze shifts, head movements) and specific prosodic patterns (rhythm, intonation, loudness) were defined using a subset of the taxonomy in [13]. This procedure provided the ECA with consistent multimodal expressiveness, in accordance with P2. In addition, as Eloona was meant to embody an online help system, she was placed at the top of the EH window in a fixed position; when she was not answering users' requests she did not disappear; she just turned her head away from them.

The study presented here offers an opportunity to test the usefulness of the two principles proposed above for designing embodied conversational assistants that will enhance the effectiveness of user support and be well accepted by users.

2.4 Measures and Working Hypotheses

To assess participants' affective responses to the presence of Eloona, behavior-based measures were collected in addition to verbal and non verbal judgments, as recommended in [11]. Participants' gaze movements were recorded throughout the experiment (at 60 Hz), using a head-mounted eye tracker (ASL-501) which allows free head and chest movements without loss of gaze points and precision. As voluntary eye movements express shifts of visual attention, they are valuable indicators of users' engagement and involvement in interaction with an ECA [9]. Physiological measures, such as heart rate or galvanic skin response [10] were excluded as they are more intrusive than eye tracking in standard HCI environments.

To assess the actual influence of Eloona's presence on help consultation and participants' learning Flash concepts and procedures, post-test marks and qualitative evaluation of scenario realization were used. A software platform was developed for recording participants' interactions with Flash and either help system. Logs include time-stamped user and system events, mouse positions and clicks, screen copies, gaze samples. They can be "replayed" and annotated semi-automatically. An in-depth analysis of participants' interactions and gaze activity using this platform is in progress.

Analyses of collected data aim at validating two main hypotheses:

H1: The presence of an embodied assistant is capable of improving the effectiveness of semantic and procedural knowledge acquisition by increasing novice users' motivation and self-confidence.

H2: An embodied online help system designed according to principles P1 and P2 will be well accepted by novice users; its presence will be perceived as having the potential to increase help effectiveness significantly.

End users' subjective judgments on innovative applications or devices are often influenced by individual attitudes towards novelty, and behaviors in experimental situations may differ from those in real contexts of use. Besides, attitudes and behaviors are liable to evolve with usage. To minimize the possible influence of these factors on results, young participants (20 years of age on average) with extensive practice of Internet and electronic games were selected. In addition, online help to an attractive software tool was chosen so that participants' attention and interest would be likely to focus on the interaction with Flash rather than on the ECA's assessment.

3. MAIN RESULTS

We have not performed any statistical analysis on data. The number of participants is much too small to obtain meaningful reliable results, in view of the cognitive complexity of experimental tasks.

3.1 Verbal and Nonverbal Subjective Judgments

According to verbal questionnaires, 16 participants preferred the embodied help system (EH) to the unembodied one (UH). 5 participants only rated UH higher than EH, and one participant gave preference to standard (textual and graphical) help facilities. Non verbal judgments in Sam first line are also very positive: 19 participants were very satisfied or satisfied with Eloona's presence, the feelings of the 3 remaining ones being neutral. During debriefings, 3 participants refined their positive judgments in this way: the presence of an ECA having the potential to stimulate help consultation and increase novice users' self-confidence and trust in the system, it is most useful in the early stages of unfamiliar software discovery; its usefulness decreases as practice develops.

Most participants were of the opinion that the presence of an ECA could significantly increase the efficiency, usability and appeal of online help for the general public, as shown in table 1.

Results in table 2 indicate that a large majority of participants was amused (rather than irritated) by Eloona's presence which, according to them, did not disrupt interactive animation creation but reduced their stress and increased their motivation for discovering Flash. They judged her attractive despite the lack of realism of her behavior. Their perception of her personality agreed with the intents and expectations that had guided its design. In addition, their feelings about Eloona and her presence evolved positively in the course of the session.

These results contribute to validating hypothesis H2 and principles P1 and P2 proposed as guidelines for designing embodied assistants that will meet with wide user acceptance and be perceived as having the potential to increase help effectiveness. Participants' attitude towards Eloona's presence also supports the design requirements (inferred from [4] and [5]) that online help ECAs should meet to be well received by users. These requirements which the Microsoft Companion does not comply with may explain its failure to gain general user acceptance.

Table 1. Participants' (22) opinions: effects of Eloona's presence on help efficiency and user friendliness (7-level Lickert scales)

ECA's effects on help:	Number of participants per level						
	Negative → Positive						
Efficiency				3	10	6	3
Conviviality				1	3	7	11

Table 2. Participants' (22) subjective judgments: Eloona's presence and personality (7-level semantic differentiator scales)

Differentiators	Number of participants per level						
Eloona's presence							
Irritating → Amusing					7	8	7
Disruptive → Stimulating		1		4	7	5	5
Eloona's appearance, behavior and personality							
Dull → Attractive			1	4	9	4	4
Reserved → Affable-nice			2	1	11	6	2
Cold → Friendly				3	10	5	4
Passive → Cooperative			3	4	3	6	6
Artificial → Realistic		1	5	2	7	5	2
Unpleasant → Likeable				1	8	7	6
Evolution of participants' judgments in the course of the session							
Negative → Positive			1	9	7	2	3

3.2 Performances and Behaviors

Scenarios have been designed so that most semantic and procedural knowledge is to be acquired during execution of the first scenario. Therefore, analyses of participants' performances and behaviors have been focused on the first scenario, and participants have been divided into two groups according to the help system they had used first, EH or UH. To elicit the actual influence of Eloona's presence

on online help effectiveness, performances and behaviors of participants in each group, EHG and UHG, were compared.

For most measures, average values computed over each group are not sensibly different; similar values have been observed for:

- Duration of the first scenario execution: about 40 min.
- Number of interactions with Flash and the help system (from annotated interaction traces): 83% with Flash, 10% with EH/UH, and the remainder (7%) with Windows.
- Quality of scenario realization and post-test marks.

The only noticeable difference between the two groups is the average number of oral help message activations per participant: 22 for EHG vs 16 for UHG. This result suggests that the presence of an ECA may encourage help consultation for novice users.

Analysis of annotated eye tracking data indicates that, from the beginning of scenario execution to the end, EHG participants glanced at Eloona both while she was talking and while she was silent with her head slightly turned away from the user. Each participant focused their visual attention on her 75 times on average; 46 fixations, that is, 13% of her total talking time (241 sec.), occurred while she was speaking. Fixation duration was longer while she was talking than when she was silent (428 vs 336 ms). These objective measures indicate that Eloona actually succeeded in arousing participants' interest and maintaining it throughout the first scenario execution, although they were intent on carrying out the scenario using Flash; this result confirms the positive subjective judgments expressed in questionnaires and interviews.

As shown by eye tracking data, Eloona undoubtedly attracted participants' attention and interest, which validates H2. Nevertheless, her influence on their activities and performances seems to have been limited. Her presence did not interfere with the quality and duration of scenario execution. However, it had no noticeable effect on Flash semantic and procedural knowledge acquisition. It only encouraged help consultation. These results are at variance with H1. Nevertheless, observation of novice users' behaviors and activities over a longer time span might be necessary for getting perceivable effects of the presence of an ECA on learning new concepts and skills.

4. CONCLUSION

An empirical study has been presented which aims at assessing the possible effects of embodiment on online help effectiveness and attraction. 22 undergraduate students who were unfamiliar with Flash, a tool for designing multimedia animations, created two simple Flash animations using two multimodal online help agents, EH and UH, one per animation. Both help agents used the same database of speech and graphics messages; EH was represented by a talking head while UH was not embodied. EH and UH presentation order was counterbalanced between participants.

Subjective judgments elicited through verbal and nonverbal questionnaires indicate that the presence of the ECA was well accepted by participants and its influence on help effectiveness perceived as positive. In addition, analysis of eye tracking data indicates that the ECA actually attracted their attention and interest since participants glanced at it often, from the beginning to the end of the animation creation. Contrastingly, its presence had no perceivable effect on concept and skill learning (according to post-test marks) and task achievement (according to analysis of annotated interaction traces). It only encouraged help consultation.

These results validate the two design principles proposed for creating embodied online help agents that will be well accepted by novice users in the general public. However, further research is needed to determine whether the presence of embodied help agents does actually facilitate and/or improve conceptual and procedural knowledge acquisition.

5. REFERENCES

[1] Bradley, M.M., Lang, P.J. 1994. Measuring emotion: the self-fassessment manikin and the semantic differential. Journal of Behavioral Therapy and Experimental Psychiatry, 25: 49-59.

[2] Breton, G., Bouville, C., Pelé, D. 2001. FaceEngine a 3D facial animation engine for real time applications. Proc. Web3D Conf., Paderborn, Germany, pp. 15-22.

[3] Buisine, S., Abrilian, S., Martin, J.-C. 2004. Evaluation of Multimodal Behaviours of Embodied Agents. In C. Pélachaud, Z. Ruttkay (Eds.), From brows to trust: Evaluating embodied conversational agents, Kluwer, Part III, ch. 8, pp. 217-238.

[4] Capobianco, A., Carbonell, N. 2001. Contextual online help: elicitation of human experts' strategies. In Proc. HCI International. vol. 2, pp. 824-828.

[5] Carroll, J.M., Smith-Kerber, P.L., Ford, J.R., Mazur-Rimetz, S.A. 1987. The minimal manual. HCI, 3(2): 123-153.

[6] Darves, C., Oviatt, S. 2004. Talking to digital fish: Designing effective conversational interfaces for educational software. In C. Pélachaud, Z. Ruttkay (Eds.), From brows to trust: Evaluating embodied conversational agents, Kluwer, Part IV, ch. 10.

[7] Eichner, T., Prendinger, H., André, E., Ishizuka, M. 2007. Attentive presentation agents. Proc. 7th Int. Conf. on Intelligent Virtual Agents (IVA'07), Springer LNCS 4722, pp 283-295.

[8] Krämer, N., Simons, N., Kopp, S. 2007. The effects of an embodied conversational agent's nonverbal behavior on user's evaluation and behavioral mimicry. Proc. 7th Int. Conf. on Intelligent Virtual Agents (IVA'07), Springer LNCS 4722, pp. 238-251.

[9] Ma, C., Prendinger, H., Ishizuka, M. 2005. Eye Movement as an Indicator of Users' Involvement with Embodied Interfaces at the Low Level. Proc. AISB'05 Symp. on 'Conversational Informatics for Supporting Social Intelligence and Interaction - Situational and Environmental Information Enforcing Involvement in Conversation', pp. 136-143.

[10] Mori, J. Prendinger, H., Ishizuka, M. 2003. Evaluation of an Embodied Conversational Agent with Affective Behavior. Proc. Workshop on Embodied Conversational Characters as Individuals at AAMAS'03, pp. 58-61.

[11] Picard, R.W., Daily, S.B. 2005. Evaluating Affective Interactions: Alternatives to Asking What Users Feel**.** CHI Workshop on Evaluating Affective Interfaces: Innovative Approaches. http://affect.media.mit.edu/publications.php

[12] Piwek, P., Hernault, H., Prendinger, H., Ishizuka, M. 2007. T2D: Generating dialogues between virtual agents automatically from text. Proc. 7th Int. Conf. on Intelligent Virtual Agents (IVA'07), Springer LNCS 4722, pp 161-174.

[13] Poggi, I., Pélachaud, C., de Rosis, F. 2000. Eye Communication in a Conversational 3D Synthetic Agent}, In E. André, (Ed.), Artificial Intelligence Communications, the European Journal on Artificial Intelligence, 13 (3): 169-182.

[14] Ruttkay, Z., Dormann, C., Noot, H. 2002. Evaluating ECAs. What and How? AAMAS'02 Workshop on Embodied Conversational Agents. Let's specify and evaluate them. http://wwwhome.cs.utwente.nl/~zsofi/publist.html.

[15] Wang, N., Johnson, W.L., Rizzo, P., Shaw, E., Mayer, R.E. 2005. Experimental Evaluation of Polite Interaction Tactics for Pedagogical Agents. Proc. ACM Int. Conf. on Intelligent User Interfaces (IUI'05), pp. 12-19.