Embodied Conversational Agents for Voice-Biometric Interfaces

Álvaro Hernández-Trapote
The Signal Processing Applications Group,
Universidad Politécnica de Madrid
Ciudad Universitaria s/n, Madrid, 28040, Spain
+34-91 336 72 80
alvaro@gaps.ssr.upm.es

Beatriz López-Mencía
The Signal Processing Applications Group,
Universidad Politécnica de Madrid
Ciudad Universitaria s/n, Madrid, 28040, Spain
+34-91 336 72 80
beatriz@gaps.ssr.upm.es

David Díaz
The Signal Processing Applications Group,
Universidad Politécnica de Madrid
Ciudad Universitaria s/n, Madrid, 28040, Spain
+34-91 336 72 80
dpardo@gaps.ssr.upm.es

Rubén Fernández-Pozo
The Signal Processing Applications Group,
Universidad Politécnica de Madrid
Ciudad Universitaria s/n, Madrid, 28040, Spain
+34-91 336 72 80
ruben@gaps.ssr.upm.es

Javier Caminero
Multilinguism & Speech Technology Group,
Telefónica I+D, Emilio Vargas 6, Madrid, 28043, Spain
fjcg@tid.es

ABSTRACT
In this article we present a research scheme which aims to analyze the use of Embodied Conversational Agent (ECA) technology to improve the robustness and acceptability of speaker enrolment and verification dialogues designed to provide secure access through natural and intuitive speaker recognition. In order to find out the possible effects of the visual information channel provided by the ECA, tests were carried out in which users were divided into two groups, each interacting with a different interface (metaphor): an ECA Metaphor group -with an ECA-, and a VOICE Metaphor group -without an ECA-. Our evaluation methodology is based on the ITU-T P.851 recommendation for spoken dialogue system evaluation, which we have complemented to cover particular aspects with regard to the two major extra elements we have incorporated: secure access and an ECA. Our results suggest that likeability-type factors and system capabilities are perceived more positively by the ECA metaphor users than by the VOICE metaphor users. However, the ECA's presence seems to intensify users' privacy concerns.

Categories and Subject Descriptors
H5.2 [**User Interfaces**]: *Evaluation/methodology*.

General Terms
Experimentation, Security, Human Factors, Standardization, Verification.

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

Keywords
Multimodal evaluation, Embodied Conversational Agent, biometrics interfaces, voice authentication.

1. INTRODUCTION
Biometric identity verification is a reality today. However, as has happened with many other technologies that involve human-computer interaction, this field requires a large amount of user-centered studies before becoming widespread and commonly used. Angela Sasse [1] provides an excellent overview of current research on the usability of biometric authentication systems and other security mechanisms. The fact is that usability is becoming a well established area of work in the HCI (Human Computer Interaction) domain, but security and usability have often been considered as conflicting design goals. Questions are being raised regarding the amount of effort and knowledge required of users by some security systems [2], for instance the role of trust in technology mediated interactions [1], the importance of people's attitude towards privacy [3], the need to safeguard privacy when monitoring an individual's behavior closely, or the underlying perceptions of the public concerning security [1], [3]. These and other similar issues have made user-centred criteria a major concern for the biometrics community.

The research we present in this paper is intended to contribute to the improvement of voice-based biometric interfaces, more specifically, of identity verification through speech. Enrolment and training dialogues are interesting to study [4], especially with regard to user trust in security. Factors such as privacy, user confidence, or pleasantness are especially relevant for adequately completing the task and at the same time satisfying the user. Unfortunately, few studies in the literature look at such factors in depth, and most are concerned solely with dialogue management and task success rates. Our objective is to study these aspects and try to reduce the negative effects they may have on users' acceptance of the system. We hope to do so by improving the

intuitiveness, naturalness, and efficiency of the spoken dialogue interface through the introduction of an animated character displaying especially designed gestures.

A general problem with spoken language dialogue systems (SLDSs) that acquires special relevance in user authentication dialogues is robustness. Speech recognition errors are hard to recover from, and recovery strategies can cause confusion among users as the dialogue takes unexpected twists. Turn management is also tricky and users are often not sure when they should speak. For this purpose ECAs may convey supra-linguistic information by performing gestures, including some designed as visual cues specifically to smoothen the flow of the dialogue making it seem more "natural" (for instance, by marking turn transitions), and others characterizing expectations, mental processes (e.g., how well the system is understanding the user) and emotions (e.g., using empathic strategies to control user frustration when errors occur) – [5], [6], [7] and [8].

Our application scenario, designed in order to have a realistic, though simulated, experimental framework, is a mobile internet service where users check the state of various home appliances using mobile phones (simulated on a computer screen). Access to the simulated application is granted using voice authentication technology. This main task in our test system: users are asked to enroll with a speaker recognition system, and then must verify their identity. We are interested in the effect an embedded ECA might have on the performance and user acceptance of biometric systems. It is generally considered that there is a trade-off between security and usability in these systems [9], and we wish to see whether ECAs may allow simultaneous improvement of both.

The article is organized as follows: In section 2 we present the verification and enrolment dialogue strategies and ECA behavioral schemes associated with them. Section 3 sketches the basic outer structure of a user-centered acceptability assessment frame we are currently developing and which has guided us when preparing our evaluation framework. In section 4 we outline the test procedure, and we present the main results we have obtained in section 5. Finally, in section 6 we draw conclusions and some open lines of research.

2. GESTURES FOR ROBUST ECA INTERFACES

We put together a set of dialogue strategies to prevent user frustration when problems arise in standard dialogue and also in enrollment and verification dialogues. Then we designed specific ECA behavior for each identified case. This section presents specific discussions on non-verbal communication through ECA gestures and their semantic interpretation and impact on four specific situations (see Table 1).

Table 1. Dialogue stages and associated ECA behavior

ENROLLMENT AND VERIFICATION DIALOGUE		
Dialogue stage	**Description (when it occurs)**	**ECA behavior (movements, gestures and other cues)**
Take Turn	The system starts to speak	Look straight at the camera, raise hand into gesture space. Camera zooms in. Light gets brighter.
Give Turn	The system prepares to listen to the user	Look straight at the camera, raise eyebrows. Camera zooms out. Lights dim.
Verification error	When the user identity hasn't been positively verified (False rejection)	Smile (and express remorse for not having been able to verify the user).
Wrong sequence of numbers recognized	The system "believes" to have "understood" a sequence of numbers uttered by the user, but it is not the one requested	Lean head sideways and down, raise inner eyebrow, eyebrow of sadness (remorse). Then opening eyes, and smile slightly (show interest).
Marking the tempo	Visual cue indicating the tempo with which the sequence of numbers (which the user is asked to repeat) is given.	Hand beat gesture for each successive number.

2.1 Turn Management

ECA body language and expressiveness may be exploited to help the flow the dialogue [10]. In particular, turn changing can be smoother with facial feedback provided by avatars [11].

Turn management basically involves taking and giving turn. Dialogue fluency improves and fewer errors occur if alternate system and user turns flow in orderly succession with the user knowing when it is her turn to speak. It is important to point out that we have not allowed barge-in (the speech recognizer is inactive while the system is speaking). This makes for a less flexible dialogue than may be desirable, but in certain situations such as recognition error spirals [12] it may be advisable not to allow the user to interrupt while the system is trying to reach a stable, mutually understood dialogue state, especially if the user's perception of reliability in identity authentication rests partly on how much under control the dialogue is seen to be.

Our ECA strategy is as follows: When it's the ECA's turn the camera zooms in slightly and the light becomes brighter. While the ECA approaches it raises a hand into the gesture space to 'announce' that it is going to speak (see Figure 1). When it's the user's turn the camera zooms out, lights dim and the ECA raises its eyebrows to invite the user to speak. Hopefully the user will learn to associate different gestures, camera shots and levels of light intensity with each of the turns.

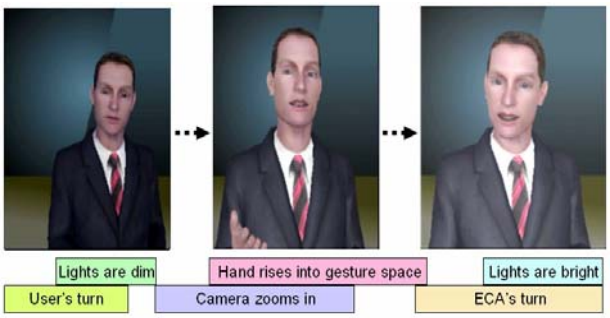

Figure 1. Visual sequence: turn transition from user to ECA.

2.2 Verification Errors

When the system is unable to verify the identity of the user –a typical problem with voice authentication (called *false rejection* if, as is the case in our tests, the user is not an impostor)– she may become nervous and, as a consequence, more prone to failure in the next verification attempt (because then her voice is strained and acquires a different quality than that which the system knows from the training stage (i.e., the enrollment stage). To partly avoid this problem our ECA doesn't tell the user that the system couldn't recognize her. Instead, the ECA kindly asks the user for another voice sample, making it simply seem that another sample is necessary as a normal part of the process. By hiding the fact that a verification error has occurred we hope to keep the user in a calm mood. The corresponding gestural strategy for the ECA is simply to remain smiling while requesting another voice sample.

2.3 Wrong sequence of numbers recognized

In order to avoid fraudulent access it is common for voice authentication applications to implement such strategies as requesting a different random sequence of numbers every time. Then, in addition to performing speaker recognition, the system performs speech recognition to find out what sequence of numbers the user has uttered. This is then compared with the requested sequence, and if it is found not to be the same, the user is rejected (this is a security measure aimed at avoiding fraudulent access using voice recordings). Such a strategy can lead to an increase in the number of rejections of genuine users, and so, in turn, increase frustration levels in such users. We have adopted the same recognition scheme and we use the ECA to try to empathize with the user in this unfortunate situation. We hope to achieve this by implementing a gesture to express remorse for not having been able to identify her, followed by an expression of interest in order to keep the user confident for the next verification attempt (see Table 1). The idea is for the system to take the blame for "misunderstanding." The "remorse" gesture is based on a gesture given in [13].

2.4 Marking the tempo

A common situation in speaker verification dialogue is that during training (enrollment) users repeat the sequence of numbers slowly, but once they acquire familiarity with the system they tend to repeat the requested sequence of numbers at a significantly higher pace. This can be a source of errors because verification algorithms perform better when a similar tempo of speech is followed in the training phase and in the verification attempts. The idea is, then, to implement an ECA strategy to try to get users to follow the same constant tempo when repeating the requested number sequence in both enrolment and verification, but without telling them, lest the system seem overly cumbersome to use. For this purpose our ECA marks the tempo with one beat of the hand for each number of the sequence [14] (see Figure 2).

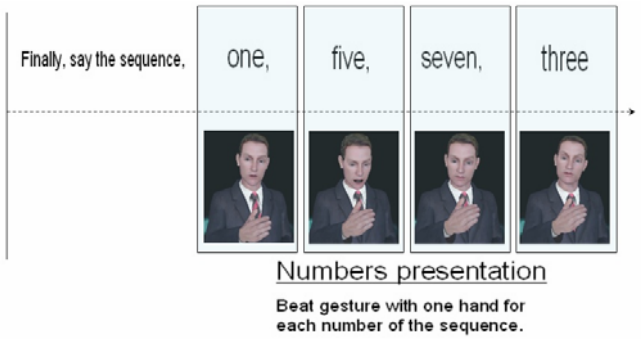

Figure 2. ECA speech and gesture lines for number sequence presentation.

3. QUALITY EVALUATION AND QUESTIONNAIRES

As far as we know, there is no standard procedure to evaluate enrollment and verification dialogues for speaker authentication systems with a spoken dialogue interface guiding the process. Our approach is to combine system and interaction performance and event data registered automatically with user's responses to questionnaires. It has been inspired by Möller [15] and PARADISE [16]. What we have done is to follow the ITU P.851 recommendation [17] on questionnaire design for the subjective assessments, and Möller's objective parameters [18] for quantifying the system and interaction performance and event data. In order to evaluate our system, we have decided to expand these previous works. Hence we include dimensions (in the form of sets o questions) that we have seen appropriate for evaluating user perceptions related with secure access and training/verification process using an ECA. At this point, and leaning on those previous references, we work with a reasonable conceptual scheme in which we have selected three classes of factors that may affect acceptability in our evaluation frame (see Figure 3):

Usefulness (as perceived by the user): This class involves all aspects relating to how well the user believes the system is suited to the pursuit of the goals she would expect or want to achieve by using it. To evaluate a dialogue system a relevant question would be, for instance, how well users believe the system understands them. And for a voice authentication system, how well users believe the system can recognize them.

Likeability: This class includes all factors that have to do with the experience of using the system. For instance, usability-related factors such as pleasantness, dialogue clarity, and ease of use, as well as emotions and other sensations.

Rejection factors: This class is qualitatively different from the other two. While in the latter the user's response may have a positive or a negative valence, rejection factors can only be negative. We believe that when rejection elements are present they may affect user acceptance in a different way to how negative values on likeability factors such as ease of use do. For this reason we choose to study them separately.

We have focused on certain aspects of privacy and security that are important in secure access systems. Namely, fear that unauthorized people may manage to access the system, fear that the biometric data may be misused, feeling observed and concerns about impersonation.

Notwithstanding our categorization, factors in different classes may, of course, be interrelated. This is what the arrows in the figure mean.

We have also analyzed how user expectations related to security and privacy evolve through use of the system (other studies, such as Jokinen's [19], have focused primarily on user expectations). We do this by repeating certain questions at different stages of the test, before and after the training and verification phases.

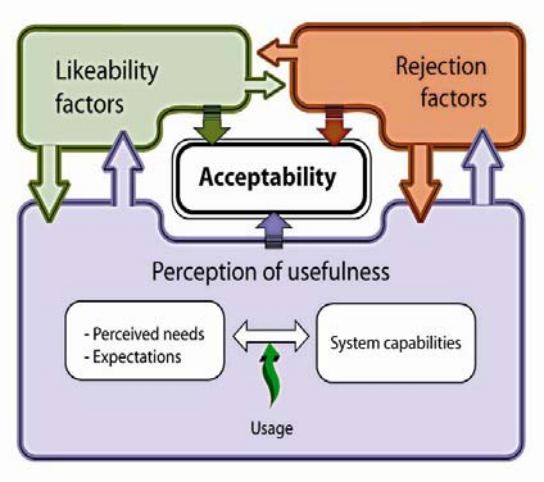

Figure 3. User acceptance-oriented evaluation frame.

4. EXPERIMENTAL SETUP

4.1 System Implementation

The architecture of the test environment is based on web technology, with which we simulate a mobile phone interface (see Figure 4). The ECA character was created by Haptek [20]. Speech recognition technology was provided by Nuance Communications [21]. Dialogues were implemented with Java Applet technology. Nuance's speech recognition engine provides a useful Java API that allows access to different grammars and adjusting a range of parameters.

Interaction parameters (utterance durations, number of turns, number of recognition errors, etc.) were recorded automatically during the test interactions. User questionnaires were implemented using HTML forms.

4.2 Description of the experiment

We tested the system with 16 undergraduate and graduate students (7 female and 9 male), aged 19 to 33, divided into two groups (8 users in each group), one to test the system with the ECA interface (or interaction metaphor: it is meant to look as if a human-looking character were in front of the user carrying out the tasks) and the other without ECA (VOICE metaphor: designed to feel to the user like a phone call with a distant agent).

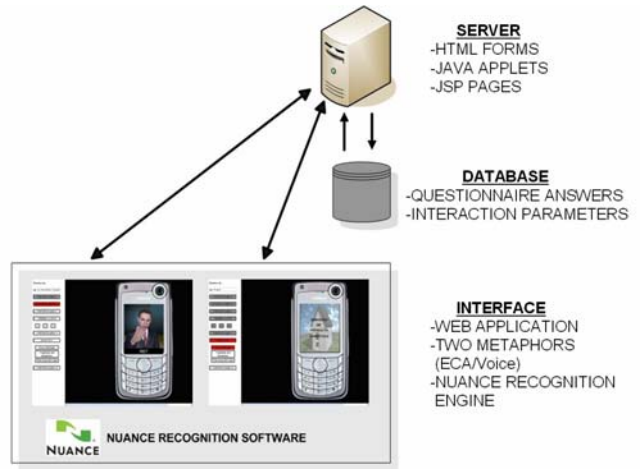

Figure 4. System Architecture

Testing was carried out in a small meeting room. Each user was seated at the head of a long table in front of a 15" screen. Two different views of the user interacting with the system were video-recorded to provide us with visual data to inspect and annotate the subject's behavior. A frontal view was taken from the top edge of the user's screen, and a lateral view was recorded from a wide-angle position to the right of the user. Both views were taken with Logitech Quickcam Pro 4000 webcams. The users interacted with the system using a headset microphone, and the system prompts are played through two small speakers. Half of the users interacted with a system that only produced spoken dialogue; the other half encountered an interface that included an ECA. All user-system dialogue was in Spanish. The evaluation was designed so that users could carry out the test with minimal intervention on the part of experimenter. The stages of the test are as follows:

1) Brief explanation: The user is told what the general purpose (to "evaluate automatic dialogue systems") and methodology of the evaluation are, as well as the tasks that lie ahead for him/her.

2) Opening questionnaire to learn about the user's prior experience and expectations.

3) Training phase: The user is asked to enroll in a secure access system, which requires interacting in guided dialogue with an application that registers his/her voice traits. (The system asks the user to repeat four four-digit sequences.)

4) Post-enrolment questionnaire to capture the user's opinions on the form of access and related aspects such as privacy and security.

5) Verification phase: The user does three successive verification exercises. In each he/she is required to repeat a random four-digit sequence (up to three times, in the event of verification failures). The outcome of each exercise is predetermined (there is no real verification going on, even though speech recognition *is* real). The idea is to let the user feel various situations that can arise during verification: In the exercise the system reacts as if it had successfully recognized the user at the first attempt; in the second the user is rejected after three failed attempts; and in the third, the user is granted access at the second attempt.

6) Post-verification questionnaire: Similar to the post-enrolment questionnaire, to see if users' opinions change after using the secure access system.

7) Dialogue phase: Users are asked to find out the state (on/off) of three household devices. The present paper does not focus on the interactions produced in this application as it lacks the secure access component.

8) Final questionnaire: To obtain the user's overall impression of the system, its main elements and the most important aspects of using it. Some questions are the same as in previous questionnaires, so that we may observe how user perceptions evolve throughout the various stages of system use.

5. RESULTS

We have obtained the results detailed in this section by a) comparing performance and questionnaire responses in the ECA metaphor group of users with those in the VOICE metaphor group; and b) observing how performance and responses to certain questions evolve throughout the test. We used two sample t-tests, setting the significance level at 5% (p=0.05). Questionnaire responses were collected on Likert-type 5-point response formats. User comments were also collected and compared to the findings in a) and b).

In Table 2 we summarize the main observations in which we found differences between the ECA and the VOICE-only interfaces, as regards user-system interaction.

Table 2. Summary of observations comparing user-system interaction with ECA and VOICE-only interfaces.

Interaction quality indicators	**ECA vs. VOICE-only**
Objective performance indicators	
Barge-in attempts	Fewer in the ECA case
Time-outs	Fewer in the ECA case
False rejections	Fewer in the ECA case
User turn duration	Shorter in the ECA case
Users' subjective impressions	*(results from both quantitative and qualitative analyses reported)*
Comprehension of which is the current stage of the dialogue	Better in the ECA case
Sense of control of the interaction (by the user)	Higher in the ECA case
Speed of task resolution	Higher in the ECA case
Efficiency of using the system	Higher in the ECA case
Sense of amusement	*Increases* throughout the interaction, more with the *ECA* interface
Pleasantness	*Decreases* throughout the interaction only for the *VOICE* interface
Privacy and security concerns	*Increase* throughout the interaction only for the *ECA* interface

We now discuss in greater detail the main findings obtained from these comparative analyses, focusing successively on each of the three quality evaluation categories introduced in Section 3: usefulness, likeability and rejection factors.

5.1 Perceived usefulness

In this section we look at how parameters related to the users' perception of system usefulness were affected by interaction features designed to make dialogue flow better and so gain in efficiency and clarity. We focus on two important aspects: turn management and tempo. (Both of these interaction elements were described in Section 2.)

5.1.1 Visual cues for turn switching

Users' perception that "*Dialoguing with the system led quickly to solve the task proposed*" (1 - totally disagree ... 5 - totally agree) was on average greater in the ECA group (4.2) than in the VOICE group (3.2) ($t(12)=3,16$; $p=0.004$). This is not just a subjective impression induced by the presence of the ECA, which would make it an instance of the persona effect. A close examination of the ECA-supported dialogues shows that users easily learn when it is their turn to speak to the system. This helps prevent most of the typically observed failed barge-in attempts and time-outs, which we found occurred more often for our VOICE metaphor users. Some of these users said they had felt confused at certain stages of the dialogue (e.g., *"between tasks there were silences and I didn't know if I was supposed to say anything," "a couple of times I think I spoke too early and that's why the system didn't get what I said," "it would be better if some sort of visual sign told you when the system is ready to listen"*).

We also found consistent differences between the two groups of users in task duration and number of turns taken, which are, of course, two important efficiency indicators. However, none were statistically significant. This may be due to the small size of our test groups. Nevertheless, before we test the system with more users it is reasonable to explain our findings as a combination of a persona effect with the fact that ECA-metaphor users learn to interact with the system more easily, feel more in control, and actually experience a more coordinated dialogue than VOICE-metaphor users.

Thus, it seems our visual feedback channel featuring an ECA displaying contextual dialogue management cues may be providing supra-linguistic information that users are able to interpret correctly, leading to improved coordination, which in turn increases the users' impression of the dialogue being fast, efficient and under control.

But what are these visual cues that appear to be so useful? Our findings suggest that the visual information strategy for turn-switching that we have implemented –involving a combination of gestures and lighting and camera zoom effects– may be creating a "proxemic-code" that helps avoid the complicated, problem-laden interaction patterns reported in [11], where user-ECA interaction suffers from rather severe coordination problems. Moreover, we observed no negative reactions, so users seem to accept proxemic shifts as a "natural" element of the interaction.

However, we shouldn't lose sight of the fact that, apart from a specific turn-giving gesture, the visual cues we have designed (the camera and lighting shifts) basically only involve the way the avatar is presented on screen. Other visual turn-switching markers

that don't involve the ECA (such as specially designed signs or color schemes) might work just as well. And then, the fact that acoustic markers were not used for the VOICE only metaphor limits the conclusions we may draw from our results. In future tests we plan, on the one hand, to use purely gestural ECA turn-switching cues, and on the other, to introduce purely iconic or abstract visual cues and then compare users' response to these with the results we are reporting in this paper to see which strategy works best and seems more natural to users. Only then can we really say whether differences are due to the ECA's behavior or to the presence of other turn-markers.

5.1.2 Visual cues to mark the tempo of speech

Unfortunately, we were unable to find any clear indication that users follow the pace marked with the ECA's hand gesture. Overall, some users followed a more uniform pace than others in both test groups. Furthermore, average overall UTD (User Turn Duration) was significantly shorter (μ_{ECA} = 21 secs., μ_{voice} = 26 secs.) for the ECA Metaphor group than for the VOICE Metaphor group in the training phase (t(13)=-1.90; p=0.040). Also, average UTD for the first verification attempt is significantly shorter (μ_{ECA} = 4.6 s, μ_{voice} = 5.1 s) for the ECA Metaphor group than for the VOICE Metaphor group (t(13)=-2.22; p=0.023). However, ECA Metaphor users suffered a lower number of rejections (a total of three rejections with the ECA metaphor, all for the same user and same number sequence, compared to four rejections among three users of the VOICE metaphor).

We hope a forthcoming analysis of the video recordings will give us further clues regarding the homogeneity of the duration and rhythm of individual user utterances.

5.2 Likeability Factors

A key element in speaker training and verification dialogues is that users feel comfortable during the interaction. Problematic situations include false rejections of users, which can cause frustration, which in turn can negatively affect the rest of the interaction. ITU-P P.851 factors related with pleasantness, amusement, and encouragement display a significant evolution throughout the tests.

The average grade users award to the *amusement* value of spoken dialogue interaction grows significantly in the ECA Metaphor users from the first questionnaire (2.9) to the last (3.6) (t=-2.39; p=0.024) (The precise question was: *"Compared to other ways of interacting with a system (e.g., pressing buttons to choose options from menus), do you reckon spoken dialogue can be more fun or more tedious?"* (1 - much more tedious … 5 - much more fun) In contrast, the average *pleasantness* score for VOICE metaphor users falls from the first questionnaire (3.6) to the last (3.3) (t=2.05; p=0.040). (In this case the precise question was: *"Compared to other ways of interacting with a system (e.g., pressing buttons to choose options from menus), do you reckon spoken dialogue can be more pleasant or more unpleasant?"* (1 – much more unpleasant … 5- much more pleasant).

5.3 Rejection Factors

A major concern in identity verification systems is privacy. Therefore, "personifying" with an ECA a system designed to capture sensitive information, as voice features are, requires special care. These are the findings in our study that bear on this issue:

Responses to the question *"Would you feel uncomfortable using the remote control system for home devices because you would feel your privacy was being encroached on?"* (1 - no, not at all … 5 - yes, very much so) evolved significantly in the ECA metaphor group, averaging 2.5 in the first questionnaire and 3.3 in the last (t =-2.05; p=0.040). Similarly, for the question *"Would you have security concerns using the system, perhaps because you fear that unauthorized people might manage to remotely control your home devices?"* (same scale): replies averaged 2.5 in the first questionnaire and 3.5 in the last (t=-3.06; p=0.009).

These results are in accordance with previous work [7] in which they studied the effect an ECA could have on users interacting with a biometric authentication application. They found that the mere presence of an ECA (without any specifically designed gestures and with little expressiveness) can negatively affect users' perception of loss of privacy. However, our new findings seem clearer, suggesting that a more active ECA has a greater negative impact on the users' perception of security and privacy. This could be either because the user feels observed or because an animated figure makes the system look less serious and therefore less trustworthy. We need to continue testing to clarify this point.

5.4 Overall acceptability

We have seen that there are differences between the two interaction metaphors (ECA and VOICE) regarding likeability, perception of usefulness and rejection factors. In particular, we found that users with the ECA tend to have greater privacy concerns than those without it. On the other hand, interaction with an ECA seems to be more pleasant. Since we found no significant differences regarding factors that might bear on user acceptance (such as overall impression and intention of future use), we may speculate that there is some sort of balance of likeability and rejection factors in our acceptability evaluation frame (described in section 3). However, this is only a vague notion and further testing is needed to propose and refine models with which to describe the relationships between specific factors and overall system acceptability.

6. CONCLUSIONS & FUTURE WORK

In this paper we have presented a study of gestural and other visual strategies to improve the fluency and robustness of human interaction with the spoken language dialogue interface of a voice authentication system, to which we have added an embodied conversational agent. We have focused on specific dialogue phases including (tempo of) information delivery, turn-taking and error recovery. We have also presented the design and results of an empirical test we have carried out to evaluate the effects of these strategies on interaction performance and the user's experience.

We have seen that adding specific ECA gestures and 'camera movements' to mark turn changes may improve dialogue flow and prevent barge-ins and related problems (compared to the voice-only interface). Users seem to be able to learn our proxemic code and accept it rather naturally. To determine how much of the improvement is attributable to the avatar's gestures and how much to the extra visual cues we introduced, and whether acoustic cues could be devised that would work just as well, we must carry out more specific tests.

Our findings also suggest that our evaluation frame can be useful in showing certain likeability and rejection factors might cancel

each other out in terms of the effect they have on user acceptance. We have observed that interaction with our ECA is more enjoyable but increases privacy concerns, while, overall, no noticeable difference in acceptance was observed between the two test groups. However, with our data we cannot determine a precise relationship.

Many questions open up before us. For instance, why are ECA users more concerned about privacy? Is it because of the way the ECA behaves? Because it seems more natural, as if there were a real person in the interface, so users feel observed? Or does this effect depend primarily on whether the ECA is present or not (and not on its expressiveness)? Also, our research on how to mark tempo needs further refining to see if it is possible to devise strategies to influence the way users speak to the system.

We hope also to be able to widen and refine our ECA's behavioral repertoire (which now only consists in a relatively small set of predefined gesture sequences that may be performed in succession) so that believability can be maintained, indeed improved, in broader and more flexible dialogue contexts.

We plan to perform further user tests with this experimental set-up shortly, after which we will analyze all the gathered information, including the video recordings (what we have presented here is a first batch of results that don't fully exploit the possibilities of our dialogue and gesture strategies, or our acceptability evaluation frame). We expect the videos will help us study the reactions of users to the 'emotional' behavior of the ECA.

7. ACKNOWLEDGMENTS

The activities described in this paper were funded by the Spanish Ministry of Science and Technology as part of the TEC2006-13170-C02-02 project.

8. REFERENCES

[1] Sasse, M.A. 2004. Usability and trust in information systems. Cyber Trust & Crime Prevention Project. University College London.

[2] Zimmermann, P.R. 1995. The official PGP User's Guide, Cambridge MA: MIT press.

[3] Biovision,. Roadmap for Biometric in Europe to 2010. Available at: http://ftp.cwi.nl/CWIreports/PNA/PNA-E0303.pdf (Accessed: 2008, August 13)

[4] Dialogues Spotlight Research Team, 2000. Dialogues for Speaker Verification/Operator Hand-Over: Apology Strategies, Security Dat, Insult Rate and Completion Procedures. University of Edinbutgh. Available at: http://spotlight.ccir.ed.ac.uk/public_documents/technology_reports/No.5%20Verification.pdf (Accesed: 2008, August 13)

[5] Poggi, I., Pelachaud, C. and Caldognetto, E.M. 2003. Gestural Mind Markers in ECAs, *Gesture Workshop 2003*, pp 338-349.

[6] Leßmann, N., Kranstedt, A. and Wachsmuth, I. 2004. Towards a cognitively motivated processing of turn-taking signals for the embodied conversational agent Max, *Proceedings Workshop W12, AAMAS* 2004, New York, 57 - 65.

[7] Hernández A., López B., Díaz D., Fernández R., Hernández L., and Caminero J. 2007. A person in the interface: effects on user perceptions of multibiometrics, Workshop on Embodied Language Processing, in the 45th Annual Meeting of the Association for Computational Linguistics, ACL, pp. 33-40, Prague.

[8] López B., Hernández A., Díaz D., Fernández R. Hernández L. and Torre D. 2007. Design and validation of ECA gestures to improve dialogue system robustness. Workshop on Embodied Language Processing, in the 45th Annual Meeting of the Association for Computational Linguistics, ACL, pp. 33-40, Prague.

[9] Whitten A. and Tygar D. 1999. Why Johnny can't encrypt: a usability evaluation of PGP 5.0. Proceedings of the 8th USENIX Security Symposium, Washington DC.

[10] Bickmore, T., Cassell, J., Van Kuppevelt, J., Dybkjaer, L. and Bernsen, N. 2004. Chapter Social Dialogue with Embodied Conversational Agents, (eds.), *Natural, Intelligent and Effective Interaction with Multimodal Dialogue Systems*. Kluwer Academic.

[11] White, M., Foster, M.E., Oberlander, J. and Brown, A. 2005. Using facial feedback to enhance turn-taking in a multimodal dialogue system, *Proceedings of HCI International 2005*, Las Vegas.

[12] Bulyko, I., Kirchhoff, K., Ostendorf, M. and Goldberg, J. 2005. Error correction detection and response generation in a spoken dialogue system, *Speech Communication* 45, pp 271-288.

[13] Pelachaud C. 2003. Overview of representation languages for ECAs. Tech. rep., Paris VIII, IUT Montreuil.

[14] Cassell J., Nakano Y.I., Bickmore T.W., Sidner C.L., and Rich C. 2001. Non-verbal cues for discourse structure, in Proceedings of the 39th Annual Meeting on Association For Computational Linguistics.

[15] Möller S., Smeele P., Boland H., and Krebber J. 2007. Evaluating spoken dialogue systems according to de-facto standards: A case study, Computer Speech & Language 21 26-53.

[16] Walker, M. A., Litman, D. J., Kamm, C. A. and Abella, A., 1997. PARADISE: A Framework for Evaluating Spoken Dialogue Agents. In: Proc. of the 35th Annual Meeting of the Assoc. for Computational Linguistics (ACL/EACL 97), Morgan Kaufmann, USA-San Francisco, 271-280.

[17] ITU-T P.851. 1999. Subjective Quality Evaluation of Telephone Services Based on Spoken Dialogue Systems, International Telecommunication Union (ITU), Geneva.

[18] Möller, S. 2005.Parameters for quantifying the interaction with spoken dialogue telephone services", In SIGdial6-2005, 166-177.

[19] Jokinen K. and Hurtig T. 2006. User Expectations and Real Experience on a Multimodal Interactive System, In INTERSPEECH-2006, paper 1815-Tue2A3O.2.

[20] Haptek, http://www.haptek.com (Accesed: 2008, August 13)

[21] Nuance Communications, Speech Recognition Technology, http://www.nuance.com (Accesed: 2008, August 13)