Chapter 1

EVALUATION OF MULTIMODAL BEHAVIOUR OF EMBODIED AGENTS

Cooperation between Speech and Gestures

Stéphanie Buisine, Sarkis Abrilian, and Jean-Claude Martin

> *They define being and body as one, and if any one else says that what is not a body exists they altogether despise him, and will hear of nothing but body.*
>
> —Plato, Sophist

Abstract   Individuality of Embodied Conversational Agents (ECAs) may depend on both the look of the agent and the way it combines different modalities such as speech and gesture. In this chapter, we describe a study in which male and female users had to listen to three short technical presentations made by ECAs. Three multimodal strategies of ECAs for using arm gestures with speech were compared: redundancy, complementarity, and speech-specialization. These strategies were randomly attributed to different-looking 2D ECAs, in order to test independently the effects of multimodal strategy and ECA's appearance. The variables we examined were subjective impressions and recall performance. Multimodal strategies proved to influence subjective ratings of quality of explanation, in particular for male users. On the other hand, appearance affected likeability, but also recall performance. These results stress the importance of both multimodal strategy and appearance to ensure pleasantness and effectiveness of presentation ECAs.

Keywords: Embodied conversational agent, evaluation, multimodal behaviour, redundancy, complementarity.

1.   Introduction

In order to make Embodied Conversational Agents (ECAs) more believable (Nijholt (2001)) and more comfortable (Ball and Breese (2000)), attempts are made to give them some aspects of emotions and personality during the interaction with human users (see Ball and Breese (2000) for a review; Workshops AAMAS (2002) and (2003)). Personality contributes to a large extent to defining ECAs as individuals: extraversion, agreeableness or friendliness are some personality traits that have been most studied. They affect all verbal and nonverbal modalities of communication: content of speech, intonation, facial expression, body posture, arm movements, etc.

Personality can be given to ECAs whatever their function. In assistance tasks, some ECAs (André et al. (2000)) combine specific behaviours depending on their personality (on the dimensions of extraversion and agreeableness) and presentation acts, which are not based on individual characteristics. To increase again ECAs' believability, we could also imagine to associate presentation acts themselves to individual strategies. In human behaviour, speech-accompanying arm movements can be considered as an integral part of individual communicative style (Kendon (1980)) and their occurrences could depend on the tactic of expression temporarily preferred by the speaking person (McNeill (1987), quoted by Rimé and Schiaratura (1991)).

During presentation tasks, ECAs have to relate speech and pictorial information. In such a context, cooperation between modalities observed in humans could be used to specify ECAs' behaviour. In social sciences, spontaneous gestures produced by a speaker were mostly studied for themselves (see Goldin-Meadow (1999a) for a review). Authors classically tried to observe and classify these gestures independently of the context and the speech content. The categorizations that emerged from these works show different levels of granularity but there seems to be a consensus on the following categories (see for example McNeill (1992)):

- *Emblems* are gestures that have a signification per se, for example waving the hand to say hello.

- *Iconic gestures* capture aspects of the semantic content, for example when the speaker mimes an action or symbolizes an object with his hands.

- *Metaphoric gestures* are pictorial gestures like iconics but displaying rather an abstract content, for example shrugging the shoulders to say "I don't know".

- *Deictic gestures* designate something in the conversational space, for example pointing at an object.

- *Beat gestures* are movements along with the rhythm of speech.

However, these categories do not detail to which extent the meanings conveyed by speech and gestures cooperate in the discourse. Simultaneous speech and gestures were related in some studies (e.g., Goldin-Meadow et al. (1999b)), but only in terms of match/mismatch of information. The framework provided by this field of research appears to be inadequate to the study of cooperation between modalities for ECAs.

On the other hand, the development of multimodal interfaces raised new needs in terms of analysis of human multimodal behaviour. Thus, on the basis of a survey of video corpora, we have proposed a taxonomy for the cooperation between modalities. The following types of cooperation are extracted from this taxonomy (see Martin et al. (2001) for more details):

- *Redundancy*: modalities cooperating by redundancy produce the same information.

- *Complementarity*: different chunks of information are produced by each modality and have to be merged.

- *Specialization*: a specific kind of information is always produced by the same modality.

In a presentation context, redundancy consists in giving verbal information and repeating it either with an iconic gesture or a deictic gesture towards an object. Although not explicitly named, this kind of strategy seems to be most frequently adopted for animated presenters or pedagogical agents (André et al. (2000); Rickel and Johnson (1999)). Conversely, cooperation by complementarity enables a decrease in the amount of information given by each modality. For example, the ECA talks about an object and gives information (e.g., shape or size) by hand gesture without mentioning this information by speech (Cassell et al. (2001)).

Some other presentation agents could be designed to give verbally the whole content of the presentation. This happens when the agent is embodied as an animated face without any body (e.g., Pelachaud et al. (2002)). A fully-embodied agent could also display no semantic content through gestures. In this case, modalities cooperate by speech-specialization. This type of cooperation corresponds to the 'elaborate speech-style', which is likely to occur in humans when the discourse content is distant from personal experience, conventional, abstract, and

objective (Rimé and Schiaratura (1991)). This strategy also constitutes a kind of control condition in comparison to redundancy and complementarity.

The primary goal of this study was to determine whether individual multimodal strategies, when exhibited by ECAs, would be perceived by a human listener and/or would have an impact on the effectiveness of the presentation. In these cases, what strategy would be the best one? We decided to test the effect of three multimodal strategies — cooperation by *redundancy*, *complementarity* and *speech-specialization* — in ECAs short presentations. We have selected these three strategies as they are rather different from one another and thus one could expect significant results when comparing them (although we did not make any preliminary hypothesis about which one would be perceived best). Another important issue in such a context is the influence of ECA's look on the effectiveness of presentation. As a secondary goal, we decided to test the effects of ECA's appearance independently from its multimodal strategy. Thus, the three selected strategies were randomly attributed to three different-looking ECAs. We investigated the impact of these two factors on two kinds of variables: subjective impressions of users (in a post-experimental questionnaire) and recall performance of the information provided in the presentations. Finally, we included in the questionnaire items about ECAs' personality, in order to test whether multimodal strategy and/or appearance influenced users' perception of ECAs' personality. In order to fully control the parameters of the ECAs' behaviour, the users could not interact with them. Thus, the users' task consisted in listening to three short technical explanations (60 to 75 seconds), trying to recall the maximum of information, and then filling out a questionnaire.

Next section presents the experimental setting. The results are described in section 3 and discussed in section 4. A few concluding remarks are presented in section 5.

2.   Experimental Setting

In this section, we present our methodology in details.

2.1   Participants

Two groups of users from our laboratory participated in the experiment: 9 male adults (age range 23 to 51, $mean = 30.7$) and 9 female adults (age range 22 to 50, $mean = 29.2$). These two groups did not differ in age ($F(1/16) = 0.129$; N.S.).

2.2   Apparatus

Animations were presented on a 19" computer screen (1024 × 768 resolution) and loudspeakers were used for speech synthesis with IBM ViaVoice[1]. In addition to speech synthesis, the text of the ECA's presentation was displayed sentence by sentence on the top of the screen (see Figure 1.1; the initial text was in French).

Figure 1.1.   Lea presenting a software with a redundant strategy. Other examples of Lea's behaviour can be seen on figure 1.4.

2.3   Scenarios

The presentations were three short technical explanations, dealing with the functioning of a video-editing software, a remote control for video-projector and a copy machine. The main difficulty lay in ambiguities of position, colour and shape of keys or menu items which are on the three objects. These objects were thus particularly relevant to study multimodal spatial references. They also involved similar functional behaviours, and were of the same complexity.

The explanations addressed on the position of buttons or menu items, on their function, etc. The ECAs appeared in front of a black background and a whiteboard. Each explanation was associated with a single picture displayed on this whiteboard (see Figures 1.1 to 1.3).
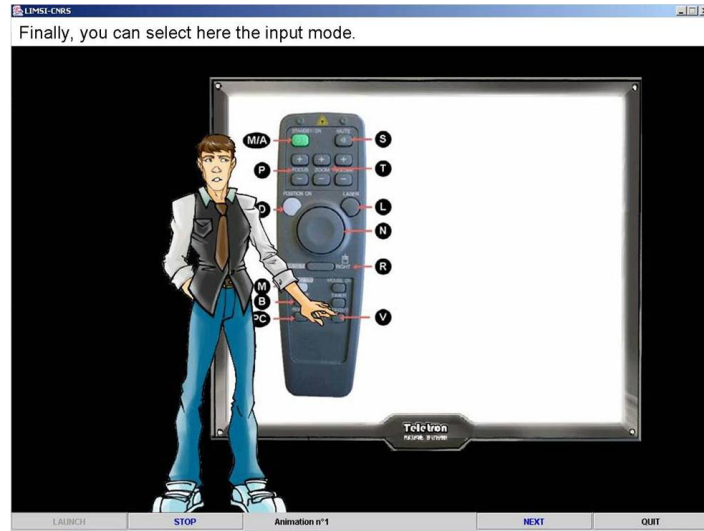
Figure 1.2. Marco presenting the remote control with a complementary strategy.
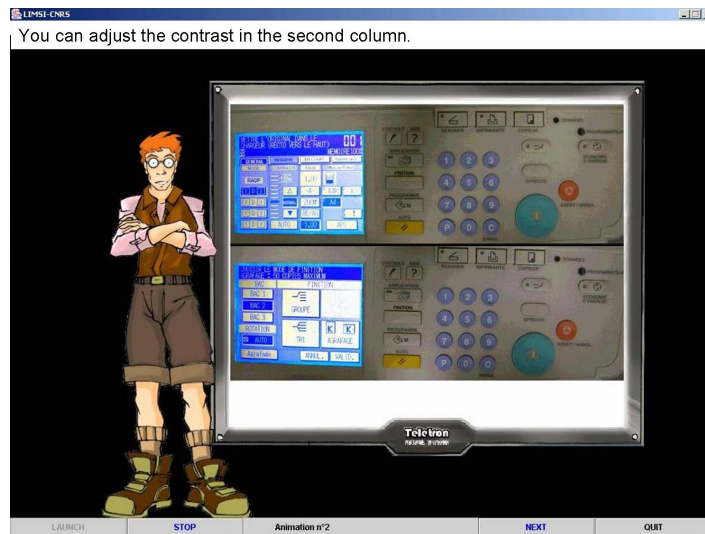

Figure 1.3. Julien presenting the copy machine with a speech-specialized strategy.

2.4 Independent Variables

The primary variable tested was the *multimodal strategy* of the ECAs. It had the following three values:

- *Cooperation by redundancy*: relevant information (e.g., position, shape, size of items) was given both by speech and arm gesture (deictic gesture towards the picture or iconic gesture when possible, see Figure 1.1).

- *Cooperation by complementarity*: half of relevant information was given by speech, and the other half was given by gesture (deictic gesture towards the picture or iconic gesture, see Figure 1.2).

- *Cooperation by speech-specialization*: all information was given by speech. Gestures did not convey any semantic content (see Figure 1.3).

The *appearance* of the ECAs was the second variable investigated in this experiment. We used three 2D cartoon-like Limsi Embodied Agents that we have developed. The 2D ECAs technology we used was described by Abrilian et al. (2002). Multimodal behaviour of all ECAs was specified using a low-level XML language. In this experiment, we used one female ECA and two male ECAs, namely Lea, Marco and Julien (see Figures 1.1 to 1.3). A demonstration is available on the Web[2].

Combinations between ECAs' appearance, multimodal strategy and content of presentation were determined by means of a Latin square design (Myers (1979)). Each ECA used each strategy and presented each object the same number of times across each group of users. For example, Figure 1.4 shows Lea presenting the remote control with the three different strategies. Such a design enables investigating the three variables with less expenditure of time (each user saw 3 presentations) than complete factorial designs would involve (27 presentations). It also removes some sources of error variance such as repetition effects. However, with this design, tests of interactions between these three variables are impossible to extract. We could only test the effect of ECA's appearance and multimodal strategy independently.

Finally, the influence of users' *gender* on dependent variables was tested. The two groups were paired regarding the Latin-squared combinations. Additional variables such as the content of the presentations or the order of presentations were considered as subsidiary variables. The presentations were equivalent in duration for the three contents (75 seconds for redundant and speech-specialized scenarios, 60 seconds for complementary scenarios). The presentation order of the three explanations, of the three strategies and of the three ECAs were neutralized across each group of users.
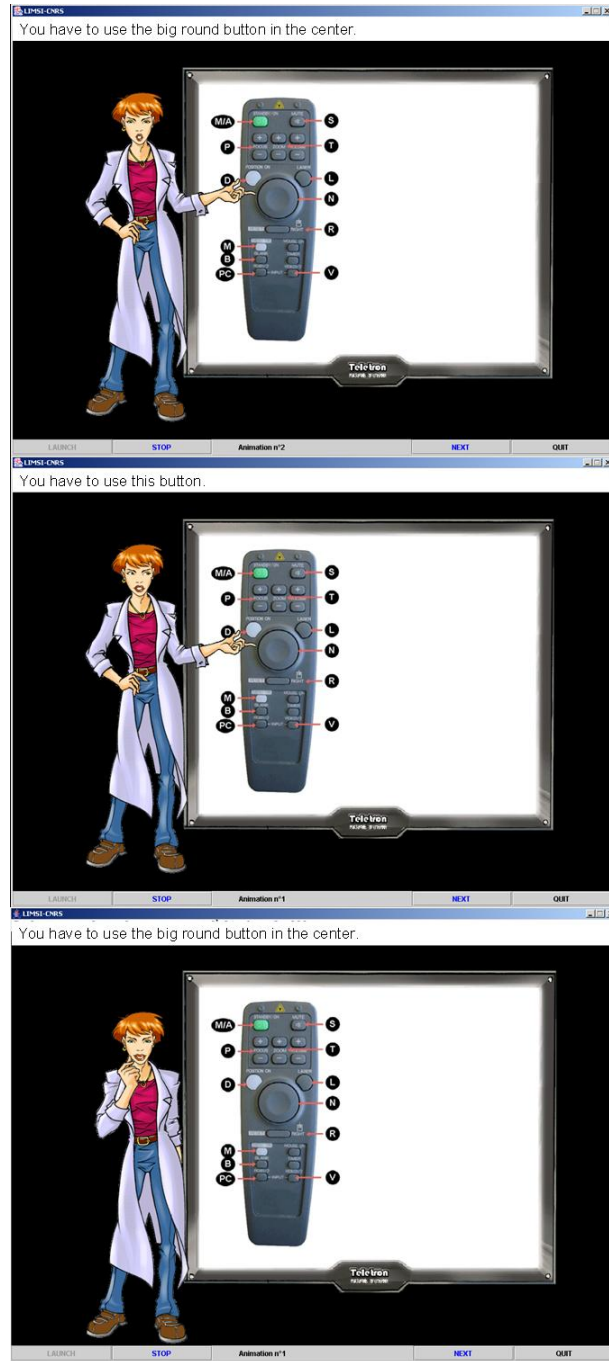
Figure 1.4. Each ECA (Lea in this screenshot) was tested with the three strategies: redundant (upper window), complementary (middle window) and speech-specialized (lower window).

2.5   Generation of Multimodal Behaviour

In this section, we present the way we specified the ECAs' behaviour whatever their appearance. All the animations were made manually. We first present the simple specifications we used for the animations that were common to the three strategies. Then, we describe the rules underlying each strategy, which were the focus of this study.

2.5.1   Common Animations.   Each feature of the ECA was manually animated in accordance with the content of the discourse. Lip movements, periodic eye blinks, and eyebrow movements were appropriately inserted in order to have a natural-looking animation. The ECAs also periodically turned the head towards the whiteboard, and emphasis was displayed via eyebrows on certain words (e.g., "on the **right**", or "the **blue** button"). Voice intonation was set to neutral.

The gestural modality was of prior importance in this study. We made sure that the number of gestures was exactly the same for all strategies so that we could compare them — any difference in users' reactions to the three strategies could not be attributable to variations in the *amount* of gesticulation. The rate of semantic gestures (deictic or iconic) among arm/hand movements was maximal in redundant scenarios, intermediate in complementary scenarios, and null in speech-specialized scenarios.

Hand shapes and movements for non-semantic gestures (e.g., laying the hand on the hip, moving the arm downwards, touching one's chin, folding the arms, etc.) were selected in our database according to the naturalness of their combination with each specific utterance. Since no intonation specifications were included, strokes of all gestures were placed manually in the speech course.

2.5.2   Rules for generating Redundant Multimodal Behaviour.   Redundant presentations were created by including the following rules in ECAs' animations:

- Speech: for items of interest, absolute localization (e.g., "on the top left side") was used whenever it was possible; otherwise the ECA used relative localization (e.g., "just below, you will find..."). Shape, colour and size of items were given whenever it was a discriminative feature.

- Hand and arm gestures: shape and size were displayed via an iconic gesture when possible (with both hands). A deictic gesture was used for every object. Finger or palm hand shape was selected as a function of the precision required (size of the item to be des-

ignated). Non-semantic gestures (as described above) were used when no other gesture was possible.

- Gaze: the ECA glanced at target items for 0.4 second at the beginning of every deictic gesture.

- Eyebrows: shape of big objects was not only displayed with speech and gestures, but also via raised eyebrows.

- Locomotion: if needed, the ECA moved closer to the target item before deictic gesture.

2.5.3 Rules for generating Complementary Multimodal Behaviour.
The following rules define complementary presentations:

- Speech: in comparison with redundant scenarios, information concerning localization, shape, colour or size was given for half of the items.

- Hand and arm gestures: deictic or iconic gestures were used every time the information was not given by speech. Non-semantic gestures were used the rest of the time.

- Gaze: the ECA glanced at target items for 0.4 second at the beginning of every deictic gesture.

- Locomotion: if needed, the ECA moved closer to the target item before deictic gesture.

2.5.4 Rules for generating Speech-specialized Multimodal Behaviour.
In speech-specialized presentations, ECAs were animated as follows:

- Speech: the same information as in redundant scenarios was given by speech (localization, shape, colour, size of items).

- Hand and arm gestures: only non-semantic gestures (as described in section 2.5.1) were displayed.

2.6 Dependent Variables

In this section we describe the variables we investigated and how they were collected.

2.6.1   Subjective Variables.   The users filled out a questionnaire in which they had to grade the three ECAs for the following questions:

- Which ECA gave the best explanation?

- Which ECA do you trust the most?

- Which ECA is the most likeable?

- Did the ECAs have the same personality? Which one had the strongest personality? (in French, the expression "strong personality" corresponds more or less to extraversion).

- Which ECA was the most expressive?

The users could also add free comments, and were particularly prompted to explicit their observations about the way each ECA gave explanations.

2.6.2   Recall Performance.   After viewing the presentations, the users were given the three pictures used in the experiment. On this basis, they had to recall the maximum of information they remembered. The experimenter marked out the performance (between 0 and 10) according to the number of information recalled (e.g., "this is the start button" counts for one information).

2.7   Data Analysis

Subjective variables as well as performance data were submitted to analysis of variance with user's gender as the between-user factor. For each dependent variable, the analysis was successively performed using ECA's strategy and ECA's appearance as the within-user factor. By way of control, the effects of the content of explanation were also tested. All the analyses were performed with SPSS[3].

3.   Results

The results described in this section will be discussed globally in the next section.

3.1   Subjective Variables

3.1.1   Quality of Explanation.   The main effect of ECA's strategy on ratings of quality of explanation proved to be significant ($F(2/32) = 5.469$; $p = 0.009$; see Figure 1.5). Indeed, ECAs with a redundant or a complementary strategy obtained equivalent ratings
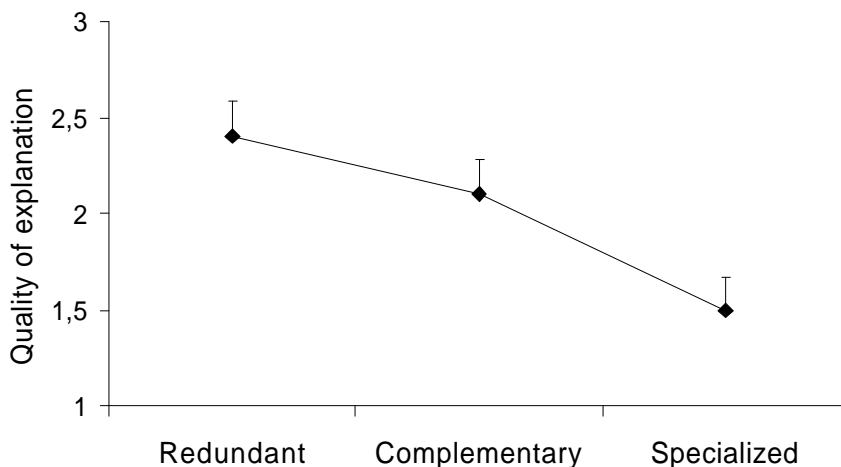
Figure 1.5. Ratings of the quality of explanation as a function of ECA's multimodal strategy.

$(F(1/16) = 1.000$; N.S.) but were both rated better than ECAs with a speech-specialized strategy (respectively $F(1/16) = 13.474$; $p = 0.002$, and $F(1/16) = 4.102$; $p = 0.060$).

The interaction between strategy and user's gender was also significant $(F(2/32) = 4.980$; $p = 0.013$; see Figure 1.6): the strategy effect was significant for male users $(F(2/16) = 19.000$; $p < 0.001$) but not for female users $(F(2/16) = 0.757$; N.S.). Ratings of male users could thus be considered as responsible for the previous main effect. Male users rated the ECAs with a redundant strategy better than the others $(F(1/8) = 12.000$; $p = 0.009$ for complementary strategy and $F(1/8) = 100.000$; $p < 0.001$ for speech-specialized strategy). They also tended to rate complementary strategy better than speech-specialized strategy $(F(1/8) = 4.000$; $p = 0.081$).

No effect of ECA's appearance or content of presentation was observed.

3.1.2   Trust.   No main effect of ECA's strategy arose in subjective ratings of trust, but an interaction between strategy and user's gender appeared $(F(2/32) = 3.735$; $p = 0.035$). In a similar way as for quality of explanation, the effect of ECA's strategy tended to be significant for male users $(F(2/16) = 2.868$; $p = 0.086$), whereas it was not for female users $(F(2/16) = 2.500$; N.S.).
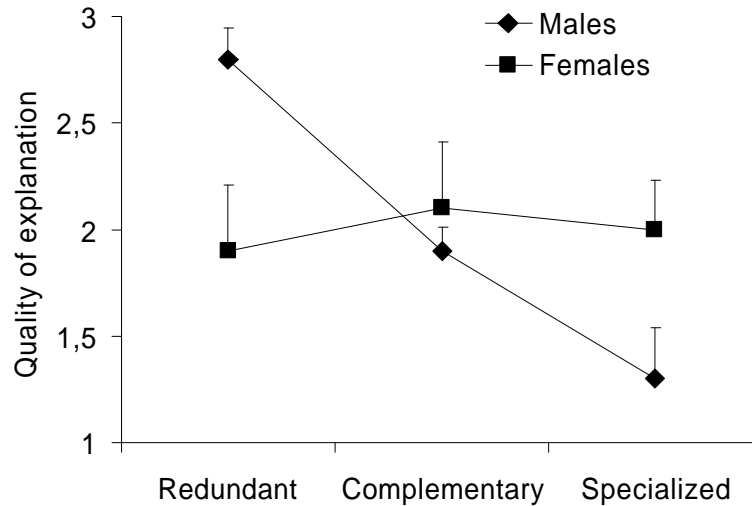
Figure 1.6. Ratings of the quality of explanation as a function of ECA's multimodal strategy and user's gender.

A positive linear correlation was found between this variable and ratings of quality of explanation (Pearson's correlation between 0.630 and 0.757, $p < 0.005$ for the three strategies). This result not only confirms that the interaction effect was of the same kind for the two variables, but also shows that ratings of trust were linked to ratings of quality of explanation.

No effect of ECA's appearance or content of explanation was observed on ratings of trust.

3.1.3 Likeability. Analyses on this variable yielded no effect of ECA's strategy, but a main effect of appearance proved to be significant ($F(2/32) = 3.328$; $p = 0.049$; see Figure 1.7). It showed that no preference arose between Marco and Lea ($F(1/16) = 0.471$; N.S.), but Julien appeared less likeable than Marco ($F(1/16) = 6.479$; $p = 0.022$) and than Lea (in trend: $F(1/16) = 3.390$; $p = 0.084$). This effect did not vary with user's gender. Moreover, if Marco and Julien's scores are combined, no interaction between ECA's gender and user's gender appears.

3.1.4 Personality and Expressiveness. No effect of ECA's strategy or appearance was observed on these variables.
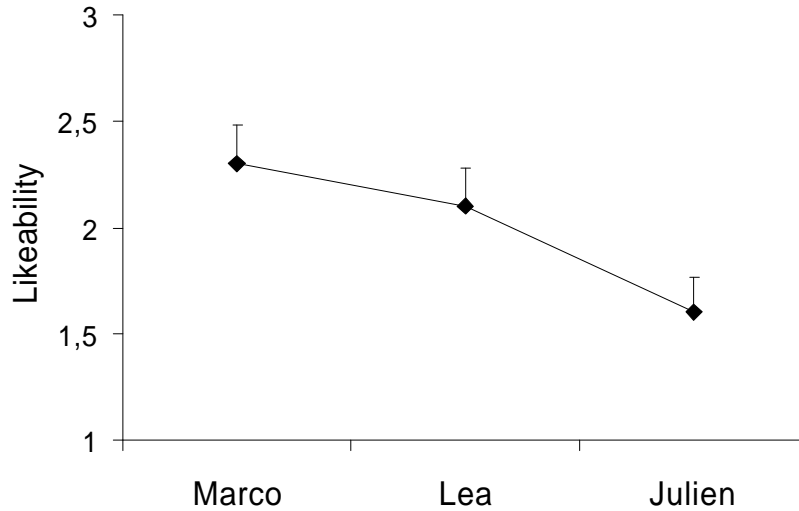
Figure 1.7. Ratings of likeability as a function of ECA's appearance.

3.2 Recall Performance

The average performance was 6.45/10. A main effect of user's gender on the amount of information recalled was significant in trend ($F(1/16) = 4.174$; $p = 0.058$), suggesting that female users recalled slightly more information (7.1/10) than male users (5.8/10).

ECA's strategy did not influence recall performance, but a main effect of ECA's appearance neared significance ($F(2/32) = 3.215$; $p = 0.053$; see Figure 1.8), suggesting that recall was slightly better when Marco had given the explanation, and slightly worse with Julien — recall with Lea being intermediate. This decrease of performance seems to follow the ratings of likeability, but no significant correlation between these two variables was found.

Concerning the influence of the content of explanation, no main effect arose, but an interaction between content and user's gender proved to be significant ($F(2/32) = 5.150$; $p = 0.012$). The effect of the content of explanation on recall performance was significant for female users ($F(2/16) = 9.838$; $p = 0.002$) but not for male users ($F(2/16) = 0.683$; N.S.). Actually, female users recalled more information about the copy machine than the two other objects. This effect, which constitutes a bias in our experiment, could come from a better previous familiarity of females with this object, although our two groups of users were homogeneous regarding socio-professional category.
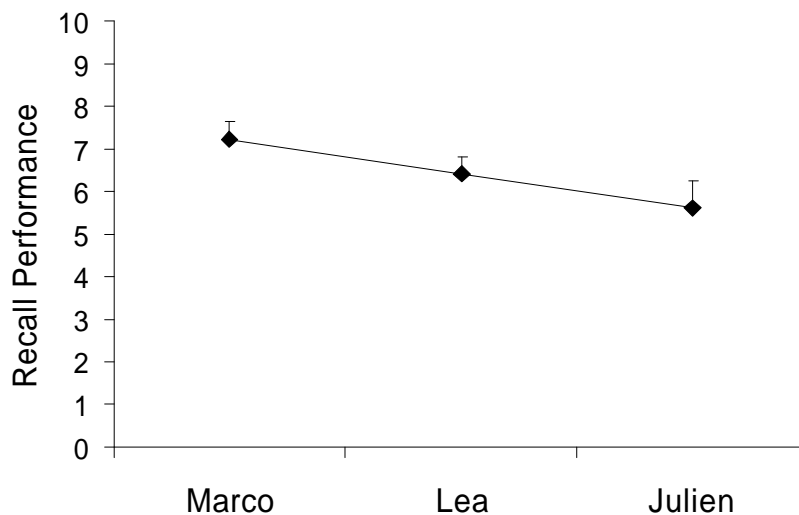
Figure 1.8.   Recall performance as a function of ECA's appearance.

4.   Discussion

Table 1.1 summarizes the main results of this experiment.

4.1   Effects of Multimodal Strategies

The main goal of this experiment was to study the effect of multi-modal strategies of ECAs. Before discussing the results, we would like to emphasize that these strategies were hardly consciously noticed by the users. The analysis of free comments given after the experiment shows that only 10 users (5 males, 5 females) from the 18 reported that they had observed differences in how the three ECAs gave explanations. Moreover, they noticed that some ECAs made deictic gestures, but nobody mentioned differences between redundant and complementary strategies. This is consistent with Rimé's figure-ground model (Rimé and Schiaratura (1991)) in which the speaker's nonverbal behaviour is usually at the periphery of the listener's attention.

The effect of multimodal strategies on ratings of quality of explanation was globally significant. However, considering the interaction with user's gender, this main effect proved to be produced by ratings of male users only. For this group of users, the preference for redundant ECAs was clear, though unconscious as underlined above. In contrast, ratings of female users yielded no preferences among strategies. This gender

Table 1.1. Summary of our results: our two main independent variables are presented in column and dependent variables are listed in raw.

	Multimodal Strategy	ECA's Appearance
Quality of Explanation	Main effect: $redundant = complementary$ $redundant > specialized$ $complementary > specialized$ Interaction with gender: *Effect of strategies for males, no effect for females.*	no effect
Trust	Interaction with gender: *Effect of strategies for males, no effect for females.* Correlation between trust and quality of explanation.	no effect
Likeability	no effect	Main effect: $Marco = Lea$ $Marco > Julien$ $Lea > Julien$ No gender effect (ECA or user).
Personality, Expressiveness	no effect	no effect
Recall Performance	no effect	Main effect (trend): $Marco > Lea > Julien$ No correlation between performance and likeability.

difference was unexpected. Before interpreting this result, we may point out that the number of users we tested may cast doubt on interaction effects. Indeed, we may consider that we had a fair number of users to test main effects, but the interactions arisen from our data will surely have to be confirmed in further experiments.

Nevertheless, the interaction we obtained raises interesting hypotheses on gender differences. We cannot assume that females were less focused on the ECAs than males. Indeed, our female users made a lot of com-

ments about ECAs' appearance and did not notice fewer differences in ECAs' strategies than males did.

The literature on recognition of nonverbal behaviours cannot explain either our result, because it usually reports that women have greater decoding skills than males (Feldman et al. (1991)). Besides, no gender differences have been described in biological motion recognition (see Giese and Poggio (2003)).

Finally, we could tentatively explain this result by the well-known cognitive differences between men and women (e.g., visual-spatial vs. auditory-verbal preferences, see Kimura (1999)). However, our protocol was too different from classical cognitive studies to claim that the same processes were involved. Thus we will conduct further experiments not only to verify our result with a greater number of users, but also to relate it to a cognitive model.

This gender difference is not clarified either by performance data, since ECA's strategy had no effect on user's recall in our experiment. Similar pattern of results (effect on subjective but not on objective variables) was previously found for example with the persona effect (van Mulken et al. (1998)). The fact that ECA's strategy influenced subjective variables without affecting performance does not in any way detract from the importance of these multimodal strategies. Indeed, we think that subjective variables remain a crucial factor of engagement and determine, to a certain extent, the success of such multimedia tools.

Ratings of trust yielded the same kind of interaction between ECA's strategy and user's gender. Actually, trust proved to be linked to the perceived quality of explanation. This result could be confirmed by more indirect questions, such as: "Would you buy a mobile phone from this ECA?" If it is confirmed, the influence of multimodal strategy on trust could be of interest in applications where trust is required (e.g., e-commerce).

4.2   Effects of ECAs' Appearance

The ECA's appearance had no effect either on ratings of quality of explanation or ratings of trust. However, it had a significant effect on likeability, which was independent of user's gender. This result showed that Marco and Lea were preferred to Julien. Marco's smile happened to be designed broader than the smile of the other ECAs, and this was appreciated by the users, as they indicated after the experiment. Comments about Lea were more contradictory, because of her white coat: some users found her nicer and more serious; some others found her too strict. The influence of ECAs' clothes on their evaluation was previously

mentioned in some empirical research (McBreen et al. (2001)). Finally, the fact that Julien's eyes were not so visible through his quite opaque glasses was negatively perceived by most of the users. Besides, his position at rest consisted in having his arms folded, and several users found it unpleasant.

ECA's appearance also tended to influence recall performance of the users. Although this result lacks statistical significance, it warns us about the consequences of ECA's design not only on user's satisfaction, but also on the effectiveness of the application. Performance was not shown to be correlated to ratings of likeability. In a similar way, Moreno et al. (2002) found that pedagogical efficacy of ECAs varied with their appearance, but they failed to find a link with any subjective variable (likeability, comprehensibility, credibility, quality of presentation, and synchronization of speech and animation). Further experiments are thus needed to confirm and interpret the influence of ECA's appearance on recall performance.

4.3   Additional Results

No effect of multimodal strategy or appearance of ECAs arose in perceived personality or expressiveness. Comments given by users at the end of the experiment indicated that three dimensions influenced their judgments for these variables: ECA's appearance, amount of movements, and voice. The importance of this last parameter was emphasized in recent research (Chapter by Darves and Oviatt), but it was not controlled in our experiment: we used only one male voice and one female voice from IBM ViaVoice speech synthesis. It should also be noticed that 4 users (1 male and 3 females) did not find any personality differences between the three ECAs.

Finally, the bias produced by the content of presentation (better recall for females about one of the objects) could possibly explain the overall better performance of female users (obtained in trend).

5.   Conclusions and Future Directions

Our results stress the importance of both multimodal strategy and appearance to ensure the design of pleasant and effective presentation ECAs. As highlighted by Table 1.1, multimodal strategies and ECAs' look did not influence the same variables. We then could suspect these two factors to be independent. However, a factorial design would be necessary to validate this assumption.

Taken as a whole, males and females subjective ratings showed no preference between redundant and complementary scenarios. The ad-

vantage of complementary strategy lies in the possible reduction of the amount of information transmitted by each modality: it enables avoiding both an overload of verbal information and an exaggerated gesticulation, which can be perceived as unnatural (Cassell and Stone (1999)). As a consequence, complementary scenarios could also save presentation time (to provide the same information, complementary scenarios were 20 % shorter than redundant and specialized scenarios in our experiment). However, if it is confirmed that male users find redundant strategies better, it could be interesting to use redundancy when target users are males or when the duration of presentation matters little. Benefits of redundancy in pedagogical applications were previously observed (e.g., Craig et al. (2002); Moreno and Mayer (2002)), but they concerned multimedia presentations (addition of text to auditory material) rather than multimodal behaviour of ECAs. In humans, teachers' hand gestures were shown to be useful in a math classroom (Goldin-Meadow et al. (1999b)), but the redundant or complementary nature of these gestures was not investigated.

Our findings about multimodal behaviour of ECAs might not be generalized to other contexts. This experiment investigated only a presentation task with some spatial aspects — positions of items were crucial. The importance of multimodal strategies might be lowered in a more narrative or conversational context. But it could also be increased in other situations, for example when the data to process are more complex. We might even hypothesize that multimodal strategies could yield differences in performance in more complex tasks.

Users' comments about ECAs' appearance suggested avoiding teacher-like features (such as a white coat), avoiding behaviours such as folding arms, and keeping eyes and gaze clearly visible. Conversely, a cartoonish broad smile seemed to be a predominant factor of likeability. Dramatized characters, because of the emotions they display, have previously been claimed to make better interface ECAs than do more realistic and human-like characters (Kohar and Ginn (1997)).

In the near future, we will carry out further experiments within the same methodological framework, in order to complement this study with data on more users. We also intend to improve our 2D ECAs technology by going up from manual specification of behaviour to higher-level specification language. Such a language should include rules for synchronizing not only gestures to speech, but all the modalities (e.g., for the role of eyebrow movements, see Chapter by Krahmer and Swerts). It could also be interesting to include different speech intonations, different energies and temporal patterns in movements, and some idiosyncratic gestures.

2D ECAs with individual behaviour can be of interest for mobile applications, but the design of 3D ECAs should also be considered.

We also suggest building ECA's individuality from corpora of individual human behaviours. We believe that ECAs look as if they came from the same mould because they are usually specified by the same set of general psycholinguistic rules. So far, both the literature on individual multimodal behaviour and the automatic extraction of context-dependent and individual rules from corpora annotation were neglected in the field of ECAs.

More experimental results could lead to recommendations for ECA design in various application areas such as games or educational tools, which could also include teams of ECAs having each their own multimodal behaviour. One issue will be the granularity of such design guidelines which should not be too specific in order to be useful to ECA designers.

Acknowledgments

Notes

1. http://www-3.ibm.com/software/speech/ (last accessed 2003-11)
2. http://www.limsi.fr/Individu/martin/research/projects/lea/ (last accessed 2003-11)
3. http://www.spss.com/ (last accessed 2003-11)
4. http://www.niceproject.com/ (last accessed 2003-11)

References

AAMAS (2002) Marriott, A., Pelachaud, C., Rist, T., Ruttkay, Zs., and Vilhjalmsson, H. (Eds.) *Proceedings of Workshop on Embodied Conversational Agents — Let's Specify and Evaluate them!*. AAMAS02, Bologna, Italy.

AAMAS (2003) Pelachaud, C., Marriott, A., and Ruttkay, Zs. (Eds.) *Proceedings of Workshop on Embodied Conversational Characters as Individuals.* AAMAS03, Melbourne, Australia.

Abrilian, S., Buisine, S., Rendu, C., and Martin, J.-C. (2002) Specifying Cooperation between Modalities in Lifelike Animated Agents.

In: *Proc. PRICAI02 Workshop on Lifelike Animated Agents: Tools, Functions, and Applications*, pp. 3-8, Tokyo, Japan.

André, E., Rist, T., Van Mulken, S., Klesen, M., and Baldes, S. (2000) The automated design of believable dialogues for animated presentation teams. In: Cassell, J., Prevost, S., Churchill, E. (Eds.) *Embodied Conversational Agents*, pp. 220-255, MIT Press, Cambridge.

Ball, G. and Breese, J. (2000) Emotion and personality in a conversational character. In: Cassell, J., Prevost, S., Churchill, E. (Eds.) *Embodied Conversational Agents*, pp. 189-219, MIT Press, Cambridge.

Cassell, J., Bickmore, T., Vilhjalmsson, H., and Yan, H. (2001) More than just a pretty face: Conversational protocols and the affordances of embodiment. *Knowledge-Based Systems*, 14, pp. 55-64.

Cassell, J. and Stone, M. (1999) Living hand to mouth: Psychological theories about speech and gesture in interactive dialogue systems. In: *Proc. of AAAI99*, pp. 34-42, North Falmouth, MA.

Craig, S. D., Gholson, B., and Driscoll, D. (2002) Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features, and redundancy. *Journal of Educational Psychology*, 94, pp. 428-434.

Darves, C. and Oviatt, S. (this book) Designing conversational interfaces for educational software.

Feldman, R. S., Philippot, P., and Custrini, R. J. (1991) Social competence and nonverbal behavior. In: Feldman, R.S. and Rimé, B. (Eds.) *Fundamentals of Nonverbal Behavior*, pp. 329-350, Cambridge University Press.

Giese, M.A. and Poggio, T. (2003) Neural mechanisms for recognition of biological movements. *Nature Reviews*, 4, pp. 179-192.

Goldin-Meadow, S. (1999a) The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, 3(11), pp. 419-429.

Goldin-Meadow, S., Kim, S., and Singer, M. (1999b) What the teacher's hands tell the student's mind about math. *Journal of Educational Psychology*, 91, pp. 720-730.

Kendon, A. (1980) Gesticulation and speech: two aspects of the process of utterance. In: Key, M. R. (Ed.) *The relationship of Verbal and Nonverbal Communication*, pp. 207-228, Mouton Publishers.

Kimura, D. (1999). *Sex and Cognition*. MIT Press, Cambridge.

Kohar, H. and Ginn, I. (1997) Mediators: Guides through online TV services. In: *Electronic Proc. CHI97*.
URL: http://www.acm.org/sigchi/chi97/proceedings/demo/hk.htm/

Krahmer, E. and Swerts, M. (this book) More about brows.

Martin, J. C., Grimard, S., and Alexandri, K. (2001) On the annotation of the multimodal behavior and computation of cooperation between

modalities. In: *Proc. AAMAS01 Workshop on Representing, Annotating, and Evaluating Non-Verbal and Verbal Communicative Acts to Achieve Contextual Embodied Agents*, pp. 1-7, Montreal, Canada.

McBreen, H., Anderson, J., and Jack, M. (2001) Evaluating 3D embodied conversational agents in contrasting VRML retail applications. In: *Proc. Int. Conf. on Autonomous Agents Workshop on Multimodal Communication and Context in Embodied Agents*, pp. 83-87, Montreal, Canada.

McNeill, D. (1987). *Psycholinguistics: A new approach*. Harper and Row, New York.

McNeill, D. (1992). *Hand and Mind*, University of Chicago Press.

Moreno, K. N., Klettke, B., Nibbaragandla, K., and Graesser, A. C. (2002) Perceived characteristics and pedagogical efficacy of animated conversational agents. In: *Proc. of ITS02*, pp. 963-971, Biarritz, France and San Sebastian, Spain.

Moreno, R. and Mayer, R. E. (2002) Verbal redundancy in multimedia learning: when reading helps listening. *Journal of Educational Psychology*, 94, pp. 156-163.

van Mulken, S. André, E., and Müller, J. (1998) The Persona effect: How substantial is it? In: *Proc. of HCI98*, pp. 53-66, Berlin, Germany.

Myers, J.L. (1979). *Fundamentals of Experimental Design*. Third Edition, Allyn and Bacon Inc, Boston.

Nijholt, A. (2001) Towards multi-modal emotion display in embodied agents. In: *Proc. of Artificial Neural Networks and Expert Systems 2001*, pp. 229-231, Dunedin, New Zealand.

Pelachaud, C., Carofiglio, V., De Carolis, B., De Rosis, F., and Poggi, I. (2002) Embodied contextual agent in information delivering application. In: *Proc. of AAMAS02*, pp. 758-765, Bologna, Italy.

Rickel, J. and Johnson, W. L. (1999) Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13, pp. 343-382.

Rimé, B. and Schiaratura, L. (1991) Gesture and speech. In: Feldman, R.S. and Rimé, B. (Eds.) *Fundamentals of Nonverbal Behavior*, pp. 239-284, Cambridge University Press.