

Can Embodied Conversational Agents Induce Empathy In Users?

Fiorella de Rosis, Addolorata Cavalluzzi, Irene Mazzotta and Nicole Novelli

Department of Informatics, University of Bari
Via Orabona 4, 70126 Bari, Italy
<http://www.di.uniba.it/intint/>

{derosis, cavalluzzi, mazzotta, novielli}@di.uniba.it

Abstract

In this paper, we discuss the results of a study which was aimed at investigating which forms of empathy may be induced by ECAs on users, how empathy can be measured and which aspects of the ECA's behaviour may increase this effect. The study was performed with a Wizard of Oz tool which enabled us to vary easily the agent's behaviour, and involved subjects with different backgrounds.

1. Introduction

Embodied Conversational Agents (ECAs) are seen as a new metaphor of human-computer interaction which should give the users the illusion of cooperating with a human partner rather than just 'using a tool'. The more these agents succeed in achieving this goal, the more their users are expected to show some sign of 'social relationship' with them: ECAs should be equipped to notice these signs and to respond appropriately. Although a number of evaluation studies have been produced, which describe how users see ECAs and how their vision is influenced by variations in the agent characteristics (see Ruttkay and Pelachaud, in press, for a recent review), the exact nature of the relationship between users and ECAs is still unclear. The Stanford group formulated in the famous *media equation*, the hypothesis that social science theories may be applied in this domain (Nass et al, 2000): recently, the need to specify the applicability conditions of this hypothesis and its rationale was advanced by several authors. Some studies proved that human interaction with technology is not exactly the same as the human-human one, and that humans tend to automatically adapt their dialog style when they are aware of interacting with a tool (Oviatt and Adams, 2000; Darves and Oviatt, 2002; Coulston et al, 2004). This finding brought to propose organizing Wizard of Oz studies to investigate the nature of interaction with technology, either in natural language (Dahlback et al, 1993) or with artificial agents: the first corpora of dialogs collected with these studies contributed to elucidate how the user behaviour changes

according to the interaction condition and the application domain.

We have been working in the last four years at an ECA which is aimed at promoting appropriate eating habits. To design this system, we integrated knowledge derived from psychological theories about health promotion with analysis of a corpus of human-human dialogs in which the 'client' had serious smoking, drinking or eating problems. In the first prototype of our system, the ECA tried to emulate the behaviour of the 'human therapist', the underlying hypothesis being that the human-ECA relationship should aim at mirroring the human-human one (de Rosis et al, 2003). To test whether this hypothesis was reasonable, we subsequently designed and prototyped a tool to perform Wizard of Oz studies with our ECAs in different conditions. The idea was to employ this tool as an iterative design method for our health promotion dialogs. In this paper, we discuss the first results of a study which was aimed at investigating, in particular, which forms of empathy may be induced by ECAs on users, how empathy can be measured and which aspects of the ECA's behaviour may increase this effect.

2. Background

Empathy is a quite fuzzy concept: it implies listening skill and emotional intelligence, with the ability to identify with and understand another's situation, feelings and motives. Empathy therefore implies an active attitude, requires some kind of cognitive evaluation of the interlocutor's situation, may occur

even in absence of any expression of emotion by the ‘empathizing interlocutor’ and may be either sincere or simulated (Poggi, 2004). Vaknin attributes to this concept a meaning which goes beyond pure emotion transmission, by claiming that: “The empathor empathizes not only with the empathee’s emotions but also with his physical state and other parameters of existence” (Vaknin, website). By accepting this definition, we take empathy (in a broad sense) as *the process of entering into a warm social relationship with someone else, of being in a way involved in her goals and feelings*: a concept closely related to friendship.

The need, for an ECA, to show empathy towards the user has been broadly investigated. Cassel and Bickmore worked at endowing REA with the ability to apply some of the strategies which are applied by humans to facilitate trust and collaboration: increase *intimacy* and *common ground* over the course of the conversation, decrease *interpersonal distance*, use *non explicit ways* of achieving conversational goals and display *expertise*. These abilities were implemented by means of variations in the agent’s language, the main of them being: (i) to introduce small talk to facilitate intimacy and build common ground; (ii) to induce emotional contagion by verbal and nonverbal affect expression and (iii) to increase credibility by means of expert’s jargon (Cassell and Bickmore, 2003). The effectiveness of these techniques was demonstrated by a small experiment, in which the user ratings of REA were measured by a questionnaire with Likert scales. Results of this experiment showed that the effectiveness of these techniques depends on the subject’s personality (introvert vs extrovert) and on their level of initiative.

Although an increase in the overall effectiveness of interaction induced by an empathic attitude of the agent could be proved by these studies, much less clear was whether and how the users really felt (and showed) empathy for the ECA and whether feeling empathy contributed to their overall evaluation: finding a circumstantiated answer to this question is crucial for designing an ECA which is aware of the user attitude and is able to react appropriately. If we assume that the user-agent relationship is symmetrical, we may hypothesize that users display their empathic attitude towards the agent with the same forms of expression which are employed by ECAs to this aim: in particular, attempts to increase intimacy and decrease interpersonal distance, attempts to establish a common ground and use of affective language. Humorous acts may also be taken as an offer of sympathy, as indirect indices of attempt to manifest an empathic relationship with the agent: “*When the participants are in the mood for jokes, joke telling occurs naturally and there is some meta-level cooperation*” (Nijholt 2004).

Some hypothesis about the agent features which may influence the degree of social relationship the ECA induces in the user are suggested by the Perceiving and Experiencing Fictitious Characters (PEFiC) model (Hoom and Konijn, 2003), which asserts that the appraisal of characters by an observer occurs along

‘ethic’, ‘aesthetic’ and ‘epistemic’ dimensions. Ethics relates to the ‘moral appraisal’ of the character features, be they negative (e.g. violence) or positive (e.g. politeness). Aesthetics relates to its ‘physical features’ (beauty) and to personality. Epistemic relates to the ‘sense of reality’ the observers feel when interacting with the character, on whether they can ‘trust’ what they perceive. Positive and negative values of these dimensions seem to influence, respectively, the ‘involvement’ and ‘distance’ of observers towards the agent. Apparently, optimal appreciation of a character is not achieved by settling all the features to a ‘positive’ value but rather by balancing tendency to be involved and tendency to maintain distance, and therefore positive and negative features. Horn and Colleagues therefore strive for avoiding too much realism or too much ‘positive’ features, to rather employ agents whose features are “a little bad, ugly and unrealistic and that arouse some negative valence and dissimilarity with their daily practice”. Balance between involvement and distance seems to be a function of the interaction duration: the initial degree of involvement is usually higher than the initial degree of distance because most observers are open to new experiences, but the two factors seem to increase differently with time, as the observers’ desire to reach their goal becomes more urgent over time. Therefore, an ‘optimal’ appraisal would be reached when distance does not outweigh involvement and observers only start perceiving doubt, apathy or ambivalence. Comparison between the self and the character also affects appraisal: perception of *similarity* (in age, race, class and gender but also in attitudes, beliefs and physical attraction) seems to be fundamental for feeling sympathy.

Though not being synonyms, friendship and empathy are closely related concepts. *Friendship* may involve varying types and degrees of companionship, intimacy, affection and mutual assistance. It is influenced, again, by interpersonal attraction but also by rewards, which should outweigh costs such as irritation or disappointment. In advice giving dialogs, rewards are affected by the subject’s expectation (information and, maybe, also fun). Therefore, even if (as in our study) subjects are pre-informed that the ECA with which they are going to interact is still in a prototypical stage, their involvement is probably affected by the degree of satisfaction in the information received and by how pleasant they find interacting with it. The subjects’ overall evaluation of the ECA and the dialog will probably depend on their personality, their interest for the dialog topic, their previous level of information on that topic and others.

3. Our Study

As we said, in our study we took empathy in the broad sense of ‘entering into a warm *social relationship* with someone’, and aimed at studying how it might be induced in the user by an ECA. The previous Section suggested the aspects of interaction which we could take

as signs of social relationship and the factors which could be varied to induce this kind of relationship.

As we wanted to apply measuring methods that went beyond subjective ratings of the agent characteristics, we employed an experimental setting which was based on a Wizard of Oz tool. This tool enables us to perform experiments in various conditions, by varying the physical aspect of the agent, its expressivity, the dialog moves, the evaluation questionnaire and other factors. Data of various kinds may be collected: subjects may be asked to evaluate the individual agent moves as well as its overall behaviour. On the other side, the resulting corpus of human-agent dialogs enables us to perform more analytical studies of the subjects attitude by means of a linguistic analysis of their moves. The architecture of our tool is sketched in figure 1. The head-only embodied agents we employ in our experiments are built with a commercial software (Haptik, see website): their voice may be rendered with a text-to-speech (TTS) synthesizer in Italian or in English. This flexibility enables us to diversify, in a quick and easy way, the dialog content, that is the ‘moves’ the agent can pronounce. It enables us, at the same time, to employ in our experiments a gallery of characters with a more or less realistic voice, more or less emphasized facial expressions. In the study described in this paper, we manipulated these parameters in a controlled way, by setting the study conditions at every step according to the particular hypothesis we wanted to test in that step. Our application domain was that of health promotion (in particular, suggestions about diet), in which we already got a considerable experience with the evaluation of character’s monologs (Berry et al, submitted).

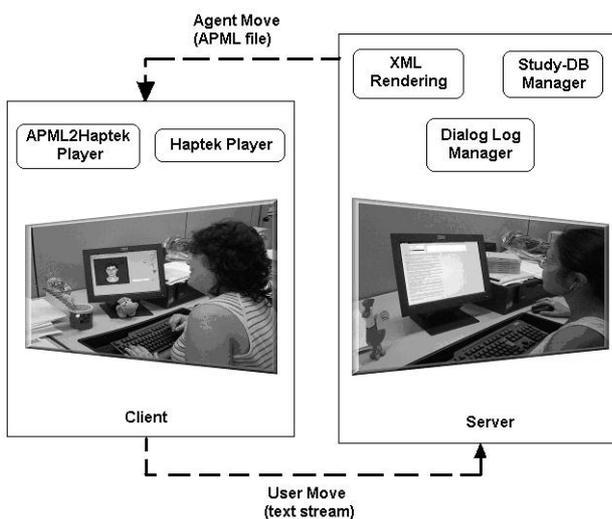


Figure 1: The Wizard of Oz experimental setting.

To insure uniformity of experimental conditions throughout the whole study, we had to establish some rules the wizard was requested to follow. After every subject move, the wizard selected her next move so as to respect a well defined dialog plan and to insure, at the same time, internal coherence in every dialog. This was

achieved by a careful preliminary training of the wizard and by employing the same wizard with all subjects. The plan applied by the wizard was defined after Prochaska and Di Clemente’s Transactional Model of Change (Prochaska et al, 1992). In this model, the dialog between a ‘therapist’ and a ‘client’ proceeds according to a strategy in which the presumed ‘stage of change’ of the client (from a presumably wrong to a more ‘correct’ behaviour) is considered, to adapt dynamically the dialog plans applied. These plans include a phase of *Situation Assessment*, which is aimed at understanding the client situation in the considered domain. This initial phase is followed by several ones:

- *Validate lack of readiness*, to verify whether the subject is really not intending to take action in the foreseeable future,
- *Clarify: decision is yours*, to explain that an effective change of behaviour requires an intentional change,
- *Encourage re-evaluation of current behaviour*, to try to reduce the subjects’ resistance to think and talk about their risk behaviour,
- *Encourage self-exploration*, to promote the subjects’ reflection on their living style and the reason why they are adopting it and
- *Explain and personalize risk*, to inform the subjects about short and long term effects of their behavior on their health, by adapting this analysis to their goals and priorities.

We employed an head-only character with a rather realistic and pleasant aspect (figure 2) and with two kinds of voices: a mechanical and not much natural one (produced with the Microsoft TTS in Italian) and a much more natural one (produced with Loquendo: see website).

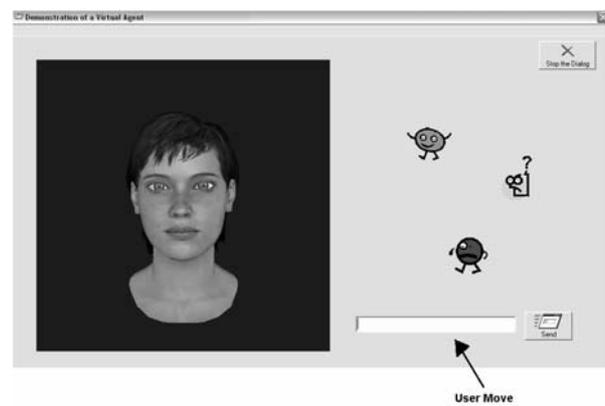


Figure 2: The Haptik character

During the dialog, the subject could evaluate every single agent move by clicking on one of the icons at the right side of the window, which indicate, respectively, whether the expression was considered as ‘nice’, ‘unclear’ or ‘bad’. At the end of the experiment, a final questionnaire was displayed on the video, to collect an evaluation of several features of the message and of the

agent, each with a Likert scale from 1 to 6. Items in this questionnaire measured various aspects of the PEFiC model: how much *credible* was the message and *sincere* the agent (ethics), how much *likable* (aesthetics) and *natural*, *intelligent* and *competent* (epistemics) was the agent, how much *plausible*, *clear*, *useful* and *persuasive* (relevance) was the message.

Dialogs were stored in a log at the end of interaction with every subject, to be analysed from a ‘qualitative’ and deeper viewpoint. We defined, first of all, two measures of the subject attitude during the dialog:

- *Level of involvement*, which is a function of the average number of subject moves in a dialog and of their average length, and
- *Degree of initiative*, as a function of the ratio between questions raised by the subject and overall moves.

These measures were integrated with a set of ‘signs of empathy’ that we drew from a linguistic analysis of the subject moves. These signs enabled us to evaluate the *degree and kind of social relationship* of the subject with the agent and to assess the relation between overall evaluation of the agent and the dialog, level of involvement, degree of initiative and forms of expression of social relationship.

3.1 Some Results

We describe the results of 6 tests, with 5 subjects in each of them and 30 subjects overall (Table 1). The tests were considered as steps of an ‘iterative design’ of our ECA: therefore, in designing every step we considered the results of the previous ones to find out the main limits of the ECA and revise its behaviour so as to avoid them (as we will see, we were not always successful in these attempts). After the first three tests, we could stabilize the agent moves and behaviour and we changed the background of our subjects to evaluate the possible role played by this factor.

Table 1: tests performed

Test ID	Ag move available	Subject background	Agent behavior
T1	53	Degree in humanities	‘cold’ style; Microsoft TTS
T2	53	Degree in humanities	‘warm’ style; Microsoft TTS
T3	84	Degree in humanities	intermediate style; ‘social’ agent moves added; Microsoft TTS
T4	84	Student in CS	as in T3; Loquendo TTS
T5	84	PhD student in CS	as in T3; Loquendo TTS
T6	84	PhD student in CS	as in T3; Loquendo TTS

In T1 and T2, we wanted to compare the effect of a ‘cold’ vs ‘warm’ style of the agent behavior. A cold style was obtained by enabling the agent to talk at the ‘third person’ (“*One should eat at least five portions of vegetables a day*”), to use ‘scientific’ arguments (“*Vitamins A and C purify the blood and enable growth and regeneration of tissues*”), to employ a formal style (“*Do you believe your weight is right or would you*

want to change it?”) and to hide any form of emotion in its facial expression. A warm style was obtained by addressing the sentences directly at the user (“*You should eat...*”), by using more emotion-evocative arguments (“*Vitamins A and C help you to get a healthier aspect and a brighter skin*”), a less formal style (“*Maybe I’m a bit indiscreet: but tell me, do you believe your weight is right or would you want to change it?*”) and by showing negative and positive emotional expressions in the agent face when appropriate. The number of alternative moves among which the wizard could select increased, from 53 (in T1 and T2) to 84 (in T3-T6): to overcome the limits we had discovered in the first two tests, in the subsequent ones we added to the set among which the wizard could select the next character’s move some information on topics which were frequently asked by subjects in the previous tests. Essentially, we introduced some generic answers and comments to make the dialog more ‘fluid’ and some answers to questions concerning the agent which (as we will see in the next Section) were rather frequently raised by the subjects. Subjects in T1-T3 were recruited among young people with a training in humanities, while subjects in T4-T6 had a background in computer science (BsC students in T4 and PhD students in T5, T6).

a. quantitative evaluations

A pre-test questionnaire enabled us to verify that the six groups of subjects were comparable in their level of knowledge, habits and interest for healthy eating, and in the importance given to it. They belonged to the same age group (23 to 26) and were equi-distributed in gender. The length of the dialogs (in n of adjoint pairs¹) ranged from 9 to 60 and increased only slightly with the number of overall moves among which the wizard could choose her answers (22.4 for T1&T2, vs 25.5 in T3-T6). The average length of moves for every subject ranged from 29 to 95 characters.

The message received, in the three experiments, a better rating than the agent. In the Likert scale from 1 to 6, it was considered as rather *clear* (3.7 on the average), *plausible* (3.7) and reasonably *useful* (3.4) but not much *persuasive* (2.1). The agent was considered as rather *likable* (3.6), reasonably *intelligent* (3.1) but not much *competent* (2.5) and not much *natural* (2.1). While the message ratings were a bit more favourable in the warm style condition (T2) than in the cold one (3.6 vs 2.8), the agent ratings were similar in the two conditions (2.8 vs 2.7).

A multiple regression analysis (Table 2) shows that the message rating is associated positively with the ratings in the initial questionnaire and the percentage of subject moves tagged as ‘social’ (see next Section). On the contrary, it is correlated negatively dialog duration (n. of moves), average length (in characters) of subject

¹ An *adjoin pair* is a couple of adjacent wizard-subject moves in the dialog.

moves and percentage of their questions in a dialog. This shows that the subjects' evaluation of the message was not associated positively (as we expected) with the variables we defined for measuring their degree of involvement and of initiative. However, the table shows, as well, that only a small part of the overall variability of the dependent variable is explained by the independent variables considered in the study. We will attempt an interpretation of these findings in the next Section.

Table 2: Least square estimates of multiple regression for variable: Message Rating

Least Square Estimate				
variable	coefficient	st. error	t-value	one-sided p
intercept	1.53	1.78	0.87	0.20
Initial quest. rating	0.40	0.35	1.15	0.13
n of moves	-0.47	2.07	0.23	0.41
av char / move	-0.05	0.17	0.28	0.39
% questions	-1.13	1.00	1.12	0.14
% of social moves	1.43	1.89	0.76	0.23
St error of estimate			97.5	
R squared			0.21	

Table 3 shows that the percentage of social moves in a dialog is associated positively with the subjects involvement (n of moves and their length), while it is correlated negatively with their level of initiative (% of questions): in this case, the value of R squared is higher than in Table 2.

Table 3: Least square estimates of multiple regression for variable: % of 'Social' Moves

Least Square Estimate				
variable	coefficient	st. error	t-value	one-sided p
intercept	-13.3	7.9	1.7	0.05
n of moves	0.53	0.21	2.47	0.01
av char / move	0.07	0.01	5.64	0.0000
% questions	-0.17	0.10	1.67	0.05
St error of estimate			12.1	
R squared			0.63	

Table 4: Role of subjects' background

	T1-T3 (humanities)	T4-T6 (comp.science)	F, p
Av n. of adjacent pairs in a dialog	27.7	20.1	F=3.9; p=.05
Av n. of characters in subjects' moves	56.4	40.6	F=7.3 p=.01
Subjects' questions/moves	.22	.39	F=5.3 p=.02
% of social moves	.37	.24	F=4.44 p=.04
Message rating	3.2	3.2	
Agent rating	2.7	3	

The subjects background was the factor which mostly influenced their behaviour: as shown in Table 4, computer scientists (T4-T6) made shorter dialogs, with

shorter moves, a larger proportion of questions and a lower proportion of social moves than subjects with a background in humanities.

b. qualitative evaluations

Overall, our corpus included 721 subject moves, that we labelled manually to identify those of them which showed some sign of the subject attitude to establish a social relationship with the ECA. The following are the language features that we considered as signs of this kind. For each of them, we provide some examples of adjoint pairs² which are extracted from the logs of our experiments and translated from Italian³. The examples come from all our tests and some of them belong to several classes.

a. Friendly self-introduction

The first move of the ECA is to briefly introduce herself by describing her name and role. The subject sometimes answers by briefly introducing self as well, as in the following examples⁴:

Oz: Hi. My name is Valentina. I'm here to suggest you how to improve your diet.

S: Hi, my name is Simone and I care about my diet. or

S: Hi, my name is Isa and I'm curious to get some information about education to healthy eating

b. Familiar expressions

Some subjects employ a familiar language:

Oz: Are you attracted by sweets?

S: I'm crazy for them. or

Oz: What do you think of a beautiful dish of warm and crispy fried food?

S: I think that, once in a while, this won't do any harm, especially if it is fried with the oil of papà.

or introduce dialectal expressions or proverbs:

Oz: I know: somebody may think that eating, and maybe living, in a messy way is pleasant, and maybe they are right. But, in the long term, negative effects may occur.

S: Somebody says that one day as a lion is better than a hundred days as a sheep.

or argue informally about the suggestion received

Oz: There seems to be a common agreement on the idea that limiting the amount of fat, in particular the 'saturated' one, is a fundamental rule of healthy dieting.

S: But this takes away the pleasure of eating!

c. Personal information

Providing personal information even when not requested may be seen as a sign of intimacy, as in the following examples:

Oz: Did you ever desire to change your diet?

S: Yes, I did it sometimes and got very good result. But now, both because of my indolence and of my stressing daily rhythms, I can't force myself through a

² We call 'adjoint pair' a couple of adjacent moves in the dialog.

³ We had to leave out some examples including very 'vivid' expressions because of the difficulty of translating them into English.

⁴ Oz stays for 'Wizard', S for 'Subject'

more rigid food regimen. or

Oz: Do you like sweets? Do you ever stop in front of the display window of a beautiful bakery?

S: *Very much! I'm greedy!* or

Oz: Do you remember what you ate yesterday?

S: *Yesterday I overdid a bit, as I went to a birthday party.*

d. *Humor and irony*

As we said in Section 2, answering with humorous forms to the agent's questions or suggestions is a sign of 'offer of sympathy'; for example:

Oz: I understand that organizing yourself so as to eat correctly may not be easy, especially if you work or study and nobody may help you in preparing food. You must find the time to make the market and cook a varied meal, while preparing a sandwich or eating what comes across is certainly much quicker.

S: *I'm disabled at 90%* or

Oz: Do you like sweets? Do you ever stop in front of the display window of a beautiful bakery?

S: *I don't only stop: I go in!* or

Oz: I know we risk to enter into private issues. But did you ever try to ask yourself which are the reasons of your eating habits?

S: *Unbridled life, with light aversion towards healthy food.*

e. *Questions about the agent's life*

These may be seen as signs of attempts, by the subject, to induce the agent to reciprocate manifestations of intimacy and decrease interpersonal distance:

Oz: What did you eat at lunch?

S: *Meet-stuffed peppers. How about you?* or

Oz: What did you eat at breakfast?

S: *Only two 'espressini'⁵ today. How about you?*

Oz: Maybe you forget I'm only an artificial agent

Subject: *So you don't eat? How do you feed yourself?* or

Oz: I can't eat, so I don't follow any particular diet.

S: *But if you don't follow any diet, how can you advice others about their diets?*

f. *Benevolent or polemic comments*

These may be seen as signs of involvement or disappointment; for instance:

S: *Apparently you don't know much about the properties of legumes*

Oz: Unfortunately I'm not an expert in this field.

S: *I appreciate your sincerity.* or

(after an agent's suggestion)

S: *OK: quite intelligent answer.*

Oz: I'm sorry, I'm not much expert in this domain.

S2: *OK: but try to get more informed, right?* or

(after a generic answer of the agent)

S: *It seems like if you are using a roundabout expression to answer the simple and precise question I raised.*

g. *Requests to carry on interaction*

If, when the agent tries to close the dialog, the subject asks to carry it on, this may be seen as signs of engagement:

Oz: My compliments. Good bye.

S: *What to you do? You leave me?*

Oz: Yes

S: *You are very rude! You interrupt our conversation without any real reason. I'll leave you, as you don't wish to talk with me any more.* or

Oz: Goodbye. It was really pleasant to interact with you. Come back when you wish.

S1: *But I would like to chat a bit more with you.*

While we found signs of social relationship in 33 % of the moves of our subjects, we could not understand which factors in the ECA's behaviour may increase the likelihood of establishing this relationship. None of the factors we considered seems to produce a positive effect: not the use of an empathic language (the 'warm' condition), not the use of a more 'natural' voice (with the Loquendo TTS), not the extension of the agent's ability to show signs of social relationship on its side, by talking about self or commenting on the subject's problems. Rather, the opportunity of establishing this relationship seems to be favoured, in our studies, mainly from the subjects' personality and background. In particular, we could check that subjects with a training in humanities were more open and ready to be involved in the dialog, while computer scientists had, in the large majority of cases, an attitude aimed at challenging the character, at discovering its limits: and they kept a rather indifferent attitude during the whole dialog. What was clear, in any case, was that noticing a cold reaction, by the agent, to some attempt to establish a friendly relationship with it was a source of strong disappointment by all the subjects: this suggests the need to endow the ECA with the ability to recognize the various forms of socio-relational attempts we discussed in this Section and to react appropriately.

4 CONCLUSIONS

Wizard of Oz evaluation studies may be seen as a method for the iterative design of conversational characters. Although the number of subjects in every group was too small, in our study, to come to any statistically significant conclusion, we drew, from every step of our experiment, new hints on how to revise the subsequent version. Persuasiveness of the message did not increase significantly though, probably because the arguments we employed (long term effects of a correct/incorrect diet on health) were not very strong for the young subjects involved in our studies.

Wizard of Oz simulations have clear advantages as a method to collect knowledge about human-technology interactions but also some limits, at least in our experience. Even if the wizard is trained to apply the same dialog strategy to all subjects, the number of available moves cannot be too large if uniformity in her behaviour through the whole experiment must be

⁵ An expression used in Bari, to denote a particular kind of preparing coffee

insured. This does not enable representing the wide range of opportunities for social relationship that particularly extroverted subjects offer with their moves.

For instance, even in T-T6, in which the number of available moves was larger, we could not include humour, dialectal expressions and other forms of ‘manifestations of friendship’ by the agent. Therefore, the agent was not much effective in communicating a sense of ‘reciprocity of liking’. It is then likely that the ‘neutral’ or ‘serious’ answers of the character to the subject attempts to manifest an empathic relationship might have contributed to induce a sense of irritation or disappointment in some of them. It is also likely that subjects who appeared to be more ‘engaged’ in the dialog were those who, in a way, were trying to challenge the character, in order to check the limits of the dialog it was able to entertain with the user. The combination of these two factors might explain the inverse relationship we found between overall evaluation of the dialog and degree of involvement of the user: however, due to the limited number of cases in our studies, these may be considered only as preliminary findings that we should verify in our future work.

Another question issue of this method is whether two levels of social relationship may be hypothesized for subjects involved in the study. The subjects know that they are part of an experiment to which they participate on a voluntary basis and whose goal is to assess positive and negative aspects of the behaviour of an ECA. As such, they interact with the agent and establish some form of relationship with it. But, at the same time, they establish an *indirect* relationship with the study designer, who will read and evaluate the transcripts of their dialogs. Therefore, their behavior may be influenced, either positively or negatively, by this meta-level relationship. Picard raised the question of whether “*when users perceive an expression of ‘emotion feeling’ in a machine, they attribute it to the designer of the software (to ‘implicit people’ behind the machine) or to the machine itself*” (Picard, 2002). One may extend this question, by asking oneself how much the behaviour of subjects interacting with ECAs in WoZ studies is, in fact, influenced by their desire to appear as serious, fun, competent, and so on, that is by the goal to give, in some form, a ‘good’ image of self to the agent designer.

A final consideration about the association between social relationship and application context. If it is clear that ‘friendship’ is a natural requirement of any entertaining domain for an ECA (such as game playing), it is likely that producing involvement in other applications (such as ‘advice giving’ in health promotion) would probably require enhancing different aspects of this relationship. In this context, probably ‘trust’, ‘confidence’ and ‘esteem’ are much more influent factors than friendship. Therefore, among the three dimensions of PEFiC, ethic and epistemic are probably much more influent than aesthetic. This might explain why comments about facial expressions (which were enabled, in our experiments, by the icons

associated with individual moves) were introduced only infrequently and only in ‘abnormal’ cases (that is, in case of really unnatural expressions) and why the condition ‘warm expression’ (in T2) improved the ratings assigned to the message but not to the character.

5 FUTURE WORK

To some researchers, classical methods of interaction design (including user requirement analysis, task analysis, scenarios, storyboards) should be applied, even if with some revision, in designing ECAs which fit the user needs in specific application domains: “*Do not augment realism, augment relevance*” is the password of supporters of this idea (Hoorn et al, 2003). We share this proposal, and claim that WoZ studies may be a useful method and tool in this iterative design process, especially when dialog simulation aspects rather than graphical aspects of the character have to be evaluated.

We learned a lot from our, even initial experience of iterative prototyping of health promotion dialogs. We initiated our studies with the belief that a key requirement of dialog simulation was the recognition of the emotional state of the users and of their stage of change. This is true, especially when the user problems are serious and therefore produce a strong emotional state (as in the case of natural dialogs with a therapist about drinking and smoking that we examined in another work: Carofiglio et al, 2004). On the contrary, we understood that, when the user problems are less serious, different kinds of emotions emerge in interaction: rather than strong ‘individual’ emotions like fear, joy, anxiety, relief etc, softer ‘social’ emotions like sympathy or antipathy, tenderness, contempt, sense of belonging (Poggi and Magno Caldognetto, 2003). To increase the effectiveness of advice-giving, the ability to recognize the degree of involvement of the user and to manifest reciprocity of social relationship seems to be more important than displaying realistic facial expressions of emotions. This opens complex problems, like recognising and generating humorous acts, formulating moves in a ‘familiar’ style, adding the ability to talk about ‘self’ and so on: and this, as everybody knows, is a typical category of ‘open problems’ in ECA’s design and implementation. .

There are two immediate steps forward for the research described in this paper: on one side, we wish to employ the corpus of dialog we collected so far to understand how a model of the ‘social’ attitude of users may be built dynamically during the dialog, by means of linguistic analysis of their moves. To this aim, we will process this corpus with knowledge discovery methods, to build a Bayesian Network with which to interpret the language features. On the other side, we wish to compare whether and how the user behaviour is influenced by the interaction modality: to this aim, we will build a new version of our WoZ tool in which the subjects will be able to interact by speech with the ECA.

Acknowledgements

This work was financed, in part, by HUMAINE, the European Human-Machine Interaction Network on Emotion (EC Contract 507422). We thank Berardina De Carolis and Giuseppe Clarizio for cooperating in the implementation of the WoZ tool, Loquendo for providing us their software in the scope of a scientific cooperation and Haptik Inc. for assisting us in the application of their software.

References

Berry, D.C., Butler L. and de Rosis, F.: Evaluating GRETA. The importance of consistency of behaviour in a multimodal animated agent. Submitted.

Cassell, J. and Bickmore, T.: Negotiated collusion: modelling social language and its relationship effects in intelligent agents. *User Modelling and User-Adapted Interaction*, 13, 1-2, 2003.

Coulston, R., Oviatt, S. and Darves, C. Amplitude Convergence in Children's Conversational Speech With Animated Personas. *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002)*. J. Hansen and B. Pellom (Eds). 2002.

Dahlback, N., Joensuu, A. and Ahrenberg, L. Wizard of Oz Studies. Why And How. *Proceedings of the Int Workshop on IUI*, 1993.

Darves, C. and Oviatt, S. Adaptation of Users' Spoken Dialogue Patterns in a Conversational Interface. *Proceedings of the 7th International Conference on Spoken Language Processing (ICSLP2002)*. J. Hansen and B. Pellom (Eds). 2002.

de Rosis, F., Pelachaud, C., Poggi, I., Carofiglio, V., De Carolis, B. (ed.) From Greta's Mind to her Face: Modeling the Dynamics of Affective States in a Conversational Embodied Agent. *International Journal of Human-Computer Studies*. B.R.Gaines. Vol.59 (1-2) (2003).

Haptik website: <http://www.haptik.com>

Loquendo website: <http://www.loquendo.com/>

Horn J. F. and Konijn, E.A. Perceiving and Experiencing Fictitious Characters: An

integrative account. *Japanese Psychological Research*, 45, 4, 2003

Horn, J.F., Konijn, E. A. and van der Veer, G.C. Virtual reality. Do not augment realism, augment relevance. *Upgrade*, 4,1,2003.

Nass, C., Isbister, K. and Lee, E: Truth is beauty: Researching Embodied Conversational Agents. In *J. Cassell, J. Sullivan, S. Prevost and E Churchill: Embodied Conversational Agents*. The MIT Press, 2000.

Nijholt, A. Observations on humorous act construction. *Proceedings of EMCSR 2004*.

<http://wwwhome.cs.utwente.nl/~anijholt/artikelen/emcsr2004.pdf>

Oviatt, S. and Adams, B. Designing and Evaluating Conversational Interfaces With Animated Characters. In *J Cassell, J Sullivan, S Prevost and E Churchill: Embodied Conversational Agents*. The MIT Press, 2000.

Picard, R. What does it mean for a computer to 'have' emotions? In *R. Trappl, P. Petta and S. Pays (Eds): Emotions in humans and artefacts*. A Bradford Book, MIT Press, 2002.

Poggi, I. Emotions from mind to mind. In A Paiva (Ed): *Proceedings of the Workshop on "Empathic Agents"*. AAMAS 2004.

Poggi, I. and Magno-Caldognetto, E. Il parlato emotivo. Aspetti cognitivi, linguistici e fonetici. *Proceedings of the Conference "Il Parlato italiano"*. D'Auria, Naples, 2003

Prochaska, J.O., Di Clemente, C. and Norcross, J.C.: In search of how people change: applications to addictive behaviour. *American Psychologist*, 47, 1992.

Ruttkay, Z. and Pelachaud, C. (Eds). From Brows Till trust: Evaluating Embodied Conversational Agents. Kluwer, in press.

Stock, O. and Strapparava, C. An experiment in automated humorous output production. *Proceedings of Intelligent User Interfaces*, 2003.

Vaknin, S.: On Empathy.
<http://samvak.tripod.com/empathy.html>