Modeling The User Attitude Towards An ECA

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

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

We discuss the results of a Wizard of Oz study which was aimed at investigating which forms of empathy may be induced by ECAs on users and how empathy can be measured through linguistic analysis. We describe how we built a model of the user's empathic attitude by learning and revising a bayesian network from the corpus of data collected with the study.

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, 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 organize 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 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 from linguistic analysis of the user moves and how a Bayesian network model of the user's empathic attitude can be built from the collected corpus of data.

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. It therefore requires some kind of cognitive evaluation of the interlocutor's situation and may occur even in absence of any expression of emotion by the 'empathizing interlocutor' (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 employed 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 the following variations in the agent's language: (i) introducing small talk to facilitate intimacy and build common ground; (ii) inducing emotional contagion by verbal and nonverbal affect expression and (iii) increasing credibility by means of expert's jargon (Cassell and Bickmore, 2003). Although an increase in the overall effectiveness of interaction induced by an empathic attitude of the agent could be proved by the evaluation studies performed in this Project, 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).

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. 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 head-only embodied agents we employ in our experiments are built with a commercial software (Haptek, 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 the dialog content, that is the 'moves' the agent can pronounce and to employ a gallery of characters with a more or less realistic voice and 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, in press).

3.1 Method

To insure uniformity of experimental conditions throughout the whole study, we established 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. We employed an head-only character with a rather realistic and pleasant aspect (figure 1) 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). 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 how much *credible, plausible, clear, useful* and *persuasive* was the message and how much *sincere*, *likable*, *natural*, *intelligent* and *competent* was the agent.

Figure 1: The character employed in our Woz studies

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

- *Level of involvement*: a function of the average number of subject moves in a dialog and of their average length, and
- *Degree of initiative*: 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.

3.2 Main Results

We performed 6 tests, with 5 subjects in each of them (Table 1). These 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. 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.

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 pairs[1]) 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. A multiple regression analysis showed that the message rating was 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 was correlated negatively with the dialog duration (n. of moves), the average length (in characters) of subject moves and the percentage of their questions in a dialog. This showed 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. The percentage of social moves in a dialog was associated positively with the subjects involvement (n of moves and their length), while it was correlated negatively with their level of initiative (% of questions. The subjects background was the factor which mostly influenced their behaviour: computer scientists 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. More detailed data on this quantitative analysis of our corpus may be found in (de Rosis et al, 2005).

Table 1: tests performed

Test ID	*Ag move available*	*Subject background*	*Agent behaviour*
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

Overall, our corpus included 708 subject moves, that we asked three raters to label 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 an example of adjoint pair which was extracted from the logs of our experiments and translated from Italian: some pairs 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 subjects sometimes answer by briefly introducing themselfs as well, as in the following example[2]:

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

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

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

[2] Oz stays for 'Wizard', S for 'Subject'

education to healthy eating

b. Familiar expressions

Some subjects employ a familiar language in their moves:

Oz: Are you attracted by sweets?

S: *I'm crazy for them.*

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: 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

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 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?*

f. Benevolent or polemic comments

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

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

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

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 a sign of engagement:

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.*

4. Labelling The Corpus

Overal, 237 'non neutral' moves were extracted from the corpus, to be examined by three independent raters. We considered the label of a move as 'agreed' when at least two raters gave it the same value. Agreement rates for the various categories are shown in Table 2. There was a high agreement rate for 'Friendly self-introduction', 'Talks about self', 'Questions about the agent', 'Irony', 'Friendly farewell' and 'Comments', and a rather low agreement on the 'Familiar language' label. Signs of positive attitude were more frequent but less agreed than those of negative attitude. This was probably due to a fuzzy definition of the mentioned categories: in particular, as raters were allowed to label a move with more than one tag, 'Familiar Language' and 'Attitude' were considered as more general categories than the other ones and therefore these tags frequently co-occurred with other ones.

Table 2: agreement rates

	Agreement rate	Positive cases
Familiar language	*.33*	59 %
Friendly self-introduction	.98	4 %
Talks about self	.73	33 %
Questions about agent	.70	31 %
Irony	.84	5 %
Friendly farewell	.93	7 %
Favourable comments	.82	5%
Neutral comments	.68	81 %
Negative comments	.86	12 %
Positive attitude	*.33*	52 %
Neutral attitude	*.24*	20 %
Negative attitude	.75	16 %

5. Learning a model of the user's empathic attitude

We applied the K2 Bayesian network learning algorithm (Cooper and Herskovitz, 1992) which is employed in Bayesware (http://www.bayesware.com) to the corpus of 708 dialog moves. This algorithm is appropriate to our case, as all our variables are discrete or may be discretized easily (move length), the number of nodes is small and there are no missing values in our data. Variables introduced belonged to the following categories:

a. *Stable and known subject characteristics*: background (in humanities or in computer science) and gender;

b. *Dynamically evolving, unknown subject characteristics*: empathy towards the ECA;

c. *Single move, easily recognizable characteristics:* context (type of previous agent's move) and subject's move length;

d. *'hidden' subject move characteristics:* these are the signs of subject's empathy described in Section 3.2 (friendly self-introduction, irony and so on);

e. *'observable' linguistic features of the subject move:* we developed a very simple parsing method, to recognize only six categories of 'words' or expressions: question marks, personal words about 'self' or about

‘the other’, friendly self-introduction and farewell expressions, and a few more.

We tested several kinds of models, by setting a search order of variables which considered their semantics and the intended use of the model: stable and known subject characteristics were taken as network roots, linguistic features as leaves and other variables as intermediate nodes. We initially learned automatically both the structure and the parameters of the network, and subsequently refined interactively the ‘best’ model found in terms of MLL (maximum logarithmic likelihood) by introducing a few links between some nodes, to avoid problems in evidence propagation due to d-separation properties or to too long paths between observable nodes and the main node to monitor (friendship). The resulting structure is shown in figure 2. We plan to employ the model learnt to build a dynamic belief network in which every time slice corresponds to a user move, the prior of the user’s gender (Gend) and background (Back) do not change from time to time, and the only temporal link between two consecutive time slices is assigned to friendship (Frship). The model will develop as follows:

- *at the beginning of interaction*, the BN is initialized by propagating evidence about stable user characteristics (gender and background); the probability of friendship takes a prior value which corresponds to this category of users;
- *after every user move*, the length of the user move (Lengh) is calculated, some of its linguistic features are analysed with the parser and these results are introduced and propagated in the network together with evidence about the agent’s move (Contx);
- *the new probability of the ‘friendship’ node* is read and contributes to planning the next agent move.

Combining results of parsing with knowledge about the subject, the context and the previous friendship attitude of the user enables us to reduce the diffuculty in the recognition of the user attitude by means of linguistic analysis which was acknowledged by several researchers (for instance, Gill and Oberlander, 2002; Carberry et al, 2002; Litman et al, 2003). As a next step of our research, we wish to test the predictive value of our model when used dynamically during the dialog to guess how the user friendship towards the ECA evolves.

Figure 2: outline of the user model

6. Conclusions

We learned a lot from our 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. This is true 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). On the contrary, 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 expressions of emotions in the agent's face. This opens complex problems, like recognising and responding to humorous acts (Stock and Strapparava, 2003), 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.

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 Haptek Inc. for assisting us in the application of their software.

References

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

S Carberry, L.Lambert and L.Schroeder. Towards recognising and conveying an attitude of doubt via natural language. *Applied Artificial Intlligence, 2002.*

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

G.F.Cooper and E.Herskovitz. A Bayesian method for the induction of probabilistic networks from data. *Machine Learning* , 9, 1992.

R. Coulston, S. Oviatt and C. Darves. 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.

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

C. Darves and S. Oviatt. 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.

F de Rosis, C Pelachaud, I Poggi, V Carofiglio and B De Carolis. 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).

F. de Rosis, A.Cavalluzzi, I.Mazzotta and N.Novielli: Can embodied conversational agents induce empathy in users? *AISB Virtual Social Characters Symposium, 2005.*

Haptek website: http://www.haptek.com

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

D. Litman, K.Forbes and S.Silliman: Towards emotion prediction in spoken tutoring dialogues. *Proceedings of HLT/NAACL* 2003.

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

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

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

A.Gill and J. Oberlander: Taking care of the linguistic features of extraversion. *24th Annual Conference of the Cognitive Sciene Society,* 2002.

S. Oviatt and B. Adams. 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.

R. Picard. 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.

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

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

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

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