humaine

D8e

Proceedings of the Workshop

"WP8 Emotion in Communication"

Place: ITC-Irst Povo, Trento, Italy, November 17-18, 2005

Oliviero Stock and WP8 members

ist

information
society
technologies

Date: 20th December 2005

Table of Contents

WORKSHOP PROGRAMME

Thursday, November 17 - 2005:

Morning session

09:00 – 09.30   **Welcome - Opening Address**
                       Roddy Cowie and Oliviero Stock

09:30 – 10:45   **Presentations:**

                       Emotional and Non-Emotional Persuasion
                       Maria Miceli, Fiorella de Rosis, Isabella Poggi

                       Persuasive goals of Irony, Humour and Ridicule in a political trial
                       Isabella Poggi, Federica Cavicchio, Emanuela Magno Caldognetto

                       On Emotion-Rich Persuasion Artifices
                       Irene Mazzotta, Fiorella de Rosis

11:15 – 12:30   **Presentations:**

                       Design of a Pilot Experiment for the Evaluation of Automated Persuasion systems
                       Marco Guerini, Shlomo Hareli, Shay Tzafrir

                       Affective weight of lexicon as an element for creative language production
                       Oliviero Stock, Carlo Strapparava, Alessandro Valitutti

                       Cantoche and WP8 Exemplar
                       Benoit Morel

Afternoon session

13:45 – 15:25   **Presentations:**

Backchannels and other Listener Actions in Argumentative Dialogue
Dirk Heylen

Backchannels Through Gaze as Indicators of Persuasive Success
Elisabetta Bevacqua, Maurizio Mancini, Christopher Peters, Catherine Pelachaud

More than just a pretty wording - Politeness in ECAs
Matthias Rehm, Elisabeth André

Proposing new ways of resolving online conflicts: an intelligent facilitation of forgiveness in CMC
Asimina Vasalou, Jeremy Pitt, Guillaume Piolle

15:45 – 17:00   **Demos of initial prototypes:**

Automated Persuasion: The Promoter prototype
Marco Guerini, Oliviero Stock, Massimo Zancanaro

Atelier Bari
Fiorella de Rosis and colleagues

Automated Humor, initial steps
Oliviero Stock, Carlo Strapparava, Alessandro Valitutti

17:00 – 18:00   **Special session on ethical issues**:

Ethical issues arising from workshop presentations confronted with Applied Ethics and Principlism
Paolo Petta and Sabine Döring

Friday, November 18 - 2005:

09:00 – 10:00   **Invited Talk:**

                       **The Ontological Semantics of Emotion, Humor, and
                       Persuasion**
                       Victor Raskin (Purdue University)

10:00 – 10:45   **Posters and Coffee:**

                       **Affective Dialogues with an ECA: from information-oriented
                       to persuasion-oriented**
                       Fiorella de Rosis, Giuseppe Clarizio, Ian O'Neill, Philip Hanna

                       Affective and Persuasive Guide
                       Mei Yii Lim, Ruth Aylett, and Christian Martyn Jones

                       **An emotional approach to designing self-awareness
                       mechanisms for CMC**
                       Asimina Vasalou, Jeremy Pitt

10:45 – 11:45   **Relations across WP's**
                       coordinator: Roddy Cowie

11:45 – 13:00   **Operational meeting & conclusive panel**

ORAL PRESENTATIONS

Emotional and Non-Emotional Persuasion

Maria Miceli
Institute of Cognitive Sciences and
Technologies
National Research Council
Via San Martino della Battaglia 44 – 00185
Roma
maria.miceli@istc.cnr.it

Fiorella de Rosis
Department of Informatics
University of Bari
Via Orabona 4 – 70126 Bari
derosis@di.uniba.it

IsabellaPoggi
Department of Education
University Roma Tre
Via d. Castro Pretorio 20 - 00185 Roma
poggi@uniroma3.it

A relevant issue in the domain of natural argumentation and persuasion is the interaction (synergic or conflicting) between 'rational' or 'cognitive' modes of persuasion and 'irrational' or 'emotional' ones. This work provides a model of general persuasion and emotional persuasion.

Our model takes the persuader's P perspective, thus focusing on P's theory of the recipient's R mind, and P's planning strategies for influencing R, that is, for changing R's mental state so as to make her intend to do a certain action or plan. We also try to circumscribe the notion of persuasion in relation to such criteria as 'success' versus mere 'attempt' at persuasion, 'accidental' versus 'intentional', 'communicative' versus 'non communicative', 'manipulative' versus 'non manipulative', and 'coercive' versus 'non coercive' persuasion.

A qualifying feature of our model is the attempt to integrate emotional and non emotional persuasion. Emotional persuasion is viewed as just a sub-case of general persuasion. As its non emotional counterpart, emotional persuasion is aimed at generating, activating, or increasing the strength of, R's goals, so as to induce in R some intention instrumental to such goals. The specificity of emotional persuasion lies in the means used for accomplishing this task. That is, when using an 'emotional' strategy, P tries to generate, activate, etc. R's goals through the medium of either R's emotions or R's beliefs and goals *about* her emotions.

While showing how the appeal to emotions is grounded on the strict and manifold relationship between emotions and goals (which is, so to say, 'exploited' by a persuader), we identify two general modes of emotional persuasion: *persuasion through actual arousal of emotions* and *persuasion through appeal to expected emotions*. In persuasion *through arousal of emotions*, P's intention to modify R's beliefs or their strength is a means for P's superordinate goal to arouse an emotion in R, which in turn is a means for P's further super-goal to generate a goal in R, and then an intention instrumental to it. For instance P's saying to R 'How disgustingly fat you are!' is meant to provoke R's shame, which should generate R's goal of not losing her face, and induce, as a means for this goal, her intention to go on a diet. In persuasion *through appeal to expected emotions*, P's intention to modify R's beliefs or their strength is a means for P's super-goal to activate or increase the strength of R's *goal of (not) feeling a certain emotion*, and to induce in R an intention instrumental to this goal. For instance, P's saying to R 'If you are kind to John, you will feel at peace with your conscience (or you will not feel guilty)' is meant to activate R's goal to feel at peace with her conscience (or not to feel guilty), in order to induce in R the intention to be kind to John as a means for it.

We argue that the rational/irrational, as well as the argumentative/non argumentative dimension, do not allow to distinguish such forms of persuasion from the non emotional ones. Actually, one mode of emotional persuasion, the appeal to expected emotions, can be perfectly rational, as long as 'rational' implies the correct processing of the information available, the derivability of conclusions from premises, and the production of plausible means-ends relationships. An appeal to expected emotions is 'structurally' indistinguishable from any other 'argument from consequences' or, in our terms, 'intention generation by acting on pre-existing goals'. The only difference resides in the *content* of the goal on which P acts: in the appeal to expected emotions, this content is precisely that of 'feeling' a certain emotion rather than having a certain state of the world true.

The other mode of emotional persuasion – persuasion through arousal of emotions – works very differently from the former, in that the aroused emotion (say, shame) can directly produce a certain goal (say, to save one's face), independent of R's reasoning and planning about means-ends relationship. Therefore, this is no doubt a form of non argumentative persuasion. However, in this context 'non argumentative' should be made equal to 'a-rational', rather than 'irrational' (as long as 'irrational' implies going *against* the dictates of reason). Moreover, it should be stressed that the direct production of a goal through emotional arousal is just one step which is generally included in a more complex persuasion strategy expecting a very 'rational' planning and behaviour on R's part. That is, once a certain goal (like saving one's face) is emotionally produced, R's reasoning and planning can be, and generally are, called into play in view of its achievement. (For instance the goal to save face, together with the belief that being in shape is a means for achieving it, is likely to generate R's goal of being in shape, and the consequent planning as a means for achieving it.) Thus, even such a form of persuasion is partially based on the recipient's reasoning and planning abilities, which testifies to the constant mingling and intertwining of rational and a-rational ingredients in most persuasion strategies.

We propose a method to formalize and represent persuasion strategies as oriented graphs, and show how emotional and non emotional strategies (and also emotional and non emotional components in the same strategy) may interact with, and strengthen, each other. We finally address the role of uncertainty in persuasion strategies and show how it can be represented in persuasion graphs. Most of our working examples will refer to the domain of healthy eating, in which a differentiated corpus of examples may be found, ranging from advertising to messages produced by scientific agencies delegated to the promotion of a correct behaviour in this domain.

Persuasive goals of Irony, Humour and Ridicule in a political trial

Isabella Poggi
Dipartimento di Scienze dell'Educazione
Università Roma Tre, Roma
poggi@uniroma3.it
Tel. 06/49229296

Federica Cavicchio
Istituto di Scienze e Tecnologie della
Cognizione del CNR
Sezione di Fonetica e Dialettologia, Padova
cavicchio@pd.istc.cnr.it
Tel. 049/8274418

Emanuela Magno Caldognetto
Istituto di Scienze e Tecnologie della
Cognizione del CNR
Sezione di Fonetica e Dialettologia, Padova
magno@pd.istc.cnr.it
Tel. 049/8274418

In a judicial debate, a typical goal of a judge is to assess the credibility of accused and witnesses. As to witnesses, s/he must assess if they tell the truth both about events, mainly by assessing how certain they are of what they say (if by direct witnessing, or only by having heard, or by their own inferences). As to the accused, a judge must assess not only factual states of affairs, but also, and more importantly, his/her intention and awareness of committing a misdeed. A trial is then, as acknowledged by classical rhetoric, which distinguished among deliberative, judiciary and epideictic rhetoric, a typical case of persuasion. The attorney general, for example, must persuade the judge that the accused did know that s/he was committing some transgression, while the accused aims, in turn, to convince that if s/he caused some damage to someone or transgress some norms s/he was not aware of this.

But there are trials in which this aim is pursued also with the arms of irony, humour, and ridicule. A such case is a very important and famous trial, "Mani pulite", held in Italy in 1993, which had a very high political importance because many politicians were charged of corruption.

Giglioli et al. (1999) put it, this trial was a "degradation ritual" (Goffmann, 1967) in that a whole political class was impeached: a situation in which men of a formerly very high status and high social and political power were publicly shamed.

In this process, the general attorney, Antonio Di Pietro, and the counsel for the defence, Spazzali, often use irony while interrogating the accused and the witnesses; but their questioning provide examples of humour and ridiculisation too. Conversely, also the accused, Cirino Pomicino, sometimes makes use of ridicule and irony in response to the accursers' communicative behaviour.

In this work we provide a definition of irony, self-irony, ridicule, humour and involuntary humour. We provide examples of these behaviours by analysing some fragments of the "Mani Pulite" trial. The fragments are analysed both as to their multimodal cues (how the overall multimodal behaviour of an Agent reveals the humorous or ironic import of his move), and as to their persuasive goals.

To describe multimodality, Magno Caldognetto and Poggi (2001) suggest a method: the Multimodal Score, a procedure to transcribe and analyse the multimodal signals classified separately and in their mutual interaction. This method allows us to transcribe on five parallel lines, like in a musical score, the communication items transmitted at the same time in five modalities: *speech, prosody, gesture, facial(mouth, gaze, eyes, eyebrows), head and body posture*, by labelling signals on five different levels. In the Score, each signal of each modality goes through five levels of analysis:

- *description*: the gesture or movement is described on the basis of its perceptual characteristics. For example, in the gesture line it is possible to describe gesture both in a word transcription ("right hand draw an arch with the index") or in a codified transcription systems, as cheremes (the minimal unit of gestural communication, the "phonemes" of gesture, Stokoe 1980). Descriptions of facial and body movements are, for example: "eyebrow raising", "wide shut eyes" etc.
- *descriptive typology*: the gesture or movement is classified on the basis of a typology of gestures, including also self touch and not communicative case such as "hands at rest".
- *meaning*: the movement analysis is paraphrased with words or phrases;
- *meaning typology*: the meaning of each movement or gesture is classified on the basis of a semantic taxonomy that distinguishes Information about the World, the Speaker' Identity and the Speaker's Mind;
- *semantic function*: by comparing the gesture or movement with the coproduced speech signal, five different "functions" are distinguished, that is five kinds of relationships between them: **repetition**, if it bears the same meaning, **addition** if it adds information to word meaning, **substitution** if it replaces a word that is not uttered at all, **contradiction** if it communicates something opposite to what said by words, or **no relationship**, if it makes part of a different communicative plan.

The Multimodal Score method is implemented in M. Kipp's ANVIL (ANnotation of VIdeo and Language, 2001). This application, here presented in 4.0 version, is used by ISTC-CNR of Padua. At the Institute we are developing, with respect to the visual display planned by M. Kipp, more analytic evaluations linked to the acoustic analysis through PRAAT (Boersma, 1996).

Concerning the analysis of the persuasive goals of these strategies, it is carried on in terms of a goal and belief model of persuasion. From this analysis it results that both in the attorneys and the accused, humour, ridicule and irony have persuasive goals. They all have, as often these strategies do, an aggressive goal, in that they all aim at abasing the image of the victim, they typically aim at blowing on his reputation; but while the very bulk of these strategies in the attorneys is to persuade the judge that the accused is not credible in showing innocent or unaware of what he was doing, the accused uses them mainly to stress how the attorney is taking advantage of his new power coming from the media visibility of the trial. The use of irony, humour and ridicule was quite justified by the very goal of re-establishing a balance between the politicians who, in a sense, did not deserve such power, and the people. In different judicial contexts the use of irony, humour and ridicule might be less frequent because they could be viewed as a politically incorrect and pointlessly cruel behaviour if addressed to laypeople in such an uncomfortable situation as a trial.
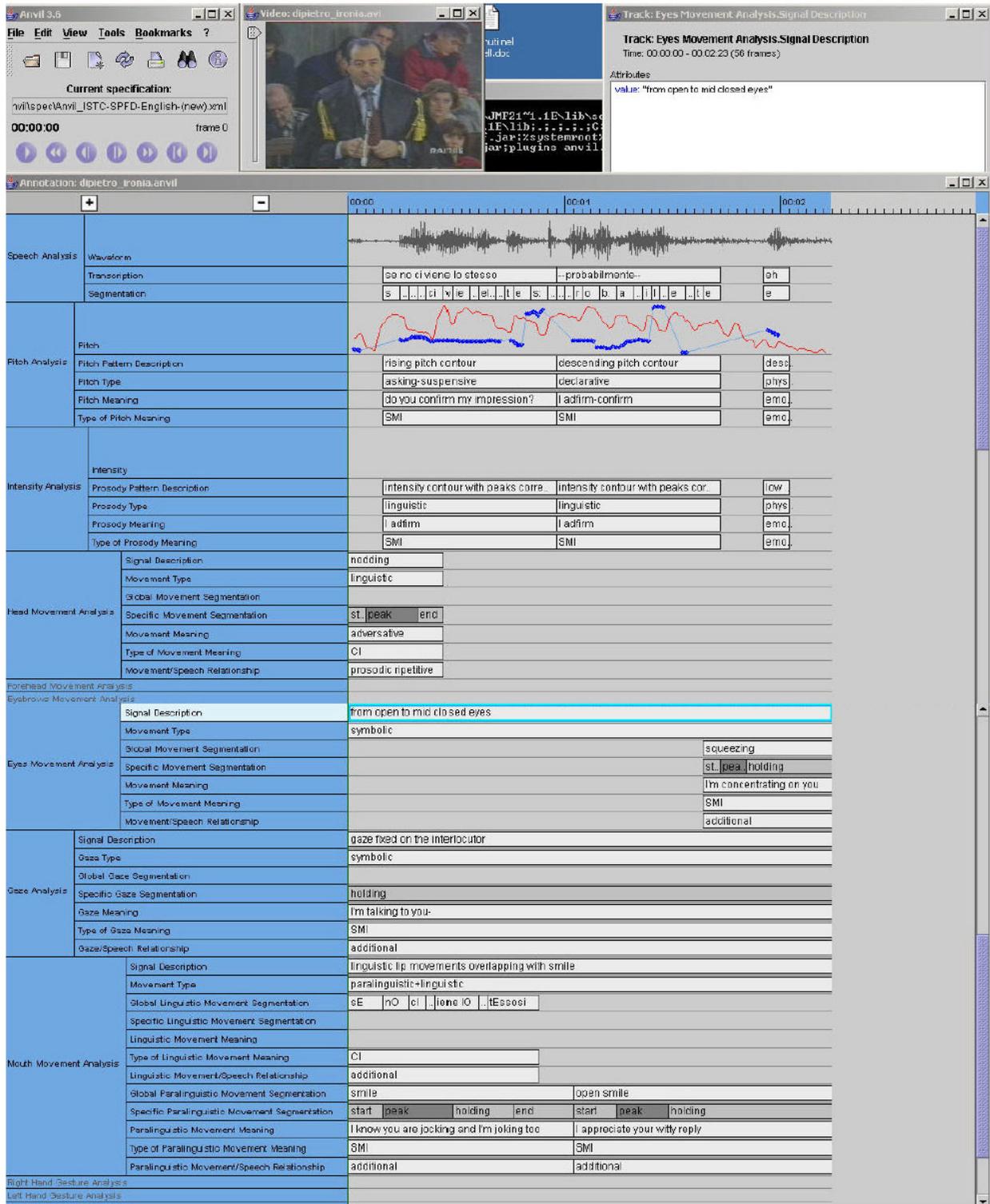
Fig. 1. An example of Multimodal Score in ANVIL display.

References

Boersma P., (1996) "PRAAT, a System for Doing Phonetics by Computer", Glot. International 5 (9/10), pp. 341-345 (PRAAT web site: http://www.fon.hum.uva.nl/praat/).

Giglioli P.P., Cavicchioli S. & Fele G., Rituali di degradazione. Anatomia del processo Cusani. Bologna: Il Mulino 1997.

Goffmann E., Interaction ritual. New York: Doubleday Anchor Books 1967

Kipp M., "From Human Gesture to Synthetic Action", in Pelachaud C. and Poggi I. (Eds.), Multimodal Communication and Context in Embodied Agents. Proceedings of the Workshop W7 at the 5th International Conference on Autonomous Agents, Montreal, Canada, 29 May 2001, pp .9-14.

Magno Caldognetto E. e Poggi I., "Dall'analisi della multimodalità quotidiana alla costruzione di agenti animati con facce parlanti ed espressive", in: Cosi P. e Magno Caldognetto E., (Eds.), Multimodalità e multimedialità della comunicazione, Atti delle XI Giornate di Studio del G.F.S., Padova 29-30 novembre, 1 dicembre 2000, Padova, Unipress, 2001, pp. 47-53.

Stokoe, W. C., (1980) "Sign Language Structure", Annual Review of Anthropology, 9, pp. 365-390.

On Emotion-Rich Persuasion Artifices

Irene Mazzotta and Fiorella de Rosis
Intelligent Interfaces, Department of Informatics, University of Bari
Via Orabona 4, 70126 Bari, Italy
{mazzotta, derosis}@di.uniba.it
http://www.di.uniba.it/intint/

Introduction

A few years ago, we proposed a probabilistic formalism to represent knowledge in Walton and Reed's argumentation schemas: we demonstrated how this kind of knowledge base could be used, at the same time, to generate receiver-tailored persuasion messages and to respond to subsequent 'critical questions' (Carofiglio, 2004). We subsequently cooperated with ISTC-CNR in studying how this formalism might be employed, as well, to represent some form of 'a-rational' persuasion (see Miceli et al, in this Workshop) in which 'purely rational' arguments are combined with 'emotional' ones. In the two mentioned steps of our research, we worked on examples of daily life rather than on an organic corpus of data. To root the range of envisaged strategies on concrete data, we subsequently decided to collect a more systematic corpus: this paper describes how this corpus was collected and analyzed and draws some conclusions on the validity of bayesian networks in representing this body of knowledge.

A small corpus of persuasion messages about healthy eating

Eating habits are the result of cultural, psychological, educational and life style factors. As such, they consolidate in time and, when wrong, are quite difficult to modify. Information media are masters in employing tricky arguments to persuade the population to consume products of doubtful healthiness: according to the local culture, irony or sex are the strings most frequently touched by advertizing. Attempting to contrast this pressure to persuade the population to adopt more 'appropriate' habits, by employing only 'rational' and 'scientific' arguments is probably not effective. This is, therefore, one of the domains in which artifices are justified, if not needed. Of course, as suggested by Walton (1992), attention should be paid to insure that arguments are relevant and strong: this is a subjective judgement which depends on the persuader's personality and on the context.

To assess which artifices are employed by people with no particular competence on healthy eating education, we collected a corpus of persuasion messages with a website (http://www.di.uniba.it/intint/H-persuasion-bi.html). A scenario was presented initially to describe the 'situation' which the subjects participating to the study (the 'persuaders') should imagine to find in:

"Mary, one of your best friends, is a 25 year old girl who follows a wrong diet. She does not eat much fruits and vegetables while tends to overeat meat, sweets and pasta. Try to persuade her to eat more fruits and vegetables and, in doing so, don't forget that Mary is famous for her obstinacy!

You know the following fact: eating fruits and vegetables is good for health. They are good sources of vitamin A and C, which are important for growth and repair of body tissues, to cleanse the blood and

give resistance against colds. Moreover, various epidemiological studies proved that a diet rich in vitamin A and C decreases the risk of coronary heart diseases and stomach cancer.

In addition consider that health is very important for Mary: she likes sports, undergoes periodical check ups and looks at TV programs about health care. Mary would have enough free time to cook vegetables and delicious fruit dishes.

Please, use this information to write a text (from 5 to 10 lines) to argue about your thesis."

The scenario was formulated to suggest the following keypoints in formulating the strategy:

    a.   Friendship relation between persuader P and recipient R (Mary in the previous example)
    b.   R's personality (to be obstinate)
    c.   R's goal (to be in good health, in the previous scenario)
    d.   R's living habits (makes sports, undergoes checkups, looks at specialized TV programs)
    e.   Holding of conditions to make the action p possible (R has time to prepare vegetables)
    f.   Relationship between desired action (eat vegetables) and likelihood to achieve R's goal.

Cognitive dissonance was implicitly assumed in R's mind. The hypothesis was that, in conditions of cognitive coherence, the intention to perform some action should be a consequence of a set of beliefs, goals and conditions which make the action possible. In the scenario, on the contrary, premises were presumed to be true while the consequence was not. This case of cognitive dissonance was similar to the smoking example originally formulated by Festinger (1957). Four variations of this scenario were displayed randomly to the subject, which differed for two variables:

employing positive vs negative arguments and mentioning consequences on health vs appearance. The previous scenario corresponds to the 'positive arguments' and 'mentioning consequences on health' condition. We collected, overall, 33 messages from subjects aged between 23 and 63 year of age, of both genders, with various backgrounds. We will examine them very shortly to single out variations in strategies and to reason on their possible formalization, while we plan to extend this description at the Workshop.

Outline of Corpus Analysis

The hypothesis behind persuasion is that cognitive dissonance, being psychologically uncomfortable, motivates the persons to reduce it and leads them to avoid information likely to increase it (Harmon-Jones and Mills, 1999). The persuasion process may therefore aim at reducing the dissonance by increasing the desirability of the outcome and/or by reminding information (unknown or avoided) about the activity-outcome relation. In addition, as it has been claimed that being aware of the dissonance may produce a motivation that results in genuine cognitive changes, specific reference to inconsistency between the referents' beliefs and goal and their behaviour can be made, to strengthen the persuasion power of a message. 'Rational' and 'a-rational' arguments may be employed to produce the mentioned effects. The combination of the various factors produces a large variety of 'artifices' whose validity depends on the specific situation. Let us try to make some examples from our corpus of data.

We start from the hypothesis that, in conditions of cognitive coherence, the following implication holds

$$[(\text{V-Goal R q}) \wedge (\text{A-Goal R q}) \wedge (\text{Bel R } (\text{Do}(R,p) \rightarrow \Diamond q)) \wedge \text{Bel R CanDo}(R,p))] \Rightarrow \text{IntToDo}(R,p)$$

where p is a variable denoting an action (e.g.: p = to eat vegetables); q is a formula denoting states of the world that may include agents such as R (e.g.: q = R is in good health); $\Diamond$ q denotes states of the world which are holding in a more or less near future; Bel, Int, A-Goal, V-Goal are modal operators to denote the various aspects of the mental state of agent R which are relevant in persuasion processes, that is, respectively, beliefs, intentions, active-goals and valued-goals. In particular (V-Goal R q) means "Goal q is of high value to R"; (A-Goal R q) means "Goal q is active in R's mind"; (Bel R (Do(R,p)→ $\Diamond$q)) means "R believes that doing p implies achieving q"; ((Bel R CanDo(R,p)) means "R believes that conditions hold to make p"; IntToDo(R p) means "R has the intention to make p" (Miceli et al, in press)

The main goal (the claim) of the persuasion message is to recommend the activity by acting on the intention to perform it: IntToDo(R p). This goal may be achieved by combining various techniques which aim at reducing the cognitive dissonance by taking, as a their target, different items in the previous implication:

1) in attempting to increase the desirability of the outcome, the target is (V-Goal R q) but also (A-Goal R q);

2) in attempting to remind information about activity-outcome relationship, the target is (Bel R (Do(R,p)→ $\Diamond$q));

3) in attempting to prove that conditions exist for performing the activity, the target is: Bel R CanDo(R,p)

4) in attempting to produce a motivation based on evoking the cognitive dissonance, the target is the inconsistency between the receivers' beliefs and goal and their behaviour.

We analyzed the texts in our corpus by trying to find out whether and how each of these techniques was implemented. Let us start with a first example[1]:

Mary, I believe you should eat more fruits and vegetables. Aim: recommend the activity. Target: IntToDo(R p)
By making sport, you should know that vegetables are good for health! They strengthen muscles and bones as they are rich in minerals. Aim: remind information about activity-outcome relationship. Target: Bel R (Do(R,p)→ $\Diamond$q))

[1] Translated form Italian: we apologize with our subjects for the bad translation of their very rich texts!

The rational strategy adopted is enriched by exploiting evidence about the referent which is in favour of the belief ('by making sport, you should know…").

Especially after making sport, a good quantity of fresh season fruit tonifies and rehydrates the body after the big toil!

Same aim and target as in the previous sentence.

Emotional items are introduced in the style ('fresh season fruit', 'big toil': 'faticaccia', in Italian)

Without counting the benefits of vitamins A and C for skin and hair!

Always the same aim and target

Maybe you might get rid of some portion of meat or sweets, to leave more space to fruits and vegetables!

The persuader suggests a plan to implement the activity.

This text is very simple: it is a nearly rational message that we will take as a reference schema in our next analysis. Other items may be added to this basic schema, with different purposes:

a. *to increase the desirability of the goal* (…you pretend you care for your health!, …a person like you, who cares so much for her health!)

b. *to prove that conditions exist for making the activity* (…as you have time at your disposal, …you may find some excellent vegetables and fresh fruits…)

c. *to evoke the dissonance* (…And you, who care so much for being well, you don't think to that?)

However, very few of the messages in the corpus were formulated according to the 'nearly-rational' schema. This occurred primarily when the persuader's background was scientific (computer science inparticular). On the contrary, the majority of subjects with a humanistic background employed much more refined and emotional techniques. A few examples:

- the recommend the activity section is usually introduced at the beginning of texts which tend to be rational, while it is introduced only subsequently in less rational texts, after preparing the subject to receive the message. In some cases, the activity is not recommended explicitly; the recommendation is substituted with the description of some tempting consequences of the activity: …tomorrow you invite me at home for dinner and we only eat vegetables and fruits, OK?

- the *desirability of the goal* and its activation may be dealt with in an emotional way: …try to think, Mary, to how much more beautiful and healthy you might appear and be!

- *proofs that conditions exist for making the activity* may be given, as well, in emotional form: …you, who have time and may enjoy in preparing food…

In addition to the techniques mentioned so far, other persuasion strategies were employed by our subjects:

- they *introduced higher-order goals* in the text, such as 'to live in a natural way', 'to satisfy gluttony', 'to enjoy', 'to make friends' that the activity might contribute to achieve: …you would contribute to the life of biological peasants, …you may always enjoy in preparing gorgeous vegetable meals.

- they *appealed* in a more or less explicit way *to emotions*: …here is the sagacity of experienced women: you have the creative intelligence on your side… (*pride*), …I would be delighted to meet you and discuss pleasantly with you… (*attraction*).

Conclusions

In this short paper we described how we collected and analyzed a corpus of persuasion texts in a particular domain: healthy dietary behaviour. With this analysis, we wanted to single out the strategies that were adopted by our subjects in producing a persuasive text: the preliminary results we have got proved that purely rational strategies were employed very unfrequently and that emotional elements could be found everywhere, in various forms. As proposed in a previous paper, (Carofiglio, 2004), rational strategies may be formalized with bayesian networks; however, representing a-rational strategies with this formalism without loosing richness of employed artifices will be much more difficult.

Acknowledgements

Main References

V.Carofiglio: Modelling argumentation with belief networks. Workshop on Computational Models of Natural Arguments (CMNA'04).Valencia, August 2004.

L.Festinger. A theory of cognitive dissonance. Evanston, IL. 1957

E.Harmon-Jones and J.Mills. Cognitive dissonance: progress on a pivotal theory in social psychology. Americal Psychological Association Books, 1999.

P. E. Jorgensen: Affect, persuasion and communication processes. In P. A. Andersen and L. K. Guerrero (Eds): Handbook of Communication and Emotion. Academic Press, 1998.

M. Miceli, F. de Rosis and I. Poggi: Emotional and non emotional persuasion. Applied Artificial Intelligence: an International Journal. In press

D. Walton: The place of emotion in argument. The Pennsylvania State University Press, 1992.

Design of an Experiment for the Evaluation of the Promoter Persuasion System

Marco Guerini (ITC-Irst)
ITC-irst, Istituto per la Ricerca Scientifica
e Tecnologica
I-38050 Povo Trento Italy
guerini@itc.it

Shlomo Hareli
University of Haifa
Haifa 31905, Israel
shareli@gsb.haifa.ac.il

Shay Tzafrir
University of Haifa
Haifa 31905, Israel
stzafrir@research.haifa.ac.il

Abstract

Persuasive interfaces need to be evaluated: they have to be effective with real users. Evaluation studies up to now have focused mainly on *usability*. Evaluation of *effectiveness* is not straightforward, specific methodologies have to be used/defined.

In this paper we describe the design of an evaluation experiment of a persuasive system, for multimodal Computer Human Communication (CHC), called *Promoter*. The overall aim of the experiment is to investigate the role of emotions in persuasive HCI.

General description

Since there are no well defined methodologies for measuring the "persuasiveness" of a system, and in particular the role of emotions, yet, we faced a complex problem. We decided to ground our work on experiments/measures coming mainly from Social Psychology.

The design of this evaluation experiment serves two purposes:

- An evaluation study of the specific automated persuasion system (*Promoter*)
- A general model for evaluation of automated persuasion systems.

The Specific Question we address in the experiment is: *Can display of emotion enhance the persuasive force of a message in HCI*?

1) For **message generation** we focused on Content Ordering Meta-Strategies (COMS). COMS are persuasive strategies that allow to state, given basic persuasive strategies (pieces of information), how to compose them to create a complex message.

2) The theory used for **emotion displaying** by the *Promoter* prototype is the *Fitting principle* which suggests that a message that is accompanied by an emotional expression that fits the valance of the message will be more persuasive than the same message emitted with no emotion or an emotional expression that contradicts the message. For example: a congratulation made with an angry face results less effective than the same congratulation made with an happy face.

Overall structure of the experiment

The subjects are conducted to a room where there is a computer. They are told that they will have to watch a short clip showing a synthetic face advertising a new exhibition at the Hecht museum. As a c*over story* we lead them to believe that they have to judge an actual advertisement campaign without letting them know that the ECA is being tested.

We want the experiment to be in the context of intelligent systems that provide "personalized" messages (even if in the case of this experiment they will be pre-recorded clips). We will use a fake on-line pre-questionnaire to give to the subjects the impression of a personalized message, by saying that the answer to this pre-questionnaire will be used for the generation of a tailored message. In particular the Need For Cognition (NFC) questionnaire will be exploited so that the answers can be used in the analysis step of the experiment.

After the presentation of the message the subjects must fill in a questionnaire. Specific evaluation of some questions will be taken into consideration depending on the persuasive strategy (COMS) being tested. For example: we could expect that with some strategies the reliability of the source is rated much more since the aim of these strategies is exactly this.

Aims Conditions of the experiment

The pilot experiment will be run with one COMS strategy on 3 different conditions (between subjects, each person gets only one condition) to understand the role of emotions (according to the fitting principle) in persuasive communication. The 3 conditions are:

- Fitting
- Unfitting
- Neutral, no emotion, just message

Structure of the questionnaire

The structure of the questionnaire is divided in five main parts, each investigating a specific topic.

- *Attitude change*: list of questions, to be rated. E.g.: How interesting the museum seems to you? How good you think it is? Would you go to the exhibition if you had a chance? Etc.
- *Memorization*, to test, among others things, the degree of involvement of the subject in the topic.
- *Source credibility*: list of adjectives referring to the source, to be rated. E.g.: clever, competent, credible, expert, informative, etc.
- *Information Quality*: list of questions, to be rated. E.g.: How much did you trust the information provided? How helpful was the information? How insightful was the information?
- *Manipulation check*: for emotion recognition (according to Ekman theory we want to understand how much the subjects recognized the emotions).

Expected Results & Future Work

We expect that the fitting message will be more persuasive than the other messages. The unfitting message will be the less persuasive relative to the other messages.

In particular, as a pilot experiment we will use a very simple COMS that generates messages made with positive pieces of information only.

We can expect, from studies already done in the HHI and HCI fields, that each persuasive strategy alone works (each COMS, the fitting principle).

What we want is to obtain a broader result, not just the replication of studies already done. We want to test is the effectiveness of the *interaction* of different strategies (between fitting principle and COMS strategies): there are no results about their mixing.

Since *Promoter* uses four main content-ordering persuasive strategies, the experiment will be replicated conceptually three times, one for each COMS strategy.

Affective Weight of Lexicon as an Element for Creative Language Production

Oliviero Stock
ITC-irst, Istituto per la Ricerca
Scientifica e Tecnologica
I-38050 Povo Trento Italy
stock@itc.it

Carlo Strapparava
ITC-irst, Istituto per la Ricerca
Scientifica e Tecnologica
I-38050 Povo Trento Italy
strappa@itc.it

Alessandro Valitutti
ITC-irst, Istituto per la Ricerca
Scientifica e Tecnologica
I-38050 Povo Trento Italy
alvalitug@itc.it

Introduction

All words can potentially convey affective meaning. Each of them, even those more apparently neutral, can evoke pleasant or painful experiences. While some words have emotional meaning with respect to the individual story, for many others the affective power is part of the collective imagination (e.g. words "mum", "ghost", "war" etc.). Therefore, it is interesting to individuate a way to measure the affective power of a generic term. To this aim, we studied the use of words in textual productions, and in particular their co-occurrences with the words in which the affective meaning is explicit. As claimed by Ortony et al. [1], we have to distinguish between words directly referring to emotional states (e.g. "fear", "cheerful") and those having only an indirect reference that depends on the context (e.g. words that indicate possible emotional causes as "monster" or emotional responses as "cry"). We call the former *direct affective words* and the latter *indirect affective words*.

The main contributions of this work consist on (i) the organization of the direct affective words and synsets inside WordNet-Affect, an affective lexical resource based on an extension of WordNet, and on (ii) a selection function (named *affective weight*) based on a semantic similarity mechanism automatically acquired in an unsupervised way from a large corpus of texts (100 millions of words), in order to individuate the indirect affective lexicon. Then we employed these two resources for the automated generation of evaluative expressions. For example, given in input the noun \teacher" and the indication of a positive valence, system generates noun phrase "encouraging teacher", while with the input noun "war" and negative valence, it returns "oppressive warfare".

A careful selection of affective words plays a crucial role in communication. In particular, the automated generation of evaluative expressions with a bias on some valence orientation are a key component for automatic personalized advertisement and computational humor [2].

WordNet-Affect and Emotional Hierarchy

WordNet-Affect is an extension of WordNet database [3], including a subset of synsets suitable to represent affective concepts. Similarly to our method for domain labels [4], we assigned to a number of WordNet synsets one or more affective labels (*a-labels*). In particular, the affective concepts representing emotional state are individuated by synsets marked with the a-label emotion. There are also other a-labels for those concepts representing moods, situations eliciting emotions, or emotional

responses. WordNet-Affect is freely available for research purpose at http://wndomains.itc.it. See [5] for a complete description of the resource.

Recently, we extended WordNet-Affect with a set of additional a-labels, hierarchically organized, in order to specialize synsets with a-label emotion. Hierarchical structure of new a-labels was modeled on the WordNet hyperonym relation. In a second stage, we introduced some modifications, in order to distinguish synsets according to emotional valence. For example, the synset fear#1 includes hyperonym horror#1 (with negative valence), but also frisson#1 (with positive valence).

Another important property for affective lexicon concerning mainly adjectival interpretation is the stative/causative dimension [6]. An emotional adjective is said *causative* if it refers to some emotion that is caused by the entity represented by the modified noun (e.g. "amusing movie"). In a similar way, an emotional adjective is said *stative* if it refers to the emotion owned or felt by the subject denoted by the modified noun (e.g. "cheerful/happy boy").

Affective Weight

In order to characterize indirect affective lexicon, we consider the notion of semantic similarity between words, and chose to implement it with Latent Semantic Analysis (LSA). LSA yields a vector space model with reduced dimensionality for representing in a homogeneous way the meaning of words, word sets and texts [7]. The dimensions of the latent semantic space are the most explicative principal components of the terms-by-documents matrix. For the present work, we run LSA on the British National Corpus, a very large (over 100 million words) corpus of modern English, both spoken and written. Moreover, LSA provides a formalization of semantic similarity between words and word sets in terms of their co-occurrence frequence in texts. This allows us to introduce the notion of *affective weight*.

We define affective weight of a given term respect to an emotional category as semantic similarity between the term and the set of (direct affective) words included in the emotional category. For example, the term "problem" has high affective weight with respect to the positive emotional category identification, to the negative category negative concern, to the ambiguous valence tagged category ambiguous hope, and to the neutral apathy.

Evaluative Expression Generation

Among the possible application of WordNet-Affect and of affective weight function, we chose to implement a procedure for the automated generation of evaluative expressions. These expressions are composed by a part referring to the evaluated object (named *target*) and a part expressing the affective evaluation on it. For example, target can be represented by a noun and the evaluation by an adjective, in order to generate an expression consisting of a noun phrase. The procedure gets in input a generic term and a _xed value of valence and creates the corresponding LSA-vector. Then, the system selects the emotional category with the input valence and the maximum value of affective weight. Finally,

depending on the type of required expression, target-term and and evaluative-term are selected and the corresponding expression is composed.

For example, if we give in input the verb "shoot" with negative valence, the system individuates the emotional category horror. Then, it extracts the target-noun "gun" and the causative evaluative adjective "frightening" and finally generates noun phrase "frightening gun".

Possible Applications

Creative Humorous Test bed. WordNet-Affect and the affective-weight function can be useful for computer assisted creativity. The automated generation of evaluative expressions with a bias on some valence orientation are at the basis of various possible applications such as automatic personalized advertisement, computational humor [2] and persuasive communication. In particular an automatic selection of affective words, given same constraints, plays a crucial role in the *Creative Humorous Test bed*, developed in the context of the Humaine project.

Verbal Expressivity of Embodied Conversational Agents. Emotions expression by synthetic characters is considered now a key element for their believability. Intelligent dynamic words selection is crucial for realizing appropriate and expressive conversations.

Sentiment Analysis. The emotional weight function can be employed in text analysis as a sentiment analysis technique (e.g. text categorization according to affective relevance, opinion exploration for market analysis, etc.)

References

[1] Ortony, A., Clore, G.L., Foss, M.A.: The psychological foundations of the affective lexicon. Journal of Personality and Social Psychology 53 (1987) 751{766.

[2] Stock, O., Strapparava, C.: Getting serious about the development of computational humour. In: Proceedings of the 8th International Joint Conference on Artificial Intelligence (IJCAI-03), Acapulco, Mexico (2003)

[3] Fellbaum, C.: WordNet. An Electronic Lexical Database. The MIT Press (1998)

[4] Magnini, B., Cavagli_a, G.: Integrating subject _eld codes into wordnet. In: Proceedings of LREC-2000, the Second International Conference on Language Resources and Evaluation, Athens, Greece (2000) 1413{1418

[5] Strapparava, C., Valitutti, A.: WordNet-A_ect: an affective extension of WordNet. In: Proc. of 4th International Conference on Language Resources and Evaluation (LREC 2004), Lisbon (2004) 1083 { 1086

[6] Goy, A.: Lexical semantics of emotional adjectives. In Feist, S., Fix, S., Hay, J., Moore, J., eds.: Proceedings of Student Conference in Linguistics. MIT Working Papers in Linguistics 37. MIT Press (2000)

[7] Landauer, T.K., Foltz, P.W., Laham, D.: Introduction to latent semantic analysis. Discourse Processes 25 (1998) 259{284

Cantoche exemplar proposal WP8

Benoit Morel
CEO Cantoche
bmorel@cantoche.com

The recent dramatic increase of online e-administration procedures and services is the guarantee that, finally, the Public Services have decided to strongly invest in the NTIC and specifically e-services for the benefit of its users! As a key innovating public actor in the field of technologies, the city of Issy-les-Moulineaux (close to Paris, France) has confirmed its leadership while it organizes in December 2005 the very first 100% online election for the renewal its "conseils de quartiers".

So as to promote and explain this election and the online procedure, the city has hired a virtual citizen, Sam, who is in charge of the (viral) communication surrounding the project. Sam is a cartoonish style character in order to break the austere image of the politicians and to create a friendly relationship with the user. Sam is powered by the Living Actor™ technology.

Cantoche built a web based application where Sam is acting by playing different scenarios depending of the user's choice.

- The first mission for Sam is to answer general questions that a citizen can have regarding the local life. The choice is large because the citizen has a variety of needs.
- The second mission is persuading the visitor to vote on-line. It's new, it's easy.
- The third mission is helping the visitor to be a real citizen by inviting other people to go to vote on-line. The visitor can "send" by mail Sam with the scenario of his choice to a friend. This is viral campaign

You can discover this environment with Sam at www.issyquartiers.com. You can also watch the Sam interview by a real journalist in the report published at www.issy.com.

Cantoche created an evaluation system to measure how many people sent a scenario, which scenario has been selected and how many people accepted the invitation.

Cantoche suggests starting from this experience for a potential exemplar as the environment already exists. We can duplicate this environment for every culture (languages, characters, scenarios) and measure the persuasion strength of a virtual character. We can define different methodologies for evaluation according to WP9. The work consists to:

- Choose the most appropriate character for each culture
- Write the scenarios in different languages
- Define the methodologies for evaluation
- Create several focus group

Backchannels and other Listener Actions in Argumentative Dialogue

Dirk Heylen

Human Media Interaction
University of Twente
heylen@ewi.utwente.nl
+31-53-4893745

The general assumption underlying day-to-day, face-to-face conversations is that they, in some way or another, comply with the Gricean maxims that govern cooperative conversational behavior: say what you want to say and express yourself clearly. Conversation always involves politics. Talking to some one, always involves trying to get something done. When conversing, we try to get our interlocutors to engage with us in a joint project. We utter words, so that they will be heart, listened to, understood and that the other engages in the proper acts that we are soliciting: attending, understanding and reacting appropriately: answering our questions, believing our statements, taking up our orders, etcetera. Conversational actions are undertaken to engage the addressees in taking actions in turn. When advising people, we hope people take the advise to heart. When cheering them up, we hope they feel happy for a while. When we argue in favour of a certain proposition, we hope that the other will become convinced of what we believe to be true.

In persuasive or argumentative conversations, just like with any for of action, it is important for speakers to check to what degree there actions are successful on all levels. In the case of conversations these checks involve monitoring the interlocutor to see his or her uptake of the joint projects proposed by the speaker. Through typical behaviours involving gaze[2], facial expressions, head movements, posture and vocal backchannels listeners show that they are engaged in the conversation, paying attention, showing they are interested in what is being said. In part, the speaker constructs his utterances in such a way that they elicit the required response. [Goodwin, 1981] has investigated some of the ways in which speakers and hearers establish mutual orientation.

> [S]peakers who do not obtain the gaze of a hearer may perform specific actions, such as the production of restarts and pauses. After such phrasal breaks, nongazing hearers generally begin to move their gaze toward the speaker; if they do not, the speaker may continue to produce phrasal breaks until he obtains an appropriate response.

Another case where speakers elicit responses of hearers during their turn involves the request for backchanneling.

> Many instances of backchanneling were assumed to be internally motived; i.e. the listener backchanneled when he or she felt like it. Microanalysis of *speaker* head movements in relation

---

[2] Goodwin nicely investigates gaze of hearers. "It will be argued that one way in which a nonspeaking party can indicate whether he is acting as a hearer is by gazing at the speaker."

to listener head movements reveals that what were heretofore presumed to be spontaneous, internally motivated, listener responses are actually responses to the speaker's nonverbal requests for feedback. These requests are in the form of up-and-down nods, and listeners recognize and respond to such requests in a fraction of a second. [Goodwin, 1981]

On a slightly higher level, [van Eemeren and Grootendorst, 2004] discuss a similar form of interaction between speakers and hearers.

In ordinary conversational situations some speech acts by speakers are specifically calculated to elicit from listeners certain verbal (and possibly also non-verbal) responses in which they indicate understanding and (in particular) acceptance. In our view this applies pre-eminently to the argumentation advanced during a discussion or debate. This means that to a certain extent arguments in debates are designed to achieve precisely defined *verbally externalized illocutionary and perlocutionary effects* that are immediately related to the speech acts performed.

In the context of the AMI project, we, the Human Media Interaction group in Twente, have been collecting and analysing conversational data paying special attention to listener behaviours and how these shape the design of the speaker's utterance. We have to a large degree focussed on argumentative discourse. So far, our work has involved the following activities.

1. Defining an annotation schema for argumentative discourse [Rienks et al., 2005] and annotating a large part of the AMI hub data with the schema.
2. Defining an annotation schema for emotions and other mental states to apply on the same data. Originally our brief was to annotate the emotions in the meetings. However, we felt it more appropriate to include other mental states as well (curiosity, level of attention, comprehension, certainty, confidence, belief and disbelief, etcetera) as we thought they would be more informative for further meeting analyses. It also turned out during the trial annotations that participants during the meetings showed far more signs of these mental states than they displayed emotions.
3. Annotating a few meetings with a detailed dialogue annotation scheme based on [Bunt, 2000] and analysing the verbal and non-verbal behaviours that characterize both speaker and listener actions [Petukhova, 2005].
4. Refining theoretical models for joint actions and backchanneling [Heylen, 2005a], [Heylen, 2005b] and analyzing some nonverbal behaviours in this framework.

With the analysis of the annotated material, we can see detect some interesting ways in which cognitive mental states and emotional states are elicited and displayed in meetings. At this stage we can combine the information from the argumentation level, the discourse level, the mental state level and the behavioural level (the signals) to find out more about the relation between argumentative discourse moves, the emotional and mental reactions of the participants to these moves and how speakers adapt to these reactions, the way the argumentative moves are embedded in the dialogue structure and the behaviours that accompany these functions. This not only provides information about the details of

argumentative discourse processes but also about more general level characteristics of meeting participants, complementing our earlier research on dominance in meetings ([Rienks & Heylen, 2005]).

In the context of Humaine, we are looking at the analysis of backchannel behaviours more in general to arrive at models to implement in the Sensitive Artificial Listener exemplar. Together with Isabella Poggi, Mark Schröder and Hannes Pirker, the Paris group (Catherine Pelachaud and Elisabetta Bevacqua), we are working on specifying the architectural requirements enabling an ECA system to generate backchannel utterances in appropriate places. Future collaboration between Paris and Twente will involve deriving computational models of backchanneling behaviour from the data collected in Twente and implementing and testing versions of Greta engaging in argumentative dialogue.

In this talk we will report on the various strands of research 1 to 4 mentioned above and on the way they combine together.

References

[Bunt, 2000] Bunt, H. Dynamic interpretation and dialogue theory. In *The structure of multimodal dialgoue II*, M. Taylor, F. Neel and D. Bouwhuis (eds). John Benjamins, Amsterdam.

[Goodwin, 1981] Goodwin, C. *Conversational Organization* Interaction between speakers and hearers. Academic Press.

[Heylen, 2005a] Heylen, D. *Challenges ahead. Head movements as social acts during conversations.* In: Virtual Social Agents. AISB Workshop.

[Heylen, 2005b] Heylen, D. A Closer Look at Gaze. AAMAS Workshop on Creating Bonds.

[Petukhova, 2005] Petukohova, V. *Multidimensional interaction of multimodal dialogue acts in meetings.* Msc Thesis. (Tilburg University)

[Reidsma et al. to appear] Reidsma, D., Heylen, D, Ordelman, R. *Annotating Mental States in Meetings. Some Trials*. AMI Report, to appear.

[Rienks & Heylen, 2005] Rienks, R. and D. Heylen. *Automatic dominance detection in meetings using support vector machines*. Proceedings of MLMI 2005 (in press). Springer Verlag.

[Rienks et al, 2005] Rienks, R., D. Heylen and E. van der Weijden. *Argument Diagramming of Meeting Conversations* International Workshop on Multimodal Multiparty Meeting Processing. ICMI 2005, Trento.

[Van Eemeren & Grootendorst, 2004] van Eemeren, F. and R. Grootendorst. *A systematic theory of argumentation: the pragma-dialectical approach*. Cambridge University Press.

Backchannels Through Gaze as Indicators of Persuasive Success

Elisabetta Bevacqua, Maurizio Mancini, Christopher Peters, Catherine Pelachaud
University of Paris 8
e.bevacqua / m.mancini / c.peters / c.pelachaud@iut.univ-paris8.fr

Introduction

A fundamental behaviour and essential first step in establishing interest, agreement, trust and beliefs is eye contact, which has also been referred to as 'touching somebody with your eyes'. Its importance is outlined somewhat by everyday phrases such as "I don't trust him; He won't even look me in the eye". It is an important ability for a persuasive ECA establish backchannels with the other, to make eye contact *properly*, to be able to interpret the looking behaviours of others. By 'properly', we mean that eye contact should be engaged in the correct amounts to be persuasive (the eyes should not be excessively averted, nor should gaze give the sense of being threatening or dominating).

In this work, we consider reactive backchannels through attention and interest, as signalled through the gaze behaviour of a virtual listener, in order to define a speaker that alters its own gaze behaviours during discourse for the purposes of persuasion. The work presented here has its emphasis on persuasion and is part of a larger body of work, bridging WP6 and WP8, on modelling backchannels and engagement. For this work, we concentrate on one aspect of the model based on signal analysis – semantic and cognitive aspects are left for future work.

Metrics

We use a number of metrics that aid with the interpretation and generation of ECA gaze behaviours based on virtual visual sensing of the others looking behaviour. *Attention Level* is the amount of visual attention one is perceived to be paying at an instant of time, based on their eye, head and body directions. *Interest Level* is the amount of attention perceived to be paying over a period of time and is a measurement of the others interest in what the speaker has to say. *Effectiveness* indicates how well communication is progressing – that is, how the listener is perceived to be behaving (e.g. attending, looking elsewhere, not interested) when the speaker is saying something important.

Summary of Process

The process that uses these metrics is summarised as follows: The listener shows its interest in the conversation through looking behaviours, nodding and paraverbals. The speaker interprets the interest of the listener in different ways and obtains information on how effective it is in persuading the other. This drives changes in the speaker's behaviour, which has to fulfil the goals of emphasising the most important parts of the discourse, ensuring there is the proper amount of eye contact to be persuasive and

maintaining the listeners interest in the conversation. Our efforts in this work require three main ECA capabilities:

1. Listener backchannel

During a conversation, two streams of information are exchanged at the same time between participants (the speaker and listener). The first one is concerned with the topic of the conversation, while the second provides information on the successfulness of the communication. Through linguistic and gestural signals, called backchannel feedbacks, the listener can show his level of engagement in the conversation.

We will base our work on the taxonomy suggested by Isabella Poggi [POGGI05]: listeners produce backchannels feedback to signal interest, understanding, believability, agreement, and attention towards what the speaker is saying. A backchannel can be positive or negative depending on whether or not the listener communicates that he is attentive, understanding, believing, interested or agreeing.

In particular, we focus on interest and attention on a signal level and not on a cognitive level, meaning that the listener's feedback is triggered by the speaker's behaviour and not by what the speaker is saying. Listener's behaviour produced to show feedback can be categorized as follows:

- gaze,
- paraverbals, short sounds or words like "mh", "yes", "m-mh",
- head movements, like nods, jerks, shakes, tilts,
- facial expressions,
- other gestures such as hand and body movements.

Our research focuses primarily on looking behaviour. By analysing the listener's gaze behaviour, we compute his level of attention and interest (see next Section). As pointed out in [SIDNER04] the more that people share looking behaviour, the more they are involved in and coordinate conversation. So if the listener starts to frequently avert his eyes from the speaker, it means that he is probably losing interest and that he is not paying attention anymore.

2. Speaker interpreting listener behaviour

The speaker perceives and interprets the behaviour of the listener in order to calculate the effectiveness and progression of its persuasion attempts. Eye, head and body directions of the listener are perceived by the speaker and are weighted to form *attention levels*. An attention level is a measurement of how much attention one is perceived to be paying at a particular instant. Attention levels over time form an *attention profile* and are integrated to produce *interest level* metrics. These describe how much attention the other has been paying over a time period and are vital in establishing how successfully ones attempt at persuasion is progressing. When compared with the importance of the speaker's utterances, the speaker determines the *effectiveness* of its communication, which is used to alter or generate the speaker's behaviour towards the listener. Effectiveness therefore frames the actions of the listener in the context of what the speaker is saying.

3. Speaker generating / altering its behaviour

In our model, the speaker alters his behaviour in order to try to be more persuasive and obtain the listener's attention. The kind of behaviour variation we can achieve during the speaker generation is based on the speaker's gaze. Our model of gaze has a certain number of timing constraints, for example, the maximum amount of time the speaker could look at the listener without interruption or the maximum amount of time the speaker and listener can engage in mutually attention. These constraints arise from the analysis of gaze behaviour of two subjects during conversation, as explained in [PELACHAUD03]. We have developed a model that considers effectiveness of conversation as another constraint for the system. During behaviour generation, the level of effectiveness is constantly calculated and compared to a desired level. If it is too low, then the speaker increases his tendency to look at the listener and vice-versa. After that, temporal constraints are applied and the final behaviour is decided by a probabilistic model [PELACHAUD03].

Process

The overall process proceeds as follows. The speaker creates a communication plan detailing looking behaviour towards the listener during its utterances. This is distributed over the speech act and particularly at important times in speech. The listener acts in predefined way, either paying attention and showing interest towards what the speaker has to say, or not. The speaker formulates a theory about how much attention the listener is paying and how effective its communication is based on whether the listener maintains eye contact when the speaker is looking at it. If the speaker perceives communication not to be effective, it alters its eye contact durations and allocation during its speech while trying to constrain its looking behaviour so as not to look too much. If effectiveness continues to drop, the communication attempt is abandoned.

Future Work

At the moment, there is no cognitive analysis of the speaker's speech that allows the listener to understand the other agent's goal and belief. In our future work we aim to implement an interpretation module, through which the listener will be able to understand the speaker's goal and belief and compare them with his own. If what the other participant is saying fits the listener's goal and belief, then he will agree with the speaker and provide the proper feedback. Therefore the backchannel generated by the listener will be trigged not only by the behaviour of the other participant but also by his real intentions.

In the future, the speaker will create a plan for the next part of the discourse to be said (for example the next sentence). The plan will be a curve representing the importance of conversation through time which will be coded into APML as tags corresponding to utterances. Depending on the plan, different kinds of behaviours could be activated, for example raising the voice level, adding verbal signals like "are you listening to me?", or looking longer at the listener. The actual level of effectiveness of conversation will be considered and the plan will be modified accordingly. The speaker's animation will be then generated by instantiating behaviours depending on the plan.

References

SIDNER04 C. L. Sidner, C. D. Kidd, C. Lee, and N. Lesh. Where to look: A study of human-robot interaction. In *Intelligent User Interfaces Conference*, pages 78–84. ACM Press, 2004.

POGGI05 I. Poggi, Backchannel: from Humans to Embodied Agents, AISB, University of Hertfordshire, Hatfield, England, 12-15 April 2005.

PELACHAUD03 C. Pelachaud and M. Bilvi. Modelling gaze behavior for conversational agents. In proceedings of the IVA 2003 conference. Springer LINAI Series, 2

More than just a pretty wording – Politeness in ECAs

Matthias Rehm
Institute of Computer Science
University of Augsburg, Germany
rehm@informatik.uni-augsburg.de

Elisabeth André
Institute of Computer Science
University of Augsburg, Germany
andre@informatik.uni-augsburg.de

Brown and Levinson (1987) established a theory of verbal politeness that can be interpreted as traffic rules of social interaction. Focusing on verbal strategies of politeness, many contextual features like emotional influence or non-verbal behaviour were neglected. In a face-to-face setting, those features play important roles. If a criticism e.g. is wrapped up in a joke the critic should exhibit this in her non-verbal behaviours as well otherwise the criticism might get even more severe by being taken literally.

Endowing a technical artefact like an embodied conversational agent with the ability for emotion-based polite interaction behaviours, we essentially have to consider three different steps on our way:

1. get information about non-verbal means of politeness
2. model an emotion-based strategy selection
3. generate verbal and non-verbal politeness behaviors

The Augsburg SEMMEL corpus of non-verbal polite interactions

Due to the sparse literature on non-verbal means of politeness, the Augsburg SEMMEL (Strategy Extraction for MultiModal Eca controL) corpus of non-verbal polite interactions was collected. This multimodal corpus consists of staged interactions of inherently face threatening situations. We devised a scenario that forced the participants to use their (unconscious) knowledge of politeness strategies by confronting them with an inherently face threatening situation. Criticizing someone is a prototypical example of such a situation. Therefore, we chose seminar talks with subsequent discussion to provide for a more or less "natural" situation for the participants. The focus was on the criticism given by the audience to the speakers on their performance. Students were divided into two groups: audience and speakers. The speakers were asked to give a five minute talk about one of their hobbies.

The results indicate that gestures are indeed used to strengthen the effect of verbal acts of politeness (Rehm and André, 2005). In particular, vagueness as a means of politeness is not only reflected by verbal utterances, but also by gestures. Iconic and deictic gestures were overwhelmingly used in more direct criticism while there was a high frequency of metaphoric gestures in off record strategies. Obviously, our subjects did not attempt at compensating for the vagueness of their speech by using more concrete gestures.

Emotion-based selection of politeness strategies

Starting from ideas of Walker et al. (1997) we introduce a model for augmenting Brown and Levinson's strategy-selection mechanism by information about the emotional context of the interaction. Walker et

al. consider the speaker's emotional state to parameterize the acoustic realization of the speech act. However, the emotional state is not calculated due to situational or pre-defined personality factors and it has no effect on the choice of strategies at all. Our work starts from the assumption that the perceived threat resulting from a speech act heavily depends on the user's emotional state. Consequently, the emotional state is a factor that emerges during the interaction and dynamically influences the ongoing dialogue. We represent emotions using a dimensional model which characterizes emotions by the two orthogonal dimensions valence and arousal. Brown and Levinson mention situational factors that may influence the variables employed to calculate the impact of a given threat but leave aside the question of how this influence is reflected. We treat the emotional state of the user as such a situational factor integrating this information in the weight calculation (André et al., 2004).

Generating verbal and non-verbal behaviour

Having selected an appropriate strategy for the current context, the agent's verbal and non-verbal behaviour has to be generated. Brown and Levinson's work offers a rich source of information on the verbal realization of politeness strategies. For a number of exemplary strategies, we have implemented a mechanism which takes a neutral utterance and transfers it into a polite form. The generated utterance has then to be enriched by gestural behaviour relying on the insights gained from the above mentioned user study. The result of this process is an APML-structured document which can be used to create the expressive behaviour of e.g. the Greta agent.

The evaluation of the methods is pending at the moment and after an overview of the above mentioned points, we would like to discuss possible persuasive scenarios for this evaluation. The scenario will include two versions of the agent one exhibiting politeness behaviour and one without such behaviour. Although Brown and Levinson describe the verbal strategies as universal, their realisation is nevertheless culture dependent. This is especially interesting for gestural behaviour. To capture a possible culture-dependency, the evaluation will be replicated by at University of Paris VIII.

References

André, E., Rehm, E., Minker, W.. and Bühler, D. *"Endowing Spoken Language Dialogue System with Emotional Intelligence"*, Affective Dialogue Systems , Elisabeth André, Laila Dybkjaer, Wolfgang Minker and Paul Heisterkamp, 178-187, Springer, 2004

Brown, P., and Levinson, S. C. Politeness – Some universals in language use. Cambridge University Press, Cambridge, 1987.

Rehm, M., and André, E. *"Informing the Design of Embodied Conversational Agents by Analyzing Multimodal Politeness Behaviors in Human-Human Communication."* In *AISB Symposium for Conversational Informatics*, 2005.

Walker, M. A., Cahn, J. E., Whittaker, S. J. *"Improvising Linguistic Style: Social and Affective Bases for Agent Personality."* Proceedings of the First International Conference on Autonomous Agents

(Agents'97), 1997.

Proposing new ways of resolving online conflicts:
an intelligent facilitation of forgiveness in CMC

Asimina Vasalou, Jeremy Pitt
Imperial College London
Electrical and Electronic Engineering Department
Intelligent Systems and Networks Group
Exhibition Road
London SW7 2BT
+44 (0)20 7594 6187
{a.vasalou, j.pitt}@imperial.ac.uk

Guillaume Piolle
Imperial College London
Department of Computing
180 Queen's Gate
London SW7 2AZ
+44 (0)20 7594 8298
guillaume.piolle@imperial.ac.uk

Abstract

Trust and reputation mechanisms have in part replaced online policing by allowing members to document the quality of their interaction with their peers. This paper proposes an integration of forgiveness into those systems by offering both a theoretical and operational model of forgiveness. Motivating this proposal is the belief that forgiveness can work as a possible prosocial mechanism, which in the short-term may help resolve a given conflict and in the long-term may add to an increasingly prosocial and homeostatic environment.

Introduction

In human societies, when violating a norm the offender is usually 'punished' both emotionally (e.g. experiencing embarrassment) and practically (e.g. by prosecution). The threat of those two implications persistently evoked by physical markers, (e.g. people watching, the presence of law enforcement officials) works *preventively* so that a sense of general social order is maintained within the community. Online societies differ from physical societies, both in how the emotional and practical implications are perceived. To begin with, anonymity and the absence of a physical self weaken the impact of the emotional consequences that an offender experiences as a result of his/her offence (e.g. shame or embarrassment). To add to this, the presence of an active policing force is not visible until the member's behaviour has reached what is considered to be by-law illegal. Therefore, one of the problems identified through these two points is the need for an intermediate mechanism that will signal the offender early on and that will also inform the community about milder offences where punitive legal action against the offender cannot be taken.

Trust and reputation mechanisms have been widely adopted in addressing this issue. These mechanisms have empowered members of online communities by allowing them to appraise and capture the granularity of their fellow members' actions (e.g. through ratings). However, in doing so, the designers of those systems have placed emphasis and value on the quantitative appraisal that usually follows an offence while neglecting to account for the qualitative appraisal that often makes repair between two members possible [3]. In human-human interactions, a violation of norms is unavoidable but not unforgivable.

There are strong incentives for considering forgiveness as a possible reparative mechanism in online environments. For example, the act of issuing forgiveness alone is known to stimulate the offender into voluntary actions of repair. Moreover, punishing the offender for an action they did not intentionally perform (e.g. bad ratings for accidentally delivering the wrong product) often results in anger and low-compliancy behaviors [1]. In the context of online communities this could possibly motivate the member to withdraw from the community due to the unjust treatment. Even more, one's judgment can be often misguided and construed on false information. In this case, a system supporting irreversible judgments is both unfair and unethical. Finally, although forgiveness does not necessarily mean that trust is automatically regained, it often provides closure, which may alleviate the aggression that has resulted from a disrupted interaction. This is further demonstrated by the physical well-being of those who tend to issue forgiveness more frequently [5].

Given the many benefits of forgiveness, we believe it is important to integrate such a mechanism into current trust and reputation frameworks. This paper gives a general overview of our research plan and progress on this topic. Section 2, gives a summary of our forgiveness model, inspired by theories of psychology. Section 3, describes the architecture and decision maker driving the theory, built on a fuzzy logic inference system. Finally, Section 4 discusses our future work on the subject.

Theoretical model

In this section we present the definition of forgiveness that will be later used as a basis for the operational model (see Section 3). According to this definition, forgiveness is proposed to result from a *number of prosocial motivational changes which reverse one's initial desire to adopt negative strategies towards the offender* (i.e. revenge, avoidance). In sum, as depicted in fig. 1, forgiveness replaces malevolent motivations towards the offender with constructive and positive behaviors which work to reverse the initial censure [2].
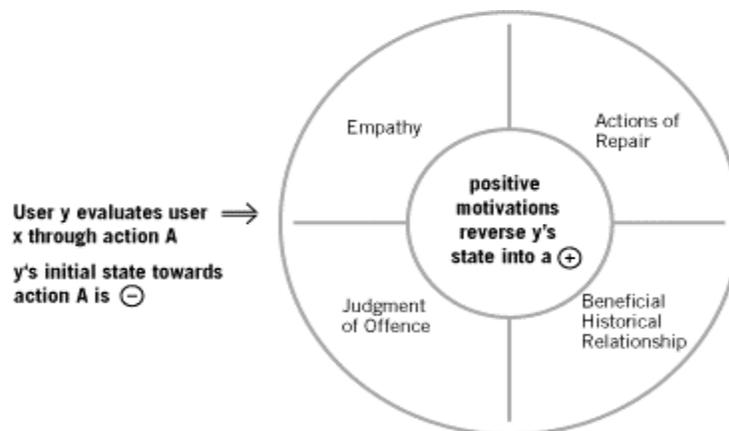

Fig. 1: Theoretical Model

We consider four main motivations that add up to form forgiveness. Those are the judgment of offence, empathy, actions of repair and a beneficial historical relationship between offender-victim. Each motivation has its own constituent motivators that increase or decrease its impact on the overall forgiveness decision. Following, is a more detailed analysis.

1) **Judgment of Offence:** In judging an offence, its *severity* is at first assessed. More severe offences elicit harsher judgments. Next, a historical trail of one's past behaviors is compared against the current offence. Together, *frequency* and severity of past acts impact one's inclination to forgive. To add to that, apparent *intent* leads towards more negative attributions, while low intent actions lead to more positive attributions.

2) **Actions of Repair:** *Apology* and restitution following an offence is known to facilitate and even predict forgiveness. Similarly, it is also possible to *reverse* one's immoral act by performing a good deed that outweighs the impact of the offence.

3) **Historical Relationship:** Prior familiarity and a *relationship of commitment* with the transgressor increase the likelihood of forgiveness. Good friends or successful business partners rely on a richer and mutually-rewarding history that fosters a propensity towards forgiveness.

4) **Empathy:** Empathy, one's emotional response towards another's affect is regarded as a mediator appeasing the victim and facilitating forgiveness. Empathy is encouraged by a number of motivators such as the offender's *visible regret through signs of embarrassment*, a *prior familiarity* with the offender, *similarity* with the offender and the victim's innate *propensity to embarrassment*.

Next, Section 3 gives a general overview of the architecture and decision maker supporting the theory given above.

Implementation

The implementation of the forgiveness mechanism includes two components, written in java. The first one is the input conversion layer of the system, in which a member's successful or non-successful collaborations are each stored as separate java objects. The second component is a decision maker that receives numerical values from the input conversion layer as its input in order to make a forgiveness decision. The decision maker is implemented using fuzzy logic inference systems based on the Takagi-Sugeno model [3]. We go on to discuss the functionality and integration of these two core components in more detail.

• **The Input Conversion Layer:** Following a collaboration with another member, a user reports on his/her experience. If the experience was positive, then the user's report is stored in a *Collaboration Report* object. The *Collaboration Report* object encloses the identity of the user, a timestamp, and a measure of the utility (benefit) of the collaboration as reported. In contrast, if the collaboration experience was negative, the user's report is stored into an *Offence Appraisal* object. The *Offence Appraisal* object contains the kind of offence, the identity of the offender, a timestamp and the motivations that characterize the offence which are the intent and severity of the offence, apology and reparative actions offered by the offender and his/her visible acknowledgement. The embarrassment propensity motivation is a constant value that is stored in the input conversion layer. The motivations that rely on historical data such as offence frequency, historical relationship, similarity and prior familiarity between the offender-victim are computed separately in the input conversion layer to then

be later passed as signals to the decision maker. Upon completing the interaction, both the *Collaboration Report* and the *Offence Appraisal* objects are stored so that each user builds up a history over time.

• **The Decision Maker:** The forgiveness decision maker goes into effect only when the *Offence Appraisal* object is instantiated i.e. a user has violated a norm. Next, as depicted in Figure 2, the input conversion layer extracts the information inherent to the current offence from the *Offence Appraisal* object. This extracted information, along with the constant and computed historical values (e.g. embarrassment propensity, frequency), is transformed into the corresponding signals $j_0 - j_9$ that are then used as input for the decision maker. Finally, the decision maker computes an output value "d" on the basis of these signals using a set of fuzzy rules built on the grounds of the theory described in Section 2. This final value constitutes the forgiveness recommendation (forgiveness = true if d>0.5).
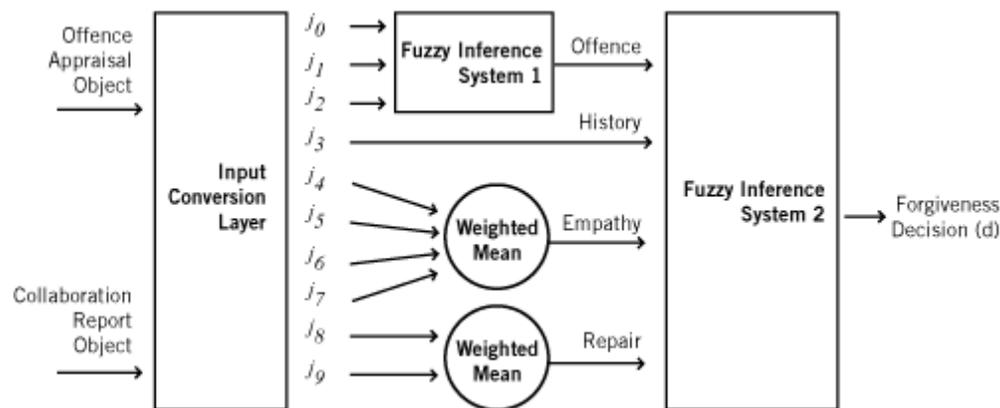

Fig. 2: Forgiveness Implementation

Future work

Our future work on this subject mainly involves three topics. First, we are developing user-centred tools for the collection of the qualitative motivation values (e.g. intent) along with a set of formulas to compute the quantitative motivations (e.g. frequency). Second, we are designing the forgiveness decision, as it will be presented to the user. Finally, we intend to validate our work through a series of experiments that will test the impact of our proposal by capturing the user's decision to forgive while using our tool during an online conflict.

References

[1] Kelln, B.R.C. and Ellard, J.H. An equity theory analysis of the impact of forgiveness and retribution on transgressor compliance. *Personality and Social Psychology Bulletin*, 25, (1999) 864-872.

[2] McCullough, M. Forgiveness who does it and how do they do it. (2001) Retrieved on 10.10.2005, http://www.psychologicalscience.org/members/journal_issues/cd/cdir1061.pdf.

[3] Sugeno, M. *Industrial Application of Fuzzy Control*, Elsevier Science Pub. Co, 1985.

[4] Vasalou and Pitt. Reinventing Forgiveness: A Formal Investigation of Moral Facilitation. *Proceedings of Third iTrust International Conference (iTrust '05)* (Rocquencourt, France, May 23-26, 2005) Springer-Link, 2005, 146-160.

[5] Worthington, E.L., Jr. and Scherer, M. Forgiveness is an emotion-focused coping strategy that can reduce health risks and promote health resilience: Theory, review, and hypotheses. *Psychology and Health*, 19, (2004), 385-405.

POSTERS

Affective Dialogues with an ECA:
from information-oriented to persuasion-oriented

Fiorella de Rosis and Giuseppe Clarizio   *Ian O'Neill and Philip Hanna*
University of Bari, Italy   Queen's University of Belfast, Ireland

In this poster, we plan to describe our ongoing work on persuasion dialogues.

The dialog simulator includes a '*user modeling*' component, in which the aspects of the users which may be relevant to select an appropriate persuasion strategy and to react to critical questions will be represented: their background, their scale of values, their beliefs in the domain to which persuasion applies and their social attitude towards the agent. Effects produced by the user moves on the agent's mind are represented in an *agent model.*

The dialog simulator will be based on the '*information state*' approach and will employ a knowledge base including a repertory of persuasion strategies that we built from a corpus collected in the first stage of this research.

The recently initiated cooperation between our Universities (in the scope of the HUMAINE Interchange Program) is aimed at the following goals:

1. to integrate an embodied agent into the Queen's Communicator at Belfast (a speech-based dialog), so as to express affective components of the agent in its face and speech;

2. to study how the 'error repair' function of the Queen's Communicator may be employed to respond intelligently to some simple user objections about information provided in the persuasion attempt;

3. to begin to investigate how a dialogue simulator based on the information state approach and representing the agent's goals and plans may be designed so as to include some persuasion component.

We plan to describe these ideas andfirst results in a poster, which will be integrated with a video showingthe first results produced in the scope of the Interchange Program.

Affective and Persuasive Guide

Mei Yii Lim, Ruth Aylett, and Christian Martyn Jones
School of Mathematical and Computer Sciences,
Heriot Watt University,
Edinburgh, EH14 4AS, Scotland
{myl, ruth, cmj}@macs.hw.ac.uk

Abstract

The Affective Guide System is a context-aware system, including an 'intelligent emotional guide with attitude', offering the user with a multi-modal interaction interface. This research moves away from the concept of a guide that has it reciting facts about places or events to that of an 'intelligent emotional guide with attitude'.

Introduction

Recently, many research projects have explored the new possibilities of locationaware systems to provide guidance to users in their everyday activities. Cyberguide [1] and MARS [5] are examples of existing mobile context-aware applications. However, none of these systems employ a life-like animated character. C-MAP [9] and Geist [2], on the other hand, attempts to use life-like characters as personal guides in virtual tours. [6] proposed storytelling from a virtual guide perspective. All these systems share a common goal, that is to provide user with context-aware information. However, an Affective Model and Affective Interaction are missing! Therefore, Affective Guide System (AGS) will be a tourist guidance system, with a new feature - the 'intelligent emotional guide with attitude' !

The System

The AGS is proposed to address the frustration that usually occurs in interaction with an emotionless computerised system. It consists of two emotional virtual agents each possessing a contrasting personality, presenting users with different versions of stories of the same event or place. By telling story based on its past experiences and its own ideology, the guide tries to invoke empathy in the user, persuading them to think in the way it thinks. This will allow them to enquire, reflect, evaluate and use the sources of information critically to arrive at a conclusion, a learning strategy targeted by the UK National Curriculum for History [8].

A multi-sensory system will be integrated with the PDA, using wireless communication. The virtual tour guide will first extract some information from the user and plan a route based on this information. The tour guide will draw the user's attention to landmarks, describing them in accordance with the user's interest, applying a story telling technique which links the memory and interests of the guide as well as the visitor to the spatial location. The user will interact through the grahical user interface and

receive a response by means of text, graphics or audio. The system will principally consists of a server that performs the processing and provides location-related information, the guide memories - both emotional and semantic, and the user profiles.

Emotion Models

The main emphasis of this research is the development of an emotion model that expresses personality. The 'Psi' model [4] serves as the basis for its design. Functionally, the agent will perceive the environment continuously and generate intentions based on the acquired information and user needs. These intentions together with some built in motivators - level of competence and level of uncertainty are stored in a memory of intentions. Next, the agent selects an intention considering the current situation and adapting its behavior to its internal states and the environmental circumstances. Each execution of intention will produce a feedback into the system and recovery will be performed when necessary.
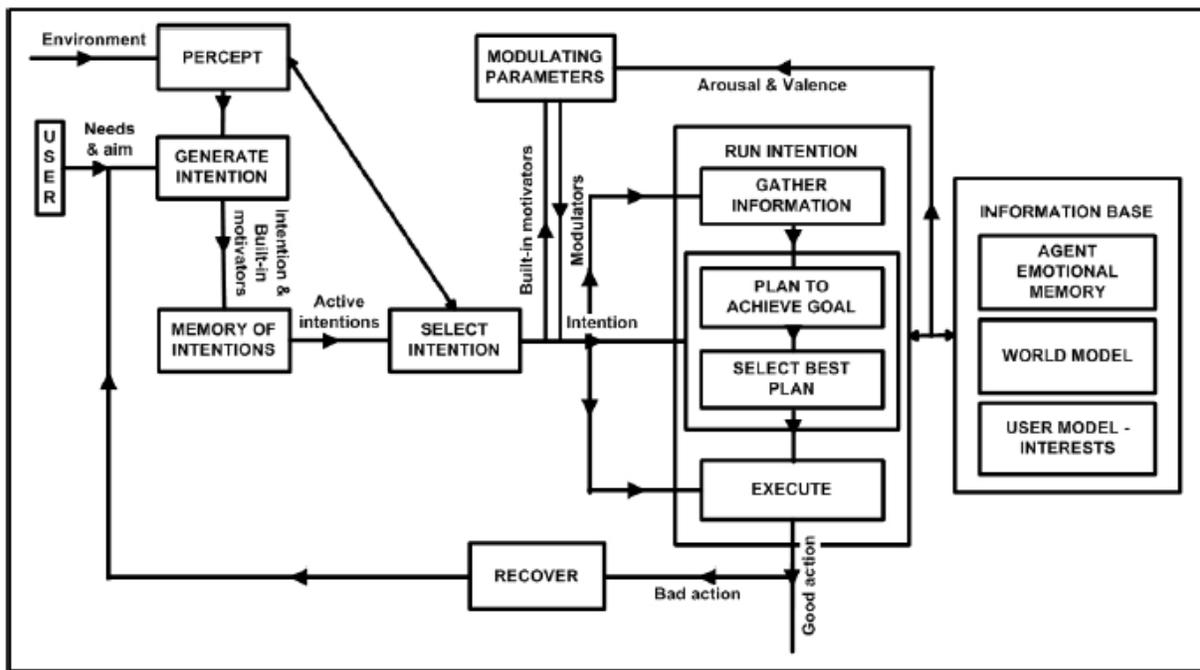

Fig. 1. The Emergent Affective and Personality Model

The idea of emerging emotions resulting from modulation of behavior is expected to give more colors and variation to the agent's expressions. The guide's emotional memory is a simulation of its past experiences and holds the 'valence' and 'arousal' tags, the two elements that usually exist in emotional events. When interacting with the user, the guide will be engaged in meaningful reconstruction of its own past [3], at the same time presenting facts about the site of attraction. This recollective experience

is related to the evocation of previously experienced emotions through the activation of the emotion tags. It results in re-experiencing of emotions, though there might be a slight variation due to the input from the user. User's responses form positive and negative stimuli to the agent that gives hints about his/her interest. In terms of storytelling, this will mean that when the user adopts the guide's perspective by agreeing, the agent's level of competence increases which leads to a more detailed explanation about the subject or related subjects. Similarly, when the user disagrees, the agent's level of competences decreases and the agent tends to focus on only the important points of the subject without much elaboration. That is, the emotion elicited determines the information retrieved and how it is retrieved. Besides emerging emotions, personality also emerges and the resulting personality affects the way the guide behave and story presentation. For more information, please refer to [7].

References

[1] G. D. Abowd, C. G. Atkeson, H. Hong, S. Long, R. Kooper, and M. Pinkerton. Cyberguide: A mobile context-aware tour guide. *Wireless Networks*, 3(5):421–433, 1997.

[2] N. Braun. Storytelling & conversation to improve the fun factor in software applications. In Mark A. Blythe, Andrew F. Monk, Kees Overbeeke, and Peter C. Wright, editors, *Funology, From Usability to Enjoyment*, Dordrecht, April 2003. Kluwer Academic Publishers.

[3] K. Dautenhahn. The art of designing socially intelligent agents – science, fiction and the human in the loop, July 07 1999.

[4] D. D¨orner and K. Hille. Articial souls: Motivated emotional robots. In *Proceedings of the International Conference on Systems, Man and Cybernetics*, pages 3828–3832, 1995.

[5] T. H¨ollerer, S. Feiner, T. Terauchi, G. Rashid, and D. Hallaway. Exploring MARS: developing indoor and outdoor user interfaces to a mobile augmented reality system. *Computers and Graphics*, 23(6):779–785, December 1999.

[6] J. Ibanez, R. Aylett, and R. Ruiz-Rodarte. Storytelling in virtual environments from a virtual guide perspective. *Virtual Reality*, 2003.

[7] Mei Yii Lim, Ruth Aylett, and Christian Martyn Jones. Emergent affective and personality model. *The 5th International Working Conference on Intelligent Virtual Agents (IVA05), LNAI* 3661:371–380, 2005.

[8] NHC. History: The level description. 2004. http://www.ncaction.org.uk/subjects/history/levels.htm.

[9] Y. Sumi, T. Etani, S. Fels, N. Simone, K. Kobayashi, and K. Mase. C-map: Building a context-aware mobile assistant for exhibition tours. *The First Kyoto Meeting on Social Interaction and Communityware*, June 1998.

An emotional approach to designing self-awareness mechanisms for CMC

Asimina Vasalou
Imperial College London
Electrical and Electronic Engineering Department
Intelligent Systems and Networks Group
Exhibition Road
London SW7 2BT
+44 (0)20 7594 6187
a.vasalou@imperial.ac.uk

Jeremy Pitt
Imperial College London
Electrical and Electronic Engineering Department
Intelligent Systems and Networks Group
Exhibition Road
London SW7 2BT
+44 (0)20 7594 6187
j.pitt@imperial.ac.uk

Abstract

The distance and anonymity that is inherent in computer-mediated communication often cultivates non-compliance and uninhibited behaviors as it shields an offender from the emotional consequences that follow a given offence. In this paper, we consider theories of shame and embarrassment for the design of two distinct self-awareness mechanisms with the underlying objective of increasing behaviour that complies with a set of norms between online members.

Introduction

Human societies offer a set of norms that its members are expected to adhere to. When a member of the community violates a norm, there are practical consequences (e.g. banishment from the community) that follow, as well as emotional consequences (e.g. shame and embarrassment). The experience of the emotional consequences that follow norm-violating acts i.e. shame and embarrassment, has been repeatedly shown to be more acute when self-awareness is increased while in the absence of self-awareness, the same emotions are experienced more weakly (see [6] for a detailed overview). As a result of the self-awareness evoked by our everyday interactions, in face-to-face communication people avoid the experience of shameful and embarrassing acts by putting effort into the image they project and maintain towards others.

In computer-mediated communication however, the salience of awareness that is otherwise created by physical and social cues (e.g. facial and corporal expressions, social distance between participants), is reduced. Therefore, online, one would expect the emotional consequences that usually follow norm violations to be experienced less acutely, offering fewer incentives for compliance. This phenomenon is indeed seen in action through the observation of online human behaviour: One of the less desirable effects of computer-mediated communication has been the emergence of uninhibited and uncontrolled behaviour, in many cases shockingly unlike what is practised face-to-face (e.g.[1]).

In addressing this problem, it is tempting to propose the use of stable identities, where accountability and awareness of one's actions is created by linking the action-violation to the perpetrator. However, research has shown that the inherent anonymity and physical detachment online, offers many benefits (e.g. [5]). Consequently, there is a need to find alternative solutions that enhance accountability but at the same time do not compromise the benefits of online anonymity.

In this short paper, we describe solutions for increasing self-awareness by utilizing theories of shame and embarrassment. These theories suggest clear ways of bringing attention onto the perpetrator and his/her actions without compromising anonymity so that the emotional consequences that follow norm violations are made possible in computer-mediated communication. In Section two, we outline the design and function of two distinct self-awareness mechanisms and then relate them to theory. Section three briefly describes future plans for experimentally evaluating these two mechanisms.

Self-awareness mechanisms

Avatar self-awareness mechanism

Following an action-violation, human actors express the emotions of shame and embarrassment with the display of facial and corporal signals such as blushing, avoiding eye gaze contact, turning one's head away and a drooping body posture [3]. In the first self-awareness mechanism, we built these emotional signals into an avatar. During the user's first interaction with the system, the user has the choice of selecting a male or female avatar. The avatar is 2-D and black/white to avoid any cultural identification as the objective is that the user identify as much as possible with their avatar representation. While the user is participating in a given interaction with another member, the avatar is in an idle state. Following an action-violation performed by the user against another member of the community, the avatar displays the embarrassed-shameful expression.
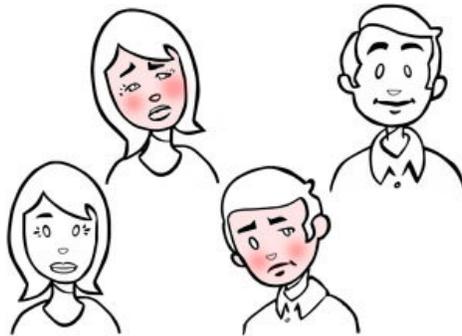

Fig. 1: Avatar self-awareness mechanism

The *avatar based self-awareness mechanism* is linked back to theory in two ways:

- The presence of an observer, who witnesses the full-fledged emotion expression such as the blush for example, adds even further to the feeling of exposure that already follows the experience of both emotions i.e. the observer, now knows that the offender has experienced a negative emotion as a result of the offender's inappropriate behaviour. This knowledge is the cause for an even more painful and vivid emotional experience [4]. We therefore expect that the user will have an emotional experience due to his/her apparent exposure within the community as the avatar's expression is visible for all to see.

- Embarrassment can be induced intentionally by 'accusing' the offender of looking embarrassed [1]. Consequently, it is expected that a user who has seamlessly identified with his/her avatar, will interpret the avatar's emotion display as an 'accusation' of what his/her own façade looks like, therefore resulting in a form of intentional embarrassment.

Text-based self-awareness mechanism

The second self-awareness mechanism we have designed is text-based. Following an action-violation performed by the user against another member of the community, the interface displays two concurrent text messages onscreen. The first message informs the user of the decision they just made and the consequences it carries for his fellow member. The second message notifies the user that his community, including the member he has offended, has been informed of his/her decision to 'defect'.
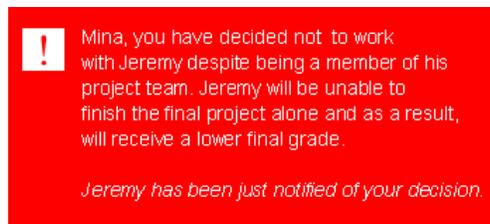

Fig. 2: Text-based self-awareness mechanism

The *text-based self-awareness mechanism* is linked back to theory in two ways:

- It has been proven difficult for offenders to appraise their action-violation if at the time of the offence attention is not closely focused on the action itself. A good example illustrating this phenomenon occurs during multi-tasking where attention is divided among many tasks so that offenders are often unable to experience the emotional consequences resulting from their offensive behaviour. Similar to multitasking, the consequences of one member's action on another are rarely visible online, therefore inhibiting this appraisal process. Consequently, the self-awareness mechanism described here is expected to engage the user into an appraisal of the action-violation by focusing attention on his/her recent behaviour, finally creating the basis for an emotional experience.

- Similar to the arguments made in 2.1, an offender who knows that others are aware of his/her action-violation, will feel increasingly exposed. Consequently, we expect the offender to experience an emotional reaction as a result of his/her notification of exposure in the community.

Future work

We are currently planning a controlled experiment that will test the strength of the two self-awareness mechanisms. The experiment will be built on a trust game between two participants, A and B. Participant A will be simulated and will always choose to collaborate with participant B. Participant B

will be given incentives to defect, despite the knowledge that by this decision, participant A is penalized. Following the decision-defection, participant B will be treated in one of the three following conditions: the avatar self-awareness mechanism, the text-based self-awareness mechanism or the control condition. At the end of the experiment, participant B will fill out two questionnaires. The first one will measure the self-awareness experienced following the defection and the other will request the participant to report the degrees of experienced emotion.

We expect to find that participants of the two self-awareness conditions will report higher scores of self-awareness and experienced shame or embarrassment. However, past experience indicates that theory and practice often do no coincide or directly translate into each other in computer-mediated communication. Ultimately, our expectations' validity is to be put to the test through our future experimental work plan.

Acknowledgments

References

[1] Drummond, P.D., Camacho, L., Formentin, N., Heffernan, T.D., Williams, F. and Zekas, T.E. The impact of verbal feedback about blushing on social discomfort and facial blood flow during embarrassing tasks. *Behaviour Research and Therapy*, 41 (2003), 413-425.

[2] Grady, D. Faking pain and suffering on the Internet. *The New York Times*, 1998.

[3] Keltner, D. and Buswell, B. N.: Embarrassment. Its distinct form and appeasement functions. *Psychological Bulletin*, 122, (1997), 250-270.

[4] Miller, R.E. *Embarrassment. Poise and peril in everyday life*. Guilford Press, New York, 1996.

[5] Stritzke, W. G. K., Nguyen, A., and Durkin, K. Shyness and Computer-Mediated Communication. A Self-Presentational Theory Perspective. *Media Psychology*, *6*, 1, (2004), 1-22.

Tracy, J. L., and Robins, R. W. Putting the self into self-conscious emotions. A theoretical model. *Psychological Inquiry*, *15*, (2004), 103-125.