D8a

Report on Basic Cues and Open Research Topics in Communication and Emotions

Oliviero Stock and WP8 members

Date: 11 October 2004
<table>
<thead>
<tr>
<th>IST project contract no.</th>
<th>507422</th>
</tr>
</thead>
<tbody>
<tr>
<td>Project title</td>
<td>HUMAINE Human-Machine Interaction Network on Emotions</td>
</tr>
<tr>
<td>Contractual date of delivery</td>
<td></td>
</tr>
<tr>
<td>Actual date of delivery</td>
<td>October 11, 2004</td>
</tr>
<tr>
<td>Deliverable number</td>
<td>D 8a</td>
</tr>
<tr>
<td>Deliverable title</td>
<td>Report on Basic Cues and Open Research Topics in Communication and Emotions</td>
</tr>
<tr>
<td>Type</td>
<td>Report</td>
</tr>
<tr>
<td>Number of pages</td>
<td>81</td>
</tr>
<tr>
<td>WP contributing to the deliverable</td>
<td>WP 8</td>
</tr>
<tr>
<td>Task responsible</td>
<td>O. Stock</td>
</tr>
<tr>
<td>Author(s)</td>
<td>See page 5</td>
</tr>
<tr>
<td>EC Project Officer</td>
<td>Philippe Gelin</td>
</tr>
</tbody>
</table>

Address of author: stock@itc.it
# Table of Contents

THE STATUS OF THIS REPORT .................................................................5

1. PERSUASION: DEFINITION AND HUMAN SCIENCES APPROACHES ........7

2. PERSUASION: COMPUTATIONAL ISSUES ....................................15

3. COMPUTATIONAL HUMOUR .........................................................23

4. LIES AND EMOTION .................................................................31

5. EMOTIONS AND POLITENESS .......................................................37

6. IMPORTANCE OF THE EYES AND GAZE .....................................42

7. NARRATIVE AND STORY-TELLING ............................................48

8. RESPONSE GENERATION IN DIALOGUE SYSTEMS .......................55

9. DESCRIBING THE FEELINGS THAT MUSIC COMMUNICATES: A DIMENSIONAL APPROACH .................................................................60

10. EXPERIMENTAL ISSUES: EMOTIONAL EXPRESSIONS, HUMOR AND CREDIBILITY IN PERSUASION .................................................................66

11. EMBODIED CONVERSATIONAL AGENTS FOR PERSUASION IN DIFFERENT APPLICATIONS ........................................................................71

12. APPLICATIONS OF EMOTION-ORIENTED SYSTEMS: TAKING A LONG LOOK .........................................................................................76
The status of this report

The HUMAINE Technical Annex identifies a common pattern that is followed by most of the project’s workpackages:

There will be a workshop exploring the thematic area. It will be designed to ensure both that the key issues in it are understood throughout the network, and that the way work proceeds reflects issues affecting other areas. Workshops are held as early as practical limitations allow. There will also be a sustained process leading to the development of an exemplar designed to embody good practice in the area, by showing how a key problem in the area can be solved in a principled way; and also how work focused on that area can integrate with work focused on the other areas.

The practical limitations noted make it impossible for all workpackages to follow the pattern exactly, most immediately because not all workshops can be held before decisions about exemplars need to be taken (at the second plenary session in month 16). Workpackage 8, Emotion in Communication and Persuasion, is a case in point: the workshop is not scheduled until month 24.

This deliverable represents a practical response to the practical problem. In effect, it brings forward one of the key functions of a workshop. Key participants in WP 8 present position papers on key issues as they would in a workshop. That provides a way to promote understanding of the key issues relevant to the workpackage, both among people associated with the workpackage, and between them and people with no direct connection to it. A key part of promoting understanding is acknowledging the diversity of views within the community that HUMAINE aims to integrate, and for that reason, the papers deliberately reflect the contributors’ own positions rather than attempting to define a consensual view.

The overall intention was to cover all the main topics in WP8, and the table of contents shows that this intention has been broadly achieved. Furthermore we have adopted a common structure for most contributions: definition of the topic, state of the art, present challenges. Where appropriate joint subgroups have been formed already in this phase of elaboration of contributions.

The papers show clearly the range of approaches and aspirations that there are in the area, and thereby help to define both the outcomes towards which the workpackage might aim, and the tasks involved in achieving co-ordinated progress towards them.

The following persons have contributed to the report:

Elisabeth André (University of Augsburg), Ruth Aylett (University of Salford), Aaron Ben-Ze’ev (University of Haifa), Valeria Carofilio (Department of Informatics, University of Bari), Roddy Cowie (School of Psychology, The Queen’s University of Belfast), Fiorella de Rosis (Department of Informatics, University of Bari), Marco Guerini (ITC-irst), Philip Hanna (School of Computer Science, The Queen’s University of Belfast), Shlomo Hareli (University of Haifa), Xingkun Liu (School of Computer Science, The Queen’s University of Belfast), Edelle McMahon (School of Psychology, The Queen’s University of Belfast), Benoît Morel (Chairman and CEO Cantoche), Ian O’Neill (School of Computer Science, The Queen’s University of Belfast), Catherine Pelachaud (University of Paris 8), Matthias Rehm (University of Augsburg), Thurid Spieß (University of Augsburg), Oliviero Stock (ITC-irst),
Carlo Strapparava (ITC-irst), Shay Tzafrir (University of Haifa), Alessandro Valitutti (ITC-irst)

The institutions that have contributed are:

Cantoche
ITC-irst
Queen’s University of Belfast
University of Augsburg
University of Bari
University of Haifa
University of Paris 8
University of Salford
1. Persuasion: definition and human sciences approaches

Marco Guerini\(^1\), Oliviero Stock\(^1\), Valeria Carofiglio\(^2\) and Fiorella de Rosis\(^2\),

\(^1\)ITC-irst, Trento
\(^2\)Department of Informatics, University of Bari

Definition of the field of investigation and applied prospects

Persuasion is a pervasive phenomenon in human communication. The semantic amplitude of the term is very extended: it has been defined in many different formal ways, e.g. attitude and behaviour inducement. While there are lots of theories explaining different facets of persuasion, only few have hardly tried to categorize all its aspects in a comprehensive and computationally oriented way.

In this paper, we deal with the different definitions of persuasion and classify them. This is a preparatory phase for expressing challenges for a computational treatment of this concept and foundational for investigating requirements in HCI systems.

Having a unified vocabulary/formalism, it will be possible to compare different theoretical researches. On the constructive side, it will allow clarifying which elements are to be used in diverse persuasive systems (according to their influencing aims and contexts of use). On the evaluation side, it will be possible to highlight the possible ethical issues raised by these systems, in a more systematic way.

State of the art

Models of persuasion

Historically, lots of definitions of persuasion have been given. Most of them have a common core that addresses methodologies aiming at changing, by means of communication, the mental state of the receiver.\(^1\) In these definitions, much attention is paid to the cognitive aspects of persuasive processes, that is, to the importance that the receiver attributes to the source (the persuader) and to the message (the contents communicated), and how they are processed. But, these definitions, despite the common core, are quite different one from another. They can be roughly divided in four groups, according to which elements, involved in the process of persuasion, they focus on. The first three definitions broadly refer to the beliefs of persuader and persuadee, and how they are related. The fourth category still studies the relations among persuader and persuadee but focusing on the goals sets of the two participants.

---

\(^1\) See for example wordnet definition: “the act of persuading (or attempting to persuade); communication intended to induce belief or action”
1. Definitions focusing on the *why* of persuasion. They explain why persuasion can work, making reference to the “heuristics” through which the receiver may be persuaded. See for example [Chaiken, 80; Cialdini, 93].

2. Definitions focusing on the *how* of persuasion. They account for the *way in which persuasion occurs*, that is, for what are the cognitive devices and the inferential routes that operate in the receiver to be persuaded. For example, according to Petty and Cacioppo [Petty & Cacioppo, 86], persuasion processes should be based on the hypothesis that "when a person is motivated and able to process a message carefully, generated cognitions are likely to reflect evaluative inferences or cognitive elaborations about the quality of the message content. ... however... when individuals lack sufficient motivation or ability to process elaboratively, peripheral-route processing is likely to occur...". Here, central and peripheral routes represent the rational vs. affective – contextual factors that lead the receiver to the conclusion proposed by the source.

3. Definitions focusing on the *communicative goal of persuasion*: (behaviour, attitude or action inducement).
   
   a. Behaviour inducement: changing, in a stable and persistent manner, the way an agent acts (for example in response to certain events or state of affairs in the world).

   b. Attitude inducement: changing, in a stable and persistent manner, the way an agent evaluate events, state of affairs, objects.

   c. Action inducement: changing a particular planned action of an agent.

   All these models study the relationship between the beliefs of the receiver and those of the persuader, but not the relationship between their goals.

4. According to Poggi [Poggi, 04], in order to understand the process of persuasion it is necessary to study also the goals of the receiver and their relationship to the persuader's goals. In her work, Poggi proposes an analysis of persuasion in terms of a model based on the notions of goal and belief and shows how logical, emotional and contextual factors can be simultaneously involved in the process of persuasion.

**Definitions of related concepts**

**Argumentation** = intuitively, when talking about the relation between argumentation and persuasion, a dichotomy between these two concepts is put forward. The former is seen as a process that involves “rational elements”, while the second uses a-rational elements like emotions. In our view, instead, argumentation is a resource for persuasion because:

   a. Planning of persuasive messages involves a ‘rational’ activity, even when emotion inducement is employed as a means to increase the persuasion strength. On the other side, the way persuasion is performed (items selected, their order of presentation, their ‘surface’ formulation) also depends on the emotional state of the persuader.

   b. Argumentation is concerned with the goal of making the receiver believe a certain proposition (influence his mental state) and, apart from coercion, the
only way to make someone doing something (persuasion) is to change his beliefs [Castelfranchi, 96].

Persuasion, since it includes a-rational elements as well, is a “superset” of argumentation, but this does not rule out that there is a role for emotion within argumentation [Miceli et al.]: through arousal of emotions (see Rhetorics) or through appeal to expected emotions. In classical argumentation, though, these problems are not addressed since emotional argumentation is often considered as some sort of ‘deceptive’ argumentation [Grasso et al., 00].

A better distinction between argumentation and persuasion can be drawn considering their different focus of attention: while the former is focused on the correctness of the message (its being a valid argument) the latter is more concerned with its effectiveness. The point is that an argument can be valid but not effective, or, on the contrary, can be effective but not valid.

**Natural Argumentation** = Natural argumentation comes closer to persuasion, as it is also concerned, for example, with the problem of the adequacy - effectiveness - of the message [Fiedler & Horacek, 02]. Even in professional settings, such as juridical argumentation, extra-rational elements can play a major role [Lodder, 99]. Recent works have studied applications of natural argumentation [Walton & Reed, 02; Das, 02]: argumentation- based text generation has been proposed by Zukerman [Zukerman et al., 00], relying on a Bayesian approach. Negotiation has also been widely investigated and modelled in a computational framework; see for instance [Kraus et al., 98; Parson & Jennings, 96].

**Coercion** = using force to “persuade” someone to do something he is not willing to do. Obviously coercion falls out of our definition of persuasion.

**Rhetorics** = the study of the ways of using language effectively. This area of studies concerns the linguistic means of persuasion (one of the main means, but not the only one).

**Social Influence** = affecting or changing how someone behaves or thinks (by changing his mental state). Social Influence is a superset of persuasion since the definition of persuasion restricts the field of its coverage by making reference to both the concepts of “aim” and “communication”. The term “aim” indicates that persuasion is an intentional process: there is persuasion only when there is the intention to produce a change in the mental state of the receiver². The term “communication” rules out those effects of (social) influence that, for example, are caused by mere exposition to repeated stimuli.

**Dimensions characterising persuasion**

There are several dimensions of persuasion that can be used for structuring broad areas of study.

1. Audience specific vs. universal
   a. The first definition of persuasion given by Perelman [Perelman & Olbrechts-Tyteca, 69] claims that what characterizes persuasion is its being audience specific, namely its capacity of adapting the topic to the specific listeners.
   b. Instead Cialdini [Cialdini, 93] takes the opposite position: all the strategies he analyses are meant to be universal (since they use cognitive patterns of the receiver which are common to everybody).

---

² This means that, from a computational point of view, persuasion can be seen as a goal directed process.
2. Monological vs. dialogical
   a. Perelman’s analysis of persuasion, since concerned with rhetorics (how to create effective discourses), is more involved with monological interaction.
   b. Cialdini’s analysis of persuasion, since concerned mainly with selling agents, is more involved with dialogical interaction (e.g. “foot in the door” strategy, “door in the face” strategy, etc.).
3. Domain specific vs. universal
   Some strategies are typically domain specific (like “fake discount” strategies to sell more) while other are not domain specific (like “resort to fear” strategy with impressionable persons)
4. Just language vs. multimodality
   This distinction is relevant especially with the coming of the new media, and for HCI is an opening field of research.

**Elements used in persuasion**

In persuasion processes there are different elements playing a role:
1. The cognitive state of the participants (beliefs and goals of both agents)
2. Their social relations (social power, shared goals, etc.)
3. Their emotional state
4. The context in which the interaction takes place.

Here we just sketch how emotions can be used to affect decision making. The proposed framework is similar to the one in [Gmytrasiewicz & Lisetti, 01]. Yet, while Gmytrasiewicz and Lisetti focus on how the emotions of an agent can change his own behaviour, we focus on how emotional elements can be used to increase or diminish the effectiveness of a persuasive message. There are four dimensions to be considered:
1. The current emotional state of persuadee (how it affects strategy selection by persuader)
2. The emotional state expressed by persuader (which emotion persuader has to display in order to maximize the persuasive force of the message)
3. The emotional state possibly produced on persuadee through the message (the induced emotional state could be non desirable and in general it must be taken into consideration for the subsequent interaction)
4. The current emotional state of persuader (how it affects his strategy selection)

Issues 2 and 4 above reflect the difference between what is felt and what is expressed. It is still a matter of debate if to consider issue 4 in modelling persuasive interfaces (while this is the case for the simulation of emotional agents); The two main standpoints are:
   a. A perfect persuasive agent should be emotion neutral; he just has to display the most effective emotion for the current persuasive goal.
   b. For persuader, to feel emotions is a good way to handle unpredicted situations and a resource for coming with responses to persuadee moves.

**Present challenges**

To meet the first goal of our research (to put some order in the persuasion research field) we have to model, from a computational point of view, the concept of persuasion. We converge with the social influence definition adopting Perelman characterization [Perelman & Olbrechts- Tyteca, 69] that persuasion is a skill that human beings use in order to make their
partners perform certain actions or collaborate in various activities, see also [Moulin et al., 02].

For us persuading a (human or artificial) agent implies planning how to modify its predispositions to certain actions, its/his/her complex of beliefs and judgments (see also the concept of “argumentation”). We do not conceive persuasion as characterized by his functioning or by the elements it uses. In fact, according to the work developed by linguists, philosophers and cognitive psychologists persuasion may appeal to both the informational and the emotional route of the recipients [Petty & Cacioppo, 86; Sillince & Minors, 91]. So the distinctions between peripheral/central route and rational/emotional elements, is not relevant for persuasion definition (but fundamental for modelling its functioning).

In defining persuasion we prefer to differentiate a “broad” definition of persuasion (behaviour inducement) from a “narrow” one (action inducement). Another distinction we will give is between the weak notion (capturing the idea that persuadee is not already planning to perform the required action/behaviour) and the strong notion (capturing the idea that persuadee has also some barriers against the required action/behaviour).

<table>
<thead>
<tr>
<th>SOCIAL INFLUENCE</th>
<th>No barriers</th>
<th>Barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action inducement</td>
<td>Weak Persuasion (narrow def.)</td>
<td>Strong Persuasion (narrow def.)</td>
</tr>
<tr>
<td>Behaviour Inducement</td>
<td>Weak Persuasion (broad def.)</td>
<td>Strong Persuasion (broad def.)</td>
</tr>
</tbody>
</table>

We propose two different formalizations of persuasion, or better of “persuasive-goal”, for action inducement (see formulae below). The two formulae, when y is an autonomous agent, are equivalent.

A1) PERSUADE (x y ay) GOAL (x DOES (y ay))

A2) PERSUADE (x y ay) GOAL (x INTEND (y ay))

Following the second definition it is important to note that in persuasion it is presupposed that the receiver is not already planning to perform the required action ay. In a more strict definition, adherent to the intuitive notion of persuasion, it can also be presupposed that the receiver has some barriers (motives) against ay. Persuasion is then concerned with the problem of finding means to overcome these barriers by conveying appropriate beliefs in the mind of y.

The relation between persuasion and dissuasion is non-trivial. Although “not performing an action” is a form of acting (considering dissuasion as persuasion to not perform a given action) the situation is not totally symmetrical. This becomes clear in analyzing the notion of intention that, at least from a psychological point of view, has three cases. The intention of performing ay (see 1), the intention of not performing ay (see 2), and the lack of intention (see 3).

1. INTEND (y ay)
2. INTEND (\(y \neg ay\))

3. \(\neg\)INTEND (\(y ay\))

Given this three cases it is important to note that while (h) is true, this is not the case for the contrary (k):

(h) INTEND (\(y \neg ay\)) \(\neg\)INTEND (\(y ay\))

(k) \(\neg\)INTEND (\(y ay\)) INTEND (\(y \neg ay\))

The two cases of persuasion, given these premises, can be modelled this way:

a) WEAK notion: from the lack of the intention to the intention

\(\neg\)INTEND (\(y ay\)) INTEND (\(y ay\))

b) STRONG notion: from the contrary intention to the intention

INTEND (\(y \neg ay\)) INTEND (\(y ay\))

In the strong case there is more resistance because these barriers require a double passage: first making \(y\) to drop the contrary intention of not performing \(ay\) (so going back to the case of the lack of intention) and then to adopt the intention of performing it. It can be also argued that if there’s a lack of intention it is because the pros and cons of \(ay\) are balanced in \(y\) beliefs set (or not computed) while if there is a contrary intention is because the cons of \(ay\) are stronger (so persuasion is harder to be reached).

The concept of dissuasion can be similarly divided along two cases: the weak case (from the intention to the lack of the intention) and a strong case (from the intention to the contrary intention)

To meet the second goal of our research (develop models for persuasive systems) we first have to individuate, in a systematical way, the typology of the various persuasive strategies coming from the human sciences approaches. The main typologies we propose:

1. **Monological strategies**: strategies that affect the way a message is formed. They are involved in the planning of a single communicative move. We have three typologies:

   a. Single-Strategies: conveying a single content, a claim (like “this action has this positive outcome”) \((inventio\) of old rhetoric)  
   
   b. Abstract-Strategies: modifying the Single-strategies features (like “if the user is tired then express the contents on positive outcomes with an happy mood”) \((elocution\) of old rhetoric)
c. Meta-Strategies: disposing the content of the usable strategies, for a persuasive message, in the most effective way (like “if the user is impressionable then illustrate the negative outcomes of not performing the action first, then show the positive outcomes of performing the action”). (dispositio of old rhetoric)

2. Dialogical strategies:

a. The ones we call “Unfolding-Strategies” are strategies that describe how an interaction should unfold to reach the persuasive end (e.g. the “foot in the door” strategy). These strategies involve the planning of more than one communicative move.

b. Recovering-Strategies that describe (i) how to recover a step of an Unfolding-Strategy if something goes wrong during the interaction, (ii) how to respond to critical questions and so on.

Finally we have to solve many key problems for approaching the construction of a persuasive system: defining new methods for representing knowledge, for reasoning on it and for generating natural language monologs or dialogs.

For examples in strategy selection we have to: (i) define how to compute, at every system move, the various degrees of strength an usable persuasion strategy may have (depending on the goal which motivates the persuasion process, on the plan of the persuader and on the characteristics of the recipient, like, for example, emotional states). (ii) Define how to compute the persuasive force of a complex message composed of several single strategies/claims. (iii) Describe, for dialogs interactions, the decisional process that bring persuader to chose among the several categories of recovery strategies (depending on the kind of objection the recipient makes). (iv) How to generate arguments in natural language and how to simulate undermining of the argument which supports a given conclusion, in a persuasion dialog.

References


2. Persuasion: Computational Issues

Valeria Carofiglio\textsuperscript{1}, Fiorella de Rosis\textsuperscript{1}, MarcoGuerini\textsuperscript{2} and Oliviero Stock\textsuperscript{2}

\textsuperscript{1}Department of Informatics, University of Bari

\textsuperscript{2}ITC-irst, Trento

Definition of the field of investigation and applied prospects

Persuasion is an emerging topic in the field of Human Computer Interaction: persuasion functions will improve the effectiveness of intelligent interfaces, especially in applied fields where interfaces have to induce a particular intention/behaviour on the part of the user. In these scenarios the "argumentative way" is not enough. For intention and behaviour adoption, often, it is not enough to have valid arguments proposed by the artificial agent: what really matters is the overall impact of the message. Within this prospect emotions and social elements are key features that characterize persuasion for effective interactions.

A useful framework for our study is the artificial intelligence research on BDI&E (belief-desire-intention-emotions) agents [Rao \textit{et al.}, 91] and theories of how intentions and commitments are produced in these agents [Jennings \textit{et al.}, 98]. Within this framework various (emotional and non-emotional) persuasion and argumentation strategies can be investigated, pointing on different methodological aspects: (i) how to model persuasion, and persuasion strategies using the tips coming from the human sciences approaches. (ii) How to model the interaction of emotional processes with persuasion ones. (iii) How to treat the uncertainty (typical of real world interactions) in the reasoning process behind persuasion and (iv) how to simulate the various steps of persuasive and argumentative monologs and dialogs, according to the characteristics of the two interlocutors.

In our view, the models used to account for monological and dialogical interactions, and the solution proposed, will be tightly interconnected, sharing large parts of the theory. These frameworks will be built upon two commonly agreed assumptions:

(i) When emotional arguments are reasonable, that is when they are not fallacious, they are often instances of practical reasoning, directed towards a conclusion describing a prudent course of action, in a set of particular circumstances. This means that emotional arguments may be seen as a kind of goal-directed, knowledge-based argumentation, exactly as non-emotional ones.

(ii) Moreover, persuasive emotional arguments rely on the choice of suitable initial premises for convincing the respondent, through the arguer’s ability to establish an empathic relationship with him, that is to make plausible inferences about the respondent’s intentions and commitments.

Assumptions (i) and (ii) give us the opportunity to handle emotional and non-emotional arguments in a similar way.
State of the art

There is an increasing research interest in HCI on providing systems with deep reasoning capabilities on persuasive features. Here we will point out some systems, or researches, that are moving toward this direction and classify them according to the definitions of persuasion pointed out in the paper “Persuasion: definition and human sciences approaches”:

1) SYSTEMS FOCUSING ON THE COMMUNICATIVE GOAL OF PERSUASION: STOP is one of the most well known systems for behaviour inducement that exploited persuasion [Reiter et al., 03a; Reiter et al., 03b]. STOP is (mainly) a natural language generation system, employed for real human settings. Its aim is to induce users to stop smoking. The authors, as pointed out in [Reiter et al., 03a], had a hard time in evaluating it because of the problem of assessing a real change in behaviour and because of the particularly awkward nature of smoking addiction.

2) SYSTEMS FOCUSING ON ARGUMENTATIVE ASPECTS OF PERSUASION: some systems, such as the one proposed in [Zukerman et al., 00], use argumentation strategies in the generation of persuasive messages. Zukerman is concerned with the abstract form of the unfolding of the argument (e.g. reductio ad absurdum, inference to the best explanation, reasoning by cases). In general, though, logical reasoning is just one resource to support persuasion. Existing computational model of persuasion (e.g. [Grasso et al., 00], [Walton et al., 02] and [Zuckerman et al., 99]) are built on the seminal work developed by linguists, philosophers and cognitive psychologists: Toulmin [Toulmin, 58] first of all, but also Perelman and Olbrechts-Tyteca [Perelman et al., 69] to mention only a few. The study of these theories/systems enlightened the limits of applying a purely logical reasoning to this domain and the need, on one side, of considering uncertainty [Zuckerman et al., 00] and on the other side of introducing argumentation schemes more refined than logical modus ponens [Walton, 00].

3) SYSTEMS FOCUSING ON THE “HOW” OF PERSUASION: One of the more recent subjects of interest in this trend of research on natural argumentation concerns widening the persuasion modes from considering 'rational' or 'cognitive' arguments to appealing to values and emotional states ([Sillince et al., 91; Grasso et al., 00; Guerini et al., 03; Fogg, 04]). Carofiglio and de Rosis [Carofiglio & de Rosis, 03] also focus on emotions as a core element for affective message generation. Their model, since their main concern is dialogical argumentation, uses one persuasive strategy per time.

4) SYSTEMS FOCUSING ON MULTIMODAL ASPECTS OF PERSUASION: many systems using ECAs address the multimodal aspects of persuasion, but, in general, more on the perceptual interface side than the planning one (in which persuasion plays a major role).

Finally, it should also be noted that there are commercial tools claiming to have persuasive aims, but they have only hardwired persuasive features. These are the main focus of “Captology”, the term introduced by Fogg [Fogg, 02] with reference to persuasive technologies.

Present challenges

Our long term aim is to develop a framework which will be the basis for building effective and usable tools for simulating persuasion in HCI. Many challenges are to be taken into consideration:
1. Integrated models of emotion manipulation and beliefs and goal induction are necessary. The approach will emphasize BDI&E (belief-desire-intention-emotions) agents and models of how intentions and commitments are produced (and induced) in such agents. These models will be used not only to describe the characteristics of persuadee and how these characteristics are affected by a persuasive message (i.e. how the information about the user attitudes is used to select the ‘best’ argumentation move the system can make, in a given situation). They also will be used to describe the process of persuasion itself (i.e. the planning of a persuasive message). Obviously these models will have different degree of complexity depending on the kind of interaction modelled (static user model for monological interaction may suffice, dynamic user models are instead necessary for dialogical situations).

2. To simulate natural argumentation and (emotional) persuasion we need to define new methods for representing knowledge, for reasoning on it and for generating natural language and multimodal messages (both in monological and dialogical situations). We will try to represent in a unique formalism/framework persuasive strategies coming from different human science researches. Starting from the continuum which characterizes the various (emotional and non-emotional) persuasion modes, we will propose a framework which tries to unify the various items of this continuum (see for example [Guerini et al., 03]). We will then investigate various (emotional and non-emotional) persuasion and argumentation strategies, (like those in Walton and Toulmin [Prakken et al., 03; Toulmin, 58]) and will propose a method to formalise them by representing the various sources of uncertainty and incomplete knowledge they may include. We will finally investigate related aspects, fundamental for dialog interactions, like critical questions, counter-arguments and so on, and how they can possibly influence monological production.

3. To handle the problem of uncertainty, to model the concept of effectiveness of a message and to foster the process of choosing the best strategy to be used at every interaction, we need to furnish models of measurement of the strength of persuasive strategies, and of other related concepts such as: argumentation strength [Sillince & Minors, 91], probative weight [Walton, 00], dialectical relevance [Walton, 99] and impact [Zukerman et al., 99; Zukerman, 01]. Various methods have been proposed for representing uncertainty in this domain: for instance BIAS [Zukerman et al., 99; Zukerman, 01] and [Carofiglio, 04]. Both the cited examples may be employed in a sort of hypothetical reasoning, to plan a combination of data that can produce a desired emotional impact on the claim. This emotional impact of the message may be evaluated by integrating cognitive emotion activation models with persuasion and argumentation models [Carofiglio & de Rosis, 03].

4. Multimodality: The realization of a persuasion message requires the expression in a communication language. In most approaches natural language is the main modality, but there is an increasing research on the role that could be played by music, kinetic typography, ECA’s, and so on. In fact, the natural language generation processes can be usefully combined with other output modalities. For instance a talking head may express the mood of the message originator, a music theme may emphasize a given emotional aspect, or simply, a relevant image can be combined with the produced text. Multimodality poses lots of challenges: for example, the question whether an ECA should be credible or realistic (it can be argued that, with children, cartoon style ECAs are more credible - and persuasively effective - than realistic one).
5. Persuasive systems need to be evaluated. That is, it is not sufficient that they are theoretically sound: they also have to be effective with real users. Evaluation is not straightforward at all: it is necessary to point out carefully all the variables that can affect the effectiveness of the system, and how they can correlate (context of use, scenario of the interaction, typology of the user, required task, persuasive strategies at hand, and so on). Specific evaluation methodologies have to be defined.

6. “Indirect aspects” like attention and memorisation can affect the effectiveness of persuasive messages. E.g. if the attention of the user is low, or there are key concepts persuader wants to stress, then persuadee’s attention has to be focused or enhanced by using various means. Among them we consider of high importance the use of irony or affectively “coloured” terms. Similar considerations can be made about memorisation.

Requirements for a persuasive system

For a persuasive system we have to meet the following requirements: given a communicative goal of the system, (a) revise the user image (b) identify and prepare the system next move and, finally (c) generate a message (natural language utterance, with, possibly, multimodal features). To complete each of the steps above, the system activates an Argument Builder (AB), which performs tasks that are specific to the current phase.

The communicative goal (that represents the intentional aspect of persuasion as a goal directed process) can be triggered, for example, (i) by a communicative move of the user - an argument, a question or a sentence – or (ii) by an external event - like the user acting in the real world –.

A) Revise the user image

If there is a feedback structure, the system has to revise the user image. The way how new evidence is integrated in the user image depends on the method employed for representing knowledge and reasoning. For example, argumentation (and also persuasion) is often modeled in the framework of ‘defeasible reasoning’. According to classical terminology, a rebutting defeater is any reason which directly denies the claim of the argument, while an undercutting defeater is a reason which undermines the validity of the warrant [Toulmin, 58]. Undercutting and rebutting defeaters have similar functions; however, the former attacks the connection between data and claim, while the latter questions the facts which support the conclusion. The user may therefore contradict, question or build an argument i) against a datum included in an argument proposed by the system or ii) against an instance of the scheme. In the first case, premises for the argument are in focus; in the second one, the warrant is considered. The user may propose, as well, a rebuttal for the system’s arguer.

In our modeling of the persuasion process, structural differences between handling emotional and non-emotional arguments do not exist: in both cases, general forms of arguments (schemes) may be applied to specific circumstances. Warrants associated with a given scheme define general conditions needed to derive a conclusion with that scheme, while an instance of the scheme satisfies all the conditions expressed by the warrant. However, in the case of emotional arguments, the affective decision-support system component of the AB has to guess the affective state of the user, to estimate the expected validity of an argument.

B) Identify and prepare the system’s move

Identify and preparing the system’s next move: persuasion must induce attitude change, which entails affective (emotion-based) and rational change. As we said, our long term aim is to define a framework that allows us to represent the core notions of Walton’s argumentation
schemes. In this framework, information about the user attitudes (a user image model) will implement one of the rhetorical aspects of arguments: it will inform about the ‘best’ argumentation move the system can make in a given situation. In this context we will propose measuring methods for various concepts. For instance, argumentation strength, probative weight, dialectical relevance and (both emotional and rational) impact. On the other side, the way persuasion is performed (items selected, their order of presentation, their ‘surface’ formulation) also depends on the emotional state of the persuader (the system). The affective decision-support system component of the AB, deals with this problem.

For dialog systems there are additional requirements: we expect that, given an (emotional or non-emotional) argument formulated by the system, the user may make several kinds of objections (what Walton calls ‘critical questions’) and that, depending on the kind of objection the user makes, several categories of recovery strategies can be proposed, again, by the System:

1) **The user contradicts datum**: The recovery strategy consists in arguing about the defeating datum, to show the argument which supports it.
2) **The user questions the datum**: If the datum is questioned, the recovery strategy consists in providing information about the questioned fact. The recovery strategy therefore consists in a back-chaining (like in the case of contradicting a datum), but is realized in linguistically different terms.
3) **The user proposes an argument against the datum**: In this case, the recovery strategy employs a what-if type of reasoning. The most convenient argument pro the datum is proposed.
4) **The user contradicts/questions the warrant**: The recovery strategy consists in proposing a backing for the defeated warrant.
5) **The user rebuts the argument**: The focus comes back to the main claim and a new argumentation strategy for it is explored. The argument is based on a set of premises, or points of departure for the argumentation: the more the premises accepted by the recipient, the more successful the argument will tend to be.

In an ideal argumentation system, a set of principles for choosing among several argumentation schemes (in preparing system’s move) should be defined. Gilbert suggests the following ordered lists:

1) Choose the scheme that the user exploited most recently: these schemes are probably preferred and more easily accepted by the user.
2) Identify a preponderance of characteristics from user’s belief, which may be indicative of preferred schemes. For example, if the user’s image includes many empirical data, this may be indicative of the validity of an argumentation “from sign”.
3) Choose the schemes whose premises are accepted by the user.

Of course, the first principle is preferable if the system is able to recognize the kind of argumentation schemes the user employs in her communicative moves. In this case, Gilbert’s strategies may be implemented by observing the past user interaction: the more a given scheme was employed by the user in her moves, the more its probability of being favorite by the user increases.

C) **Message generation**

Natural Language Generation is the field of Artificial Intelligence concerned with the creation of systems capable of producing - automatically - texts in natural language (like English or Italian), starting from an underlying, non linguistic, representation of the information [see
Reiter & Dale, 97]. NLG includes several (usually sequential) steps [Reiter et al., 00] but we can distinguish two main levels, the strategic and the tactic one.

The strategic level includes content selection and text planning (rhetorical structuring and ordering of material) starting from (i) the communicative goal, (ii) a knowledge source, (iii) the, possibly revised, user model and (iv) a discourse history, if in dialogical interaction.

The content selection and text planning phase are quite overlapping with phase B of our description (Identifying and preparing the System’s move) where the “what to say” is specified.

The tactic level of NLG includes linguistic realisation (starting from the specification output from the strategic level, sentences are formed, on the basis of grammar and lexical choices and constraints). Once the syntactic structure is defined (actually this operation is not always performed first), the other main tasks of this level are:

1) **Lexicalization**: the choice of terms (nouns, verbs, adjectives, …), in agreement with the syntactic structure, to express the information contained in the output of the first level. These decisions about lexicon, to be effective, can be influenced by the user image, and by traits of the chosen terms. Let us consider, as an example, what we can call “emotional lexicalization”: depending on the user emotional state, the system can decide to extract, from an affective lexical resource (a resource that contains words referring to emotions, moods, emotion-related cognitive states) the term with the most appropriate features. So, for example in an educational scenario concerned with driving behaviour, it can decide whether to use “to kick the bucket”, “to die” or “to decease” when talking about negative outcomes of not using the helmet.

2) **Referring expression generation**: the choice of the way in which to refer to the various entities in the message or in the context (e.g. using pronouns).

3) Decision about what to say explicitly: if necessary, the system should leave implicit parts of a given argument (e.g. the warrant)

4) **Aggregation**: The choice of merging sentences to avoid too short and fragmented texts. This is a decision about the structure of a sentence and, as the other tasks, is linked to the problem of how to measure argumentation strength, probative weight, dialectical relevance and (both emotional and rational) impact.

5) Decision about the order of the components of a given argument: Depending on the user image, an argument may be more effective if the warrant is presented as first and then the other parts of the argument follow, or the inverse.

Persuasive features can have an impact on both strategic and tactic levels since the effectiveness of a message can be enhanced by appropriate selection of the content and text planning (see, for example, [Guerini et al., 04]) as well as with proper linguistic choices (lexicalization).

Multimodal Generation addresses similar problems as NLG, but goes “a step further” since the message is communicated across more than one mode, as is the case of ECA’s. The term “mode” refers primarily to the human perception involved in information acquisition (vision, audition, …) or to the cognitive and physical process in producing a message allocated on more than one medium.
Persuasive strategies will influence not only the content of the message but specifically its form. In the case of multimodality, an effective message may involve the use of a combination of coordinated means like graphics and synthesized language, each of them possibly including emotional expressions. For instance, ECA’s can help conveying a more persuasive message than a simple spoken message.

References


3. COMPUTATIONAL HUMOUR

Oliviero Stock, Carlo Strapparava and Alessandro Valitutti

Definition of the field of investigation

Computational humour is an emerging computational field in artificial intelligence that deals with building systems capable of inducing amusement and affecting the emotional state of users. Humour is an essential element in communication: it is strictly related to themes such as entertainment, fun, emotions, aesthetic pleasure, motivation, attention, engagement, creativity, and motivation. While it is generally considered merely a way to induce amusement, humour provides an important way to influence the mental state of people to improve their activity. Even though humour is a very complex capability to reproduce, it is realistic to model some types of humour production and to aim at implementing this capability in computational systems.

Humour is a powerful generator of emotions. As such it has an impact on people's psychological state, directs their attention [Kitayama and Niedenthal, 94], influences the processes of memorization [Kahneman, 73] and of decision-making [Ishii, 93], and creates desires and emotions. Actually, emotions are an extraordinary instrument of motivation and persuasion because those who are capable of transmitting and evoking them, have the power to influence other people's opinions and behaviour. Humour, therefore, allows a conscious and constructive use of the affective states generated by it. The affective induction through verbal language is particularly interesting and humour is one of the most effective ways of achieving it: the purposeful use of humorous techniques enables us to induce positive emotions and mood and to exploit their cognitive and behavioural effects. For example, the persuasive effect of humour and emotions is well known and widely employed in advertising. Advertisements have to be both short and meaningful, and are able to convey information and emotions at the same time.

Humour acts not only upon emotions, but also on human beliefs. A joke plays on the beliefs and expectations of the hearer. By infringing on them, it causes surprise and then hilarity. Jesting at beliefs and opinions, humour induces irony and accustoms people not to take themselves too seriously. Sometimes simple wit can sweep away a negative outlook that place limits on people desires and abilities, and makes people overcome self-concern and pessimism that often prevents them from pursuing more ambitious goals and objectives. Humour is the healthy way of creating ‘distance’ between one's self and the problem, a way of standing back and looking at the problem with perspective. Humour reveals new aspects, disarms and relaxes. It is also infectious, it is an important way to communicate ideas, it helps getting and keeping people's attention and it helps remembering.

Finally, humour encourages creativity. The change of perspective caused by humorous situations induces new ways of interpreting the same event. By stripping away clichés and commonplaces, and stressing their inconsistency, people can be more open to new ideas and points of view. Creativity redraws the space of possibilities and delivers unexpected solutions to problems. Actually, creative stimuli constitute one of the most effective impulses for human activity.
In this context, computational humour deserves particular attention because of its potential to change computers into an extraordinary creative and motivational tool for human activity. Machines equipped with humorous capabilities become able to play an active role in inducing users’ emotions and beliefs, and in providing a motivational support.

State of the Art

Humor has been studied since the ancient times and in the Twentieth Century saw a lot of developments. Theories can be classified along three main groups.

- **Cognitive**, where the main concepts are incongruity and contrast and the focus is on the stimulus. Many approaches are based on this. Incongruity theories focus on the element of surprise. They state that humor is created out of a conflict between what is expected and what actually occurs in the joke. This accounts for the most obvious features of a large part of humor phenomena: ambiguity or double meaning. Among the most interesting authors we like to mention Arthur Köstler that in his work about creativity [Köstler, 64] gives particular emphasis to the humor production process. In particular he proposes a theory based on bisociation, that makes “a distinction between the routine skills of thinking on a single plane, as it were, and the creative act, which, as I shall try to show, always operates on more than one plane.” The collision of these two ‘habitually, incompatible frames of reference’ produces bisociation.

- **Social**, where the main concepts are superiority, hostility, derision, disparagement and the focus is on interpersonal effects. The theories tend to be rather crude and account more for categories of humor like ridicule and funny visual inappropriateness.

- **Psychoanalytical**, where the main concepts are relief, release, liberation, sublimation, and the focus is on the audiences reaction. The relaxation of inner censorship and the release of energy that derives from it produce an intense pleasure that tends to repeat itself. The first question is of course when is the situation favourable for giving raise to this phenomenon. Freud (1905) states what are the most favorable conditions for the production of comic pleasure:
  - A generally cheerful mood in which one is “inclined to laugh”;
  - Expectation of the comic, by being attuned to comic pleasure;
  - Encouragement coming from any other pleasurable accompanying circumstance.

In contrast, major interference factors to humor are:

  - The kind of mental activity with which a person is occupied at the moment;
  - Attention is focused on the comparison from which the comic may emerge;
  - The situation gives raise at the same time to a release of strong affect.

The conditions have been discussed in the literature and should be taken into account seriously when dealing with computational interactive systems.

The field of computational verbal humor was initiated by Victor Raskin (1985). His first approach relies on the concept of script, a large chunk of information, typically commonsense stereotypical information, evoked by a word or word combination. His main hypothesis states that a text can be recognized as a single joke-carrying text if both of the following conditions are satisfied: (i) the text is compatible, fully or in part, with two different scripts (i.e. partially overlapping scripts). (ii) the two scripts with which the text is compatible are opposite. The semantic mechanisms of humor build on combinatorial rules of script-based semantic theory, according to Raskin. Essential is the relation of script oppositeness, that holds in a kind of
script ontology. To give raise to the humorous effect, a trigger, obvious or implied, is there to release the oppositeness relation among scripts: fundamentally it is based either on ambiguity or on contradiction. Clearly, the punch line of a joke implies or contains the trigger.

**Preliminary computational efforts**

So far only very limited effort has been put on building computational humour prototypes. Indeed very few working prototypes that process humorous text and/or simulate humour mechanisms exist. Mostly they are concerned with rather simple tasks.

There has been a considerable amount of research on linguistics of humour and on theories of semantics and pragmatics of humour [Attardo, 94; Attardo & Raskin, 91; Giora & Fein 99; Giora 02]; however, most of the work has not been formal enough to be used directly for computational humour modelling. An effort toward formalization of forced reinterpretation jokes has been presented by Ritchie (2002).

Within the artificial intelligence community, most writing on humour has been speculative [Minsky, 80; Hofstadter et al., 89]. Minsky made some preliminary remarks about how humour could be viewed from the artificial intelligence/cognitive science perspective, refining Freud's notion that humour is a way of bypassing our mental "censors" which control inappropriate thoughts and feelings [Freud, 05]. Utsumi (1996) outlines a logical analysis of irony, but this work has not been implemented. Among other works: De Palma and Weiner (1992) have worked on knowledge representation of riddles, Katz (1993) attempted to develop a neural model of humour. Ephratt (1990) has constructed a program that parses a limited range of ambiguous sentences and detects alternative humorous readings. A formalization, based on a cognitive approach (the belief-desire-intention model), distinguishing between real and fictional humour has been provided by Mele (2002).

An important attempt to create a computational humour prototype is the work of Binsted and Ritchie (1997). They have devised a formal model of the semantic and syntactic regularities underlying some of the simplest types of punning riddles. A punning riddle is a question-answer riddle that uses phonological ambiguity. The three main strategies used to create phonological ambiguity are syllable substitution, word substitution and metathesis.

Jester [Goldberg et al., 01] is an on-line joke recommending system using collaborative filtering in order to recommend jokes adapted to the user’s ‘sense of humour’. It does not implement any linguistic technique dealing with humour; rather it uses statistical techniques to recommend jokes based on the user's ratings of a set of sample jokes.

Many computational approaches try to deal with the incongruity theory at various level of refinement [Köstler, 64; Raskin, 85; Attardo, 94]. The incongruity theory focuses on the element of surprise. It states that humour is created out of a conflict between what is expected and what actually occurs in the joke. This accounts for the most obvious features of a large part of humour phenomena: ambiguity or double meaning.

An important working prototype mainly based on incongruity theories was developed in the context of the EU project HAHAcronym (IST-2000-30039) [Stock & Strapparava, 2003]. The project was intended to convince people about the potential of computational humour, through the demonstration of a working prototype and an assessment of scenarios where humour can add something to existing information technologies. A visible result to be checked with human evaluators was at the basis of the deal. A situation of practical interest was proposed, where there was no domain restriction and many components were present, but simpler than in
more extended scenarios. Indeed the HAHAcronym prototype was a system that makes fun of existing acronyms, or, starting from concepts provided by the user, produces a new acronym, constrained to be a word of the given language. And, of course, it had to be funny.

A relevant aspect to be taken into account is how humour is appreciated by different individuals. Personality studies regarding this specific theme give important indications Ruch (2002). One option will also be to develop humor for conversational systems, based on embodied agents. The work of Nijholt (2002), Andre and Rist (2000), and Cassell (2001) could provide the starting point for introducing dynamic humor.

Specific workshops concerned with Computational Humour have taken place in recent years and have drawn together most of the community active in the field. The proceedings of the most comprehensive events are [Holstijn & Nijholt, 96] and [Stock, Strapparava & Nijholt 02]. Ritchie (2001) has published a survey of the state of the art in the field.

Several workshops on computational humour have taken place in recent times:

- International Workshop on Computational Humour (Automatic Interpretation and Generation of Verbal Humour) September 11-14, 1996 University of Twente, Enschede, The Netherlands (http://wwwhome.cs.utwente.nl/~joris/IWCH);
- AISB 99 Symposium on AI and Creative Language: Stories and Humour, April 6-9, 1999, University of Edinburgh, UK (http://www.csl.sony.co.jp/person/kimb/aisb99);
- The April Fools’ Day workshop on Computational Humour, April 15-16, 2002, ITC-irst Trento, Italy (http://haha.ite.it/FAWCHindex.html);

There are some commercially available humour systems. ELOL (Electronic Laughing Out Loud) is an application that sends users a daily joke chosen along customized patterns. The system uses collaborative filtering (a statistical user modeling technique, used by Amazon.com, among others) to determine user's tastes, using users' rankings of the jokes they receive. ELOL is the first Internet application of NetCustomize, a 60 million-dollar-capital corporation funded by venture capital and investment funds (http://www.elol.com/).

Many humour mailing lists exist, several of them with commercial goals, such as Twisted Humour http://www.twistedhumor.com/ of Joke-a-Day http://www.rayowens.com/. These lists claim subscribers in the hundreds of thousands.

**Present challenges**

A deep modelling of humour in all of its facets is not something for the near future; the phenomena are very complex, humour is one of the most sophisticated forms of human intelligence. It is AI-complete: the problem of modelling it is as difficult to solve as the most difficult Artificial Intelligence problems. But some steps can be followed to achieve results. Here below we raise a number of questions/issues for future developments in the field.

**Theoretical issues**

At the moment, there are no general models of humour suitable for computational treatment. Psychologists have typically differentiated between two facets of research: (a) humour appreciation (the ability to understand and enjoy humorous messages), and (b) humour
production (the ability to communicate humorously in context). Although the latter seems to be easier to realize in the next few years, there are still a number of difficulties.

In order to be successfully humorous, a computational system should be able to: recognize situations appropriate for humour; choose a suitable kind of humour for the situation; generate an appropriately humorous output; and, if there is some form of interaction or control, evaluate the feedback. To make progress in this area, we have to pay attention not only to the humorous message but also to the receiver. Therefore, the system must contain a user model including characteristics of the receiver such as personality traits, gender, socio-economic status, ethnic background, etc.

Besides features of the receiver, the system has to be able to model temporarily variable features such as moods and transient emotional states. It is important to have a correlation between the contextual information representing the current situation in which both the system and the receiver are situated. On the basis of that knowledge, the system can decide if and when to generate some humorous output.

**Effects of humour**

As claimed by Ritchie [2001b], “what is more difficult at present is to produce concrete implementations which are useful or which yield interesting insights.” Then, how might we direct our efforts in this direction?

One way can consist of exploring the potential role of humour in the user interface, in particular embodied conversational agents. Can computational humorous functionalities make natural language interfaces much friendlier and make their imperfections more acceptable for users?

To what extent is it possible to use the power of humour for influencing the affective and cognitive state of the receiver? In which contexts is it useful to provoke surprise or to change the focus of attention?

There are also situations in which “inappropriate” humour induces negative emotions (such as embarrassment or indignation). Could this be exploited to direct motivation or to change the current beliefs of the user?

One possible applied scenario conveying most of these issues is a system for the creation of humorous advertisement. An important reason for choosing the advertisement domain is the fact that it is sufficiently simple (generally ads consist of short expressions) and, at the same time, they have a sufficiently rich semantic content. To create a humorous advertisement, it is necessary to have knowledge of the product to advertise and of the potential consumers. Some examples generated by the humorous creative environment are: “Thirst come, thirst served” [advertisement for a soft drink], “Come join us at Growman’s mortuary—where thousands are dying to get in....” [advertisement for a mortuary], “A chat has nine lives” [advertisement for a chat].

As claimed by Harel et al. (see section 10 in this deliverable), affective communication (in particular, affective induction) can be used in order to directly influence others’ attitudes, as consequence of the persuasive message processing. Affective communication could also be employed in order to influence the message processing itself. This influence consists of different effects. One is the possibility of directing the receiver’s attention on the message content with the aim of the promotion of the conscious processing and the use of persuasive
mechanisms based on argumentation. Other possible effects are the reduction of the “emotional interference” that some emotions (such as anger) have on reasoning, and finally the increase of the message trustworthiness. Two possible ways for realizing these functionalities seem to be humour and affective expression. In particular, it is interesting to study a possible integration of these two processes and to verify if it could allow us to develop more effective computational based persuasion models.

The relation between computational humour and creativity is another interesting field of development. Research has shown [McGhee, 71; Ziv, 84] that the use of humour stimulates creativity. Generally, the tools for computer-aided creativity are mostly employed in image processing (e.g. in architecture or in computer graphics), but there is an increasing interest in developing similar systems centred on verbal language. In particular, investigations on computer-aided creative writing show a strong analogy with computer-aided design. How could we characterize a computational humour program as a tool for encouraging the expression of creative ideas?

Resources

Even considering only systems that generate humorous messages without analyzing user responses, is it possible to play with human expectations to generate surprise and then hilarity? A required step is to collect a wide amount of common sense linguistic resources and to manipulate the stereotypical knowledge in order to infringe the expectations of the potential user. That knowledge must include familiar expressions such as proverbs, idioms, clichés, quotations, movie titles, etc. NLP techniques have to be explored providing humorous variations of familiar expressions: assonances and rhymes, wordplays, synonymy and antonymy, etc.

Evaluation methodology

Development of an evaluation methodology is crucial in order to calibrate the humorous strategies and parameters, and finally to generate effective humorous messages. Probably, it would be overambitious to provide the system itself with the capability to dynamically detect the humour appreciation of the user.

A first step to this end is to investigate the “picture” that the receiver gathers about the characteristics of the sender, as claimed by Ruch [2002]. In this case we are talking about an artificial agent. That representation can considerably influence the way in which the humour content of the message is appreciated. In order to have control of this additional aspect, it is necessary to consider the appropriate user interface, and to organize the system possible behaviours (type of humour, type of output, feedback to the user) and to integrate them into a model of personality. That configuration may allow the system to perform a more empathetic interaction with the user.

References


4. Lies and Emotion

Matthias Rehm\textsuperscript{1}, Elisabeth André\textsuperscript{1}, Fiorella de Rosis\textsuperscript{2} and Thurid Spieß\textsuperscript{1}

\textsuperscript{1}University of Augsburg, Germany
\textsuperscript{2}University of Bari, Italy

Definition of the field of investigation and applied prospects

In human-human communication, emotions are the number-one topic that people lie about and studies show that up to 30\% of social interaction longer than 10 minutes contain such deceptions, [DePaulo et al, 96]. Endowing technical systems like embodied conversational agents with the ability to detect, represent, generate and/or show emotions, it is thus indispensable to investigate the crucial questions how to handle false emotional expressions from the user and how and when to create false emotional expressions in the ECA.

As users tend to treat computers as humans, they often behave in a way that they consider as socially desirable and decide not to tell the system their real thoughts, see [Reeves & Nass, 96]. For instance, [Bosma & André, 04] observed that some users tried to be polite when their tutoring agent asked them whether they liked the system even though it was quite obvious that they were actually frustrated. It is clear that a tutoring system needs to be sensitive of the student’s actual emotional state in order to achieve its pedagogical goals. On the other hand, it does not make much sense to simulate a tutor’s “true” emotions within a learning application. Rather, the system should intentionally express emotions with the goal of motivating the students and thus increasing the learning effect.

State of the art

Lies and deception are not very popular topics and thus the literature about this phenomenon is sparse. The most fundamental and influential work was presented by Ekman and colleagues who precisely described the clues to deception that can be found in the human face and body. According to [Ekman, 92], there are at least four ways in which facial expressions may vary if they accompany lies and deceptions: microexpressions, masks, timing, asymmetry.

1 Micro-expressions:

A false emotion is displayed but the felt emotion is unconsciously expressed for the fraction of a second. The detection of such micro-expressions is possible for a trained observer.

2 Masks:

The felt emotion (e.g., disgust) is masked by a not corresponding facial expression, in general by a smile. Because we are not able to control all of our facial muscles such a masking smile is in some way deficient. Thus, it reveals at least in part the original emotion.

3 Timing:

Facial expressions accompanying true emotions do not last for a very long time. Thus,
the longer an expression lasts the more likely it is that it is accompanying a lie. A special case seems to be surprise, where elongated on- and offset times are a good indicator of a false emotion.

4 Asymmetry:

Creating facial expressions voluntarily like it occurs during lying and deceiving, those expressions tend to be displayed in an asymmetrical way, i.e., there is more action on one side of the face than on the other.

Most studied cues to lies or deception so far are visual. In the voice, however, hints to deceptive speech can also be found as lying causes emotional arousal in the liar. But in speech, we have to distinguish between ‘able’ liars and ‘bad liars’ [Anolli & Ciceri, 97]. Able liars can hide their emotions very efficiently e.g. by not only convincing the person they are lying to but also themselves of the truth of the lie. Bad liars may also succeed, but their behavior differs whether their counterpart is compliant or suspicious. When lying to someone who is compliant the verbal behavior is usually highly controlled using a more exact articulation, a deeper voice, less repetitions and corrections and more pauses. Lying to a suspicious person, on the other hand, puts the liar into a state of high emotional arousal characterised by a higher and more variable pitch, many repetitions and corrections.

In the area of human-computer interfaces, especially in the area of embodied conversational agents, lies and deception are nearly non-existing topics. [McKenzie et al, 03] describe deceiving agents as training partners for the military domain (check point). But they present no implemented system, just some ideas on the benefits for such a training scenario. Castelfranchi and Poggi developed a theory of deception in communication [Castelfranchi & Poggi, 93], which helps in modeling the decision to deceive and the choice of the deception strategy which is appropriate in a given context. This theory grounded prototyping of a deception modeling tool in which both the 'deceiver' and the 'receiver of the message' are modeled [Carofiglio et al, 01; de Rosi et al, 03a]. [Carofiglio et al, 01] model some deceptive strategies by belief networks that may be incorporated into conversational agents. More literature can be found in the area of conversational systems [Lee & Wilks, 97] and of multi-agent systems where different strategies of deception and their effects are examined. [Ward & Hexmoor, 03] as well as [Castelfranchi et al, 98] both present multi-agent simulations as testbeds for examining the effects, benefits as well as damages, of different forms of deception on the interactions of agents.

Besides work on the simulation of deceptive agents, there are various attempts to model virtual agents that deliberately oppress or express emotions since the social or pedagogical situation requires it. The Cosmo system, where the agent’s pedagogical goals drive the selection and sequencing of emotive behaviors, is one example [Lester et al, 00]. For instance, a congratulatory act triggers a motivational goal to express admiration that is conveyed with applause. To convey appropriate emotive behaviors, agents such as Cosmo need to appraise events not only from their own perspective but also from the perspective of others. Pelachaud and colleagues developed an agent that deliberately plans whether to exhibit a certain emotion or not [de Rosi et al, 03b]. A similar approach has been developed by Prendinger and colleagues who considered the social variables distance and power in order to control emotional displays of agents [Prendinger & Ishizuka, 03]. For instance, if the social distance between an agent and its conversational partner is high, the agent would not show anger to the full extent.

Other researchers have been concentrating on the opposite problem – namely how to identify the user “true” emotions (as opposed to those deliberately expressed). Most of these
approaches focus on the analysis of bio signals which are - compared to external means of expression - hardly consciously controllable by the user and thus largely circumvent the artefact of social masking. Values measured distinguish between emotions include the heart rate, skin conductivity, skin temperature, muscle activity and respiration rate. Examples of well-known applications are lie detectors.

[Bosma & André, 04] describe an attempt to reveal the user's intention from ambiguous dialogue acts by interpreting the emotional state of the user. Especially short utterances tend to be highly ambiguous when solely the linguistic data is considered. An utterance like “right” may be interpreted as a confirmation as well as a rejection, if intended cynically, and so may the absence of an utterance. The recognition of emotions in [Bosma & André, 04] is based on physiological user input. Integration of the multiple input modalities, i.e., language and emotion, is achieved by extending the approach of [Johnston & Bangalore, 00] who employ finite-state transducers for this task. It is important to note, however, that the objective of their approach was not to find out whether the user is lying. Rather they aimed at resolving ambiguities by considering his or her emotional state.

**Present challenges**

Work done so far on lies and deception in agents focuses on the effect of false information that is exchanged to achieve a predefined goal. In everyday life this is not the general case in which deception is applied. People tend to lie primarily about their emotions. Thus, endowing a computational system with the ability to detect or show emotions it is inevitable to regard this aspect of human communication giving rise to two main research questions: (i) How to handle false emotional expressions from the user and (ii) How and when to create false ECA emotional expressions (because it is expected).

Approaches so far handled deception as an ability of the agent to achieve its goals in a complex environment or to make the user believe a false statement is true. These are more or less anti-social lies, in that they are employed to solely benefit the liar. By far more frequent in human communication are so called social lies, e.g., about one's emotions. Social lies are employed to protect the face of others or the relationship to others. They are highly situation and context dependent. To give an illustrative example consider the conversation between two friends where one just bought some clothes and asks for the opinion of the other. Being polite and not wanting to strain the relationship for such a trivial reason the friend will most likely give a positive feedback regardless of his real opinion. But if the new clothes were bought for the important job interview on the next morning a more reliable answer can be expected and will be given. Social lies are more fuzzy in that they are not necessarily employed to achieve a specific well-defined goal but are more or less used to protect others or the relationship with others.

*Handling false emotional expressions from the user:* Recognition of lies or more subtle forms of deception of the user requires processing expression signs and interpreting them in the light of the context in which communication occurs. Cognitive models of the user and representation of context features help in reducing the high level of uncertainty in this recognition process, by providing a (often tentative, approximate) answer to questions such as: "Is the user expression plausible, credible, consistent with his or her past behavior?" or "Would the user receive any advantage in not saying the truth, in the present situation?" and others. The same kind of models might help in deciding whether to generate a false emotional expression in the ECA before actually generating it.
Acoustic analysis of the user’s speech can support the automatic recognition of deception. However, this has not received much scientific attention yet (apart from approaches to emotion recognition that use speech material from actors pretending emotions which is somehow also deceptive speech). Automatic lie detection from speech can be based upon recognizing the underlying emotion which is e.g. fear to be caught as a liar or a high stress level, but the automatic extraction of prosodic features specific to lies, which an on-going research project of Hirschberg and colleagues [Hirschberg, 2004] deals with, still needs to be accomplished.

Because social lies are employed in specific situations to protect others or the relationship to others, recognizing the current situation aids detecting false emotional expressions from the user. Depending on the interaction context, the system has to react to the user's false emotional reaction. In the case of uncritical situations where the user is just polite to the agent and employs a white lie, the system can safely opt to ignore this lie. Or the information about the user's lie might be used to update the user model in favor of the user furthering positive emotions on the agent's side for being taken seriously as an interaction partner. In critical situations on the other hand like e.g., in medical advising it might be dangerous to ignore the user's false affirmations that he understood correctly. A more elaborate and perhaps simpler explanation might be the right reaction in this case.

The examples emphasize the strong situation dependence of the reaction to users lies. Thus, one of the main challenges consists of integrating context representations and cognitive user models to assess the situation at hand and trigger an appropriate reaction like ignore, update user model, or elaborate.

Creating false ECA emotional expressions: Creating false emotional expressions in ECAs can be divided into two subparts, (i) Decision when to create a false emotional expression, and (ii) How to realize a false emotional expression. As mentioned before, approaches so far dealt with the strategic value of deception to achieve the agent's goals. In the case of creating false emotional expressions the agent again will have to assess the current interaction situation to decide on an appropriate emotional reaction. It remains to be seen, if the existing deception planning algorithms can be extended to incorporate reasoning about the social dimensions of the situation that might require the use of deceptive expressions.

But dealing with ECAs does not only require to solve the question how to strategically or in a socially appropriate way create false emotional expressions but also the question if and how the agent's lying attempt influences the generation of appropriate nonverbal behaviours like facial expressions or body movements. According to [Ekman, 92], concealing one's real emotion perfectly behind a faked emotion is nearly impossible. Masking smiles e.g., cannot completely override the muscular program of the original emotion because not every facial muscle can be consciously controlled. As a consequence, such a mask will always include expression segments of one's felt emotion. Although there might be scenarios where it is not desirable that an agent exhibits such clues to deceive, to create an agent that employs social lies in a believable way, the agent's nonverbal behaviour has to be influenced by the fact that it is lying and thus the agent has to exhibit these clues to deceive. In some applications like games it is even indispensable because an agent capable of perfectly concealing its felt emotions will only result in user frustration.

Modeling the influence of a lying attempt on the agent's communicative behaviours makes only sense if people are able to interpret the resulting behavior in an appropriate way and will not just ignore them or attribute them to a malfunctioning of the system or a bad design of the agent. To test users interpretation of faked facial expressions in an agent that adhere to
Ekman's clues, we devised the GAMBLE testbed that allows us to investigate in a principled way (i) if users react at all to clues that are described in the literature to indicate a lie when an ECA exhibits these clues, and (ii) if this is the case, how they react to and interpret these clues, e.g., as a malfunction of the system, as an affront, or as a useful feature that makes the interaction more engaging. Moreover, by controlling the agent's expressive behaviour we will be able to examine how different expressive behaviours of the agent exert an influence on the user's affective states.

References


5. Emotions and Politeness

Elisabeth André and Matthias Rehm

University of Augsburg, Germany

Definition of the field of investigation and applied prospects

The concept of a virtual character promotes the idea that humans, rather than interacting with tools prefer to interact with an artefact that possesses some human-like qualities at least in a large number of application domains. If it is true - as [Reeves & Nass, 96] Media Equation suggests – that people respond to computers as if they were humans, then there are good chances that people are also willing to form social relationships with virtual personalities. That is, a virtual character is not just another interface gadget. It may become a companion and even a friend to the user. A prerequisite for this vision to come true is that the virtual characters have a great deal of social competence that manifests itself in a number of different abilities.

One aspect of social interaction is the use of politeness strategies as they are described in detail in Brown and Levinson’s seminal work [Brown & Levinson, 87]. People maintain positive (self image) and negative face (wants and desires), which are continuously threatened during interactions, e.g., by commands or criticism on one’s behavior. Such speech acts are called face threatening acts (FTAs). People try to redress such undesirable acts, e.g., by referring to the good looks of the addressee before asking her for a favour.

There is an obvious relationship between FTAs and the emotional function of conversational behaviours. For instance, Brown and Levinson suggest a number of strategies that aim at lowering the face threat by verbally expressing emotions, such as empathy or sorrow. The use of additional means of expression may even increase this effect. This role of emotions is also addressed by [Bou-Franch, 02] as another level of non-verbal behaviour to redress face threats by deliberately modifying emotional reactions. On the other hand, an agent may deliberately suppress emotions, such as anger, to reduce face threats. Furthermore, an agent may exhibit certain emotions that are in conflict with its utterance or extremely exaggerated. In this case the face threat is alleviated by making the utterance appear funny. [Ekman, 1992] describes how people learn to use display rules for the expression of emotions. A prominent example is a study about Japanese subjects watching TV. Being alone they show the facial expression of joy and happiness. In company, this emotional display is suppressed.

A number of research groups have applied the sociolinguistic theories of [Brown & Levinson, 87] to the problem of modelling politeness in a dialogue, e.g. [André et al, 04; Johnson et al, 04; Moore et al, 04]. The driving force behind the implementation of politeness strategies is to improve the user’s affective state by a deliberate use of communication tactics.

State of the art

According to [Brown & Levinson, 87], politeness strategies are communicative devices for redressing the threats inherent in verbal and nonverbal utterances. Positive politeness aims at protecting the individual’s desire to be evaluated positively, for example by expressing
admiration for the addressee. Negative politeness accounts for the individual’s desire to act free from impositions, for example, by emphasizing that the final decision is up to the addressee.

[Walker et al, 97] have shown how the Brown and Levinson approach may be successfully applied to the implementation of simulated dialogues between conversational agents. Depending on the expected threat to the user’s basic desires, different dialogue styles are realized. Walker and colleagues choose one of the following four strategies:

1. Do the speech act directly
2. Orient the realization of the act to the hearer’s desire for approval (positive politeness)
3. Orient the realization of the act to the hearer’s desire for autonomy (negative politeness)
4. Do the act off record by hinting facts and/or ensuring that the interpretation is ambiguous.

There is no doubt that the perceived threat resulting from a speech act heavily depends on the user’s emotional state. For instance, if the user is already rather irritated due to communication problems, a proposal by the agent to input a long identification number is rather likely to be perceived as an impingement. Furthermore, knowledge about the causes for the user’s emotions should guide the selection of politeness strategies. Consequently, the emotional state is a factor that emerges during the interaction and dynamically influences the ongoing dialogue.

Walker and colleagues 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. Instead, it is set as a fixed parameter for a given agent and only affects the acoustic output which becomes more variable in this way. Moreover, they don’t attempt at manipulating the hearer’s emotional state by the deliberate choice of dialogue strategies.

Recent approaches focus on the situational factors that should be considered when selecting specific strategies to mitigate face threats.

[Johnson et al, 04] question whether there is an absolute ranking of politeness strategies as the theory by Brown and Levinson indicates and suggest to treat positive and negative politeness separately. That is the speaker decides whether to mitigate a positive or a negative face threat and selects re-dressive strategies depending on the type of face threat. To simulate such a behavior, Johnson and colleagues coupled a natural language generator with a politeness module.

[André et al, 04] focus on the causes for the user’s emotions as an important situational factor. For instance, a student that feels frustrated after repeatedly providing wrong answers to a tutoring system might interpret an autonomy-oriented strategy, such as “Why not try it again?”, as pure irony. Since the event is not desirable to the user, but the user is responsible for the failure, the system would rather go for an approval-oriented strategy. On the other hand, if the failure is obviously caused by the system, the user might feel offended if the system applies an approval-oriented strategy and tries to convey the impression that they are collaborators, for instance by uttering: “Don’t worry! Together, we will manage this!”. Here, an excuse would be more appropriate. To consider the user’s emotions when deciding between several communicative strategies, André and colleagues integrate a cognitive model of emotions with a theory of politeness.
[Porayska-Pomsta & Mellish, 04] make use of Brown’s Levinson’s model in order to motivate linguistic variations of a natural language generator. In order to apply the framework to an educational application, a number of changes of the original theory were necessary. First of all, the authors noticed that a teacher’s utterance may at the same time accommodate for the student’s need for autonomy as well as his or her need for approval. Consequently, they decided to characterize their strategies in terms of the degree to which the two needs were addressed: <Aut,App>. Furthermore, they relied on a larger set of situational factors derived from educational literature, dialogue analyses as well as interviews with teachers. To model the bidirectional dependencies between the autonomy and approval values on the one hand and the situational factors on the other hand, the authors employed Bayesian networks. An empirical study was performed in order to find out to what degree the system generated strategies coincide with those suggested by human teachers.

While the approaches above concentrate on linguistic strategies to portray politeness in dialogue systems, other approaches concentrate on the oppression of emotions as a sign of politeness.

[Prendinger & Ishizuka, 01] consider Brown’s and Levinson’s social variables distance and power in order to control emotional displays of agents. For instance, if the social distance between an agent and its conversational partner is high, the agent would not show anger to the full extent. This behavior can be interpreted as an attempt to reduce the face threat for the conversational partner. Even though Prendinger and Ishizuka make use of Brown and Levinson’s social variables, they do not explicitly relate the suppression of emotions to face threats.

The question of whether to show an emotion or not has also been handled by [Pelachaud et al, 02]. However, while Prendinger restricts himself to deciding whether to display an emotion or not, Pelachaud and colleagues also vary the propositional contents of the utterance.

Another question is, how users respond to systems that simulate behaviours of politeness common in human-human dialogue.

[Bickmore & Cassell, 00] describe how small talk is utilized to build up common ground between an embodied conversational agent and the user based on an extension of Brown’s and Levinson’s theory of politeness. An empirical evaluation of their system indicates that an agent capable of social dialogue may indeed improve the user’s perception of the interaction.

[Alexandris & Fotinea, 04] conducted an empirical study for a Speech Technology application that was developed for customer service in Modern Greek. The study revealed that dialogues in Modern Greek with discourse particles indicating positive politeness are perceived as friendlier and more natural. Interestingly, other forms of politeness that are common in other languages, such as negative politeness, are often perceived as unnatural.

While the empirical work presented above was based on explicit system ratings by human users, [Prendinger et al, 03] showed by measuring the user’s skin conductivity that users seem to be less stressed if the agent apologizes for delays.

**Present challenges**

Even though the theory by Brown and Levinson seems to provide a promising basis for the implementation of politeness strategies, it also creates a number of problems, see [Knapp &
Daly, 03] for a more detailed discussion. First of all, the linear ordering of politeness strategies from direct over approval-oriented and autonomy-oriented to off-record may lead to inconsistencies. There is no doubt that the perceived politeness of any strategy may drastically vary depending on a number of context factors, such as the user’s personality, his or her ability to perform a certain task etc. Indeed, a number of studies revealed that autonomy-oriented strategies are not always conceived as more polite than approval-oriented strategies. Furthermore, speakers tend to use several combinations of politeness strategies within one utterance. Even a dialogue act that is aimed at sustaining negative face of the hearer can be employed in an approval-oriented strategy. It is also questionable whether indirectness and vagueness are actually useful means to redress face threats. In some cases, vague system utterances might even increase the user’s negative emotional state (instead of mitigating it).

There is sparse information on non-verbal means related to politeness strategies. Previous work has concentrated for the most part on the linguistic aspects of FTAs, i.e., on verbal means to deliver and redress FTAs. But FTAs are inherently multi-modal. Dressing up a threat in a joke usually only works if the speaker shows in his whole appearance (facial expression, body posture) that he is telling a joke. Otherwise the threat might be even more severe than it is.

First studies indicate that the appropriate use of theories of politeness may have a positive influence on the user's emotional state. Nevertheless, it is still rather unclear which strategies of politeness to employ in which context. Furthermore, standardized measurements are missing for determining whether the coded strategies actually lead to an improved perception of the interaction experience.

Finally, it is a great challenge to adapt politeness behaviours to a specific culture. For instance, people in individualistic cultures seem to display their emotions more overtly which is usually perceived as impolite in collectivist cultures. On the contrary, people from collectivist cultures tend to suppress emotions that most likely conflict with the mood of the group.

References


6. Importance of the Eyes and Gaze

Catherine Pelachaud

University of Paris 8, France

Gaze is a salient behaviour. It has been shown to play an important role in social hierarchies [AC76]. For instance, studies of gorillas have shown that visual attention is often directed towards the more dominant members of the group, the alpha males, to form an attentional structure. Attention is directed upwards to more dominant animals resulting in social cohesion and a dominance hierarchy; in this way infants pay attention to their mothers, mothers to their mates and males to more dominant males [Cha67]. Such an attention structure may also apply to humans, as noted by Argyle and Cook [AC76]. Gaze also plays an important role in signalling interpersonal attitudes and emotions. It can be used as a threat signal. In experiments conducted by Exline and Yellin [EY69], rhesus monkeys that were stared at attacked in 47% of trials. Eyespots are used by some species as a form of defence against predators; for example, butterflies will often expose their eye-spots when predators approach in order to deter or mislead their attack. There is evidence that human infants pay attention to the direction of gaze of others from as young as 3 months [FJBS00] and also, that infants are faster in making saccadic eye movements to peripheral targets cued by the eye gaze of a central face [SB75]. Finally, there is an adult propensity to follow the gaze of others which seems to be related to an ability to attribute intentions to the gazer.

Gaze for social interaction

Gaze awareness and eye contact are extremely important during face-to-face interactions. Not only does gaze help to direct practical tasks, such as turn-taking in conversation, but it also conveys a range of social signs. People who use frequent eye contact may be perceived as being more attentive, friendly, cooperative, confident, mature and sincere, while those who avoid it often give impressions of coldness, pessimism, defensiveness, evasiveness, indifference and submissiveness. Experiments draw attention to the persuasive nature of gaze: people using eye contact received more jobs offers after interviewing, more help after asking for it and teachers who use eye contact have been found to have more productive students. Perhaps on insight into the nature of gaze is hinted at by the fact that individuals have been found to look more at each other when they are cooperating than when they are competing [Arg88].

Garau et al. investigated the importance of eye gaze in humanoid avatars representing people engaged in conversation. They tested whether having an avatar would improve the quality of communication compared to an audio-only case, and also, in the case where an avatar was present, if a gaze model related to the conversation improving the quality of communication. It was found that gaze behaviours related to the conversational context provide a marked improvement for human interaction over an avatar that just exhibits random gaze patterns.

Colburn et al. [CCD00] have investigated the use of eye gaze in avatars by presenting participants with the three visual stimuli: a blank screen, a fixed-gaze avatar and an avatar with a functioning eye gaze model. They have found that viewers respond to avatars that have natural eye-gaze patterns by changing their own eye-gaze patterns. Participants were also
found to look at the screen more in the case where the avatar was present and its gaze model was active.

As Garau et al. notes, these results suggest that an informed eye gaze model motivates participants to pay more attention to the avatar during conversation.

**Generation of gaze behaviours**

Cassell and Thorisson [CT99], conducted a number of studies on user interaction with autonomous human-like agents. Importantly, they demonstrated that non-verbal communication behaviours, such as gaze, are very important for conversational plausibility.

Cassell and Vilhjálmsson [CV99] present the BodyChat system, which automates the animation of communicative behaviours based on context analysis and discourse theory. It allows users to communicate via a text interface while the avatars automatically animate with appropriate gaze, salutations, turn taking and facial expressions. Cassell et al. [CVB01] have also presented BEAT, a toolkit for the animation of expressive behaviours. The toolkit automatically selects appropriate gestures, facial expressions, pauses for an input text and provides synchronisation information that is necessary to animate the behaviours in conjunction with the character’s speech. The toolkit’s gaze generator is based on an algorithm from Cassell and Torres [CTP99] that accounts for the relationship of gaze behaviour to turn-taking and information structure.

Chopra-Khullar [CK99] presents a computational framework for generating realtime visual attention behaviour in a simulated human agent based on observations from psychology, human factors and computer vision. The AVA, or Automated Visual Attending, system is composed of a three-level attention hierarchy, allowing eye movement patterns to be combined and interact with each other and with attention capturing events. At the top of hierarchy are intentional tasks. Tasks are supplied to the system in the form of text, for example pick up the newspaper. The locations of objects to look at in these deliberate tasks are queued into an Intention List. This top level of the hierarchy has precedence over the lower levels. The second level of the hierarchy handles exogenous behaviours, or behaviours invoked by stimuli external to the character. Peripheral motion and flashing lights are the main stimuli considered; peripheral motion locations are entered into a peripheral motion list or Plist. The final level, with lowest precedence, consists of idle viewing. Idle viewing mode is activated when there are no tasks at hand and attention is being captured in a bottom-up manner. In this mode, a simplified image analysis technique is used to select directions of interest. An attention arbitrator is used to select between competing behaviours in the hierarchy and direct the agent’s gaze in an appropriate manner. The arbitrator works by considering the attention entries in the Intention and Periphery lists. If there are entries only in the Intention List, then they are looked at in FIFO order. If there are entries in both the Intention List and Plist, irrelevant stimuli in the Plist are looked at probabilistically. If there are no entries in either the Intention List or the PList, then idle looking commences. Within each list, competition between entries is equal; that is, behaviours of the same type have an equal probability of being attended to.

Gillies [Gil01] presents a simulation of what a character is attending to in the environment in order to generate, among other behaviours, navigation and eye movements. Attending behaviours are generated in an object-centric manner; ray-casting is used for occlusion testing and retinal images are not considered. Objects contain basic properties that are defined as comms and agents. They also contain features that represent more abstract concepts such as
beauty and interest. Each object has a variable number of features and new features can be added by the user. Every feature has a corresponding probability that determines its chance of being looked at. The system does not attempt to attribute any meaning to object properties;

actors are programmed to show more interest in some properties than in others. Attention is modelled through the interactions of numerous attention agents that are controlled and arbitrated by a central attention manager. In contrast to a request queuing methodology, agents send attention requests to the attention manager. Four types of attending behaviour are implemented: immediate, monitor, search and undirected. Request types are prioritised in that order. Undirected looking has the lowest priority and is activated only if the character has nothing else to attend to.

Computational model of gaze behaviors

Various systems [BES97; CAS94; CTP99; THO02; CAS99] simulate face-to-face conversation with a user. Such systems combine several modules for the perception and generation of audio and visual signals. Conversational spoken dialogue modules (i.e., speech recognition and natural language understanding) are integrated with the analysis and recognition of nonverbal signals such as facial expressions, eye and hand movements. This audio and visual information is used to emulate turn-taking protocols [Bes97; CPBSABDPSS94; CTP99; Tho02; CAS99], to call for the user’s attention [WRLKT96] and to indicate objects of interest in the conversation [LSCVF00; Tho97;Bes97]. These systems produce context-sensitive facial expressions, gaze and pointing gestures.

On the other hand [CCD00, FOMSH02, LBB02] use a statistical model to drive eye movements. In particular, the model of Colburn et al. [CCD00] uses hierarchical state machines to compute gaze for both one-on-one conversation than multi-party interactions. On the other hand Fukayama et al. [FOMSH02] use a two-state Markov model which outputs gaze points in the space derived from three gaze parameters (amount of gaze, mean duration of gaze and gaze points while averted). These three parameters have been selected based on gaze perception studies. While the previous researches focused more on eye gaze as communication channel, Lee et al. [LBB02] an eye movement model based on empirical studies of saccades and statistical models of eye-tracking data. An eye saccade model is provided for both talking and listening modes. The eye movement is very realistic but no information on the communication functions of gaze drives the model. Most models presented so far concentrate either on the communicative aspects of gaze or on a statistical model. In this paper we propose a method that combines both approaches to get a more natural as well as meaningful gaze behavior.

The lexicon of gaze

Poggi et al [PPR00] have proposed a method to build a lexicon of gaze. A lexicon is constituted of (meaning, signal) pairs. On the signal side, Poggi et al [PPR00] have defined a set of parameters to describe eye movement. On the meaning side, each element is described following the taxonomy of communicative functions as proposed by Isabella Poggi [PogTA]. Four broad classes are being distinguished based on the speaker’s communicative intention. These classes are [PogTA]: information about the speaker's beliefs, information about the speaker's intentions, information about the speaker's affective state and meta-cognitive information about the speaker's mental actions.
The meaning side of gaze

Poggi et al [PPR00] propose a gaze model based on the communicative functions model defined by Poggi [PogTA]. This model predicts what should be the value of gaze in order to have a given meaning in a given conversational context. For example if at a point of her speech, the agent wants to emphasize a given word, the model will output that the agent should gaze at her conversant. But using only this model creates a very deterministic behavior model: at every communicative function associated with a meaning corresponds all the time the same signals. This model also does not take into account the duration that a given signal remains on the face. Indeed, this model is event-driven: it is only when a communicative function is specified that the associated signals are computed and that the corresponding behaviors may vary. Such a model used by itself has several drawbacks: first of all it does not take into account the past nor the current gaze behaviors to compute the new one, neither does it consider the duration gaze states by S and L have lasted.

To embed this model into temporal considerations as well as to compensate somehow missing factors in this gaze model (such as social and culture aspects) Pelachaud and Bilvi [PB03] have developed a statistical model. The previously developed model is used to compute what should be the communicative gaze behavior; the gaze behavior outputted by this model is then probabilistically modified. The probabilistic model is not simply a random function, rather it is a statistical model defined with constraints. This model has been built using data reported in [CP02]. This data corresponds to interactions between two subjects lasting between 20 and 30 minute. A number of behaviors (vocalic behaviors, gaze, smiles and laughter, head nods, back channels, posture, illustrator gestures, and adaptor gestures) have been coded every 1/10th of second. Analysis of this data was done having in mind to establish two sets of rules [CP02]: The first one, called `sequence rules', refers to the time a behavior change occurs and its relation with other behaviors (does breaking mutual gaze happened by having both conversants breaking the gaze simultaneously or one after the other); while the second set of rules, called `distributional rules' refers to probabilistic analysis of the data (what is the probability to have mutual gaze and mutual smile). The gaze model comprises two main steps:

1. **Communicative prediction:** First it applies the communicative function model as introduced in (Poggi, to appear) to compute the gaze model so as to convey a given meaning.

2. **Statistical prediction:** The second step is to compute the final gaze behavior using a statistical model and considering information such as: what is the gaze behavior for the Speaker (S) and the Listener (L) that was computed in step one of our algorithm, in which gaze behavior S and L were previously, the durations of the current gaze of S and of L.

The model introduces several temporal gaze parameters, namely:

- Maximum time of mutual gaze
- Maximum time S (resp. L) looks at L (resp. S)
- Maximum time S (resp. L) looks away from L (resp. S)

By specifying different values for these temporal parameters, the model is able to simulate various gaze behavior; such as the speaker barely looks at the listener or that both agents look at each other quite a lot (“mutual gaze” is set to a high value).
References


7. Narrative and story-telling

Ruth Aylett
University of Salford, UK

Definition of the field of investigation and applied prospects

Narrative has become an increasingly important research area in the recent period as its importance as a human mechanism for structuring experience and understanding the world has become recognised. As a research field it goes back to Aristotle [20], and there is a rich literature of narrative theory in the humanities, but it has moved out of the artistic and creative areas, in which it has always been important, into a number of others in which the application of technology seeks to extend the power of narrative in new ways [7].

Education and training have always been fields in which narrative has played an important role: “Ask people to tell you about their most significant learning, and they will tell you a story. All deep learning takes the form of a story. All powerful education is an adventure that can be narrated.” (Michael L. Umphrey [http://www.edheritage.org/tools/alert.htm]) Because children have a well developed understanding of narrative [31], teachers often use narrative to teach children difficult concepts and to bring structure to the curriculum [13]. For this reason, children have been the target users of many narratively-based e-learning systems developed by researchers. Narrative has also been seen as a powerful method for adding knowledge to interactive graphical environments in general, sometimes through the medium of virtual guides.

However it has also been shown that experts in any field tend to embody their knowledge in the form of narrative [14]. Thus work in organizational knowledge-management has recently focused on narrative as one of the significant components of organizational memory and an important dissemination mechanism from experts to novices in business environments, as Deming argues:

“It was ever present, albeit low-key and usually unrecognized. It was embodied in the unstated assumptions from which optimization proceeded – the underlying narrative of how the business works. And in practice, it played a central role in decision-making: although good business cases are developed through the use of numbers, they are usually approved on the basis of a story. So despite its lowly official status, narrative remained pervasive, and ready to fill whatever space it was allotted. This is hardly surprising, since narrative is universal in all cultures including our own, because this is how human beings think.” [10].

The ability to express and structure experience has also led to the use of narrative approaches in systems aimed at therapeutic use – for the support of mothers of young cancer victims (Carmen’s Bright IDEAS) or for the articulation of the emotions of depressed adolescents.

Finally, computer games and electronic entertainment in particular have investigated the use of narrative to create a more immersive and engaging experience and to move beyond existing genres, as for example in The Sims.
State of the art

It is apparent that narrative theories have been heavily influenced by the idea that narrative must be authored. The relevant works and theories of Greek philosophy [16], literary critics [1,17], cinema critics [2, 3, 4] and classic theatrical dramaturges, all converge towards an authorial view on narrative. However, a characteristic of the narratives of current interest is that the role of the subject (i.e. the user) to whom the narrative is communicated is, in terms of interaction, “active” in the unfolding of the narrative as opposed to its “passive” role in most of other classical narrative media (i.e. the spectator). The distinction between the two terms "user" and "spectator" refers to two fundamentally different roles and set of characteristics in a narrative or interactive display. Whereas the definition of a spectator is common regardless of media consideration, the definition of a user varies within particular contexts. Such a distinction between spectator and user implies that a differentiation must also be made between authorial and interactive approaches to narrative. On one hand, narrative is seen as an artefact that can be studied, involving non-interactive spectators, whereas, on the other hand, it could be perceived as the dynamic process resulting from the interaction between characters and its impact on the user (the ‘Storification’ process).

Many research groups are dealing with the problems that this relatively new area provides. Some groups are trying to obtain methods and mechanisms to automatically generate narratives (this is the plot based approach), while others are more concentrated on trying to obtain believable characters (character based approach). A few try to reconcile both approaches.

A wide number of different types of systems are being built by researchers, many with different purposes. In some cases [29, 30, 35, 34, 32, 36] the objective is virtual drama, with embodied characters playing the role of virtual actors. Computer based storytelling is also a flourishing area [33]. Here there is a more legitimate place for the concept of a narrator. Some researchers work with story pieces which are small parts of stories, while other researchers work with simple events. People working with story pieces usually approach the problem in one of two ways: connecting these pieces in a hypertext-like network (and then navigating through this network to get a story) or using a narrative theory that specifying overall story structures, in order to combine the story pieces. The granularity of the information used to construct stories is a determinant factor. In general, the larger the granules used to build with, the easier it is to build. The smaller the granule, the more precise and smooth the building can be. Systems with large granules are easier to maintain and their related story generation process is simpler, but the generated stories are more "predefined". Systems with small granules are more difficult to maintain and their related story generation process is more complex (at least if the story is expected to show narrative devices as climax, tension, etc.), but the generated stories are more surprising.

As pointed out in [37], from an artificial intelligence point of view, two different approaches have been taken to try to solve the narrative generation problem. The knowledge based approach makes an a priori attempt to capture the rules for successfully solving or navigating a domain, the behavior based approach instead relies on a set of lower level competences which are each "experts" at solving one small part of the larger problem domain. In this sense, work which tries to apply Propps ideas to generate narrative is using a knowledge based approach, while that giving every character autonomy in order to get an emergent narrative from their interactions is using a behavior based approach.

From a more linguistic point of view, as pointed out in [21], previous work on story generation has generally taken one of two approaches: structuralist or transformational.
Structuralists use real-world story structures such as canned story sequences and story grammars to generate stories, while transformationalists believe that story-telling expertise can be encoded by rules, or narrative goals that are applied to story elements such as settings and characters. More surprising and emergent stories are obtained by using: a character based approach, events, small granules, a behavior based approach or transformationalism.

Much of the work dealing with the first global set of approaches, that is, these which obtain more predefined stories, try to order story pieces into a coherent whole story. This work relies on a long tradition of work that suggests that traditional narrative, both oral and written, has well-defined high-level structures. In this sense, the theories of Polti, Propp, Aarne, Thompson and Branigan are especially interesting. Polti, a French literary critic, published an analysis of literary plots [27] where he claimed that the totality of all literature and drama can be reduced to thirty-six basic dramatic situations (for example: rivalry of superior and inferior, crime pursued by vengeance, obstacles to love). Propp [23] proposed a set of thirty-one narrative characteristics to provide a method for understanding and cataloguing Russian fairy tales. He broke up a large number of fairy tales into components and then made a comparison of the tales according to their components. Propp created two levels of these components or categories, the thirty-one major characteristics, with each one including one or more subcategories. His work offers detailed patterns of narrative events.

Aarne created a tale type and motif index for the comparative study of folk tales. After Aarne's death, this work was taken up by Thompson [19], who expanded the tale type and motif index to include tale types and motifs from many cultures around the world. A tale type is made up of a one or more motifs. Branigan researched the area of narrative structure, proposing a narrative schema [6] consisting of: introduction of setting and characters, explanation of a state of affairs, initiating event, emotional response or statement of a goal by the protagonist, complicating emotions, outcome, and reactions to outcome. The application of Propp's morphological approach to interactive storytelling was suggested in [22] and have been used, for example, in [26], where the authors propose an automated narration system based on a structuring method that construct narratives by combining story pieces. The generated structure does not depend on the content of the story pieces, but on a semiotic definition of the several functional parts found within a story.

Brooks [37] uses Branigan's structures as a general guide in developing a framework for structuring metlinear stories. He proposed the term *metlinear narrative* for a collection of small related story pieces designed to be arranged in many different ways, to tell many different linear stories from different points of view, with the aid of a story-engine which sequences the story pieces. Brooks suggested that a writing tool which offers the author knowledgeable feedback about narrative construction and context during the creative process is essential to the task of creating metlinear narratives of significant dimension. In particular he designed AgentStories [37], a software tool consisting of a set of environments for authoring pieces of stories, authoring the relationships between the many story pieces, and for designing an abstract narrative structure for sequencing those pieces.

On the other hand, an example of the works dealing with the second global set of approaches (that is, these which obtain more surprising stories) is the Oz project [9]. From the Oz team's point of view, in order to build characters that have the "illusion of life", they will need to have broad capabilities to interact with complex environments. This led Oz to develop a research philosophy and technology with strong affinities to behavioral artificial intelligence.

Terminal Time [40] was a system that combined historical events, ideological rhetoric, familiar forms of TV documentary, consumer polls and artificial intelligence algorithms to
create hybrid cinematic experiences for mass audiences that were different every time. Through an audience response measuring device connected to a computer, viewing audiences responded to periodic questions reminiscent of marketing polls. Their answers to this questions allowed the computer program to create historical narratives that attempted to mirror and often exaggerate their biases and desires. The engine used the past 1000 years of world history as "fuel" for creating these custom-made historical documentaries.

The Kyoto Tour Guide [38], was a project aiming to develop an agent integrated in an on-line 3D website tour, a digital version of Kyoto. It explored ways to make an agent's narrative effectively adaptive to different groups who take the tour, like an expert human tour guide. The authors focused in developing storytelling strategies that produce an engaging experience for tour takers. They derived a list of target abilities for the agent by researching the behavior of actual tour guides in Kyoto. They found that tour guides made use of illustrative stories frequently, supplementing the rich visual environment of the city with explanations of how Japanese people, both past and present made use of these settings.

MAKEBELIEVE [21] is a story generation agent that generated short fictional text of 5 to 20 lines when the user supplied the first line of the story. MAKEBELIEVE used a subset of sentences extracted from Open Mind Commonsense (OMCS) [39] that describe simple commonsense causations. The authors reported that, although the system does not incorporate plot devices such as motifs, climax, tension, etc., many users involved in the evaluation nonetheless felt that these devices were present in the generated stories.

In [20], Dautenhahn explores how theories on autobiographical memory can be applied to a virtual environment (VE). She classifies narrative agents into four types: Type 0 (an agent which is always telling a single story), Type I (an agent which can tell a great variety of stories, but the stories are not situated to the conversational context, i.e. the agent randomly selects a single story from its story-base and tells it in exactly the same way as it was stored), Type II (an agent which selects the story which fits the current context best), Type III (an agent that tells and listens to stories, 'understands' them in the sense that it is able to interpret the meaning and content of the story and finds in its own story-base the most similar story which is then adapted in order to produce an appropriate response) and Type IV (an autobiographic agent, whose story-telling ability is linked to a living, autonomous agent, a personality).

**Present challenges**

The key theoretical challenge faced by work in this area is to find solutions to the ‘narrative paradox’ [18] which bedevils work on interactive narrative. While interactivity supposes user freedom to choose actions and interactions, narrative requires structure (classically, a beginning a middle and an end) which clashes badly with this freedom. The paradox is felt most acutely in participative narrative in which the user plays a role in creating a story, and least acutely where the user is made the author (thus merging author and spectator). Intermediate solutions include the use of pre-defined plan-trees for characters [29], the development of very large numbers of alternative story fragments [8], branching narratives, episodic interaction using concepts like Boal’s ‘Spectators’ [5] and character-based emergent narrative [28].

A further set of engineering challenges lies in the integrative effect of narrative systems, which frequently demand virtual characters with affective architectures (WP7) and expressive behaviour (WP6), and means of detecting the affective engagement of the user in the narrative process (WP4). There are also challenging issues of evaluation (WP9).
References


[28] Louchart S, Aylett R. Narrative theories and emergent interactive narrative. NILE02. 2002; 1-8


[33] Szilas, Nicolas. Interactive Drama: Beyond Linear Narrative. AAAI, 1999

[34] Izzo Gary (1997) The art of Play, the new genre of interactive theatre, Heinemann Portsmouth. 1997


8. Response Generation in Dialogue Systems

Ian O’Neill, Philip Hanna and Xingkun Liu

School of Computer Science, The Queen’s University of Belfast

Definition of the field of investigation and applied prospects

Advanced spoken dialogue systems are typically used to help telephone callers complete a transaction in a well-defined business domain. For example, a user may make a train timetable enquiry or book accommodation. Advanced dialogue systems have the advantage that they allow mixed-initiative interactions – that is, they permit the caller to express his/her intentions using whatever words seem most appropriate and to provide more information or less information than the system requested. The dialogue manager in a spoken language system determines the system’s response to the user’s utterances. The system may confirm what the user has said, check to see if the user’s request can be fulfilled (e.g. is there a train on the requested day at the requested time?), or ask for more information. The system may perform a combination of confirming, validating and requesting in a single dialogue turn, in much the same manner as a human dialogue partner. In mixed initiative dialogues that deal with more than one domain (e.g. enquiries about accommodation and events, and supporting exchanges about payment and addresses), the system must be able to identify the topic of the dialogue – since in a mixed initiative exchange the user is free to provide whatever information he or she deems appropriate at the particular dialogue turn (e.g. the user may start to ask about events while booking accommodation). Having identified the (ongoing) topic of the dialogue, the system must then apply appropriate dialogue management expertise.

State of the art

A dialogue manager is only one of several main modules that are required to implement an end-to-end dialogue system. As a whole, the spoken dialogue system must

- recognise the words that the user says,
- attempt to determine the intention behind the words,
- decide how to respond, potentially using information from a database,
- have the ‘conceptual’ response generated as a well-formed natural language phrase,
- and utter the phrase as synthesised or concatenated speech.

Thus the dialogue manager requires a means of communicating with the other, probably third-party modules that it needs for support – and whose implementation details need not concern the DM developers. These components typically include: an automatic speech recogniser, a semantic parser, a database ‘back-end’, a natural language generator, and a text-to-speech engine.

The DARPA (Defense Advanced Research Projects Agency) Communicator architecture is a recent solution to the problem of facilitating interaction between the different dialogue system modules. Communicator systems use a ‘hub-and-spoke’ architecture: each module seeks services from and provides services to the other modules by communicating with them through a central software router, the Galaxy hub, developed by the Spoken Language
Systems group at MIT [http://www.sls.csail.mit.edu/sls/technologies/galaxy.shtml], subsequently released as an open source package in collaboration with the MITRE Corporation, and now available on SourceForge [http://communicator.sourceforge.net/]. Java, along with C and C++ is supported in the API (application program interface) to the Galaxy hub. Thus an object-based dialogue manager written in Java, for example, can interact via the Galaxy hub with the third-party modules that enable the system as a whole to hear and understand the user, to interact with a database and utter a response. The relationship between the dialogue manager, the Galaxy hub, and the other system modules is represented schematically in Figure 1. Of immediate concern to a DM are the parser – which creates text-based semantic representations of key concepts in the user’s utterance, and requires DM developers to create appropriate semantic grammars for it to work with – and the natural language generator (NLG) – which also requires rules to translate semantic representations emerging from the DM into natural language utterances.

![DARPA Communicator architecture](image)

**Figure 1. DARPA Communicator architecture**

A number of approaches have been taken to development of the DM itself.

1. **A object-oriented and frame-based approach to dialogue management**

   In an OO implementation, routine functionality, such as turn-taking and confirming what the user has said, is made available through inheritance to more specialised dialogue components. The attraction of object-orientation is that it can be used to separate generic dialogue behaviour from domain-specific behaviour. Maintainability is therefore enhanced: generic behaviour need be implemented only once; new agents can inherit the generic dialogue behaviour and possibly refine the specialised behaviour of existing experts: e.g. a TheatreExpert can be introduced as a subclass of an EventExpert. Generic behaviour will typically include confirmation strategies: confirm information that is new to the system, query information that has been changed or negated, etc.

   The more specialised components encapsulate ‘expert rules’ for enquiry processing in a particular business domain (e.g. “if a location and a hotel have been requested, then check what class of hotel is required”) or rules that represent know-how that can be used in a number of business domains (e.g. how to elicit credit card details). In order to enable mixed initiative interactions across domains, the system’s behaviour is modelled as a collaboration between a cohort of ‘agents’ (AccommodationExpert, TheatreExpert, etc.). Each agent encapsulates a skillset for what may be a substantial dialogue or subdialogue relevant to its domain.

   Domain spotting can be used to enable the system to choose the most appropriate expert for a particular dialogue task. In a domain spotting system, each agent can give a score to parse tags that are relevant to its area of expertise. When it comes to formulating a response to a user utterance, a ‘domain spotter’ component passes dialogue control to the agent that scores
the user’s utterance highest, or else the domain spotter asks the user to choose (‘Is that a theatre or a cinema enquiry?’), if the user’s utterance is ambiguous (‘I’d like to go to see Grease.’) The selected agent – a CinemaExpert or a TheatreExpert, say – then becomes the handling agent.

An OO implementation such as this can be frame-based: in a frame-based, phrase-spotting system any relevant information will be used by the handling agent to populate slots in the current dialogue frame. The dialogue frame is a set of attributes that corresponds to the frame of slots that must typically be filled to complete a transaction in a particular domain – accommodation booking or payments, say. Attributes in a dialogue frame may comprise not just attribute values (e.g. the name of a show, the number of tickets required) but also information about the attribute values (has the user repeated this value, how confirmed is the value, what has the system decided to do about the value – e.g explicitly confirm the name of the show.) Dialogue frames evolved over several turns constitute a discourse history – the system’s memory of what the user said and what intentions the system formulated. This helps the system make sense of the user’s responses: e.g. if the user answers “Yes” in response to the system’s observation “So, you’re staying in Dublin” (an implicit confirmation, which the system will have recorded as its intention towards the value Dublin in the discourse history, then Dublin can be regarded as the confirmed location.)

The Queen’s Communicator [O’Neill & McTear, 00; O’Neill & McTear, 02; O’Neill et al., 04], is an example of a mixed initiative (either the user or the system can lead the dialogue) object-oriented, and frame-based dialogue system that uses the Communicator architecture based on the Galaxy software hub. Other mixed initiative Communicator systems are the CU Communicator [http://csrl.colorado.edu/begin/web/cumove/cumonomunicator.html] and the CMU Communicator [http://fife.speech.cs.cmu.edu/Communicator/index.html], both of which are available for download as complete systems (including source code).

2. Finite State Dialogue Management

Finite state dialogue management is particularly well suited to relatively straightforward dialogue scenarios, where the dialogue passes through a number of predictable stages or states, at each of which the user has a fixed number of choices. Buying a pizza is indicative of the sort of scenario to which this approach is suited: "What kind of topping would you like – cheese and tomato, pepperoni or vegetarian?" – and the user must choose one of the options offered, whereupon a new question and further options (sometimes yes/no questions) are presented: "Do you want a salad with that?"

This makes for an easily managed dialogue, but has the disadvantages that the developer must design the dialogue flow, which the user must follow step by step, usually supplying one piece of information at a time – even if he or she might want to supply several pieces of information in one dialogue turn. It can also be problematic if the user wants to change a value several turns after having originally supplied it. Changing information may require having to go back several states and then having to redo several states (or else the developer must explicitly incorporate shortcuts between states).

Nonetheless, where a tightly controlled dialogue flow is required, and the dialogue can be completed in a relatively small number of moves, a finite state approach to dialogue management can be very effective. The CSLU Toolkit from the Oregon Graduate Institute offers a drag-and-drop development environment for finite state dialogues, and incorporates a range of lip-synced animated characters for a multimodal dialogue experience. The Toolkit is
available for download and comes with a suite of very useful tutorials [http://speech.bme.ogi.edu/toolkit/docs/2.0/apps/rad/quickstart.html].

3. Intelligent problem solving agents

One of the most powerful approaches to dialogue management is that based on intelligent problem solving agents. The approach is typified by the work currently being conducted by the team at Rochester University in the US [http://www.cs.rochester.edu/research/trips/]. Here dialogues are regarded as a series of objectives and subobjectives, which can be accomplished by enacting recipes, which in turn entail the use of resources. [Blaylock et al., 03]

The advantage of this approach is that the system, in formulating its responses to user utterances, is able to take into account the evolving dialogue context and the emerging aims and objectives of the user. Such a system, while of course operating within certain contextual limits (logistics involving transportation networks, co-ordination of emergency services, etc.) is capable of recommending solutions to the user in the light of his or her emerging aims and objectives, and of warning the user of inconsistencies in their objectives – in a tourist information system the user might, for example, be alerted to the fact that they are trying to book a show at Belfast’s Grand Opera House on the day after their visit to Belfast ends.

However, development of such a dialogue system is a rather more complex business than development of those based on the previous dialogue models: now the developer must consider not only system responses to individual user utterances, but also the potential relationships between the individual discourse segments.

Present challenges

The dialogue management strategies described so far are largely emotion-free. The Queen’s Communicator, for instance, currently has no means of ascertaining the emotional state of the user, nor has it strategies for inducing emotional states in the user. The system cannot use signs of emotion to vary its responses sympathetically; its generic confirmation strategy is a mechanistic one. However, even a frame-based system, given some measure of the user’s emotional state – calm, confused, agitated, etc. – could potentially vary its confirmation strategies – confirming user input less frequently with a relaxed user, providing more explanatory and more frequent confirmations for a confused user, and so on.

Developers in the area of spoken dialogue systems are now beginning to realise the importance of the emotion-aware and affective human computer interface. The animated characters supplied with the CSLU Toolkit, can be given an emotional state (a facial expression selected from a list by the user) to accompany a particular utterance. The team at the Center for Spoken Language Research at the University of Colorado is now taking steps not only to use animated agents that express emotion, but to incorporate face tracking and emotion perception [http://cslr.colorado.edu/beginweb/perceptive_animated_interfaces/perceptive_animated_interfaces.html], to enable, in the long term, much more naturalistic human-computer dialogues.

Nevertheless, emotional awareness remains only one factor in establishing a natural pattern of user input and appropriate system response. The three approaches to dialogue management outlined under State of the Art have their own particular strengths, and might usefully be combined in one system, giving a continuum of strategies, from finite state, through frame-based, to problem-solving agent. A finite state approach might be deployed for finely
controlled, system-led dialogue segments, where clearly defined options need to be presented to the user – perhaps because the system has been having difficulty in understanding the user when the latter has been given the option of less constrained input. Frame-based systems respond to user-supplied information by eliciting further information needed to fill the remaining slots of the current transaction frame. They are thus well suited to routine information gathering roles. However, a purely frame-based approach is limited by the fact that the system lacks knowledge of the user’s overarching objectives: how, for example, will information supplied for one transaction affect the information needed for another transaction; must the user be alerted to any conflicts that might arise and be made aware of possible solutions? Here an intelligent problem-solving agent could usefully establish the context in which routine frame-filling tasks, and routine elicitation of information occur. Signs of emotion – alongside more established metrics like speech recognition confidence [Wessel et al., 98; Evermann & Woodland, 00], percentage ‘no parses’, and the proportion of turns spent by the user on repairing misunderstandings – could play a useful role in determining the most appropriate dialogue strategy. Should a problem-solving agent assist a confused user; or might a ‘limited options’, finite state dialogue manager assist a frustrated user who is having difficulty making him/herself understood? Indeed, would an emotion-sensitive system have the ‘telephone manner’ or ‘presentational style’ that would make the user feel at ease and help establish a productive dialogue?

References

Note: web references are included directly in the text.


9. DESCRIBING THE FEELINGS THAT MUSIC COMMUNICATES: A DIMENSIONAL APPROACH

Edelle McMahon  Roddy Cowie

School of Psychology, Queen’s University, Belfast

Elisabeth André

Institut für Informatik Universität Augsburg, Augsburg

Intuitively it seems obvious that music conveys emotion. Because it conveys emotion, it can influence people, and contribute to communication and persuasion. However, a range of technical problems have to be solved before that potential can be put to use in computer-human interaction. Research in the area exists (e.g., Sloboda 1985, Scherer 1995, Bresin 2004), but it seems fair to say that it remains at an early stage of development.

One of the key problems is how to describe what music conveys emotionally. The traditional ‘basic emotion’ categories are still widely used (e.g. Juslin & Laukka, 2003), but their lack of subtlety is perhaps even more striking in the context of music than elsewhere. This study explores one natural avenue, which is to use representational techniques based on a small number of dimensions to capture the emotional effects of music. An attraction of the approach is that the same techniques have been used to capture the emotional content of both faces and voices (Cowie et al, 2001). Hence they offer the prospect of integrating research on music and emotion into the wider field of emotion research. Another attraction is the suggestion that at least some dimensions may have correlates that can be measured physiologically. That has prompted the idea that the music a person hears could be geared to his or her state using dimension-oriented control systems (Healey, Picard & Dabek 1998, Kim and André 2004).

Modern approaches to dimensional representation originated with Schlossberg (1954), who used statistical techniques to show that most of the emotional content of faces can be captured by considering them as points in a space with two dimensions. There are various ways of naming the dimensions: we have called them activation and evaluation (Cowie et al, 2000). Activation measures how dynamic an emotional state is. For instance, exhilaration involves a very high level of activation, boredom involves a very low one. Evaluation is a global measure of the positive or negative feeling associated with an emotional state. For instance, happiness involves a very positive evaluation, despair involves a very negative one. A third dimension, related to power or control, also emerges from analyses, but it accounts for much less variance than the first two.

The main aim of this paper is to show that dimensional representation based on activation and evaluation can be used to capture broad aspects of the emotional content of music.

The dimensional representations are obtained using a tool called FeelTrace, which has been developed at Queen’s University, Belfast. FeelTrace allows users to describe emotional experiences using a combination of emotion labels and colour which were chosen to give users a continuous input device capable of recording emotional interpretations easily and reliably (Cowie, et. al, 2000).
Two variants of the dimensional representation were considered. The Cartesian variant assumes that the activation and evaluation axes represent psychologically distinct dimensions. The second variant, the circumplex, has been advocated by many authors, notably James Russell (1997). When strong emotions are plotted in activation-evaluation space, they tend to form a circle. From that and other evidence, Russell and others have proposed that the space of emotional states is naturally circular. The centre of the space represents a neutral state, and extreme emotions, such as rage or bliss, are located at the periphery. That suggests an emotion’s position in that space can be represented using polar co-ordinates, with the distance between the central point and a particular emotion’s location representing the intensity of the emotion, and the angle between it and an arbitrarily chosen axis expressing what we have called emotional orientation.

FeelTrace translates those ideas into a computer program that allows users to track the perceived emotional content of a stimulus as it occurs over time. The 2-dimensional ‘activation-evaluation’ space is represented by a circle on a computer screen. It is illustrated in Figure 1. Raters convey their sense of emotional content in real time by moving the cursor around inside the circle. The system automatically records the co-ordinates of the cursor at any time.

![Figure 1 The FeelTrace Circumplex](image)

The system includes additional features which help to stabilise responses. The colour of the cursor is keyed to its position using a method of colour coding introduced by Plutchik, which subjects find intuitive (Cowie et. al, 2000). When the cursor is in a position that indicates very positive emotions, it is green. It is red when in positions that indicate very negative emotions, yellow in positions that indicate active emotions and blue in positions that indicate passive emotions. The coding provides easily-used feedback on the meaning of the current cursor position. In addition, ‘landmark’ words are presented on the screen. Each has an empirically derived co-ordinate which places it on the 2-D space (Cowie et. al, 2000).
Ten excerpts of music were chosen to study, for specific reasons. They were extracts, lasting from 1 to 2.5 minutes, from the following orchestral works:

Simple Cases

1. Gnomus, from *Pictures at an Exhibition*, by Mussorgsky/Ravel.
2. Catacombae, from *Pictures at an Exhibition*, by Mussorgsky/Ravel.
3. Falls, from the sound track of *The Mission*, by Ennio Morricone.
4. Arrival of the Queen of Sheba, by Handel.

These were chosen because they were judged to express a single emotion.

Intermediate Cases

5. Baba Yaga’s Cottage, from *Pictures at an Exhibition*, by Mussorgsky/Ravel.
6. Remorse, from the sound track of *The Mission*, by Ennio Morricone

These were chosen because they were judged to divide into 2 units with different emotional tones.

Complex Cases

7. Great Gate of Kiev, from *Pictures at an Exhibition*, by Mussorgsky/Ravel
9. Moderately Moving, from *Symphony No. 2 in C Minor*, By Bruchner.

These were chosen because they were judged to divide into 3 or more units with different emotional tones. The prediction to be tested was that units judged by the experimenters to differ in emotional tone would receive statistically different FeelTrace ratings.

Ten participants took part in the experiment on a voluntary basis. All were musicians of varying degrees of ability.

Visual inspection of the data was provided by colour graphs of the form illustrated in figures 2 and 3. They use the X-axis to represent time, the Y-axis to represents the intensity of the perceived emotion, and colour to represent emotional orientation.

Figure 2 shows ratings of a single piece (The Great Gate of Kiev) by different raters. It illustrates the core findings, which were borne out statistically using analyses of variance – that there was a good degree of consistency between the participants with respect to emotional orientation; and that different parts of the excerpts were consistently rated differently.
Figure 2 Graphical representation of output for “Great Gate of Kiev”

Figure 3 shows ratings of contrasting pieces, Catacombe (top panel), with a sombre, negative tone; and Finlandia (lower panel) with extreme variations from section to section.

Figure 3 Graphical representation of output for contrasting pieces

The results provide a preliminary indication that dimensional representation provides a reasonably powerful way of expressing the emotional content of music, not only in homogeneous pieces, but also within pieces that change in their emotional tone.

Carrying out the experiments highlights deficiencies in the details of the approach, though. FeelTracing speech, one rarely feels that the absence of a ‘power/control’ dimension is a major problem. FeelTracing ‘The Great Gate of Kiev’, the deficiency is glaring: the sense of power and confidence is at the very core of the music. Owens (2001) has developed a version of FeelTrace which allows users to work in all three dimensions, but it is ergonomically awkward. Also, some very common musical effects cry out for a representation involving two
emotional centres rather than one – typically one centre associated with sadness, the other with intense pleasure (‘sweet sorrow’, in Shakespeare’s phrase).

These observations highlight one of the reasons for thinking that music should be taken very seriously in the context of research on emotion and communication. Music seems somehow to achieve a purity and intensity of emotional expression that language, cluttered with semantics and syntax, rarely does. Hence questions about the nature of the feeling it evokes seem to be sharper, and it may pay dividends to study them before re-engaging with the cluttered world of language.

We do not argue that dimensional representations are necessarily the best way to express emotional content in music. However, we do believe that trying to express that content in a dimensional framework is a revealing exercise, and one well worth pursuing.

References


Sunjung Kim & Elisabeth André, A Generate and Sense Approach to Automated Music Composition, Proceedings of the 9th International Conference on Intelligent User Interfaces, 2004, 268--270, ACM Press, Funchal, Madeira, Portugal

Owens, M. (2001) B Sc Thesis, School of Psychology, Queen’s University, Belfast


10. Experimental issues: emotional expressions, humor and credibility in persuasion.

Shlomo Hareli, Shay Tzafrir and Aaron Ben-Ze’ev

University of Haifa

Definition of the field of investigation and applied prospects

One of the goals of communication is persuasion. Persuasion can be best viewed as the formation, reinforcement or change of other’s attitudes by means of communication [Smith & Mackie, 99]. Persuasion is typically a device employed to change behavior for different purposes such as buying, education and propaganda. Within the realm of automated communication systems the question of persuasion is a central issue for many applications such as applications used for preventive medicine, social action and edutainment. For example, the system STOP is a generation system employed in a real human setting with the aim at inducing the user to stop smoking [Reiter, Robertson & Osman, 03]. One of the main goals of such applications is to achieve effective persuasion. One mean by which persuasion can be achieved is by using emotional communication. There are at least two central ways by which emotion can be involved in persuasive communication. Often persuaders use messages that induce specific types of emotions in perceivers. These emotions are chosen so that they will be conducive to the aims of the persuader. For example, threats of possible illness can be used to induce fear in smokers as means of persuasive attempts to make the receivers of such a message stop smoking. Likewise, guilt or pity arousing messages can promote in people positive attitudes toward donating money for different goals. Although receivers are often persuaded via their own emotions, the emotions expressed by the source of a persuasion message can also affect the extent to which these receivers are persuaded. One way by which that can occur is by the fact that other’s emotions often elicit in perceivers similar emotions, i.e., emotional contagion [Hatfield, Cacioppo, & Rapson, 94]. However, another way by which perceiving other’s emotions can affect persuasion is by enhancing the perceived reliability of the message conveyed. This can increase persuasion because perceived reliability of messages is known to be positively related to persuasion [Hovland, Janis, & Kelley, 53]. Given that this Workpackage has a specific computational aspect and that its aims are to help understand some of the critical issues for building persuasive or otherwise emotion inducing computational systems, the purpose of the proposed project is to suggest a framework for the understanding to role of emotions in persuasion, to study it empirically and make use of the conclusions from this study in the design of persuasive systems.

State of the art

As mentioned above, emotions can be involved in persuasion at least in two different ways. In the following we will review briefly what is known about each of these ways by which emotions are involved in persuasion.

While, past research almost totally ignored the question of how the expression of emotions by the source of a message affects persuasion, much research documented the role that the emotions of the receiver of a persuasive message play in his/her persuasion. In particular, research documented how the elicitation of particular emotions in the receiver can enhance
persuasion and under what conditions such an enhancement is achieved (e.g., Coulter & Pinto, 95; Leventhal, 70). For example, under certain condition fear can be effective in achieving persuasion in contexts such as the prevention of behaviors that put health at risk [Hovland, Janis, & Kelley, 53]. Such messages are most successful if the level of fear elicited by the message is not too strong and the message also contains suggestions concerning how to avoid the threat that led to the fear [Smith & Mackie, 99].

Although, emotions can affect persuasion in such a direct way, emotions are known to be involved in persuasion not only directly by causing an attitude change that is in accord with the nature of the emotion elicited in the receiver, but also indirectly by affecting the way the receiver processes the message. The way emotions potentially affect processing is best captured by the Elaboration Likelihood Model (ELM) suggested by Petty & Cacioppo (86). The ELM model claims that there are two routes to persuasion: the central route and the peripheral route. The central route is most effective when the receiver is motivated to think about the message and has the ability to think about the message thoroughly. If the person cares about the issue and has access to the message with a minimum of distraction, then that person will elaborate on the message. Under such circumstances, if a persuasion attempt uses strong arguments, i.e., arguments that are based on good reasons, such arguments are more likely to lead to persuasion if they elicit in the receiver favorable thoughts about the message. In contrast, weak arguments, i.e., arguments that use less sound reasons, are less likely to lead to persuasion because they are expected mostly to lead to unfavorable thoughts. When the receiver is persuaded via the central route, change of attitude is longer lasting and more immune to further contradicting persuasive attempts. If, however, the receiver is unable and/or unmotivated to process the information contained in the message, persuasion can still occur via the peripheral route were simple cues related to the message or the messenger can affect persuasion. For example, a receiver that invests little effort in thinking about the message may be persuaded by it just because the source is known to be an expert and regardless of the strength of the arguments contained in that message. If persuasion occurs via this route, persuasion is less immune to counterattacks and lasts a shorter period of time.

According to the ELM, any aspect of the communication can serve one or more functions in persuasion. It can be perceived as an argument, and/or it can be perceived as a peripheral cue and/or it can affect the depth of processing of the message by the receiver. For example, a nice looking woman in an advertisement of a health product can serve as a type of argument that is weight as much seriously by the receiver and hence works via the central route. It can also serve as a peripheral cue that elicits positive emotions towards the product due to the pleasant feelings aroused by the mere presence of the attractive women. In such an instance no deep processing is involved. Finally, it can capture the attention of the receiver to the advertisement and cause him or her to process its content more deeply. In considering the question of how emotions can affect persuasion otherwise than by directly being linked with a certain attitude, the ELM suggests the possibility that emotions will affect persuasion by determining the level of processing. In accordance with this claim, research indicates that emotions arising in receivers often affect persuasion indirectly by affecting a receiver’s level of processing of the incoming message. For example, research has shown that anger sometimes reduces the ability of receivers to think carefully about a message. In such a case, the effectiveness of appeals that demand careful processing is undermined whereas the effectiveness of appeals that require little processing is increased [Bodenhausen, Sheppard, & Kramer, 94].

In the following, a research program is suggested. This research is targeted at the filling of some gaps in the relevant literature as described above. In addition, it will also enable a better understanding of how emotions can be used effectively in the design of persuasion systems.
Present challenges

As the above review indicates, a thorough understanding of the role of the expression of emotions in persuasion is missing. Here we suggest a theoretical framework that enables such an understanding and an empirical exploration of this framework. In considering what is the role of emotions expressed by the source of a message in the persuasion of a receiver we assume that such emotions affect the extent to which the receiver finds the message reliable. Thus, in our view, the function of emotional expressions of a source in persuasion is to serve as credibility enhancers (i.e., increase the perceived trustworthiness of the message). This assumption is based on the fact that emotions are commonly regarded as serving a communicative function, both intra-and inter-individually [e.g., Clore, 94; Oatley, 92]. Given that a basic requirement from any communication system is that it must be reliable emotions have to be reliable messages on average in order to survive as communication devices. [Zahavi, 75, 77a, 77b, 97]. In accord with this claim, it is further assumed that emotions are perceived as relatively reliable indications of one’s inner state including attitudes and positions. Hence, in the context of persuasion, communications that include emotional expressions are more persuasive to the extent that the emotion that accompanies that communication is in fit with the content of the message. In contrast, communications that lacks any accompanying emotional expression or communications that are accompanied by emotions that do not fit the content of the message, are perceived as less reliable and hence are less persuasive. For example, a message of a complaint is expected to be seen as more reliable if it is voiced angrily than if it is wrapped by an expression of indifference or a positive emotion. Research conducted in the context of complaints indeed indicates that this is the case [Hareli & Harush, 04]. The assumed connection between reliability and persuasion is in line with research indicating that source credibility is a factor that enhances persuasion [Hovland, Janis, & Kelley, 53]. It is therefore suggested that persuasive agents will achieve higher degrees of persuasion if the transmission of a message will include emotional expressions that are in fit with the content of the message conveyed and to the extent that the expression resembles a genuine expression of the relevant emotion.

We plan to study these assumptions by having participants attend to different recorded communications in which the fit between the content of the message and the accompanying emotion expressed by the source is manipulated. These participants will have then to rate persuasion level and intended behavior. At a later stage, we plan to replicate this study while using artificial persuasion agents designed to convey different messages while emitting different emotional expressions. Participants will have to evaluate these messages, including the extent to which they are persuasive. In this context we suggest the use of Embodied Conversational Agents (ECA). These ECA’s could be set to act according to the emotion-message fitness hypothesis described above. This will enable to test and implement the framework suggested above in the design of effective persuasion agents. For example, the Embodied Conversational Agent, Greta created by Pelachaud, Carogilio, De Carolis, & de Rosis, (02) could be modified to be used for that purpose. Possible applications to be considered in this context are automated selling agents, automated service agents or automated educators, all of which having to emit effective persuasive messages.

As mentioned above, another significant aspect of the involvement of emotions in persuasion is the use of induction of emotions in perceivers as means for enhanced persuasion. To that aim, we plan to focus on humor as an example of an emotion induction system and implement the use of computational humor systems to induce emotions in perceivers of persuasive messages. In the first stage we plan an empirical test of the role that humor plays in persuasion. To date, an understanding of the role of humor in persuasion is missing. In
exploring this issue, we plan an empirical study in which we will test the effect of humorous content incorporated in persuasive messages on persuasion. Here too the ELM serves as a conceptual framework. In line with this model, we assume that humor can have at least three different roles in persuasion. First, humor can serve as a cue to a certain feature of the source that enhances persuasion. For example, a successful joke can cause the source to seem likable or trustworthy and hence persuasive (it could, however, also lead to negative judgments about the source). Under certain conditions humor can be part of the argument. For example, a receiver may think deeply about the humorous part in the message and find it convincing. Finally, humorous messages can either enhance processing by increasing attention to the message or decrease processing by setting a positive atmosphere. Positive mood is known to decrease processing especially when messages are negative [Isen, 87].

At later stages we plan to implement the understanding stemming from this research in the design of persuasion systems that will use humor as means of inducing emotions in receivers of persuasion messages. In the context of the induction of humor as means for the enhancement of persuasion, we plan to use the computational humor system designed by Stock & Strapparava (02).

References


First International Joint Conference on Autonomous Agents & Multi-Agent Systems (AAMAS), Bologna, Italy, July 2002


11. Embodied Conversational Agents for persuasion in different applications

Benoît Morel
Chairman and CEO Cantoche

Introduction

Today companies, universities, but also individuals interact with a computer as it enables access to more information and numerous exchanges despite cultural differences. The man-machine relationship is pushed toward more interactivity and increased personalization as the same way we do in the real life. Computers are taking on a variety of roles as persuaders, including roles of influence that traditionally were filled by teachers, coaches, clergy, therapists, doctors, and salespeople, among others [Foggs,03]. With the arrival of the Internet and new technologies, we are experiencing more and more a real invasion of the virtual into our living habits and this tendency is growing also because the computer is everywhere: Personal and office computer, Personal Digital Assistant, Mobile phone, interactive booth, television, Cars assistant, home automation. Just to give an example of this increasing evolution, the e-commerce represented 95.7 billions of dollars in US in 2003 and expect to grow to 229.8 billions of dollars in 2008 who represents a +140% variation [Forrester Research,03].

Despite this impressive growth, the user is not totally satisfied. Regarding the e-commerce, still in the US, the figures show that the average visitor/buyer is only 3.4% [Forrester Research,02] and about 50% of the buyers stop their transactions on the web and this proportion doesn’t reduce: 45.9 % in Q2 2003 and 57% in Q2 2004 [Double Click, 04].

These figures are disappointing and show a real issue in the Man-Machine interaction comparing with the real life. The reasons are numerous: The most frequently cited barriers to online shopping include product pricing (cited by 77 percent of individuals), potential return hassles (67 percent), concerns about credit card security (65 percent), and personal privacy issues such as worries about junk e-mail (58 percent). Difficulties in navigating a merchant's site were cited as a barrier by 35 percent of individuals who have yet to make an online purchase, and 48 percent of those who have purchased something online. One-in-four buyers also complained about the amount of time it took to receive their orders [Intermarket Group,00] and the online shopper needs to be very persuasive to satisfy his client and convert him in a real buyer.

The analogy with the real life is interesting. The way people interact with technology is the way they interact with each other [Picard, 04]. The user, in front of the computer, reacts like he does with a human. If we stay with the e-commerce example, the buyer enters a shop with the purpose to satisfy her needs or simply her curiosity. A vendor plays the role of companion: he will analyse her needs, interpret them and orient the client in the shop, reassuring and consulting her as well as satisfying her needs. The goal of the shop owner is to sell and to establish a relationship of trust between its clients and its brand. A real exchange is established in order to come close to the client’s expectations. Good expertise and adaptation to the client will increase satisfaction both of vendor and of buyer [Morel,04]. Trustworthiness
is a key factor in this interaction. So we can think that computing technology that is viewed as trustworthy (truthful, fair and unbiased) will have increased powers of persuasion [Fogg, 03].

The e-commerce is not the only application where persuasion is necessary and obvious to satisfy its goal and, if trustworthy is a key factor for e-commerce applications, other persuasive factors are more benefits for others. Virtually, we will establish relationships with emissaries of commercial sites, vendors in online shops, professors in e-Learning environments, nurses and doctors on health sites, or just off-line office companions, speakers for institutions or communities that want to get into contact with us.

**Persuasive ECAs**

There are several ways to have persuasive technology and Cantoche works on embodied conversational agents (ECAs) in order to make the relationship with the user more friendly and to motivate him by creating emotions. Characters can express social roles, emotions, and organized personalities that match learning goals, company brands, and transaction needs. Characters can increase the trust users place in online experiences, in part because they make online experiences easier [Reeves,04]. The most important attribute of virtual agent technology is its ability to conduct dialog with the user, determine a user’s needs, and provide an appropriate response, much in the same manner that a real person does [Hickernell,002]. Everything in an ECA can help to be more persuasive on a Man-Machine relationship:

The first thing the user sees is the graphical representation of the character: The first approach is to work on cartoon characters that facilitate the establishment of a playful relationship with the user. Cartoon representations allow us to "play" with the form and the proportions of the characters. According to the target audience, a character like Mickey can be made an accomplice with soft rounded forms, or it can appear more aggressive like Pokemon [Montigneaux, 02]. The second approach is to create a realistic or semi-realistic character. The more realistic appearance avoids any possible ambiguity and allows for professional discourse.

Beside the graphical appearance, the behavior of the character will give it a strong personality. Always remember that character is all that matters in the making of great comedians, in animation, and in live action [Jones, 90] and like actors, animators communicate through the language of the movements [Lord, 98].

The voice associated with the ECA is an important factor too. Some voices are recognizable like Bugs Bunny voice but also the tone of the voice allows creating an emotion. Sometimes, it’s impossible to add a voice and so a dialog balloon is a good device to supply the sound. Like in the comics, we can create emotions by just changing the shape of this balloon.

Finally, the verbal dialogue of the character allows the user to have a full emotional interaction. Instead of dialogue, David Freeman, a famous game writer and designer, created a new interesting world “EmotioneeringTM”. Emotioneering entails much more than great dialogue and its goal is to move the user through an interlocking sequence of emotional experiences. Emotioneering is the vast body of techniques that can create, for a player or participant, a breadth and depth of emotions in a game or other interactive experience, or that can immerse a game player or interactive participant in a world or a role [Freeman, 04].
How an ECA can be persuasive depending of the applications

The e-commerce we described in the beginning of this document is not the only market where persuasion using ECAs is necessary. Virtual characters, that can conduct emotioneering and push interactivity with the user, benefit to other areas as well. Here are examples of the main applications where the ECAs market exists and few directions of how they can be persuasive regarding the application fields:

Sales/Marketing : By 2005, major vendors of CRM suites will either acquire virtual agent technology or partner with best-in-class vendors to enhance their collaborative CRM capabilities across sales, marketing, and service products [Hickernell, 02]. AVIS has launched in early 2002 a website dedicated to its partners who wished to offer on-line car-booking services. « Lola has been hired to create an interpersonal relationship between the user and the site, just like what the relationship is between a customer in the real world and an agent in any AVIS booking travel agency. Lola has been trained to fulfil this assistance's mission: she assists users as soon as they need help, and she is available at any time, 24 hours per day! », says Olivier Sellem, Avis France's distribution and e-commerce manager. Lola helps the user filling a form in order to receive a quote. Before each request is processed by Avis’ Central Bookings system, Lola checks the user’s information in the form, and informs him if some fields need to be re-entered. In this case, she gives many advices to make sure the user will get the best conditions for his quote by suggesting a better car for example. Once the users’ quote is returned, Lola explains what options are offered for the payment and invites the user to enter his name and address in the final reservation form.

Support and assistance: It’s one of the most obvious use for an ECA where the character needs to be very helpful. Laura was hired by EDF for a long-term contract in December 2002 in order to give advice and provide suggestions for maximizing home comfort though effective use of electricity on the site www.maison.edf.fr. The EDF team trained Laura to accompany and assist Internet users in their visit to a virtual home. With each click on one of the elements of the environment (water heater, hi-fi system, radiator, TV, etc.) Laura gives advice on how to consume electricity more intelligently. Laura uses a lot of different ways to be persuasive like changing clothes depending of the room she is, playing quiz with the user,... Laura makes learning fun for people of all ages!

External communication: Organisations with existing personalities embedded in their brands should consider extending branded personalities to Virtual Agents to exploit brand affinity among customers. Even if an organisation doesn’t have its own branded personality, it’s possible to communicate by its representation. This is illustrated by the case of Packard Bell who, in February 2002, launched their new line of computers, namely Internet Dre@m M@chine, together with Netissimo and Intel. For promotion, they wrote an interactive script played by Capt’n Surf, a sort of superhero who praises the power of the machine. The superhero design of the character facilitated the transmission of the Packard Bell message and added to be persuasive.

E-Learning: Virtual instructors have the role to bridge the gap between asynchronous and synchronous) learning systems. The goal is not just to provide applications using just a cyber teacher but several studies show that that several characters can be more persuasive. Software can help students learn better when a virtual teacher is accompanied by a virtual student. That way, the teacher can occasionally direct questions to the virtual student, and the real student does not feel picked on all the time. And the virtual student creates an illusion of a classroom setting, in which the real student can receive praise from both the teacher and a virtual peer [Nass, 04].
Internal communication: That’s the main market Cantoche actually develops. Customers see with the avatar a very good cyber employee in order to motivate the real ones. The company Valeo hired the character Kiny to persuade the employees to use the new Intranet web site. Kiny is integrated in the computers and appears to promote a new Intranet features. Kiny uses several gestures and an adaptive dialog with humor in order to create user’s emotions. Once the user opens the Intranet site, Kiny serves the role of assistant (tour guide portal, contextual help at the user request,..) and messenger (broadcasts messages and announcements). Kiny speaks several languages depending of the user.

Present challenges

As described above, an ECA (or more) can create a lot of emotions for the users and can be very persuasive for the user depending of the application. In this WP, Cantoche wants to continue working with the Humaine group by experienced different ways of persuasions depending of the situation. We have entered an era of persuasive technology, of interactive computing systems designed to change people’s attitudes and behaviors. [Fogg,03] but we need to find the best way to do it and avoid the failures. For instance, the interaction with a supposedly realistic humanoid character still leaves the human viewer with a more eerie and disturbing impression than a corresponding interaction with an intentionally non-realistic character [Picard,01]. An animated paperclip winking at you every time you click on it to go away is analogue at a person who insists on winking at you every time you ask them to leave your office [Picard,01]. Persuasion can be on different ways but needs to be used with care. The risk is to give the user the impression to be manipulated and it’s not the goal.

References

[Freeman, 04] D. Freeman, Creating emotion in games, New Riders, Indianapolis. 2004


[Hickernell, 02] T. Hickernell, Intelligent self-service agents are “virtually” there, Media Group. 2002


[Reeves, 04] B. Reeves, *The benefits of Interactive Online Characters*, Stanford University. 2004
12. Applications of emotion-oriented systems: taking a long look

Roddy Cowie
Queen's University, Belfast, UK

A core feature of HUMAINE is that it takes a strategic approach to the development of emotion-oriented systems. In that respect, it contrasts with the more tactical approach that predominates in related areas. The reason for emphasising strategy is that dealing effectively with emotion seems likely to depend of solving a wide range of problems, at very different levels, in ways that cohere.

Part of a strategic approach should be to take a structured view of potential applications. Existing reviews have offered lists of rather specific applications [1]. This paper offers an attempt to do something more general, and to group applications into broad types. The scheme was introduced in a recent talk [2]. It divides application types into two broad groups of three, corresponding to relatively short and relatively long term types of goal. The cutoff is associated with a watershed, which is both the single most important application and a key to other possibilities.

One of the results is to highlight questions about research in the particular area of communication and persuasion. In many ways, the concepts do not sit easily within a structure that deals reasonably well with other areas covered in HUMAINE. But as often happens, engaging with the difficulty produces interesting ideas.

Types of application that are attainable in the short term

It is difficult to carry out technological research without having in mind some applications that seem to be attainable in the reasonably near future – for the sake of argument, five to ten years. For emotion-oriented technology, there seem to be three broad types of application in that category.

Trouble shooting

At present, probably the most active application area is detecting troublesome emotions - in callers using automatic exchanges, pilots, drivers, etc. [3],[4],[5],[6]. These applications have clear links to older research areas concerned with the detection of stress [7] and lying [8]. Many extensions could be imagined, e.g. detecting emotions that are not so much troublesome as short of a relevant ideal – for instance, picking out staff at a counter or in an exchange who fail to show sufficient warmth towards customers.

Trouble-spotting applications are an obvious target early in the development of emotion-oriented systems, partly because it looks as if rather simple systems could achieve useful results, and partly because it fits a long-established stereotype that considers emotions primarily in a negative way – as departures from the natural human ideal of rationality. However, the area is fraught with problems, some ethical, some to do with doubts about the level of performance that simple systems can actually achieve. These are explored at more length in HUMAINE deliverable D10a [9].
Affective selection

Affective selection involves detecting emotion-related responses to various stimuli, and using the responses to make choices that reflect the user’s preferences. For instance, Kim and Andre proposed adjusting ambient music to suit the listener’s mood [10], and Aizawa [11] proposed to archive video footage of times when brain signals indicate strong emotions. There are many natural extensions, such as adjusting colouring to suit a user’s mood, or our own proposal to use emotion in locating the types of holiday that a user might favour.

Affective loops

Höök [12] coined this term for a type of technology that goes back at least as far as the drum. It is illustrated in the figure below. The user acts; the system reacts; the result affects the user emotionally, and encourages the user to take actions that continue the cycle.

A key kind of affective loop occurs in conversation. Not all conversations act as positive feedback loops for emotion, in fact many damp it very effectively. But some types of conversation are highly effective at promoting relatively uninhibited emotionality. It is an interesting theoretical challenge to understand what they are.

The watershed: Really natural language processing

‘Really natural language processing’ seems as good a phrase as any to describe something that contemporary language technology does not allow. It is what system achieves when it allows people to interact with it by talking as they would talk to another person, and gives responses that they can assimilate in the way they would assimilate another person’s responses. At present, the only way for a person to access the information and capacities encapsulated in a computer system is to adopt a mode of communication which is very far from that – one that is engineered to suit the computer’s very limited communicative skills. That constitutes an enormous bottleneck in people’s access to the potential of computing. Access would be transformed if people could use really natural language to communicate with computers as well as with each other.

It seems very likely that emotion is one of the areas where progress is needed to achieve that. Emotional colouring is an integral part of person-to-person exchanges, and there seem to be rules governing the emotional aspects of interaction much as there are rules for its content
match the general emotional tone of the other speaker (emotional convergence)

pick up topics that interest the other speaker

avoid topics or styles of speech that cause the other distress or boredom

repair ill feeling

As in related areas, the rules define default behaviours, which are broadly speaking co-operative. Violating them does not produce an ill-formed formula: instead it signals that the agent is choosing not to adopt the default co-operative attitude. That is a natural reading of the way people react to other people who are for some reason unable to follow the rules, such as the deaf [17], and it explains why people with disabilities that affect speech production place such value on being enabled to give it appropriate emotional colouring [18].

That analysis fits informal evidence that people react badly to systems which are sophisticated about other rules of communication, but ignore the emotional ones. The violation means that the systems are felt – at a level reason cannot easily over-ride – to be displaying an antagonistic or unpleasant emotional stance. It is unlikely that speech interfaces subject to that kind of problem will be widely accepted – and hence, achieving a level of emotion-sensitivity that allows them to be observed is integral to achieving really natural language communication.

The proposition so far could be summed up ‘without emotional sophistication, no really natural language processing’. However, there is an obvious converse: ‘without really natural language processing, no emotional sophistication’. Language is not the only channel by which human beings express their emotional orientations, but it is a key channel. The point can be obscured if we think of emotion as primarily a matter of a global state that envelops a person at certain times. Perhaps that kind of state can be detected reasonably well using only non-verbal cues. But emotion is usually intimately bound up with particular people or things or events or courses of action. Genuine emotional sophistication depends on being able to discern what an emotion is bound to in the person’s environment or mindscape, and that is very difficult without well-developed linguistic communication. An interesting parallel is that a pet may well register its owner’s nervousness, but not (despite some owners’ claims) the root of it, and hence the things that could be done to help.

For these reasons, it seems reasonable to see the pairing of linguistic and emotional sophistication as a key goal for information technology as a whole – and emotion-oriented computing as something much more than a minority interest aimed at a few esoteric applications.

There are applications that depend on combined linguistic and emotional sophistication. The next section takes up that issue. However, there is another practical reason for companies to be interested. Suppose that two systems carry out essentially the same function, but in one the communication involved is easy and pleasant, and in the other it requires special skills, and a degree of emotional toughness. The company that markets or uses the second is unlikely to continue long in business.
Types of applications that are primarily long term prospects

This section looks forward to applications that depend on combined linguistic and emotional sophistication, and which will therefore at least not mature until that combination has been achieved.

Many of the applications that open up may depend on naturalness rather than any conceptual link to emotion. Obvious examples in that category are systems that provide access for people who find it difficult to adapt their communicative styles in the way that contemporary systems require them to, because of age, intellectual limitations, or the fact that they are in difficult environments. However, there are also major areas where emotion as such is at least implicated.

Uncovering feelings

This term refers to something that is currently done – often crudely – by human interviewers, for a variety of purposes: teasing out the systems of values and dispositions that surround some kind of person or object or event in a user’s mind. That is very different from attaching a label to a brief emotional state, and to do it well needs sophistication about language as well as about emotion. It is an essential key to a range of services – market or political research, careers advice, politics, non-directive counselling, personalised entertainment, etc.

Facilitating learning

This is what good teachers do – not just presenting information to a learner, but taking account of the emotional issues – boredom, excitement, pride, humiliation – that make learning likely to succeed or fail. It probably depends on a fair degree of ability to uncover feelings unobtrusively. It is not confined to the classroom – it applies to manuals as much as multiplication tables.

Moulding

The term is meant to convey that the system sets out to change users’ outlook or values or priorities rather than simply to extend their knowledge. It is separated from facilitating learning mainly because the two are very different ethically: moulding has obvious attractions for sales or politics, but it raises major ethical concerns. Emotionally sophisticated persuaders with no conscience, but infinite patience, are a nightmare.

Three implications of this overview should be noted. First, not many of the areas are the sole preserve of emotion oriented technology. More often, its function is to provide techniques that enhance systems with many other elements. Second, not many of the applications centre on ‘episodic’ emotions, that is, relatively short lived states where emotion dominates the subject’s experience. Those that do are the short term ones. The key issue in the longer term is the kind of emotional colouring that pervades mental life, even when we are being thoroughly rational. Third, generally speaking, the point of introducing emotion is to rehumanise functions that for various reasons we either have handed over to machines, or might want to hand over. The level of emotional competence that systems need to show for that purpose is not dramatic, and it probably does not need to be present all the time. Nevertheless, it seems a safe bet that systems which are able to recognise and use relevant emotional colouring will obliterately systems that are not.
How do communication and persuasion fit in?

The terms communication and persuasion do not correspond to individual items in the sequence outlined above. Broadly speaking, they refer to goals that can be achieved to various levels with various kinds of resource. Trying to locate these goals within a cohesive framework raises interesting issues.

Communication is effective exchange of information. It is linked to emotion because, as noted above, effective exchanges of information involving humans are very likely to depend on taking proper account of emotions – just as really natural language processing does. It would not be unreasonable to subsume all but the short term applications above under the heading ‘emotion-dependent communication’. That reflects the point made at the end of the last section, that long-term applications are likely to involve emotion as a means to an end which is not specifically emotion-related, in conjunction with techniques that are not specifically emotion-dependent.

What is distinctive about the term communication is that it invites a perspective concerned with the transmission of information rather than the modality. Taking that perspective makes it clear that there are multiple ways of taking into account the emotional nature of human communicators. Many of these could be practically available before really natural language processing becomes a full-fledged reality, and that suggests an interesting picture of the approach to the long-term ideal of communication in a really natural language context.

A basic emotional requirement for effective exchange of information is that the person must engage. The affective loop conception captures various ways of achieving that which is already widely used, for instance in interactive museums. Developing the strategy is therefore an approach with immediate practical potential. However, it can also be seen as preparatory to a longer term objective, which is understanding how to engage users in conversation when the necessary speech and linguistic technologies become available.

Really natural language processing requires both speech and linguistic technologies, but for historical reasons, the research tradition associated with the term ‘communication’ has dealt in more depth with the linguistic component [Berger & Br]. That research tradition highlights a range of ways in which language might be used to take account of the emotional issues in communication. Many of these have the potential to be used before the difficult issues surrounding speech are resolved satisfactorily.

Another strand of research on communication is its awareness of factors that lie outside the realm of really natural language processing. Many contextual factors can make people more or less receptive to communication. Music and visual pattern are the easiest to manipulate, and again, they may offer opportunities to make progress in the relatively short term. In the longer term, they are a subset of a domain that is no less complex than really natural language processing – that is, environmental influence on communication. Everyone knows that certain things are easier to communicate in an atmosphere of candlelight and soft music. It is not so clear how that element of human communication could be transferred to interface design.

In sum, focusing on communication can be seen partly as focusing on certain ways to approach goals that are shared throughout emotion-oriented computing; and partly as acknowledging that the long term future of emotion-related techniques is facilitate achieving goals whose character is not specifically emotional.
On the surface, persuasion is a very different matter. Persuasion is leading a person to a position that he or she would not have chosen at the beginning of the process. It is intimately bound up with emotion, because the most cost-effective techniques of persuasion involve bypassing argument by activating processes rooted in emotion. People can be persuaded by a single photograph that appeals to emotion where a philosopher’s most cogent argument would have precisely no effect.

Unfortunately, in reality, there are no clear lines. Communication depends on inducing some kind of receptivity. To call an interaction persuasion rather than communication is to say that the process of inducing receptivity crosses a moral line.

It is a vexed issue how artificial systems should position themselves with respect to that line. Commercial interests would certainly pay to have systems that knew how to overcome clients’ unreasonable opposition to buying their bargain holidays, PC’s, fashion t-shirts, etc.. That may sound obviously dangerous, but how to distinguish it from systems which increase the likelihood that people will do things that are entirely reasonable and beneficial, such as encouraging HUMAINE partners (on behalf of the co-ordinator) to ensure that deliverables are submitted on time?

At that point, application takes an unlikely turn. Philosophy is not known as an applied discipline. Nevertheless, it is hard to avoid the conclusion that philosophical scrutiny would have to be part of a any commercialisation of emotion-oriented technology where persuasion was involved. It is quite pleasing to imagine a brass plate in a prestigious district of London declaring bearing the inscription

Socrates, Kant and Wittgenstein Inc: Consulting philosophers.

References


[2] Cowie R Piecing together the emotion jigsaw Invited paper to MLMI04, Martigny, June 2004 (see www.idiap.ch/events/workshop-mlmi04/)


