Emotion in Human-Agent Interfaces

Peter Wallis, Roger Moore, Petra Fagerberg, Marc Cavazza and Yorick Wilks

COMPANIONS Consortium: State Of The Art Papers 3, January 2006

Recent years have seen an upsurge of interest in the implications and applications of emotional behaviour in the human-computer interface (Picard 1997). In particular, research has focused on interactive systems involving visual and/or vocal modes of communication, where the user and/or the system may exhibit both positive and negative expressive behaviour in the course of an ongoing dialogue. Such behaviour can lend naturalness to an otherwise artificial situation, but it can also be exploited to enhance the communicative effectiveness of a human-agent interface; for example, a system may detect the emotional state of a user (from the look on their face, or from the tone of their voice) and act accordingly, or it may signal a problem by adopting suitable emotional behaviour itself (by generating appropriate facial and/or vocal expressions).

Current research in this area revolves around virtual manifestations of human-agent interfaces known as 'Embodied Conversational Agents' (ECAs): speech-enabled animated characters that can act for example as a receptionist, a tour guide or a personal tutor (Cassell et al 2000, Pelachaud and Poggi 2001, Beskow et al 2005). The ability of an ECA to detect and exhibit emotional behaviour is believed to be an important requirement for achieving effective and naturalistic human-computer interaction (Bates 1994).

Emotion and Behaviour

The formal study of emotion in human (and animal) behaviour has a long history, from the early observational work of Charles Darwin (1872) up to the recent emergence of 'Affective Science' (Davidson et al 2003) as a cohesive discipline. Over that period, three main categories of psychological model of human emotion have emerged.

The earliest 'discrete' theories of emotion (stemming from Darwin's work) hypothesised the existence of a small number of basic emotions, such as happiness, sadness, fear, anger, surprise and disgust (Ekman 1999). In such theories, it is supposed that these emotions are based on specific physiological response patterns to external stimuli.

Another early model of emotion is the 'dimensional' approach (Wundt 1874) in which a wide variety of emotions are mapped into a low-dimensional space. For example, the coordinates of a hypothesised two-dimensional representation of emotion would reflect subjective aspects of behaviour such as feelings of positivity vs. negativity and active vs. passive. Cowie et al (2001) use such a scheme as the basis for FEELTRACE, a computer-based tool for emotion annotation.

The third, and most recent, theoretical view of emotion is the 'componential' model. This approach emphasises the variability of different affective states, and links the production of an emotion to the appraisal of a situation with respect to an organism's needs and goals (Scherer 2001).

Both dimensional and appraisal-based approaches have been sources of inspiration for affective interface systems. The former have been generally used to augment traditional user interfaces, while the latter have been used to define agent's emotion in terms of their internal cognitive mechanisms (e.g. anticipation of plan success or failure, see (Gratch, 1999)).

Previous Relevant EU-funded initiatives

SAFIRA (IST-1999-11683) was one of the first EU-funded projects specifically dedicated to affective computing. Its objectives were to create a framework to enrich interactions and applications with an affective dimension. It planned to implement a toolkit for affective computing combining a set of components addressing affective knowledge acquisition, representation, reasoning, planning, communication and expression. Another objective of the project was to verify under which conditions the hypothesis that emotion, as well as other affective phenomena, contributes to improve rationality and general intelligent behaviour of the synthetic characters, thus leading to more believable interactions between humans and computers. The project has released a toolkit and has generated a number of demonstrators: "FantasyA", a novel kind of affective computer game, the "Influencing Machine", an affective ambient intelligence system and "James the Butler", an affective ECA. James is a Personal Sales Assistant (PSA) that employs intelligence and adaptive reasoning methods to provide active, collaborative assistance to a customer of an online store. James is an agent that amplifies or modifies the

motivational state of an agent and its perceived bodily state. It has the ability to perceive and produce the visual (animated expressions), verbal and non-verbal signals and regulate the flow of information between service agents, the interface agent and the user. These capabilities enable James to engage in complex interactions with customers via natural social communication rather than complex command languages, or direct manipulations.

NECA (IST-2000-28580) was dedicated to multi-modal communication with animated synthetic personalities. A particular focus in the project lies on communication between animated characters that exhibit credible personality traits and affective behavior. The key challenge of the project is the fruitful combination of different research strands including situation-based generation of natural language and speech, semiotics of non-verbal expression in situated social communication, and the modelling of emotions and personality. The project developed several demonstrators, among which the eShowroom, that featured ECA as car sales assistants. In this demonstrator, personality profiles could be defined for agents that determined inter-agent dialogues.

MAGICSTER (IST-1999-29078) was concerned with the development of believable conversational interface agents which make use of gaze, facial expression, gesture and body posture as well as speech in a synchronised fashion. The project also intends to evaluate the use of conversational agents in laboratory conditions to determine which aspects of the embodied agent are relevant to various situations of human-computer interaction. Finally, the project aimed to develop and document the agent architecture and components to enable other research and development teams to prototype and evaluate new versions of the agent interface in new domains and for novel tasks. MAGICSTER contributed important developments to the field of affective ECA, such as the APML markup language. The project produced several demonstrators of emotional dialogues with ECA (Cavalluzzi et al., 2003) some integrating state-of-the-art ECA (the "Greta" agent) with state-of-the-art dialogue toolkits (the TRINDI system). The emotional model was derived from the OCC approach.

PF-STAR (IST-2001-37599) addressed three crucial areas of user interfaces: technologies for speech-to-speech translation, the detection and expressions of emotional states, and core speech technologies for children. The project has produced a variety of results in the field of affective dialogue; on the relation between users' emotional states and dialogue progression (Batliner et al., 2004) and on how emotions influence articulatory patterns with application to MPEG-4 encoding of talking faces (Beskow, 2004). The system has produced several prototypes and toolkits to develop talking heads supporting emotional display (Cosi et al., 2004).

VICTEC (IST-2001-33310) developed ECA as part of a tutoring system educating children on bullying issues. It is based on the generation of empathy towards virtual actors being bullied by fellow children as part of an interactive narrative. The FearNot! System developed as part of the project includes NL communication features supporting the recognition of speech acts to give advice to the virtual actors. The prototype has been used to stage real-world evaluations in schools.

NICE (IST-2001-35293) was a Human Language Technology project developing Multimodal conversation with virtual characters in an edutainment context. It was dedicated to children and teenagers, introducing them to a virtual character of H.C. Andersen. NICE implemented Multimodal dialogue based on speech and gesture recognition.

HUMAINE (IST-2004-507422). The rapid growth of academic research in affective computing has resulted in the formation of an EU-funded network of excellence: 'Human-Machine Interaction Network on Emotion' (HUMAINE) funded under the Framework 6 IST Programme. Started in 2004, HUMAINE involves around thirty different laboratories and incorporates research across a wide number of disciplines and applications. The project's research is organised into six thematic areas:

- theories of emotion
- from physical signals to emotionally significant features and vice versa
- patterns of signs that convey emotion in interactions
- functions of emotion-related elements in communication and persuasion
- emotion in cognition and action
- usability of emotion-oriented systems

HUMAINE aims to contribute to "the development of systems that can register, model and/or influence human emotional and emotion-related states and processes".
Computers obviously do not experience emotion in the same way as people, but less obviously perhaps,

emotion plays an essential role in the everyday doings of human beings. At first blush, the emotions that come to mind are those Cowie and Schröder (2005) call **episodic emotions** and include fear, happiness, anger and so on. The categories of emotion of primary relevance to our day to day doings are however the **pervasive emotions**. These include moods (cheerful, gloomy, irritable, listless, depressed, buoyant) Interpersonal stances (distant, cold, warm, supportive, contemptuous) attitudes (liking, loving, hating, valuing, desiring) and affect dispositions (nervous, anxious, reckless, morose, hostile). These emotions are pervasive in the sense that the decisions we make every minute are influenced by some underlying mechanism that surfaces in behaviour that is described in these terms.
Cowie and Schröder talk of the watershed issue for HUMAINE being an understanding of 'really natural language processing' and emphasise the fact that a conversation with a machine is an affective loop for which, like playing a drum (to use Höok's example) the user behaviour should be reflected in the user experience. Direct applications of an understanding of emotion in dialog would include things like trouble shooting in automated call handling systems - if the agent could recognise when the caller has (or was about to have) a negative experience, what conversational strategies would relieve or avoid the situation? This issue has been addressed directly (see Brahnam (2005) for example) but is there some underlying explanation for why and when we experience emotion, and why and when we expect others - including synthetic characters and ECA - to experience emotion?

Emotion and Interaction

As outlined above, the HUMAINE umbrella includes many different approaches to studying emotion, some of which are based on a strategy that has proved so successful in computational linguistics over the last decade. By tagging a corpus of speech or text - or in the case of ECA, of video - with markers indicating the emotion being expressed at that point, machine learning techniques or statistical analysis can, it is felt, tell us something important about the nature of emotion in human interaction. Marking up corpora for emotion is a much more dubious task than marking up for part-of-speech, and its consistency and value has not yet been shown. .
However, significant activity has been dedicated to emotion markup langages and other a priori definitions of emotional categories that would be used to control the animation of virtual agents

Mel Slater and colleagues (2000) have designed a system, Acting in Virtual Reality, where the user expresses her emotions by changing the characteristics of a drawn face. The user can influence the eye brows and the mouth of that face. The mouth can express happy, neutral and sad, while the eye brows can express surprised, neutral and angry. By interacting with both the mouth and the eye brows at the same time the user can create more complex expressions. It is also possible for the user to affect some body parts of her avatar. The emotions are carried out by the user's avatar in a virtual rehearsal system. The system was set up to be used by actors to see if they could rehearse a play in virtual reality that in the end was going to be performed on a real stage. The actors who were not previously familiar with each other met in the virtual reality four times for a one hour rehearsal each time. Then they met a fifth time for a live performance in front of an audience. Even though the actors did not think the system could replace real rehearsal they all learnt to master the program and to use its qualities. One of the actors compared it to talking on the telephone which is not like a real life meeting but still very effective and interesting.
ExMS is a system where the user explicitly states her emotions (Persson 2003). ExMS is an avatar-based messaging system where users can create short pieces of animated film to send to each other. The idea is that each user chooses an avatar that she can identify herself with and by using the library of animated expressions specific to her character she can express feelings, reactions and moods in the messages she sends to her friends. One disadvantage with ExMS was that the avatars had so much character in themselves so that it sometimes was hard for people to see themselves and their own expressions represented through their avatars.
Another communicational system where users explicitly state the emotion they want to communicate is used in CHATAKO, a speech synthesis system developed to assist people with communication problems (Iida et al. 2000). In CHATAKO the user writes what she wants to say and then chooses if she wants to say it with a female or a male voice and what emotional value she wants that voice to have. The prototype has three emotions to choose from; joy, anger and sadness.
These two systems are examples of applications that support emotional communication. The first two are creative and fun and CHATAKO is an important solution for people with speech problems. However, they do not fulfill the physicality and ambiguity of an affective loop that we want to create. Even if the expressivity in Acting in Virtual Realty and CHATAKO was to be extended on there were other problems with personality and open interpretation experienced in ExMS where there were more expressions to choose from.

The principal applications making use of emotional ECA consist of Information Access (e.g. user emotion recognition in SMARTKOM, (Streit et al., 2004)), Training and Tutoring systems, as well as entertainment systems (i.e. interactive storytelling featuring ECA as virtual actors).
We can identify from the state-of-the-art a set of important research problems that characterise ECA-based affective interfaces. These problems can be listed as:

- *Empathy* is an important concept to characterize the quality of the relation established between users and ECA. The benefits of Empathy in communication have been reviewed by (Martinovski et al., 2005) who also studied the linguistic manifestations of empathy (or its rejection) in dialogue. It should also be noted that depending on the application, empathy can be exerted by the ECA towards the human (Prendinger et al., 2004) or by the human towards the ECA (as in the FearNot! System introduced above).
- *Politeness* is a recurring topic in ECA. Politeness can be an important part of a information access system or a tutoring system, although in some cases the tutoring system precisely consists in acquiring elements of politeness and proper behaviour, for instance in another language and culture, as in the Tactical Language Training System developed at the University of Southern California. It should be noted that both sides of politeness are explored: agents politeness (Johnson et al., 2004) and users politeness (or rudeness, see (Kopp et al., 2005)).
- *Humour*. The importance of humour for ECA has been discussed as early as 2002 by Nihjolt (2002), who related its implementation to the use of emotional modelling (e.g. using the OCC model). ECA typically involve multimodal humour through co-ordinated non-verbal behaviour and humorous language generation. Humour is however not limited to "jokes"; in some cases humour can also be represented by witticisms and punchlines, for instance within an interactive storytelling context featuring virtual actors (Cavazza et al., 2004) (Cavazza and Charles, 2005).
- *Naturalness*, which can be defined as the endeavour to provide believable communication rather than "system prompts" uttered by an agent, tends to be taken for granted in recently developed systems and to disappear behind more specific relational modes, such as those listed above.

.

Emotion in Language and Speech

According to Scherer (2003), systematic research into the effect of emotion on the voice started in the 1960s when psychiatrists took an interest in diagnosing affective states from vocal expressions. Since then, psychologists, linguists, phoneticians and engineers have also become involved, culminating in 2000 with an international workshop on 'Voice and Emotion' and a subsequent special issue of the journal Speech Communication (Douglas-Cowie et al 2003).
One of the first demonstrations of adding emotion to synthetic speech resulted from the work conducted by Ian Murray at Dundee University (Murray and Arnott 1993). Murray developed a system called 'HAMLET' which used rules to alter voice, pitch and timing of the commercially available DECTalk speech synthesiser. DECTalk is an example of a 'formant-based' text-to-speech system which has the characteristic that it provides parametric control over the content and characteristics of the output voice, hence it is easy to manipulate in order to introduce expressive behaviour. Similar work took place at MIT (Cahn 1990) and other early work combined the generation of both vocal and facial expression (Henton and Litwinowicz 1994). In his review of emotional speech synthesis, Schröder (2001) concluded that it was not yet usable in many real life applications due to the restriction of using a few basic emotions, the simplicity of the models for intonation and timing and the lack of naturalness of formant-based speech synthesisers.
Recent years have seen a significant upsurge of interest in the recognition of emotion in speech, particularly for applications such as call-monitoring for 'Interactive Voice `Response' (IVR) systems, e.g. to detect a complaint. As in speech synthesis, much of the relevant research is directed towards the identification of a few basic emotions such as happiness, anger and sadness. Such investigations fall under the wider heading of the recognition of 'Speech under Stress' (Moore 1996); a broad area of research that encompasses the full range of external and internal 'stressors' that can condition the speech production process (for example, cognitive load, physical and mechanical stressors, physiological constraints, psychological stress and emotion). Interest has also been fuelled by recent coverage in the popular press of how financial institutions such as insurance companies are beginning to use so-called 'Voice Stress Analysis' (VSA) on incoming telephone calls in order to detect (and deter) deception - although it must be stated that scientific opinion is divided about the effectiveness of such devices (Haddad et al 2002).
Many studies on the recognition of emotion in speech have investigated the use of prosodic features, i.e. the influence of emotion on the patterning of the pitch, intonation and timing in speech signals. More

general stress analysis techniques also use spectral features, and much experimental work has been done using a variety of different pattern classifiers such as linear discriminant analysis (LDA), artificial neural networks (ANNs) or Gaussian mixture models (GMMs). In general, such research has shown that, given sufficient training data, it is possible to classify the emotional content of speech with a reasonable degree of accuracy, e.g. ~65% for four basic emotions (Yildirim et al 2004). In a recent study, Rigoll et al (2005) demonstrated an ability to classify utterances taken from an automotive application with an accuracy of 74% using acoustic information only, 60% using linguistic information only, and 83% using both (for seven basic emotions).

Emotional Content and Dialogue Management

Emotion as a computational topic in artificial intelligence goes back to the speculations of Sloman (REF) and the implementation of Colby's PARRY (Colby 74), the first robust conversational agent in the early 1970's which had a simple screen dialogue interface. PARRY had internal variables FEAR and ANGER which rose and fell according to keywords in the human dialogue partner's input, and whose values selected a response at each turn from the alternatives available as responses to the given input. From the late 1980s there was a general consensus in the natural language processing community that assembling the knowledge about Language and the world was going to be a key problem for any useful natural language system. The solution, for those interested in whole systems, was to have their conversational agent play a constrained role. Hence, PARRY played the role of a patent with paranoia, and Weizenbaum's ELIZA (Weizenbaum 1966) program played the role of a Rogerian psychologist. Later work also focused on roles that, at first blush, could be construed to have no emotional content such as a virtual teacher (Zukerman 2001) or virtual assistants in the form of the ever patient library assistant or (un)smiling butler (AskJeeves). As discussed above however, the pervasive emotions influence our moment by moment decisions, and in order to make virtual tutors and advisors believable, these virtual humans must have some understanding of the human emotional landscape. Capturing such information is key to creating believable characters of any kind. Those coming from a background of dialog systems have however have tended to see the issue as one of **politeness**.

Walker et al (97) introduce Linguistic Style Improvisation in which texts are generated by decomposing a conversational goal into a series of speech-acts. Parameters to this planning process are social distance and power relations that must be accounted for in the management of politeness. REA (Cassell 99) is a virtual real-estate agent that shows real humans through virtual houses. In order to do this REA must manage her interpersonal relationship with the human. The role of real-estate agent is more flexible than the formal roles of teacher or Rogerian psychologist, and the changing role is reflected in REA's measure of social distance. Franco (Estival 03) is an ECA designed as a virtual assistant in a military command and control centre. Such an agent must not only be able to convey information, it must do it in a manner that obeys the rules of social relations. Once again the approach taken was to see the problem as one of politeness (Wallis 01). Recent work within the scope of HUMAINE has embraced the notion of politeness and Rehm and André (05) are marking up video with politeness markers.

A more recent trend has been the adoption of more sophisticated models of emotions based on cognitive modelling, such as the OCC model (Ortony et al., 1988) and the Oatley model (Oatley and Johnson-Laird, 1987). This can be explained as a consequence of the development in the past few years of more sophisticated ECA comprising cognitive, affective and communicative modules.

One illustration of this approach is given by research carried out at the Institute of Creative Technologies (ICT) of the University of Southern California, in particular the Mission Rehearsal Exercise (MRE) project. This military training application is dedicated to crisis situations in peacekeeping operation and features emotional ECA throughout the cast of virtual actors, from military personnel reacting to the trainee's orders or shifting blame on fellow virtual agents, to civilians caught in dramatic situations as part of the mission being rehearsed. The MRE system features emotional ECA that incorporate user emotion detection, ECA emotional modelling related to the agent's cognitive model and a complete spoken dialogue system. The system has been used to experiment the role of emotion in dialogue (Traum et al., 2004). In addition, its various components are sufficiently integrated to support the processing of global communication situations such as social judgements and attribution of responsibility (Mao and Gratch, 2004), modelling how agents cope with situations (Marsella and Gratch, 2003), etc.

Several projects have beed dedicated to emotional ECA at the Center for Advanced Research in Technology for Education (CARTE) of the University of Southern California, with specific emphasis on ECA with social skills, politeness, etc. For instance, the tactical language training system (Johnson et al.,

2005) develops tools to support individualized language learning, and apply them to the acquisition of tactical languages: subsets of linguistic, gestural, and cultural knowledge and skills necessary to accomplish specific tasks. Other projects have been dedicated to the impact of emotions in learners using Interactive Tutoring Systems featuring ECA.

The "Façade" system (Mateas and Stern, 2004) is an interactive narrative application, in which the user communicates through Multimodal dialogue with artificial actors evolving in a virtual 3D world. The main mode of user interaction is through (written) dialogue with the virtual characters; the multimodal aspects corresponding to the physical interactions resulting from user navigation in the virtual world (which select deictic references to virtual world objects or characters). The user input is interpreted as speech acts that in turn affect the emotional status of the virtual actors, altering the course of the narrative.

Other recent research has developed emotional models to assist the control of user-system dialogue itself, and this shall be of particular relevance to future ECA systems, especially those based on spoken dialogue systems.

- Bosma and Andre (2004) have introduced a method to disambiguate dialogue acts using emotional data, with the goal of improving the communication between users and ECA. The underlying hypothesis is that the meaning of utterances in dialogue is closely correlated to the user's emotional status, and that the emotional status can even assist in determining the actual meaning of two textually identical utterances. Their system processes natural language input through finite-state systems that contain variable weights for various dialogue acts and obtains these weights from a Bayesian network computing the user's emotional status (using application history as well as input from physiological sensors).
- Riccardi and Hakkani-Tur (2005) have studied the grounding of emotions in spoken dialogue systems in the context of AT&T's "How May I Help You System" (Gorin et al., 2002). This work essentially addresses the impact of the user state on machine dialogue strategies and machine's performance. Although the emotional model is restricted to a positive/negative rating (towards the system), this work demonstrated that the recognition of emotional patterns (including temporal patterns) could be used to improve dialogue strategies and to predict speech recognition errors.

Developing dialog systems for ECA that more than simply demonstrate the possibility takes considerable resources. For instance, one long running project at the University of Bielefeld that has developed Max (Kopp et al, 05). A recent incarnation of Max has been running for 18 months as a guide in the Heinz Nixdorf Museums Forum in Paderborn, Germany. Max incorporates an emotion system which keeps track of, for example, obscene or politically incorrect wordings in the user input. Repeated insults will put the agent in an extremely bad mood which eventually results in Max leaving the scene. This abusive behaviour is not bought on by a lack of politeness on behalf of the agent; it appears to be spontaneous and extremely common amongst dialog systems in public places. These recent findings close the loop between the emphasis on politeness in the earliest dialogue systems (PERRY, ELIZA) It seems something is missing from our common understanding of what natural language processing is about.

The Companions approach to Emotion

The paradigm we advocate is somewhat different from the popular corpus and mark-up approach, and rests on investigating and modelling not human-human behaviour involving emotions but on the behaviour of humans to machines, which is now an established and attested type of behaviour, and what behaviour in humans towards machines we should be seeking to elicit in a cooperative interface like a COMPANION (Wilks 2004). This interaction type is quite distinct (de Angeli 01) and human-human dialogue corpora may not be relevant to its modelling.

The intention is to embrace the notion of an 'affective loop' and develop a series of spoken language dialog systems that provide varying degrees and kinds of (pervasive) emotional feedback. Rather than "understanding" what is being said, or grounding information state updates (for example, Kreutel 2000) the primary aim will be to provide feedback that engages the user in the process of conversation. In other words, the COMPANIONS approach aims at grounding emotional aspects in the semantic content of human-agent communication, rather than resorting to an a priori ontology of affects.

From a technical perspective, a further challenge consists in developing a proper HLT approach supporting this process of engagement through dialogue. This approach should include semantic principles underlying the automatic processing of various aspects of human-ECA bonding, including

politeness, empathy, humour, etc. These semantic principles should also underlie the representation of dialogue acts in a way which would be compatible with the experiments in dialogue control planned in COMPANIONS.

Like playing a drum, the user's emotional response should have an obvious link, via the system, with the user's actions. Measures of success might include the extent to which a user treats the system as if it is human. For example, does the system allow the user to "willingly suspend their disbelief" or not, and how often does a system get used voluntarily. In the context of HUMAINE's stated aim of trouble shooting in automated call handling systems, it is quite shocking just how much verbal abuse conversational agents receive (de Angeli 05) and the extent to which systems can "push back" as part of this affective loop is an open question.

Bibliography

Ask Jeeves, http://www.ask.com

Bates J (1994). The role of emotion in believable agents, Communications of the ACM, 37(7), 122-125.

Batliner, A., Hacker, C., Steidl, S. Nöth, E. Haas, J. (2004). From Emotion to Interaction: Lessons from Real Human-Machine-Dialogues. In E. Andre', L. Dybkiaer, W. Minker, P. Heisterkamp (eds.) Affective Dialogue Systems. Springer Verlag, Berlin, Germany, 2004

Beskow, J. Cerrato, L., Granstrom, B. House, D. Nordenberg, M. Nordstrand, M. Svanfeldt,G., (2004). Expressive Animated Agents for Affective Dialog Systems. In E. Andre', L. Dybkiaer, W. Minker, P. Heisterkamp (eds.) Affective Dialogue Systems. Springer Verlag, Berlin, Germany, 2004

Beskow J, Edlund JG and Nordstrand M (2005). A model for multi-modal dialogue system output applied to an animated talking head. In Minker, W, Bühler, D and Dybkjaer, L (eds) Spoken Multimodal Human-Computer Dialogue in Mobile Environments, Text, Speech and Language Technology, 29, Dordrecht, The Netherlands, Kluwer Academic Publishers.

Bosma, W. and Andre, E (2004). Exploiting Emotions to disambiguate Dialogue Acts, Proceedings of the 9th international conference on Intelligent User Interface, IUI 2004, Funchal, Madeira, ACM Press pp. 85-92.

Brahnam, Sheryl, Strategies for handling customer abuse of ECAs in:Abuse: the darker side of Human-Computer Interaction (INTERACT '05), edited by Antonella De Angeli, Sheryl Brahnam and Peter Wallis, Rome, September 2005.

Cavalluzzi, A., De Carolis, B., Carofiglio, V., Grassano, G (2003). Emotional Dialogs with an Embodied Agent. *User Modeling*: 86-95

Cavazza M., Martin O., Charles F., Mead S.J., Marichal X. and Nandi A., (2004). Multi-modal Acting in Mixed Reality Interactive Storytelling. IEEE Multimedia, July-September 2004, Vol. 11, Issue 3.

Cavazza, M. and Charles, F., (2005). Dialogue Generation in Character-based Interactive Storytelling. AAAI First Annual Artificial Intelligence and Interactive Digital Entertainment Conference, Marina del Rey, California, USA, AAAI Press.

Cahn JE (1990). The generation of affect in synthesised speech, journal of the American Voice Input/Output Society, 8, 1-19.

Cassell, J., Bickmore, T., Billinghurst, M., Campbell, L., Chang, K., Vilhjalmsson, H. and Yan, H. Embodiment in Conversational Interfaces: Rea in "Proceedings of the CHI'99 Conference", Pittsburgh, PA 520-527 1999

Cassell J, Sullivan J, Prevost S and Churchill E (2000). Embodied Conversational Agents, MIT Press Cambridge.

Colby (Kenneth Mark) Roger C. Parkinson and Bill Faught Pattern-Matching Rules for the Recognition of Natural Language Dialogue Expressions American Journal of Computational Linguistics, 1974

Cosi, P. Fusaro , A. Grigoletto, D. Tisato, G. (2004). Data-Driven Tools for Designing Talking Heads Exploiting Emotional Attitudes", In E. Andre', L. Dybkiaer, W. Minker, P. Heisterkamp (eds.) Affective Dialogue Systems. Springer Verlag, Berlin, Germany, pp. 101-112.

Cowie R, Douglas-Cowie E, Tsapatsoulis N, Votsis G, Kollias S, Fellenz W and Taylor J (2001). Emotion recognition in human-computer interaction. IEEE Signal Processing Magazine, 18, pp.32-80.

Cowie R. and Marc Schroder Piecing Together the Emotion Jigsaw in "MLMI 2004, LNCS 3361" edited by S. Bengio and H. Bourlard Springer-Verlag Berlin 305--317 http://emotion-research.net/aboutHUMAINE/ 2005

Damasio, A. R. Descartes' Error: Emotion, Reason and the Human Brain, Grosset/Putnam, New York. 1994

Darwin C (1872). The Expression of Emotions in Man and Animals, John Murray, London.

Davidson R, Scherer KR and Goldsmith H Eds. (2003). Handbook of Affective Sciences, Oxfrod University Press.

de Angeli (Antonella) Graham I. Johnson and Lynne Coventry The unfriendly user: exploring social reactions to

chatterbots in "Proceedings of The International Conference on Affective Human Factors Design" edited by Helander, Khalid and Tham, ASEAN Academic Press London, 2001

de Angeli (Antonella) Stupid Computer! Abuse and Social Identity in "Abuse: the darker side of Human-Computer Interaction (INTERACT '05)" edited by Antonella De Angeli and Sheryl Brahnam and Peter Wallis, Rome (http://www.agentabuse.org/) September, 2005

Douglas-Cowie E, Cowie R and Campbell N Eds. (2003). Speech and emotion, Speech Communication (special issue), 40(1-3).

Eisna, Roos Usable Technology for Older People: Inclusive and Appropriate (UTOPIA) http://www.computing.dundee.ac.uk/projects/UTOPIA/

Ekman P (1999). Basic emotions, Handbook of Cognition and Emotion, Dalgleish T and power M (Eds.), John Wiley, New York, 301-320.

Estival (Dominique) Spoken Dialogue for Virtual Advisers in a semi-immersive Command and Control Environment in "SIGdial Workshop on Discourse and Dialogue", Sapporo, Japan (citeseer.ifi.unizh.ch/646935.html) July 2003.

Gorin, A.L., Abella, A., Alonso, T. Riccardi, G. and Wright, J.H. (2002), Automated Natural Spoken Dialog, IEEE Computer Magazine, vol. 35 (4) pp. 51-56.

Gratch, J. (1999). Why you should buy an emotional planner. Proceedings of the Agents'99 Workshop on Emotion-based Agent Architectures (EBAA'99)

Gratch, J. and Marsella, S. Evaluating a Computational Model of Emotion Journal of Autonomous Agents and Multiagent Systems (Special Issue on the Best of AAMAS 2004) 2005

Haddad D et al (2002). Investigation and evaluation of voice stress analysis technology, US Dept. Justice, AFRL Report No.193832.

Henton C and Litwinowicz PC (1994). Saying and seeing it with feeling: techniques for synthesizing visible, emotional speech, Proc. 2nd ESCA/IEEE Workshop on Speech Synthesis, 73-76.

Iida, A., Campbell, N., Iga, S., Higuchi, F., and Yasumura, M. (2000) A Speech Synthesis System with Emotion for Assisting Communication, ISCA Workshop on Speech and Emotion, Belfast.

Johnson, W.L., Rizzo, P., Bosma, W., Kole, S., Ghijsen, M., van Welbergen, H. (2004) Generating Socially Appropriate Tutorial Dialog. In E. Andre', L. Dybkiaer, W. Minker, P. Heisterkamp (eds.) Affective Dialogue Systems. Springer Verlag, Berlin, Germany, 2004, pp. 254-264.

Johnson, W.L., Vilhjalmsson, H. and Samtani, P. (2005). The Tactical Language Training System (Technical Demonstration), , The First Conference on Artificial Intelligence and Interactive Digital Entertainment, Marina del Rey, CA, AAAI Press.

Kopp (Stefan), Lars Gesellensetter , Nicole Kramer and Ipke Wachsmuth A Conversational Agent as Museum Guide - Design and Evaluation of a Real-World Application in "5th International working conference on Intelligent Virtual Characters" (http://iva05.unipi.gr/index.html) 2005

Kreutel J. and Colin Matheson Modelling Dialogue using Multiple Inferences over Information States in "Proceedings of ICOS-2, 2nd Workshop on Inference in Computational Semantics", Dagstuhl, 2000

Mao, W. and Gratch, J. (2004). Social Judgment in Multiagent Interactions, in: Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems (AAMAS 2004), New York, 2004.

Marsella, S. and Gratch, J. (2003). Modeling Coping Behavior in Virtual Humans: “Don't Worry, Be Happy," in: 2nd International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2003), Melbourne, Australia, July 2003

Martinovski, B., Traum, D., and Marsella, S. (2005). Rejection of empathy and its linguistic manifestations in Procedings of Conference on Formal and Informal Negotiation, FINEXIN, Canada: Ottawa.

Mateas, M. and Stern, A. (2004). Natural Language Processing In Facade: Surface-text Processing. Technologies for Interactive Digital Storytelling and Entertainment (TIDSE 2004), Darmstadt, Germany.

Mei Si, Stacy C. Marsella and David V. Pynadath Thespian: Using Multi-Agent Fitting to Craft Interactive Drama in The Fourth International Joint Conference on Autonomous Agents and Multi Agent Systems (AAMAS 2005), Utrecht, The Netherlands.

Moore RK Ed. (1996). Speech under stress, Speech Communication (special issue), 20(1-2).

Muller, T.J., Hartholt, A., Marsella, S., Gratch, J., Traum, D.R. Do You Want To Talk About It? A First Step Towards Emotion Integrated Dialogue. In: E. Andre, L. Dybkjae r, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems: Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004

Murray IR and Arnott JL (1993). Towards the simulation of emotion in synthetic speech: a review of the literature on human vocal emotion, J. Acoustical Society of America, 93, pp.1097-1108.

Nijholt, A. (2002). Embodied Agents: A New Impetus to Humor Research. In: The April Fools Day Workshop on

Computational Humour, O. Stock, C. Strapparava & A. Nijholt (eds.), Proceedings Twente Workshop on Language Technology 20 (TWLT 20), ISSN 0929-0672, ITC-IRST, Trento,Italy, pp. 101-111.

Oatley, K. and Johnson-Laird, P.N. (1987). Towards a Cognitive Theory of Emotions, Cognition and Emotion 1(1):29-50.

Ortony, A., Clore, G., Collins, A. (1988) The Cognitive Structure of the Emotions, Cambridge University Press, Cambridge.

Pelachaud C and Poggi I (2001). Towards believable interactive embodied agents, Fifth International Conference on Autonomous Agents workshop on Multimodal Communication and Context in Embodied Agents, Montreal.

Persson, P. EXMS: an animated and avatar-based messaging system for expressive peer communication, In Proceedings of the 2003 international ACM SIGGROUP conference on Supporting group work, pp. 31-39, Sanibel Island, Florida, USA. 2003.

Prendinger, H., Dohi, H., Wang, H., Mayer, S. and Ishizuka, M. (2004). Empathic Embodied Interfaces: Addressing Users' Affective State: Embodied Interfaces That Address Users' Physiological State. In: E. Andre, L. Dybkjae r, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems: Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, pp. 53-64

Picard, R. Affective Computing, MIT Press, Cambridge, MA. 1997

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

Riccardi, G. and Hakkani-Tür, D. (2005). Grounding Emotions in Human-Machine Conversational Systems. In: M. Maybury, O. Stock, W. Wahlster (Eds.) Intelligent Technologies for Interactive Entertainment: First International Conference, INTETAIN 2005, Madonna di Campiglio, Italy, Springer-Verlag, pp. 144-154.

Rigoll G, Müller R and Schuller B (2005). Speech emotion recognition exploiting acoustic and linguistic information sources, Proc. SPECOM, Patras, Greece, 61-67.

Scherer KR, Schorr A and Johnstone T Eds. (2001). Appraisal Processes in Emotion: Theory, Methods, Research, Oxford University Press, New York and Oxford.

Scherer KR (2003). Vocal communication of emotion: a review of research paradigms, Speech Communication 40, 227-256.

Schröder M (2001). Emotional speech synthesis: a review, Proc. Eurospeech, Allborg, Denmark, 561-564.

Slater, M., Steed, H. A. and Gaurau, P. M. (2000) Acting in Virtual Reality, In Proceedings of the third international conference on Collaborative virtual environments, pp. 103-110, San Francisco, California, United States.

Streit, M., Batliner, A. and Portele, T (2004). Cognitive-Model-Based Interpretation of Emotions in a Multi-modal Dialog System. In: E. Andre, L. Dybkjae r, W. Minker, P. Heisterkamp (Eds.), Affective Dialogue Systems: Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004, pp. 65-76.

Sundström (Petra) Exploring the Affective Loop, PhLic Thesis, Stockholm University, Sweden, 2005.

Turunen (Markku) Esa-Pekka Salonen, Mikko Hartikainen, Jaakko Hakulinen, Kristiina Jokinen, Jyrki Rissanene, Kari Kanto, Antti Kerminen, William Blacm, Allan Ramsay, Adam Funk, Andrew Conroy, Paul Thompson, Mark Stairmand, Björn Gambäck, Magnus Sahlgren, Fredrik Olssom, Maria Cheadle, Preben Hansen, and Stina Nylander AthosMail: a Multilingual Adaptive Spoken Dialogue System for the E-mail Domain in "Robust and Adaptive Information Processing for Mobile Speech Interfaces (DUMAS Final Workshop)" edited by Björn Gambäck and Kristiina Jokinen (http://citeseer.ist.psu.edu/729314.html) August, 2004

Walker, M., Cahn, J., and Whittaker, S. (1997). Improvising Linguistic Style: Social and Affective Bases for Agent Personality . In Proceedings of the Conference on Autonomous Agents, AGENTS97.

Wallis, P., Helen Mitchard, Damian O'Dea and Jyotsna Das, Dialog Modelling for a Conversational Agent in "AI2001: Advances in Artificial Intelligence, 14th Australian Joint Conference on Artificial Intelligence" edited by Markus Stumptner, Dan Corbett and Mike Brooks, Adelaide Australia, Springer (LNAI 2256) 2001

Weizenbaum (Joseph) ELIZA - A Computer Program For the Study of Natural Language Communication Between Man and Machine Communications of the ACM vol 9(1) January 1966

Wilks, Y. (2004) Companions : a new paradigm for agents. Proc. International AMI Workshop, IDIAP, Martigny,CH.

Wundt W (1874). Grundzüge der Physiologischen Psychologie, Engelmann, Leipzig.

Yildirim S et al (2004). An acoustic study of emotions expressed in speech', Proc. InterSpeech, Jeju, Korea.

Zukerman (Ingrid) and Richard McConachy WISHFUL: A Discourse Planning System that Considers a User's Inferences Computational Intelligence vol 17(1) Blackwell Publishers Boston MA & Oxford UK, Februrary 2001