Animating Expressive Characters for Social Interactions

Symposium of the AISB'02 Convention

Imperial College, London, UK

4th - 5th April, 2002

Ruth Aylett and Lola Cañamero
Symposium Chairs

Programme Committee:

Elisabeth André , University of Augsburg, Germany
Yasmine Arafa, Imperial College, UK
Ruth Aylett, University of Salford, UK
Daniel Ballin, BTExact Technologies, Radical Multimedia Lab, UK
Cynthia Breazeal, MIT, USA
Paul Brna, University of Leeds, UK
Lola Cañamero, University of Hertfordshire, UK
Kerstin Dautenhahn, University of Hertfordshire, UK
Fiorella De Rosis, University of Bari, Italy
Adrian Hilton, University of Surrey, UK
Ana Paiva, INESC, Portugal
Catherine Pelachaud, Universita di Roma La Sapienza, Italy
Paolo Petta, OEFAI, Austria
John Vince, Bournemouth University, UK
Thomas Wehrle, University of Geneva, Switzerland
Adrian Woolard, BBC Imagineering, UK

Table of Contents

Preface

The ability to express and recognise emotions is a fundamental aspect of social interaction. The importance of endowing artefacts (for example synthetic characters or robots) with these capabilities is nowadays widely acknowledged in different research areas such as affective computing, socially intelligent agents, computer animation, or virtual environments. Researchers in all these areas are however confronted with the problem of how to make the emotional displays of artefacts and characters believable and acceptable to humans. This can involve not only generating appropriate expressions and behavioural displays - explored in animated film for many years - but also endowing artefacts with underlying models of personality and emotions that support the coherence and autonomy of their emotional displays and interactions.

Thus this symposium concerns 'animation' not only from a graphical perspective, but more generally in the human sense: making characters 'life-like', externally but also 'internally': giving them an 'anima'. The aim of this symposium is to bring together researchers from different disciplines (including psychology, animal behaviour, the arts, computer graphics and animation as well as those mentioned above) to reflect on this common problem from different perspectives and to gain new insights from this multi-disciplinary feedback. Researchers from multiple disciplines (e.g., AI, arts, biology, ethology, humanities, neurosciences, philosophy, psychology, social sciences) are invited to participate.

This symposium covers a domain of active research in a number of different communities. Expressing and recognising emotions is studied in the social agent community because it is known to be fundamental to establishing a social relationship; it is studied in computer animation because it is also a basic requirement for the believability of animated characters and for human engagement in the narratives in which they are involved. It has also been studied for many hundreds of years in expressive arts and for most of the 20thC, and into the 21stC, in the psychology of emotion.

With the growth of synthetic characters in virtual environments and on the web, as well as of the introduction of domestic or entertainment robots, this topic is receiving an increased importance. Who could accept or trust an Annanova who read the news about September 11th with the same emotion as a report on the England football team qualifying for the World Cup? If users are to accept and interact with artefacts such as intelligent synthetic characters and domestic robots, then appropriate emotional interaction is vital.

However the different communities that study emotional expression and recognition do so in very different ways, and often do not interact with each other. Within computing there is a gulf between graphics and animation researchers who are concerned about exterior expressiveness, and workers in AI and cognitive science who build computer models of the internals of such artefacts. Meanwhile communities entirely outside of computing have ideas known little or not at all within it but of major potential use, as witness the recent use of Laban analysis from dance choreography in behavioural animation.

We therefore hope that this symposium will provide a rare opportunity for people from these varied communities to meet and interact and would argue that the very diversity of concepts and approaches should produce many new ideas and new feedback paths between the different disciplines.

For further information, please visit the symposium website:

http://homepages.feis.herts.ac.uk/~comqlc/aecsi02

Ruth Aylett and Lola Cañamero

Symposium Chairs

Pygmalion Revisited: Taming and Training the Virtual Social Animal

Robert Burke
Lionhead, UK

Abstract

Recent work in the Synthetic Characters group, led by Bruce Blumberg at the MIT Media Lab, has focused on integrating concepts from learning and planning into a group of socially-aware virtual creatures. One contribution to that work was a mechanism, based on an ethological model, that gave a creature the ability to predict events that are likely to happen in the future, so that the creature could plan its actions based on those expectations. Some of the other concurrent work has allowed us to augment the creatures' social sense. Social systems are dynamic, constantly evolving and adapting to the circumstances in which their members find themselves. In this talk, I will describe how some of the learning, planning and training techniques work, and how they are useful for our creatures in a social setting. By social, we mean both that these creatures co-exist with "real-world" participants, as well as with other virtual creatures.

Facial Expressions in Social Interactions: Beyond Basic Emotions

Susanne Kaiser & Thomas Wehrle
Geneva Emotion Research Group, University of Geneva, Switzerland
Geneva Interaction Lab, University of Geneva, Switzerland
Susanne.Kaiser@pse.unige,ch ; Thomas.Wehrle@pse.unige.ch

Abstract

As can be seen from the many contributions to this symposium and related ones, there exist an increasing number of research groups that are developing computer interfaces with synthetic facial displays. These researchers attempt to use these facial displays as a new modality that should make the interaction more efficient (or more natural, intuitive, appealing, etc.), while lessening the cognitive load. In addition, several researchers point out that automatic interpretation of gestures and facial expressions would also improve man-machine interaction. Facial expression is usually synthesized or predicted on the basis of a given emotion. The prototypical expressions for basic emotions (happiness, sadness, surprise, disgust, anger, and fear) postulated by discrete emotion psychologists are rather consistently produced and interpreted among different cultures, and can be used as icons to represent an emotion. However, these facial expressions are actually rarely observable in every day affective episodes, because of the complexity and multi-functional nature of facial expression and a high inter-individual variance.

In the first part of the presentation, we would like to sketch the different functions of facial expressions in human-human and human-computer interactions from a psychological point of view. When considering spontaneous interactions, it is very difficult to identify whether a facial expression is an indicator of an emotion or whether it is a communicative signal. A smile or a frown, for instance, can have different meanings. It can be a *speech-regulation signal* (e.g., a back-channel signal), a *speech-related signal* (illustrator), a means for *signalling relationship* (e.g., when a couple is discussing a controversial topic, a smile can indicate that although they disagree on the topic there is no "danger" for the relationship), an indicator for *cognitive processes* (e.g., frowning while concentratedly attending to a problem, or when a difficulty is encountered in a task), or the *expression of an emotion* (affect display). Affect displays that occur during an interaction can refer to the interaction partner (e.g., becoming angry with the other) but it can also refer to other persons or themes the interaction partners are talking about (e.g., sharing the anger about something).

In the second part, we suggest an alternative way to predict, synthesize and understand facial behaviour, which is based on *componential appraisal theory*. Appraisal theorists following a componential approach share the assumption that a) emotions are elicited by a cognitive evaluation (appraisal) of antecedent situations and events and that b) the patterning of the reactions in the different response components (physiology, expression, action tendencies, feeling) is determined by the outcome of this evaluation process. For these appraisal theorists the complexity and variability of different emotional feelings can be explained without resorting to a notion of basic emotions. They argue that there are a large number of highly differentiated emotional states, of which the current emotion labels capture only clusters or central tendencies of regularly recurring ones, referred to as *modal* emotions. In line with this reasoning, facial expressions are not seen as the "readout" of motor programs but as indicators of mental states and evaluation processes. To illustrate our appraisal-based approach, we present an empirical setting, which allows us to collect behavioural data in human-computer interactions as well as some examples of empirical studies.

Finally we aim at illustrating how an appraisal-based approach to the understanding of the relation between emotion and facial expressions might be instrumental for a) facial expression synthesis (animated intelligent agents), and b) automatic expression recognition (decoding the emotional state of the user). From the beginning of the early eighties appraisal theories have given important inputs to emotion synthesis and affective user modelling. However, most of these applications only use appraisal theory to implement the "cognitive" component of an emotional interface. The outcome of the appraisal process is then mapped into an emotion category, which determines the synthesis of the respective facial expression pattern. This might be an unnecessary and complicated procedure that also reduces the available information and might bias the analyses.

Suspending Disbelief: How to Make People Feel They ARE Walking with Dinosaurs

Mike Milne
Framestore, UK

How Are You ? Synthetic Personalities for Edutainment.

Tim Barker
School of Computing
Staffordshire University, UK.
email: timothy.barker@staffs.ac.uk

Abstract

A Virtual Learning Environment (VLE) was introduced in schools in Leeds to help overcome educational and underlying social problems. Studies revealed that the VLE 'note-taking' functionality was inadequate so a user-centred design of an Animated Pedagogical Agent was undertaken. This led to the development of a hybrid reactive/deliberative agent-based prototype to help motivate and engage students through the use of appropriate behaviour-based affective qualities. Subsequent evaluation in schools showed that students perceived the intended agent affect. Furthermore, a correlation was also observed between student achievement and student affective perceptions. It is therefore claimed that future agents could adapt affectively partly based upon prior and subsequent student achievement.

"Well he acts like he has genuine emotions, of course he's programmed that way to make it easier for us to talk to him but as to whether or not he has real feelings is something I don't think anyone can truthfully answer."

Dr. Poole, "2001: A Space Odyssey"
[speaking of HAL] (Kubrick, 1968)

1. Preamble

The question "How are you ?" is commonly used to elicit a description of an interlocutor's well-being prior to engaging in conversation. It suggests a degree of empathy, curiosity and personable nature of the enquirer. Synthetic personalities, in the domain of Education, aiming to provide such qualities without utilising underlying affective cognitive models are described herein.

2. Background

Research was undertaken in conjunction with Chapeltown and Harehills Assisted Learning Computer School (CHALCS) and other local Leeds schools. CHALCS is an out-of-hours school established in 1987 to directly address educational and corresponding social problems in a deprived inner-city area of Leeds. The project involved the introduction of a Virtual Learning Environment (VLE) for Advanced Level Physics (Barker and Pilkington, 2000).

As the VLE was used with students the tool integral to WebCT (the off-the-shelf VLE) was thought by students, teachers and observers to be unsupportive of the *process* of student's summarisation as well as offering few facilities and utilising uninspiring Human-Computer Interaction techniques.

One further motivating factor for the research was the notion of Social Learning Systems proposed by Chan (1996) which are *"environments where multiple participants, either computer simulations or real human agents, work at the same computer or across connected machines"*. It is a student-centred model which sees the student at the centre of a range of learning supports. However of concern here is an essential part of a Social Learning System, namely the "virtual" or artificial Learning Companion.

3. Supporting Student's Note-Taking

Studies were undertaken at CHALCS and a 'feeder' school, Notre Dame Sixth Form College, in order to ascertain qualities of an improved 'take notes' facility. This resulted in recommendations such as a 'scratch pad' for jotting ideas and a process model describing the stages which students typically negotiate when note taking (Barker and Pilkington, 2000). A new tool was therefore envisaged which would provide an artificial Learning Companion (Chan and Baskin 1988) with which students could collaborate on the task of note-taking. Previous research in the field of Learning Companions includes that of Aimeur and Frasson (1996), Chan and Baskin (1988, 1990), Goodman et al.(1998), Dillenbourg and Self (1992), Uresti (2000) and Brna et al. (2001). This research is novel in the context of these efforts in attempting to supplement a *commercial* VLE with an artificial Learning Companion. It was further hypothesised, based on previous research (Lester et al., 1997), that in order to further *engage* students in the new process of *interactive* note-taking the Learning Companion should be based upon work in Animated Pedagogical Agents (Johnson and Rickel, 2000) which would help motivate students and therefore serve as a form of *Edutainment.*

4. The Philosophy of Affectations

Consequently an Agent was designed in a user-centred approach utilising the Wizard-of-Oz (WoZ) technique (Barker and Pilkington, 2001). Central to the design of the proposed two agents and to the thesis of this research is a notion of seven 'affectations' (Barker and Pilkington, 2001) of distinct personality types ('dominant' and 'passive') utilising "emergent emotions" to ascribe intended 'feelings' as hopefully perceived by students. Picard (1997) describes "*emergent emotions*" as "*those which are attributed to systems based on their observed emotional behaviour – especially when the system which is behaving has no explicit internal mechanism or representation for emotions*".

Table 1. Example 'moves' for Genie.

Moves	Animations	Speech
me take over	move to centre of screen, cloud magically appears (+ cymbal noise), rubs hands together, raises eyebrows, returns to original position	"I'm going next !"
you wrong !	hands out, shake head, scratch head, hands shoot out, mouth open aghast	"you could have done better !"
why ?	hands shoot out to sides, mouth open aghast, one hand on hip, other on side of face with eyebrows raised	"tell me why you did that !"
yes	raised eyebrows, smile, clutching hands	"yes, definitely !"
no	hands out to side and shake head	"no way !"
well done	move to centre of screen, applaud then return to original position	"Clever you !" "Well done !"

A Genie was prototyped for the WoZ study to demonstrate the 'dominant' affectations. Table 1 shows some of it's final 'moves' including the animation sequences and corresponding speech acts. However, these are a refined version of those utilised in the WoZ study as they were redesigned based upon results of the study.

5. Implementing an Artificial Learing Companion

Barker and Pilkington (2001) describe the architecture of the Summary Intelligent Learning Assistant (SILA). This description includes accounts of the conceptual units of SILA such as the environment (i.e. WebCT and summary), the sensors, the reactive layer (containing the dialogue model), deliberative layers (e.g. to condense phrases), the world model (containing the summary representation) and the actuators. In addition to development using Visual BASIC and Microsoft Agent (also described therein) Microsoft Agent Scripting Helper (MASH) was utilised to aid authoring of the kinds of animations and speech utterances shown in Table 1. Microsoft Agent was chosen for speed of development in the prototype SILA (ProSILA) as it provides all the necessary core services to quickly implement an Animated Agent, such as fundamental animation sequences, a Text-To-Speech engine, voice recognition and an API for Visual BASIC.

Figure 1 shows the Genie making a contribution in the SILA environment, pointing towards the new addition and telling the student to "look at this !". It can be seen that the top window in Figure 1 shows the WebCT Physics course in a web browser and the bottom window shows the current notes. Between the two windows can be seen formatting tools and information then at the very top of the SILA window can be seen menu commands, for example, to provide file and browser functionality, assistant selection (the 'dominant' Genie or 'passive' Peedy the Parrot) and help.

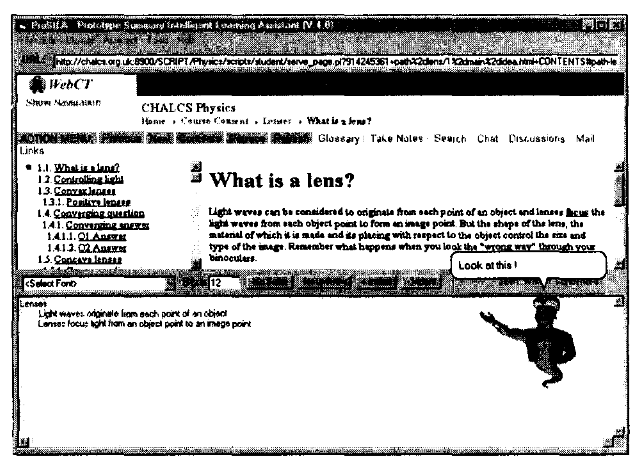
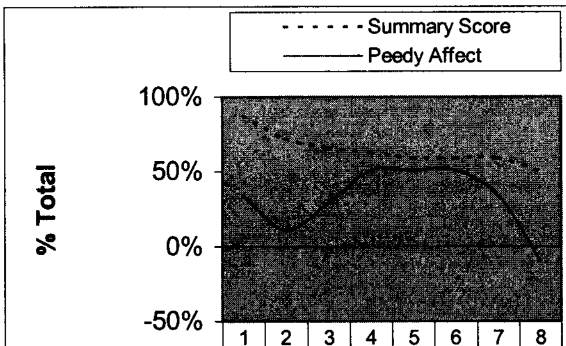


Figure 1. Genie contributing to the summary in SILA.

The student, then, interacts with the two agents by right mouse clicking over them (or speaking) to select context-sensitive commands. This results in the dialogue model advancing to the next state which may necessitate an action by either the student or the agent, such as the addition of a phrase in the summary document. This interaction continues until either party is satisfied that adequate notes have been produced.

6. Evaluating SILA

SILA was evaluated with students utilising a 'cooperative evaluation' methodology (Dix et al., 1997). The computer screen was videoed and one pre and one post semi-structured interview carried out. Also, a log of interactions and expert's summary scores were obtained. Subsequent analysis of the data employed both quantitative and qualitative techniques. This resulted in an interesting emerging correlation of perceived agent affect and student achievement, see Figure 2, suggesting a relationship between the two measures. That is, high achievers and low achievers perceive less positive affect whilst average students perceive most positive affect.



	1	2	3	4	5	6	7	8
Summary Score	88%	72%	66%	63%	59%	59%	59%	50%
Peedy Affect	34%	11%	29%	51%	51%	51%	34%	-9%

Student

Figure 2. Peedy affectations totals ranked by summary score for each student.

In addition student's self-reports indicated a difference in perception of affect between the two assistants. Peedy was seen as portraying more positive affect along the seven employed affectations than that of Genie thus demonstrating a successful design. Furthermore a number of suggestions for further work emerged from these evaluations such as having the assistants display differing behaviours for the same 'moves' and improving the explanation facility so that any phrase in the student notes could be elaborated upon by SILA and not just the most recent addition.

7. Preliminary Conclusions

The correlation of affect and achievement leads to a possibility of future agents partly adapting affectively to students on the basis of their performance, for instance in summarisation, or potentially in other cognitive tasks. For example, for students who perform well at summarisation the agents may act more 'dominantly' and for those who perform less effectively then the agents could act more 'submissively'. That is, 'dominant' agents will use terse language and be hasty whereas 'passive' agents would be more polite and more patient. Additionally, the agents can be seen to be displaying some of the "empathic characteristics" as elucidated by Brna et al. (2001) and validated in the evaluations although they have no underlying model of affect. Therefore one conclusion is that a valid cognitive model of an artificial Agent's internal 'emotional' state is not necessarily a precursor to creating valid emotional couplings between student and Learning Companion. This is in-keeping with results found by Reeves and Nass (1996). That is, personality is conveyed to subjects through such simple means as appropriate language and simple animation without resorting to 'traditional' AI techniques.

8. Future Research Directions

Current research (Kapoor et al., 2001) is exploring Learning Companions which recognise student's affective states across mixed modalities thus circumventing reliance on error-prone student self-reporting (e.g. Abou-Jaude, 1999). This would allow an agent to assess the student's affective state which, together with other information such as summarisation achievement, would result in the agent adapting its reasoning and affective behaviour accordingly. These effective affective responses to student states need not be 'traditional' cognitive-like models but situated state-based automata as currently utilised in SILA's dialogue model although more 'traditional' AI techniques could still be employed in the more deliberative cognitive tasks such as summarisation. This hybrid deliberative and reactive approach with its combination of emotional and behavioural 'intelligence' in the Agent would lead to a more holistic approach to the aim of creating socially meaningful student-agent interactions in-keeping with current ideas concerning human cognitive abilities emerging from, for example, Neuroscience (Damasio, 1994). Unfortunately, for now, the question "How are you ?" is practically going unasked. However, it seems particularly in *Synthetic Personalities for Edutainment.*

References

S. Abou-Jaoude, C. Frasson. Integrating a Believable Layer into Traditional ITS. *Proceedings of Artificial Intelligence in Education*, Le Mans, France, IOS Press, 1999.

E. Aimeur, C. Frasson. Analyzing A New Learning Strategy According to Different Knowledge Levels. *Computers in Education*, 27(2): 115-127, 1996.

T. Barker, R.M. Pilkington. An Investigation and Design of Networked Learning in Inner-City Leeds. *Proceedings of Networked Learning 2000: Innovative Approaches to Lifelong Learning and Higher Education Through the Internet*, Lancaster, UK, 2000.

T. Barker, R.M. Pilkington. Simulated Affectations of an Animated Pedagogical Agent. *Proceedings of AISB'01: Symposium on emotion, cognition and affective computing*, University of York, UK, 2001.

P. Brna, B. Cooper, L. Razmerita. Marching to the Wrong Distant Drum: Pedagogic Agents, Emotion and Student Modelling. *Proceedings of Workshop on Attitude, Personality and Emotions in User-Adapted Interaction in conjunction with User Modelling*, Southofen, Germany, 2001.

T.W. Chan, A.B. Baskin. Studying With the Prince: The Computer as a Learning Companion. *Proceedings of Intelligent Tutoring Systems*, Montreal, Canada, 1988.

T.W. Chan, A.B. Baskin. Learning Companion Systems. *Intelligent Tutoring Systems*, 6-33, C. Frasson, G. Gauthier (Ed.), Ablex, New Jersey, 1990.

T.W. Chan. Learning Companion Systems, Social Learning Systems and the Global Social Learning Club. *International Journal of Artificial Intelligence in Education*, 7(2):125-159, 1996.

A.R. Damasio. *Descartes' Error*. G.P. Punam's Sons, New York, 1994.

Proceedings of Intelligent Tutoring Systems, Montreal, Canada, 1992.

A. Dix, J. Finlay, G, Abowd, R, Beale. *Human-Computer Interaction*. (Second ed.) Pearson Education, 1997.

B. Goodman, A. Soller, F. Linton, R. Gaimari. Encouraging Student Reflection and Articulation using a Learning Companion. *International Journal of Artificial Intelligence in Education*, 9:237-255, 1998.

W.L. Johnson, J.W. Rickel. Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments. *International Journal of Artificial Intelligence in Education*, 11, 2000.

A. Kapoor, S. Mota, R.W. Picard. Towards a Learning Companion that Recognizes Affect. *Proceedings of Emotional and Intelligent II: The Tangled Knot of Social Cognition*, North Falmouth, MA, 2001.

S. Kubrick. 2001: A Space Odyssey. (Video ed.), AOL Time Warner, 1968.

J.C. Lester, S.A. Converse, B.A. Stone, S.E. Kahler, S.T. Barlow. Animated Pedagogical Agents and Problem-Solving Effectiveness: A Large-Scale Empirical Evaluation. *Proceedings of Artificial Intelligence in Education*, Kobe, Japan, 1997.

R.W. Picard. *Affective Computing.* MIT Press, 1997.

B. Reeves, C. Nass. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* Cambridge University Press, 1996.

J.A.R. Uresti. Should I Teach My Computer Peer ? Some Issues in Teaching a Learning Companion. *Proceedings of Intelligent Tutoring Systems*, 103-112, Montreal, Canada, Springer, 2000.

Mixed Emotion Modeling

Valeria Carofiglio, Fiorella de Rosis and Roberto Grassano

Department of Informatics; University of Bari
Via Orabona 4, 70126 Bari, Italy, +39 080 5443284
{derosis, carofiglio, grassano@di.uniba.it}
http://aos2.uniba.it:8080/IntInt.html

Abstract

We describe a method to simulate emotion triggering in Artificial Conversational Agents through Dynamic Belief Networks. We discuss the main advantages of this method (mixed emotion modeling, time decay, adaptation to personality factors and to the social context) and contrast it with alternative approaches to the same problem.

1 Introduction

In the 'Pure vs Mixed Emotions' Section of her book on 'Affective Computing', Rosalind Picard introduces the following example: after winning a Marathon, a professional runner described *"feeling tremendously **happy** for winning the race, **surprised** because she believed she would not win, somewhat **sad** that the race was over and a bit **fearful** because during the race she had acute abdominal pain"* (Picard, 2000). A runner's friend, when assisting to that Marathon, was probably *"**happy for her** because she won the race, although a bit **envious** for not being able to participate to it and **sorry for** seing her so tired, at the end"*. How is it that these two persons referred to feel this different mixture of emotions? Clearly, the main source of difference is due to the different structure of beliefs and goals of their minds. In the runner, the *intensity* of **fear** during the race was probably related, at the same time, to the *importance* she assigned to her goal of winning it and to variations in the *probability* of achieving this goal, that she dynamically revised during the race. The importance of this goal also affected the intensity of **happiness** for achieving it, while **surprise** was probably due to a difference between the likelihood she initially attached to achieving the goal and the final result. The **sadness** that the event was over might be a mixed emotion in its turn, some combination of **nostalgia** for a pleasant past event and **hope** to be again in a similar situation, in the future. The mixing of emotions in the runner's friend was probably due to a mixing of goals of approximately equal weight: **happy-for** is due to his desire of achieving 'the good of his friend' (conditioned to her winning the race) **sorry-for** to his desire of 'preserving her from bad' (illness, in this case) and **envy** to his desire of 'dominating her', in a way.

So: differences between the two persons, in the example, are due to differences in the beliefs, the goals they want to achieve, the weights they assign to achieving them and the structure of links between beliefs and goals; variations of these measures with time seem to govern cognitively-generated emotions. We claim that *Dynamic Belief Networks* are a mathematical tool that enables representing appropriately the dynamic arousal, evolution and disappearing of emotions, and the way these phenomena are related to personality factors and to the social context. We illustrated in detail the logics behind this modeling method in another paper, in which we discussed the advantages it offers in driving the affective behavior of an Embodied Conversational Agent that we are building in the scope of the European Project 'Magicster'[1] (de Rosis et al, submitted). In this short paper, we summarise the main features of our emotion modeling method, to focus our discussion on how the two metaphors of mixing emotions proposed by Rosalind Picard may be modeled with this formalism. We will restrict our analysis to goal-based emotions, in Ortony's classification (Ortony, in press).

2 The way emotions may mix

Picard evokes the *generative mechanism* as the key factor for distinguishing between emotions that may *coexist* (by mixing according to a 'tub of water' metaphor) and emotions that *switch* from each other in time (by mixing according to the 'microwave oven' metaphor): *"to the extent that two emotions have non overlapping generative mechanisms,.. they can coexist in time"*. Co-existence might be due, as well, to differences in the time of decay between emotions generated by the same mechanism: for instance, 'primary' emotions, like *fear,* and cognitively-generated ones, like *anticipation* (Picard, 2000).

To represent the two ways emotions may mix up, modeling formalisms should therefore enable representing their generative mechanism, the intensity with which they are triggered and the decay of this intensity with time. We claim that this representation power may be achieved with dynamic belief networks (DBNs). We briefly describe, in Section 3, what are DBNs and how they may be employed to represent emotion triggering. We then focus our analysis on the differences in the generative mechanism of emotions that may be represented with DBNs, with the aim of establishing a correspondence between the differences that might occur in these mechanisms and the two mixing metaphors proposed by Picard.

3 Emotion triggering with DBNs.

As we anticipated in the Introduction, our departure point is that emotions are triggered by the belief that a particular important goal may be achieved or threatened.

[1] Information Societies Technology Programme IST-1999-29078

So, our simulation is focused on the *change* in the belief about the achievement (or threatening) of goals of an Agent A, over time. In our monitoring system, the cognitive state of A is modeled at the time instants {T, T+1, T+2, ...}. Events occurred during the time interval (T,T+1) are observed, to construct a probabilistic model of the new state and reason about emotions that might be triggered by these events. A well known formalism for representing dynamic phenomena in conditions of uncertainty is that of *dynamic belief networks* (DBNs). DBNs are based on the idea that time is divided into time slices, each representing the state of the modelled world at a particular instant or interval; this state is described by means of a static belief network. Two consecutive time slices are linked by arrows between the corresponding domain variables we want to monitor. When something changes in the world, the network is extended for an additional time slice. As a consequence, its structure may change over time in a number of ways (Pearl, 2000).

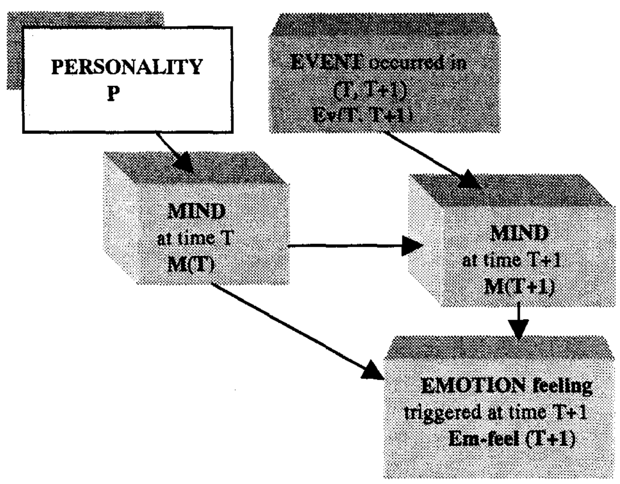


Figure 1: Outline of our Emotion Monitoring System

Figure 1 shows the general structure of our model, that includes the following static components:

- M(T) represents A's Mind at time T, with its beliefs about the world and its goals;
- Ev(T,T+1) represents the event occurred in the time interval (T, T+1), with its causes and consequences;

- M(T+1) represents A's Mind at time T+1;
- Em-feel(T+1) represents the fact that a particular emotion is activated, in A, at time T+1.

M(T+1) depends on M(T) and the event occurred in the interval (T,T+1). The feeling of emotions depends on both M(T) and M(T+1). We calculate the intensity of emotions as a function of two parameters: (i) the *uncertainty* in A's beliefs about the world and, in particular, about the possibility that some important goal is achieved or threatened, and (ii) the *utility* assigned to achievement of this goal. More in depth, if:

- Bel_A $Ach(G_i)$ is A's belief that the goal G_i will be achieved;
- $P(Bel_A\ Ach(G_i))$ and $P^*(Bel_A\ Ach(G_i))$ are the probabilities that A attaches to this belief, respectively before and after the event Ev occurred;
- $W_A\ (Ach(G_i))$ is the weight that A attaches to achieving G_i,

then, according to the utility theory (Pearl, 1990), the *variation of intensity in the emotion* (*ΔIe*) may be calculated as follows:

$$\Delta Ie=[P^*(Bel_A\ Ach(G_i))-P(Bel_A Ach(G_i))] * W_A(Ach(G_i))$$

In other words, ΔIe is the product of the change in the probability that G_i will be achieved, times the weight of this goal.

4 Goal-Based Emotion Categories

'Fortune-of-others' emotions (*sorry-for, happy-for, envy and gloating*) may be represented as points in the two-dimensional space ('desirability of the event', 'empathy of the reaction'): happy-for and envy apply to 'desirable' events while 'sorry-for' and 'gloating' apply to 'undesirable' ones; happy-for and sorry-for are driven by an empathic attitude, while gloating and envy are driven by a contrasting one. **Figure 2** shows the dynamic belief network that models how *happy-for* and *envy* may be triggered in an Agent A who assists to her friend U running the Marathon. This model shows that *happy-for* is triggered by an increasing probability of the belief that the high-level goal of 'desiring the good of others' will be achieved (Ach-GoodOf U). The intensity of this emotion depends on the variation of this probability, that is produced when evidence about some desirable occurred event (the runner comes to the end of the race) is propagated. It depends, as well, on the weight the Agent attaches to achieving that goal; this weight is, in its turn, a function of the Agent's personality (high for *altruistic* people, low for *egoistic* ones). The emotion of *sorry for* is triggered by the same nodes as *happy-for*, when these nodes have opposite values because the occurring event is *undesirable* (for instance, the runner has a suffering aspect): the high-level goal involved in this case is that of 'preserving others from bad' (Ach-PreserveBad U). *Gloating* is related to envy by a similar triggering mechanism: the high-level goal involved is, in this case, that of 'desiring the bad or others' and 'being preserved from bad', at the same time.

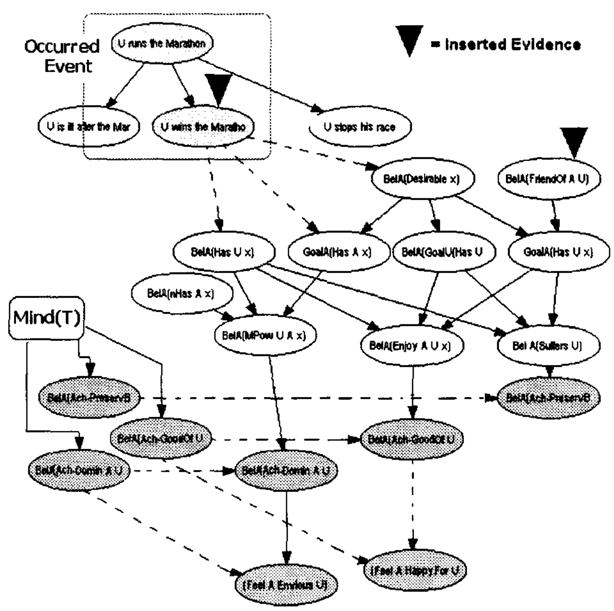


Figure 2: the DBN for envy and happy-for

Which 'fortune of others' emotions may mix up, and how? According to our model, *happy-for* and *envy* are examples of potentially 'coexisting' emotions: their generative mechanisms are different in that *the nodes that trigger them*, in the BN, *belong to two nonoverlapping sets*. For instance, the node Goal A (Has A x) does not affect achievement of the goal 'desiring the good of others'. So, Agents who are moderately altruistic and moderately dominant may be moderately envious and moderately happy-for at the same time, when they come to know that a desirable event occurred to a friend. For similar reasons, *sorry-for* and *gloating* may coexist, in particular circumstances. Does the same mixing metaphor apply to other emotion combinations, in this set? Probably not. The generative mechanisms of *happy-for* and *sorry-for*, for instance, are the same: as we said, they are triggered by a different value of desirability of the occurring event and, therefore, by *different values of common nodes* in the BN; therefore, they cannot coexist. However, it might happen that the same event *activates*, in two different time instants, *different kinds of beliefs*. Persons may see positive and negative consequences of the same event, at different times (*"he is going fast! Oh, but he seems to be very tired!"*): they will, in this case, switch more or less rapidly between *happy-for* and *sorry-for* emotions, according to how rapid is the activation of new beliefs about the consequences of the event. Switching in time may occur from **Prospect-based** emotions (*fear, hope*) to **Well being** (*distress, joy*) or **Confirmation** emotions (*disappointment, relief*). Belief about achievement or threatening of high-level goals ('Desiring the Good of Self' or 'Preserving Self from Bad') are involved in these cases. Switching from one emotion to another is due to a change in the probability of the belief that a (desirable or undesirable) event will occur, is occurring or occurred; this change may be due to observation of different consequences of this event at different times. In the Marathon example, switching from **fear** to **joy** to **relief** is closely related to the probability with which the runner believes she will win the race; this probability changes from time to time, according to the progress of the race. A 'microwave oven' metaphor may therefore be applied to mixing of emotions originating from the same belief observed at different time instants, in the three categories. The same happens for emotions within the Prospect-based, the Well-being or the Confirmation category, which cannot coexist but between which the Agent may switch. For instance, the Marathon runner was, in different time instants, **hopeful** to win the race and **fearful** about its possible consequences on her health. According to similar considerations, one may conjecture about mixing of emotions belonging to different categories: for instance, **happy** and **happy-for** may coexist, as well as **sorry-for** and **relief**, ... and so on.

5 Power of DBNs in Representing the Mixing of Emotions

DBNs enable representing the three kinds of situations that produce mixing of emotions:

- *concurrent triggering* of emotions occurs when an event has *different consequences* whose beliefs co-occur in the Agent's mind and when these beliefs influence, in their turn, the achievement of different but *compatible* goals[2];
- *switching* among different (and possibly contrasting) emotions occurs when the probability that the Agent attaches to a given belief changes with time or when new beliefs are activated in different time instants: these beliefs may be consequences of the same or of different events.

As our DBNs are cognitive models of the way *all* types of emotions are elicited, we do not make a distinction between primary and cognitively-generated emotions. However, we represent time decay of goals by introducing a link between two nodes representing the same goal at different time instants; the strength of this link may be varied, to simulate various forms of time decay. This enables us to formalise, as well, the case in *which coexistence is due to the fact that, after the same stimulus, emotions tend to take more or less time to disappear* and therefore may coexist for part of the time.

6 Related work

We are assisting to a flourishing of proposals on how to model emotion elicitation, expression and recognition: these methods all build upon psychological theories (Or-

[2] Compatibility of goals is defined according to psychological plausibility: for instance, it is infrequent to attach a *very high* importance to 'dominating others' and to 'desiring their good', at the same time. Medium weights to both goals might be, on the contrary, plausible.

tony et al, 1988; Elliott and Siegle, 1993; Oatley and Johnson-Laird, 1987, to cite only the main of them) and differ in the mathematical methods that are employed in formalising them. The main ingredients of emotion triggering models may be found in these theories. Oatley and Johnson-Laird suggest that "the functions of emotion modes are both to enable one priority to be exchanged for another in the system of multiple goals, and to maintain this priority until it is satisfied or abandoned" (Oatley and Johnson-Laird, 1986). Roseman establishes a relationship between 'appraisals', that are defined in terms of motivational and situational state, and emotion raising, and represents 'problem type expectedness' and 'control potential' in terms of beliefs (cited in Picard, 2000). Picard suggests that affective computational models should include a representation of 'response decay' and of the influence of personality and temperament.

As far as the methods for formalising these theories are concerned, two main approaches are prevailing: the first of them follows the line traced by Ortony in 1988, by proposing *mathematical functions* to combine a large number of (numerical) parameters in a measure of emotion intensity. In Affective Reasoner, Elliott lists the variables that may influence the intensity of an agent's affective state (Elliott and Siegle, 1993), without proposing any specific 'emotion-intensity calculation function'. The computational model of Affective Reasoner has been recently developed to formalise emotional reasoning in Emile (Gratch and Marsella, 2001). The main idea behind this system is to link emotion elicitation to appraisal of the state of plans in memory rather than directly to events. Emotion intensities are measured as a function of the goal importance and of the probability of goal attainment which is evaluated, in its turn, by an analysis of the plans that could bring that goal about. In Prendinger et al (in press), this function is a logarithmic combination of the sum of exponentials of a number of variables: the intensity of *joy* depends on the desirability of the triggering event; the intensity of *happy-for* is a function of the Interlocutor's happiness and of the 'degree of positive attitude' of the Agent towards the Interlocutor, and so on. If the number of variables is low, the function's value is only influenced by the 'strongest' variable. Personality does not influence emotion elicitation but only the 'filtering' process, in which it is decided whether the emotion should be expressed or suppressed. The main limit of this method is, in our view, in the combination of several heterogeneous measures in a unique function: when different scales are employed, the effect of their combination and the relative importance of each of them cannot be foreseen precisely; in addition, the range of variability of the function may go outside the established range of the variable it should measure, with the consequent need of 'normalising' the obtained value or of 'cutting' values out of this range.

The second approach to affective modeling is based on representing the cognitive aspects of emotion triggering. Belief Networks (BNs) seems to be a ideal formalism, in this case. BNs have first been applied to model expression and recognition of emotions from the Agent's external expression (voice, face, etc: Ball and Breese, 2000). DBNs have the additional advantage of enabling representation of the dynamic aspects of this phenomenon. They have been applied, recently, to model affective aspects of the User's mind in computer games (Conati, in press).

Our modeling method follows this second line of research. The main difference from the cited experiences is in the 'grain size' of knowledge represented. Rather than representing directly the effect of events on the Agent's mind, we build a fine-grained cognitive structure in which the appraisal of what happens in the environment is represented in terms of its effects on the Agent's system of beliefs and goals. We only employ two measures (uncertainty of beliefs and utility of goals), while Elliott proposes a different measure for each variable. However, there is a close relationship between the aspects of the phenomenon that are represented in the two cases. Elliott's 'importance of achieving a goal/not having a goal blocked' corresponds to the 'weights' we attach to goal achievement or threatening. All variables associated with our 'belief' nodes correspond to Elliott's 'simulation event' variables (for instance, the 'appealingness of a situation') and 'stable relationship variables' (for instance, 'how desirable an event is', 'how friend is the Agent with her Interlocutor'). Rather than attaching to them an integer value in some established scale (as Elliott proposes), we discretize these variables in a limited number of values and attach a probability distribution to each of them. In addition, by attaching conditional probability distributions to the network's links, we define a strength of these links in probabilistic terms.

7 Final remarks

We discusseded, in this short paper, the ability of our formalism in representing the way emotions may mix. Other advantages of this formalism are in the possibility of representing the following aspects of emotion elicitation:

- *temperament and personality*: how these factors affect the Agent's propensity to feel and to show emotions, by influencing the weights it assigns to achieving its goals;
- *social context:* how the Agent's relationship with the context influences emotion triggering by influencing its beliefs and goals;
- *dynamics of the Agent's state:* how the Agent's affective state changes over time, either as a consequence of 'endogenous' or of 'exogenous' factors;
- *response decay:* how emotion felt by the Agent decays, in absence of any new specific stimulus.

Of course, a model is always a shallow representation of a complex reality: the historical evolution of models in physics is a proof of the progressive advancement of knowledge in that domain. Refinement of a model requires, however, comparison with accurate experimental data in addition to a progress of mathematical techniques. This is true also for the domain of emotion modeling: in particular, refining parameters in BNs and the way they change with the context requires testing the model in well defined settings (which is probably of the same difficulty of experiments in modern physics). We plan to revise and refine the structure of our DBNs and their parameters through a systematic simulation of the kind of dialogs we want our Artificial Agent to make and through an evaluation of the 'believability' of these dialogs. More in general, if interest towards affective aspects of Human-Computer Interaction is not a transient reality, we will hopefully assist to a progressive refinement of these models, through an interleaving of theory formulation and experimental validation.

References

G Ball and J Breese: Emotion and personality in a conversational agent. In S Prevost, J Cassell, J Sullivan and E Churchill (Eds): Embodied Conversational Agents. MIT Press, 2000.

C Conati: Probabilistic assessment of user's emotions in educational games. *Submitted to Applied Artificial Intelligence.*

B De Carolis, C.Pelachaud, F.de Rosis and V.Carofiglio: Interactive Information Presentation by an Embodied Animated Agent. *IPNMD, Trento, 2001.*

F de Rosis, C Pelachaud, I Poggi, V Carofiglio and B De Carolis: From Greta's mind to her face: modeling the dynamics of affective states in a Conversational Embodied Agent. *Submitted to IJHCS.*

C.Elliott and G. Siegle: Variables influencing the intensity of simulated affective states. *Proceedings of the AAAI Spring Symposium on Mental States '93, 58-67, 1993.*

J.Gratch and S. Marsella: Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents. *Proceedings of the 5th International Conference on Autonomous Agents, Montreal, Canada, June 2001.*

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

A.Ortony, G.L. Clore and A. Collins: The cognitive structure of emotions. *Cambridge University Press, 1988.*

A.Ortony: On making believable emotional agents believable. In R. Trappl and P. Petta (Eds): Emotions in humans and artifacts. *MIT Press, Cambridge, MA, in press.*

H Prendinger, S Descamps and M Ishizuka: Scripting affective comunication with life.like characters. *Submitted to Applied Artificial Intelligence.*

A.J Pearl: Probabilistic reasoning in intelligent systems: Networks of plausible inference. *Morgan Kaufman Publishers, 2000.*

R Picard: Affective Computing. *The MIT Press, 2000.*

Expressing Emotion Through Body Movement: A Component Process Approach

Mark Coulson
School of Health and Social Sciences
Middlesex University
Queensway
Enfield EN3 4SF
United Kingdom
m.coulson@mdx.ac.uk

Abstract

This paper outlines a computational model which generates animated figures capable of expressing emotion through movements of the musculo-skeletal system. The model is an implementation of an appraisal-based theory of emotion (Scherer, 1987, 2001) which views emotion and emotional expression as arising independently and simultaneously from the outcome of a series of appraisals, called Stimulus Evaluation Checks (SECs). Expressions are the temporal unfolding of functionally significant consequences of a series of appraisals which simultaneously give rise to emotional experience. The strengths of such an approach and the nature of the implementation, along with the challenges it presents are described.

1 Introduction

The human body represents a potentially important medium for the expression of emotion. Bodies are large, infinitely poseable and easily perceived, yet have not received the same degree of attention that facial and vocal expression have attracted, and remain the 'poor relation' of emotional expression and communication. The understanding of bodily expression of emotion is limited despite the obvious uses of animated characters for instruction, communication and entertainment.

There are several reasons for this lack of knowledge and research. First, bodies are complex objects which exhibit a large number of degrees of freedom (defined as joint rotations about one or more axes) and an enormous variety of postures and movements. The systematic investigation of the resulting high dimensional space is a considerable challenge, and although several attempts at dimension reduction have been made (see for example Montepare et. al.'s 1999, dimensions of form, tempo, force and direction), none of these has been related to actual joint rotations. Hence there exists a gap between perceptual gestalts and actual bodily configurations or patterns of movement which hampers attempts at simulation.

Second, the primary functions of bodies are locomotion and manipulation rather than expression. Despite being capable of expressing emotion, bodies have not evolved to do so. Whereas movements of the facial musculature may be directly interpreted as expressive, perception of the body's ability to express emotion is likely to be indirect. Movements are functional actions and responses to stimuli and situations which may (or may not) have emotional consequences. The affective content of a sequence of body movements is therefore perceived as epiphenomenal rather than a direct expression of the emotion. Also, as movements which may be considered expressive are embedded within ambulatory or manipulative ones, the signal to noise ratio of bodily expression of emotion is poor relative to facial and vocal channels.

Third, there is a great deal of individual variation in the way in which people express and perceive emotion from body movement (Wallbott, 1998). While there is considerable evidence that the vast majority of people produce very similar facial expressions when experiencing emotion, the same has not generally been found for body posture.

Fourth, 'body language' is generally seen as more culturally distinct than universal, and as such the meaning of postures and movement is held to vary across cultures to a degree which makes universal recognition unlikely. In the face of tremendous success in the field of universal recognition of facial expression, body movement appears to be a less attractive domain for simulation.

Despite these challenges, the simulation of expressive movements has application to a wide variety of fields and represents a worthy pursuit. The next section outlines theories of emotion, and identifies how adopting a particular perspective may overcome some of the difficulties arising from individual and cultural variation in expression. Following the theoretical discussion I outline the nature of the implementation and conclude by discussing challenges and outstanding issues.

2 A theoretical basis for emotional expression

A theory of emotion is a necessary pre-requisite in any attempt to simulate emotional expression, although as will be seen the theoretical position adopted here permits emotional expressions to be simulated without the necessity for identifying an agent's emotion in terms of lexical tokens. Three general classes of emotion theory are briefly outlined, following which a more detailed account of the approach adopted here is provided. Although the descriptions omit a great deal of detail, they capture the essence of the competing theoretical approaches. In each, it is assumed that the nature, experience and expression of emotion are determined by the same set of theoretical constructs.

2.1 The classical view of emotions

Emotions are qualitatively different states which are adequately described by linguistic tokens (fear, anger, sadness etc.). Phenomenological experience, physiological response, action readiness and expressive behaviours are closely associated as patterns of response to stimuli or events. Emotions are associated with specific expressions, although these may be modified or inhibited according to personal and cultural standards. This position is most closely associated with researchers such as Paul Ekman (e.g. Ekman, 1993) who in a series of studies over the past 30 years has convincingly demonstrated that facial expressions for at least six 'basic' emotions (anger, disgust, fear, happiness, sadness and surprise) are recognised at similarly high levels in a wide variety of cultures.

2.2 The dimensional view of emotions

Emotions result from the combination of two or more orthogonal dimensions. The first two are generally identified with valence and arousal (Reisenzein, 1994), or rotated through 45 degrees and labelled positive and negative affect (Watson, Wiese, Vaidya, & Tellegen, 1999). Both variants result in a circumplex structure, with emotion tokens occupying locations on the circumference of a circle bisected by the two dimensions. Emotion tokens are still valid descriptions of emotion, but do not represent qualitatively different categories or states, being instead varying combinations of an underlying set of dimensions.

2.3 The component process view of emotions

Emotions are the end result, and a phenomenological by-product, of a linear and cyclic series of cognitive appraisals. The appraisal system continuously evaluates the environment in terms of its significance for the organism's ongoing activity. Appraisal outcomes have immediate and direct effects on all aspects of the organism, and the temporal unfolding of these gives rise to expression. Appraisal outcomes therefore determine expression and emotion separately, and although certain aspects of emotion are captured by linguistic tokens, these are only rough approximations. What follows is a description of an implemented component process approach to the bodily expression of emotion.

3 Applying a component process approach

The most carefully thought through version of the component process approach is that of Klaus Scherer (1987, 2001). Under Scherer's model there are four main objectives of the appraisal system (Relevance, Implications, Coping, and Normative Standards) each of which is further subdivided into specific Stimulus Evaluation Checks (SECs). The evaluation proceeds by processing each SEC in turn, cyclically, so that once the sequence finishes it starts again from the beginning, and both emotional experience and overt behaviour and expression arise independently from this sequence (see figure 1).

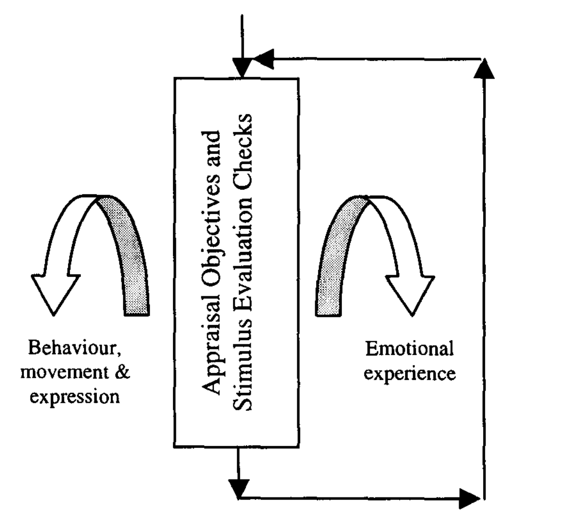


Figure 1: The linear and cyclic process of stimulus evaluation

As an example, consider the appraisal of a sudden loud noise in the middle of the night (see figure 2 for a graphical account of the following description).

According to Scherer's theory, the three SECs covering the first appraisal objective, Relevance, would process the stimulus as novel, intrinsically unpleasant, and relevant to ongoing activity (Novelty check, Intrinsic Pleasantness check, Goal Relevance check). The Implications objective would appraise the stimulus as being caused by an external agent or event, with an as yet unknown likelihood that action will be required (Causal Attribution check, Outcome Probability check), and being discrepant from expectations, hindering rather than helping, and possibly urgent (Expectation discrepancy check, Goal/need conduciveness check, Urgency check).

The Coping Potential SECs classify the stimulus as being potentially controllable, and the individual may or may not perceive him/herself as having the power to enforce this control (Control check, Power check). He/she may also feel capable or otherwise of adjusting to the stimulus without further action (Adjustment check). Finally, the Normative Standards checks determine whether the stimulus violates expected codes of personal (internal) or external behavior. The cumulative effect of all the checks would be to yield an emotion which one might describe as fear, but whose expression has developed along with the feeling, and cannot therefore be said to have been caused by it, or even to be an expression of it

Several important points arise from this example. First, emotion arises as the cumulative and interactive effect of all SECs. Second, there is no point at which the cycle is complete - stimuli are constantly re-evaluated and initial outcomes may be revised as more information is made available. Third, not all SECs necessarily contribute to all emotions (in the example given above, violations of Normative Standards are probably not relevant to fear responses). Finally, although an emotion profile may leave certain SECs open, these may still affect the nature of the emotion. For instance, the Adjustment SEC is open for 'fear' in figure 2, but a high value for this might reflect adaptive fear (scared, but able to do something about it) whereas a low value might reflect blind terror. The lexical token 'fear' captures the essence of the experience, but the full account lies only in the pattern of SEC outcomes.

4 Implementing a component process approach

The result of each SEC may have an immediate and functional effect on all aspects of the organism as it responds to the demands of the appraisal. In terms of body movement, a specific SEC outcome may specify a particular posture which is a response to the situation as so far appraised. For instance a positive novelty check results in an orienting response - weight transfer is neutral, head and upper torso turn to face the stimulus. This response is a direct effect of the appraisal, in this example a positive novelty check indicates there is something new and potentially important in the environment which should be attended to.

As the sequence unfolds, the appropriate posture will change, and the body will move to achieve this. It is assumed that the speed of appraisal is generally faster than the body's rate of movement, so the target posture may not be attained before being altered by subsequent SEC outcomes. The result of this is a continuously moving body trying to keep up with the demands of the incoming stream of appraisal outcomes.

From an implementation point of view the task is to identify those SEC outcomes which lead to functionally significant postures, define these postures in terms of a set of joint rotations, and code the expressive response as a dynamic series of transitions between postures. Defining the postures is achieved by analysing the demands made on an individual by the sum of all possible SEC outcomes up to a specific point in the sequence. These demands reflect cognitive/informational processes (e.g. leaning backwards widens the perceptual field, leaning forwards narrows it) and simple affective ones (e.g. positively valenced outcomes result in approach movements, negative ones in withdrawal). Some of the challenges associated with creating adequate postural versions of such demands analyses are addressed in section 6.

5 Details of the implementation

All SECs are input to the program by the user or read in as previously defined patterns. The pattern of outcomes may be designed to represent a specific emotion, or may be any arbitrary set. The time at which each SEC triggers may also be specified, with the restriction that SECs must occur in linear sequence. Scherer argues that each SEC must reach 'preliminary closure' before the next SEC can begin processing the stimulus, so some flexibility in terms of when each SEC triggers is required.

Certain SECs have scalar values (e.g. Expectation Discrepancy, Goal Relevance, Urgency) whereas others are qualitative categories (e.g. Causal Attribution, which is defined in terms of Ortony, Clore & Collins' 1988, scheme whereby causes are either the self, another agent, or an event or object). The magnitude of a scalar value may influence the nature of the posture and may also affect the speed with which the figure moves. For instance, a highly unpleasant and discrepant stimulus is likely to result in a more rapid response than one which is only mildly unpleasant and slightly discrepant. SEC outcomes can therefore qualitatively specify a new posture towards which the body will move and/or quantitatively affect the speed of movement.

Target postures are defined in terms of a 17-segment model described by 54 variables representing rotations about the major joints as well as the position of the hips. The BVH format is used to represent figures (see www.biovision.com/bvh.html for details) as this is simple to manipulate and can be read into Curious Labs' POSER, the package used to generate animations. In this format, the hips form the top level of a hierarchy with all other segments attached directly or indirectly. Once the three dimensional location of the hips has been defined, the positions of the other segments are a function of this and the specific joint rotations affecting each segment. For example, the position of a hand is defined by the location of the hips plus rotations about the waist (twist and bend), chest (bend and lean), shoulder (twist, adduct/abduct, forwards/backwards), and elbow (twist and bend).

Once the pattern and timing of SECs has been defined and parameters describing the target postures for all SEC outcomes read in, the program generates 100 frames of animation representing smooth interpolations across the target postures defined by the SEC outcomes. The output is then imported into POSER, and animations produced from this.

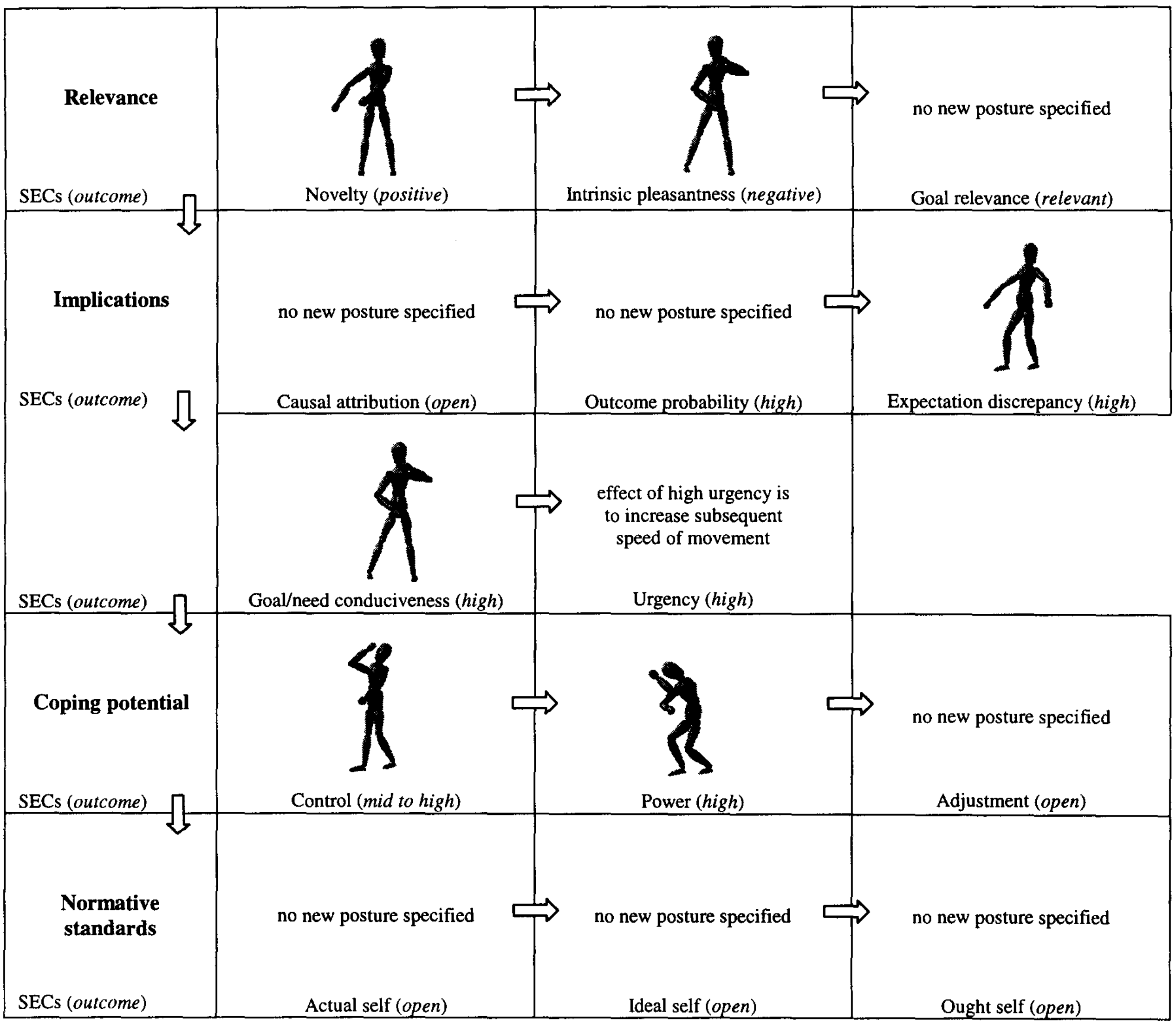


Figure 2: A series of target postures for a sequence of appraisals leading to fear. Arrows indicate the progress of the appraisal. Possible values of each SEC are described in Section 3. The posture indicated in each box is that towards which the body will move following preliminary closure of the SEC - as the process of appraisal is faster than movement, these postures will not always be attained and the diagram provides only an indication of the animation produced by this sequence of appraisals.

6 Challenges

The current version of the implementation provides a starting point for the evaluation of both the component process approach and the effectiveness of the modelling. Data from human participants collected in our lab suggest that the movements as currently modelled do not communicate emotion to the same level as has been found in studies using video or point light displays of actors expressing different emotions (de Meijer, 1989; Montepare, Koff, Zaitchik, & Albert, 1999; Schouwstra & Hoogstraten, 1995; Wallbott, 1998), and that much work is still required. Three main challenges present, and each is examined below.

6.1 Specification of functionally significant postures resulting from SEC outcomes

The exact posture specified by each SEC outcome is determined in terms of its functional significance, and a given series of SEC outcomes will always specify the same posture. There is a sense in which this is a limitation in the model's flexibility – a single specific posture is never the only functionally appropriate response to an appraisal. However, the ever-changing nature of the response to a stimulus means that an approximate identification is probably sufficient, and this loss of flexibility may not introduce any significant inaccuracies into the simulations.

More difficult is the production of appropriate postures for SEC outcomes further down the chain of appraisal. To the extent that responses to later appraisals depend on earlier ones, the number of potential postures propagates. While this may reflect the complexity and flexibility of normal response, it requires careful analysis and the continual refinement of postures. The current implementation includes 24 target postures arising from the 14 SECs, but the number of postures is likely to increase as evaluation proceeds.

6.2 Transitions between states - realistic movement and inverse kinematics

Currently all transitions between successive postures are handled by smoothly interpolating joint rotations from current to target posture across a set number of frames. While this offers a simple means of creating animations and experimental stimuli, it is a simplification for two reasons. First, speed of movement as defined by angular velocity about an axis of rotation does not remain constant across a typical movement. Onset tends to be rapid, and offset more gentle. A more realistic model will need to take this into account.

Second, transfer of the mass centre is generally achieved by lifting the foot and stepping forwards or backwards. This action is relatively slow to initiate and may arise as a consequence of other movements which affect the body's balance as opposed to being an integral part of them as currently modelled. Sequencing movements and modelling the nature of weight transfer more realistically will clearly add to the naturalness of the animations.

6.3 Adequacy of theoretical assumptions

The component process approach is an ambitious and complex attempt to explain the cognitive nature of emotion, and repeated testing of its predictions and assumptions can only strengthen the model. The approach has been shown to have utility in studies of facial (Wehrle, Kaiser, Schmidt, & Scherer, 2000) and vocal (Banse & Scherer, 1996) expression, and the framework has been continually refined and developed over the past 15 years. Its detail and complexity offers an excellent basis for the investigation of emotional expression through body movement.

Conversely, the process of simulation offers an ideal means of testing the adequacy of theoretical assumptions. The current implementation models the Causal Attribution SEC in terms of Ortony et. al.'s analysis of the cognitive structure of emotions (see above), as this SEC is somewhat underspecified in the current version of the theory. Additionally, Normative Standards are modelled as arising from three models of the self; the actual, the ideal and the ought self, as outlined by Higgins (1987). Rather than distinguish between internal and external normative standards as is the case in the current version of Scherer's theory, the model views standards as equalling, exceeding, or falling short of the actual, ideal and ought selves. Higgins has argued that patterns of discrepancy between the different conceptions give rise to a variety of emotional responses.

7 Conclusions

The strength of the component process approach lies in its definition of expression in terms of an individual's evaluation of a situation as opposed to a direct effect of an imprecisely defined lexical token. Expressions can therefore be simulated as a direct consequence of cognitive appraisals, and although this rests upon the theoretical view of emotion outlined above, the generation of animated affective characters does not need to simultaneously model those characters' actual emotions. If a stimulus or event can be described in terms of the SEC outcomes it will provoke, then an emotionally expressive response can theoretically be generated. Modelling these SEC outcomes may be a simpler task than modelling emotions. Indeed, removing emotion words from the analysis of expression may remove an artificial constraint on the degree to which expressions across all channels appear to be individually and culturally dependent. As Wierzbicka (1992) has persuasively argued, the use of emotion words to analyse emotion may be seriously misguided, and an approach based on utilizing fundamental components such as the SECs described by Scherer offers a way in which this may be avoided.

Ongoing development of the implementation is focussing on the practical, biomechanical and theoretical aspects outlined above in addition to testing human participants' attributions of emotion to the model's output.

Acknowledgements

I should like to acknowledge the assistance and advice of Dr Rebecca Gould and Stephen Nunn in the preparation and delivery of this paper.

References

Banse, R., & Scherer, K. R. (1996). Acoustic profiles in vocal emotion expression. *Journal of Personality and Social Psychology, 70*(3), 614-636.

Ekman, P. (1993). Facial expression and emotion. *American Psychologist,, 48*, 384-392.

Higgins, E. T. (1987). Self-discrepancy: A theory relating self and affect. *Psychological Review, 94*, 319-340.

de Meijer, M. (1989). The contribution of general features of body movement to the attribution of emotions. *Journal of Nonverbal Behavior, 13*, 247-268.

Montepare, J., Koff, E., Zaitchik, D., & Albert, M. (1999). The use of body movements and gestures as cues to emotions in younger and older adults. *Journal of Nonverbal Behavior, 23*(2), 133-152.

Ortony, A., Clore, G. L., & Collins, A. (1988). *The cognitive structure of emotions*. Cambridge: Cambridge University Press.

Reisenzein, R. (1994). Pleasure-arousal theory and the intensity of emotions. *Journal of Personality and Social Psychology, 67*, 525-539.

Scherer, K. R. (1987). Toward a dynamic theory of emotion: The component process model of affective states. *Unpublished manuscript, University of Geneva.*

Scherer, K. R. (2001). Appraisal considered as a process of multi-level sequential checking. In K. R. Scherer, A. Schorr, & T. Johnstone (Eds.), *Appraisal processes in emotion: Theory, Methods, Research* . New York & Oxford: Oxford University Press.

Schouwstra, S. J., & Hoogstraten, J. (1995). Head position and spinal position as determinants of perceived emotional state. *Perceptual and Motor Skills, 81*, 673-674.

Wallbott, H. G. (1998). Bodily expression of emotion. *European Journal of Social Psychology, 28*(6), 879-896.

Watson, D., Wiese, D., Vaidya, J., & Tellegen, A. (1999). The two general activation systems of affect: Structural findings, evolutionary considerations, and psychobiological evidence. *Journal of Personality and Social Psychology, 76*(5), 820-830.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. *Journal of Personality and Social Psychology, 78*(1), 105-119.

Wierzbicka, A. (1992). Talking about emotions: Semantics, culture, and cognition. *Cognition and Emotion, 6*, 285-319.

Do TV's Listen Better than Dogs?

The Effects of an Animated Character as an Interface Intermediary for Voice-Controlled Interaction

Elmo M.A. Diederiks

Philips Research Laboratories; Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands

elmo.diederiks@philips.com

Abstract

An animated dog, named Bello, has been proposed as a concept to facilitate voice control. Such an on-screen character can help to make the user-system interaction more social. It can prevent people from feeling uncomfortable while talking to a system and it can act as a clear focus point on the screen. Furthermore, the recognition capabilities of a dog are comparable to those of current voice control systems. In order to validate these assumptions, the Bello concept has been compared with an animated character in the form of a butler and a concept without an animated character. The results of this experiment indicate that an animated character is not essential, but the Bello concept helps to provide a better first-time experience: people feel less silly. Bello is furthermore perceived as more reliable after a period of use. However, the appropriate form of an animated character seems be dependent on the application.

1 Introduction

Voice control is an increasingly feasible and important input modality for many products. Speech dictation programs and voice-dial for cell phones are application examples. However, there are still basic problems in voice controlled user-system interaction. Many people feel uncomfortable and less self-assured in using voice control. Moreover, people often do not know what they can say and how they should say it. Additionally, they do not always have a clear notion of the system states and thus can loose track of the status of the interaction. Finally, voice control systems will always make recognition errors, which could lower the level of acceptance of these systems (Kamm & Helander 1997).

Animated characters are increasingly used in user interfaces. Research is ongoing with respect to what the beneficial applications of animated characters are, where they can make the interaction more pleasing and easy, and what their effects are on user-system interaction (for example Cassell & Bickmore 2001; Koda & Maes 1996; Lester et al. 1997; but also see Dehn & van Mulken 2000 for an overview). This paper briefly explains how an animated dog can make it more natural to control a system by voice. Moreover, it describes an experiment in which the added value of animated characters in this new application domain has been evaluated.

2 The Bello Concept

An animated dog, named *Bello*, has been developed to facilitate voice-controlled interaction for screen-based systems, such as a television set. The concept builds upon the conversation metaphor (Hutchins 1989), in the sense that Bello acts as intermediary between system and user. Users merely need to command Bello what it should do, for example "Bello, start CNN" or "Switch off". In order to provide appropriate feedback to the user, animations with sound have been designed for several states and state transitions of the voice control system. Figure 1 gives an impression of the animations.

Figure 1: A few different expressions of Bello.

During normal operation Bello is not visible on the screen. When summoned Bello appears at the bottom of the screen to indicate that the system is listening. Bello is accompanied by textual feedback on the recognised commands (see Figure 2). If no command has been recognised for some time Bello slowly falls asleep and then automatically disappears from the screen. Bello can also explicitly be removed from the screen by saying "Down".

Figure 2: Bello with textual feedback.

2.1 The benefits of the concept

It has formerly been argued that the Bello concept has a number of anticipated benefits (Diederiks & van de Sluis 2001). For clarification purposes, the evaluated benefits are summarised in the following paragraphs. The first four are benefits of an animated character in general, while the last two advantages are specific for the metaphor of a dog.

2.1.1 Comfort
People realise that a system or device is not a living creature and talking to it seems to be a bit silly. Providing the system with a face by means of an animated character can compensate for this mismatch between interaction modality and the interaction partner.

2.1.2 Focus point
An animated character at the edge of the screen acts as clear focus point for attention and interaction, both because of its motion and its social nature (Reeves & Nass 1996; Fogg & Tseng 1999).

2.1.3 System behaviour
An animated character can enhance the social dimension of the interaction and communicate the system feedback in a comprehensible manner. The interaction can be based on social rules by means of expressive behaviour that is known from daily life (Reeves & Nass 1996).

2.1.4 Fun
A system for the home should be easy to use, but it should also be fun! An animated character can make the interaction with such a system more enjoyable.

2.1.5 Use of commands
Dogs are known to respond to a small set of simple voice commands, rather then to full sentences. These recognition capabilities of a dog are similar to those of current *command and control* voice recognition systems.

2.1.6 Error acceptance
A dog does not always respond perfectly to commands. Current voice control systems also do not work error free. Where technical systems are expected to work perfectly, dogs are allowed to make mistakes.

3 The experiment
In a Wizard-of-Oz experiment it was *explored* how users perceive the above-mentioned anticipated benefits, both initially and after a brief period of use. This was chosen to explore possible habituation effects. For example, fun can be more important at first time use, while efficiency can be more important if people are more experienced in using a system. The Wizard-of-Oz approach allowed control of the number (and kind) of errors.

The Bello concept was compared with two other concepts: the *Plain* and the *Jeeves* concept. All three concepts provide textual feedback on what was recognised by the system.

The Plain concept does not make use of an animated character. It provides basic visual and non-speech audio feedback in the form of simple electronic chimes.

Figure 3: The Plain concept with textual feedback.

The Jeeves concept deploys an animated character in the form of a butler. This metaphor was chosen because it fits the concept of intermediary reasonably well. The concept provides feedback by means of animations, accompanied by pre-recorded speech. Jeeves should have the advantages of using an animated character.

Figure 4: Jeeves with textual feedback.

The Bello concept should have the benefits both of using an animated and of using the metaphor of a dog.

3.1 Method
3.1.1 Participants
36 Philips employees, aged between 20 and 55 *(M=33.6, SD=9.6)*, participated in the experiment. 12 subjects had a university level of education, 14 a middle level of education and 10 a lower level of education. Half of the subjects were male and half of them were female.

3.1.2 Manipulation
The concept used (Plain, Jeeves or Bello) was manipulated as a between-subject variable, such that each of the three concepts was used by a total of 12 subjects. Gender, age and education level were distributed evenly over the three concepts.

Recognition errors were generated on purpose such that the same errors occurred at the same moment for all subjects. The number of errors was based on experiences with a speech recognition system (speaker independent with 25% error rate).

3.1.3 Measures
Due to the lack of a useful validated instrument, a custom questionnaire was used. The subjects were asked to indicate the level to which they agreed with various statements on a 7-point scale (1='I totally disagree', 7='I totally agree'). Four statements were used for each anticipated benefit.

3.1.4 Apparatus
A voice control television set was simulated using a Pentium II Windows PC and a 27-inch Philips SVGA television used at a resolution of 800x600 pixels. Five television channels with two programs each were simulated using a still-screen grab and an audio grab. Such a still image could bias the perception of the subjects, but it was assumed that this bias would be on average equal for the three concepts.

3.1.5 Procedure
The subjects were recruited in the premises of the test room, so it would be minimal effort to participate. One week before the start of the experiment an email with a basic explanation was sent as a reminder. To minimise cross-participant influence, the subjects were urged not to talk about the experiment to anyone for the duration of the experiment.

Each subject worked with one of the three concepts for five consecutive days (Monday to Friday) and had to perform 6 to 8 different basic tasks every day. The tasks were provided on a piece of paper with some basic instructions. The tasks were the same for all subjects and took about 5 minutes to complete. To flatten the learning curve, the required command words were provided along side every task. Additionally, all the available command words were presented on a card situated on top of the television set. On Monday and Friday they were asked to fill in the questionnaire, which took on average an additional 10 minutes. After completing the second questionnaire the subjects were given a small gift as a surprise in appreciation for their participation.

The subjects were seated in a chair at 2 metres distance from the television set. The experiment facilitator performed the speech recognition from behind a curtain by simply listening to the commands uttered by the subjects and subsequently pressing the right button on the keyboard of the PC. This resulted in the required system behaviour of the television set. The subjects were (kept) unaware of this Wizard-of-Oz approach.

3.2 Results

The responses were analysed using ANOVA's ($\alpha=0.05$). These have been conducted for the measures of Monday, the measures of Friday and for the difference between these two sets of measures. The homogeneity of variance has been tested with Levene's tests. The normality has been analysed using Boxplots. The results should be interpreted with special care if the variance is not homogeneous or if the distribution is not normal.

Due to the limitations of the measurement instrument, the results cannot be used to falsify the anticipated benefits. They however *do* provide a good indication.

3.2.1 Comfort

Initially, Bello makes people feel significantly less silly *(Monday: $M=2.58$, $SD=1.08$)* compared to interaction without Bello *(Monday: $M=3.71$, $SD=1.49$), ($F(1,34)=5.39$, $p=.026$)*. After the five days this difference is levelled out and people do not feel silly using voice control independent of the concept used (Figure 5).

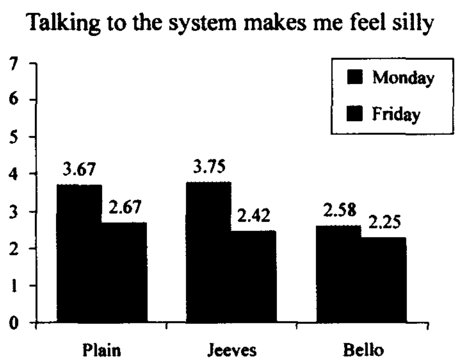


Figure 5: The mean scores of the statement "Talking to the system makes me feel silly".

3.2.2 Focus Point

The concept has a significant influence on whether people feel that they are interacting with an entity rather than with the television itself *(Monday: $F(2,33)=13.48$, $p=.000$; Friday: $F(2,33)=12.54$, $p=.000$)*. This difference is mainly caused by the Plain concept *(Monday: $M=1.92$, $SD=1.16$; Friday: $M=1.67$, $SD=0.78$)*, which scores significantly lower compared to the other two concepts *(Monday: $M=4.67$, $SD=1.83$; Friday: $M=4.58$, $SD=2.10$), (Monday: $F(1,34)=22.30$, $p=.000$; Friday: $F(1,34)=21.33$, $p=.000$)*. However, the assumption for homogeneity of variance does not hold.

3.2.3 System Behaviour

The behaviour of Jeeves initially has a trend to be less acceptable *(Monday: $M=4.75$, $SD=1.76$)* compared to the Plain and Bello concepts *(Monday: $M=5.58$, $SD=0.83$), ($F(1,34)=3.77$, $p=.060$)*. In the end, the three concepts were considered to behave equally acceptable.

3.2.4 Fun

In general people find it enjoyable to interact with the system. After a short period of use the Jeeves concept trends to be perceived as more inefficient *(Friday: $M=3.33$, $SD=1.50$)* than the other two concepts *(Friday: $M=2.50$, $SD=1.02$), ($F(1,34)=3.88$, $p=.057$)*.

3.2.5 Use of Commands

People prefer to use short commands in case of all three concepts. The Bello concept scores slightly higher whilst the Jeeves concept scores slightly lower. The concept used does also not influence whether people expect that only short commands can be used.

3.2.6 Error Acceptance

The Monday measures do not show a difference between the concepts with respect to the perceived reliability of the system. On Friday the system with the Bello concept is perceived significantly more reliable *(Friday: $M=2.33$, $SD=1.15$)* than the other two concepts *(Friday: $M=3.42$, $SD=1.28$), ($F(1,34)=6.08$, $p=.019$)* (Figure 6).

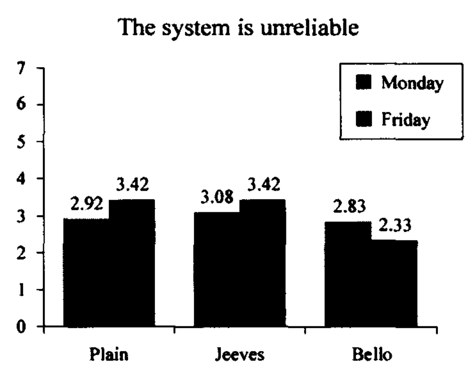


Figure 6: The mean scores of the statement "The system is unreliable".
(Lower scores indicate a higher perceived reliability).

The perceived number of errors is the same for all three concepts. These system errors are furthermore perceived as acceptable, independent of the concept used.

4 Discussion

People feel comfortable while using voice control, which is in line with a recent field test of a voice-controlled CE-device at our research laboratories. Initially, however, people feel less silly using the Bello concept. The Jeeves concept was expected to display the same effect. This (lack of) effect could be induced by the behaviour of Jeeves initially being perceived as less acceptable, suggesting that the more human appearance of Jeeves raises too high expectations. It seems that the metaphor of a dog fits the system behaviour better than a butler does, which *is* in line with the expectations.

The concept has a strong effect on whether people feel they are interacting with a different entity than the television itself, as was anticipated.

People think it is fun to use voice control independent of the concept used. It is likely that voice control is rather new to them and therefore exciting.

People *prefer* to use short commands to interact with the system, independent of the concept used. The use of an animated character or specifically the form of a dog does not influence this, nor does it influence whether people *expect* the system to listen to short commands. It was however anticipated that the use of commands would be more appropriate in case of the Bello concept. The preference to use commands seems to eliminate the effect of the (in)appropriateness of using commands.

People perceive the same number of errors, and find these errors acceptable, independent of the concept used. Despite these findings, Bello is, after a period of use, perceived as more reliable than the other two concepts that are perceived equally reliable. This also supports the assumption that the metaphor of a dog is better suited as animated character to facilitate voice control by commands. A more human form (e.g. Jeeves) would perhaps better fit a natural language recognition system.

5 Conclusion

This paper has briefly described how an animated character could facilitate voice-controlled interaction and how the metaphor of a dog could help to create realistic expectations about the capabilities of current voice control systems. The advantages of using an animated dog have been evaluated in an experiment.

The results of this experiment indicate that an animated character is not indispensable for voice-controlled interaction. However, people feel less silly using the Bello concept if they use voice control for the first time. Such an initial effect can be quite helpful in lowering the threshold to start using voice control.

The behaviour of Jeeves was initially perceived as less acceptable, whilst the Bello concept is perceived as more reliable after a period of use. The appropriate shape of the animated character seems dependent on the application. A wrong metaphor can easily neutralise the advantages of using an animated character or even have a negative effect. Due to this application domain dependence the endeavour to standardise tools or methodology, as advocated by Dehn & van Mulken (2000), may prove to be more difficult than anticipated.

TV's might listen better than dogs. However, interacting with a dog by voice is for the moment more natural. An animated character in the shape of a dog can indeed facilitate voice-controlled interaction: it helps to provide a better first time experience and to make the system appear more reliable.

Acknowledgments

The Bello concept has been co-developed with Richard van de Sluis. The experiment set-up and the analysis of the results have been greatly supported by Gerard Hollemans and Boris de Ruyter.

References

J. Casell, T. Bickmore. Relational Agents: A Model and Implementation of Building User Trust. *CHI '01 Conference Proceedings*, ACM Press, pp. 396-403, 2001.

D.M. Dehn, S. van Mulken. The impact of animated interface agents: a review of empirical research. *International Journal of Human-Computer Studies*, 52, pp. 1-22, 2000.

E. Diederiks, R. van de Sluis. Bello, An Animated Character facilitating Voice Control. M. Hirose (ed.), *Human-Computer Interaction INTERACT '01*, IOS Press, pp. 686-687, 2001.

B.J. Fogg, H. Tseng. The Elements of Computer Credibility. *CHI '99 Conference Proceedings*, ACM Press, pp. 80-87, 1999.

E. Hutchins. Metaphors for Interface Design. M. Taylor, F. Néel, D. Bouwhuis (eds.), *The Structure of Multimodal Dialogue*, Elsevier Science Publishers, pp. 11-28, 1989.

C. Kamm, M. Helander. Design Issues for Interfaces Using Voice Input. M. Helander, T.K. Landauer, P. Prabhu (eds.), *Handbook for HCI*, Elsevier Science Publishers, pp. 1043-1059, 1997.

T. Koda, P. Maes. Agents with Faces: The Effects of Personification of Agents. *Proceedings of HCI'96*, pp. 239-245, 1996.

J.C. Lester et al. The Persona Effect: Affective Impact of Animated Pedagogical Agents. *CHI'97 Conference Proceedings*, ACM Press, pp. 359-366, 1997.

B. Reeves, C. Nass. *The Media Equation: How people treat computers, television and new media like real people and places*, CSLI Publications & Cambridge University Press, 1996.

Anthropomorphism and Robotics

Brian R. Duffy

Media Lab Europe, Sugar House Lane, Bellevue, Dublin 8, Ireland

brd@media.mit.edu

Abstract

Should a robot be constrained to look like a human? A difficult question, but extremely relevant at a stage where technology is starting to provide us with some quite robust solutions to technical problems that have constrained robot development over the years. This paper discusses the role that anthropomorphism plays in robotics research and proposes that research on human-robot interaction or social robotics is the primary motivation for incorporating human aspects in robotics.

1 Introduction

Robotic research has evolved through a few basic stages when considering the impact of computers as control mechanisms. The most notable example of the application of Classical AI in the late 1960's was the robot Shakey (Nilsson, 1984). This highlighted the inability of computationally intensive centralised control systems to deal with real-world unpredictability and complexity. A rebelling against the strongly deliberative approach resulted in highly reactive techniques where little to no internal representation of the environment was maintained (Brooks, 1986). Issues regarding the difficulty in realising explicit goals have limited the development of purely reactive control approaches. In recent years, one could say that following the emergent intelligence experiments with multiple simple reactive robots (Cao *et al., 1997*), the idea of developing more explicit social functionality between a collective of robots gained momentum (Duffy, 2000). While humanoid robotics has arguably been *the* ultimate robot research, the implications of this social dimension to robotics have reinforced the design and building of human-like robots (Honda Asimo, PINO). In the extreme, the building of strongly human-like robots constrains the functionality that a robot could otherwise have. The role of anthropomorphism in robotics is not to build an artificial human, but rather to take advantage of it as a mechanism through which social interaction can be facilitated. Once robots come out of the washing machine and start moving around our physical and social spaces, their role and our dealings with them will change significantly. It is in embracing their inherent advantage in being machines rather than seeing this as a disadvantage that will lead to their success.

2 The Humanoid

However disparate the means by which humanoid robot research has evolved, most are concerned with the human form and corresponding motor and perceptual functionality required to realise a physical and functional anthropomorphic robot. The exterior design of the robot has therefore aimed to define its functionality and facilitate its assertion of its identity as distinct from that of a mere object. Such research in building an artificial human has been fundamentally driven by two distinct motivations:

1. Addressing the engineering issues of building an artificial entity capable of performing in people environments with a similar capacity to a human
2. Building mechanisms whereby computational models can be implemented and tested in order to better understand human beings.

These areas stem from either the perspective of being inspired by humans in order to realise sophisticated machines, or use machines to understand humans. Research centring on the development of a humanoid machine from a mechanical engineering approach (Honda Asimo, PINO) have lead to impressive results. Similarly, the development of symbolic machines by which artificial mechanisms of "thought" are developed are still playing "catch-up" with the human brain. Yet a freely moving autonomous entity reacting to diverse sensory information through control mechanisms, with a sufficient degree of complexity, can still simulate a degree of intelligent behaviour.

It is proposed that success in designing a robot's functionality is not to attempt to make it as multi-functional and versatile as ourselves, but rather to employ those characteristics and attributes that are intrinsically "machine-like" to address capabilities that we do not have and require machines to do. It is important to make the distinction between the issues of whether one views the human body as inadequate as could be perceived in aspects of cyborg research, or if one aims to incorporate functionality on a robot that we simply require machines to perform anyway.

An interesting third viewpoint is to let the robot "be all it can" (i.e. employ technologies that provide functionality that humans do not have) with the humanlike features contributing to the social scenarios that arise when this robot is interacting with people, rather than being the paradigms through which the robot is constrained in its functionality and appearance. This social interaction paradigm is the only real justification for building robots that look like humans. The ascription of human features and metaphors, to facilitate interaction between man and machine in social scenarios, are important. Work to date on investigating the social interaction between people and robots employing human face-like features have demonstrated the importance of such analogies (Hara & Kobayashi, 1995; WE-3RIV, 2001; Brezeal, 2000). On the other hand, if functional requirements for a robot *as a machine* outside any social context, efficient design would most likely require the physical form of the robot to constitute something completely different from the human form.

The general objective of Social Robot research is to build a robot that can engage in social interaction scenarios, often with humans, in a familiar and compelling manner providing social communicative functionality that is natural and intuitive. From a scientific perspective, the use of such terms as "familiar", "compelling", "natural" and "intuitive" are about as difficult to deal with as the notion of "anthropomorphism" (see section 4). It is the psychological affiliation with the object that presents interesting challenges. This paper argues that these underlying perceptions hint at characteristics for a social robot, which eases the engineering task of creating flawlessly realistic humanoids. The *Placebo* project (Tony Dunne & Fiona Raby) demonstrates the development of personal attachments to objects and has effectively already highlighted the magnitude of flexibility available for robots strongly endowed with social functionality and features.

2.1 Human-robot relations

Recent robot research has studied the relations between humans and robots (Breazeal, 2000; Iida *et al.*, 1999; Duffy, 2002). These relationships will enable humans to interact meaningfully with a robot. The designer now has to strive towards facilitating the relationship between the human and the robot often with a view towards defining the scenario of the interaction. The physical construction of the robot plays an important role in such social contexts. The simple notion of a robot having two eyes in what could be characterised as a head, would intuitively facilitate where a person focuses when speaking to the robot. As highlighted in the PINO research, "the aesthetic element [plays] a pivotal role in establishing harmonious co-existence between the consumer and the product" (PINO). Such physical attributes as size, weight, proportions and motion capabilities are basic observable modalities that define this relationship (PINO is approximately 75cm tall).

Anthropomorphism is prevalent in robotics because of one's tendency to need such familiarity. Robotics research focuses on building humanoids not because a humanoid is the most efficient design for *any* given task, but rather because of one's tendency and need to anthropomorphise. Searle already points out the dangers present through confounding reasoning and rationalising because of this tendency to anthropomorphise (1992). Effectively, anthropomorphism obstructs the fact that the human form is not the ideal for a machine. However, objectively, the humanoid form best helps us interact with machines *because* of the fact that we anthropomorphise. But this does not necessarily mean that one should build seamlessly perfect-looking human-machines as anthropomorphism allows roboticists be heuristic in robot design (i.e. cartoon characters).

And then the robot looks *back* at you. This adds a further dimension to the interaction. Maybe this is what we mean by artificial life, a *perceived* notion of consciousness artificially attained through anthropomorphism. This intimate social scenario between a human and a physical robot will become a very important issue in social robot development.

3 The Role of the Social Robot

What role will such a robot have? While a dreaded question, the answer is actually quite straightforward. It is not the definition of a specific task that warrants the research, but rather any situation that requires people to interact with a machine. The application scenarios that naturally stem from this can include a museum guide, a corporate building butler, or a post office clerk. In fact, the fundamental issue is to ascertain what the robot, *as a machine*, is inherently *good* at.

Some argue that the design and construction of strong humanoid form and functionality is such that robots can be successfully integrated into human society where their primary mobile functions will be to negotiate the same obstacles encountered by humans in daily life. This is a weak argument when the proliferation of wheelchair access can minimise the necessity for two legs over the more efficient wheeled chassis for a robot. Even work in hazardous environments such as nuclear plants does not require a human friendly set-up but rather can have a "robot-friendly" configuration for highly specialised machines.

Contemporary interaction between humans and technology is dictated by machine-specific operations. Thus the user is required to conform to the operational procedures and learn the "language" of the machine. Machines do not provide intuitively understandable interfaces with the consequence that many users of a machine are unable to utilise its complete functionality.

In order to address such interaction problems between people and machines, intelligent socially capable machines must be developed which are able to analyse and understand various natural and temporally coordinated input modalities such as intentions of others which can be inferred from communication (through language, and facial and other gestures such as eye and body behaviours). Through an analysis of the particular context and the fusion of different modalities, faulty and incomplete perceptions may be meaningfully interpreted. (see also Wahlster *et al.*, 2001). Psychological experiments are currently exploring how motion can be employed to realise such complex notions as perceived emotional states and deliberative reasoning (Duffy, 2002). These all require an investigation of the role of anthropomorphism in robotics.

4 Anthropomorphism

Anthropomorphism (from the Greek word *anthropos* for man, and *morphe*, form/structure) can be viewed as the tendency to attribute human characteristics to inanimate objects, animals and others with a view to helping us rationalise a situation. It is attributing cognitive or emotional states to something based on observation in order to rationalise another's behaviour in a given social environment. This phenomenon has been exploited from religion in Homer's gods of ancient Greece to recent animation films like "Chicken Run" (Aardman Animations, 2000) or "Antz" (DreamWorks & SKG/PDI, 1998). Accordingly, it has been open to criticism with a first example found in *Fragments* where Xenophanes (c. 560-478BC) wrote a scathing critique of the Homeric texts:

> *"But mortals suppose that the gods are born (as they themselves are), and that they wear man's clothing and have human voice and body. But if cattle or lions had hands, so as to paint with their hands and produce works of art as men do, they would paint their gods and give them bodies in form like their own-horses like horses, cattle like cattle"* (Xenophanes in Lesher, 1992).

Xenophanes' claims articulate an important facet of attributing nonhuman entities with human characteristics. Anthropomorphic activity can complicate simple distinctions between human and nonhuman, while on the other hand, it helps us cope with unfamiliarities and intangibles.

From a psychological perspective, anthropomorphism has not been rigorously studied to any major extent where only a few have seen it as worthy of study in its own right (e.g. Caporael, 1986; Eddy *et al.*, 1993; Tamir & Zohar, 1991), rather than the common view that it is a hindrance to science. Anthropomorphism is considered a hindrance when confounded *with* scientific observation rather than been studied more objectively and taken advantage of. Anthropomorphism *is* a complex and subtle phenomenon (Eddy *et al.*, 1993).

4.1 Use and Misuse of Anthropomorphism

One way of understanding an entity's behaviour is to ascribe mental states to it by simply anthropomorphising it. This involves ascribing human mental characteristics with often little or no reference to the entity's real competences. Eddy *et al.* (1993) suggest that people anthropomorphise animals depending on:

> *"(1) the degree of physical similarity between themselves and the species in question (e.g. primates,) and (2) the degree to which they have formed an attachment bond with a particular animal (e.g. dogs and cats)"*

Caporael (1986) suggests that anthropomorphism is a *"'default schema' applied to non-social objects"*. Eddy points out that familiarity increases the tendency to anthropomorphise (Eddy et al., 1993).

Watt points out that "[a]nthropomorphism is part of what makes us see others as minds rather than as bodies" (Watt, 1998). From a methodological perspective, using such anthropomorphic projection is seen as bad scientific practice, particularly in biology and ethology. Similarly, in cognitive science, Searle (1992) comments:

> *"Prior to Darwin, it was common to anthropomorphise plant behaviour and say such things as the plant turns its leaves towards the sun to aid in its survival. The plant 'wants' to survive and flourish and 'to do so' it follows the sun".*

Searle then tries to provide a more functional explanation:

> *"Plants that turn their leaves towards the sun are more likely to survive than plants that do not".*

Is it possible to remove anthropos from our science when we cannot directly view something from another entity's perspective? Watt points out that *"while we may sympathise with Searle's criticism of anthropomorphism in cognitive science, it may well be endemic"*. Eddy *et al.* (1993) note that it is *"almost irresistible"*. Krementsov and Todes (1991) comment that *"the long history of anthropomorphic metaphors, however, may testify to their inevitability"*. If we are unable to decontextualise our perceptions of a given situation from ourselves, then condemning anthropomorphism won't help. Caporael (1986) proposes that if we are therefore unable to remove anthropomorphism from science, we should at least "set traps for it" in order to be aware of its presence in scientific assessment.

Molloy (2001) highlights that *"[w]hilst particular scientific discourses find the ascription of human characteristics to nonhuman entities interminably fallacious, anthropomorphism also emerges where taxonomic legitimacy of the classification 'human' is under threat"*. Kennedy goes as far as to say that anthropomorphic interpretation '*is a drag on the scientific study of the causal mechanisms*' (Kennedy, 1990). Building social robots forces a new perspective on this. When people are the motivation for social robot research, then people have an influence on how the robot is realised.

An area where anthropomorphism has become a strong issue is in Human Computer Interaction research (Nass & Moon, 2000; Bates, *et al.* 1992; André *et al.*, 1999; Reilly, 1996) with few doubting its strengths. With regard to realising synthetic personalities in artificial intelligence, the assumption has generally been that the creation of even crude computer "personalities" necessarily requires considerable computing power and realistic human-like representations. On the contrary, investigations using simple scripting of text has demonstrated that "even minimal cues can mindlessly evoke a wide range of scripts, with strong attitudinal and behavioural consequences" (Nass & Moon, 2000). The anthropomorphic design of human-machine interfaces is inevitable. The important criterion is to seek a balance between people's expectations and the machines capabilities.

4.2 Two Kinds?

If anthropomorphism refers to the tendency to attribute human characteristics to an entity, two strongly related situations may arise. The first is where we see something and rationalise its behaviour based on humanlike analogies, i.e. "the robot looks lost", "it's thinking". This would correspond to Watt's projective anthropomorphism where we project our interpretations of what is happening onto the entity strictly from an observer perspective: *"external animals and systems become chimerae through the superposition of aspects of the observer"* (Watt, 1998). We draw humanlike assumptions about an entity's behaviour. We may ascribe stronger mentalistic notions than actually exist. It involves rationalising behaviour through mechanisms that we can personally relate to.

The second is the use of anthropomorphic paradigms to augment the functionality and behavioural characteristics of something (both anticipatory and actual) in order that we can relate to and rationalise its actions with *greater ease*. The use of humanlike features for social interaction with people (i.e. Kismet (Breazeal, 2000)) facilitates our social understanding. It is the explicit designing of anthropomorphic features (i.e. two eyes and a mouth) to facilitate the social interaction. These two aspects are not distinct, but rather could be viewed as being two points along the anthropomorphism continuum. This highlights that anthropomorphism is fundamentally observer dependent.

Anthropomorphism can also lead to misunderstanding when the metaphor is either misinterpreted or carried too far. It can lead to an observer who may not be aware of the complexity of a given situation and call the robot "stupid". The observer can similarly attribute behavioural characteristics synonymous with human behaviours to systems that could, in fact, be "trying" to perform something else entirely. A typical example in autonomous mobile robots is a wandering behaviour misconstrued as "the robot is looking for something!"

Watts proposes a notion of "introjective anthropomorphism" in which the *"observer comes to be, in part, a chimera with the observed system"*. This "introjective" notion does not constitute a form of anthropomorphism as it involves the modification of the person's behaviour to the observed subject's, i.e. meowing at a cat. Not even if the subject is another person could this be termed anthropomorphic, as anthropomorphism inherently refers to the ascription of humanlike characteristics to nonhuman entities.

4.3 Optimal Anthropomorphism?

Is there a notion of "optimal anthropomorphism"? What is the ideal set of human features that could supplement and augment a robots social functionality? When does anthropomorphism go too far? Using real world robots poses many interesting problems. Currently a robot's physical similarity to a person is only starting to embrace basic humanlike attributes, and predominantly the physical aspect in humanoid research. A robot not embracing the anthropomorphic paradigm in some form is likely to result in a persistent bias against people ascribing mental states to them. Molloy proposes that *"[f]amiliarity, fortunately, offers us a way out of this trap – we can in principle learn to see computers as minds"* (Molloy, 2001).

Research in HCI and robotics has recently started to address what supplementary modalities to physical construction could be employed for the development of social relationships between a physical robot and people. Important arenas include expressive faces (Hara & Kobayashi, 1995; Breazeal, 2000; Cozzi *et al.*, 2001) often highlighting the importance of making eye contact and incorporating face and eye tracking systems. Interestingly, it can be argued that the most successful implementation of expressive facial features is through more mechanistic and iconic heads such as Anthropos (Duffy, 2002) and Kismet (Brezeal, 2000). Strong human-like facial construction in robotics (Hara & Kobayashi, 1995) has to contend with the minute subtleties in facial expression, a feat by no means trivial. Contrarily, it can be argued that successful highly humanlike facial expression is an issue of resolution.

An important question is how does one manage anthropomorphism. The previous examples demonstrate two methodologies that employ either a visually iconic or a strongly realistic humanlike construction (i.e. with synthetic skin and hair) for facial gestures in order to portray artificial emotion states. The iconic head defines the degree through which anthropomorphism is employed through the robot's construction and functional capabilities and therefore constrains and effectively manages the degree of anthropomorphism employed. Building mannequin-like robotic heads where the objective is to hide the "robotic" element as much as possible and blur the issue as to whether one is talking to a machine or a person results in effectively unconstrained anthropomorphism and a fragile manipulation of robot-human social interaction.

The question of optimal anthropomorphism is an extremely difficult one, but if we look to try and break it down within a particular context, we may begin to understand its usability and advantages as a conceptual paradigm. It remains to be explored and hints that robot designers must bootstrap it through psychological investigations.

5 Conclusions

The criticism of anthropocentricity is an important issue in cognitive science. Is it appropriate? Should we clarify the distinction between research on *human-like* artificial intelligence and other forms? While anthropomorphism inherently relates to humans, is the phenomenon itself really restricted to our own species? If we refrained from seeking to replicate human intelligence, could we succeed more quickly in achieving a notion of "robot intelligence"? This extreme perspective does highlight one aspect of how we perceive our environment. But there is an obvious counter-argument. Can we realistically decontextualise ourselves from our perceptions? Can we be objective in our view of the environment? Isn't the use of pattern recognition a powerful tool to rationalise for example the behaviour of a robot as it moves about the environment? "It's looking for food!", "it's dancing!". Emergent intelligence takes this observability aspect a stage further. With the implementation of relatively simplistic sensory-motor mappings as found in the Braitenburg vehicles (Braitenberg, 1984) and in emergent intelligence experiments with collectives of simple robots (Cao *et al.*, 1997), it is the observer that has the overall perspective of the system. It is our anthropomorphisation of the robot system that often dictates how we define its behaviour, and consequently what the system is.

The perspectives discussed here have tended towards the pragmatic by viewing the robot as a functional tool employing such humanlike qualities as personality, gestures, expressions and even emotions to facilitate its role in human society. The recent film AI (Warner Brothers, 2001) highlights the extended role that robot research is romantically perceived as aiming to acheive. Even the robot PINO (Yamasaki *et al.*, 2000) is portrayed as addressing "its genesis in purely human terms" and analogies with Pinocchio are drawn, most notably in the name. But, does the robot *itself* have a wish to become a human boy? Or do *we* have this wish for the robot? We should be asking why.

While anthropomorphism is clearly a very complex notion, it intuitively provides us with very powerful physical and social features that will no doubt be implemented to a greater extent in social robotics research in the near future.

6 References

André, E., Klesen, M., Gebhard, P., Allen, S., Thomas, R., (1999) Integrating models of personality and emotions into lifelike characters, *Proc. workshop on Affect in Interactions – Towards a new Generation of Interfaces, 3rd i3 Annual Conference, Oct., Sienna, Italy,* 136-149.

Bates, J., Loyall, A.B., Reilly, W.S., (1992) An architecture for action, emotion, and social behaviour, CMU-CS-92-144, Carnegie Mellon University.

Braitenburg, V. (1984), *Vehicles - experiments in synthetic psychology.* MIT Press

Breazeal, C. (2000), Sociable Machines: Expressive Social Exchange Between Humans and Robots, *Sc.D. dissertation, Department of Electrical Engineering and Computer Science, MIT.*

Brooks, R., Breazeal, C., Marjanovic, M., Scassellati, B., Williamson, M., (1998) The Cog Project: Building a Humanoid Robot, *in C. Nehaniv, ed., Computation for Metaphors, Analogy and Agents,* Vol.1562, *Springer Lecture Notes in Artificial Intelligence*

Cao, Y.U., Fukunaga, A.S., Kahng, A.B. (1997) Cooperative Mobile Robotics: Antecedents and Directions, *Autonomous Robots,* 4, P1-23,

Caporael, L. R. (1986). Anthropomorphism and mechanomorphism: Two faces of the human machine. *Computers in Human Behavior,* 2(3):215–234.

Cozzi, A., Flickner, M., Mao, J., Vaithyanathan, S., (2001) A Comparison of Classifiers for Real-Time Eye Detection, *ICANN, P993-999*

Duffy, B.R., (2002) The Anthropos Project: http://www.mle.ie/research/socialrobot/index.html

Duffy, B.R., Collier, R.W., O'Hare, G. M. P., Rooney, C.F.B., O'Donoghue, R.P.S. (1999) Social Robotics: Reality and Virtuality in Agent-Based Robotics, *Bar-Ilan Symposium on the Foundations of Artificial Intelligence: Bridging Theory and Practice (BISFAI), Israel*

Duffy, B. The Social Robot. (2000) *Ph.D. Thesis, Department of Computer Science, University College Dublin.*

Eddy, T.J., Gallup, G.G., Jr. & Povinelli, D.J. (1993). Attribution of cognitive states to animals: Anthropomorphism in comparative perspective. *Journal of Social Issues,* 49, 87-101.

Gallup, G.G., Jr., Marino, L.A. & Eddy, T.J. (1997). Anthropomorphism and the evolution of social intelligence: A comparative approach. In R.W. Mitchell, N.S. Thompson & H.L. Miles (Eds.), *Anthropomorphism, Anecdotes, and Animals.* (pp. 77-91). Albany, NY: State University of New York Press.

Hara, F. Kobayashi, H., (1995) Use of Face Robot for Human-Computer Communication, *Proceedings of International Conference on System, Man and Cybernetics,* p10.

Haraway, D. (1992) Otherworldly Conversations; Terran Topics; Local Terms. *Science as Culture, 3*(14), 64-98.

Iida, F., Ayai, H., Hara, F., (1999) Behaviour learning of a face robot using human natural instruction, *8th International Workshop on Robot and Human Interaction, Sept., Pisa, Italy*

Kennedy, J.S. (1992) *The New Anthropomorphism.* Cambridge Press.

Krementsov, N. L., & Todes, D. P. (1991). On Metaphors, Animals, and Us, *Journal of Social Issues, 47*(3), 67-81.

Lesher, J.H. (1992), *Xenophanes of Colophon: Fragments: A Text and Translation With a Commentary,* Univ of Toronto Pr; ISBN: 0802059902

Nass, C., Moon, Y. (2000) Machines and mindlessness: Social responses to computers, *Journal of Social Issues, 56*(1), 81-103

Nilsson N.J. (1984), Shakey the robot, *SRI A.I. Center Technical Note 323, April 1984*

PINO http://www.symbio.jst.go.jp/~yamasaki/

Rao, A.S., Georgeff, M.P. (1991), Modelling Rational Agents within a BDI Architecture, *Prin. Of Knowl. Rep. & Reas., San Mateo, CA.*

Reilly, W.S.N., Believable social and emotional agents, PhD Thesis, CMU-CS-96-138, Carnegie Mellon University

Searle, J. R. (1992). *The Rediscovery of the Mind.* Cambridge, Massachusetts: MIT Press.

Tamir, P., & Zohar, A. (1991). Anthropomorphism and Teleology in Reasoning about Biological Phenomena, *Science Education, 75*(1), 57-67.

Wahlster, W., Reithinger, N., Blocher, A., (2001) SmartKom: Multimodal Communication with a Life-Like Character, Eurospeech, Aalborg, Denmark.

Watt, S. (1998) Seeing this as people: anthropomorphism and common-sense psychology, *Ph.D Thesis, Knowledge Media Institute, The Open University.*

WE-3RIV: http://www.humanoid.waseda.ac.jp/

Yamasaki F., Miyashita T., Matsui T., Kitano H., (2000) PINO the Humanoid: A basic architecture, *Proc. of The Fourth International Workshop on RoboCup, August 31, Melbourne, Australia*

Face Value: Towards Emotionally Expressive Avatars

Marc Fabri*; David J. Moore*; Dave J. Hobbs$
* Leeds Metropolitan University; School of Computing; ISLE Research Group
$ University of Bradford; School of Informatics
M.Fabri@lmu.ac.uk; D.Moore@lmu.ac.uk; D.Hobbs@bradford.ac.uk

Abstract

This paper investigates how inhabitants of collaborative virtual environments (CVEs) can communicate with each other through channels other than speech, and it is primarily concerned with the perception of facial expressions of emotion in CVEs. We outline our experimental work on emotion recognition and investigate to what extent findings from real life socio-psychological research can be applied to quasi face-to-face encounters in three-dimensional virtual environments.

1 Introduction

Natural human communication is based on speech, facial expressions, body posture and gestures. While speech is an obvious instrument for mediating our thoughts and ideas, social intercourse also depends heavily on the expressions, and movements of the body (Morris et al, 1979). Such socio-emotional content is vital for building relationships that go beyond the purely factual and task-oriented communication usually encountered in a business environment. Indeed, social psychologists argue that more than 65% of the information exchanged during a face-to-face conversation is carried on the nonverbal band (Knapp, 1978; Argyle, 1988). Therefore, it is expected to be beneficial to provide such communication channels in collaborative virtual environments (CVEs) in some way.

In CVEs, inhabitants are usually represented by humanoid embodiments, generally referred to as avatars. Since each avatar is both part of the perceived environment and represents the user that is doing the perceiving (Slater and Wilbur, 1997), inhabitants potentially develop a strong sense of mutual awareness. The avatar can provide direct feedback about one particular user's actions, degree of attention and interactive abilities, to the other inhabitants. It becomes an effective interaction device.

Potential applications for CVE systems are all areas where people cannot come together physically, but wish to discuss, collaborate on, or even dispute certain matters. We are in particular concerned with the use of CVE technology in Distance Learning systems (see Fabri and Gerhard, 2000, for a detailed discussion). Interaction between those involved in the learning process is important for mutual reflection on actions and problem solutions, motivation and stimulation (Laurillard, 1993; Garrison, 1993; Moore, 1993). More specifically, the ability to show emotions, empathy and understanding through facial expressions and body language is central to ensuring the quality of tutor-learner interaction (Knapp, 1978; Cooper et al, 2000).

Recent findings in psychology and neurology suggest that emotions are also an important factor in decision-making, problem solving, cognition and intelligence in general (Dittrich, 1993; Damásio, 1994; Picard, 1997; Lisetti and Schiano, 2000). We therefore argue that computer-based educational technologies ought to emulate this in some way.

2 The expression of emotion

Various channels are available to express emotion: voice, the face, gaze, gesture, or posture. Of the nonverbal channels, the face is the most immediate indicator for the emotional state of a person (Ekman and Friesen, 1975a). As socially active humans, we usually have an informal understanding of what emotion is and what different emotions there are. There is also a formal research tradition, which has investigated the nature of emotion systematically. Major figures in scientific research have contributed, most notably the philosopher Rene Descartes (1641), biologist Charles Darwin (1872), and more recently psychologists William James and Paul Ekman.

Ekman et al (1972) found that there are six universal facial expressions, corresponding to the following emotions: *Surprise*, *Anger*, *Fear*, *Happiness*, *Disgust/Contempt*, and *Sadness*. The categorisation is widely accepted, and considerable research has shown that these basic emotions have a clear meaning across cultures (Zebrowitz, 1997; Ekman, 1999). Indeed, it is held that expression and, to an extent, recognition of these six emotions has an innate basis.

This naturally developed skill to "read" facial expressions is considered highly beneficial to communication in CVEs. We believe that the emotionally expressive virtual face of an interlocutor's avatar can aid the communication process and provides information that would otherwise be difficult to mediate.

To comprehensively describe the visible muscle movement in the face, Ekman and Friesen (1975a; 1978) developed the Facial Action Coding System (FACS). It is based on highly detailed anatomical studies of human faces. A facial expression is a high level description of facial motions, which can be decomposed into certain muscular activities, i.e. relaxation or contraction, called Action Units (AUs). The figure below shows variations of the anger category (from Ekman and Friesen, 1975b):

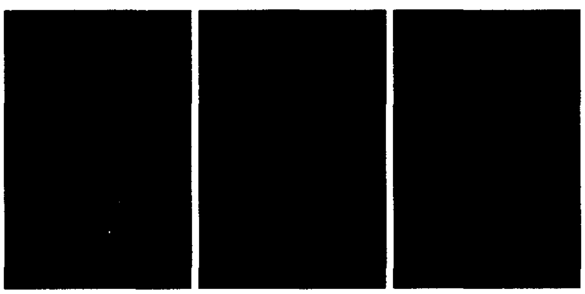


Figure 1: Variations of *Anger* (FACS photographs)

3 Modelling expressive faces

Interest in modelling the human face has been strong in the computer graphics community since the 1980s. Platt and Badler (1981) developed the first muscle-based model of an animated face, using geometric deformation operators to control a large number of muscle units. Parke and Waters (1985), later Terzopoulus and Waters (1993) further developed this by modelling the anatomical nature of facial muscles and the elastic nature of human skin, resulting in a dynamic muscle model that is controllable via fewer parameters. Improved versions of muscle and skin based models are in use today.

The approach we have chosen is feature-based and less complex than a realistic simulation of real-life physiology. This is sufficient, and in fact preferable, as it allows us at the same time to establish what the most distinctive and essential features of a facial expression are. Our face model is based on H-Anim (2002), the specification proposed by an international panel that develops the Virtual Reality Modeling Language (VRML). Figure 2 shows the controllable parameters of the virtual head, modelled in VRML:

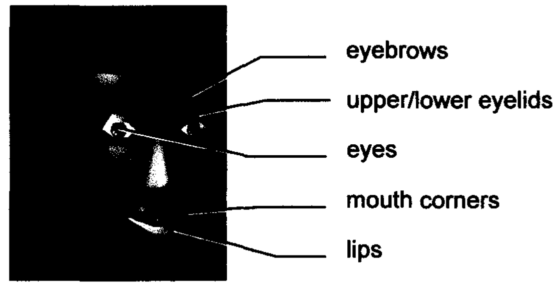


Figure 2: Controllable features of the virtual head

Clearly, these parameters alone do not allow representation of all facial expressions possible in real life. However, not the entire set of FACS action units needs to be reproduced to achieve the level of detail envisaged for the current face model. There is evidence that approaches aiming to reproduce the human face in detail may in fact be wasteful (Benford et al, 1995; Donath, 2001).

Donath (2001) warns that because the face is so highly expressive and we are so adept in reading (into) it, any level of detail in 3D facial rendering could potentially provoke the interpretation of various social messages. If these messages are unintentional, the face is arguably hindering communication more than it is helping. Also, exaggerated or particularly distinctive faces may convey emotions more efficiently than normal faces (Zebrowitz, 1997; Bartneck 2001), a detail regularly taken advantage of by caricaturists.

Further, the human perception system can recognise physiognomic clues, in particular facial expressions, from very few visual stimuli (Dittrich, 1993), and our head model is designed to display merely these distinctive facial clues. Figure 3 shows the same variations of *anger* as Figure 1, now depicted by the virtual head:

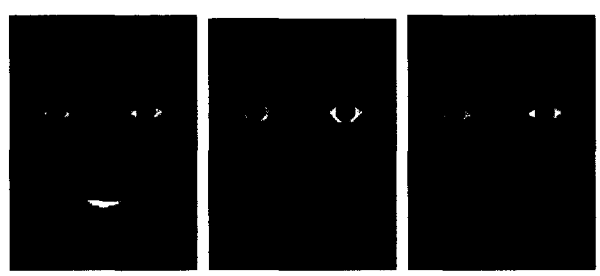


Figure 3: Variations of *Anger* (virtual head model)

We regard the virtual head with its limited, but human-like expressive abilities as a potentially effective and efficient means to convey emotions in virtual environments. Further, we consider the reduced set of action units and the resulting facial animation control parameters as being sufficient to express, in a readily recognisable manner, the six universal facial expressions of emotion (plus the neutral expression).

4 Experimental study

An experiment was carried out to test above assumptions. Screenshots of the virtual head depicting variations of emotions were shown to subjects. Subjects were also to look at photographs of people showing corresponding expressions, taken from a FACS databank (Ekman and Friesen, 1975b). The task was to assign each expression to one of the aforementioned categories. Two further categories were "Other..." and "Don't know". During the study, subject responses were automatically recorded and compiled. In an attempt to equalise possible individual differences in recognition skills we chose a repeated measures design, i.e. subjects constituted their own control group.

29 subjects took part in the experiment, 17 female and 12 male, with an average age of 30 (ranging from 22 to 51 years old). None had classified facial expressions or used FACS before. None of the subjects worked in facial animation, although some were familiar with 3D modelling techniques in general. Each subject was shown 28 FACS photographs and 28 corresponding virtual head images, in a randomly generated order that was the same for all subjects. Each of the six emotion categories was represented in 4 variations, plus 4 variations of the neutral face. The variations were defined not by intensity, but by differences in expression of the same emotion. The controllable parameters of the virtual head were adjusted to correspond with the natural faces.

All virtual head images depicted the same male model throughout, whereas FACS photographs showed several people, expressing a varying number of emotions (21 images showing male persons, 8 female). Photographs were selected from the FACS databank solely based on their high recognition rates. This was believed to be the most appropriate method, aiming to avoid the introduction of factors that would potentially disturb results, such as gender, age or ethnicity.

5 Results

Statistical analysis showed that the recognition accuracy, or *distinctness* as labelled by Bartneck (2001), for FACS photographs (78.6% overall) is significantly higher than those for virtual heads (62,2%). A Mann-Whitney test at significance level 0.01 confirmed this. Recognition rates vary across emotion categories, as well as between the two conditions, illustrated clearly in Figure 4.

Interestingly, subjects who achieved better results did so homogenously between virtual and FACS images. Lower scoring subjects were more likely to fail recognising virtual heads, rather than FACS photographs.

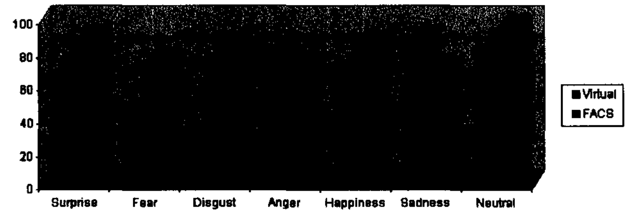


Figure 4: Summary of recognition rates

A closer look at recognition rates of particular emotions reveals that all but disgust have at least one FACS/virtual head pair with comparably high results, i.e. recognition was as successful with the virtual head as it was with the directly corresponding FACS photographs. Figure 5 shows results for the highest scoring virtual head per category (see Figure 7 for the artefacts these recognition rates are based on):

Figure 5: Recognition rates for selected images

6 Categorisation error analysis

We conducted a detailed analysis of the errors subjects made when asked to categorise facial expressions, shown in Table 1. Rows give per cent occurrence of each response. Confusion values above 10% are shaded light grey, above 20% dark grey, above 30% black.:

Table 1: Categorisation confusion matrix

	Response [Virtual / FACS]							Other/
	Surprise	Fear	Disgust	Anger	Happiness	Sadness	Neutral	Don't know
[illegible]	.67 .85	.06 .07	.00 .00	.00 .01	[illegible] .00	.00 .00	.01 .00	.03 .08
[illegible]	.15 .19	.41 .73	.00 .04	[illegible] .00	.03 .00	.03 .00	.02 .00	.06 .03
[illegible]	.01 .02	.02 .00	.22 .77	[illegible] .14	.01 .00	.04 .00	.10 .01	[illegible] .07
[illegible]	.03 .04	.00 .04	.00 .03	.77 .72	.02 .00	.03 .03	.11 .05	.05 .09
[illegible]	.01 .00	.01 .00	.01 .00	.01 .00	.64 .84	.03 .00	[illegible] .15	.04 .02
[illegible]	.06 .00	.09 .10	.00 .00	.00 .01	.01 .01	.85 .66	.03 .09	.01 .07
[illegible]	.03 .00	.03 .00	.01 .00	.00 .01	.00 .02	.11 .01	.78 .94	.04 .02

6.1 Disgust

The majority of errors were made in the category Disgust, an emotion frequently confused with Anger. When examining results for virtual heads only, anger was picked nearly twice as often (39%) as disgust (22%). Further, with faces showing disgust, subjects often felt unable to select any given category and instead picked "Don't know" or suggested an alternative emotion. The alternatives included the terms *aggressiveness, hatred, irritation,* or *self-righteousness.*

Ekman and Friesen (1975a) describe disgust (or contempt) as an emotion that often carries an element of condescension toward the object of contempt. People feeling disgusted by other people tend to feel morally superior to them. Our observations confirm this tendency. Where "other" was selected instead of the expected "disgust", the suggested alternative was often in line with the psychologists' interpretation.

6.2 Fear and Surprise

The error matrix further reveals that *Fear* was often mistaken for *Surprise,* a tendency also observed in several of Ekman's studies (1999). Incidentally, the experience and therefore expression of fear and surprise often happens simultaneously, e.g. when fear is felt suddenly due to an unexpected threat (Ekman and Friesen, 1975a). The *appearance* of fear and surprise is also similar, with fear generally producing a more tense facial expression. However, fear differs from surprise in three ways:

1. Whilst surprise is not necessarily pleasant or unpleasant, even mild fear is unpleasant.
2. Something familiar can induce fear, but hardly surprise (for example a visit to the dentist).
3. Whilst surprise usually disappears as soon as it is clear what the surprising event was, fear can last much longer, even when the nature of the event is fully known.

These indicators allow differentiating whether a person is afraid or surprised. All three have to do with *context* and *timing* of the fear-inspiring event – factors that are not perceivable from still images. In accordance with this, Poggi and Pelachaud (2000) found that emotional information is not only contained in the facial expression itself, but also in the *performatives* of a communicative act: suggesting, warning, ordering, imploring, approving and praising.

Bartneck (2001) observed that recognition rates for still images of facial expressions where higher when shown in a game context, compared to images shown out of context, as was the case in this study.

In other words, the inherent meaning and subjective interpretation of an emotional expression depend on the situation in which it is shown. This strongly suggests that when using emotionally expressive avatars in CVEs, recognition rates will be higher since there will be both a context and triggers for the emotion display.

6.3 Fear and Anger

The relationship between *Fear* and *Anger* is similar to that between fear and surprise. Both can occur simultaneously, and their appearance often blends. What is striking is that all confusions were made with virtual faces. This may suggest that the fear category contained some particularly unsuitable examples of modelled facial expressions. An examination of results showed that one virtual head expression of *fear* in particular was regularly mistaken for *anger,* shown below on the left, as expression (a) in Figure 6:

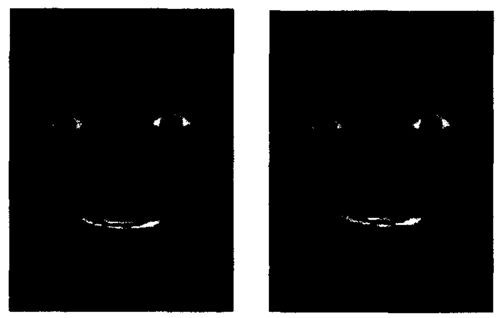


Figure 6: Fear expression (a), left. Alternative (b), right

Expression (a) displays characteristic *fear* eyes. The lower eyelid is visibly drawn up and appears to be very tensed. Both eyebrows are slightly raised and drawn together. The lower area of the face also shows characteristics of fear, namely a slightly opened mouth with stretched lips that are drawn together. In contrast, an angry mouth has the lips either pressed firmly together or open in a squarish shape, as if to shout.

However, despite these seemingly characteristic clues, 18 out of 29 times expression (a) was categorised as *anger.* Presumably, the main reason for this is the eyebrow: In anger, as in fear, eyebrows can be drawn together. But unlike the fearful face which shows raised eyebrows, angry faces typically feature a lowered brow. The eyebrows in this example are only slightly raised from the relaxed position - obviously not enough to give the desired impression. Another confusing factor may be the furrowed shape of the eyebrows. A straight line or arched brows are more typical for fear.

Consider expression (b) in Figure 6, which is identical to (a) apart from the eyebrows. They are now raised and arched, a detail that changes the facial expression significantly.

7 Choice of emotion categories

Results also indicate that limiting the number of categories might have had a negative effect on the recognition scores because of the number of "Other" and "Don't know" responses. This could probably be avoided in the future by allowing more categories, or alternatively by offering a range of suitable descriptions for an emotion category (e.g. *Joy*, *Cheerfulness* and *Delight*, to complement *Happiness*).

Overall, it has to be noted that many of the terms suggested by subjects as their "Other..." choice were actually very similar to the emotion category expected, confirming that the facial expressions in those cases were not necessarily badly depicted. This does however highlight the importance of having a well-defined vocabulary when investigating emotions - a problem that is not new to the research community and that has been discussed at length over the years (see Ekman et al, 1972 for an early comparison of dimensions vs. categories; Zebrowitz, 1997; Lisetti and Schiano, 2000; Wehrle and Kaiser, 2000)

8 Conclusions and further work

We showed that, when applying the FACS model to virtual face representations, emotions can effectively be visualised with a limited number of facial features. Virtual face recognition rates are, for the most part, comparable to those of their corresponding natural faces.

However, the study has also shown that the approach is not guaranteed to work for all expressions, or all variations of a particular emotion category. Not surprisingly, the critical parameters here are similar to those identified by real-life social psychology:

1. No categorisation system can ever be complete. Although accepted categories exist, emotions can vary in intensity and inevitably there is a subjective element to recognition. When modelling or animating facial features, such ambiguity in interpretation can be minimised by focussing on, and emphasising the most distinctive visual clues of a particular emotion.

2. *Context* plays a crucial role in emotion expression and recognition. Effective, accurate mediation of emotion is closely linked with the situation and other, related communicative signals. A reliable interpretation of facial expressions without taking cognisance of the context in which they are displayed is often not possible. It is hoped that emotion confusion can be avoided when facial expressions are applied to real-time interaction in CVEs.

3. Likewise, further work on emotion recognition in a real-time virtual reality setting has to consider the effects *timing* has on emotion display and interpretation. For example, showing surprise over a period of, say, a minute would - at the very least - send confusing or contradictory signals.

Further work will focus on those facial expressions that have proven to be most distinctive, i.e. scored the highest recognition rates (see Figure 7). To establish the possible role emotions can play for real-time interaction in CVEs, we will deploy emotionally expressive avatars that can be controlled, and mutually perceived, by CVE inhabitants.

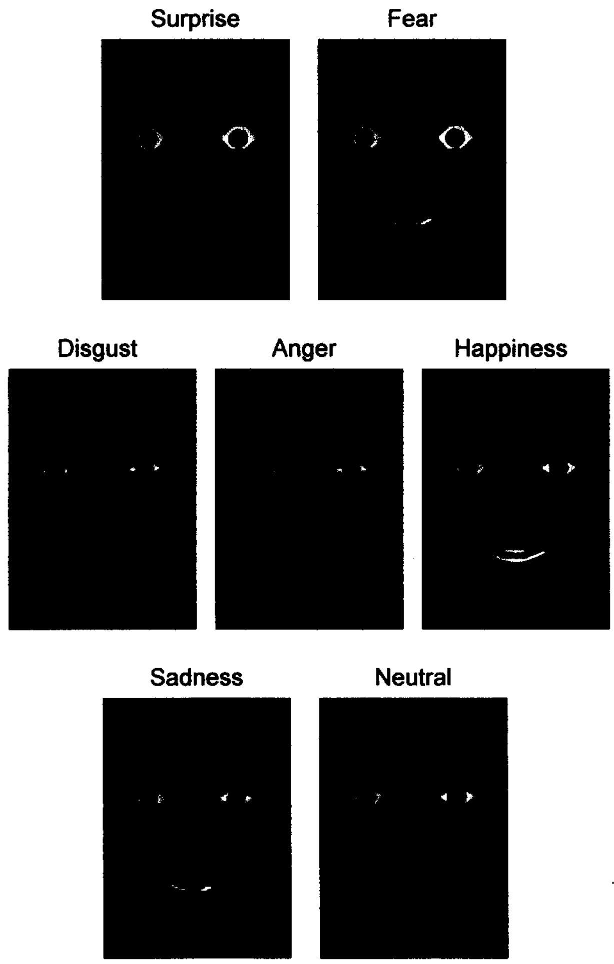


Figure 7: Most distinctive expression in each category

Acknowledgements

The Millennium Festival Committee, as part of the *CyberAxis* virtual gallery project, has funded part of this work. Original virtual head geometry copyright Geometrek (www.geometrek.com). Photographs from CD-Rom *Pictures of Facial Affect* (Ekman and Friesen, 1975b) used with permission.

References

M. Argyle. *Bodily Communication.* Methuen & Co, New York, 1988.

C. Bartneck. Affective Expressions of Machines. *User Modeling and User-Adapted Interaction,* 11:279-295, Kluwer Academic Publishers, 2001.

S.D. Benford, J. Bowers, L.E. Fahlén, C.M. Greenhalgh, and D. Snowdon. User Embodiment in Collaborative Virtual Environments. *Human Factors in Computing Systems.* CHI 1995 Conference Proceedings, ACM Press, 1995.

B. Cooper, P. Brna, and A. Martins. Effective Affective in Intelligent Systems: Building on Evidence of Empathy in Teaching and Learning. A. Paiva (Ed.) *Affective Interactions*, LNAI 1814:21-34, Springer, 2000.

A.R. Damásio. *Descarte's Error: Emotion, Reason and the Human Brain*, Avon Books, New York, 1994.

C. Darwin. *The Expression of the Emotions in Man and Animals*, Reprint, Oxford University Press, 1872.

R. Descartes. *Meditations and Other Metaphysical Writings*, Reprint, Penguin, 1641.

W.H. Dittrich. Action categories and the perception of biological motion. *Perception*, 22:15-22, 1993.

J. Donath. Mediated Faces. M. Beynon, C.L. Nehaniv, K. Dautenhahn (Eds.) *Cognitive Technology: Instruments of Mind*, Cognitive Technology Conference Proceedings, Warwick, United Kingdom, 2001.

P. Ekman. Facial Expressions. T. Dalgleish and M. Power (Eds.) *Handbook of Cognition and Emotion.* John Wiley & Sons, New York, 1999.

P. Ekman and W.V. Friesen. *Unmasking the Face*, Prentice-Hall, New Jersey, 1975a.

P. Ekman and W.V. Friesen. *Pictures of Facial Affect.* (Databank on CD-Rom), University of California, Department of Psychology, San Francisco, 1975b.

P. Ekman and W.V. Friesen. *Facial Action Coding System.* Consulting Psychologists Press, 1978.

P. Ekman, W.V. Friesen, and P. Ellsworth. *Emotion in the Human Face.* Pergamon Press, New York, 1972.

M. Fabri and M. Gerhard. The Virtual Student. G. Orange and D. Hobbs (Eds.) *International Perspectives on Tele-Education and Virtual Learning Environments*, Ashgate, 2000.

R. Garrison. Quality and access in distance education: theoretical considerations. D. Keegan (Ed.) *Theoretical Principles of Distance Education.* Routledge, New York, 1993.

H-Anim. *Specification for a Standard VRML Humanoid.* H-Anim Working Group, http://www.h-anim.org, 2002.

M.L. Knapp. *Nonverbal Communication in Human Interaction.* Holt Rinehart Winston, New York, 1978.

D. Laurillard. *Rethinking University Teaching.* Routledge, London, 1993.

C.L. Lisetti and D.J. Schiano. Facial Expression Recognition: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect. *Pragmatics and Cognition.* 8(1):185-235, 2000.

M.G. Moore. Three Types of Interaction. , K. Harry, M. John, and D. Keegan (Eds.) *Distance Education: New Perspectives.* Routledge, London, 1993.

D. Morris, P. Collett, P. Marsh, and M. O'Shaughnessy. *Gestures, their Origin and Distribution*, Jonathan Cape Ltd, London, 1979.

R. Picard. *Affective Computing.* MIT Press, 1997.

S.M. Platt and N.I. Badler. Animating facial expression. *ACM SIGGRAPH Conference Proceedings.* 15(3):245-252, ACM Press, 1981.

I. Poggi and C. Pelachaud. Emotional Meaning and Expression in Animated Faces. A. Paiva (Ed.) *Affective Interactions*, LNAI 1814:182-195, Springer, 2000.

M. Slater and S. Wilbur. Speculations on the Role of Presence in Virtual Environments. *Presence.* 6(6):603-616, MIT Press, 1997.

D. Terzopoulos and K. Waters. Analysis and synthesis of facial image sequences using physical and anatomical models. *Pattern Analysis and Machine Intelligence.* 15(6):569-579, 1993.

T. Wehrle and S. Kaiser. Emotion and Facial Expression. A. Paiva (Ed.) *Affective Interactions.* LNAI 1814:49-63, Springer, 2000.

L.A. Zebrowitz *Reading Faces: Window to the Soul?* Westview Press, Boulder, Colorado, 1997.

Toward the Nature of Animation: An Architectural Approach

Carl B. Frankel
Stanford University, Stanford California
carlf@psych.stanford.edu

Abstract

The enterprise to endow artifacts with animation will benefit from architecting artifacts to have the substance of intentionality and emotionality. Toward understanding what it is to have the substance of intentions and emotions, this paper proposes an evolvable progression of control architectures that display progressively greater intentional and emotional capability. The progression exhibits the most rudimentary form of explicit intentionality from the point at which architectures include hedonics preference, a rudimentary form of emotionality. Incorporating additional intentionality yields multiple features of emotionality. The proposed progression thereby gives constructive substance to the inextricability of intentionality, emotionality and animation.

1 Introduction

What does it mean to be animated? In humans at least, animation is inextricably linked to having and pursuing intentions and also to having emotions. As such, it is reasonable to think that the animation of artifacts will be strengthened to the extent that artifacts embody the substance of intention and emotion, rather than the mere surfaces of goal orientation and of emotional display. While intentionality and emotionality are philosophical and psychological problems, the problem of giving embodiment to intentions and emotions is, in fact, as much architectural as it is philosophical and psychological.

This paper takes a concretely architectural approach that is informed by the central philosophical issues and psychological data. The result is a plausibly evolvable progression of control architectures with progressively more complete intentionality. This progression culminates in a control architecture that tightly fits the wealth of empirical data on human emotionality. This paper thus outlines a concrete architectural proposal for what it means to be animated with intentions and emotions.

2 Foundations

Construal of intentionality remains contentious. Mainstream approaches to intentionality start from Husserl's (1901/1970) observation that intentionality is characterized by explicit mental representations both of desires and goals and of "aboutness", the ways that behavior is about pursuing a future in which desires and goals are realized. From this common starting point, opinions diverge greatly as to the basis for inferring the existence of mental representations. Heider (1958) infers intentionality to be present when an organism uses rational and efficient means to accomplish a particular outcome even as circumstances change. For Dennett (1987), satisfaction of these criteria are the basis for taking an intentional stance even toward single celled organisms, side-stepping the issue both of mental representations and of their explicitness. By contrast, Searle (1997) believes that a mental representation is not explicit unless it is available in conscious awareness and self-report. The Buddhist analysis narrows the aperture even further, seeing self-report itself often to be a form of conditioned response. In this line of thinking, behavior is not fully or truly intentional unless the volitional process itself is fully and accurately exposed to conscious inspection. The range of mainstream ontologies of intentionality is so broad as potentially either to include amoebas or to exclude all but bodhisattvas.

An architectural construal of explicitness. If mental representations and their explicitiness are taken first and foremost to be an architectural problem, and only secondarily to be a phenomenological problem, a plausible solution emerges. In a phenomenological treatment, 'explicit' is taken to mean explicit in awareness. However, such narrow usage for 'explicit' leaves many unexplained gradations of intentionality in the creatures that phylogenetically precede humans. In this paper's architectural treatment, explicit is taken more broadly, to mean that there is information content that is explicitly represented *qua* information, that is, as a representation of the value of something that can vary. Information values can be given explicit representation *qua* information, for example, by being stored in a buffer, a register or a persistent memory. So recorded, the value of the information is stable for some period of time, during which it may be validly sampled.[1] With 'explicit' construed in this broader sense, intentionality unfolds in stages, from primitive to human, as registers and memories are added that make progressively more capable intentionality.

CogAff. Sloman (1999) has restricted his attention to evolvable progressions of architectures. In evolvable progressions, more complex architectures must extend simpler architectures, such that the new increment of complexity in each extension confers adaptive advantages that clearly offset the increment's costs. As a result of working within this restriction, Sloman proposes a three category ontology, CogAff, for the capabilities of architectures. *Reactive* architectures are hardwired to evaluate impinging contingencies and to deploy

[1] By contrast, information is implicit when it exists in a continuously changing transmission channel. This distinction between explicit and implicit is directly analogous to the Verilog hardware description language's distinction between `reg` and `wire` data types.

hardwired responses. *Deliberative* architectures explicitly consider possible futures, in the course of selecting and deploying responses. *Meta-managing* architectures explicitly consider their own operation in order to deploy improved patterns of response.

3 Progressively Intentionality

Originally a categorization of architectures' capabilities, CogAff can also be used to distinguish structurally among forms of explicitness of information in the control pathways of control architectures. Control architectures (or equivalently, self-regulating architectures) are distinguished by control pathways that feedback error signals, in order iteratively to pursue goals. When both goals and the future that goals are about creating are entirely implicit, a control architecture is fully reactive. When both goals and at least some aspects of possible futures are explicitly registered, a control architecture is deliberative. When goals and their aboutness are taken to be objects of thought in their own right in the course of adjusting the individual's own performance, a control architecture is meta-managing. CogAff thus characterizes a spectrum of intentionality in control architectures, from the fully implicit through the fully explicit to the fully self-aware. (Further analysis of meta-management is beyond the scope of this paper.)

Stateless control. The simplest possible control architecture attempts always to accomplish the same goal, no matter how circumstances change (see Figure 1). In general, such an architecture must attend to relevant information about circumstances, interpret that information, determine how the

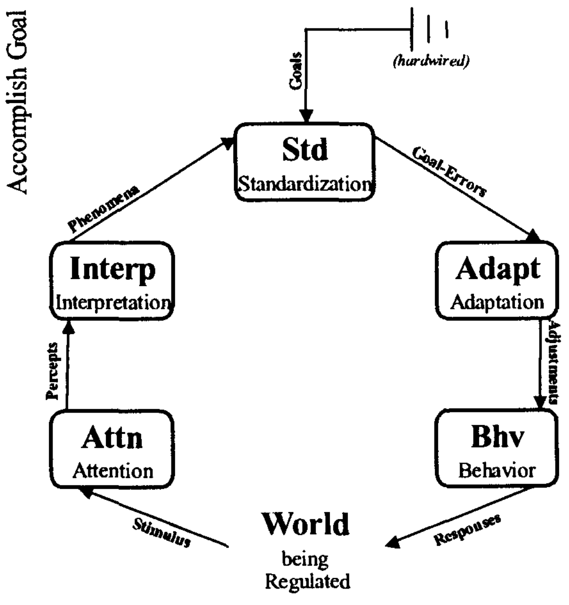


Figure 1: Stateless Control

current state-of-affairs diverges from the hardwired goal, decide what must be adjusted in current circumstances, and select a behavior that accomplishes the adjustment. An amoeba moving toward a region of favored temperature in a gradient is an example of such a system, wherein both the targeted ambient temperature and a reproductively successful future are fully implicit in the organism's chemistry.

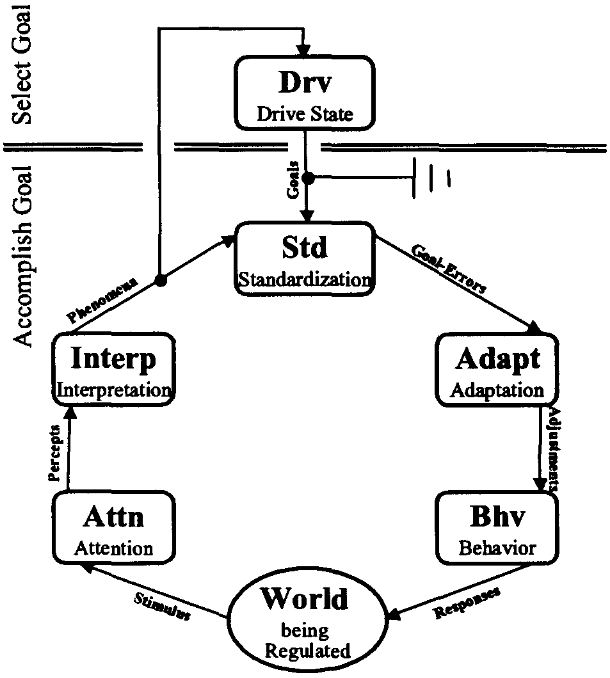


Figure 2: Drive

Drive. An important increment of adaptive advantage is conferred by adding a single element of explicit state to the control pathway that *drives* goal selection (see Figure 2) For example, Scheutz (2001) has evaluated an architecture that maintains a hunger state, and is driven to pursue food only when hungry. The hunger state explicitly represents the need for food, different from statically hardwired reactions like blinkng. This architecture takes a step toward explicit intentionality, since the goal to pursue food is the explicit result of the hunger drive state. Aboutness, however, remains implicit, since no explicit representation of future satiety need be represented to seek food when currently hungry.

Arbitrated drives. Scheutz's hunger state can be augmented by additional states, *e.g.*, sexual arousal, *e.g.*, mortal peril avoidance, *e.g.*, territoriality, that reflect an organism's multiple drives (see Figure 3). The individual can then select to seek food, to pursue a mate or to avoid a predator as these states explicitly represent the need to do so. When multiple drives impinge, contention can occur for sensors and effectors. The simplest architectural augmentation to resolve contentions is an arbitrator with fixed priorities. The arbitrator is a precursor of explicit aboutness; however, in this fixed-priority incarnation, no deliberation among explicit possible futures is involved, and so aboutness is still implicit and reactive.

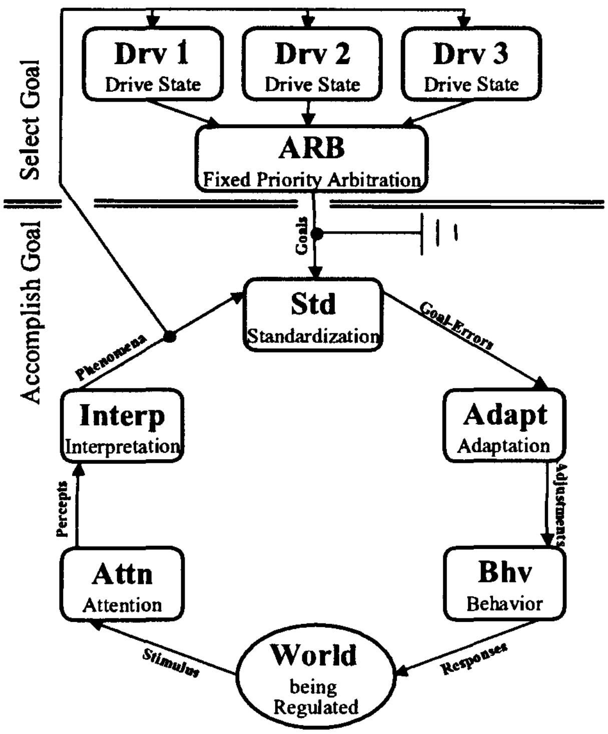


Figure 3: Arbitrated Drives

Intentional hedonic preference. The rigidity of fixed priority arbitration can be overcome by replacing fixed priorities with a dynamic preference ranking, based upon expected future hedonics (see Figure 4). The multiple drives in Figure 3 are subsumed into a discrete classification system that proposes palettes of relevant goals. However, rather than arbitrating goals based upon fixed priorities, goals are dynamically arbitrated based on the future hedonics expected to the situated result. Cabanac's (2000) preliminary data indicates that reptiles, birds and mammals exhibit hedonic preference selection, but not fish or amphibia.

Even though hedonics offer a very limited representation of possible futures, nonetheless, to the extent that expected hedonic valence adequately estimates likely future harm and benefit, dynamic preference deliberation allows targeted subordination of drives to fit current circumstances and demands. As such, *hedonic preference is the simplest control architecture in which both goals and aboutness are explicitly registered information.* Having both discrete classification and hedonics, it is also the first proto-emotional architecture.

Intentional husbandry of bandwidth. The rudimentary intentionality and emotionality of the hedonic deliberation architecture is made much richer by adding features that manage the allocation of processing bandwidth (see Figure 5). (*a*) The individual gives salience to the emotionally most urgent impinging contingency first as a result of interrupt prioritization. Because interrupts are generated by positive feedback events, there is rapid and complete switching of activities in case of emergencies. (*b*) While

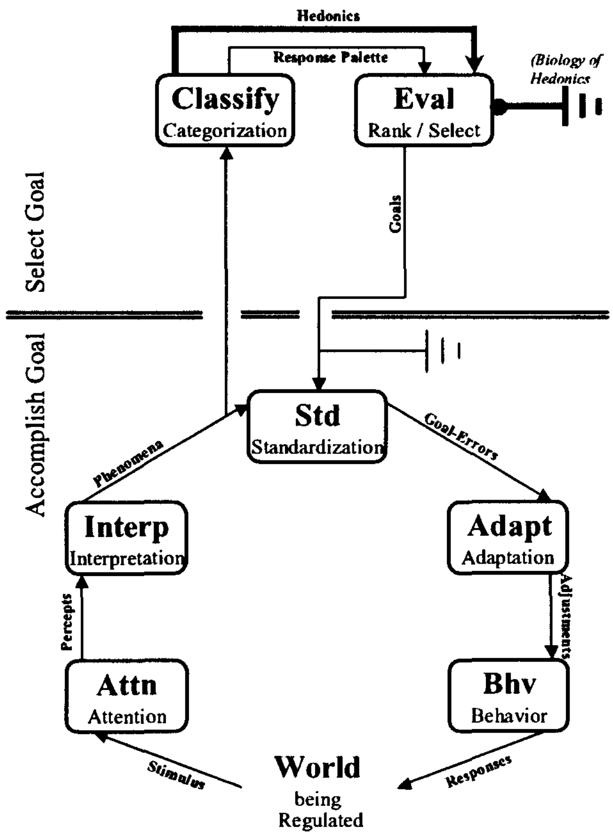


Figure 4: Intentional Hedonic Preference

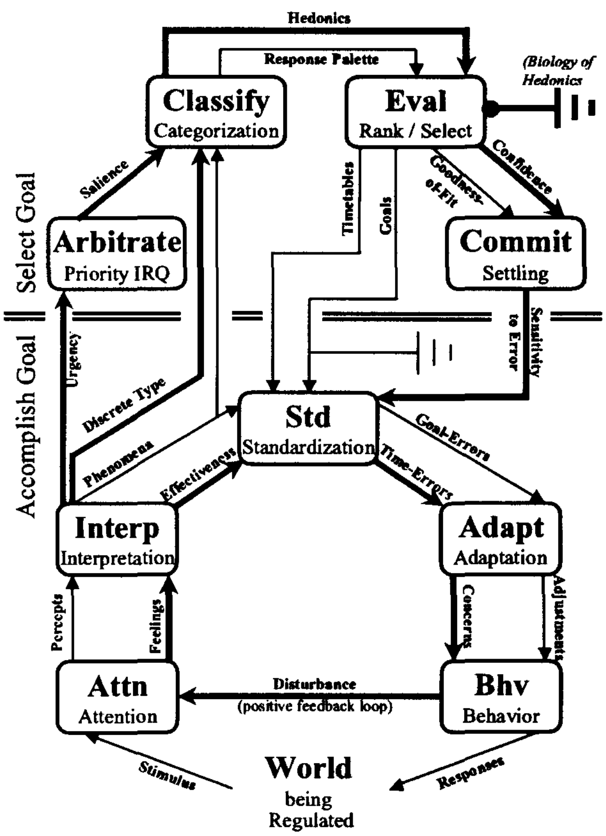


Figure 5: Intentional Husbandry of Bandwidth

positive feedback creates advantageous nonlinearity of response, it also creates the dangers both of excessive error scan and consequent uncontrollable reactivity. To harness the positive feedback loop, commitment to selected responses dampens sensitivity to error, which dynamically increases the hysteresis of onset of positive feedback events. As selected responses are believed to be both more appropriate and more realizable, emotional confidence increases. With increased confidence comes increased commitment. Sensitivity to error is thereby decreased and so also is the likelihood of a cascade of positive feedback events. Conversely, when anxiety increases, so also does error scan, decreasing hysteresis and increasing perturbability.

Intentional sustainability. The emotional future that is likely in an individual's adaptive niche is explicitly represented by accruing summary expectancies that tailor the individual's patterns of response toward maintaining a tenable position in her or his niche (See Figure 6). Pursuing unattainable levels of positive emotion (benefit) and avoiding inevitable levels of negative emotion (harm) wastes resources and jeopardizes sustainability. Using summary expectancies accrued from experience and observation of emotions, as well as education and acculturation about emotions, more wasteful options are rejected, while more tenable options are considered.

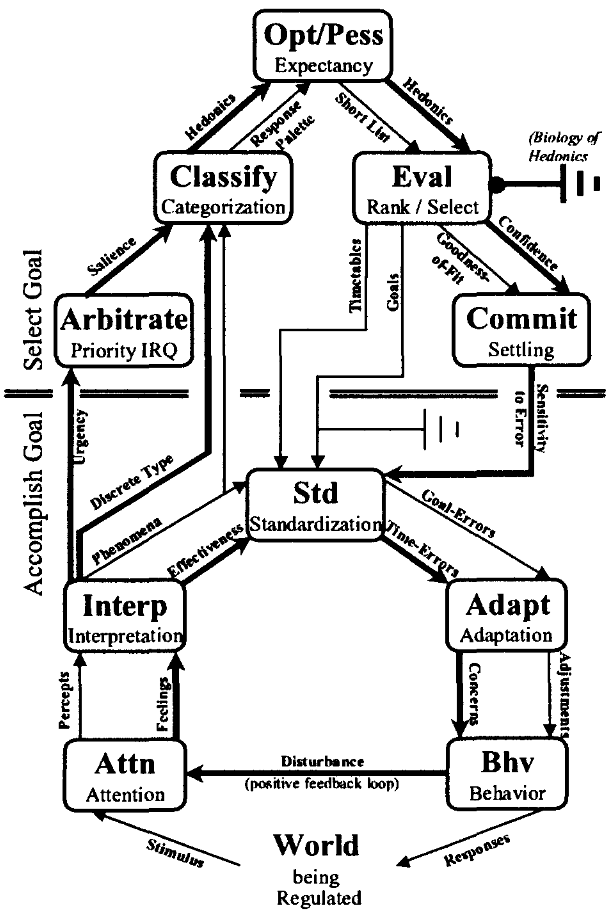


Figure 6: Emotional Intentionality

Adding summary emotional expectancies, dispositional optimism and pessimism, produces the emotionally governed adaptive control architecture, proposed and documented by Frankel and Ray (2000) to make a tight fit across the broad empirical literature on emotion.

4 Conclusion

The fundamental connection binding animation, emotion and intention has long been held as a content of human intuition. For example, whether from the biblical Hebrew 'lev', the ancient Greek 'kardia' or the ancient Chinese 'hsin', the heart has long been understood to be the animating organ both of emotion and of intention.

The succession of architectures sketched in this paper proposes a rich, constructive embodiment of the intentional and emotional underpinnings of animation. *Philosophy.* Crucial for understanding the nature of animation, the architectures are grounded directly in the central terms from the discussion of intentionality. The progression illuminates a new use of 'explicit' that characterizes specific gradations of intentionality, from amoebae to primates. *Psychology.* Emotion is a common feature of animation. With the introduction of hedonics, the architectures start to make a fit with the data on human emotions. The last architecture, emotionally governed adaptive control, makes a tight fit. *Evolution.* Animation includes the broad range from tropisms to self-conscious volitional acts. The architectures form an evolvable progression from reactive tropisms through emotionally deliberative behavior, and thus offer a plausible narrative in which natural selection favors intentionality and emotionality. *Control architecture.* The architectures in the progression are textbook models from control theory.

Much work remains to extend this model to cover self-conscious volition. Nonetheless, the present succession of architectures suggests a pathway for architects of animated artifacts to produce the foundations of the substance of intentionality and emotionality.

References

Cabanac, M. (2000) The phylogeny of emotion. Plenary address at ISRE 2000, International Society for Research in Emotion. Quebec City, QC, Canada. August.

Dennett, D. C. (1987) *The intentional stance.* Cambridge, MA: MIT Press.

Frankel, C. B. & Ray, R. D. (2000). Emotion, intention and the architecture of adaptively competent information processing. In Proceedings, AISB2000. Society for the Study of Artificial Intelligence and the Simulation of Behaviour. Birmingham, England. April.

Heider, F. (1958) *The psychology of interpersonal relations.* New York: John Wiley.

Husserl, E. (1901/1970) Logical investigations. J. N. Findlay (Tr.). New York: Humanities.

Scheutz, M. (2001) The evolution of simple affective states in multi-agent environments. In, *Proceedings, AAAI 2000,* Seattle, WA.

Searle, J. R. (1997) Minds, brains and programs. In J. Hoagland (Ed.), *Mind design 2: Philosophy, psychology, artificial intelligence, 2nd Ed.,* (pp. 183-204). Cambridge, MA: MIT Press.

Sloman, A. (1999) What sort of architecture is required for a human-like agent? In M. Wooldridge and A. Rao (Eds.), *Foundations of rational agency.* New York: Kluwer.

Autonomous Secondary Gaze Behaviours

Marco Gillies†; Neil Dodgson¥; Daniel Ballin ‡

† Department of Computer Science, UCL@Adastral Park, Ross 1, Adastral Park, Ipswich, IP5 3RE, Suffolk, U.K.

¥University of Cambridge Computer Laboratory, William Gates Building, JJ Thomson Avenue, Cambridge CB3 0FD, UK

‡ Radical Multimedia Laboratory, BTexact Technologies, Ross 1, Adastral Park, Ipswich, IP5 3RE, Suffolk, U.K.

†marco.gillies@bt.com; ¥neil.dodgson@cl.cam.ac.uk; ‡daniel.ballin@bt.com

Abstract

In this paper we describe secondary behaviour, this is behaviour that is generated autonomously for an avatar. The user will control various aspects of the avatars behaviour but a truly expressive avatar must produce more complex behaviour than a user could specify in real time. Secondary behaviour provides some of this expressive behaviour autonomously. However, though it is produced autonomously it must produce behaviour that is appropriate to the actions that the user is controlling (the primary behaviour) and it must produce behaviour that corresponds to what the user wants. We describe an architecture which achieves these to aims by tagging the primary behaviour with messages to be sent to the secondary behaviour and by allowing the user to design various aspects of the secondary behaviour before starting to use the avatar. We have implemented this general architecture in a system which adds gaze behaviour to user designed actions.

1 Introduction

In general when a user controls an avatar they will give a sequence of commands that will be transformed into the avatar's motion. The method of giving commands will vary greatly, from simple mouse and keyboard interfaces to various forms of body tracking. Though purely text based interfaces to avatars are rare, there is one exception, which is that some avatars are controlled by the user typing text into an interface, that the avatar then speaks. The level of control also varies greatly, the system might map the users motion directly onto the avatar or the input might be a much more discrete set of higher level commands, for example a command to walk to a particular position.

What these methods have in common is they are incomplete. They cannot specify the entire behaviour of the avatar. Even full body tracking systems cannot capture the full detail of the user's motion and expression (of course this would normally not be desirable, part of the appeal of using avatars is that they add some thing new to the action of the user whether it is in graphical appearance or behaviour). As discussed above this leaves various aspects of the avatars behaviour that must be determined by the system.

2 Primary and Secondary Behaviour

We divide the avatars behaviour into two types: primary behaviour that is explicitly specified by the user input, and secondary behaviour that is automatically generated by the system. Primary behaviour should correspond to large scale goal directed actions, such as moving around or manipulating objects. They should be the sort of actions that people consciously decide to do, thus making them easy for users to specify. Secondary behaviour on the other hand should be smaller scale, and should tend to correpsond to more sub-conscious behaviour such as non-verbal communication (or at least behaviour that is not the main focus of the user's action). For example, a primary behaviour would be invoked if the user requests the avatar to pick up a telephone and to start talking. Secondary behaviour accompanying this might be a head scratch or fiddling with the telephone cord.

We envisage that secondary behaviour should be controlled by a number of independent behavioural agents. They would produce their behaviour continually, in parallel with any user controlled behaviour. Though the secondary behaviour is not explicitly specified by the user it should be influenced by the user's input. In particular the details of the secondary behaviour should depend on what primary actions the avatar is performing.

2.1 Motivation

However graphically appealing avatars can be they will not become successful if their animation and behaviour is not compelling. It is thus very important that the behaviour of avatars is expressive. For this to be possible the behaviour must capture the nuance and complexity of human non-verbal behaviour. This is not possible solely through the user controlling the avatar directly. Most input devices will not provide enough information to animate details such as the character's gaze behaviour, gesture or facial movements. A mouse and keyboard interface is far to impoverished to control all of these in real time. Even a full body tracking system can miss significant information. Also full body tracking is not suitable for cases where the user does not want their body language directly mapped onto the avatar, for example, to hide their feelings or to make the avatar's behaviour more stylised or more expressive. Directly controlling the character's expressive behaviour would also be a large cognitive load on the user who is likely to want to concentrate on the task in hand. They would probably not pay much attention to the details thus resulting in avatars that are not expressive in practice. Finally, even if the user could control the avatar's expressive behaviour directly much of the behaviour is sub-conscious and so the user might not know how to produce appropriate animations if they are not a skilled animator. All of these reasons indicate that expressive behaviour should be generated autonomously. However, it is still important that the behaviour should be relevant to what the character is doing, i.e. what the user has commanded it to do. Thus we divide the avatars behaviour into primary and secondary behaviour and ensure that the secondary behaviour is autonomous but influenced by the primary behaviour.

2.2 User design of secondary behaviour

One aspect of secondary behaviour that we consider very important is that the user should be able to control many aspects of the avatar's secondary behaviour. This should be done by giving the user tools with which to shape the secondary behaviour before starting to use the character. This allows us to harness human creativity in the process of producing expressive behaviour. People can be excellent at creating the subtleties of human expression and the ability to harness this can add a lot to an expressive agent. This form of user input can also provide a large degree of individuality to an avatar. It can be particularly important when the avatar or environment is stylised so that what is needed is not realistic behaviour but some form of stylised behaviour (cartoonish, exaggerated etc.). As this design is done before the avatar is used it does not force the user to spend time controlling the secondary behaviour in real time but still allows them control over the avatars behaviour.

3 Previous Work

Vihljámsson and Cassel (1998) discuss the importance of autonomous behaviours for avatars, which are equivalent to what we call secondary behaviour. They use these autonomous behaviours for conversation in their BodyChat system. Similar ideas are present in BEAT (Cassell, Vilhjámsson and Bickmore 2001), which generates conversational gesture from text. This work differs from ours in that we are looking at graphical control of physical, non-conversational primary behaviour. Another related system is the multi-level control of Blumberg and Galyean (1995), this allows the user to control some aspects of a characters behaviour, at various levels of abstraction with the system generating others, for example, wagging the tail of their dog character. Finally, somewhat related is the current interest in transforming pieces of motion to express new emotions or personality (e.g. Rose, Bodenheimer and Cohen 1998, Amaya, Bruderlin and Calvert 1996 and Polchroniadis 2000). Though these do not really count as autonomous behaviours they do add secondary features to a motion.

4 General Architecture

Figure 1 gives an overview of the architecture that is being proposed for primary and secondary behaviour. The primary behaviour is controlled by direct user commands. The Secondary behaviour is a separate module (or set of modules) that is not directly influenced by user input and which acts to a large degree autonomously.

The secondary behaviour module is controlled by a number of tags attached to the primary behaviour. These tags are attached to particular points in the primary behaviour and indicate that a message should be sent to the secondary behaviour when that particular point occurs in the primary behaviour. For example, in a conversational system a tag could be attached to the point at which the avatar stops speaking and this could result in various secondary actions being requested from the secondary behaviour module, for example, looking at the conversational partner.

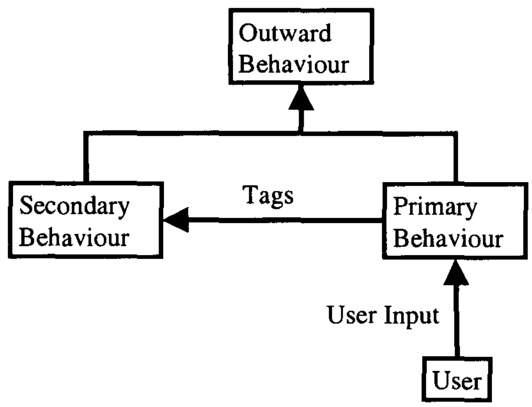


Figure 1: The relationship between primary and secondary behaviour.

These tags can be controlled by the user designing the secondary behaviour. There are two aspects that can be controlled, the points in the primary behaviour at which the tags are placed and the message that is sent by a tag. There are also two types of user who will edit these tags. The first will be designing the secondary behaviour system in general. They will design how the secondary behaviour relates to the primary behaviour and so will mostly add tags to the primary behaviour. They will also edit the messages contained in the tag to some degree. The second type of user will create the secondary behaviour for a particular avatar and will add details to the system designed by the first user. They will mostly edit the messages contained in the tags. For example, the first user might add a tag requesting that the avatar should look at the partner at the end of an utterance while the second user might indicate whether this should be a brief glance with just the avatars eyes or whether the avatar should orient itself towards the partner with its head and shoulders and look at the partner for a longer time.

5 Example: Eye Gaze

We have implemented an example of this general architecture for generating eye gaze while an avatar obeys commands given by the user. The simulation of gaze behaviour has been studied extensively for conversation, for example, Vilhájmsson and Cassell 1998, Colburn, Cohen and Drucker 2000 and Garau, Slater and Sasse 2001. We have looked instead at gaze behaviour in non-social situations. This has been studied by Chopra-Khullar and Badler but they did not investigate in detail how to integrate simulation of gaze with user control of the avatars actions. We focus on creating tool by which a user without programming knowledge can create both primary actions that the avatar can perform as requested by the user and gaze behaviour that will accompany these primary actions.

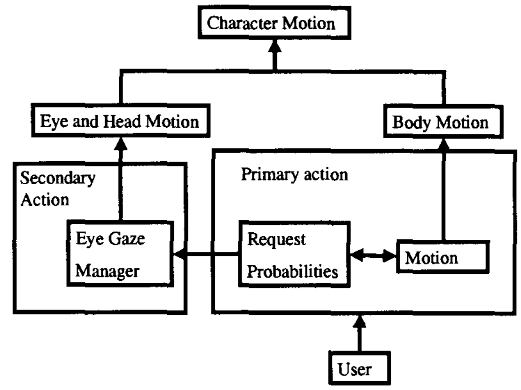


Figure 1: The Gaze Behaviour

5.1 Primary behaviour

The aim of our primary behaviour is to supply a framework for actions that can both be easily designed by a user and also easily invoked by the user in real time. When creating an action the user starts with a piece of motion, this motion is a specific example of the action to be designed. The user then specifies targets of the motion (objects that the character interacts with during the motion). For example, a target for a drinking motion would be a cup. Targets can also be objects that the character looks at but does not touch, for example, the character might be drinking in a cafe talking to a friend. This friend might be a target so the character would look at her.
The user can invoke these actions by specifying targets, which is just done by clicking on them. When an action is invoked the original motion is transformed so the character interacts with the targets correctly. For example, if the action consists of picking up and drinking a cup the motion will be transformed so that while picking up the cup the hand will move to the correct position of the specified cup rather than the position that it moved to in the original motion. The motion is transformed with techniques similar to those used by Polichroniadis (Polichroniadis 2000). The full details of the method is described in Gillies 2001.

5.2 Gaze behaviour

The secondary behaviour consists of gaze shifts which are controlled by an eye gaze manager (described in more detail in Gillies and Dodgson 2002a). This receives requests for a gaze in a particular direction or at a particular object from the primary action. These requests can be of different types, for example, an immediate request causes the avatar to look at the target as soon as the manager receives the request while a monitor request just causes the avatar to look at the target occasionally until it is told to stop by another request. The eye gaze manager animates the avatar as looking at the target of these requests. When no requests are sent the manager continues to produce gaze behaviour, generating requests itself.

The eye gaze manager is controlled by a number of parameters that influence the avatar's gaze behaviour. For example, observing people we noticed that they vary their horizontal angle of gaze but kept their vertical angle relatively constant. Thus we introduce two parameters to control the characters behaviour, a preferred vertical gaze angle and a probability of maintaining this angle. These parameters can be altered by the user to change the avatars behaviour. These parameters can be set in advance and allow the user to alter the avatars behaviour in general without having to do anything while actually using the avatar.

Some parameters can also be set for individual requests. For example, the length of gaze is controlled by a number of parameters for the character, however, it can also be set by parameters of a request. This allows two levels of control. The parameters of the gaze manager can change the avatar's gaze behaviour in general, while the request parameters change how it behaves in particular circumstances. These parameters allows different avatars to have a range of different gaze behaviours using the same architecture.

Finally, the avatar's gaze behaviour can be influenced by external events, for example, various objects can capture the avatar's attention, for example, moving objects. Also when there are no requests the avatar will look at objects in the environment tagged as interesting.

5.3 Tagging primary behaviour

As described in section 4 above the primary behaviour is tagged with messages that are sent to the secondary behaviour module at various points in the primary behaviour. These messages are to ensure that the secondary behaviour produced is appropriate to the primary behaviour. In this case the messages consist of eye gaze requests.

When the user creates a primary action from a piece of motion she divides the motion into a number of periods. These periods correspond to meaningful sub-sections of the motion. For example, for a motion of picking up and drinking a cup of coffee the periods might be, reaching out towards the cup, picking up the cup, bringing the cup to the avatar's mouth, drinking and putting the cup back down. Tags can then be placed at the start of these periods.

The user sets the tags by specifying that particular types of requests can be sent out at the start of each period. These requests will be to look at one of the targets of the action. Various parameters of the request can also be specified, for example whether it should be a short glance or a longer gaze; whether the avatar should move its head, and how often the avatar should look at a target if the request is a monitor request. If the gaze behaviour produced were identical every time the action was invoked it would seem very mechanical if the action is invoked a number of times. To prevent this the user actually specifies probabilities of sending a particular type of request. When an action is invoked requests are generated at random using these probabilities. The probabilities are set using a number of sliders as shown in figure 2.

When the user first creates the action they can set these probabilities, however, it would also be desirable to be able to alter them to create different personalities for different avatars. Users can therefore alter the probabilities later. We separate these two editing stages by making the original designer's edits effect what later edits are possible. These original edits might just be default values that users can later change. However, some values might be inappropriate, for example, when drinking it would always be appropriate to keep the avatars head still so the original designer might want that edit to be permanent, so that users cannot later change it. The original designer can indicate that an edit is a fixed value and so it will not appear in the users dialog-box for later editing. The original designer can also set particular edits as minimum or maximum values.

5.4 Results

Figures 3 and 4 give examples of actions with eye gaze attached. The first is of an avatar drinking from a can. The underlying gaze parameters are set so that the avatar has a tendency not to look around itself and to mostly look downwards when there are no explicit requests. There are two requests tagged to the actions. The avatar looks at the can before picking up and then at the other avatar at the last frame, this time just glancing and moving its eyes without turning its head. The behaviour of not looking at the other avatar in general and when looking doing so without a head move might indicate avoiding the gaze of the other avatar. The Second example is of an action where the avatar picks up an object and puts it down somewhere else. Here the avatar looks around itself more. There are two tagged gaze requests, to look at the object as it is picked up and at the shelf as it is put down. This time, when the character does not have a request between the other two it looks at a location in the distance.

6 Further Work

Both secondary behaviour in general and our particular system have a large potential for further development. We have only implemented one particular type of secondary behaviour applied to one particular type of primary behaviour. Thus there is much potential for exploring new types of secondary behaviour, for example, gesture or facial expression. Also there is a lot of potential for applying these to other types of primary behaviour, for examples, conversation or other behaviour patterns that are too complex for to fit into the framework described here. The BEAT system aims at attaching gesture to speech (see Cassell, Vilhjámsson and Bickmore 2001). The authors have also investigated adding eye-gaze to a more complex primary behaviour, navigating an environment (described in Gillies 2001 and a forthcoming).

The tool we have described here is still a prototype and needs to be made more robust and tested by creating a wider range of actions and performing user tests. In particular we would like to develop it into a tool that can be used in shared virtual environment and assess people's perception of avatars using our secondary behaviour.

One aspect that we would like to improve is the user interface for adjusting the various parameters of the secondary behaviour. These allow the user a degree of control over how a particular avatar performs its gaze behaviour. However, these are currently edited using a large set of sliders that directly effect the parameters, some of which are rather unintuitive, we would like to provide a more sophisticated and intuitive design tool.

Though this model of eye gaze is reasonably general it is not quite sufficient to model the nuances of interpersonal eye gaze in social situations, we would therefore like to include more heuristics for social situations.

7 Conclusion

We have explored the idea of secondary behaviour as a semi-autonomous system that controls aspects of an avatars behaviour, leaving the user free to control more important aspects. We have described an application of these ideas to generating eye gaze. We think this has provided a good demonstration of our general architecture and are pleased with our initial results, however, we are keen to develop these ideas further.

Acknowledgements

The work described in section 5 was done at Cambridge University Computer Laboratory and funded but the UK Engineering and Physical Sciences Research Council. The Continuation of this work is being funded by BTExact technologies. The authors would like to thank the members of the Cambridge University Computer Lab Rainbow research group, the radical multi-media lab and the UCL computer graphics and VR group, and in particular Tony Polichroniadis, for a conversation that sparked some of these ideas.

References

Kenji Amaya, Armin Bruderlin, Tom Calvert, Emotion from Motion, in *Proceedings of Graphics Interface '96*. pp. 222-229, 1996.

Bruce Blumberg and Tinsley Galyean, Multi-Level Direction of Autnomous Creatures for Real-Time Virtual Environments, In *Proceedings of ACM SIGGRAPH* 1995

Justine Cassell, Hannes Högni Vihljámsson and Timothy Bickmore, BEAT: the Behavior Expression Animation Toolkit, *in Proceedings of ACM SIGGRAPH* 2001

R. Alex Colburn, Michael F. Cohen and Steven M. Drucker, The Role of Eye Gaze in Avatar Mediated Conversational Interfaces *Microsoft research Technical Report* MSR-TR-2000-81

Maia Garau, Mel Slater and Angela Sasse, The Impact of Eye Gaze on *Communication using Humanoid Avatars, in Proceedings of ACM SIGCHI* 2001

Marco Gillies, Practical Behavioural Animation Based on Vision and Attention, University of Cambridge Technical Report TR522 2001

Marco Gillies and Neil Dodgson, Eye Movements and Attention for Behavioural Animation, under review *Journal of Visualization and Computer Animation* 2002 (a)

Tony Polichroniadis, High Level Control of Virtual Actors, *PhD Thesis, Cambridge University Computer Laborotory*, 2000.

Charles Rose, Michael F. Cohen, Bobby Bodenheimer, Verbs and Adverbs: Multidimensional Motion Interpolation, *IEEE Computer Graphics & Applications. 18 (5)*, pp. 32-40, 1998.

Hannes Högni Vihljámsson and Justine Cassell, BodyChat: Autonomous Communicative Behaviors in Avatars, *In Proceedings of the 2nd Annual ACM International Conference on Autonomous Agents*, Minneapolis, 1998

Daniel Ballin, Matthew Lawson, Mary Lumkin, Jo Osborne, Personal Virtual Humans: Inhabiting The Talk Zone and Beyond, *BT Technology Journal*, Kluwer, Vol(20) No1, January 2002.

Daniel Ballin, Ruth Aylett, and Carlos Delgado , Towards the development of Life-Like Autonomous Characters for Interactive Media, *BCS Conference on Intelligent Agents for Mobile and Virtual Media, National Museum of Film and Photography*, April 2001.

Robert Trappl, Paolo Petta, Creating Personalities for Synthetic Actors, *LNAI* Springer, 1997

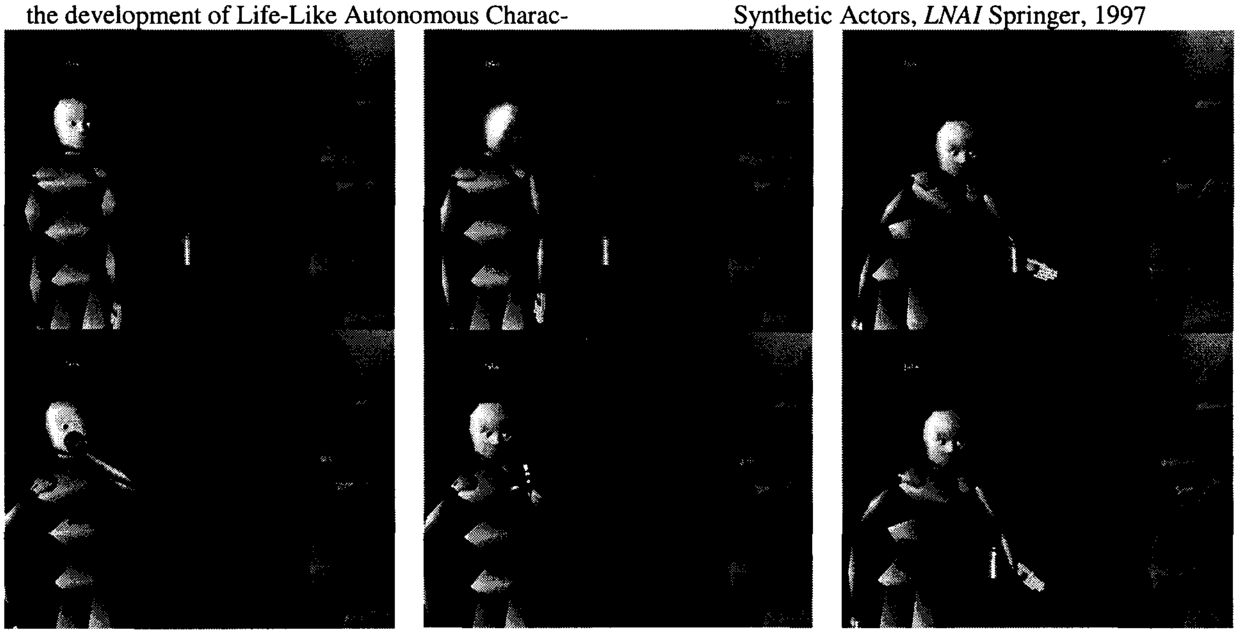


Figure 3: An action of an avatar drinking from a can

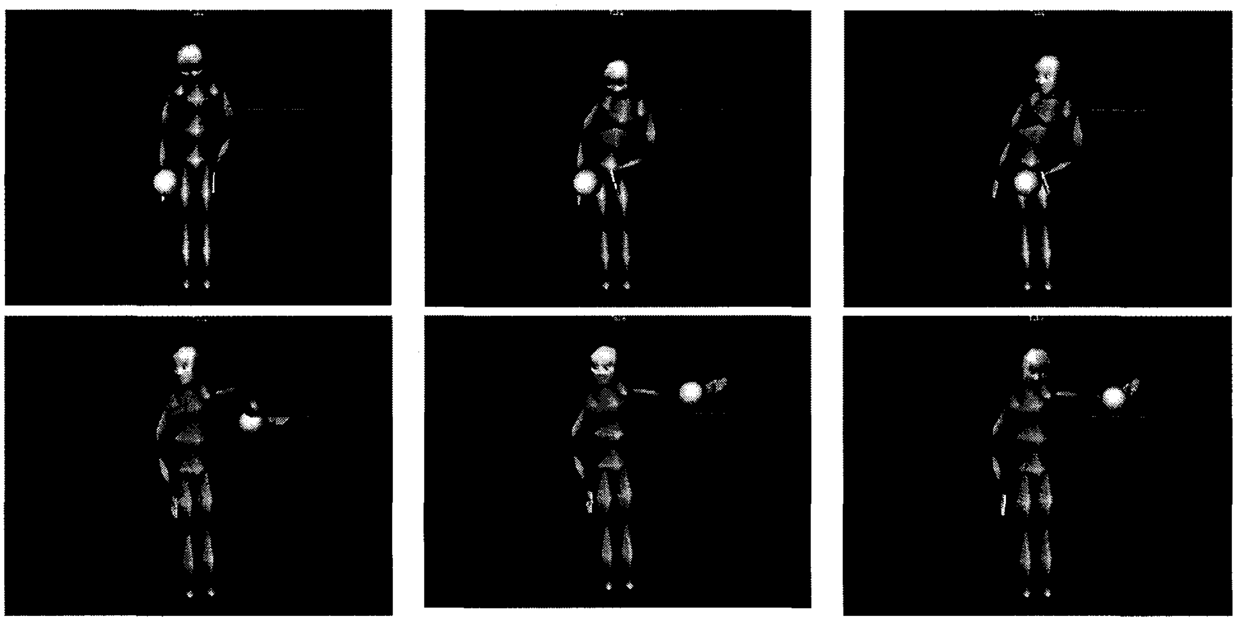


Figure 4: An action of an avatar picking up an object and putting it down somewhere else

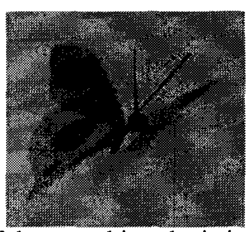
The Butterfly Project

Dr. Lizbeth Goodman; Ken Perlin
The Smartlab Centre; Central Saint Martins College of Art and Design, Southampton Row, London WC1B 4AP United Kingdom;
NYU Media Research Lab; 719 Broadway, 12th floor, New York, NY 10003
l.goodman@csm.linst.ac.uk; perlin@cat.nyu.edu

Abstract
The butterfly project uses technology to extend the boundaries of puppetry and computer graphics. Moving towards intuitive experiences with technology, this project offers the opportunity for both performers and audience to interact with a computer-generated 3D butterfly.

1 Introduction

This project is collaboration between the *Virtual Interactive Puppetry* project of the Smartlab Centre in London and the Media Research Laboratory at New York University. Various people are working on this project. The long-term purpose of The Butterfly Project is to develop tools that allow people to work with computer generated objects as though they were real, freely and intuitively defying space and scale. The short-term goal of this project is to allow a puppeteer to work with computer graphic avatars, such as a virtual butterfly,



with much of the same kinesthetic immediacy and facility with which the puppeteer can work with real puppets. Of course, virtual puppets can do things that it is very difficult to do with real puppets, such as freely change their shape, size and color, make arbitrary sounds, and appear simultaneously in different places and scales.

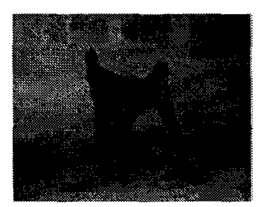
1.1 3D Perception

1.1.1 Puppeteer Stereo Vision and 3D Sound

The first goal will be to enable a creation space in which the puppeteer can perceive, simultaneously with responsive stereo vision and 3D sound, graphically generated puppets that appear to be within the puppeteer's hand/eye space. For this purpose, the puppeteer wears passive stereo glasses (alternately polarized) and wears wireless earphones. Three small LEDs are mounted on the glasses, and two video cameras are trained on the puppeteer from above. The positions of these three LEDs are continually tracked and analyzed by computer, in order to follow the position and orientation of the puppeteer's head.

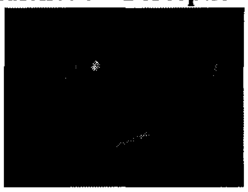
This information is used to vary the computer graphic imagery and 3D sound which the puppeteer perceives, creating the illusion that it is floating in 3D space in front of the puppeteer.

1.1.2 Miniature 3D Live Actor



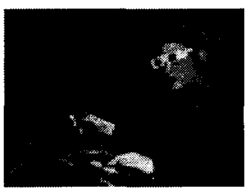
With stereo vision and 3D sound, the puppeteer sees a miniature version of a live actor or dancer. For this purpose, the actor's performance is captured by two video cameras, spaced far apart to create hyperstereo. For example, if the puppeteer is expected to operate at one tenth normal scale (thereby seeing the actor as five to six inches tall), then the two video cameras need to be separated from each other by about ten times interocular distance, or about 25 inches.

1.1.3 Audience 3D Perception



An audience in the theatre looking at the dancers and wearing passive stereo glasses will be able to perceive the dancers interacting with life-size versions of the computer-generated puppets.

1.2 Puppet Controllers



In one scenario, the puppeteer holds a Polhemus or other 6-DOF sensor in one hand, and controls a set of MIDI sliders, or some equivalent parametric controller, in the other hand. The hand that holds the sensor literally flies around the position and base orientation of the puppet. The other hand varies behaviors. In the case of a butterfly, these behaviors might include rate of wing flap, magnitude of wing flap, spin and looping. The puppeteer hears synthetically generated sound from the puppet from the proper 3D location. In the case of a butterfly, one sound might include wing flapping, which will vary in rhythm, pitch, amplitude and timbre, depending upon how the position and behaviors are set.

Because the rear projection in the live performance cannot project the animated puppet so as to be directly in front of the actor (and thereby hiding the actor from the audience), it is the puppeteer's responsibility to ensure that during the performance, the puppet never directly upstages the actor. However, the puppet can be made to appear to fly nearer to the audience than the actor, as long as it does so while vertically higher than the top of the actor's head.

1.3 Developing Software

A variety of new software is being developed for this project. For example, we are currently creating a software tool to convert the image of three LEDs, as seen by two overhead cameras, into the position/orientation of the puppeteer's head. This work will be making use of Intel's Open Source Computer Vision Library.

FABRICATING SOCIAL ABILITY IN ANIMATED AGENTS

Lynne Hall, Simon Oram
School of Computing and Technology, University of Sunderland, Sunderland, SR6 ODD
School of Computing and Mathematics, University of Northumbria, Newcastle, NE1 8ST
lynne.hall@sunderland.ac.uk; simon.oram@unn.ac.uk

Abstract

This paper discusses an animated agent that appears to exhibit social ability, using ideas from French post-structuralist social theory. This presents a departure from the cognitive models frequently used to consider agents and is based on Foucault's notion of man as a fabrication or set of fictions. Adopting this view of the individual removes the need for agents to simulate social ability. Rather, it requires that a clever imitation of such behaviour is constructed. This imitation of social ability is discussed using the example of a mobile, embodied agent that can traverse a network and organise meetings amongst a number of individuals. Early evaluations suggest that users find the agent socially able and their interactions with it enjoyable.

1 Introduction

In *The Protestant Ethic and the Spirit of Capitalism* (Weber, 2001), a genealogy is provided, from which the modern bourgeois self emerged. The subjectivity of this identity is primarily drawn from a dyadic relationship with himself and a silent other (God). From this grew a notion of individual self-reflection and the ability to guide oneself through life, making rational decisions. It is this view of the individual that has dominated agent development, with the ultimate goal of traditional AI being to produce autonomous entities capable of making rational choices (Wooley, 1992).

There has been considerable questioning of such notions of rationality, subjectivity and agency (Craib, 1993). What Weber describes as subjectivity – that is a rationality that is somehow separate and prior to identity, is for Foucault merely an effect of enlightenment metaphysics (Rabinow, 1984). These classic norms saw the construction of an individual's self through pre-defined roles or life-scripts, such as school, work, sexuality and religion. The notion of an enlightenment rationality posits that individuals created their own life-script by moving through these pre-scripted discursive structures, and from living through these structures coupled with a disengaged rational self-reflection constructed their identity (Hetherington, 1998).

These life scripts can be understood in Foucauldian terms as discourse within which the individual is bound, observed and normalised (Foucault, 1977). These discourses or bounded areas of knowledge pre-date individuals, forming the basis of worldview, social manners and mores. This notion of the individual contrasts strongly with the view of an individual based on disengaged, rational, self-reflection and coupled with Foucault's focus on the narrative rather than the rational offers potential for social agent design and development.

Foucaults' notion of man as a fabrication, living out a series of fictions with actions constrained or produced by closed discursive worlds offers a framework for social agents. Such discursive worlds would see an agent carrying out its roles, making decisions about what is to be done and about how the agent interacts socially. These decisional mechanisms result in agent actions that are configured within, and for, a particular situation.

In this paper, we discuss the development of a social agent that is situated within the workplace. The design of the discursive world creates a textual framework of utterances with which the agent can build a dynamic model of what can be said and cannot be said, what language the workplace denies and also, of what it produces. Although many narrative threads could be identified, here we focus on organisational hierarchy. From this context, language is drawn that the agent uses in its interactions with users, including passive acceptance of role, sycophancy towards those in power, dismissal of those lower within the hierarchy, etc. Other narrative threads are created through the social categories of age, gender and relationship between user and agent.

2 Social Ability and Animated Agents

Animated software agents can be characterised as entities that have some sort of form or body that they 'inhabit.' They are presented to the user as sensory representations, based on the use of graphics, movement and audio. Animation offers the opportunity to significantly improve the clarity and dramatic impact of communication through the use of emotive speech and gestures (Lester et al, 1997) and permits the agent to interact appropriately with humans (Andre, 1999).

Given their corporeality, agent systems exploit interaction potential by implementing non-verbal behaviour, such as emotions, facial expressions and mannerisms. By the careful orchestration of body movements, hand and facial gestures, as well as certain verbal nuances, the agent can convey positive and negative responses (Cassell & Thorisson, 1999b). Such responses can create rich interactive relationships between the user and the agent (Cassell & Vilhjalmsson, 1999a). However, such relationships are related not to physically sensing the agent, but rather to the user's internal conceptualisation of it. For agents to be believable they must exhibit human characteristics such as persona, empathy and sociability. For the user, the agent's internal representation or 'personality' must be consistent and match its external behaviours, for any level of credibility and believability to be assigned to the agent.

The success of animated agents remains limited, with current results offering contradictory views of how enjoyable, productive and useful such agents are for users (Dehn & van Mulken, 2000). Research related to agent's social ability frequently focuses on emotions and believability, although there is growing interest in the notion of social role awareness. Hayes-Roth (Hayes-Roth, 1998) studied role-specific interaction from the position of a master-servant relationship, with agent positions that were essentially fixed. Believability was achieved by drawing from drama and literature, with the agent having a specific, narrow, defined role.

(Prendinger & Ishizuka, 2001) propose a more dynamic use of social ability to define roles where multiple animated agents converse with one another. This conversation is based on the use of a social filter module that regulates the possible verbal expression in a given context. Social roles are augmented by context driven conventions, social relations and social distance. The agents of the Oz project (Bates, 1994), (Mateas, 1997) also have social awareness and attempt to create an *illusion of life*. The agents have social abilities that evolve through interactions, however, these abilities are defined by the characters' role within the plot. The agents exist within fictional, dramatic environments separated from the context within which they are used.

Prendinger's use of a social module that mediates conventional practices, initially appears to be similar to Foucault's notions of discourse. However, the social filter module is used in an attempt to simulate cognition and reality, whereas a Foucauldian approach attempts to fabricate fictions.

If fiction and fabrication are used to frame an agent's behaviour, this moves from an attempt to simulate real phenomena to an attempt to imitate such phenomena. Effectively, this approach leads us not to a model of what a social agent is, but rather to a position that suggests what a social agent cannot be. Agents cannot be human, they can only be at best fictions of humanity. Viewing agents as fabrication rather than simulation offers considerable scope, there is no interest in whether the agent is actually socially aware, the task at hand is to develop a system that convinces the user to think that it is. Like the Furby, an animated agent can be created using 'tricks' to imitate social ability.

3 Creating the Illusion of Social Ability

PACEO (Personal Assistant to the Chief Executive Officer) is an agent that has been developed to organise meetings within a workplace context. The target users are the hi-tech, white-collar sector who typically have the computer continuously available. The aim was to create an agent that convinced users that they were interacting with an agent with social awareness rather than actually creating an agent that had social awareness. It is not the 'personality' that is to be developed, but rather the portrayal of a fictive personality.

PACEO was developed with Microsoft Agent, using one of the standard, freely available animations. It interacts with the users within the context of a diary application similar to other such time management systems, i.e. Outlook. PACEO 'inhabited' a small LAN and roamed its domain searching for agreement with the various meeting participants.

PACEO is portrayed as an individual, autonomous entity to the users (although multiple copies of the agent (one on each machine) were actually used). It does not appear to reside on the user's machine rather it imitates the behaviour of a mobile assistant, appearing only when requested or when sent to organise a meeting by another user by another. Thus, when the agent needs to interact with the user, it graphically appears on their screen and then 'leaves.' The activity and dormancy of the agents was controlled centrally, with user events (such as request for meeting) affecting agent status. This approach was taken with the intention that users should feel that the agent was always somewhere on the system either carrying out its roles or waiting for its next orders.

The agent had access to a set of utterances (provided as text and speech) that were appropriate for specific contexts (for example, project deadlines) and users (Admin staff). These utterances and related behaviours were developed using a framework of plot, role and character (Hayes-Roth, 1998) within the narrative discourse of organisational hierarchy.

3.1 Fabricating the Users

For PACEO each user appears as a fabricated being, a 'social-body,' no different to itself, mapped as a set of social variables that the agent can access. These variables are used as a basis for differentiation between users, allowing PACEO to adapt its utterances and behaviours depending upon whom it is interacting with. Clearly, it is possible to use an wide range of variables, however, the current agent only has access to user name, age, gender and social standing within the organisation.

The key social variable is position within the organisational hierarchy. The agent itself, is at the mid-point of this hierarchy, and is a stereotype of a 30-something, heterosexual, male Personal Assistant to the Chief Executive Officer. The agent gives more respect to those above it in the organisation, yet may be quite curt to more junior staff members. Typically, PACEO is more polite to those older than it, for example, age differentials result in the agent using either the users' first / last name.

3.2 PACEO's Mood: Relating to Users

The variables: relationship to user and last time seen, express the social interaction between user and agent. PACEO has a 'memory' of the users, remembering prior interactions with them in terms of when the interaction occurred and the user's treatment of the agent. This historical relationship is critical to PACEO, as its social interaction with users is based on social variables, previous user relations and the agents' current 'mood.' Any user interaction could potentially alter its mood and subsequent interactions.

The user could interact with the agent neutrally or could adopt one of two moods in conversing with the agent: love and hate. Selection of either of these (presented as buttons in the response dialog box) would influence the general mood of the agent as well as increasing or decreasing its relationship to the user. Interacting with the agent consists of five movements: appearance, introduction, work, farewell and disappearance. Each one of these areas consists of a variety of utterances depending on the mood, the agent's memory of the user and the impact of the social variables.

The social variables that represent the user are filtered through the agent's mood. The mood manipulates the social relationship and social standing variables by either increasing or decreasing them. This has the effect of creating a more agreeable / disagreeable agent both for the subsequent user and for the current user's subsequent use of the agent. Being unpleasant to the agent (selecting the hate mood) or using a range of unpleasant utterances tends to result in the agent being less pleasant with subsequent users and in any further interactions for the offending user.

4 Evaluating PACEO

The qualitative evaluation of PACEO reported here focused on:

- Whether interaction with PACEO results in satisfied users who have had positive interactive experiences.
- Whether social engagement between user and agent occurs, that is, was the user prepared to suspend disbelief and to have empathy or social feelings about PACEO.
- Further, did the agent's apparent anthropomorphic behaviour give rise to the basis of a social relationship.

4.1 Experimental Design

The evaluation was conducted with 4 users, each of whom was allocated a role in a fictitious organisation. The users interacted with PACEO over the course of a two hour session.

A brief overview of PACEO and the diary system it mediates for the user was provided before the session began, focusing on how to interact and the potential impact on PACEO's mood due to that interaction. At all times a researcher was with the users, to give support or help, or to record users' observations. Each of the users spent their time on personal computer-based work tasks (i.e. writing reports, preparing a lecture, etc.) simulating a work environment, in addition to interacting with PACEO to organise meetings.

4.2 Results

Table 1: User Social Variables and Relationship

User	Age	Org. Status	Gender	Relationship
A	50	High	Female	Friendly
B	20	Low	Female	Unfriendly
C	35	High	Male	Friendly
D	30	Medium	Male	Friendly

Table 1 provides the social variables of each of the participants and PACEO's most typical social relationship with each user. In the case of Participant A, the agent was extremely subservient, responding using utterances that highlight the hierarchical difference in positions between itself and the user. The user commented "*He's very crawly isn't he! ... [but] ... quite amusing, though the Genie comes across as if he is taking the mickey a little bit.*" Participant C also dealt with the sycophantic agent, however, his response was "*I don't really like this character, he's a bit slimy isn't he. Even seems a little sleazy...*"

Participant B took a 'hate' approach to PACEO, finding the increasing rudeness of the agent highly entertaining: "*What a cheek... [laughter] ... Never been talked to like this by a computer before!*" Clearly, however, the agent's negative approach had an impact on the user who also stated "*The way it was talking to me, I wonder if it will actually book this meeting...*" The evolution of the interaction between Participant B and the agent showed a steady deterioration of the relationship.

Participant D had a similar social body to PACEO, but the participant found the agent boring "*Well, he's bit dull towards his friends isn't he? Would be good if he knew office gossip...*"

Our second evaluation aim was related to social engagement and the level of social ability that participants ascribed to PACEO. Notably, as the user comments identify, all users anthropomorphised the agent, referring to it as he and assigning characteristics and attributes to the agent, that it didn't possess. Participant D received a visit from PACEO after it had had a negative interaction (with Participant B) and stated, "*To be honest, I don't think he likes me, mind you, he probably doesn't like anyone..*" Participant A gave her view of PACEO: "*You know he's not real, but one would imagine that, given the right character you could actually feel for such things, you kind of know its made up, but you get drawn in*"

5 Discussion

The results we achieved from this limited evaluation of PACEO suggests that the user experience, particularly in terms of satisfaction, can be enhanced using animated agents that appear to exhibit social ability. All of the users found interacting with PACEO enjoyable, even when it was being unpleasant, finding the experience "*fun and a lot more interesting than Outlook.*" One participant commented: "*I could see this catching on in an office, with everyone messing around with this character.*"

All of our participants attributed social ability to PACEO, yet PACEO does not have social ability, it is not an emotional nor empathic entity. Rather, it is a fabrication, living out a series of fictions or actions, constrained by the organisational context. However, the Foucauldian approach adopted in this paper appears to offer considerable opportunities for the development of agents that appear to be socially able. Future work will further consider the appropriateness and potential of this approach.

Current work focuses on adopting an appropriate animation for PACEO and attempting to incorporate a more natural voice, factors that users had identified as being irritating. We are also extending the narrative threads available to PACEO, increasing available utterances and behaviours. Our intention is to evaluate PACEO in an organisational context over a period of several months. Our aim will be to determine whether users will continue to be satisfied by the interactive experience and believe in the agent's social ability over a longer period of time.

References

Andre, E. (1999), Animated Interface Agents (special issue), Applied Artificial Intelligence, 13.

Bates, J. (1994) The Role of Emotion in Believable Agents In Communications of the ACM, Vol. 37(7), 122-125.

Cassell, J., Vilhjalmsson, H. (1999a) Fully Embodied Conversational Avatars: making communicative behaviours autonomous In Autonomous Agents and Multi-Agent Systems, 45-64.

Cassell, J., Thorisson, K.R. (1999b), The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents, Applied Artificial Intelligence, 13, 519-538.

Craib, I. (1993) From the Politics of Class to the Politics of Pleasure, Open University Press, London.

Dehn, D. M., van Mulken, S. (2000), The impact of animated interface agents: a review of empirical research, International Journal of Human-Computer Studies, 52, 1-22.

Foucault, M. (1977) Discipline and Punish: The Birth of the Prison, Penguin, London.

Hayes-Roth, B. (1998), Animate Characters, In Autonomous Agents and Multi-Agent Systems, 195-230.

Hetherington, K. (1998) The Badlands of Modernity: Heterotopia and Social Ordering, Routledge, London.

Lester, J. C., Converse, S.A., Kahler, S.E., Barlow, S.T., Stone, B.A., Bhogal, R.S. (1997) The persona effect: affective impact of animated pedagogical agents In CHI '97, Amsterdam, ACM Press,

Mateas, M. (1997) An Oz-Centric Review of Interactive Drama and Believable Agents, available at http://www-2.cs.cmu.edu/afs/cs.cmu.edu/project/oz/web/papers/CMU-CS-97-156.html

Prendinger, H., Ishizuka, M. (2001), ---, International Conference on Intelligent User Interfaces.

Rabinow, P. (1984) The Foucault Reader, Penguin Books, New York.

Weber, M. (2001) The Protestant Ethic and the Spirit of Capitalism, Routledge, London.

Wooley, B. (1992) Virtual Worlds, Blackwell, Oxford.

WHAT'S IN A ROBOT'S SMILE? THE MANY MEANINGS OF POSITIVE FACIAL EXPRESSIONS

Marianne LaFrance
Yale University; Department of Psychology; New Haven, Connecticut
marianne.lafrance@yale.edu

Abstract

Human smiles are distinctive, noticeable, complex, and socially indispensable. The issue addressed in this paper concerns whether the smiles robots display or perceive can or should possess similar characteristics. Regarding the distinctive appearance of human smiles, Darwin suggested that smiles became associated with happiness because they are so different in appearance from negative expressions (Darwin, 1872/1965). Smiles are easily noticed and even subtle differences in their form are detectable when they surface for very brief periods and from considerable distance (Ekman & Hager, 1979). Human smiles are also very potent such that people have been known to go to considerable lengths to get someone else to produce one and then feel positively euphoric when they have been successful in doing so. In short, human smiling in all its various forms is probably indispensable for successful face-to-face interaction and possibly for human-robot interaction.

1 Introduction

Creators of social robots have taken note of the importance of smiling (Breazeal, 1998; Billard & Dautenhahn, 1997). For example, Breazeal's robot Kismet can show emotive expressions analogous to happiness (along with other emotions like anger, fear, disgust, excitement, sadness, and surprise). Specifically, it is programmed to respond to external visual input with changes in its facial expressions. Such emotive displays apparently demonstrate the capacity to communicate with and learn from human "caregivers." While initial tests of these capabilities show promise, nonetheless, it is also clear that expressive communication between robots and people will only work to the extent to which the displays are believable and acceptable to humans. Braezeal (2000) has phrased it this way, "For a robot, an important function of the motivation system is to regulate behavior selection so that the observable behavior appears coherent, appropriately persistent, and relevant given the internal state of the robot and the external state of the environment."

The goal of this paper is to draw on several lines of research to show that the task of creating credible smiles in robots is a formidable one. Specifically, I describe various dimensions of human smiling that need to be taken into account if robots' smiling is to be taken seriously. To be sure, smiles do reflect and generate positive affect (Ekman & Friesen, 1982) but that is not all they are or do. In fact, there is good reason to believe that human facial expression in general and smiles in particular are a good deal more diverse in form and multifaceted in function than a simple emotion read-out formulation would indicate. In what follows, I describe some of the forms smiles take and some of the range of meanings they are capable of conveying.

2 Smiles: Signs or Symbols?

The meaning of smiling seems obvious—people smile when they are happy or amused (Ekman & Friesen, 1982; Ekman, Davidson, & Friesen, 1990). Nonetheless, it is also true that people smile when the underlying emotion is anything but positive. Results show, for example, that people smile when they are embarrassed (Edelmann, Asendorpf, Contarello & Zammuner, 1989), uncomfortable (Ochanomizu, 1991), miserable (Ekman & Friesen, 1982), and apprehensive (Ickes, Patterson, Rajecki, & Tanford, 1982).

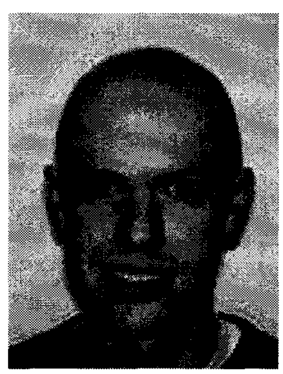


Figure 1. Deliberate smiles are often characterised by asymmetry

Emotionally incongruent smiles such as these sometimes appear indistinguishable from smiles associated with happiness and sometimes they appear on the face as a blend in which a smile is only part of the expression. In the former case, smiles, perhaps more than most other facial displays, often communicate not what one is truly feeling in the moment but rather how one wants to be seen as feeling or how one wants others to feel. Such smiles may look like the real thing because they are constructed to seem natural and spontaneous. However, closer inspection might indicate that the expression is deliberate. For example, if it is asymmetrical, that is, stronger on one side of the face than the other side or if the onset or offset of the smile is too abrupt, the indications point to deliberate display. Because spontaneous smiles and deliberate smiles sometimes look the same should not be construed as meaning the same thing or with having the same effect on self or others. In short, people often adopt smiles so that others will see *them* in a positive light. As such, smiles might more accurately be seen as *representing* rather than *reflecting* pre-existing positive affect, as communicative symbols rather than emotion signs.

3 Smile Types

There is now good reason to believe that human beings do not possess a set of basic emotions that are associated with a set of basic and unambiguous facial displays. Rather, there are families of emotion and families of associated facial expressions (Fogel, et al., 1992). Because there appear to be many types of smiles Ekman and Friesen (1982) concluded that they may be the most misunderstood of all facial expressions precisely because there are many kinds.

3.1 Duchenne and Non Duchenne Smiles

Ekman and Friesen (1982) initially distinguished between two broad smile types, namely the *felt* smile and the *false* smile. In subsequent work, these were renamed the *Duchenne* and *non-Duchenne* smiles respectively. In Duchenne smiles, lip corner contraction is accompanied by contractions of the muscles surrounding both eyes, producing a raising of the cheeks and crows feet wrinkles a the corners of the eyes.

For adults, Duchenne smiling has been linked to the experience of positive emotion. For example, Duchenne smiling is more likely to occur when people are watching pleasant as compared to unpleasant films. And Duchenne smiling is more strongly correlated with people reporting feelings of amusement, happiness, excitement, and interest (Ekman, Friesen, & Ancoli, 1980). Observers also rate Duchenne smiles as more positive compared to other smiles which is why they were initially labeled as felt smiles (Ekman, Davidson & Friesen, 1990).

In contrast, non Duchenne smiles are recognized by a single facial action, namely the contraction of the lip corners with no discernible facial actions around the eyes. The term "non Duchenne smile" replaced the "false" smile designation to allow for the fact that this type of expression is not necessarily deceptive or phoney. It is true that non Duchenne smiles are not reliably associated with the experience of amusement or happiness but it is also true that these smiles serve a number of important social functions. For example, we found that non Duchenne smiles are actually more

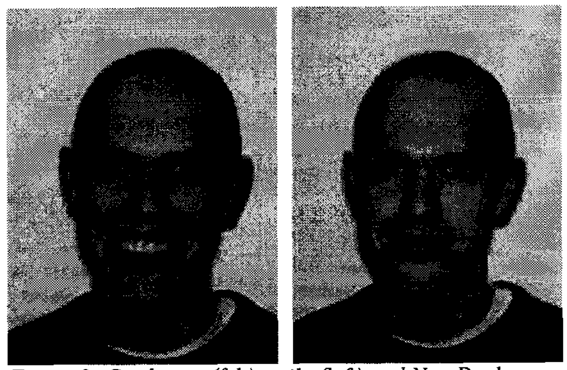


Figure 2. Duchenne (felt) smile (left) and Non Duchenne (fake) smile (right)

effective than Duchenne smiles in generating leniency for a person believed to have transgressed in some way because they convey something about a person's trustworthiness (LaFrance & Hecht, 1995). Nonetheless, the distinction between Duchenne and non Duchenne has proved to be a useful one in as much as these two smile types typically appear in different social contexts and correlate with different patterns of brain electrical activation (Ekman, Davidson, & Friesen, 1990).

Developmental psychologists have also found that even very young infants show both Duchenne and non Duchenne smiles and these different smile types are associated with distinctly different contexts (Dickson, Fogel & Messinger (1997). For example, Duchenne smiles by 10-month old infants have been found to be associated with the approach of their mothers whereas non Duchenne or basic smiles are association with the approach of strangers (Fox & Davidson, 1988). Fogel and his colleagues contend that non Duchenne or basic smiles in infants represent an "enjoyment of readiness to engage in play" rather than activated enjoyment in play itself (Fogel, Nelson-Goens, & Hsu, 2000, p. 513).

Studies of infants have also discerned two additional smile types, namely the *play smile* and the *duplay smile* (Dickson, Walker & Fogel, 1997). In play smiles, the baby's jaw drops and the lip corners rise. A duplay smile includes a lip corner raise, a jaw drop, and a cheek raise. For infants and toddlers, the play smile is more common while playing with an attentive partner, whereas the non Duchenne or basic smile occurs more frequently during toy-centered play. Duplay smiles in particular are more likely to occur during physical play with fathers and during book reading with mothers. During physical play another element is added to the mix: the mouth opening is often accompanied by a discernible inhalation (Dickson et al, 1997). Both play and duplay smiles are often accompanied by vocalizations such as laughs, squeals and yells.

In addition each smile type in 12-month-old infants co-occurs with still other behaviors. For example, duplay smiling is more likely to take place while the infant is looking at the mother and when the mother is smiling as might be the case when the mother and child are playing a game of peekaboo or tickle. That configuration appears to indicate enjoyment in the build-up to a readiness to share the enjoyment with the other who has co-created the emotional process. In sum, different patterns of co-occurring actions with different smile types appear to be associated with uniquely different emotions.

The meaning of different smiles also changes as children get older. When pre-schoolers are engaged in an achievement game, smiling appears to be a reliable readout of the child's joy or pride in accomplishment. But when the task no longer occupies the child's attention and when a familiar person is present, then the smiling appears to serve a more social communicative function, such as belittling failure at the task or seeking a comment. In other words, the connection between expression and emotion, between smiling and happiness becomes less direct with age.

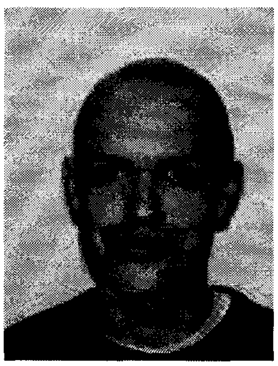


Figure3. The "dimpler" is an expression sometimes mistaken for a basic smil.e

By pre-school, children are able and willing to control their facial behaviors, be it intentionally as might be the case when they are being coy or unintentionally (Schneider, 1997). In a non-social situation, the smile of infants and pre-schoolers seems to indicate a positive emotional state associated with accomplishment or mastery. However, when someone else is present, then the smile might be used more as an invitation to mutual pleasure or a request for input.

The same distinction between emotion-based smiling and social-based smiling has also been observed for adults. For example, Fridlund and his colleagues have shown that people smile more in the company of someone else or even in the imagined company of someone else than when alone even when the experienced emotion is comparably positive (Fridlund, Sabini, Hedlund, Schaut, Shenker & Knauer, 1990). Kraut and Johnston (1979) found too that people smile more while interacting with others than when alone and that bowlers who had just succeeded in getting a strike smile more when they turned to face their friends than immediately following the success. In short, there are indications that a good deal of smiling is done in order to effect and maintain social connection rather than as an automatic manifestation of positive feeling.

3.2 Smile Blends

A retraction of the lip corners indicating a smile can co-occur with any number of other facial features that substantially change the expression and its import.

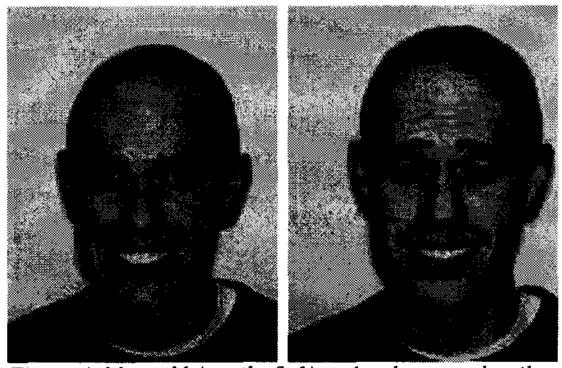


Figure 4. Miserable) smile (left) and embarrassed smile (right)

3.2.1 Miserable smiles

A case in point is the miserable smile.Sometimes, effort is put into masking a bad feeling but putting on a smile; other times, the face shows blended smiles. The lower face may show retraction of the outer corners of the mouth intimating a smile. At the same time the eyes or brows signal distress. According to Ekman (1985), the *miserable* smile simultaneously acknowledges experiencing negative emotions even while providing a comment on the state of being miserable. The message is: I am unhappy but I am going to grin and bear it. The miserable smile is a sophisticated one for it entails layering an emotion over an emotion.

3.2.2 Embarrassed Smiles

In an embarrassment smile, there is the familiar retraction of the lip corners combined with a downward or sideward gaze.

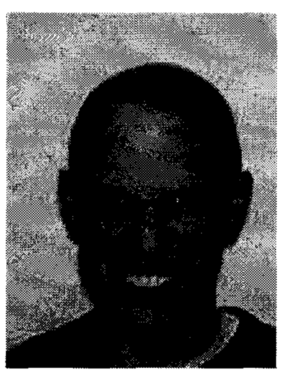


Figure5. An angry smile

3.2.3 Angry Smiles

There is also such a thing as an angry smile which combines lip retraction with lowered brows as might occur when the prevailing sentiment is "sweet revenge".

4 Smile Rules

Smiling far from being involuntary is nearly obligatory in many social contexts even if or perhaps especially if people are not feeling particularly happy. For example, anthropologists and sociologists have documented numerous instances where people smile because their role or interpersonal situation requires them to do so. For example, greetings are typically associated with smiling (Eibl-Eibesfeldt, 1989). Within the United States, Hochschild (1983) noted that many workers are required to smile as part of their jobs. For instance, airline flight attendants must smile and smile well. Thus a flight attendant is trained to "really work on her smiles" and is expected to "manage her heart" in such a way as to create a smile that will both seem as "spontaneous and sincere" (Hochschild, 1983, p. 105).

There is also evidence many social encounters have clearly understood rules about smiling. We examined this idea by asking males and females to imagine smiling or not in a number of different social contexts. As one example, they were instructed to imagine that they smiled or not in response to someone describing an achievement. We predicted that both sexes and particularly women would anticipate more negative repercussions when they did not smile (LaFrance, 1997). And that is what we found. People felt significantly less comfortable and less appropriate when they did not smile than when they did. Women in particular believed that the other's impression of them would change more if they did not smile. Other studies also find that people expect fewer rewards and higher costs when their responses to others are insufficiently positive (Stoppard & Gunn-Gruchy, 1993).

In still other social contexts, the absence of smiling would be duly noted because of implicit rules to smile unless otherwise indicated. For example, smiles are often linked with attempts by politicians and would-be suitors to be persuasive (Burgener, Jirovec, Murrell & Barton, 1992; Masters, Sullivan, Lanzetta, McHugo, & Englis, 1985).

5 Smiles as Cultural Displays

Over thirty years ago, Ekman and Friesen (1969) suggested that people do not always show what they feel. What gets expressed facially is guided by display rules that dictate to whom and in what contexts it is appropriate or inappropriate to show various expressions. In particular, cultural groups vary substantially in the degree to which they direct members as to how they should express themselves in public. Indeed there is a growing literature indicating that expressive displays vary substantially across national, ethnic, regional and social class groups. All cultural groups smile but how and where and with whom are highly variable. Ignoring group variation would be ill-advised for human and robot alike.

6 Conclusions

One of the most interesting things about smiling is the number of forms it can take, the range of contexts in which it can be found and the social uses to which it can be put. Smiles can be spontaneous and unrehearsed but human smiles are also monitored and modulated. Small smile variations can lead to substantially different outcomes. Smiles are used to influence what will happen but the form that they take, the duration they last, the speed with which they come on or leave the face will impact how they are to be interpreted. Sometimes smiles denote happiness; more often, smiling is done to maintain social relationships and to enact social rituals. In sum, if smiles on robots are to be taken as coherent and credible, researchers will need to incorporate such knowledge in addition to what is known about expressive displays as emotion signals.

References

Billard, A. & Dautenhahn, K. (1997). Grounding communication in situated, social robots (Report No. UMCS--97--9--1). Manchester, England: University of Manchester.

Breazeal, C. (1998). A motivational system for regulating human-robot interaction. Proceedings of the Fifteenth National Conference on Artificial Intelligence (AAAI-98). Menlo Park, CA: AAAI Press, 54--61.

Breazeal, C. (2000) Infant-like Social Interactions between a Robot and a Human Caregiver. Adaptive Behavior, 8, 9-75.

Bugener, S.C., Jirovec, M., Murrell, L., & Barton, D. (1992). Caregiver and environmental variables related to difficult behaviors in institutionalized demented elderly persons. Journal of Gerontology, 47, 242-249.

Darwin, C. (1965). The expression of the emotions in man and animals. New York: Appleton. (Original work published 1872).

Dickson, K.L., Fogel, A., & Messinger, d. (1997). The development of emotion from a social process view. In M. Mascolo & S. Griffen (Eds.), What develops in emotional development, (pp. 253-273). New York: Plenum Press.

Edelmann, R. J., Asendorpf, J., Contarello, A., & Zammuner, V. (1989). Self-reported expression of embarrassment in five European cultures. Journal of Cross Cultural Psychology, 20, 357-371.

Eibl-Eibesfeldt, I. (1989). Human Ethology. New York: Aldine de Gruyter.

Ekman, P. (1985). Telling lies. New York: Norton.

Ekman, P., Davidson, R. J., & Friesen, W. (1990). The Duchenne smile: Emotional expression and brain physiology II. Journal of Personality and Social Psychology, 58, 342-353.

Ekman,P., & Friesen, (1969). The repertoire of nonverbal behavior: Categories, origins, usage, and coding. Semiotica, 1, 49-98

Ekman, P., & Friesen, W. V. (1982). Felt, false, and miserable smiles. Journal of Nonverbal Behavior, 6, 238-252.

Ekman, P., & Hager, J.C. (1979). Long distance transmission of facial affect signals. Ethology and Sociobiology, 1, 77-82.

Elman, D., Schulte, D.C., & Bukoff, A. (1977). Effects of facial expression and stare duration on walking speed: Two field experiments. Environmental Psychology and Nonverbal Behavior, 2, 93-99.

Fogel, A., Nelson-Goens, G.C., & Hsu, H-C. (2000). Do different infant smiles reflect different positive emotions? Social Development, 9, 497-520.

Fogel, A., Nwokah, E., Dedo, J., Messinger, D., dicson, K.L., Matusove, E., & Holt, S.A. (1992). Social process theory of emotion: A dynamic systems approach. Social Development, 1, 122-142.

Fox, N.A., & Davidson, R.J. (1988). Patterns of brain electrical activity during the expression of discrete emotions in 10-month-old infants. Developmental Psychology, 24, 230-236.

Fridlund, A. J. (1991). Sociality of solitary smiling: Potentiation by an implicit audience. Journal of Personality and Social Psychology, 60, 229-240.

Fridlund, A.J., Sabini, J.P., Hedlund, L.E., Schaut, J.A., Shenker, J.I., & Knauer, M.J. (1990). Audience effects on solitary faces during imagery: Displaying to the people in your head. Journal of Nonverbal Behavior, 142, 113-137.

Hecht, M. A. & LaFrance, M. (1998). License or obligation to smile: Power, sex and smiling. Personality and Social Psychology Bulletin, 24, 1326 - 1336.

Hochschild, A. (1983). The managed heart: Commercialization of human feeling. Berkeley, CA: University of D.W., & Tanford, S. (1982). Behavioral and cognitive consequences of reciprocal versus compensatory responses to preinteraction expectancies. Social Cognition, 1, 160-190.

Kraut, R.E., & johnston, R.E. (1979). Social and emotional messages of smiling: An ethological approach. Journal of Personality and Social Psychology, 37, 1539-1553.

LaFrance, M. (1997). Pressure to be pleasant: Effects of sex and power on reactions to not smiling. Revue Internationale de Psychology Sociale/International Review of Social Psychology, 2, 95-108.

LaFrance, M., & Hecht, M. (1995). Why smiles generate leniency. Personality and Social Psychology Bulletin, **21**, 207-214.

LaFrance, M., & Hecht, M. A. (1999). Obliged to smile: The effect of power and gender on facial expression. In P. Philippot, R. S. Feldman, & E. J. Coats (Eds.), The social context of nonverbal behavior (pp. 45-70). Cambridge: Cambridge University Press.

Masters, R.D; Sullivan, D G; Lanzetta, J T; McHugo, G. J., & Englis, R. (1986). The facial displays of leaders: Toward an ethology of human politics. [Journal Article] Journal of Social & Biological Structures. 19, *319-343*.

Ochanomizu, U. (1991). Representation forming in Kusyo behavior. Japanese Journal of Developmental Psychology, 2, 25-31.

Schneider, K. (1997). Development of emotions and their expression in task-oriented situations in infants and preschool children. In M. Mascolo & S. Griffen (Eds.), What develops in emotional development, (pp. 109-130). New York: Plenum Press.

Stoppard, J.M., & Gunn-Gruchy, C.D. (1993). Gender, context and expression of positive emotion. Personality and Social Psychology Bulletin, 19, 143-150.

Wierzbicka, A. (1994). Emotion, language and cultural scripts. In S. Kitayama and H.R. Markus (Eds.), Emotion and culture: Empirical studies of mutual influence. Washington, DC: American Psychological Association.

An Expressive System for Endowing Robots or Animated Characters with Affective Facial Displays

Craig Latta, Nancy Alvarado, Sam S. Adams, Steve Burbeck

The IBM Thomas J. Watson Research Center; P.O. Box 218, Yorktown Heights, N.Y. 10598
crl@watson.ibm.com, nanalv@us.ibm.com, ssadams@us.ibm.com, sburbeck@us.ibm.com

Abstract

This paper describes an expressive system designed to link a model of emotion with socially appropriate expressive behaviors in an intelligent agent or robot. The system was developed to provide an affective interface to Joshua Blue, a computer simulation of an embodied mind that includes emotion and is designed to learn and interact effectively in social environments. The advantages of this implementation include the ability to accommodate different emotion models, to easily specify different expressive behaviors, and to directly inspect, continuously monitor, and intervene to change underlying affective states as they occur. The system currently uses a real-time 3D facial avatar with a simplified anatomical model, but may be readily modified to direct the expressive behaviors of more complex animated characters or robots.

1 Introduction

Expressive behavior in intelligent agents or robots serves two important purposes. First, it permits an entity to interact effectively with other expressive beings in social contexts. In many applications, this may be the primary function of the expressive behavior. Naturalness and realism, social effectiveness, and meaningfulness of displays to human observers are important concerns. Second, expressive behavior provides information to the designer about the internal states of the software. This latter function is essential to proving assertions that an emotion model has influenced the behavior of an agent or robot. In more complex systems with greater behavioral autonomy, the better the information conveyed by expressive behavior, the better a designer can understand the impact of affect on a machine's cognition, which is less directly observable.

Additionally, we felt that the utility of an expressive system would be greatly enhanced if it were able to: (1) support different models of emotion; (2) be configured to control a wide variety of behaviors with minimal human effort; and (3) allow the designer to easily inspect, monitor, or change the internal affective states of the underlying emotion system, through the same expressive interface. We designed this expressive system for use with Joshua Blue, a computer simulation of an embodied mind that must ultimately function in a wide variety of environments, virtual and physical (see Alvarado, Adams, Burbeck & Latta, 2001). These requirements made versatility an important design goal. However, we believe our approach may also enhance simpler, less general systems by making it easier to incorporate future improvements in animation techniques or robot technology that will make expressive systems more socially effective.

2 System Overview

The expressive system consists of the following components: (a) an emotion model; (b) an expressive behavior model; (c) a 3D facial avatar; (d) an emotional state display; (e) tools for editing emotional expressions. Each of these parts is described below, for the current implementation. The specifics of the avatar, emotion model and expressive behavior model can be changed with minimal impact on the remaining parts of the system, which make internal system values open to inspection and propagate changes between subsystems, all with fluid interactivity. The system is implemented in Squeak, a dynamic, open-source object system.

2.1 Emotion Model

The emotion model in the Joshua Blue system incorporates the two dimensions of affective experience, valence (appraisal as good or bad) and arousal (level of excitement), originally described by Osgood (1966) and later presented as an emotion model by Russell (1991; 1997). Since these dimensions are essentially abstractions derived from the dominant aspects of subjective experience, they are hypothesized to be present in any emotional phenomenon. This means that a variety of emotion models, from Ekman's (1992) basic emotions to simple pain-versus-pleasure models, can be readily mapped into the same two-dimensional space.

If reduction to two dimensions is undesirable, such as when finer distinctions are needed, multidimensional maps can be created in the same manner as was used to generate this simple two-dimensional system. While this complicates visualization, the system is not limited to using 2-D pointers but can support a 3-D controller. An alternative

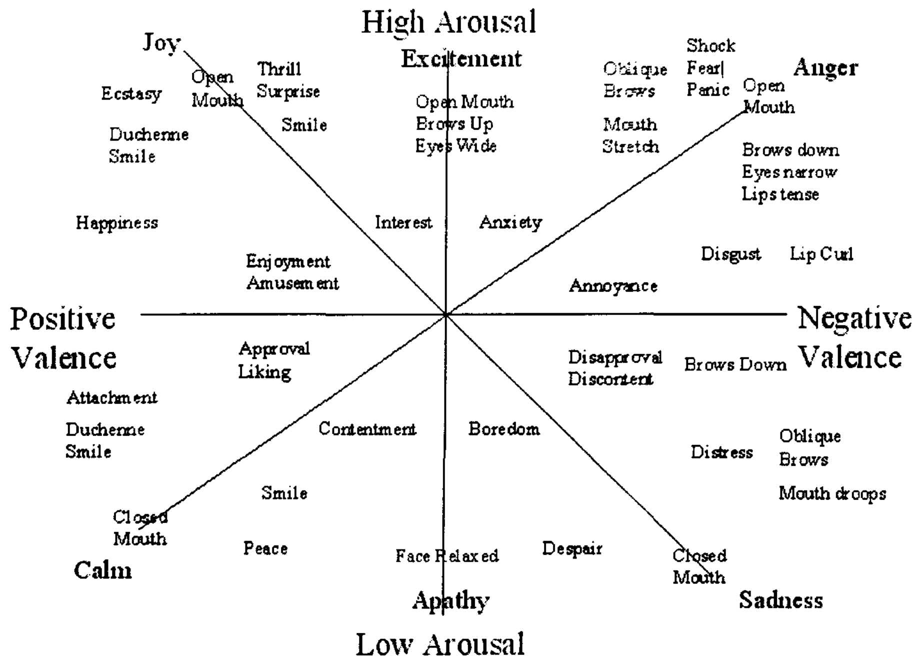


Figure 1. A Mapping of Discrete Emotions and Facial Expressions onto the Two Dimensions of Affective Experience

approach includes specifying higher dimensionality in a series of 2-D maps, each including different specified pairs of dimensions.

In this implementation, regions of the affective space are partitioned by ten equal increments along each axis. Discrete emotions are mapped to the space, as shown in Figure 1. The terms shown in Figure 1 are descriptive and suggest what occurs facially in each region. The actual specification of facial movement is defined by the expressive behavior model, described below.

2.2 Expressive Behavior Model

Facial expressions for animating the avatar were specified using the definitions of action units corresponding to basic emotional states (Ekman and Friesen, 1978). Action units for additional emotional states were identified based on empirical studies by Snodgrass (1992) and Alvarado (1996). We determined contraction values for muscle groups in each expression using action units defined by the Facial Action Coding System (Ekman & Friesen, 1978). Intensities of muscle movement (expressed as slider position) were specified for each step in the two-dimensional affective space of the emotion model. A separate map was defined for each muscle group. This specified a facial expression for each state possible in the Joshua Blue's emotion model. Basic expressions concurrent with extreme emotional arousal have been tested for reliable interpretation by Ekman and Friesen. To the extent that our versions conform to the specified basic expressions, we expect to obtain similar results in judgment studies. The less intense expressions and those expressions corresponding to states not defined as basic emotions are being tested for interpretability by human observers. Following that validation, all expressions will be tested for naturalness of transition and believability.

2.3 Facial Avatar

The current expressive display employs an animated facial avatar, rendered from a three-dimensional mesh of external facial skin points deformed under the influence of underlying musculature, as described by Waters (1987) (see Figure 2). We adapted a Smalltalk implementation of Waters' avatar (Notarfrancesco, 2001) for use as an expressive display driven by the affective states of the Joshua Blue system. A user may manually modify the tensions of 12 major facial muscle groups, by opening and manipulating sliders. While not anatomically precise, this simplified muscle model permits simulation of reasonably believable affective facial expressions. It is the first of three main elements in the expressive system's user interface, the other two being the emotional state display and modification tools.

2.4 Emotional State Display

The user may view and manipulate the procession of Joshua Blue's emotional state through the affect space over time with a specialized monitor. This monitor samples Joshua Blue's current emotional state at a user definable rate, and displays it as a point in the space. Previously displayed samples persist on the display, but change color

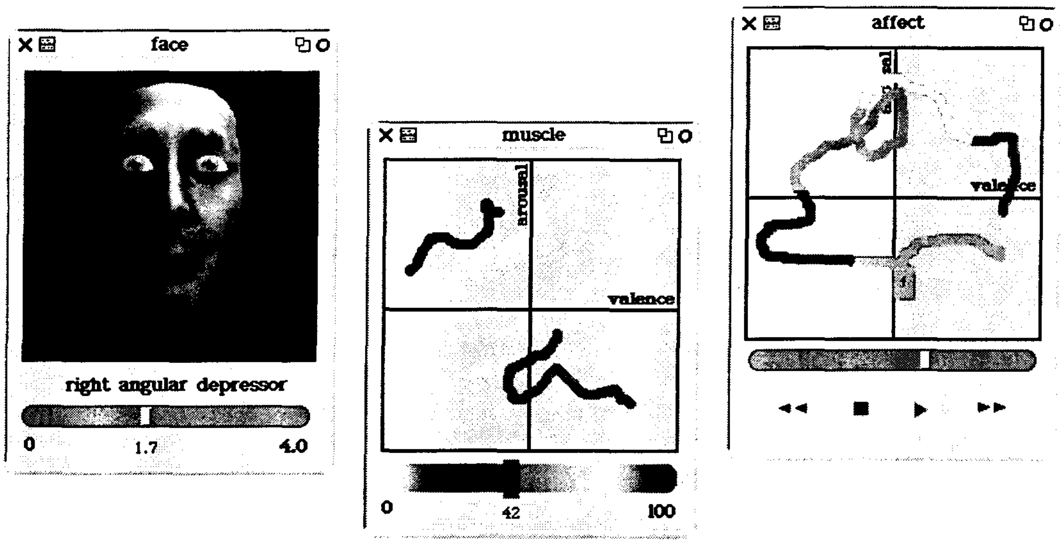


Figure 2. Facial Avatar, Affective Muscle Response Editor, and Emotional State Display

over time as they age. Over time, the monitor accumulates a "trace" of the changes in Joshua Blue's emotional state.

The monitor provides transport controls, so that the user may see the expression corresponding to a previous emotional state, and control the sampling of subsequent states. When the user clicks on a point in the space, the monitor highlights that point in outline, and the avatar takes on the expression corresponding to the associated emotional state. A slider below the affect space display provides random access to particular points in the trace's history. Finally, buttons at the bottom of the monitor provide traditional playback control, toggle the recording of new samples, and enable storage and loading of traces.

2.5 Editing Tools

In addition to rendering Joshua Blue's current emotional state as a facial expression, the avatar provides access to tools for editing muscular responses. As the user moves the pointer over the avatar, the system indicates the muscle groups muscles in the vicinity, and displays their names. The user may then select a particular muscle group, which displays its current contraction value along with a slider showing the range of possible contraction values.

Having selected a muscle group, the user may also open an editor on the mapping between affect and muscle contraction. This editor shows a particular muscle group's contraction response for each point in the affect space, indicated by color. The user may rapidly specify the response for many points by selecting a contraction value from a palette and painting its associated color in the affect space. Direct manipulation of the muscle in the avatar also selects contraction values in the muscle editor; the user is free to interpret the palette selection as a number or as a visible expressive cue. The user may open any number of muscle editors concurrently, supporting quick composition of complex facial expressions over the affect space.

3 Squeak Implementation Benefits

The Squeak Smalltalk system (see Ingalls, Kaehler, Maloney, Wallace & Kay, 1997) affords us a great deal of flexibility on several fronts. Squeak's message-oriented organization allows us to establish a strong separation of concerns, such that interdependencies between major functional parts are few. As a result, it is straightforward for us to change the emotion model, expressive behavior, and human interface subsystems independently. Since Squeak is a "late-binding" system, in which objects' message responses are determined at runtime, we may make such changes during operation. This allows us to incorporate new insights while continuing to monitor expressions and record emotional states.

The Squeak architecture is completely open. This gives us the ability to make fundamental operational changes, such as might be required when profiling performance. It also permits complete understanding of the system, both by us and by other researchers, and promotes a broad consistency throughout the system. Largely due to its openness, Squeak has an active worldwide development community. Several contributions from this community have saved us a great deal of development time, and we have been able to contribute significant works in return.

We have noticed most of these features in other systems, but only in Squeak have we found them in the same system.

4 Problems Resolved

Our attempts to produce believable facial expressions using this method led to two concerns. First, how is a transition from one affective state to another accomplished in a natural-looking manner? Second, how does the emotion model's intensity relate to the intensity of muscle movements in the facial avatar? Answers to these questions do not exist in the facial expression literature and thus must be addressed empirically. This model provides a means for doing so. An advantage of using an animated facial avatar to address such questions, rather than photographed human facial activity, is that transitions between expressions and levels of intensity can more readily be manipulated in stimuli presented to observers for evaluation. The perceived naturalness of manipulated expressions can be used to predict what is likely to be typical in human facial behavior, as well as to dictate choices about how to implement natural-looking expression in artificial faces.

Based upon previous research (reviewed by Alvarado & Jameson, 2001), it seems likely that the implementation of facial activity will depend upon whether the goal is to simulate naturalistic human expression or to produce meaningful displays for human observers. It is clear to us from many studies (c.f., Ekman & Rosenberg, 1997) that there is far from a straightforward connection between what is produced on the face in various contexts and the meaning that is interpreted by observers from such facial displays. To date, only Ekman and Friesen's basic expressions have produced consensual interpretations among human observers, while the array of behavior produced on the face is much wider than that subset of basic expressions. Wierzbicka (1999) and others have called for a component-based semantics of the face, but as yet there is no empirical support for component-based meaning or any context-independent meaning beyond Ekman and Friesen's basic facial expressions. Further, whatever might be expressed by humans in naturalistic settings is mediated by a variety of factors not typically included in expressive models, including display rules, instrumental intentions, personality, and coping style (Alvarado & Harris, 2001).

Because there is no straightforward relationship between human internal states, expressive behavior and communicative intentions, optimizing expressive behavior to attain naturalism does not necessarily maximize communication effectiveness (see Russell & Fernandez-Dols, 1997; Alvarado & Harris, 2001). One solution to this problem is to include a model of social intention that incorporates expressive display rules and instrumental goals. Otherwise, designers may be required to select one or the other goal, depending on the intended use of their agent or robot.

5 Test Results

We plan to test whether the system functions as intended, and whether the avatar displays recognizable versions of Ekman and Friesen's basic expressions. We will then use the system to explore the issues raised concerning, transitions, tradeoffs between naturalism and interpretability, and the relationship between intensity of movement and intensity of affect, as perceived by observers. This will likely constitute an ongoing research program, made possible by the flexibility of this expressive system.

6 Related Work

Several others have made implementations of expressive behavior, directed by various models of emotion. Further, the use of topographic mapping to link an emotion or motivational space with specified behaviors is also not new. For example, Breazeal's (2001) robot Kismet is directed by a topographic emotion space based upon Ekman's basic emotions, with defined behaviors corresponding to each region of a motivational space. A thorough review of the many projects with similarity to this one is beyond the scope of this paper. The main contribution of the current work is the flexible association of these components into a system where internal states can be inspected and compared with facial activity and where emotion models and the expressive output of an avatar or robot can be readily changed without redesign or extensive programming.

7 Conclusion

While the relationship between cognition and expression is not yet well understood, we believe that many useful elements for its pursuit are in hand. We have created an environment for fluid, interactive evaluation of these elements, which we hope will yield new insights into their composition and development.

References

Alvarado, N. (1996). Congruence of Meaning Between Facial Expressions of Emotion and Selected Emotion Terms. *Motivation and Emotion, 20*, 33-61.

Alvarado, N., Adams, S. S., Burbeck, S. & Latta, C. (2001). Integrating Emotion and Motivation into Intelligent Systems. Manuscript in submission.

Alvarado, N. & Harris, C. (2001). Facial Expression and Coping Style. Manuscript in submission.

Alvarado, N. & Jameson, K. (2001). Varieties of Anger: The Relation Between Emotion Terms and

Components of Anger Expressions. Manuscript in submission.

Breazeal, C. 2001. *Designing Sociable Machines*. The MIT Press.

Ekman, P. (1992). An Argument for Basic Emotions. *Cognition & Emotion*, 6, 169-200.

Ekman, P. & Friesen, W. (1978). *Facial Action Coding System* (FACS). Consulting Psychologists Press.

Ingalls, D., Kaehler, T. Maloney, J., Wallace, S., Kay, A. (1997). Back to the Future: The Story of Squeak, a Practical Smalltalk Written in Itself. *Proceedings of the 1997 ACM Conference on Object-Oriented Programming Systems, Languages, and Applications*, 318-326.

Notarfrancesco, L. (2001). An adaptation of the Waters facial animation system. Unpublished source code.

Osgood, C.E. (1966). Dimensionality of the Semantic Space for Communication Via Facial Expression. *Scandinavian Journal of Psychology, 7*, 1-30.

Russell, J. (1991). The contempt expression and the relativity thesis. *Motivation and Emotion, 15*, 149-168.

Russell, J. (1997). Reading Emotions From and Into Faces: Resurrecting a Dimensional-Contextual Perspective. In J. Russell & J.M. Fernandez-Dols (Eds.), *The Psychology of Facial Expression* (pgs 295-320). Cambridge University Press.

Snodgrass, J. (1992). Judgment of Feeling States From Facial Behavior. Reported by Russell (1997). Unpublished doctoral dissertation, University of British Columbia.

Waters, K. (1987). A muscle model for animating three-dimensional facial expressions. *IEEE Computer Graphics (SIGGRAPH'87), 21(4).*

Wierzbicka, A. (1999). *Emotions Across Languages and Cultures: Diversity and Universals.* Cambridge University Press.

Men Behaving Appropriately: Integrating the Role Passing Technique into the ALOHA System

Brian MacNamee[1]; Simon Dobbyn[2]; Padraig Cunningham[1]; Carol O'Sullivan[2]
Artificial Intelligence Group[1], Image Synthesis Group[2], Trinity College Dublin, Dublin 2, Republic of Ireland.

Email: Simon.Dobbyn@cs.tcd.ie, Brian.MacNamee@cs.tcd.ie, Padraig.Cunningham@cs.tcd.ie, Carol.OSullivan@cs.tcd.ie

Abstract

The Adaptive Level of Detail for Human Animation (ALOHA) system is a platform for animating virtual humans within a virtual environment using levels-of-detail for geometry, motion and conversational behaviour. Until now the behaviour of these humans has been determined using pre-defined scripts. This paper describes the integration of the intelligent agent based role-passing technique into the ALOHA system to allow for the creation of dynamic scenes.

1. Introduction

In the realm of Virtual Environment (VE) research, it is common for research projects to focus on one aspect of a problem, overlooking other important issues as they could distract from the core research being undertaken. Collaboration between research groups focusing on different aspects of the overall VE problem is one way of allowing groups to focus on their own particular research goals, and still pursue the goal of creating fully immersive VEs. In this paper we will discuss how two research projects, one based on level-of-detail (LOD) animation techniques and another based on Artificial Intelligence (AI) for computer games, are being combined to create dynamic virtual environments inhabited by intelligent virtual humans.

The first research effort, being undertaken by the Trinity College Image Synthesis Group (ISG), is the Adaptive Level of Detail of Human Animation (ALOHA) system, the aim of which is to animate and render virtual-humans in real-time (Giang, 2001). The system takes advantage of the limitations of the human visual system to use an LOD approach to compute less accurate models when loss of accuracy is unlikely to be noticed, and more accurate models when a model is likely to be the focus of a viewer's attention.

The second research initiative, part of the TCD GameAI project (an effort by the Machine Learning Group (MLG) to apply sophisticated AI techniques to the realm of computer games (Fairclough, 2001)), uses intelligent-agent technologies to add depth to computer controlled Non Player Characters (NPCs) (MacNamee, 2001). One of the techniques investigated in this research is *role-passing* which allows an agent in a particular situation to assume a role, which then drives the agent's behaviour for that situation. This paper will describe the integration of the role-passing technique into the ALOHA system.

The synergy between these two projects brings significant benefits to both. Firstly, by using characters implemented using the role-passing technique, dynamic scenes can be created within the ALOHA system, in which the behaviour of virtual humans is driven by their roles, rather than being scripted. Secondly, for the game AI project, the ALOHA system offers a graphically sophisticated test-bed for new agent technologies developed.

This paper will begin with an overview of the ALOHA system and its LOD approach. Following this, the intelligent-agent based work being carried out will be discussed, and the role passing technique will be introduced. A simulation example has been designed in order to test the system, and this will be described next. Finally, future and on-going work will be discussed.

2. ALOHA System Overview

The ALOHA system's aim is to animate and render virtual-humans in real-time based on an LOD approach. For example, if a user views a crowd of virtual humans from a distance, there is no need to have computationally expensive models and sophisticated animations of individual virtual humans, as the user will not perceive the difference. However, if the user zooms up closer, the realism of the virtual human's model and its motion should be improved.

The ALOHA system uses a number of LOD techniques in order to achieve this. These important techniques used are as follows:

Geometric LOD : The different geometric levels of detail for the virtual human's skin layer are implemented using subdivision surfaces. Subdivision surfaces were used in the production of *Geri's Game* (DeRose, 1998) and *A Bug's Life.* They are a mix between polygon meshes and patch surfaces, and provide some of the best attributes of each. Subdivision surfaces are constructed through recursive splitting and averaging. Because of this recursive nature, subdivision naturally accommodates LOD control through adaptive subdivision.

Animation LOD: In order for the ALOHA system to be scalable, the LOD resolver has the ability to request different animation levels of detail. This allows the LOD resolver to decide how to resolve joint angles with inverse kinematics, to decide how many frames a movement should receive, and whether to use a simple kinematic interpolation technique or a dynamic technique to simulate a motion. This results in smooth realistic animations being applied to virtual humans rated with high importance, while lower level animation techniques are applied to virtual humans in the background, taking minimal perception degradation into account.

Gesture and Conversational LOD: The Gesture and Narrative Language (GNL) group from MIT Media Lab has developed the Behaviour Expression Animation Toolkit (BEAT) which allows an animator to input typed text to be spoken by an animated human figure, and to obtain as output appropriate and synchronized non-verbal behaviours and synthesised speech in a form that can be sent to an animation system (Cassell, 2001). This toolkit automatically suggests appropriate gestures, communicative facial expressions, pauses, and intonational contours for an input text and also provides the synchronisation information required to animate behaviours in conjunction with a character's speech. Another collaborative project aims to incorporate a simpler version of this toolkit in the ALOHA system. This would allow the LOD resolver to provide realistic social interaction between characters that are rated with high importance, while a lower level of detail would be applied to characters socialising in the background.

Before beginning this project, scenes in the ALOHA system were strictly based on pre-defined scripts. In order to add the ability to create dynamic scenes populated with intelligent agents the role passing technique is being added to the system. This will provide a means to automatically drive the behaviour of virtual humans within the ALOHA system, thus creating dynamic scenes.

3. Proactive Persistent Agents using Level of Detail AI

In adventure and role-playing games there is a trend for computer controlled NPCs to be very simplistic in their behaviour. Usually, no modelling of NPCs is performed until the player reaches the location in which an NPC is based. When the player arrives at this location, NPCs typically wait to be involved in some interaction, or play through a pre-defined script. This leads to very predictable, and often jarring behaviour. For example, a player might enter a room and meet an NPC who would perform a set of actions based on some script. However, if the player were to leave that room and re-enter, the NPC would play through the same script again. In order to overcome these limitations, new models are required for implementing NPCs.

Although such models have not been widely used in computer games, a number of architectures for creating realistic characters have been developed in other settings. For example work led by Thalmann (Caicedo, 2000) and the Oz project based on interactive drama (Mateas, 1997). As part of its work the TCD Game AI project, after examining various architectures, has developed the Proactive Persistent Agent (PPA) architecture (MacNamee, 2001) which seeks specifically to overcome the limitations typically associated with agents in computer games.

Agents based on this architecture are proactive in the sense that they can take the initiative and follow their own goals, irrespective of the actions of the player. Although according to (Wooldridge, 1995) this is one of the key attributes distinguishing intelligent-agents from other software paradigms it has been largely overlooked in agent architectures used for computer games. Persistence refers to the fact that at all times, all NPCs in a virtual world are modelled *at least to some extent*, regardless of their location relative to that of the player.

As part of the PPA research effort the technique of level of detail AI (LODAI) is also being investigated. LODAI arises from similar work in the field of computer graphics and simulation (Carlson, 1997; Dingliana, 2000) where it has been used to great effect. LODAI involves controlling characters to higher or lower levels of sophistication based on their position and importance within a virtual world. Complementing the LODAI technique, role-passing has also emerged from this research and will be described next.

4. Role Passing

When intelligent agents are used in virtual environments it is often required that they behave believably in a range of different situations. For example, it might be required that, within the same simulation, an agent is found at work in an office, and then later on enjoying a drink in a bar. The kind of behaviour required of the agent, and the motivations that should drive this behaviour, are quite different in each of these situations. Inspired by (Horswill, 1999) (although what we refer to as role-passing operates at a coarser level of granularity than that discussed in Horswill's work) the technique of role-passing allows intelligent agents to take on different *roles* depending on their current situation.

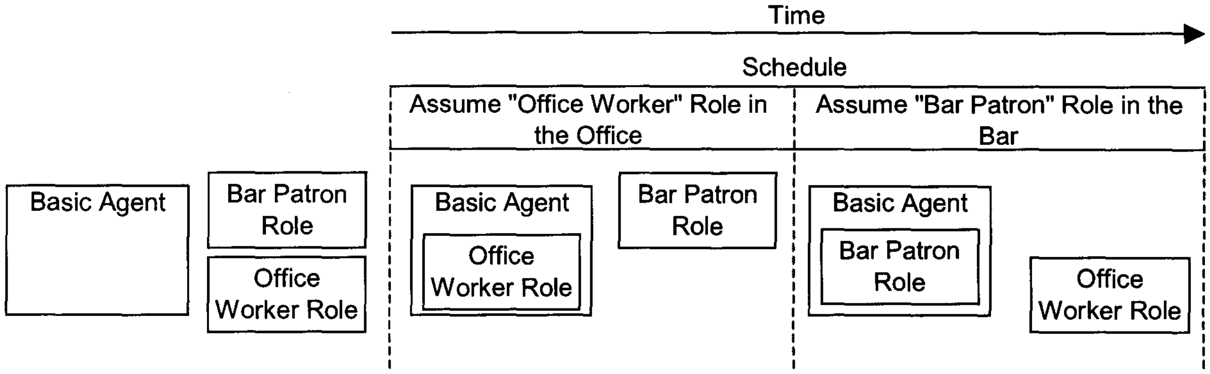


Figure 1 An illustration showing the process through which an agent assumes different roles based on a schedule.

Role-passing operates by using a schedule to layer appropriate roles on top of a very basic agent, at appropriate times within a simulation. This basic agent is capable of simple behaviours such as moving through a virtual world, using objects and interacting with other agents. The agent also has a number of very general attributes describing personality traits.

When a particular role is layered upon this basic agent, it instructs the agent on how to behave in a certain situation. The first key component of a role is a set of motivations that drives the agent. Activation levels for these motivations are extrapolated from the basic agent's personality traits. Activation of a motivation results in the agent performing a particular task, such as getting a drink, interacting with another character etc. Secondly, a role contains rules for the agent's interaction with other agents, for example, when it is appropriate for the agent to stop and interact with another character. Finally, a role contains a set of bindings to objects important to the current role.

Arising from the integration with the ALOHA system, a role also contains the animations required to render the character in that role. Figure 1 shows an illustration of the role passing process. As can be seen, at various times within a simulation, an agent can assume any one of a collection of roles as dictated by the schedule.

The main advantage of role-passing is the simplicity it lends to populating a virtual world with agents. Placing agents within a novel situation involves simply defining a new role. This eases some of the complications involved in attempting to design very general agents capable of behaving realistically in many situations, and avoids having to write completely separate agents for different roles within a single scene.

Another advantage of the role-passing technique is that it moves some way towards creating agents capable of being transferred between different applications, a major goal of intelligent agent research (Aylett, 2000). Through role-passing, the same basic agent is able to behave believably in very different situations.

Finally, role passing complements LODAI. By assuming and discarding roles as required, motivations unrelated to the current situation encountered by an agent are never considered. The result of this is that motivation levels surplus to the current situation need not be stored, and decisions not related to the current situation, are never even considered.

5. Simulation Example

As a means to test the incorporation of role-passing into the ALOHA system, a test scenario has been designed. In the test scenario, a bar scene has been created, in which a number of characters have been located.

Figure 2: Images show various aspects of a scene created using the ALOHA system with the role passing technique.

At this point, two distinctive roles are in use. The first is that of a bar patron. This entails sitting at tables, chatting with other characters and getting drinks when the character becomes thirsty. The second role is that of the barman. This role entails a special relationship with the bar objects located in the scene. The barman watches the bar objects and if he sees a customer waiting at any of them he serves them a drink. At present the agents are implemented as simple behaviour based agents (following the scheme given in (Zubek, 2001)), not exploiting the full PPA architecture.

The aim of this simulation is to test the amalgamation of the role-passing technique with the ALOHA system. This has been achieved successfully, figure 2 shows a number of screenshots showing various aspects of the bar scene. The inclusion of the role passing technique has allowed the creation of dynamic scenes, avoiding the need to use scripts.

6. Conclusion

This paper has discussed the amalgamation of two separate research areas in order to create a realistic virtual environment in which virtual humans may socially interact with each other. This collaboration allows both groups to focus on their core area of research.

With regards to the ALOHA system, further research needs to be done not only in blending different animations at the same level of detail, but also in smoothly switching between similar motions that are using different levels of detail. In the case of the role-passing system, more advanced roles and behaviours, plus more sophisticated character interaction will be added.

One final issue, which must be mentioned, is that of the evaluation of virtual worlds. Evaluation of such systems is notoriously difficult as the simulation of virtual humans is a complex task. However, a number of techniques, for example that described by Thalmann (Thalmann, 2001), have been developed and will be used to determine the success of this project. In addition, psychological evaluation techniques such as those described in (Hodgins, 1998; O'Sullivan, 2001) will also be used.

References

(Aylett, 2000) R Aylett, K Dautenhahn, J E Doran, M Luck, S Moss and Moshe Tennenholtz, *Can Models of Agents be Transferred between Different Areas?* Knowledge Engineering Review, 15(2), pp. 199-203. June 2000.

(Carlson, 1997) D A Carlson, J K Hodgins, *Simulation levels of detail for real-time animation*, Proc. Of Graphics Interface '97, pages 1-8, 1997.

(Cassell, 2001) J Cassell, H Vilhjamsson, T Bickmore, *BEAT: the Behaviour Expression Animation Toolkit*, Proc. of SIGGRAPH 2001, pp. 477-486.

(DeRose, 1998) T DeRose, M Kass, T Truong, *Subdivision Surfaces in Character Animation*, Proc. of the 25th Annual Conference on Computer Graphics, July 1998.

(Dingliana, 2000) J Dingliana, C O'Sullivan, *Graceful Degradation of Collision Handling in Physically Based Animation*, Computer Graphics Forum. Vol 19(2000), Number 3, pp 239-247.

(Fairclough, 2001) C Fairclough, M Fagan, B Mac Namee, P Cunningham, *Research Directions for AI in Computer Games*, Proc. of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science pp. 333 – 344, 2001

(Giang, 2001) T Giang, R Mooney, C Peters, C O'Sullivan, *ALOHA: Adaptive Level Of Detail for Human Animation. Towards a new Framework*, Eurographics 2000, Short Papers Programme.

(Hodgins, 1998) J K Hodgins, J F O'Brien, J Tumblin, *Perception of Human Motion with Different Geometric Models*, IEEE Transactions on Visualization and Computer Graphics, Vol. 4, No. 4, pp. 307--316, (Proc. of Eurographics 2000), October--December 1998.

(Horswill, 1999) I D Horswill, R Zubek, *Robot Architectures for Believable Game Agents*, Proceedings of the 1999 AAAI Spring Symposium on Artificial Intelligence and Computer Games, AAAI Technical Report SS-99-02.

(MacNamee, 2001) B MacNamee, P Cunningham, *Proposal for an Agent Architecture for Proactive Persistent Non Player Characters*, Proc. of the Twelfth Irish Conference on Artificial Intelligence and Cognitive Science pp. 221 – 232, 2001.

(Mateas, 1997) Michael Mateas, *An Oz-Centric Review of Interactive Drama and Believable Agents*, Technical Report CMU-CS-97-156, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. June 1997.

(O'Sullivan, 2001) C O'Sullivan, J Dingliana, *Collisions and Perception*, ACM Transactions on Graphics. (To appear, 2002)

(Thalmann, 2001) D Thalmann, *The Foundations to Build a Virtual Human Society*, Proc. of Intelligent Virtual Actors (IVA) 2001, Madrid, Spain.

(Caicedo, 2000) A. Caicedo, and D. Thalmann, *Virtual Humanoids: Let them be autonomous without losing control*, The Fourth International Conference on Computer Graphics and Artificial Intelligence. Limoges, France. May 3 - 4, 2000.

(Wooldridge, 1995) M Wooldridge, N R Jennings, *Intelligent agents: Theory and practice*, The Knowledge Engineering Review, 10(2):115--152, 1995.

(Zubek, 2001) R Zubek, *Game Agent Control Using Parallel Behaviors*, Technical Report, 2001.

AGENT CHAMELEONS: MIGRATION AND MUTATION WITHIN AND BETWEEN REAL AND VIRTUAL SPACES

G.M.P. O'Hare[1] & B.R. Duffy[2]
[1]Department of Computer Science, University College Dublin (UCD), Belfield, Dublin 4;
[2] Media Lab Europe (MLE), Sugar House Lane, Crane Street, Dublin 8
Gregory.OHare@ucd.ie; brd@media.mit.edu

Abstract

This paper introduces the concept of *Agent Chameleons* and investigates the motivation, desire and possibilities of agent migration and mutation within and between real and artificial spaces. Information spaces manifest themselves in many forms. Benefits exist where the agent/user is immersed and feels a sense of presence within such spaces. Agents often take on an embodied form. This form embodies the agent together with its associated behaviours and capabilities. The agent behaviour is inextricably mediated by a set of laws, the *agent physics laws*. This paper investigates the possibility of agents seamlessly migrating across traditional impervious inter-world boundaries and the associated possibility of agent forms mutating in order to empower itself. Such mutation may occur as the agent context changes. Context shifts can occur as the agent migrates from one world to another or indeed when environmental factors change within a given world.

1 Introduction

"If I had a world of my own, everything would be nonsense, nothing would be what it is, because everything would be what it isn't. And contrary wise, what it is it wouldn't be and what it wouldn't be it would."

Lewis Carroll, Alice in Wonderland

Increasingly we are witnessing a convergence between what were previously viewed as discrete information spaces. Historically the physical world was the first information space but has subsequently been joined by a rich array of virtual information spaces.

This paper seeks to investigate the integration of these information spaces by blurring the traditional boundaries between the Physical & the Virtual (information spaces). Ultimately we strive to support the opportunistic migration of the agent chameleon from the physical environs to that of the virtual and vice versa. The ultimate goal is seamless transition as the entity moves from one space to another.

It is our conjecture that the agent chameleon must have a sense of presence [Breazeal 2000] [Sas & O'Hare 2001] and thus be embodied within a given world. Indeed the form of an entity inextricably dictates or constrains its behaviour and capabilities within a particular environment. The optimum form is very much dependent upon its world [O'Hare 2000]. Judicious selection of appropriate forms or persona ought to empower the entity. In particular we view the mutation and evolution of agent forms as the ultimate animation of an expressive character, whereby the entire form may change rather than the mere change of limited components like the facial expression or character posture. To date this possibility has not been exploited nor delivered.

2 Objectives

The objectives of *Agent Chameleons* are as follows:

1) To investigate the choice and selection of embodiment forms;
2) To develop a mechanism whereby agent chameleons can migrate;
3) To investigate processes that support the mutation and evolution of agent forms;
4) To re-examine the concept of community and offer a redefinition based upon Collaborative Immersive Virtual Environments [O'Hare et al 2001]

3 The Vision

Within this research we seek to dismantle traditional world boundaries. We strive for a seamless migration from the physical to the virtual and the virtual to the physical. Synonymous with this migration is the obligation or possibility for *agent metamorphosis* whereby the entity changes representation such that its form is appropriate to the new *context.* The context is viewed as a tuple of task or activity and the new environment. Examples include:

- Agents adopting an avatar persona;
- Humans adopting a robotic avatar persona;
- Robots adopting a humanoid persona (HMD's)

The form of an entity inextricably dictates or constrains its behaviour and capabilities. The form is very much dependent upon its world. Appropriate forms or persona ought to empower the entity. In particular we will investigate two approaches.

(1) Augmentation of the *virtual* with the *physical*
This will involve the depiction of photorealistic avatars representing real people in virtual spaces as typified by the Blue-c Project and the ETH's Real Humans in Virtual Worlds [Staadt 2000].

(2) Augmentation of the *physical* with the *virtual*
Involving the use of VR Headsets enabling users to see virtual characters in real space. Influentail examples include the Virtual Round Table developed at GMD [Broll et al 2000a, Broll et al 2000b], and the Interactive Video Environment(IVE) of MIT [Wren et al 96]

In supporting agent immersion and migration from one world to the next we envisage an *agent chameleon* that will evolve a unique individual personality through prolonged interaction with the user. Such a trait aims to ensure agent-person familiarisation over time and across platforms (dimensions/information spaces). This strong notion of personality and character addresses the inherent embodiment issues of agent mutation. The agent effectively embodies itself depending upon the required scenario, a concept completely new to the field of artificial intelligence and embodied systems.

4 Related Work

This work will develop influential research within the arena of collaborative and immersive environments. It builds upon seminal work conducted by the Collaborative Virtual Environment (CVE) community. Notable pioneering systems that incorporate agent-based techniques include DIVA-Distributed Intelligent Virtual Agent [Vosinakis et al 1999], MAVE-Multi-agent Architecture for Virtual Environments [Cobel & Harbison 1998], STEVE [Johnson & Rickel 1999], Trilogy [Norman & Jennings 2000], ECHOES [O'Hare et al 1999] [O'Hare et al 2000a] {O'Hare et al 2000b] [Pasquarelli et al 1999], ENTER [Guinan et al 2000].

The realisation of evolvable characters in virtual environments will draw inspiration from such work as Synthetic Characters at MIT-Media Laboratory [Blumberg 1996], and work on agents as synthetic characters [Badler et al 1993], [Cassell 1994], [Doyle & Hayes-Roth 1998], [Foner 1997], [Reilly 1996].

Specifically, the agent chameleons will draw upon work undertaken by the one of the authors in mobile agents [Hristova & O'Hare 2001a] [Hristova & O'Hare 2001b] [O'Ha 2000] [O'Ha2001] [O'Hare & Jennings 1996], the synthesis of real and virtual environments (VR-Workbench, [Duffy et al 1999), Social Robots [Duffy et al 2000] [O'Hare & Duffy 2002] [O'Hare et al 1999], and adaptive social interaction [Brezzeal 1998] [Breazeal 2000].

5 Methodology

Existing research has investigated the development of social robotics, a mechanism that supports the opportunistic collaboration of teams of robots in the solution of joint goals [O'Hare & Duffy 2002],[Duffy 2000].

Within the context of this work a *Virtual Robotic Workbench* [Duffy et al 1999] has been designed and created which supports the articulation of robotic experiments, through the medium of the Internet. Key experimental descriptors were identified and instantiated. Such descriptors would permit the choice of the world within which the robots would be situated, the initial placement and number of robots, and the skill set and goal(s) associated with each robot. The Virtual Robotic Workbench serves as a visualisation metaphor depicting robotic avatars, in an accurate manner, reflecting the behaviour of the real world counterpart.

The Virtual Robotic Workbench served as the initial start point for this research and acts as a bootstrap in order to ensure the rapid delivery of proof-of-concept demonstrators. The intention in the first instance is to overlay this virtual world with *predesignated portals* or *black holes*, through which the world may be entered and exited. Thus entities migrating within a physical world could then be teleported into a virtual world and vice versa. The migration could result in an appropriate

Figure 1: A Physical World inhabited by Soccer Playing Robots

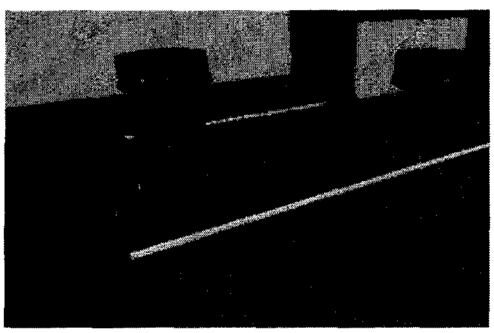


Figure 2: A Virtual World inhabited by Virtual Soccer Playing Robots

mutation of the entity thus manifesting itself differently within the new environment. A robotic avatar for example may take on the guise of a warrior in a hostile environment, whilst in a children's educational environment it may be cuddly and cute.

Two worlds are depicted in Figures 1 and 2, which respectively depict a physical and virtual representation of the same information content, namely soccer playing robots. We are currently developing physical and virtual worlds that are interconnected. Thus robots in the physical soccer scenario upon crossing the centre line of the pitch move into the other half of the pitch that is virtual with virtual ball, players and goals.

The *agent* is thus considered as an autonomous, mobile and social entity in the classic multi-agent systems sense. Above and beyond this it is considered an *agent chameleon*. The agent has at any given instance a *persona* and associated with a given persona are a given set of capabilities. The migration and mutation of agents may thus be invoked in one of two manners, firstly by the agent itself through *proactive mutation/migration* or secondly, as a result of environmental events, *reactive mutation/migration*. Both proactive and reactive responses are determined by input parameters which originate internally and externally to the agent. The former comprise the personality attitudes (e.g. aggression, friendliness, moodiness) and mentalistic attitudes (e.g. Beliefs, Desires, Intentions(BDI)) while the latter are situational or environmental.

The agent chameleon is modelled on the idea of realising an *digital friend*. This research dismantles existing constraints whereby an agent is restricted to either a particular hardware or a software environment and aims to achieve a synthesis between both the physical and the virtual information spaces. It relies on wireless technologies to migrate between platforms (802.11/Bluetooth). This work draws on work undertaken by Keegan and O'Hare 2002] in the mobile agent research and extends this through the development of strong personality functionality and a tangible presence in multiple information spaces.

As mentioned in section 4, the role of a strong sense of personality and synthetic character play an important role in Human-Agent Interaction(HAI). Personality in this work is instrumental in influencing the mutation of the agent chameleon through information spaces. An example is the persistence of a particular unique quirky behaviour as an agent mutates from one form to another. The aim is to maintain a sense of personal intimacy with the agent notwithstanding its mutation.

The application domain is analogous to bringing one's pet for a walk (mobile agent on a PDA device) and from time to time unleashing it to run an errand (agent travels via wireless protocols to other devices, performs some task, and returns at some time later). Audio, visual, and motion behavioural responses will constitute the familiarisation mechanisms to facilitate consistency in personal relationships developed between the user and the agent as it mutates across information spaces.

One can similarly draw analogies between the "Tamagochi" idea where a personal bond is developed between the user and the agent over time. This is extended greatly through the following innovative research initiatives:

1) The development of a strong sense of behavioural identity through personality and character traits exhibited by the agent to the user.
2) The mutation of forms: The agent is capable of migrating through virtual and physical information spaces through its selection of a "body" conducive to the task at hand based on its needs, desires, and options.

A richer interaction medium would thus be engendered between the user and the *collective information space*[1]. Through the realisation of a personal relationship between the agent chameleon and the user, this project aims to achieve a more tangible sense of presence in the digital world. Traditional human-computer interaction (HCI) is generally a highly structured formal medium of "conversation". This work seeks to extend from the keyboard and mouse paradigm to a notion of *personable computing* through the development of *agent friends* and to address the following crucial issues in HCI:

1) how to naturally entrain users to fully and comfortably take advantage of technology
2) how to compensate a loss of efficiency in the interaction process as the medium becomes less formal/structured (personalisation).

In parallel to the fusion of the physical and virtual spaces we also examine the fusing of the social space where characters may be virtual or real. Thus, a redefinition of the traditional community concept is necessitated [O'Hare & Byrne et al 2001], whereby new forms of *social inclusion* and *social norms* are accommodated. Further to this we will investigate *Social embodiment* and such issues: as robots delivering a form of telepresence for an associated human in a social space.

6 Conclusions

Within this paper we explore the blurring of the traditional information space boundaries, in particular the 4[th] dimension where the virtual is fused with the physical in an indistinguishable manner. Further to this we examine the decision making process that underpins the change of agent form and the potential for such processes as *agent cloning* and *agent evolution* where in a Darwinian sense through *socially situated learning* [Breazeal 2000] the agents may learn from their experience and evolve in the light of this. We postulate that the next form of evolution that must be addressed is the evolution of autonomous software entities.

[1] By Collective Information Space we refer to the Union of all spaces physical and virtual.

Acknowledgements

This work was undertaken as part of the Agent Chameleons project a collaborative project undertaken between the Department of Computer Science, University College Dublin (UCD), Media Lab Europe (MLE), Dublin and Media Lab, MIT, Cambridge,USA We gratefully acknowledge the financial support of the Higher Education Authority (HEA) Ireland.

References

B. M Blumberg,. 1996. Old Tricks, New Dogs: Ethology and Interactive Creatures. Ph.D. Dissertation, MIT, 1996.

C. Breazeal(Ferrell), and B. Scassellati, "Infant-like Social Interactions Between a Robot and a Human Caretaker", *Special issue of Adaptive Behavior on Simulation Models of Social Agents*, 1998.

C Breazeal, "Sociable Machines: Expressive Social Exchange Between Humans and Robots". Sc.D. dissertation, Department of Electrical Engineering and Computer Science, MIT, 2000.

W Broll, E., Meier, T Schardt,.: "The Mixed Reality Stage". *CAA Proceedings of the ACM SIGGRAPH 2000, 27th International Conference on Computer Graphics and Interactive Techniques*, (New Orleans, USA, July 23-28, 2000), p. 231, ACM, New York, 2000,

W. Broll, E. Meier, T. Schardt, "The Virtual Round Table - A Collaborative Augmented Multi-User Environment". *Proc. of ACM CVE 2000: The 3rd Inter'l Conf. On Collaborative Virtual Environments*, E Churchill, M. Reddy (eds.), ACM, New York, pp39-46, 2000

N. I. Badler; C.B. Phillips; and B.L. Webber,.. *Simulating Humans: Computer Graphics,* *Animation, and Control*,. Oxford University Press, 1997.

J. Cassell; C. Pelachaud, N. Badler; M. Steedman; B. Achorn; T. Becket, B. Douville, S. Prevost, and M. Stone, Animated conversation: rule-based generation of facial expression, gesture, & spoken intonation for multiple conversational agents. In *Proceedings of SIGGRAPH '94*, 1994.

J. Cobel, K. Harbison, "MAVE: A multi-agent architecture for virtual environments", in *Proc 11th Int. Conf. on Industrial and Engineering Applications of artificial intelligence and Expert Systems*, LNAI, Vol.1415, Springer Verlag, 1998.

B. R. Duffy, R. W. Collier, G. M. P. O'Hare, C. F. B. Rooney, , R. P. S O'Donoghue,. Social Robotics: Reality and Virtuality in Agent-Based Robotics, In *Proc. Bar-Ilan Symposium on the Foundations of Artificial Intelligence: Bridging Theory and Practice,Ramat Gan*, Israel, 1999.

B. R. Duffy, C. Garcia, C. F. B. Rooney, G. M. P. O'Hare, Sensor Fusion for Social Robotics *Proc. 31st International Symposium on Robotics (ISR2000)*, May 14-17, 2000, Palais des congrès, Montréal, Canada, 2000.

P. Doyle, and B. Hayes-Roth, Agents in annotated worlds. In *Proceedings of the Second International Conference on Autonomous Agents*. Minneapolis, MN: ACM Press, 1998.

B. R. Duffy,., The Social Robot, Ph.D Thesis Department of Computer Science, University College Dublin (UCD), Dublin, Ireland, 2000.

L. Foner, Entertaining agents: A sociological case study. In *Proceedings of the First International Conference on Autonomous Agents*, 122-129. Marina del Rey, CA: ACM Press, 1997.

T. Guinan, T., G. M. P.O'Hare, and N. Doikov, ENTER: The Personalisation and Contextualisation of 3-Dimensional Worlds, *Proc. 8th Euromicro Workshop on Parallel and Distributed Processing, EURO-PDP 2000*, pp 142-148, IEEE Computer Society Press, Jan. 19-21st , 2000, Rhodes, Greece, 2000.

N. Hristova, & G. M. P. O'Hare, Ad-Me: A Context-Sensitive Advertising System, *Proc. Third International Conference on Information Integration and Web-based Applications and Services (iiWAS 2001)*, 10-12 Sept. 2001, Linz, Austria,

N. Hristova, & G. M. P. O'Hare, Ad-me: Intelligent Context-Sensitive Advertising within a Mobile Tourist Guide,, *Proc 12th Irish AI and Cognitive Science Conference*., 5-7th Sept, 2001, University of Maynooth, Ireland.

L. Johnson, J. Rickel, "Virtual humans for team training in VR", *Proc. 9th World Conference on AI in Education*, pp578-585, July 1999.

S. Keegan., S., G. M. P. O'Hare, "Easi-Shop: Enabling product description services and automated shopping using mobile agents with the Ericsson Bluetooth Development Kit", submitted to 17th ACM Symposium on applied computing (SAC2002), Madrid, Spain, March 10-14, 2002.

T. J. Norman, N. R. Jennings, "Constructing a virtual training laboratory using intelligent agents", Int Journal of Continuous Engineering and Life-Long Learning, 2000.

G. M .P. O'Hare, & B. R Duffy,. The Social Robot, Wiley Publishers, (In Preparation)

G. M .P. O'Hare, & B. R Duffy, R. W. Collier, C. F. B. Rooney., R. P. S. O'Donoghue,., Agent Factory: Towards Social Robots, *Proc. First International Workshop of Central and Eastern Europe on Multi - Agent Systems (CEEMAS'99)*, St.Petersburg, Russia, 1999.

G. M. P. O'Hare, T. Delahunty, A. J. Murphy, and K. Sewell, ECHOES: A Collaborative Virtual Training Environment. *Proceedings of the Fifteenth Twente Workshop on Language Technology* May 19-21 1999 Enschede, The Netherlands, 1999.

G. M. P. O'Hare, Agents, Mobility and Virtuality: A Necessary Synergy, In *Proceedings of International ICSC Symposium on Multi-Agents and Mobile Agents in Virtual Organizations and E-Commerce (MAMA'2000)*, December 11-13, 2000, Wollongong, ICSC Academic Press, ISBN 3-906454-24-X, 2000.

G. M. P. O'Hare, Personalised and Contextualised Content Delivery for Mobile Users, *In Proc.of 1st International Conference on Universal Access in Human Computer Interaction UAHCI 2001*, August 5th-10th 2001, New Orleans, Lousiana, USA, Published as *Universal Access in HCI: Towards an Information Society for All* published by Lawrence Erlbaum Associates, Inc., 2001

G. M. P. O'Hare, and N. R. Jennings, (Editors.), *Foundations of Distributed Artificial Intelligence*, Sixth

Generation Computer Series, Wiley-Interscience Publishers, New York, ISBN 0-471-00675, 1996.

R. P. O'Rafferty, M. J. O'Grady, G. M. P. O'Hare. A Rapidly Configurable Location-Aware Information System for an Exterior Environment. *Proc. International Symposium on Handheld and Ubiquitous Computing (HUC 99),* September 27-29, 1999 Karlsruhe, Germany, 1999.

G. M. P. O'Hare, C. Sas, C. & C. Byrne, C,. Agent Mediation and Management of Virtual Communities: A Redefinition of the Traditional Community Concept, Symposium on Technology, Economic and Social Applications of Distributed Intelligence (TESADI'01, *In Proc of IEEE Int'l Conference on Systems Man and Cybernetics (SMC'01),* Tucon Arizona, Oct 7-10th, 2001, IEEE Press, 2001.

G. M. P. O'Hare, K. Sewell, A. Murphy, & T. Delahunty,. ECHOES: An Immersive Training Experience. *In Proceedings of International Conference on Adaptive Hypermedia and Adaptive Web-based Systems (AH2000).* Springer Verlag.

G. M. P. O'Hare, K. Sewell, A. Murphy, & T. Delahunty, An Agent Based Approach to Managing Collaborative Work within both a Virtual Environment and Virtual Community. *In Proceedings of the Workshop on Intelligent Agents for Computer Supported Co-operative Work: Technologies & Risks. Fourth International Conference on Autonomous Agents,* Barcelona, Spain.

A. Pasquarelli, F. de Stefani, G. M. P. O'Hare, & A. J. Murphy, ECHOES: EduCational Hypermedia OnlinE System, *Proceedings of IEEE MultiMedia Conference*, Firenze, Italy June 19th-21st 1999.

W. S. Reilly, W. S. Believable Social and Emotional Agents. Ph.D. Dissertation, CMU, 1996.

C. Sas, C., G. M. P. O'Hare, Defining and Measuring Presence in Non-Virtual Environments: An Experimental Study, In *Proc. 4th International Workshop Presence 2001*, May 21-23rd, Temple University, Philadelphia, PA, USA, 2001.

O. G. Staadt, A. Kunz, M. Meier, M. H. Gross: The Blue-C.: Integrating Real Humans into a Networked Immersive Environment. *Proc. of ACM Collaborative Virtual Environments 2000*, pp. 201-202, ACM Press, 2000.

S. Vosinakis, G. Anastassakis, T. Panayiotopoulos, "DIVA: Distributed Intelligent Virtual Agents", Proc. VA99, *Workshop on Intelligent Virtual Agents, at UK, VRSIG*, Sept 1999.

C. R. Wren, F. Sparacino, et al. TR#372: Perceptive Spaces for Performance and Entertainment: Untethered Interaction using Computer Vision and Audition *(Submitted to Applied Artificial Intelligence (AAI) Journal*, March 1996.

Vision-Based Reaching for Autonomous Virtual Humans

Christopher Peters, Carol O' Sullivan
Image Synthesis Group, Trinity College, Dublin 2, Republic of Ireland
email: {christopher.peters, carol.osullivan}@cs.tcd.ie

Abstract

A method for the generation of realistic real-time goal-directed virtual human arm motion is presented. Agents are endowed with a rudimentary synthetic vision and memory system that is used to gather and store data about objects in the vicinity. Agents then use the perceived object data rather than global database object data for the planning of reaching arm motions. Our method differs from previous attempts at goal-directed motion generation in that it uses sensory information available to the agent in order to distinguish between movements towards those objects visible to the agent, and movements towards memorised object locations. The generation of appropriate arm configurations under these circumstances is based on results from neurophysiology.

1 Introduction

This paper combines a synthetic vision and memory model, incorporating results from neurophysiology, to compute realistic arm configurations for reaching towards objects and the remembered locations of objects. We distinguish between objects that are out of view and those that are in view since neurophysiological results have shown that the computation performed by humans varies under these two conditions (see Soechting and Flanders 1989).

The method presented in this paper is applicable to real-time virtual environments for the realistic animation of agent-object interactions. For example, consider a bar scene. An agent is sitting at a table enjoying a drink. Because the glass is visible to our agent, reaching motions towards the glass are accurate. Suddenly a friend walks over, and our agent becomes distracted in a riveting conversation. During the conversation, our agent decides to reach for the glass, which is no longer inside its visual field (our agent is looking at his/her friend). Under these circumstances, the agent's reach is less accurate, since the agent is reaching towards a remembered location. It is also possible that the barman has taken the glass, or that the glass has been moved. It would be more realistic to have our agent look back at the remembered position of the glass, in order to visually establish that it is still there, and also to provide a correction to the reach in order to guide the hand towards the glass with precision.

2 Related Work

The generation of natural arm motions and implementation of synthetic vision have been applied to the field of computer graphics, albeit separately.

Lee at al. (1990) simulated lifting motions based on comfort, strength and perceived exertion factors. Tolani and Badler (1996) propose an analytic inverse kinematics based approach. Their basic strategy is to reduce the degrees-of-freedom of the arm by one in order to obtain the closed-form equations for solving the inverse kinematics problem. As the authors have pointed out, a shortcoming of this method is that it does not have a well-defined theory for generating natural looking configurations. Kondo presents an inverse kinematics algorithm based on a sensorimotor model to generate natural looking arm postures. Since the sensorimotor model is approximate, an iterative pseudo-inverse Jacobian method is used to perform a final adjustment. However, no distinction is made between visible objects and virtual (remembered) objects: the sensorimotor model for remembered objects is used in all circumstances. Nebel (2000) uses a sensitivity model from neuroscience for realistic "cautious" collision free motion of the arm.
Numerous researchers have suggested the use of a virtual model of perception in order to permit agents to

perceive their environment (see Renault et al. 1990; Tu 1996). An early example applies group behaviours to simulated creatures (Reynolds 1987). Tu and Terzopoulos (1994) implemented a realistic simulation of artificial fishes. Noser et al. (1995) proposed a navigation system for animated characters using synthetic vision and memory. Kuffner and Latombe (1999) provide real-time synthetic vision, memory and learning, and apply it to the navigation of animated characters.

3 Synthetic Vision and Memory

Our synthetic vision module is based on the model described by Noser et al. (1995). This model uses *false-colouring* and *dynamic octrees* to represent the visual memory of the character. We adopt a similar system to Kuffner and Latombe, where the speed of the memory component is increased by removing the octree structure, replacing it with a vector containing object observation information.

The process is as follows. Each object in the scene is assigned a single, unique colour that identifies it. The rendering hardware is then used to render the scene from the perspective of each agent. The frequency of this rendering may be varied, and could be used to implement attention levels for agents. In this mode, objects are rendered with flat shading in their unique false-colour. No textures or other effects are applied. The agent's viewpoint does not need to be rendered into a particularly large area: our current implementation uses 128x128 renderings. The false-coloured rendering is then scanned, and the object false-colours are extracted. These colours are then used to do a look-up of the objects in the scene, and object state information is extracted and stored with the agent in the form of observations. Essentially, an agent's visual memory consists of a list of these observations.

In our implementation, the precise position of an object in the environment is not stored as part of an observation. Rather, an approximation of the object's location in spherical coordinates with respect to the agent's viewing frame is used. During the scanning process, bounding boxes are assembled for each object based on the object's minimum and maximum x and y coordinates extracted from the view specific rendering, and the object's minimum and maximum z coordinates extracted from the z-buffer for that view. The object's position is then estimated to be the centre of this bounding box. This process has the overall effect of making accurate judgements about the positions of partially occluded objects more difficult. Also, estimates made about the distance to the centre of the object will vary depending on the obliqueness of the object with respect to the viewer.

Observations, then, are represented as tuples that are composed of the following components:

objID	globally unique identifier of the object
objAzi	azimuth of object
objEle	elevation of object
objDis	distance to object
t	time stamp

A specific object will have at most a single observation per agent. The observation will match the last perceived state of the object, although it must be noted that this may not correspond with the actual current state of the object. Also, since object coordinates are stored with respect to the viewing frame of the agent, they need to be updated whenever the direction of the view frame changes. This is not necessarily a problem; in our implementation observations are assumed to be composed of something analogous to iconic memory. Based on a number of heuristics, objects of particular interest would be converted from egocentric coordinates to world-space coordinates and placed in something more similar to a short-term or long-term memory system.

4 Reaching

The term 'sensorimotor transformation' refers to the process by which sensory stimuli are converted into motor commands (Pouget and Snyder 2000). Reaching towards a visual stimulus with the hand is an example of a sensorimotor transformation.

Research in neurophysiology suggests that there is a reasonably accurate internal representation of a target's location in a body-centred frame of reference (see Soechting and Flanders for references). The coordinates of this representation are referred to as extrinsic coordinates. There also appears to be an internal representation for the orientation of the upper arm and the forearm. These are referred to as intrinsic coordinates, and allow the calculation of the hand position.

Under normal, visually guided circumstances, the mathematically exact relationships between intrinsic and

Evolving Social Relationships with Animate Characters

Helmut Prendinger and Mitsuru Ishizuka
Department of Information and Communication Engineering
School of Engineering, University of Tokyo
7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan
{helmut,ishizuka}@miv.t.u-tokyo.ac.jp

Abstract

In order to come across as life-like or animate, characters must change their attitudes ('evolve') as a consequence of their affective interaction histories with other agents. This paper describes a model of attitude and familiarity change that allows to model simple forms of evolving social relationships. We illustrate our model by means of two web-based interaction scenarios that feature animated characters in the role of comics-style actors.

1 Introduction and Motivation

In their work on 'animate characters', Hayes-Roth and Doyle (1998) describe how to endow agents with life and personality, so that users will enjoy interacting with them. We follow their naming convention, and use *animate* rather than *animated* in order to emphasize the importance of an agent's 'inner life' (or simply, character) rather than impressive graphics for its believability and life-likeness. But even if it is true that character comes first, there is vast evidence that an agent's embodiment may significantly contribute to the likeability of agents to users (McBreen et al., 2000). However, in applications of web-based characters or even characters for mobile devices restrictions on memory and processing power may prevent the use of highly expressive animations (Wahlster et al., 2001). In those settings, the agent's believability as an *animate* character will be a key concern.

Currently, the focus of most work in the field of embodied characters is on emotion modeling and expression (André et al., 1999; Gratch, 2000; Pelachaud and Poggi, 2001; Canamero and Fredslund, 2001). Emotion modeling basically describes the appraisal process, i.e., how events are evaluated as to their emotional significance for the agent. Often, characters are also endowed with a model of personality that gives their (re)actions consistency across different situations. Emotion expression, on the other hand, relates to various communicative modalities in which emotions are expressed, such as facial display, speech, and body posture. Characters appear animate because their embodied reactions to other agents or the user are life-like (believable). However, it seems to us that there is more to being animate than visually expressing an emotion. In particular, an agent may alter its emotional reactions over time, depending on the 'affective interaction history' with another agent. If some interlocutor triggers mostly positive (negative) emotions in the agent, it might change its attitude toward the interlocutor and be biased to appraise the interlocutor's future actions in a more positive (negative) way (Ortony, 1991).

In this paper, we focus on aspects of characters' internal lives that seem relevant for developing social relationships. More specifically, we try to model the change of attitudes and familiarity assessment that underly the appraisal process of attitude-based emotions such as *attraction* and *aversion*. In short, not only do attitudes contribute to the elicitation of emotions—but induced emotions may also change a character's affective state, in particular, its attitudes and familiarity toward another agent. This is important for the design of characters that are intended to be perceived as life-like or animate over extended periods of time, such as pet-type characters.

The next section briefly reviews related work. Section 3 first sketches an influential appraisal theory, and then develops a simple model of attitude and familiarity change. In Section 4, web-based interaction scenarios featuring life-like characters are introduced that implement our approach. Section 5 concludes the paper.

2 Related Work

The problem of modeling evolving social relationships has been addressed by various researchers from different disciplines. The biology-inspired area of 'Socially Intelligent Agents' aims to describe the mechanisms underlying the process of establishing and maintaining social relationships between human agents and artificial agents (Dautenhahn, 1998). Cañamero (2001) offers a good overview of issues relating to emotional agents in social interactions. Interestingly, she hypothesizes that agents will be more acceptable to humans if they "[...] reflect the evolution of emotional interactions over time—i.e. their 'history'." (p. 23). Breazeal and Velásquez (1998) model a situation where Kismet, a robot infant, is engaged in social learning guided by a human caretaker. Cassell and Bickmore (2001) investigate the role of social language in social relationships. Specifically, they describe trust-promoting strategies between agents and humans, e.g., by the use of small talk.

3 Evolving Social Relationships

> "[...] I shall try to make a case for the claim that in addition to values being an important source of emotions, emotions are an important source of values and, more specifically, they can be the source of value in schemas." (Ortony, 1991, p. 341)

We will take Ortony's statement as a starting point for our model of evolving relationships and focus on one interpretation of values—as positive or negative attitudes toward an agent or object (liking and disliking). In the following, we will first briefly report on an interesting subset of the emotion model of Ortony *et al.* (1988), the so-called OCC model, and then describe how social relationships might change based on emotions elicited during social interaction between agents.

3.1 Modeling Emotions

According to the OCC model, emotions—or more precisely, emotion types—are simply classes of emotion eliciting conditions such as the agent's beliefs, goals, standards, and attitudes. An interesting set of emotions is crucially dependent on the agent's attitudes. The agent might be *happy for* its interlocutor if the agent likes the interlocutor and experiences *joy* over a state of affairs that it presumes to be desirable for the interlocutor. Otherwise, if the agent dislikes the interlocutor, it might *resent* its interlocutor for the same reason. A similar symmetry can be found with the *sorry for* and *gloat* emotion types. Let us now introduce some terminology by giving the specification of two other attitude-based emotion types, *attraction* and *aversion* (Ortony et al., 1988).

- A (locutor-)agent $L1$ is *attracted to* agent $L2$ with intensity δ *if* (i) $L1$ likes $L2$ with 'appealingness' degree δ_{App} and (ii) $L1$ is familiar to $L2$ with degree δ_F.
- An agent $L1$ has *aversion against* agent $L2$ with intensity δ *if* (i) $L1$ dislikes $L2$ with 'non-appealingness' degree δ_{NApp} and (ii) $L1$ is familiar to $L2$ with degree δ_F.

For simplicity, all intensity values range over $\{0, \ldots, 5\}$, with zero and five the lower and upper bounds, respectively, and are combined by means of logarithmic combination, e.g. $\delta = \log_2 \left(2^{\delta_{App}} + 2^{\delta_F}\right)$ (see Prendinger and Ishizuka (2001) for more explanation).

Our interest here is the 'opposite direction' where emotions influence our attitudes toward an interlocutor, e.g., when we (suddenly) dislike the interlocutor because we got angry with her or him, or we like an interlocutor since she or he elicited mainly positive emotions in us Ortony (1991). We will also discuss an agent's change of familiarity (social distance) due to emotions elicited in the agent Pautler and Quilici (1998).

3.2 A Model of Attitude Change

Ortony (1991) suggests the notion of *(signed) summary record* to capture our attitude toward or dispositional (dis)-liking of another person. This record stores the sign of emotions (i.e., positive or negative) that were induced in the agent by an interlocutor together with emotions' associated intensities. For instance, if the interlocutor elicits *distress* with intensity 2, *anger* with intensity 1, and *joy* with intensity 5, the summary record of the agent will contain two values, a negatively signed value of 1 (3 divided by the number of situations), and a positively signed one of 1.67. In order to compute the current intensity of an agent's (dis)liking, we simply compare the (scaled) sum of intensities of elicited positive and negative emotions (δ^σ, $\sigma \in \{+, -\}$), starting in situation s_0, the situation when the interaction starts. We will only consider the 'winning' emotional state δ_w, i.e., the most dominant emotion, although in general, multiple emotions may be elicited in each situation. If no emotion of one sign is elicited in a situation, it is set to zero.

$$\delta^\sigma(s_n) = \frac{\sum_{i=0}^{n} \delta_w^\sigma(s_i)}{n+1}$$

Positive values for the difference $\delta^+ - \delta^-$ indicate an agent's liking of an interlocutor and negative ones indicate disliking in a certain situation.[1] For simplicity, we assume perfect memory of elicited emotions, i.e., the intensity of past (winning) emotions does not decay. If the interlocutor's recent behavior is mostly 'consistent' with the agent's past experience (i.e., both have same sign), it is reasonable to update the overall intensity of the agent's attitude according to the equation above.

Ortony (1991) also considers the more interesting case where an interlocutor the agent likes as a consequence of consistent reinforcement (suddenly) induces a high-intensity emotion of the opposite sign, e.g., by making the agent very angry. He suggests three types of reactions: (i) the agent is uncertain how to construct the summary record value; (ii) the agent updates the summary value by giving a greater weight to the inconsistent information; (iii) the agent ignores the inconsistent information in the construction of a summary value. Since there is plenty of evidence that recency of the interlocutor's inconsistent

[1] Situations are actually parameterized by the agent and interlocutor, referring to time points when emotions are elicited in an agent L as a result of communicating with interlocutor I.

behavior plays a significant role in determining an attitude (Anderson, 1965), we will focus on the second type of reaction. Although the notion of 'recency' could be generalized to m latest elicited emotions, we simply refer to the very latest emotion. Here, the update rule reads as follows (parameters are omitted).

$$\delta(s_n) = \delta^{\sigma}(s_{n-1}) \times \omega_h \mp \delta_w^{\overline{\sigma}}(s_n) \times \omega_r$$

The weights ω_h and ω_r denote the weights we apply to historical and recent information, respectively. ω_h and ω_r take values from the interval $[0, 1]$ and $\omega_h + \omega_r = 1$. A greater weight of recent information is reflected by using a greater value for ω_r. By way of example, let us assume that the agent likes its interlocutor with degree 3 and then gets angry at the interlocutor with intensity 5. The new value might be computed as $3 \times 0.25 - 5 \times 0.75$, resulting in a disliking value of 3.

The crucial question is how the obtained (dis)liking value affects future interactions with the interlocutor. We consider two interpretations:

- *Momentary (dis)liking.* The new value is active for the current situation and then enters the summary record.
- *Essential (dis)liking.* The new value replaces the summary record.

An instance of momentary liking are reciprocal feedback loops where a disagreeable interlocutor's (temporary) friendliness lets the agent adopt friendly behavior due to the elicitation of a positive emotion. Essential (dis)liking typically happens when the agent finds out something very positive (negative) about the interlocutor that is crucial for its model of the interlocutor.

It is interesting to observe that the way an agent deals with inconsistent information allows to make assumptions about its personality traits along the disagreeable–agreeable dimension. For instance, if the agent's attitude changes to essential disliking if made very angry once, it might be called unforgiving. Furthermore, a subtle interaction might exist between an agent's option for momentary or essential (dis)liking and the familiarity with the interlocutor. It can be argued that the most dramatic changes happen in recent evolving relationships, whereas agents familiar with each other rather experience momentary attitude changes.

3.3 A Simple Model of Familiarity Change

We take the notion of 'familiarity' to mean the social distance an agent has toward an interlocutor (Brown and Levinson, 1987). Per definition, we are close to family and friends, and distant to strangers. The problem we want to address in the following is how change in social distance might be captured computationally.

Pautler and Quilici (1998) investigate a special form of speech acts, called 'social perlocutions', that may change the interlocutor's relationship with the agent. They argue that positive emotions elicited in the interlocutor contribute to improving the interlocutor's social relationship with the agent. Although they do not explicitly discuss relationships in terms of familiarity, we belief that this interpretation is justifiable.

Our concept of familiarity degree δ_F considers the number and intensity of induced positive emotions δ^+.

$$\begin{aligned}\delta_F(s_0) &= 0 \text{ or pre-set to some value}\\ \delta_F(s_n) &= \delta_F(s_{n-1}) + \frac{\delta_w^+(s_n)}{\pi}\end{aligned}$$

If a negative emotion is elicited, $\delta_F(s_n) = \delta_F(s_{n-1})$. π is a factor that determines how rapid a character gets familiar with another agent. Unlike a character's (dis)liking, familiarity increases monotonically, i.e., once characters are socially close, they cannot subsequently get unfamiliar. Currently, our notion of familiarity is based on the (severe) simplifying assumption that emotions are taken as the only familiarity changing factor. Cassell and Bickmore (2001), on the other hand, consider the variety and depth of topics covered by conversing agents.

4 Web-based Interaction Scenarios

We will illustrate our model by means of two web-based scenarios that feature animated characters. In the first interaction scenario the user can play the "Black Jack" game in a virtual casino, and is guided by the animated advisor "Genie". The user may either follow or not follow Genie's advice, and independently of that, win or loose a game. Our goal is to show how the user's behavior as well as the (affective) interaction history determine Genie's reaction to the user's won or lost game. Genie appears animate since he may not express *joy* over a won game if the user repeatedly refuses to follow his advice.

The second interaction scenario involves a user controlled avatar who communicates with an angel-style character, by trying to figure out her wishes. Based on a very sparse emotion model, the aim of this scenario is to demonstrate how increased (decreased) liking and familiarity may 'switch' the angel character's affective reactions to *attraction* (*aversion*) toward the user's avatar. The character may appear animate as 'secondary' emotions such as attraction only occur as a result of a longer-term interactions.

In both scenarios, the Microsoft Agent package (Microsoft, 1998) is used to embed characters into a web page based JavaScript interface. An XML-based language called MPML (Multimodal Presentation Markup Language) is employed to specify sequential and parallel behavior of multiple characters (Descamps et al., 2001). For autonomous character control, MPML provides an interface to SCREAM, a tool that allows to script a character's mental state (including goals and personality traits) and

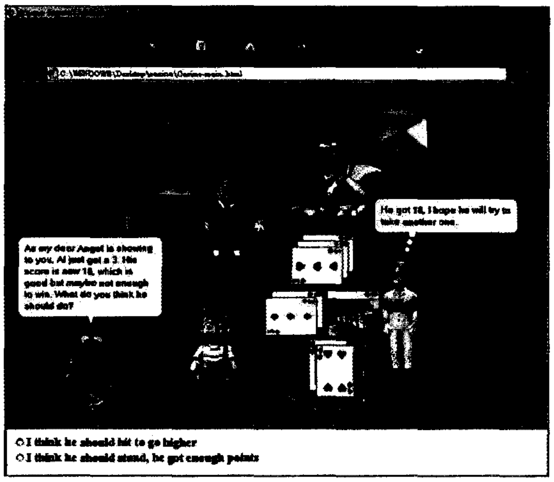


Figure 1: Casino Scenario.

mental processes such as emotion management and emotion regulation (Prendinger and Ishizuka, 2002). All characters are connected to a TTS (Text-to-Speech) engine and may express emotions by triggering pre-defined 2D animation sequences.

4.1 Playing Black Jack in a Virtual Casino

We will now watch the user playing five games of Black Jack and thereby demonstrate how Genie's mental make-up as well as the (affective) interaction history determine his behavior. Fig. 1 depicts the situation where the advisor Genie practices Black Jack with the user by commenting the game of character "Al" (Genie is the character at the bottom-left of the Internet Explorer window, and Al is the male character to the right of the dealer).

The following character profile is used (details about used intensities are omitted).

- *Genie's personality.* Agreeable and extrovert. His friendliness implies that negative emotions decay quickly whereas positive emotions remain active for a longer time.
- *Genie's goals.* He wants that the user wins with low intensity, and that the user follows his advice with high intensity.
- *Genie's social relations.* He is socially close to the user and initially slightly likes the user.

For expository reasons, we let the user *never* follow Genie's advice.

First game (user looses). Genie's winning emotional state is *distress* with high intensity, because the user did not follow his advice. However, he displays *distress* with low intensity as his agreeable personality effects a decrease in the intensity of negative emotion expression.

Second game (user looses). Genie is *sorry for* the user with high intensity, since positive ('sorry for' the user's lost game) emotions sum up (and decay slowly), which leads to an increase in Genie's liking of the user. His personality traits let him express the emotion with even higher intensity.

Third game (user looses). Genie *gloats* over the user's lost game, because at that point, the negative emotions dominate the positive ones as a consequence of the user's repeated refusal to follow Genie's advice. Hence Genie's attitude changes to slightly disliking the user which lets him experience *joy* over the user's *distress* (*gloat* with high intensity). Again, Genie's friendly personality decreases the intensity of the expressed emotion.

Fourth game (user wins). Genie's emotional state is *bad mood* with high intensity, slightly more than his *happy for* emotion (as the user wins the game this time). Here an overall, unspecific affective state (mood) is expressed with low intensity, rather than a specific emotion.

Fifth game (user wins). Genie's dominant emotion is *resent* with high intensity, because he slightly dislikes the user and consequently is feels *distress* that the user won although she or he ignored his advice. Genie expresses his emotion with reduced intensity.

An exhaustive exploration of all possible interaction patterns in the described game scenario reveals that Genie's reactions are conform at the beginning games and show more variety in the subsequent games. This can be explained by the development of Genie's attitude toward the user, depending on whether the user follows or refuses to follow Genie's advice. In effect, Genie's attitude decides, e.g., whether he is *sorry for* or *resents* the user's lost game. However, in accordance with Genie's agreeableness, his emotional reactions are mostly positive.

4.2 Interacting with Little Akko

Borrowing the idea from Fujio Akatsuka's manga series (Japanese comics) "Akko-chan's Got a Secret!", a character called 'Little Akko' (Akko-chan) plays the heroine of stories for kids. Little Akko has the power to be transformed into any person upon telling her wish to a magic mirror. By this magic, she has the power to solve many problems and even make other people happy. Figures 2 and 3 show her transformed into Little Chika, a girl whom her brother Kankichi likes. Social relationships in this comics book typically evolve in a quick and direct way and hence the stories lend themselves to easy testing of our model. We started to experiment with attitude and familiarity change based on a small set of emotion types: *joy*, *distress*, *attraction*, and *aversion*.

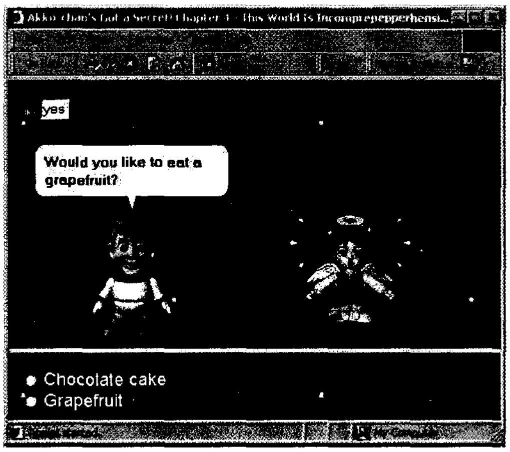


Figure 2: Angel is distressed toward Space-boy about getting a grapefruit.

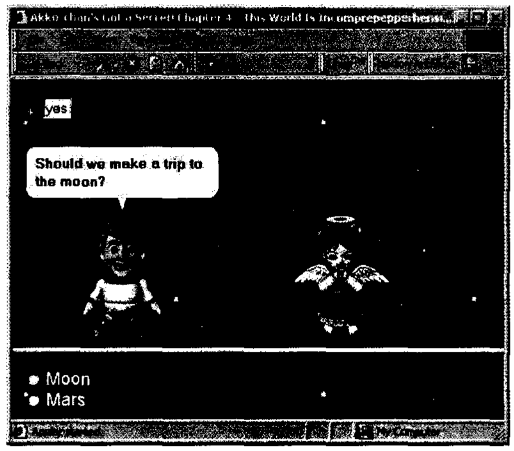


Figure 3: Angel is attracted to Space-boy after being offered a trip to the moon.

Currently, the interaction setting is fairly simple. The user can communicate with the "Angel" character (Little Akko transformed to Little Chika) by controlling an avatar, the "Space-boy" character in the role of Kankichi. By offering Little Chika items she likes, the user may increase her positive attitude and familiarity, otherwise her liking level for Kankichi goes down. Consider the following conversation.

- User may select "Strawberry milk" or "Lemon tea"
  Space-boy: Would you like to drink strawberry milk?
  Angel: Great! I like this drink.
- User may select "Chocolate cake" or "Grapefruit"
  Space-boy: Would you like to eat a grapefruit?
  Angel: No! I do not like that.
- User may select "Calculate" or "Hide and Seek"
  Space-boy: Do you want to play the Calculate Game?
  Angel: I really like that game!
- User may select "Rice" or "Noodles"
  Space-boy: Would you like to eat some rice?
  Angel: Yes! That is what I like!
- User may select "Moon" or "Mars"
  Space-boy: Should we make a trip to the moon?
  Angel: I enjoy being with you!

When Angel gets strawberry milk, she expresses *joy* as one of her goals is satisfied. After being offered a grapefruit, she shows her *distress* since she does not want this kind of dessert (see Fig. 2). However, in the conversation above, the user happens to repeatedly select items the Angel likes, which has two kinds of effects. Both the Angel's liking value toward the Space-boy and the familiarity level increase, and hence add to the intensity of the Angel's *attraction* toward the Space-boy (familiarity was incremented by 0.2 per elicited positive emotion). After the Space-boy offers the Angel a trip to the moon, her emotional state comprises two active emotions, *joy* (intensity 2) and *attraction* (intensity 3), and she expresses the emotion with the higher intensity (see Fig. 3). As we set the decay rate to a high level, all previously elicited emotions (including *distress* and *aversion*) are not part of the Angel's emotional state.

Although we believe that positive attitude and close social distance should have on the agent's emotion expression, it is not clear to us, how emotions such as *affection* or *aversion* should be instantiated by actual behavior. Currently, we use a direct way by simply letting the agent declare those emotions, e.g., *affection* as "I enjoy being with you". However, in some cases, attitude/familiarity based emotions might be used as biasing mechanisms for calculating the intensity of emotion expression, rather than emotions that are externalized by behavior.

5 Conclusion

In this paper, we try to animate characters by providing an explicit model of attitude and familiarity change. This effort toward believable and life-like characters complements research on characters' ability to express, recognize, and reason about emotion. Social interactions usually develop in some way, and we believe that change in attitude and social distance of interlocutors are key aspects of this evolution.

The most needed next step is experimentation with the proposed model. We hypothesize that (animated) characters adjusting their affective behavior as a result of the interaction history are more believable (and less boring) than characters that only consider the current interaction context. However, our model of familiarity change will have to be refined by considering the nature of the topics dealt with in a conversation. The interaction scenarios described in this paper do not allow to account for the topic parameter, as both are highly restricted and task-specific.

In the future, we hope to integrate our model to pet-type characters and web-based interactive characters that maintain believable long-term relationships with users.

Acknowledgements

We would like to express our thanks to the reviewers for their benevolent and motivating comments. This research is supported by the Research Grant (1999-2003) for the Future Program ("Mirai Kaitaku") from the Japanese Society for the Promotion of Science (JSPS).

References

N. H. Anderson. Primacy effects in impression formation using a generalized order effect paradigm. *Journal of Personality and Social Psychology*, 2:1–9, 1965.

Elisabeth André, Martin Klesen, Patrick Gebhard, Steve Allen, and Thomas Rist. Integrating models of personality and emotions into lifelike characters. In *Proceedings International Workshop on Affect in Interactions - Towards a New Generation of Interfaces*, pages 136–149, 1999.

Cynthia Breazeal-Ferrell and Juan Velásquez. Toward teaching a robot "infant" using emotive communication acts. In *Proceedings of Workshop on Socially Situated Intelligence, in conjunction with SAB-98*, 1998.

Penelope Brown and Stephen C. Levinson. *Politeness. Some Universals in Language Usage*. Cambridge University Press, 1987.

Lola Canamero. Building emotional artifacts in social worlds: Challenges and perspectives. In *Proceedings of 2001 AAAI Fall Symposium in Emotional and Intelligent II: The Tangled Knot of Social Cognition*, pages 22–30. AAAI Press, Technical Report FS-01-02, 2001.

Lola Canamero and Jacob Fredslund. I show you how I like you—Can you read my face? *IEEE Transactions on Systems, Man, and Cybernetics—Part A: Systems and Humans*, 31(5):454–459, 2001.

Justine Cassell and Timothy Bickmore. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents, 2001. User Modeling and Adaptive Interfaces. In Press.

Kerstin Dautenhahn. The art of designing socially intelligent agents – science, fiction, and the human in the loop. *Applied Artificial Intelligence, Special Issue on Socially Intelligent Agents*, 12(7–8):573–617, 1998.

Sylvain Descamps, Helmut Prendinger, and Mitsuru Ishizuka. A multimodal presentation mark-up language for enhanced affective presentation. In *Proceedings International Conference on Intelligent Multimedia and Distant Education (ICIMADE-01), Advances in Educational Technologies: Multimedia, WWW and Distance Education*, pages 9–16, 2001.

Jonathan Gratch. Émile: Marshalling passions in training and education. In *Proceedings 4th International Conference on Autonomous Agents (Agents'2000)*, pages 325–332, 2000.

Barbara Hayes-Roth and Patrick Doyle. Animate characters. *Autonomous Agents and Multi-Agent Systems*, 1 (2):195–230, 1998.

Helen McBreen, Paul Shade, Mervyn Jack, and Peter Wyard. Experimental assessment of the effectiveness of synthetic personae for multi-modal e-retail applications. In *Proceedings 4th International Conference on Autonomous Agents (Agents'2000)*, pages 39–45, 2000.

Microsoft. *Developing for Microsoft Agent*. Microsoft Press, 1998.

Andrew Ortony. Value and emotion. In W. Kessen, A. Ortony, and F. Craik, editors, *Memories, thoughts, and emotions: Essays in the honor of George Mandler*, pages 337–353. Hillsdale, NJ: Erlbaum, 1991.

Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1988.

David Pautler and Alex Quilici. A computational model of social perlocutions. In *Proceedings of COLING-ACL-98*, 1998.

Catherine Pelachaud and Isabella Poggi. Multimodal embodied agents. In *Proceedings Workshop on Multimodal Communication and Context in Embodied Agents, in conj. with Agents-01*, pages 95–100, 2001.

Helmut Prendinger and Mitsuru Ishizuka. Social role awareness in animated agents. In *Proceedings 5th International Conference on Autonomous Agents (Agents-01)*, pages 270–277, 2001.

Helmut Prendinger and Mitsuru Ishizuka. SCREAM: Scripting Emotion-based Agent Minds. In *Proceedings 1st International Joint Conference on Autonomous Agents and Multi-Agent Systems (AAMAS-02)*, 2002. To appear.

Wolfgang Wahlster, Norbert Reithinger, and Anselm Blocher. SmartKom: Multimodal communication with a life-like character. In *Proceedings of 7th European Conference on Speech Communication and Technology (Eurospeech-01)*, pages 1547–1550, 2001.

AVATAR ARENA: AN ATTEMPT TO APPLY SOCIO-PHYSIOLOGICAL CONCEPTS OF COGNITIVE CONSISTENCY IN AVATAR-AVATAR NEGOTIATION SCENARIOS

Thomas Rist, Markus Schmitt
DFKI GmbH, Stuhlsatzenhausweg 3,
D-66123 Saarbrücken, Germany
email: {rist, mschmitt}@dfki.de

Abstract

The Avatar Arena is a test-bed for the simulation of multi-character negotiation dialogues. Unlike other approaches to agent-agent negotiation processes in multi agent systems, the primary focus of our work is neither on negotiation protocols nor on optimisation issues. Rather, we are interested in the emulation of negotiation dialogues between affective characters that are embedded in a certain social context. In our current approach we try to exploit socio-physiological concepts, such as cognitive balance and dissonance. We illustrate the approach in the domain of meeting arrangement dialogues. A user can configure her avatar, send it to a virtual arena ,and have it negotiate on meeting dates with the avatars of other users.

1 Introduction

A good deal of research in the area of intelligent agent systems is driven by the idea to delegate tasks to an agent (Maes 1994). The accomplishment of delegated tasks often require an user's agent to get in contact and negotiate with agents of other users. In many cases the user will only be interested in the outcome of such a negotiation. For instance, when sending out a "bargain finder agent" the user may not care about the number of online shops visited or the details of the negotiation process - as long as a the agent was able to make a reasonable bargain buy. In this case the result of the task delegation is in the foreground. There are, however, other situations in which the user may have an interest in learning about how a certain negotiation result came about. In human-human negotiations, this is of particular interest in cases where the result of a negotiation can not be explained on the basis of a solely rational argumentation but only if the social context and the personalities of the negotiating parties need to be considered as well.

Within the context of the EU project Magicster we are currently investigating negotiation scenarios with emotional agents that are embedded in a social context. Somewhat similar to an arena, users send their delegates (avatars) to a virtual space where the avatars negotiate on behalf of their owners. Both result and process of a negotiation can be displayed to the users in form of a simulation using embodied conversational characters. Dealing with such "Avatar Arenas" requires a framework that covers the modeling of affective agents together with dynamically changing social relationships among them. While there are a number of elaborated approaches to model affect in synthetic characters (e.g., see (Paiva 2000), relatively little attention has been paid so far to the potential impact of social relationships on a character's behavior. One reason for this might be the fact that the it is often less clear what kinds of social relationships should be assumed and modeled when a human user interacts with a synthetic character. When switching to scenarios where multiple characters interact with each others (e.g., André et al. 2000), however, the need to model social relationships between the involved characters becomes more apparent, and some research groups already started to account for the social dimension in simulated conversations between animated characters. The work by Prendinger and Ishizuka (2001) deserves mentioning here. In their SCREAM system they explicitly model concepts, such as social distance, social power and thread, in order to enhance the believability of generated dialogues.

In our work, we are in particular interested in the dynamics of social relationships during negotiation dialogues. Our current working hypothesis is that socio-physiological theories of cognitive consistency, such as Heider's *Balance Theory* (1958) and Festinger's *Theory of cognitive dissonance* (1957) provide a starting point for a model that can capture some relevant aspects of group dynamics known from multi-party negotiation dialogues.

For the purpose of illustration and testing the approach, a meeting date negotiation scenario has been chosen to demonstrate the framework. Making a meeting arrangement with others is one of those everyday problems that is often delegated to personal secretaries or assistants. In our scenario we provide a group of users with the ability to have an avatar negotiate with the avatars of other users. However, unlike the various calendar tools that identify and free time slots for a group of users, we assume that the negotiation process will be driven by the avatar's personalities, their moods and emotions, and the social relationships amongst them as indicated by their owners.

2 The Avatar Arena Demonstrator

Technically speaking, an Avatar Arena can be conceived as a distributed n:1 client server architecture. While the server component provides the arena where the negotiation takes place, a client component allows the user to configure and instruct her/his avatar, and also to observe the negotiation process carried out at the server.

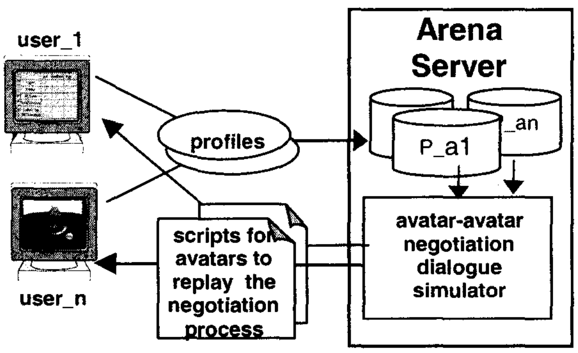


Fig. 1: Set up of the Avatar Arena Demonstrator

Building the Avatar Arena as envisaged requires a number of basic modeling tasks concerning an avatar's:

- understanding of the domain
- personal preferences and attitudes wrt. to domain concepts
- personality and affective behaviour
- social relationships wrt. to other avatars
- skills wrt. to conversation and negotiation

2.1 Modelling of domain concepts and preferences

Part of the domain modelling in the meeting appointment scenario is an ontology of meeting dates. Such ontologies have been set up in several other systems, e.g., in the Verbmobil system that aims at a simultaneous translation of spoken utterances in meeting arrangement dialogues (Wahlster 2000). However, since the primary interest of our research is not on modelling meeting appointment domains, only a rudimentary domain model has been build up. This model includes a rough characterisation of meeting dates among five attributes: time, type of activity, type of person(s) to be met (family / friends / business partners), type of temporal fixation (fixed / movable), and location of the meeting cf. Fig. 2.

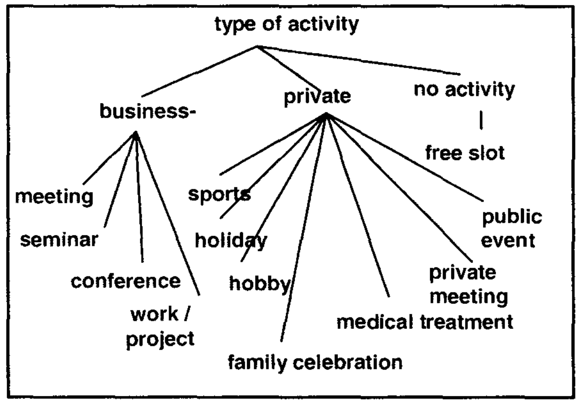
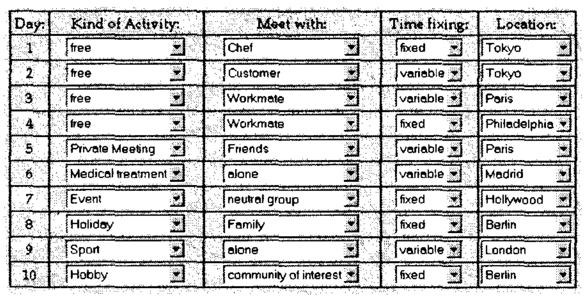


Fig. 2: Taxonomy of activity types in the "meeting date negotiation" domain.

In order to allow a user to enter already scheduled meeting dates, the client-side user interface of the Avatar Arena comprises a calendar-style input widget (cf. Fig. 3). The entries are then "read" by the user's avatar so that they can be taken into account when a new appointment has to be arranged. Thus, to a large extent, the contents of the calendar determines what an avatar can say about a the availability of a proposed meeting date, and vice versa which meeting dates the avatar may propose in a negotiation.



Day:	Kind of Activity:	Meet with:	Time fixing:	Location:
1	free	Chef	fixed	Tokyo
2	free	Customer	variable	Tokyo
3	free	Workmate	variable	Paris
4	free	Workmate	fixed	Philadelphia
5	Private Meeting	Friends	variable	Paris
6	Medical treatment	alone	variable	Madrid
7	Event	neutral group	fixed	Hollywood
8	Holiday	Family	fixed	Berlin
9	Sport	alone	variable	London
10	Hobby	community of interest	fixed	Berlin

Fig. 3: Part of the client-side UI that allows a user to specify already scheduled meeting dates.

How an avatar will behave in a negotiation, e.g., whether it will be willing to reschedule or cancel an already scheduled appointment, among other factors, depends on the importance that a user may attach to a certain activity, and this will certainly vary from one user to another. That is, in the general case there is no single importance order of meeting dates across users.

Rather, for each user we need to assess the importance of a certain meeting date.

There are several options for making such an assessment. Firstly, we could allow the user to explicitly specify an importance value for each already scheduled meeting. A more basic approach is to identify a number of general interests / value dimensions to which meeting dates can be related. We have taken such an approach and, for the purpose of our current demonstrator, we use the dimensions shown in Figure 4. Entries made in this form will be used to assess the importance of a certain meeting date. For instance, expressing a high interest in the dimension "career" would assign high importance values to activities that are related to work, such as participating in working meetings and business events. But also a private activity may become more important if the activity is carried out together with workmates or the boss.

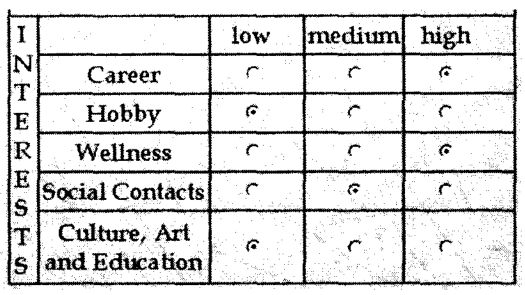


Fig. 4: Part of the client-side UI that allows a user to specify a simplified but individual interest profile.

When submitted to the Arena Server, all entries made in the client-side UI will be read by the user's avatar as input for the negotiation process with other agents. Note that an individual interest profile is not automatically revealed to all other negotiation partners. Rather, other characters will learn about the preferences of others only during the negotiation dialogue – e.g., when a character emphasises the importance of an already fixed date in order to justify its unwillingness to reschedule it.

2.2 Personality and affect

For the Avatar Arena we are especially interested in characters that represent distinguishable personalities. We assume that, in order to achieve believable characters, different personalities will be reflected in a character's conversational behaviour and negotiation style. To model the characters' personality, we orient ourselves on the so-called Five-Factor Model (McCrae & John 92). The FFM is a descriptive model, with the five dimensions (*Extraversion*, *Agreeableness*, *Conscientiousness*, *Neuroticism*, and *Openness*) being derived from a factor analysis of a large number of self- and peer reports on personality-relevant adjectives.

While there is an ongoing discussion among trait theorists on the number of traits and the relative importance of each of the traits, the dimensions *Extraversion*, and *Neuroticism* can be found in most trait theories. When dealing with social relationships among characters, such as liking and disliking others, the dimension *Agreeableness* is of importance, too. However, for the sake of simplicity, we have restricted our model and do not consider the dimensions *Conscientiousness*, and *Openness* in our current demonstrator.

Also, in our current implementation we do not yet model affective states of the characters. However, the incorporation of a model[1] of affect will be added in a further version of the demonstrator.

2.3 Social relations

As a first step for the conception of an early Avatar Arena demonstrator we rely on the concept of a cognitive configuration which can either balanced or unbalanced and which has its roots in socio-physiological theories of cognitive consistency, in our case especially in *Heider's Balance Theory* (1958) and *Festinger's theory of cognitive dissonance* (1957).

Roughly speaking, Heider's Balance Theory and related derivates start from the basic hypothesis that a good deal of inter-personal behavior and social preception is determined or at least co-determined by simple cognitive configurations which are either balanced or unbalanced. Together with the hypothesis that people tend to avoid unbalanced configurations or cognitive dissonances, one can make predictions about how a certain person may behave in certain social situations.

Balance Theory describes the phonological world of a person. It establishes a relationship between the person P perceiving the world, another person O, and an object X. To further characterise this situation, the notion of a *cognitive configuration* is introduced and in turn characterized by means of two distinct relations:

L-relations (liking / disliking) to express a person's attitude towards other persons or impersonal entities, e.g., P likes O, P likes coffee, etc.

U-relations (unit formation) to express a person's perceptive formation of cognitive units, e.g., P perceives the objects X1 and X2 being similar.

Cognitive configurations are always described from the perspective of the person P. They can be balanced or unbalanced. For instance, if P likes O, and P perceives the objects X1 and X2 being similar and at the same

[1] An advanced framework for modelling affective states of an agent is currently under development by Fiorella de Rosis' group at University of Bari - one of our consortium partners in the Magicster project.

time believes that O also perceives X1 and X2 being similar, P's perception of the situation (i.e. P's cognitive configuration) is balanced. In contrast, Fig. 5 illustrates an example of an unbalanced configuration taken from the meeting date arrangement domain. Suppose career is an important value for the parrot Peedy who likes Genie. Up to now, Peedy believed that career is also an important value for Genie. However, when talking to Genie about career opportunities, it turned out that Genie has no interest at all in issues that relate to career development. When learning about this, Peedy's cognitive configuration of the situation becomes unbalanced.

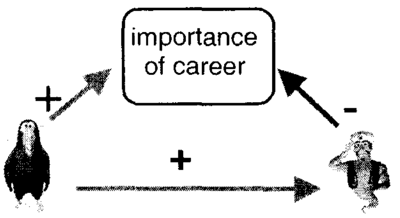


Fig. 5: Example of an unbalanced cognitive configuration.

In order to construct a matrix of L-relationships that may exist between a group of characters, we draw on sociometric concepts as introduced by Moreno (1974). In our test-bed we use the interface shown in Fig. 6 allowing a user to specify à priori social distances to the negotiation partners. For the sake of simplicity, we only distinguish between three different values (positive / negative / neutral) for each pair of characters. In the example, the user has chosen Peedy as an avatar, and indicates L-relationships with regard to the negotiation partners represented by the agents Robby, Genie, and Merlin respectively. That is, Peedy likes Robby, dislikes Genie, and has a neutral attitude towards Merlin.

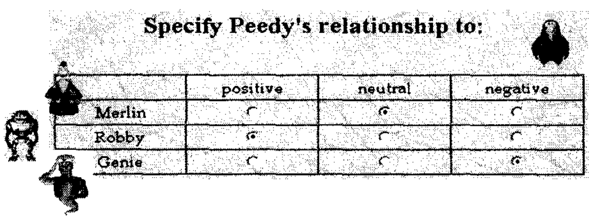


Fig. 6: Part of the client-side for specifying à priori social distances to the negotiation partners.

Interestingly, distance relationships are not necessarily symmetric. That is, a user may like another user (and thus specify that her avatar likes another user's avatar) while the other user actually dislikes her and instructs her avatar accordingly

Basically, the idea is to exploit the social context (as specified in terms of L-relationships) for the generation of more believable negotiation dialogues wrt.:

- negotiation behaviour (which goals to pursue and which strategies to apply)
- turn taking behaviour (when to take a turn, when to let others go ahead)
- content selection (what to say)
  linguistic style (how to say)

With regard to the current demonstrator our working hypothesises are:

(a) Before starting a negotiating dialogue all participating parties have a certain social distance to each other which can be represented in a sociomatrix.
(b) Characters make assumptions about the value systems of other characters so that the corresponding cognitive configurations are balanced (cf. Fig. 5).
(c) When a character discovers a mismatch between its assumption about another character's value system (i.e., when it learns about the importance values that its dialogue partner assigns to different meeting dates), this discovery may cause the experience of a dissonance and eventually triggers a change in the social distance to the other character so that a balance is achieved again.

3 Simulation of negotiation dialogues

The underlying building blocks of meeting date negotiation processes as discussed here are dialogue moves. All characters have a number of different communicative acts at their disposal. The repository of acts comprise general acts, such as inform and elaborate, as well as acts which refer more specifically to the different phases of a negotiation.

Opening phase:
- Greet, Reply-to-Greeting;
- Announce necessity to make an arrangement;
- Closing phase
- Wrap-up, Leave-taking
- Negotiation phase
- Request-proposal
- Propose-date
- Accept, Reject
- Justify-rejection
- Evaluate-justification
- Meta-communicative: Request-Turn, Give-turn, etc.

In principle, there are two fundamentally different approaches to model a multi-party negotiation. The perhaps most intuitive approach is to equip each character with a fully-fledged dialogue-manager and content planner. That is, all characters are modelled as autonomous entities that listen to contributions made by others, evaluate these contributions and react to them in the one or other way.

In contrast, one can adopt the metaphor of a director who acts out a single script that specifies the flow of the negotiation dialogue. Of course, in order to obtain a believable result, the director has to consider the knowledge and personalities of all characters and must be able to anticipate how a certain character may react at a certain stage of the negotiation.

At DFKI, we gained experiences with both approaches. While a distributed approach seems to be more natural, its implementation is quite challenging and it is much more difficult to anticipate the actual course of negotiations. Unfortunately, this unpredictability does not necessarily lead to more interesting negotiations. In contrast, the second approach can be implemented relatively easily with an existing discourse planner. Basically, such a planner must provide the decomposition of complex communicative acts into smaller units and it must be possible to distribute the decomposition results to different characters. We have adopted this approach for pragmatic reasons. The implementation of the planner is based on the Java^{TM}-based JAM agent architecture (Huber 1999). To model the knowledge used for script generation, we defined plan-operators that code a decomposition of a complex communicative goal into dialogue acts for the single agents.

Adopting the metaphor of a director who writes scripts for negotiation dialogues, the dialogue operators are formulated from a director's perspective rather than from a character-centric point of view. In particular, a decomposition of a goal may correspond to a certain dialogue pattern, such as a query-response pattern, or a comment-comment pattern. Factors such as a character's personality, affective state, and its embedding in the social context are used to formulate selection criteria for the operators. That is, we describe such factors by means of variables which can be accessed as preconditions of an operator.

A negotiation process will be initiated when a user posts the goal "negotiate-meeting-date" to the arena server. At the server, a plan operator will be applied and lead to a decomposition of this initial goal. The lowest level of this hierarchy comprises specifications of communicative acts which cannot be further decomposed by the available operators. Such leave notes may correspond to speech acts which need to be processed further by a natural language generator and a speech synthesis component. An act can also refer to non-verbal communicative expressions, such as pointing gestures, gaze, facial displays, body postures and movements. This lower level of leave notes manifests the interface between the negotiation simulator (discourse planner) and the deployed player technology. Our current demonstrator is able to generate scripts for characters that are animated with the MS Agents technology. Likewise, we can generate scripts for 3D talking heads using the Greta player that has been developed by Catherine Pelachaud's group at the University of Rome.

The screenshot shown in Figure 7 has been taken during the replay of a negotiation dialogue script which was received from the arena server where the actual simulation (script planning) was performed.

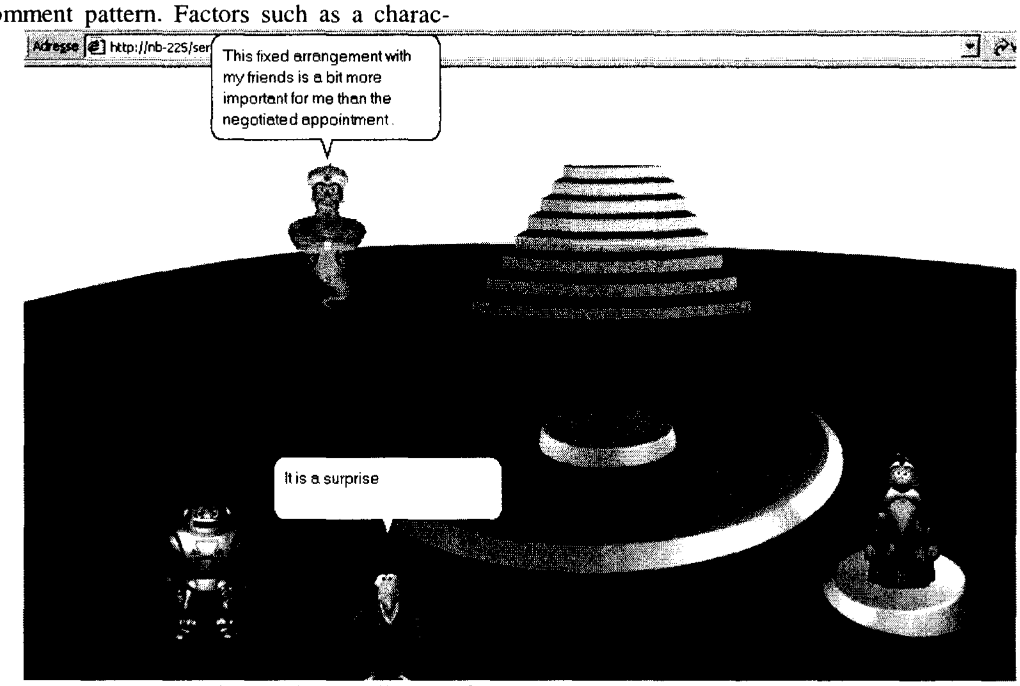


Fig. 7: Screenshot taken during the display of a generated negotiation dialogue. Avatar Peedy, who has a negative attitude towards Genie, is going to make a negative comment on Genies justification for his unwillingness to accept a certain meeting date proposal.

4 Summary and Open Issues

In this contribution we have presented some initial ideas towards an utilisation of socio-physiological concepts in the field of animated conversational characters. We are especially interested in how far such concepts can help to emulate more natural interactions and dialogues between our characters. We have adopted concepts from Heider's Balance Theory and Festinger's Theory of Cognitive Dissonance. When planning dialogue moves, our characters take into account social distances to their communication partners as well as beliefs about their own values (here importance values for meeting dates) and other character's values. Adopting the concept of balanced/ misbalanced cognitive configurations gives us an additional criteria for the planning of a character's conversational behaviour. However, we are aware of the fact that our current model is a strong simplification of the underlying theories. More recent sociological research, especially "network analysis" (Jansen 1999), suggests that much more complex modelling is required when striving for highly believable virtual personalities.

While the development of the Avatar Arena is work in progress, we already have a series of working demonstrators which differ in the granularity of the underlying models, the repertoires of negotiation and dialogue skills made available to the avatars, and also in the kind of player technology at the front end. Having a test environment that allows to compare the consequences of different models and negotiation strategies turned out to be advantageous for the exploration and validation of the deployed concepts, and last but not least, for the purpose of demonstrating the work for further discussion. We have chosen meeting date negotiation dialogues as an application domain. However, we do not claim that our demonstrators are contributions to the issue of finding a meeting date more efficiently. Rather, we are only interested in a study of the negotiation process and its variation depending on different character profiles and different social settings. We are aware of the fact that in real life, especially in a business context, people often have to hide their emotions in order to behave compliant with social norms. On the other hand, we do believe that simulation systems of this kind have high potential for educational applications which aim at training people in complex social interactions.

A major gap in our current version is the absence of affective models for the characters. However, this is on our agenda and we are confident that the overall believability of our negotiation dialogues will benefit from synergies that may emerge from an integrated model of social relations and affect.

Finally, besides some informal collecting of feedback on the demos from colleagues, no profound evaluation work, e.g. on the believability/naturalness of the generated dialogues has been carried out yet. One approach to evaluate believability would be similar to a Turing test in the sense that one would show protocols of negotiation dialogues to human subjects and have them decide on whether they believe that this was a real negotiation between human beings are a dialogue that has been generated by a computer program.

Acknowledgements

The work presented in this paper has been conducted in the EU funded project Magicster IST-1999-29078. We would like to thank our project partners for comments and suggestions on our approach.

References

E. André, T. Rist, S. van Mulken, M. Klesen, and S. Baldes, (2000). The Automated Design of Believable Dialogues for Animated Presentation Teams. In: Cassell, J., Sullivan, J., Prevost, S. and Churchill, E. (eds.): *Embodied Conversational Agents*, 220-255, Cambridge, MA: MIT Press. 2000.

P. Maes (1994). Agents that Reduce Work and Information Overload. Communications of the ACM, Vol. 37(7), pp. 31-40, 1994.

L. Festinger (1957). A Theory of Cognitive Dissonance. Stanford, CA: Stanford University Press. 1957.

F. Heider (1958). The Psychology of Interpersonal Relations. NY: Wiley. Chapter 7: Sentiment, pp.174-217. 1958

M. Huber (1999). JAM: A BDI-theoretic mobile agent architecture. Proc. of the Third Conference on *Autonomous Agents,* 236–243. New York: ACM Press. 1999.

D. Jansen (1999). Einführung in die Netzwerkanalyse. Grundlagen, Methoden, Anwendung. Opladen, 1999.

R.R. McCrae and O.P. John (1992). An introduction to the five-factor model and its applications. Special Issue: The five-factor model: Issues and applications. Journal of Personality 60: 175-215, 1992.

J.L. Moreno (1974). Die Grundlagen der Soziometrie. Wege zur Neuordnung der Gesellschaft. Opladen: Leske + Budrich. 1974.

A. Paiva (ed)(2000). Affect in Interactions: Towards a new generation of interfaces. Springer LNAI Series, Heidelberg: Springer. 2000.

H. Prendinger, H. and M. Ishizuka (2001): Social Role Awareness in Animated Agents. In: Proc. of the Fifth Conference on *Autonomous Agents,* 270–377. New York: ACM Press. 2001.

W. Wahlster, ed. (2000). Verbmobil: Foundations of Speech-to-Speech Translation. Springer, 2000.

EXCERCISES OF STYLE FOR VIRTUAL HUMANS

Zsófia Ruttkay
Center for Mathematics and Computer Science
POBox 94079, 1090 GB Amsterdam, The Netherlands
e-mail: Zsofia.Ruttkay@cwi.nl

Catherine Pelachaud
University 'La Sapienza'
via Buonarotti 12, 00185 Roma, Italy
e-mail:cath@dis.uniroma1.it

Abstract

The style of humans is a source of information about the person, her personality and enjoyment in every-day life. We wish to endow virtual characters with style of nonverbal communication. We outline the diversity in gesturing style among humans, and identify the decisive factors. We propose a computational model for defining the styled virtual characters and generate gestures accordingly. The ideas are to be demonstrated on styled facial expressions.

1 Introduction

1.1 Motivations

The title is inspired by Raymond Queneau's famous work, 'Exercises in Style' [21]. In this ingenious literary work, the French author takes a few-line banal story happening in a crowded bus, and in 99 exercises, tells it in different styles. He does it so well, that the reader can see the character acting for his eyes: how he gestures, whether he has a happy face or one of a bitter grumbler.
Another example of the power of style is 'Creature comforts', an Oscar-winning animation film [1], in which animals talk and gesture in the well-recognizable style of some human groups (of certain nationality, social status).
Thus the style is a source of information on the speaker, as well as of variety when communicating with real people. Moreover, a pioneering empirical experiment has shown that such factors as the ethnicity and the personality (introvert/extravert) of a synthetic character – even if manifested in a simple, static feature – do have consequences on their effect on the user [12].
One would like to benefit from style also when confronted with embodied conversational agents (ECAs) [3]. Even if one should not expect a virtual human to act like a blood and flesh real person, the current situation of rather puppet-like, styleless virtual characters should be improved, even if step by step.

1.2 Related work

The first steps have been taken in the direction of expressive ECAs, by endowing them with the capability of showing emotions [8]. Subtle issues like the impact of social role [5, 20] and personality [2] has been addressed. Also non-verbal signals have been used to accompany speech to make an ECA more expressive and believable. However, these works concentrate on modeling the psychological, social and communicative aspects of emotional and cognitive state. Usually the presentational issues are not dealt with as a research topic, but as a practical task for an animator, often only to make a specific application or demonstrator.
The importance of non-repetitiveness has been convincingly pointed out by Perlin [17], who (and several of his followers) used random noise to generate different instances of face and body motion, also in 'idle' state of the character. Badler and his colleagues have developed EMOTE [4], a computational framework inspired by Laban dance annotation, to modify expressiveness of hand, body and face gestures of ECAs. In the computer animation world [24] it has been recognized how important it is to 'add style' to (captured or synthesized) motion. Recently, there have been initiatives to develop XML-base markup languages to encode some of the 'human' aspects of multi-modal communication [25].

1.3 Our objectives

We are interested in making ECAs which are expressive and individual in their presentation and display. Particularly, we wish to endow ECAs with a **style in their non-verbal communicative behaviors**. A straightforward example for the possibility of using different styles is the communicative act of greeting. Factors like nationality, gender, age, physical state, mood, the unique motion habits and characteristics of the person all contribute to decide if the greeting will

be verbal and/or non-verbal, what facial expression and/or which hand be used and in what way to express the greeting.
We would like to **identify aspects and parameters of style** in human to human interaction, and rely on these findings to build styled ECAs. The presentation aspects can be considered as basically static parameters to define and tune the non-verbal presentational characteristics of an ECA: what type of gestures occur in his repertoire, how he chooses a behavior from several alternatives for a specific communicative function, in what way are the gestures presented.
We are developing a generic **computational model of gesturing style**, which makes it possible to define quickly and easily variants of an ECA animation with different, some group-specific and/or personal style. Our work is at the borderline between the 'mind' and 'body' aspects of ECAs. We are interested in finding out what and how a model of 'style' can be included.
In this article we report on ongoing work towards the above objectives. In section 2, we show how style is manifested in non-verbal communication and what are the decisive factors in using a specific style. In the following section we outline our computational model of styled ECAs. Finally, we discuss the status of our work, and outline further research tasks.

2 Gesturing styles and parameters

From the multitude of manifestations of style, we restrict our focus on style of non-verbal communication, namely facial, hand and body gestures, which accompany (sometimes replace) speech. A taxonomy of gestures and overview of research on human gesturing can be found in [9, 11]. Among the different gesture classes some iconic and emblematic gesture are typical of a **nation or ethnic group**. E.g. the two-fingers V sign for 'victory' should be made with palm outwards in England, while it can be used with both hand positions in the USA [3]. The very same gesture may convey different meaning. E.g. nodding is affirmative in most countries, but not in Bulgary and Greece. There are gestures and gesturing habits typical for **professions**, **cultural** and **social** status. People of different **sex** and **age** are likely to gesture differently.
But there is also much **individuality** in gesturing, considering the expressiveness of a gesture, the meaning and frequency of occurrence of a gesture. Humans keep moving (stretching chin or nose, rubbing hands, changing posture) even when they are 'idle': the type and frequency of 'idle motions' is often characteristic of a person and her personality [10]. While the way of gesturing is typical of a person, it is not entirely predictive and repetitive. On the one hand, a person often has several gestures for a certain communicative meaning. On the other hand, a single gesture is used with slight variations, avoiding the exact repetition of a gesture, which is often not the case with synthetic characters, making their motion looks 'unreal'.
When designing the gesturing style of an ECA, the following aspects have to be taken care of:

1. We need to define the gestures which constitute the **gesture repertoire** of the ECA, and to specify the meaning of the gestures by a **gesture dictionary**.
2. A mechanism should be given to **plan the presentation of gestures** expressing certain prescribed communicative functions: which modalities to use, and which gestures.
3. The chosen **gestures** should be tuned to achieve desired expressive characteristic.

2.1 Gesture dictionary and repertoire

To begin with, we settle our terminology. A **gesture** is some motion involving one or more of the modalities face, head, hands and body. We do not bother about the nature of the gestures, we assume that they are motions either used to express some meaning, or to fulfill biological necessities (breathing, blinking). **Basic gestures** refer to a single facial feature (eyes, eyebrows, mouth) or a single other modality (right/left arm, hand, leg and foot, ...). Gestures may be defined by composing 'basic' gestures. For example, a greet gesture may correspond to a 'head nod' and a 'smile'. Examples of basic gestures are:
eye_gesture: look_straight, look_left, look_right, ...
eyebrow_gesture: eyebrow_up, eyebrow_frown, ...
mouth_gesture: corner_mouth_up, mouth_open
head_movement_gesture: head_nod, head_shake, ...
head_direction_gesture: turn_left, turn_right, ...
basic_gesture: eye_gesture; eyebrow_gesture; ...
The compound gestures are formed from basic gestures by two operators:
& parallel composition
\+ sequential concatenation
gesture: *basic_gesure, basic_gesture + gesture, basic_gesture & gesture.*
The & operator is stronger than +. By parallel composition, the constituting gestures should start at the same time, by concatenation, sequentially, as soon as the previous one has finished. In the definition of a compound gesture, all basic gestures occurring in parallel composition should refer to different modalities or features. That is a gesture may not be composed of two basic 'eyebrow_gesture' with different values such as eyebrow_up and eyebrow_frown. While a gesture may be defined by a basic 'head_direction_gesture' (say 'turn_left') and a basic 'head_movement_gesture' (say 'nod'). The collection of all basic gestures which an ECA can use is the **gesture repertoire**.
When instantiated, the basic gesture corresponds to an animation of the involved facial features or body parts. A gesture may be instantiated in two ways which vary

in complexity. In the simplest case, instantiation of a gesture corresponds to the instantiation of the basic gestures that composed it. In a more sophisticated framework, a gesture is defined as a set of basic gestures linked by constraints [22]. Some characteristics of the gestures, like duration, intensity, on/offset time (for facial gestures), preparation/hold/withdrawal time (for hand gestures) can be specified. This framework enables the modification of gesture definition to produce variants of it with different **expressiveness**. Such a definition of gesture increases the flexibility of the creation process for animation as well as it allows the non-repetition of the final animation for each gesture instantiation [22].
Gestures may have a **communicative function** [9, 11, 18]. We use the taxonomy introduced by Poggi et al. [19] to gather communicative functions in four main groups based on the type of information they contain. Communicative functions provide information about location and properties of objects, concepts or events, about the speaker's affective (pleased, angry) and metacognitive state (thinking, listening), believes (agree / disagree, certainty) and intentions (performative, turn-allocation). Communicative functions are identified by a meaning and a signal. It is an essential characteristic of an ECA what gestures it uses to express a communicative function. There can be several alternative gestures expressing the same meaning (see Figure 1).

Commu. fct. Meaning Gesture

affective 'surprise' — eyebrows_up&mouth_open / eyebrows_up

intention 'emphasize' — eyebrows_up & head_nod / head_nod / beat / beat, head_up

Figure 1. For all communicative functions, there exist a one to many mapping between meaning and signal.

For a communicative act, the mapping between meaning and signal can be the one generally used by an ethnical or professional group, or can be idiosyncratic to the person. In each case, one has to know the person, her culture, her profession and so on, to decode the communicative function of the gesture (e.g. special way of greeting). Besides the mapping of meaning to gestures, there is difference in choosing a particular gesture from the possibilities allowed by the one-to many mapping as the selection is also characteristic of the person. E.g. right-handed people will use right-hand beats rather than left-hand beats, an 'introvert' person uses gestures less often, an Italian person uses beats more often than an Englishman to express emphasis.
We model the characteristics of usage of gestures marking the same communicative act by assigning probabilities to the individual gestures. Taking this characteristic into account too, the gestures used for a communicative act are given by a **gesture dictionary** entry of the following form:

$$communicative\_act: (Gesture_1, P_1), \ldots (Gesture_n, P_n)$$

where $Gesture_1, \ldots, Gesture_n$ are gestures, covering the alternatives of expressing the communicative function, and $P_1, \ldots P_n$ are probabilities of using the specific gesture in this role.
Different gesture dictionaries can be defined and given to set different 'gesturing code': one for gesturing habits of a culture (Italian gesturing) or profession (teacher-like gesturing), one for special gesturing of a certain person. However, one can notice that the use of a gesture or the display of a particular facial expression corresponding to a communicative act may also depend on the conversational context. Ekman [7] defines 'display rules' to embed cultural differences in the display or not display of a facial expression. Poggi and her colleagues [19] argue that the not only the expression of a gesture is cultural dependent but that it also depends on the social context defined by the personalities and cognitive capacities of the conversants as well as on the personal and role relationship existing between them. We adopt such a view here. A gesture is finally chosen not only from the 'cultural' and 'personal' dictionaries but by also considering the social context.

2.2 Style parameters

We wish to specify a language which can express all factors contributing to the non-verbal style. We identify two kinds of parameters:

1. High-level agent defining parameters
2. Low-level gesture modifying parameters

The **high-level agent defining parameters** are used to determine the gesture dictionary to be used (see Table 1). These parameters specify individual gesturing characteristics, personality, the social, cultural and biological aspects of the agent. They are static characteristics and they may be specified only once when the agent is selected. The state characteristics define time-varying physical, mental and emotional aspects. The first two aspects may be set for a certain time interval, while emotional aspects may be triggered by an event, an action or people [14] and whose display may depend on the context [19].
The final gestures, that is the characteristics of the generated motions, can be defined by **low-level gesture modifying parameters**. On one hand, the gesturing parameters (per modality such as hand gesticulation) define characteristics of the gestures of that modality; that is they are defined globally and represent a general characteristic of the agent. On the other hand, the gesture-modifying parameters may refer to a single gesture, or to all gestures of a modality within a time interval; that is they have a local influence and correspond to a dynamic aspect of the agent gesturing. The parameters are to be given in an XML-compliant Markup Language, basically to be used to annotate text to be

spoken by the ECA. The markup language is described in the next section.

In general, communicative gestures accompany speech, and are synchronized with speech. The actual frequency and timing of some repetitive gestures (e.g. blink) can be generated on the basis of biological characteristics, while others (e.g. for idle motion) may be defined on the basis of personality characteristics (e.g. if the character is nervous, he will make more frequent idle motions) or by hand.

3 Generating styled gestures

The generation of styled gestures takes a text to be spoken by the character, annotated by two types of tags. One set of tags refers to the communicative functions that accompany the text to be spoken; the other set defines the style of the agent and it influences the selection and fine-tuning of the gesture to express the communicative gestures. Both set of tags are represented by an XML-compliant Markup Language.

As noted before, the tags related to the communicative functions may be generated automatically or placed manually. We are using the APML markup language [5], developed in the framework of the European project, MagiCster[1]. The APML tags are generated automatically by the multi-modal dialog generator module of MagiCster. We are currently developing a Style Mark-up Language (SML) for the tags related to style. For the moment these tags are specified manually. Note that formally the only requirement for the APML and SML is that the gesture dictionary assigns some gesture choice to each communication functions known to the APML. The styled gestures, corresponding to an APML tag, are generated by the by the following steps:

1. Dictionary update.
2. Selection of gesture to be used.
3. Generation of expressive gesture instance.

In the **dictionary update** stage, it is checked which SML parameters have been given that have effect on the dictionary, and the current dictionary is updated. The effect of the static character-defining parameters is decided once, at the beginning of the processing of the marked up text. The probabilities in the resulting dictionary may be further modified at certain points of the text, by dynamical parameters, which have influence on style only for a subset of the text.

The generation of the current dictionary may be based on simple transparent principles. The simplest scenario is, when a hierarchy of parameters is established, and the parameters modify the current lexicon in a sequential order of hierarchy. E.g. a personal lexicon extends/overwrites cultural lexicon. The effect of state parameters can be computed locally, independent of other parameters. E.g. if the character is sad, his gesturing will be less frequent and intense. Cultural-dependent gesture will be set up. Expressiveness of gesture will depend on factors such as personality, age, emotion. Gesture type will be selected considering social, cultural and profession aspects. We are still working on the mapping of high-level tags to the choice of gesture type and its expressiveness. We will also have to face the problem of conflict that may arise due to the setting of different tags. For example, what would be the type of gestures and their expressiveness for an old person whose personality is introvert and that has the tendency to dream awake but that he currently excited due to a good news that just happened? By more complex reasoning it possible to take into account factors like situation-dependent or personality-dependent hierarchy of gesturing codes. E.g. a polite person may confirm to gesturing code expected in a situation (formal English dinner), but in the case of an individual person the personal gesturing code may be dominant in all circumstances. Also, it will depend on the culture if sadness will result in decrease or just, increase of gesturing (according to different mourning habits of cultures).

In the **gesture selection stage**, for each prescribed communicative function, a gesture will be selected to convey it. The selection will be based on the applicable entry of the current dictionary, but also taking into account possible modality conflicts [15]. In case of conflict, the selection from the alternatives in the current dictionary entry should be restricted to gestures with use free modalities (i.e. modalities not in used right now) [23]. If there is not such gestures, the conflict could be handled by resource-allocation [15]. An example of this type of conflict is the agent shaking the head to mean 'no' (gesture chosen from the cultural setting), but who also wants to emphasized her say. As head movement is already in used, choosing 'head nod' signal to denote 'emphasis' gives rise to a conflict. If it is possible to express 'emphasis' through another entry from the agent's gesture dictionary (say an eyebrow raising), the conflict does not subsist any longer. But if this is not possible, one of the two movements will have to prevail over the other ones.

[1] IST project IST-1999-29078, partners: University of Edinburgh, Division of Informatics; DFKI, Intelligent User Interfaces Department; Swedish Institute of Computer Science; University of Bari, Dipartimento di Informatica; University of Rome, Dipartimento di Informatica e Sistemistica; AvartarME

Tag	attributes	Possible values	M	P	G	E
Physical characteristic	Body	Body model file				
	Gesture repertoire	Directory of gestures			X	
	Dictionary	Dictionary file	X	X		
	Motion_manner	Smooth, hectic, angular…				X
	Handedness	Right, left	X	X		
	Hand_gesticulation	Intense, normal , minimal		X		X
	Face_gesticulation	Intense, normal, minimal		X		X
	Eye_gaze	Fixing, saccade, avoiding, neutral		X		
Personality_traits	Personality	Extravert, introvert, assertive, insecure		X		X
Social_aspects	Culture	Italian, British, American, …	X	X	X	
	Profession	Teacher, salesman,	X	X	X	
Biological_aspects	Age	Old, middle, young, child			X	X
	Gender	Female, male		X	X	X
State	Physical	Tired, sleepy, vivid, relaxed…		X		X
	Mental	Thinking, dreaming …		X		X
	Emotional	Angry, happy, …		X		X

Table 1. Character-defining parameters. M=mapping of communicative functions to gestures, P=probabilities of gestures assigned to a single communicative function, G=gesture repertoire, E=expressiveness of gestures

tag	attributes	Possible values
Gesturing	Intensity	Minimal, normal, intense
	Velocity	Fast, slow, normal
	Manner	Smooth, hectic, angular…
Timing	Gesture	Preparation phase, stroke, hold, end-phrasing
	Face	Onset, apex, offset
	Gaze	Start, end
	Start	Absolute time

Table 2. Gesture defining parameters. They decide the expressiveness of the gesture to be generated.

In the **expressive gesture generation stage**, all the static and dynamic style parameters having an effect on expressiveness of gesturing will be taken into account. E.g. an introvert person will make less articulated gestures, while a typical asymmetric eyebrow-usage will have an effect on all facial signals involving eyebrows. The effect of high-level character-defining parameters on expressiveness is given in terms of low-level gesture defining ones.

The on-the-fly generation of individual expressive gestures is based on our earlier work on the **gesture repertoire principle**: a gesture is defined in terms of constraints on control points of the parameter functions [23]. When style parameters are to be applied to the 'standard' definition of a gesture, they are expressed in terms of modifying certain constraints, e.g. the intensity of the smile will be limited as well as the speed of application and decay. The constraint framework allows the generation of different instances of a gesture, including random variants, as different solutions for the same constraint satisfaction problem.

It may happen that some modalities must be used for different gestures at the same time (like in case of speech and smile). For these modalities, a blend of the contributing gestures is to be produced, either as weighted sum of the contributing gesture parameter functions, or in a more sophisticated way, taking into account the constraints that should hold for the gestures.

4 Current status of research and further work

Up to recently, we have been working on generating facial expressions. We developed a detailed 3D realistic head model [16] and an interactive editor to specify facial expressions [13], also in terms of constraints. Using these tools and a taxonomy and descriptive data of facial expressions, we are currently implementing the system for styled facial expressions. By the time of the workshop, we will have a demonstrator showing how different facial gesturing styles can be applied to a single talking head. We will also experiment with the effect of non-determinism within a style. We wish to test if the style of facial expressions of a synthetic face is perceived as intended.

There are several issues which need to be clarified before starting to implement a more full-fledged system. Fist of all, culturally different dictionaries need to be defined from psychological and sociological studies to find out what the distinctive gesture types and manners are for certain groups. It is also a question if the gesturing of cultures could be defined in terms of lower-level parameters, like every-day and social values, living conditions. In our outlined framework, it was only the style which had effect on the choice of the gestures. However, the characteristics of the environment (noise, visibility) and the listener should have an effect on the selection of modalities and expressiveness of gestures.
Ultimately, one would like to have an ECA which manifests style also in the verbal modality. There is ongoing work to generate styled NL content [2, 6, 26], and to reflect emotional and mental state in the generation of synthetic speech [25]. On a longer term, it is a challenging task to develop ECAs which have consistent style in all their modalities.

Acknowledgements

We thank Han Noot and Isabella Poggi for their inspiring remarks on the topic.

References

1. Aardman Studios, Creature Comforts, URL: http://atomfilms.shockwave.com/af/content/atom_221
2. Ball, G., Breese, J. Emotion and personality in a conversational agent, in 3, pp. 189-219.
3. Cassell J., Sullivan J., Prevost S., Churchill E., (eds.) Embodied Conversational Agents, MIT Press, Cambridge, MA, 2000.
4. Chi D., Costa M., Zhao L., Badler N., The EMOTE Model for Effort and Shape, Proc. of Siggraph 2000, pp. 173-182.
5. De Carolis N., Pelachaud C., Poggi I., de Rosis F., Behavior planning for a reflexive agent, Proc. of IJCAI'01, 2001.
6. De Rosis, F., Grasso, F. Affective text generation, In: Paiva, A. (ed): Affective Interactions, LNAI vol. 1814. Springer-Verlag, 2000.
7. Ekman, P., Facial Expressions. In Dalgleish T. and Power T. (eds.), The Handbook of Cognition and Emotion. John Wiley & Sons, Ltd, 301-320, 1999.
8. Gratch J., Marsella S., Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents, Proc. of AA 2001, pp. 278-285. 2001.
9. Kendon A., Human gesture, In Ingold T. and Gibson K. (eds): Tools, Language and Intelligence, Cambridge University Press, 1993.
10. Marsella S., Personal communication.
11. McNeill D., *Hand and Mind: What Gestures Reveal about Thought*. The University of Chicago Press, 1992.
12. Nass C., Isbister K., Lee E-J., Truth is beauty: Researching embodied conversational agents, in [4], pp. 374-402.
13. Noot H., Ruttkay Zs., CharToon 2.0 Manual, CWI Report INS-R0004, Amsterdam, 2000.
14. Ortony A., Clore G.L. and Collins A., *The cognitive structure of emotions*. Cambridge University Press, 1988.
15. Pelachaud, C. Poggi I., Subtleties of Facial E-xpressions in Embodied Agents, Journal of Visualization and Computer Animation, to appear.
16. Pelachaud C., Magno-Caldognetto E., Zmarich C., Cosi P., An Approach to an Italian Head, Proc. of Eurospeech'01, 2001.
17. Perlin K., Real time responsive animation with personality, IEEE Transactions on Visualization and Computer Graphics, Vol. 1. No. 1. 1995.
18. Poggi I., Mind Markers. In Mueller C. and Posner R.(eds): The Semantics and Pragmatics of Everyday Gestures. Berlin Verlag Arno Spitz, 2001.
19. Poggi I., Pelachaud C. and de Rosis F., Eye communication in a conversational 3D synthetic agent. *AI Communications, Special Issue on Behavior Planning for Life-Like Characters and Avatars*, 13, 3, 169-181, 2000.
20. Prendinger H., Ishizuka M., Social role awareness in animated agents, Proc. of AA 2001, pp. 270-277. 2001.
21. Queneau R., Wright B. (Translator): Exercises in Style, New Directions 1981.
22. Ruttkay Zs., Noot H., FESINC: Facial Expression Sculpturing with INterval Constraints, Proc. of AA'01 Ws on Multimodal Communication and Context in Embodied Agents, 2001.
23. Thòrisson K.R., Layered modular action control for communicative humanoids, Computer Animation'97, IEEE Computer Society Press, p. 134-143, 1997.
24. Thomas F., Johnston O., Disney animation: The illusion of life, Abbeville Press, New York,1981.
25. Virtual Human Markup Language (VHML) http://www.vhml.org
26. Walker, M., Cahn, J., Whittaker, S. Improvising linguistic style: Social and affective bases for agent personality, Proc. of Autonomous Agents 1997.

VirtAct: The Quest for Balance

Daniel Sobral, Ana Paiva
IST / INESC-ID; Rua Alves Redol, n 9, 1000-029 Lisboa, Portugal
{Ana.Paiva,Daniel.Sobral}@gaips.inesc.pt

Abstract

This paper proposes a framework to help the creation of interactive stories. Our main goal is to enable the creation of believable characters that are able to partially follow a script. The idea is to keep a balance between scripted activity and the character's autonomous behavior, so we can achieve believability while pursuing a coherent plot structure.

1 Introduction

We aim at creating believable characters that hold the spectator and lead to the suspension of disbelief. However, even extremely well designed characters may be unable to deliver a good plot (perhaps some funny moments, but rarely a coherent global experience) if there is no thread of action allowing the creation of an empathic relationship between the characters and the audience.

The plot should evolve to an emotional climax, or the experience (as a whole) may lose all its sense. To do that we propose a framework aiming at a balance between scripted plots and emergent narratives.

2 The VirtAct Framework

This general framework will allow us to create Dramas (any possible sequence of scenes) and have them performed by a set of characters (the cast). The stories created are always partially authored (externally defined by a Human). The cast will act out the drama controlled by a Director responsible for the plot evolution.

2.1 Concept

In Character based storytelling, autonomous characters are usually defined as having some goals they have to fulfil. These goals, and the way actors manage to achieve them, is usually what makes the story. But in Plot based approaches, one has a single goal: to deliver a satisfying experience. These two approaches are often incompatible.

So, to handle the problem of keeping the plot under control and enabling synchronization among actors, we adopted the approach proposed by Mateas and Stern (2000). By defining a smaller unit of story progression, the *Beat*, the Director has a finer control over the plot. A Beat can be seen as a Take, a story point where something relevant (to the overall emotional arc) happens, the atomic unit for plot development. A scene is a sequence of related beats.

The Director evaluates, at each beat, the story progression and should decide (according to some heuristic evaluation) what should be the next scene, one that leads the plot closer to the desired emotional climax.

2.2 Implementation

A prototype following this framework was implemented using a tool for the creation of 3D Virtual Worlds: WildTangent (see Figure 1 for a snapshot of the system).

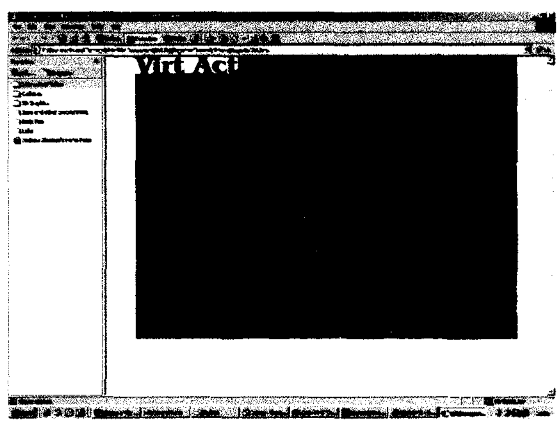


Figure 1: Snapshot of the System

The system works as follows: in a first step (Figure 2), a script is read from an external file. The script is then passed to the Director, who controls the starting of the show.

In a second step, the Director asks the WildTangent system for the necessary scene resources (set, actor's bodies, objects) and sends the appropriate scene script and related resources to every actor.

Finally, in a third step, the actors start executing behaviors that perform geometry requests to the WildTangent system (animation playing, sound playing, object positioning). All requests will be seen in the Web Browser with which the User interacts.

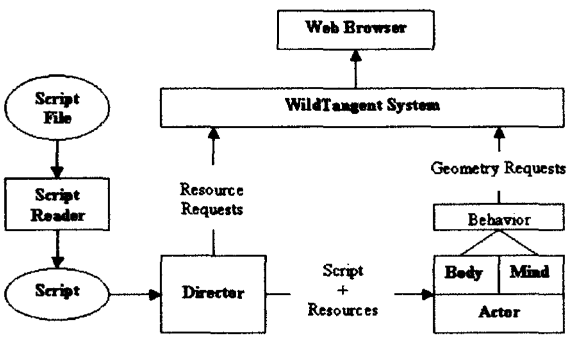


Figure 2: Application Flow

2.2.1 Behavioral Intelligence

Using Beats, the characters only have to worry about the current take they are performing. Furthermore, at each beat, the character's goals are usually very simple, and a rule-based system (represented by behavior patterns active at each beat) is enough to fulfil them.

Behaviors are therefore critical to maintain the Character's believability, and need to be created specifically. The actor's mind can be seen as a repository of transforming information, useful to behavior execution. Each behavior uses the strategy it needs: while simple behaviors can just run some scripted animations, a complex behavior like the WalkTo behavior includes path planning.

2.2.2 Liveliness by Default

The characters are always executing some behaviors. Even if they are not doing any beat behavior (any behavior relevant to the plot), they still have to "be themselves". We achieve this by creating what we named *Default behaviors*. These are behaviors designed to be always executing, while anything else is running. Some characters may even have no active part in the plot, but their default behavior allows them to keep looking "alive".

3 Results

In table 1 we can see a fragment of a first experimental script used for a short story. We can see two beats of the script:

In the first beat, Alien Emerges, sending a message to the other 2 actors. SciScared activates a behavior pattern with 2 behaviors: Be Scared and GoTo House. SciBrave is not expecting the Emerge Behavior and only runs his Default Behavior. Finally, after this beat is finished (when

Beats	Alien	SciScared	SciBrave
1	{{}, E }	{{E}, BS GTH}	
2			{{}, GTH}

Table 1: Fragment of a scene script

SciScared gets to the House) a second beat starts where SciBrave Walks to the House.

In this simple example we can already see what can be understood as being *in character*. Although they both run the same WalkTo House, SciScared is scared and runs (fast) into the house but SciBrave just walks (with normal pace). In this small example we have also used a Camera Actor, whose Default Behavior was to follow the Actor that had executed the latest beat behavior. Albeit being very simple, we have already observed some interesting results in this experience: the Camera always focused the most relevant aspects of the plot.

We observed that, using beats, only appropriated behaviors were activated. The beats allowed grouping of related behavior patterns permitting a coherent (at least, from the script author's perspective) behavior sequence.

4 Conclusions and Future Work

The major accomplishment this work is trying to achieve is to provide a balance between scripted animation and autonomous behavior, reducing the tedious job of hand-crafting animation but simultaneously keeping our characters under control.

We have only implemented scripts with a single scene. We are now working towards implementing the Script as a set of graphs, which will allow for a greater freedom in Plot development. Our work in the immediate future will also include the improvement of current characters and their behaviors.

Interactivity has not been tried yet, but we expect to achieve some results in the future using the suggested methodology.

References

Michael Mateas and Andrew Stern. Integrating plot and character in interactive drama. *Socially Intelligent Agents: The Human in the Loop, AAAI symposium*, November 2000.

Making Faces with Action Unit Morph Targets: 3-Dimensional Parameterized Models of Facial Expression

Jesse Spencer-Smith; Åse H. Innes-Ker ; Heather Wild ; James Townsend
Beckman Institute, University of Illinois Urbana-Champaign
Indiana University
jbspence@beckman.uiuc.edu; akvistin@indiana.edu; hwild@uiuc.edu; jtownsen@indiana.edu

Abstract

The purpose of this study is to create and validate parameterized models for generated life-like facial expressions in a mesh-based animated character. In this study two experimental tasks are used to validate these stimuli. The first task is a simple ratings task where participants rate computer-generated faces and photographs on perceived emotional intensity. In the second task, the faces were rated according to how accurately they convey the intended emotion. Ratings of emotional intensity and accuracy were overall lower for the female animated model. For the male animated model, overall ratings of intensity and accuracy were not significantly different. More confusions between emotions were evident for the synthetic stimuli than for the photographs.

1 Introduction

Much of our understanding of human expressions has come from research using high–intensity, unambiguous emotional expressions presented via still photographs. Research over the past four decades suggests that six basic emotional expressions — anger, disgust, fear, happiness, sadness, and surprise — are recognized reliably (e.g. Ekman, Friesen, & Ellsworth, 1982; Russell, 1994), universally (e.g. Ekman, 1971, but see Russell, 1994), and rapidly (Kirouac, & Dore, 1984; McAndrew, 1986). However, the perceptual and decisional processes underlying these findings are poorly understood. Also, work with low-intensity expressions has not been as prevalent (Hess, Blairy, & Kleck, 1997), although our everyday experience with subtle expressions suggests we are similarly adept at reading subtle emotions. Likewise, the dynamics of expressions, although widely recognized as important to our understanding of emotional expression, has received relatively limited attention (Niedenthal, Halberstadt, Margolin, & Innes-Ker, 2000; Spencer-Smith, Innes-Ker, & Townsend, 2000, November; Wehrle, Kaiser, Schmidt, & Scherer, 2000). Research in these all of these areas has historically suffered from a lack of stimuli. Here, we summarize work by Spencer-Smith et. al. (2001), and discuss possible applications.

The purpose of this study is to create a 3-dimensional, flexible, parameterized, ecologically-motivated model of facial expressions. We seek to incorporate the ability to blend movements and expressions, to produce meaningful low-intensity emotions, to allow for the creation of dynamic stimuli, and to validate the model against standard static images of expressions.

1.1 Previous Approaches

One existing approach to produce 3-dimensional expressions involves 3-dimensional capture of human expressions using highly reflective markers on key points on an actor's face, such as tip of the nose, midpoints of the lips, etc. The markers reflect back light that can be detected by cameras. Each position is then sent to a computer-generated model of the face (Guenter, Grimm, Wood, Malvar, & Pighin, 1998). Although such a technique is simple to implement and quick in calculation, it lacks fine definition. The method also does not allow for the creation of novel expressions not posed by the actor.

Another approach taken is the modeling of each facial muscle onto a model human skull. For each expression, appropriate muscle shape can be altered in the computer environment to simulate the real muscle being flexed. Although this method is physiologically sound, it is inflexible in that it is difficult to model multiple identities. New faces have to be constructed from the skull, muscles, then up to skin. The software necessary to run such models require significant overhead.

One of the most recently developed methods is the photogrammetric technique where several simultaneous shots are taken of an actor from different angles. A 3-dimensional image is calculated using the flat images. Three-dimensional morphing animates expressions (Pighin, Szeliski, & Salesin, 1999). Although this morphing technique is used most often in the facial animation industry, it still requires trained actors. And, like the earlier method using actors, there is no systematic way to generate novel images, or to exercise fine control over the intensity of expressions.

The basis for our model is a 3-dimensional mesh-based character animation package which allows for a great deal of control in conducting research in the dynamic nature of expression recognition without sacrificing ecological validity. The stimuli would also be invaluable for probing the perception of static emotional expressions since they easily allow for the necessary fine-grained control over stimuli.

Because it is of utmost importance that the stimuli reflect as closely as possible the wide range of expressiveness of humans, we opted for an approach for creating the stimuli in which expressions are built, bottom up, from the identified muscular elements of facial expressiveness. The approach is similar to that of Wehrle, Kaiser, Schmidt, & Scherer (2000). However, we created full, realistic 3-dimensional images whereas Wehrle, et al. used 2-dimensional line drawings.

As a reference for the stimulus creation, we relied on Ekman's (1978) Facial Action Coding System (FACS). The express purpose of FACS was to create a coding system for precisely cataloging facial movements in naturally occurring faces without having to resort to imprecise or emotion-related terminology. The basic element of the system is the action unit (AU). An action unit is defined as the surface (skin) structure resulting from the action of a facial muscle group that can be controlled independently from all other facial muscle groups. Expressions can be coded by detecting the presence of combinations of AU's in the face. The FACS training manual provides photographs, movie clips, precise written descriptions, and a description of musculature implicated in each AU.

In the present study, we implemented 16 AU's on simulated female and male faces using a 3-dimensional mesh-based modeling program, Poser 4 (Curious Labs, Inc., Santa Cruz, California). Poser provides realistic default female and male characters defined by polygons (3470 vertices for the female character head, 3427 for the male). Using the software, one can render images of full-body characters from arbitrary views, wrap new skin textures around faces, and arbitrarily define lighting. New characters can be created by distorting default characters by means of morph targets; morph targets describe 3-dimensional deformations to the default geometry. One morph target, for example, alters the length of the face of the character by moving down the vertices of several hundred polygons defining the face. New morph targets can be created by manually moving vertices using modeling tools provided in the program. The distortions are applied to the face by means of a parameter setting, one for each morph target. For example, at a magnitude setting of 1.0, a user-defined morph target will recreate the original distortion defined by the user; at 0.5, the deformation is less severe, and is closer to the default face. At a setting of 2.0, a morph target exaggerates the deformation. Using Ekman's (1978) FACS as a guide, our group implemented 16 of the 44 AU's in the FACS training manual as morph targets.

1.2 Research and Real-World Applications

With each AU implemented as independently controllable morph targets, it is possible to create novel expressions, to create low-magnitude expressions, and to generate dynamic displays of expressions. Multiple identities can be created using other morph targets to alter facial structure, while allowing identical expressions to be created on each character by setting equivalent AU morph target parameter settings. Lighting and viewpoint can also be set arbitrarily. Expression changes can be rendered in real time, and interactively.

The AU morph targets are flexible: they can be applied alone or in groups to create expressions, and can be used to generate novel expressions. When used in generating an expression, we found that we could manipulate perceived intensity of the expression by adjusting the magnitudes of the component AU's. By basing our AU morph targets on well-researched observations of facial movements, the morph targets are ecologically based. Most importantly, the AU morph targets allow for the creation of dynamic expression stimuli in a controlled manner.

The successful manipulation of intensity will allow investigations into low-intensity expressions which have received scant attention. Spencer-Smith, Innes-Ker and Townsend (2000, Novermber), using the AU morph targets developed in this project, have demonstrated the critical role of dynamics in perceiving low-intensity emotional expressions. Very low emotional intensity expressions were accurately identified when presented in a dynamic fashion, while accuracy with static displays was near chance.

The morph targets created in this study have been made publicly available over the web at http://www.staff.uiuc.edu/~jbspence/audown.html. Using Poser 4 software and our morph targets, researchers can create highly controlled stimuli for use in research on expressions.

1.3 Evaluation of Action Unit Morph Targets

To validate our morph target AUs, we generated versions of the six basic expressions -- anger, disgust, fear, happiness, sadness, and surprise -- and compared the resultant images to Ekman and Friesen's (1976) pictures of facial affect in three studies. In the first study, participants rated the emotional intensity and accuracy of expressions of synthesized female characters and of the photographs. To test our ability to control perceived emotional intensity, we included expressions created using both high-magnitude and low-magnitude AU morph targets. In the second study, we studied synthesized male characters using the same procedure.

2 First Experiment

2.1 Method

2.1.1 Participants
Twenty participants (seven males and 13 females) completed Experiment 1. All participants were either undergraduate or graduate students at Indiana University and all were paid for their participation.

2.1.2 Stimuli
The stimuli consisted of 119 grayscale images of faces. Seventy-seven images were photographs of faces taken from Ekman and Friesen's Pictures of Facial Affect (1976). There were photographs of eleven individuals (five males and six females) with seven expressions each. Forty-two images were generated using Poser. There were six synthesized Poser characters, female in appearance. Each character was rendered with seven different emotional expressions (see Figure 1)--anger, disgust, fear, happiness, sadness, surprise, and neutral--using the AU's listed in Table 1 as a guide. Two levels of AU morph target magnitude were used in rendering the computer-generated images: half were designed with magnitude settings of approximately 0.5, the other half with magnitude settings close to 1.0. The Poser figures were rendered in grayscale with viewpoint, camera focal length, and lighting to match the Ekman and Friesen photographs as closely as possible. The images and photographs measured 320x480 pixels.

Table 1.
Component Action Units for Expressions (Parke & Waters, 1996)

Surprise	1+2+26
Fear	1+2+4+15/16+20
Disgust	4+9+10+17
Anger	4+5+10+24/25
Happiness	6+11+12
Sadness	1+4+15

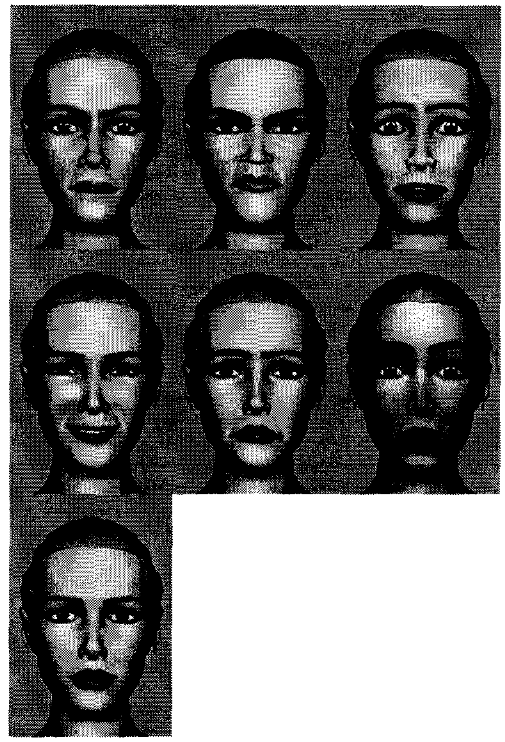


Figure 1. Female characters with expressions of anger, disgust, fear, happiness, sadness, surprise and neutrality.

2.1.3 Procedure
There were two parts to this experiment: an intensity rating task and an accuracy rating task. Participants were read a short set of instructions explaining that the goal of the experiment was to test how people perceive emotional expressions in faces. They were asked to answer quickly and spontaneously and to use the entire scale.

For the intensity rating task, each participant saw all 119 images (77 photographs and 42 computer-generated images). Images were presented in random order. Participants rated each image for each of the seven emotions (anger, disgust, fear, happiness, sadness, surprise, neutral). Question order was randomized across participants.

The participants viewed faces one at a time with unlimited viewing time, and rated how intensely each emotion was expressed by the face using a seven-point Likert scale in which 1 equalled "no intensity" and 7 equalled "high intensity".

2.2 Results and Discussion

Participants' responses for Part A of Experiment 1 are summarized in Table 2. Responses to each of the 11 actors expressing anger, disgust, fear, happiness, sadness, and surprise are averaged across the 20 participants in the top half of the table. Responses to each of the 3 Poser characters with the same expressions composed of high-magnitude AU's are summarized in the lower half. It is evident that, like the photographs, the Poser images convey multiple expressions. In expressions intended to convey happiness, for example, participants rated surprise as the second strongest emotion in both the photographs and Poser images. Overall, combining AU morph targets successfully communicated the intended emotional content. With two exceptions, the emotion rated as the most intense was the intended emotion. In the first exception, anger was rated as the most intense emotion in an expression intended to convey disgust. The distinguishing AU for disgust is AU 9, which involves a wrinkling at the base of the nose. Low polygon counts in the nasal region made clear depiction of this wrinkling difficult. We note that in the photographs, the second highest rated emotion for the disgust expression was anger. The second exception occurred with the expression intended to convey fear. In this expression, surprise was rated as the most in intense emotion. Surprise is a valence-neutral expression, meaning it is not necessarily a positive or negative expression. It is comprised of AU's 1 and 2, which are present in fear and in sadness. The high rating of surprise for this expression can be attributed to an overly high setting of these AU morph targets.

Table 2: Intensity Ratings for Ekman and Friesen (1976) Photographs and High-Magnitude Poser Female Character Expressions

Photograph	A	D	F	H	Sa	Su	N
Anger	**5.64**	4.07	2.43	1.26	2.28	2.44	1.55
Disgust	4.06	**6.12**	1.81	1.27	1.97	1.73	1.57
Fear	2.61	2.80	**5.38**	1.39	2.45	5.35	1.46
Happiness	1.19	1.10	1.12	**6.45**	1.17	2.37	1.80
Sadness	2.51	2.81	2.73	1.18	**5.49**	1.71	2.00
Surprise	1.77	2.01	3.30	2.16	1.74	**6.25**	1.65
Neutral	2.51	2.40	1.83	1.79	2.70	1.43	**4.79**
Poser							
Anger	**5.22**	4.10	3.12	1.20	3.57	2.13	1.58
Disgust	5.97	**4.70**	1.77	1.20	2.75	1.50	1.58
Fear	2.45	2.72	**4.90**	1.35	3.78	5.38	1.55
Happiness	1.27	1.22	1.33	**6.00**	1.27	2.38	2.03
Sadness	3.05	2.83	2.67	1.35	**6.00**	1.73	2.02
Surprise	1.40	1.45	2.73	3.15	1.75	**6.27**	1.75
Neutral	1.88	1.83	2.10	2.28	2.83	1.50	**5.20**

Average judged intensity varied significantly with image type (photograph, Poser with high-magnitude AU morph targets, Poser with low-magnitude morph targets), $F(2,18)=49.02$, $p<.0001$. Planned contrasts revealed that high-magnitude Poser images were rated as less intense than the photographs ($M=5.732$, $SE=0.086$), $F(1,19)=5.449$, $p<0.031$. High-magnitude Poser image ($M=5.469$, $SE=0.143$) were judged as more intense than low-magnitude Poser images ($M=4.543$, $SE=0.150$), $F(1,19)=85.0$, $p<0.0001$.

Although the intensities of the Poser high-magnitude images were judged as less intense overall than the photographs, two emotions, disgust and happiness, were primarily responsible for the difference. On the other hand, emotional intensity was successfully manipulated by reducing the magnitude of the component AU morph targets. Presumably, the magnitude of the AU morph targets comprising disgust and happiness could be raised to increase the perceived emotional intensity. This was attempted in the study of the male Poser characters in Experiment 2.

Average judged accuracy varied significantly with image type (photograph, Poser with high-magnitude AU morph targets, Poser with low-magnitude morph targets), $F(2,17)=17.83$, $p<.0001$. Planned contrasts revealed that high-magnitude Poser images ($M=5.664$, $SE=0.111$) were judged as less accurate than the photographs ($M=5.904$, $SE=0.122$), $F(1,18)=8.124$, $p<0.011$. The difference, however, was less than 0.3 on a scale of 1 to 7.

3 Second Experiment

3.1 Method

3.1.1 Participants

Twenty participants (seven males and 13 females) completed Experiment 2. All participants were either undergraduate or graduate students at Indiana University and all were paid for their participation. No participants had participated in Experiment 1.

3.1.2 Stimuli

The same actual photographs were used as in Experiment 1. The synthetic stimuli, however, were Poser male faces. Since the default geometry for the Poser female and male faces contain a different number of polygons, new AU morph targets had to be created for the Poser male faces. Examples of these stimuli appear in Figure 2.

3.1.3 Procedure

The procedure was identical to that in Experiment 1.

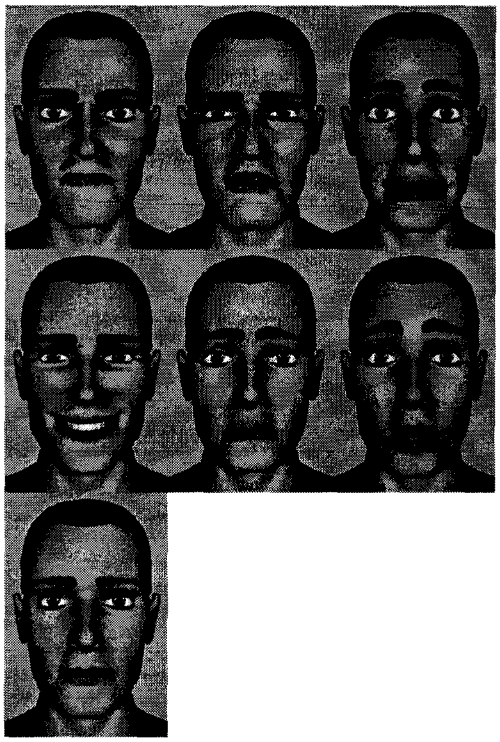


Figure 2. Sample Poser male characters with expressions of anger, disgust, fear, happiness, sadness and surprise and neutrality.

3.2 Results and discussion

Participants' responses for the intensity ratings portion of Experiment 2 are summarized in Table 3. With a single exception, the emotion rated as most intense was the intended emotion. As in Experiment 1, anger was rated as the most intense emotion in an expression intended to convey disgust. It should be noted that the confusion is also evident with the photographic stimuli. Without wrinkles to indicate movement of skin along the base of the nose, AU 9 remains difficult to capture with the current geometry. By increasing the magnitude of AU morph targets 1 and 2 (raising the eyebrows) in the surprise expression, and increasing AU 20 in the fear expression, we were able to ameliorate the confusion between fear and surprise found in Experiment 1. A similar ambiguity between fear and surprise can be seen in the photographs as well.

Average judged intensity varied significantly with image type (photograph, Poser with high-magnitude AU morph targets, Poser with low-magnitude morph targets), $F(2,18)=57.15$, $p<.0001$. Planned contrasts revealed that high-magnitude Poser images did not have a significantly different intensity from the photographs ($M=5.733$, $SE=0.106$), $F(1,19)=0.094$, $p>0.828$. High-magnitude Poser image ($M=5.705$, $SE=0.139$) were again judged as more intense than low-magnitude Poser images ($M=4.88$, $SE=0.150$), $F(1,19)=91.33$, $p<0.0001$.

Table 3: Intensity Ratings for Ekman and Friesen (1976) Photographs and High-Magnitude Poser Male Character Expressions

Photograph	A	D	F	H	Sa	Su	N
Anger	**5.45**	3.73	2.25	1.19	2.21	2.32	1.50
Disgust	3.90	**5.78**	1.81	1.23	2.12	1.91	1.39
Fear	2.07	2.63	**5.54**	1.22	2.36	5.40	1.30
Happiness	1.15	1.15	1.15	**6.40**	1.19	2.14	1.69
Sadness	2.29	2.40	2.42	1.15	**5.26**	1.46	2.20
Surprise	1.69	1.95	3.59	2.11	1.60	**6.10**	1.47
Neutral	2.06	2.02	1.51	1.75	2.66	1.36	**5.59**
Poser							
Anger	**6.37**	4.43	1.93	1.13	2.43	1.70	1.32
Disgust	5.22	**4.80**	1.92	1.15	2.67	1.67	1.42
Fear	1.67	2.13	**5.23**	1.40	2.88	5.07	1.68
Happiness	1.08	1.05	1.18	**6.40**	1.23	2.12	1.43
Sadness	1.62	1.88	2.35	1.07	**5.97**	1.53	2.30
Surprise	1.17	1.47	2.67	3.50	1.58	**5.53**	1.87
Neutral	1.47	1.47	1.37	3.27	2.18	1.32	**5.63**

Average judged accuracy varied significantly with image type (photograph, Poser with high-magnitude AU morph targets, Poser with low-magnitude morph targets), $F(2,19)=48.04$, $p<.0001$. Planned contrasts revealed that high-magnitude Poser images ($M=5.69$, $SE=0.109$) were not judged as less accurate than the photographs, as was found in Experiment 1 ($M=5.554$, $SE=0.131$), $F(1,20)=2.90$, $p=0.104$; the trend was for the Poser images being rated as more accurate than the photographs.

Acknowledgements

Christy Duffy, Chad Edwards, Kristina Ervin, Nicole Merritt, and Jae Won Paik collaborated with us in the creation of the morph targets. Funding for this project was provided by the Indiana University President's Summer Undergraduate Research Initiative.

References

Ekman, P. (1971). Universality and cultural differences in facial expressions. In J. K. Cole (Ed.), Nebraska Symposium on Motivation (Vol. 19, pp. 207-283). Lincoln: University of Nebraska Press.

Ekman, P. (1994). Strong evidence for universals in facial expressions: A reply to Russell's mistaken critique. Psychological Bulletin, 11(5), 268-287.

Ekman, P., Friesen, W.V. (1978). The Facial Action Coding System: A technique for the measurement of facial movement. Palo Alto: Consulting Psychologists Press.

Ekman, P. and Friesen, W. V. (1976). "Picture of facial affect." Palo Alto: Consulting Psychologist Press..

Ekman, P. Friesen, W.V., & Ellsworth, P. (1982). What are the relative contributions of facial behavior and contextual information to the judgment of emotion? In P. Ekman (Ed.), Emotion in the human face, 2nd ed., pp. 111-127. New York: Cambridge University Press.

Guenter, B., Grimm, C., Wood, D., Malvar, H., & Pighin, F. (1998). Making faces. SIGGGRAPH 98 Conference Proceedings, July, pp. 55-66.

Hess, U., Blairy, S. & Kleck, R. E. (1997). The intensity of emotional facial expression and decoding accuracy. Journal of Nonverbal Behavior, 21, 241-257.

Kirouac, G. & Dore, F.Y. (1984). Judgment of facial expressions of emotion as a function of exposure time. Perceptual and Motor Skills, 59, 147-150.

McAndrew, F.T. (1986). A cross-cultural study of recognition thresholds for facial expressions of emotion. Journal of Cross-Cultural Psychology, 17, 211-224.

Niedenthal, P. M., Halberstadt, J. B., Margolin, J., & Innes-Ker, A. H. (2000). Emotional state and the detection of change in facial expression of emotion. European Journal of Social Psychology, 30, 211-222.

Parke, F. I. & Waters, K. (1996). Computer Facial Animation. A. K. Peters: Wellesley, Massachusetts.

Pighin, F., Szeliski, R., & Salesin, D. (1999). Resynthesizing facial animation through 3D model-based tracking. ICCV Vision and Graphics, 1, 143-150.

Russell, J.A. (1994). Is there universal recognition of emotion from facial expression? A review of the cross-cultural studies. Psychological Bulletin, 115, 102-141.

Spencer-Smith, J., Innes-Ker, A. H., & Townsend, J. T. (2000). Dynamics of expression. Presentation at Object Perception and Memory 2000, New Orleans.

Spencer-Smith, J., Wild, H., Innes-Ker, Å., Townsend, J. T., Duffy, C., Edwards, C., Ervin, K., Paik, J. W. & Prause, N. (2001). Making faces: Creating 3-dimensional, ecologically-motivated poseable expressions. Behavior Research Methods, Instruments and Computers, 33 (2), 115-123.

Wehrle, T., Kaiser, S., Schmidt, S., & Scherer, K. R. (2000). Studying the dynamics of emotional expression using synthesized facial muscle movements. Journal of Personality and Social Psychology. 78, 105-119.

FROM VIRTUAL BODIES TO BELIEVABLE CHARACTERS

Reusable Synthetic Characters with Expressive Bodily Behaviour

Marco Vala, Ana Paiva and Mário Rui Gomes
IST / INESC-ID – Rua Alves Redol, nº 9 – 1000-029 Lisboa – Portugal
marco.vala@gaips.inesc.pt; ana.paiva@gaips.inesc.pt; mrg@inesc.pt

Abstract

This paper presents an approach for creating believable characters. We use pre-defined animations and body postures that can be combined in real-time to generate a rich set of behaviours. Moreover, parameters like speed or spatial amplitude can also be modified in real-time to influence the character's movement. Our ultimate goal is to create reusable synthetic characters that are able to express their inner-feelings using bodily expression.

1 Introduction

The last decade witnessed an impressive evolution of synthetic characters. The recent advances in technology, particularly the continuous increase of performance in computers, lead to broad use of synthetic characters in several areas. Computer simulations are a reality in many domains from warfare to rescue training. The use of synthetic characters allows safe exercises without putting anyone into harm. In education, synthetic characters have been used as tutors that are able to explain or guide a student through a task. Characters can also act as team-mates for individual training in tasks that need cooperation. In the cinema, the movie "Toy Story" opened the door to a completely new generation of movies where the actors are computer generated. Recent examples are "Shrek" and "Final Fantasy". Computer games bring us every day a full hand of new gaming experiences where synthetic characters assume the main role. As a brief example, look at "Tomb Raider" and the impact of a character like "Lara Croft", or the amusing "Guybrush Threepwood" from "The Escape from Monkey Island". Although purely fictional, these characters have a personality, likes and dislikes, friends, and many other things that drag us into the story and make us feel as a part of it. Computer games have also benefited with the development of accurate characters that simulate the reality, especially in sport simulations. Games like "FIFA" or "NBA Live" have models for most of the players and it is quite pleasant to recognize a star like "Figo" or "Kobe Bryant". Likewise, many synthetic actors start to participate in artistic performances or TV shows, and start to play an important role in these domains. The future platforms of interactive TV and TV on demand will offer a "workplace" for many virtual presenters, virtual entertainers and virtual advertisers. We are entering a new era where synthetic are entering a new era where synthetic characters need to be more human-like and, above all, more believable.

The body assumes a natural relevance in this new generation synthetic characters. Humans, for instance, express their emotions and inner feelings with facial expressions, but also with gestures and behaviours that affect the body movement. If we take a close look into the roots of animation, we quickly conclude that skilled animators use the body and the way it moves to denote personality, express emotions and appear to have a certain inner life, which is sufficient to induce believability. They are able to create characters that easily delude our eyes. Therefore, a synthetic character should be able to express its emotions through bodily animation using the appropriate gestures and behaviours.

The overall goal of our work is to provide a simple way of creating and controlling reusable bodies of synthetic characters that can express emotions. That is, to develop reusable synthetic characters with an expressive bodily behaviour.

2 Background

Traditionally, synthetic characters were created using a pure computer graphics approach in which the visual realism is the ultimate goal. Badler et al. (1993), Fua et al. (1998), Kalra et al. (1998), and Aubel and Thalman (2000) look at believability as a strict visual problem. The approach is based on detailed geometrical models and advanced animation techniques that are able to assure a good visual accuracy. However, the generated characters and movements are computational expensive, and the results are often unrealistic. Characters lack a certain inner-life that is essential to delude our eyes. Consequently, the idea that believability depends more on the characters' ability of conveying its inner-feelings lead to a different approach that relegates the visual realism to a second plane. Blumberg and Galyean (1995), Perlin and Goldberg (1996), Russell and Blumberg (1999), and Johnson et al. (1999) seek what can be called behavioural realism, and look at high-level architectures for real-time animation and interactive control. The generated characters are more expressive and more alive, but that richness is also very dependable on the animators' ability to maintain a certain behavioural coherency in the characters' library of pre-defined movements. Therefore, the characters can only express the feelings or the emotions that were previously modelled, and most of the animations cannot be reused in different characters.

In parallel, several researchers tried to find a way to change the movements in real-time to add expressiveness, personality or emotions. Some approaches propose secondary motions as a way to add naturalness to primary motions, like Perlin's work that uses periodic noise functions to add expressiveness (Perlin 1995). Others used signal analysis techniques to capture and create new motions: Unuma et al. introduce a model to describe and manipulate human periodic motions based on a Fourier Functional Model (Unuma et al. 1995); and Amaya et al. capture the difference between neutral and emotional motions, and allow the generation of new emotional motions (Amaya et al. 1996). More recently, results from movement observation have been used to define models that parameterise movements in real-time to achieve a rich set of variations: the EMOTE system proposed by Chi et Al. uses results from the Laban Movement Analysis to modify arm and torso animations into more expressive movements (Chi et al. 2000; Badler et al. 2000). However, all these approaches are still very limited and, most of the times, they are only applicable to certain body parts or to a single class of movements.

3 Our Approach

Our work is inspired mainly on Blumberg and Perlin's work, but uses some of the results of the other approaches, namely the Amaya's work on emotional transforms and how the speed and the spatial amplitude of the movement vary noticeably with different emotions.

The main idea is to use a set of pre-stored movements and modify them in real-time to reflect inner-feelings or emotions. These changes can result from the combination with specific body postures or secondary motions, or from variations of basic parameters like speed or spatial amplitude. For example, imagine a neutral walking movement that we want to modify in real-time to express sadness. The character can begin walking slowly and with smaller steps (and that denotes a variation in the speed and in the spatial amplitude of the movement), the torso could bend to the front (and that can be done by combining the movement with a body posture in which the torso is bended), and the head could occasionally turn to the left and the right alternately denoting a certain disappointment (and that can be achieved using a secondary motion that drives the head). Moreover, the movement combinations and the variation in the speed and in the spatial amplitude can be parameterised to express different degrees of sadness. For instance, and recovering the above example, we could denote more or less sadness generating a movement more or less exaggerated. Thus, if the character is very sad, the walking movement should be very slow, with very small steps, the torso highly bended and frequent head movements. On the other hand, if the character is only a little bit sad, the walking movement could be moderately slow, with small steps, the torso slightly bended and rare head movements.

4 Architecture

The system has a three-layer architecture as depicted in Figure 1.

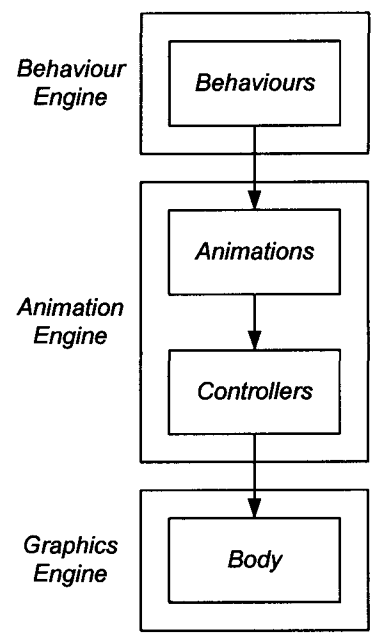


Figure 1: Architecture

The *behaviour engine* is responsible for the high-level control of the character, and it is typically implemented by the character's mind. It sends requests to the *animation engine* for activation/deactivation of animations or body postures, and for increasing/decreasing the speed and spatial amplitude.

The *animation engine* is responsible for combining all the active animations and implements a simple resource mechanism to avoid unwanted behaviours. It has direct control over the character's animations and body postures.

Finally, the *graphics engine* is responsible for maintaining the geometric model and for controlling the rendering process.

4.1 Character Model

A character is defined by a geometric model and by a set of pre-defined animations and postures that can be combined in real-time as described above. We use a simple geometric model composed by a hierarchical joint skeleton and a deformable skin layer attached to the joints. A posture reflects a certain joint configuration and an animation is the variation of the joints or a skin deformation over the time.

Using this information, the *animation engine* is capable of calculating the contribution of multiple active animations and/or postures, and updates the geometrical model every cycle to reflect the desired behaviour.

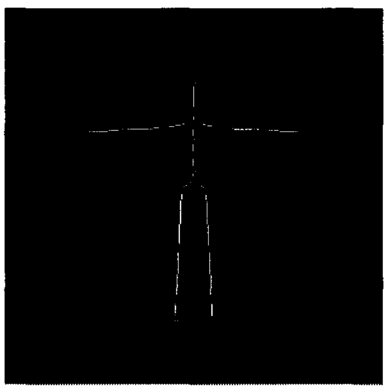


Figure 2: Hierarchical Joint Skeleton

Figure 3: Deformable Skin

4.2 Blending Animations

Typically, applications that use synthetic characters use libraries of animations that are mutually exclusive. For instance, characters used in computer games usually perform one action at a time, and that can be simply achieved using the appropriate animation for each action. However, these characters are often repetitive and predictable, unless you have a huge database of animations.

Our work combines animations and/or body postures in real-time to generate new movements that add a certain behavioural richness to the character. Basically, the *animation engine* generates a mixed movement of each active animation/posture using a linear interpolation algorithm. Each animation or posture has an associated weight that changes in real-time and that determines the contribution of that specific animation or posture to the generated movement.

This behavioural richness allows us to influence basic animations and create variations that otherwise would have to be created as independent animations. A good example is a "walk movement" and the possible body posture variations to reflect different emotional states.

4.3 Speed and Spatial Amplitude

Amaya et al. (1996) presented a work where they concluded that the speed and the spatial amplitude of the movement vary noticeably with different emotions. For instance, "sad" movements are normally slow and narrow whereas "happy" movements are fast and wide.

Our work uses these results and allows variations in the speed (increase/decrease the number of frames per unit of time) and in the joint amplitude. Therefore, it is possible to parameterise an animation in real-time to denote inner-feelings or even personality. For example, a lazy person runs more slowly than an energetic one.

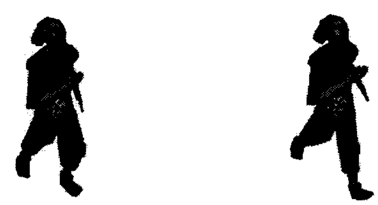


Figure 4: Lazy Run

Figure 5: Energetic Run

5 Concluding Remarks

Our work does not intend to establish any mapping between emotional states and movements or movement changes. We intend to provide characters that have the ability of modifying their movements in real-time and, therefore, that are able to express emotions. How a particular emotion is expressed is out of the work's scope.

Therefore, the main contribution of the work to the research of expressive characters for social interactions is a simple method of manipulating the character's body to achieve more natural and believable movements. The most immediate advantage is the generation of a rich set of animations using a compact database of pre-defined animations. For instance, it will be no longer necessary to create a "normal" walk, a "sad" walk, a "happy" walk, or a "tired" walk. Using a neutral walk and the right set of parameterisations and body postures we can achieve new animations similar to those produced offline. This allows a smaller dependence on the designers, and reduces the time necessary to create a character and its library of animations.

Another advantage is the possibility of visually identifying generic parameterisations for expressing a specific emotion, which could lead to the development of a broader model that establishes a matching between a theory of emotions and the parameters that affect the animations.

6 Future Work

The final version of this work will be integrated in the SAFIRA toolkit as the Bodily Expression Component. The Bodily Expression Component will be used in FantasyA, one of the SAFIRA demonstrators. FantasyA is a computer game where the characters must to be able to express emotions with the body.

Acknowledgements

This work has been partially supported by the EU funded SAFIRA project number IST-1999-11683.

Ana Paiva has also been partially supported by the POSI programme (do Quadro Comunitário de Apoio III).

References

Amaya et al. Emotion from Motion. *Graphics Interface*, 222-229, 1996

Aubel and Thalmann. Realist Deformation of Human Body Shapes. *Proceedings of Computer Animation and Simulation*, 125-135, 2000

Badler et al. Simulating Humans: Computer Graphics Animation and Control. *Oxford University Press*, 1993

Badler et al.. To Gesture or Not to Gesture: What is the Question? *Proceedings of Computer Graphics International*, 3-9 (2000)

Blumberg and Galyean. Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. *Proceedings of SIGGRAPH'95*, 47-54, 1995

Chi et al. The EMOTE Model for Effort and Shape. *Proceedings of SIGGRAPH'2000*, 173-182, 2000

Fua et al. Realistic Human Body Modeling. *Fifth International Symposium on the 3D Analysis of Human Movement*, 1998

Johnson et al. Sympathetic Interfaces: Using a Plush Toy to Direct Synthetic Characters. *Proceedings of CHI'99*, 152-158, 1999

Kalra et al. Real-Time Animation of Realistic Virtual Humans. *IEEE Computer Graphics and Applications*, 18(5):42-56, 1998

Perlin. Real time responsive animation with personality. *IEEE Transactions on Visualization and Computer Graphics*, 1(1):5-15, 1995

Perlin and Goldberg. Improv: A System for Scripting Interactive Actors in Virtual Worlds. *Proceedings of SIGGRAPH'96*, 205-216, 1996

Russell and Blumberg. Behavior-Friendly Graphics. *Proceedings of Computer Graphics International*, 1999

Unuma et al. Fourier Principles for Emotion-based Human Figure Animation. *Proceedings of SIGGRAPH'95*, 91-96, 1995