

Affective Interactions Between Expressive Characters*

Marco Gillies

University College London
Adastral Park Campus, Ross Bldg pp1
Adastral Park, Ipswich IP5 3RE, Ipswich, UK
m.gillies@ucl.ac.uk

Daniel Ballin

Broadband Applications Lab, BT plc
Ross Bldg pp4
Adastral Park, Ipswich, IP5 3RE, Ipswich, UK
daniel.ballin@bt.com

Abstract – *When people meet in virtual worlds they are represented by computer animated characters that lack a variety of expression and can seem stiff and robotic. By comparison human bodies are highly expressive; a casual observation of a group of people will reveal a large diversity of behavior, different postures, gestures and complex patterns of eye gaze. In order to make computer-mediated communication between people more like real face-to-face communication it is necessary to add an affective dimension. This paper presents Demeanour, an affective semi-autonomous system for the generation of realistic body language in avatars. Users control their avatars that in turn interact autonomously with other avatars to produce expressive behaviour. This allows people to have affectively rich interactions via their avatars.*

Keywords: Affective Computing, Graphical Agents, Computer Mediated Communication.

1 Introduction

Affective computing is computing that relates to, arises from, or deliberately influences emotions, and is now flourishing as a research field, kick started and popularized by people like Pickard [20].

Humans need emotions to survive; we use them all the time, they help us relate to others, and play an integrated part in the decisions we make. As computers become ubiquitous it was natural to ask the questions do machines need to have emotions in order to live alongside humans. We are certainly not afraid of showing our emotions to computers, people shout and show anger when their computer crashes, and people equally show delight and joy when something works. This raises the provocative question for the HCI community, how should computers react back.

As humans we can relate to virtual characters and many robots due to their anthropomorphic characteristics, therefore it is of no surprise that computational models of emotion synthesis use virtual characters as outputs, as they can express themselves in so many ways, posture, gesture, eye-gaze, gait, and facial expression.

The original goal of a good deal of research in this field was for the use of improvisational animated

characters in the entertainment industry[19]. These virtual characters would have moods and personalities and aid the author in the creation character animations. However the focus quickly shifted to interactive virtual characters, for entertainment as well as for believable characters in education. One example is the Oz project [3] developed 'The Woggles' for a Boston AI arts exhibition.

To make character look more believable many people took inspiration from Disney animators such as Thomas and Johnston [24] who said "From the earliest days, it has been the portray of emotions that has given the Disney characters the illusion of life."

The popular way to model emotion by computer scientists is to use one of the cognitive approaches such as the OCC approach[18] or Fridja's theory of emotions[11]. These psychological models treat emotion in terms of their cognitive state as opposed to the expressive behavior they portray, but have the advantage that they are easy to implement.

Martinho and Paiva used the OCC model for the implementation of an exhibit for the Lisbon Expo'98 [16], this had two virtual dolphins Tristão a shy dolphin and Isolda who was more playful. Visitors would interact with Isolda by touching a statue of a dolphin, which had touch pads located all over it. This would alter the emotional state of Isolda, who would in turn affect the emotional state of Tristão and encourage it to stop hiding in the remains of a sunken ship. This system is interesting from the point of view that the autonomous character based system still has humans in the loop.

Autonomous characters that can interact with people already exist, examples include Gandalf [25] and Rea [5] as well as [13,23]. Adding autonomy to an avatar can allow it to exhibit complex behavior without requiring the user to perform time-consuming control sequences, which may well distract from other tasks. However there are few cases where humans are influencing the behavior in multiparty semi-autonomous characters, where the characters behavior is a combination of the input from the user, the innate behavior of the character, as well the affective interactions between the characters.

FantasyA [21] is a computer game where two characters interact in an affective dual. The characters make judgments on one another based on how they feel

* 0-7803-8566-7/04/\$20.00 © 2004 IEEE.

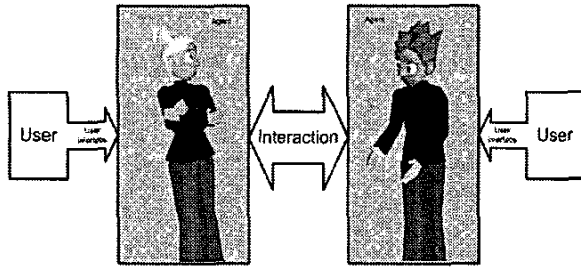


Figure 1 Human-Agent-Agent-Human interaction

towards one another. The users can influence their character with the SenToy, a tangible interface in the form of a doll, which has sensors over its body. For example by shaking the doll up and down a user can increase the character's happiness, or by covering the dolls eyes with its hands you can make it scared.

The aim of our work is to add an affective dimension to computer-mediated communication. This is done by having each user represented by a graphical agent, or avatar, which is endowed with behaviour that expresses affect. The communication has two components as shown in figure 1. The user interacts with their agent in order to determine the agents affective state. This agent then autonomously interacts with the agents of other users to produce affective communication. Having the agents interact autonomously frees the user from having to manage the details of the interaction while still maintaining control of the general affective content. This allows them to concentrate on other aspects of the communication, for example the verbal communication. This use of semi-autonomous agents as avatars in mediated conversation is based on the work of Vihljálmsón and Cassell[26] as well as our own previous work[13].

2 Overview

The Demeanour Framework for expressive avatars[14] allows people to have remote, affectively rich interactions mediated by graphical agents that act as avatars. This process is shown in figure 1. Users control their own avatar. This is done at a high level, controlling the agent's affective state. This high level control makes sure that the avatar's behavior is appropriate to the needs of the user without putting excessive demands on the user for detailed control. Lower level features of the generation of behavior are handled by the agent itself. In particular the agent handles reactions to the behavior of other agents, without need for user interaction.

This paper describes how human affective interactions are mediated by the agents and in particular how the agents themselves interact with each other based on user commands. The next section describes the psychologically based model of affective expressive behavior we use. The Demeanour framework is described in section 4. Sections 5 and 6 describe the interactions

between users and agents and between the agents themselves.

3 Affect Model

When dealing with affective communication and interaction it becomes more important to deal with the affective aspects of relationships between people, rather than emotions that are purely individual. As such the affect model used in this work is based on relationships between people, or more exactly on the attitude of one person to another. This model of attitude is based on the work of Argyle [1] and Mehrabian [17]. Though there is an enormous variety in the way that people can relate to each other Argyle identifies two fundamental dimensions that can account for a majority of non-verbal behavior, affiliation and status. Affiliation can be broadly characterized as liking or wanting a close relationship. It is associated with close postures, either physically close such as leaning forward or other close interaction such as a direct orientation or more gaze. Low affiliation or dislike is shown by more distant postures, including postures that present some sort of barrier to interaction, such as crossed arms. Status is the social superiority (dominance) or inferiority (submission) of one person relative to another. It also cover aggressive postures and postures designed to appease an aggressive individual.

For true affective interaction it is vital that agents are able to react to each other's behavior. The relationship between the attitude behavior of two people can take two forms, compensation and reciprocation. Argyle presents a model in which people have a comfortable level of affiliation with another person and will attempt to maintain it by compensating for the behavior of the other, for example, if the other person adopts a closer posture they will adopt a more distant one. Similar behavior can be observed with status, people reacting to dominant postures with submission. Conversely there are times where more affiliation generates liking and is therefore reciprocated, or where dominance is viewed as a challenge and so met with another dominant posture. Argyle suggests that reciprocation of affiliation occurs in early stages of a relationship. Status compensation tends to occur in an established hierarchy, and challenges occur outside of a hierarchy. Compensation results in what Argyle calls the equilibrium theory of non-verbal communication. Each person has a certain equilibrium value for a certain attitude such as affiliation, a certain degree of closeness with another that they are comfortable with. If the other person has behavior that is too close or distant they will attempt to compensate it with their own behavior.

Non-verbal behavior also interacts closely with the flow of conversation. Non-verbal behavior is used extensively to regulate conversation dealing with aspects such as turn-taking (determining who should speak at a given time). Gaze is a particularly important modality in this respect, Argyle and Cook[2] have done extensive

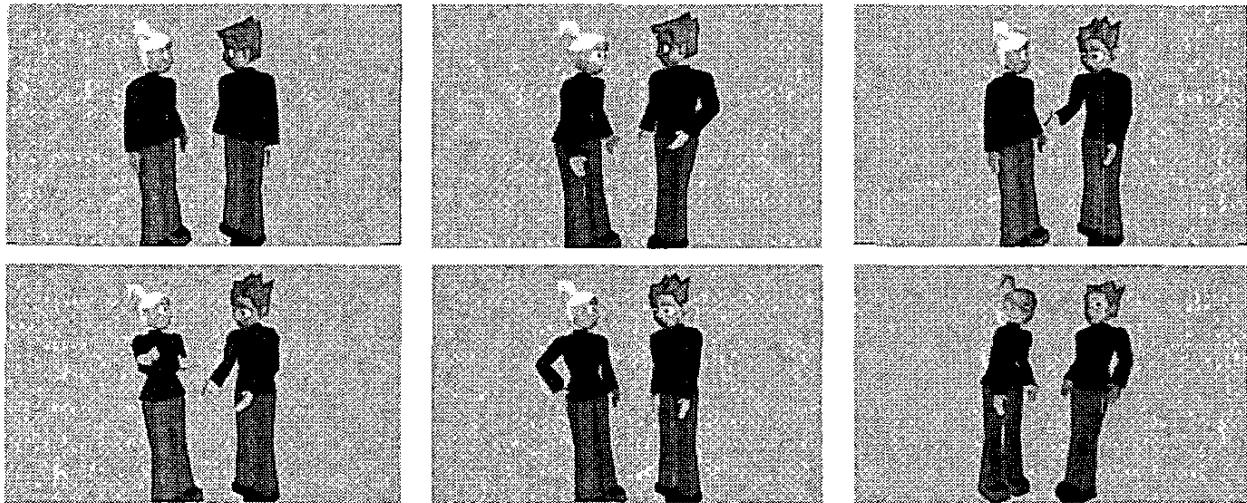


Figure 2: Interactions between the affective behaviour of agents. Clockwise from top left: neutral behaviour; female with high affiliation (close posture) and male with high status relaxed behaviour; neutral behavior with male character gesturing while talking; female character reacts affiliatively to a male character's relaxed posture; high status, space-filling female character and low status male character, and a male character talking with some high affiliation behaviour to which the female character reacts with a distant, low affiliation posture.

studies with pairs of individuals to understand levels of eye gaze, and mutual gaze, in conversations. They have produced some useful results concerning the level to which individuals will look at the other while speaking (on average 35%) and listening 75%). We have used these results to influence our model of gaze and mutual gaze in group settings. Expressive behavior is also related to other aspects of conversation, for example, gestures only ever occur during speech (whether talking or listening). Non-verbal communication also provides a back channel, feedback to the speaker from the listener, which can encourage the speaker or show disagreement, the most common back channel gesture in western culture is the head nod (and the head shake for disagreement). Figure 2 shows example of the interactions of the expressive behavior of two agents.

The affective state of the agent is expressed through the animation of its graphical "body". This builds on existing work on graphical agents with expressive behaviour, for example, Cassell *et al.*'s various systems particularly their virtual real estate agent, Rea[5]. Guye-Vuillème *et al.*[15] have demonstrated avatars with a wide range of controllable expressive behaviour. APML (Affective Presentation Markup Language) is an XML based language for defining the expressive behaviour of characters[10].

Expressive behaviour is displayed in a number of ways, or modalities, for example facial expression, tone of voice, gesture, posture and gaze. In our system we use postures (which is closely related to attitude), gaze, and, to a more limited degree, gesture (which is related to flow of conversation). research on simulation of posture includes work by Cassell *et al*[6] on posture shifts, and an attitude model similar to our affiliation by Béchariez and Thalmann[4]. Simulations of gaze for conversation

include Garau *et al*[12], and Colburn, Cohen and Drucker[19]. Vilhjálmsón and Cassell[26] use eye gaze to regulate flow of conversation and Rickel and Johnson[22] use gaze to indicate areas of interest in their virtual reality tutoring system. Gesture has been studied fairly extensively and work includes a number of systems by Cassell *et al*[5,7] and work by Chi, Costa, Zhao and Badler[8].

Posture and gesture are controlled in the same way. They are generated based on a set of basis postures and gestures. A number of these bases are chosen based on the attitude of the agent, and are interpolated together with random weights to generate new postures and gestures (using a random group term as described below). To ensure variety new bases and weights are chosen at random intervals. The postures depend on the attitude of the agent, each basis posture is attached to one or more attitude value and the weights generated for these postures and proportional to the value of their attitude. Gesture also depend on the flow of conversation, some gestures only appear while talking, while others (back-channel gestures) only appear when listening, for example head nods. These types of gestures are only turned on when talking and listening (using a switch term, as described below).

Gaze is closely connected with flow of conversation. People tend to look more at someone when they are speaking while speakers look slightly less. The gaze engine determines which agent to look at by having a probability of looking at each agent. A different value of this probability is used depending on whether the agent is talking, listening or neither. Attitude also affects gaze, for example, high affiliation results in more gaze. The base gaze values determined from speech are scaled by the agent's attitude to the other agents.

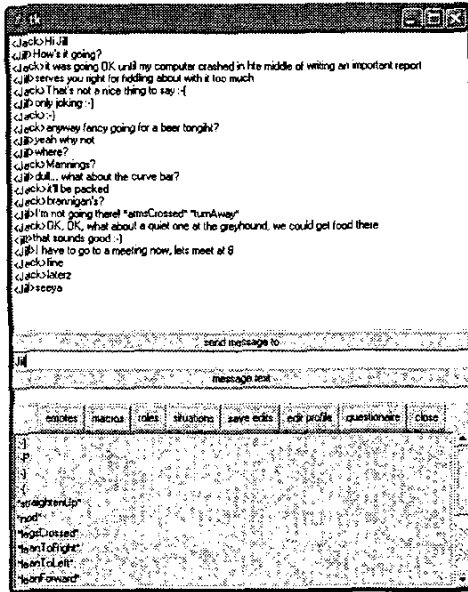


Figure 3: The Demeanour text chat interface

4 Demeanour

The work presented in this paper uses the Demeanour framework[14] for expressive behavior to implement the type of mediated affective communication described above. The behavior generated by Demeanour depends on a number of factors, such as user input, the attitudes of the agent and the behavior of other agents. These factors are mapped to a number of parameters that are passed to the underlying animation system so as to actually animate the agent. This mapping occurs in a "behavioral controller" that determines the actions of the agents. The controller consists of a number of terms which are intermediary values calculated from other terms, including input factors. Terms can have a number of types. The simplest are parameters which are single values. These can provide input from other agents, or they can provide a means for users to control the behavior of their agent. Terms can also be the combination of a number of other terms, this combination can be done in a number of ways:

- Sum of product terms combine their inputs via multiplication and addition of their values.
- Switch terms choose one of their inputs based on the value of another term. These can be used to turn a particular behaviour on or off depending on the value of a term, e.g, turning gestures on when speaking.
- Random group terms map their input to a number of outputs. The output values are each a proportion of the input, the proportions are chosen at random. These are used to create new blends of existing postures and gestures as described above.

5 Human-Agent interaction

The real time control interface to an agent should be well integrated with the main user interface of the world. We have chosen to demonstrate Demeanour using a text chat interface that is very common in 3D multi-user virtual worlds and more generally on the internet. Users have a form of conversation consisting of short messages that are sent to the other users on pressing the return key. The interface is shown in figure 3.

Interaction with the agents occurs through changing the values of the parameters of the controller. The most direct way of doing this is to edit their values via a set of sliders. Such parameters are displayed on a set of sliders that the user can use to edit the parameters. A more intuitive interface directly uses the conventions of text chat. Textual communication on the internet already has its own vocabulary to express emotion and attitude, namely emoticons or smilies :-). A smiling emoticon :-) will increase the 'friendliness' parameter while a frownie :-(will reduce it. This provides a very natural interface that does not intrude on a conversation, and is already well understood by internet users. The behavior language can define emoticons that are parsed in the text chat interface.

6 Agent-Agent interaction

In interactions between multiple agents the relationships between these agents becomes critical. Importantly, an agent can no longer be considered to have a single affective state, but a different attitude to each member of the group, with the resulting behaviour being a combination of all these attitudes. Body language can reflect an attitude to the group as a whole or may be focused on one or two members. In Demeanour interactions between agents occur via the terms of each agent's behavioral controller. Agents can have direct access to the selected terms of another's controller. In the controller, certain terms are marked as *exports* in which case they can be seen by other controllers which access them as *import* terms. When two agents start to interact the imports of one are matched, by name, to the exports of the other. For the rest of the interaction the values of the exported terms are used as the values of the imported parameters.

Additional features have been added to the Demeanour framework to deal with interaction of more than two agents. The behavioral controllers of different agents are joined with terms from one controller providing input to others. The controller must be able to generate different behavior with respect to different agents in a conversation; to achieve this each term in the controller may be evaluated independently for each agent, each of which results in a different value. To be exact, each term has a number of different types of value:

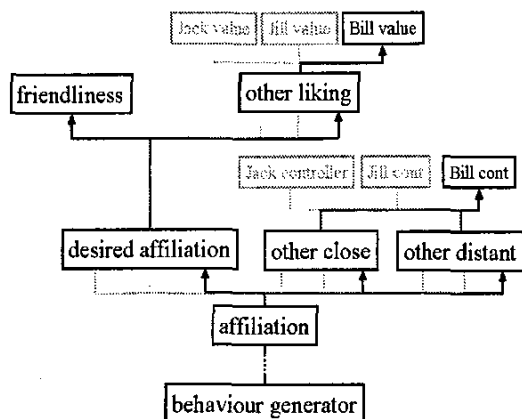


Figure 4 : Evaluating the value of a term for a particular agent.

- A default value is used for terms that are the same for all agents, for example internal personality factors. This value is also used in agent-dependent terms when there is no value specified for a particular agent.
- A value for each of the other agents in the conversation.
- The average value of all the agents.
- There are a number of roles that are specified for agents, for example the current focus of attention (see below for details). A role always stands for an agent but that agent varies. If a term is evaluated for a role the result is the its value for the agent currently occupying that role.

Figure 4 shows how terms are evaluated for a particular agent. A behavior generating module (for example, the posture or eye gaze modules) will request to evaluate a term for a given agent. How this is done depends on the type of term. As described above there are two types of term, parameters that have their own value, and compound term which are calculated as a combination of other terms. Compound terms evaluate each of the terms it depends on for the agent and then combines them as usual. Parameters on the other hand will have a different value for each agent. They might be an import term in which case their value corresponds to the value of a term in other agents' controllers. The value chosen is therefore the value from the controller of the agent we are interested in. Other parameter can also have different value for different agents. An example might be a parameter that represents how much the one agent likes a particular individual. Each agent can store parameter values for other agents, which are loaded when a conversation starts with that other agent, and so agents can remember particular relationships between each other. To give a concrete example, in figure 3 we can imagine evaluating the term "affiliation" for agent "Bill". This is a compound term that depends on "other close", "other distant" and "desired affiliation", so each of these are also evaluated for Bill. The first two terms are import terms and so they query Bill's controller for the appropriate value. Desired affiliation is a compound term



Figure 5 Affective interaction between three agents.

which depends on "other liking" and "friendliness". Other liking has a different value for each agent so Bill's specific value is used. On the other hand friendliness is a personality factor that is agent-independent so the default value is used.

Gaze is used to determine the other agents that a particular agent is reacting to. The gaze system determines a particular focus of attention at any given time, which is the agent or location currently being looked at. If the focus of attention is another agent the posture and gesture behavior will be determined with respect to that other agent. Otherwise the average values for all agents in the conversational group is used. The gaze behaviour itself evaluates the gaze probabilities separately for each other agent and uses them to determine new foci of attention.

7 Conclusions

This paper has presented methods in which human-to-human interaction can be mediated by affective agent to agent interaction. Figure 5 shows and example of interaction in a group of three agents with various attitudes to each other. The male agent with the light shirt has a high affiliation to both other agents. The female agent has a high affiliation to him, but a low affiliation to the other male agent, who in turn has a high affiliation to the female agent, and a low affiliation to the other male.

Intial results show that this is a promising new technique for internet based communications. User trials are currently being designed to test the effectiveness of the system.

References

- [1] M. Argyle, *Bodily Communication*, Routledge, 1975.
- [2] M. Argyle and M. Cook *Gaze and Mutual Gaze* Cambridge University Press 1976

- [3] J. Bates, "The Role of Emotion in Believable Agents", *Communications of the ACM*, Vol 37, Issue 7 pp122-125, July 1994.
- [4] P. Béchariez and D. Thalmann "A model of nonverbal communication and interpersonal relationship between virtual actors. Proc. Computer Animation, pp 58-67 1996
- [5] J. Cassell, T. Bickmore, L. Campbell, H. H. Vilhjálmsón and H. Yan "Embodiment in conversational interfaces: Rea" Proc ACM SIGCHI pp 520-527 1999
- [6] J. Cassell, Y. Nakano, T. Bickmore, C. Sidner and C. Rich "Non-verbal cues for discourse structure" Proc. 41st Annual meeting of the association of computational linguistics, Toulouse, France pp 106-115 2001
- [7] J. Cassell, H. H. Vilhjálmsón and T. Bickmore "BEAT: the behaviour expression animation toolkit" Proc. ACM SIGGRAPH, pp 477-486 2001
- [8] D. Chi, M. Costa, L. Zhao and N. Badler "The Emote model for effort and shape" Proc. ACM SIGGRAPH, pp173-182 2000
- [9] A. Colburn, M. Cohen and S Drucker *The role of eye gaze in avatar mediated conversational interfaces* Tech. Rep Microsoft research 2000
- [10] B. DeCarolis, C Pelachaud, I Poggi and M. Steedman "APML, a markup language for believable behaviour generation" *Life-like characters: tools, affective functions and applications*, H. Prendiger and M. Ishizuka (Eds) Springer pp 65-87 2004
- [11] N. Frijda, *Emotions*, Cambridge University Press, 1987.
- [12] M. Garau, M. Slater, S. Bee and M. A. Sasse "The impact of eye gaze on communication using humanoid avatars" Proc *ACM SIGCHI* pp 309-316 2001
- [13] M. Gillies, N. Dodgson, and D. Ballin, "Autonomous secondary gaze behaviours", Proc AISB workshop on Animating Expressive Characters for Social Interactions, Imperial College London, (R. Aylett and D. Cañamero, eds.), April 2002.
- [14] M. Gillies and D. Ballin "Integrating autonomous behaviour and user control for believable agents" Proc. Third international joint conference on Autonomous Agents and Multi-Agent Systems (AAMAS), New York 2004
- [15] A. Guye-Vuilléme, T.K. Capin, I.S. Pandzic, N. Magnenat-Thalmann, and D. Thalmann "Non-verbal communication for collaborative virtual environments" *The Virtual Reality Journal* Vol 4 pp 49-59 1999
- [16] C. Martinho and A. Paiva, "Developing Pathematic Agents", Proc. Workshop on emotion based architectures. Autonomous Agents '99. Ed. J. Velasquez, 1999.
- [17] A. Mehrabian, *Nonverbal Communication*, Aldine-Atherton, 1972.
- [18] A. Ortony, G. Clore, and A. Collins, *The cognitive structure of emotions*, Cambridge University Press 1998.
- [19] K. Perlin, A. Goldberg, "IMPROV: a system for scripting interactive actors in virtual worlds", Proc. ACM SIGGRAPH, 1996.
- [20] R. Picard, *Affective Computing*, MIT Press, 1997
- [21] R. Prada, M. Vala, A. Paiva, K. Höök, A. Bullock, "FantasyA - The Duel of Emotions", Proc. of the 4th International Working Conference on Intelligent Virtual Agents 2003
- [22] J. Rickel and W. L. Johnson "Animated agents for procedural training in virtual reality: Perception cognition and motor control" *Applied Artificial Intelligence* Vol 13 pp 343-382 1999
- [23] P. Sengers, S. Perry, and J. Smith, "Traces: Semi-autonomous avatars", unpublished, available at <http://www-2.cs.cmu.edu/~phoebe/work/publications.html>, 2000.
- [24] F. Thomas and O. Johnston, *Disney animation: The illusion of life*, Abbeville Press, New York, 1981.
- [25] K. Thórisson, "Real-time Decision Making in Multimodal Face-To-Face Communication", Proc the second ACM international conference on autonomous agents, pp 16-23, 1998
- [26] H. H. Vilhjálmsón and J. Cassell "BodyChat: Autonomous communicative behaviours in avatars" Proc. ACM International conference on autonomous agents, 1998