Agents and affect: why embodied agents need affective systems

R.S.Aylett

CVE, University of Salford,
Salford, M5 4WT, UK
r.s.aylett@salford.ac.uk

Abstract. This paper discusses why intelligent embodied agents – concentrating on the case of graphically-embodied agents – require affective systems. It discusses the risks involved in attempting to produce naturalistic expressive behaviour, and examines the specific case of facial expression generation. It considers the approach taken by the EU Framework 5 project VICTEC to facial expressions and then discusses some of the issues in affective architectures needed to drive facial expressiveness in an autonomous Intelligent Virtual Agent (IVA).

1. Introduction

Unlike software agents in general, intelligent virtual agents (IVAs) - or synthetic characters, or virtual humans - have an important characteristic in common with robots. This is embodiment, albeit in virtual form. Embodiment has a number of consequences. It means that an IVA typically interacts with its virtual world through sensors, of greater or less elaboration, and requires a perceptual component in its architecture to handle this. It raises sometimes complex control issues as limbs, trunks and heads have to move in a competent and believable fashion. However the consequence that we will concentrate on in this paper is the use of embodiment as a communication mechanism for the internal state of the agent, complementing its explicit communication mechanisms such as natural language.

Why might one want to communicate an IVA's internal state to a human user? One important reason is to support the continuing human process of inferring the intentions of an IVA – it motives and goals. This can help to produce a feeling of coherent action which is required for the user to feel that in some sense they 'understand' what an IVA is doing. We argue that a vital component of this process for the user is recognising the emotional state of the IVA and relating it to their own affective state. If in turn the IVA is able to recognise the affective state of the user and perform an equivalent integrative process, then one could speak of an 'affective loop' [INVOLVE] between user and IVA.

Affective computing is a new but growing area within computer science, often dated to the seminal work of Picard in 1987 [Picard 87]. It encompasses a large number of research topics: some human-centred - sensing human affect and affective responses, modelling the user state; and some technological - from affective wearable computers to synthesising emotional systems on machines. Affect impinges directly upon work with IVAs at the latter end of this list: in the expression of affective state, and in the synthesis of the states that are to be expressed.

Where software agents in general can often be thought of as communicating mostly with each other, embodied agents are more and more expected to function in the human social context. Emotional expressiveness in such a context is not an optional add-on but a fundamental requirement for competence and acceptability [Sloman & Croucher 1981]. Consider for example the acceptability of a virtual newsreader who announces the collapse of the World Trade Centre in the same equable manner as the winner of a multi-million pound lottery.

Moreover once an agent is embodied, the issue of trust is no longer merely a rational calculation, but is based on the perceived personality of the agent, in turn intimately connected to its emotional expressiveness. Thus an IVA that offers the 'human touch' can only do so if its expressive behaviour is consistent and appropriate. For example, selling property, as does REA [Cassell et al 99], cannot be carried out in the same way as selling newspapers. The difference is not merely about the length of the dialogue between user and IVA but about the affective engagement of the user in the process, in turn partly dependent on the feeling that the IVA 'cares' about the seriousness of the transaction. Other potential IVA applications have even greater requirements for expressive accuracy – consider for example the level of social immersion required of a user in an interactive narrative. Here, any feeling that the IVAs are 'wooden' or 'robotic', and not emotionally engaged with other IVAs and with their own choices, may completely undermine the user's own engagement in the story-telling process.

2. Expressive behaviour

The main channels or modes available to an IVA for affective display are the obvious visual ones of facial expression, posture, and gesture, plus in a few cases, voice. In this section we will discuss facial expression as a way of illustrating the main issues as this is the mode that has been researched most up to now. Less work has been carried out on posture (Badler's EMOTE system [Chi et al 00] is a seminal work here) and there is no agreed method for describing or classifying gesture as yet. Work in expressive voice is at a still earlier stage.

First it is worth highlighting an important generic issue for all modes – how far naturalism is an appropriate or realistic aim. In the graphics world, a driving force has been the attempt to achieve photo-realism, firstly in graphical worlds, and more recently in graphical characters. The 2001 animated film "Final Fantasy- The Spirits Within" is one example of such a search for photo-realism. On the other hand, many researchers in IVAs are actually striving for believability [Bates 94], where an IVA is perceived by the user as a compelling and engaging character.

The history of animation demonstrates that believability does not require naturalism – Mickey Mouse, probably the most successful animated character ever, looks nothing like a real mouse. A comparison between the films "Toy Story" and

“Final Fantasy” demonstrates that in some cases a non-naturalistic character can be more believable than a naturalistic one – certainly Buzz Lightyear in the former has made a far greater impact than any of the characters in the latter.

An implicit assumption behind the striving for photo-realism is the reverse argument that a more naturalistic character must also be a more believable one. However seminal work by Mori [Mori 82] in relation to robots, discussed in [Dautenhahn 02] in the context of the ‘life-like agents hypothesis’, demonstrates that this is not always true. Mori predicted that the more life-like a robot became, that is, the more naturalistically human, the more *familiar* it would also become. However, he argues that just short of 100% naturalism, there is actually a sharp drop in familiarity – or we might now say believability; a drop so deep that it forms what he calls ‘the uncanny valley’. An explanation of this effect suggests that a nearly-naturalistic character is capable of invoking standard human-to-human responses up to the point where these heightened expectations are jolted by some minor inconsistency. This jolt is experienced as a highly negative reaction to the IVA.

Mori actually broke down his overall curve into two separate ones for movement and appearance, and suggested that in fact the appearance curve was the dominant one of the pair. Since real-time interactive behaviour is a good deal more challenging than either static photo-realism or pre-rendered animation, the uncanny value is a persistent risk in the drive to naturalistic IVAs.

2.1 Facial expression

The risks of near-naturalism are perhaps not so widely perceived among those researching facial expressiveness for IVAs. Certainly much of this work draws directly upon human psychology and the categorisation of human facial expressions.

The Facial Action Coding System (FACS) is a method of describing facial movement based on an anatomical analysis of facial action. This was systemised by Ekman and Friesen, [Ekman & Friesen 78], but in fact goes all the way back to a French anatomist of the 19thC – Duchenne de Boulogne; the author of “Mecanisme de la physionomie humaine”. Duchenne had a patient with facial paralysis and was able to produce particular expressions on this man’s face by administering an electrical shock to different facial muscles. He then photographed the results to produce a catalogue of human expressions.

Fig. 2. GRETA [Pelachaud et al 01]

Fig. 1. Examples of the Duchenne smile

Affective facial expressions are defined in FACS not directly by muscle values but by 46 Action Units (AUs), relating to whole groups of muscles. Duchenne identified a particular type of smile ascribed to an unqualified feeling of happiness – still called the *Duchenne smile* to differentiate it from all the other types of smiles that humans display [LaFrance 02]. A Duchenne smile (Figure 1) is produced by two AUs or muscle groups. One group around the mouth, called the zygomatic pulls up the lip corners; the other, around the eyes, called the orbicularis, produces a lifting of the cheeks, narrowing of the eye opening and that gathering of the skin around the eye called 'crows feet wrinkles'. Later research has shown that while a human subject can consciously control the mouth AU, the orbicularis only responds to a genuine emotion of happiness.

FACS has been used for some time to visually code for affective expressions in human subjects from video. It can be used to define the six 'primitive emotions' identified by Ekman [Ekman 82]– anger, fear, disgust, joy, sadness and surprise – as recognised across all human cultures. It has also heavily influenced the design of the MPEG4 facial animation system with its Facial Definition Parameters (FDPs) and Facial Animation Parameters (FAPs). This in turn has provided a mechanism for researchers to produce facial animation in IVAs.

A good example of the use of MPEG4 as a basis for an expressive facial display in an IVA can be seen in GRETA (Figure 2) [Pelachaud et al 01]. This uses expressive folds and wrinkles in the model to achieve an equivalent effects to FACS using MPEG4 FAPs and XML language annotation to invoke the desired effect for a particular piece of speech output.

It can be considered perhaps the best such 'talking head' in the current state of the art, yet even the small –scale illustration of Figure 2 shows that it some way from photo-realism. It tackles a problem of naturalistic facial expression not so far mentioned – the dynamism of affective facial expression. While primitive emotions may be recognisable across cultures from photographs, no real-world face holds such am emotion for more than a fleeting second before moving on to the next. GRETA combines affective expressions with many other factors such as glance and head movement, yet the overall effect is still some way from that of a real human face.

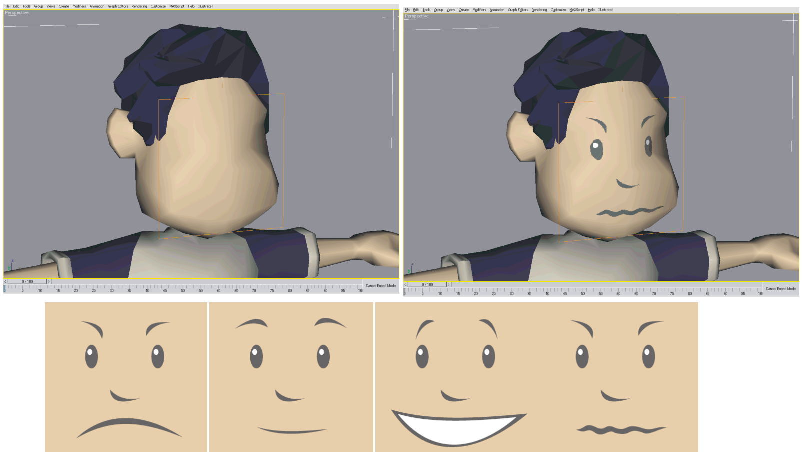

Fig. 3. VICTEC facial expressions and cartoon-like character design

Producing a naturalistic dynamic in real;-time interaction is an extremely challenging problem.

2.2 Avoiding naturalism

The IVAs being developed in the EU FP5 project VICTEC – Virtual ICT with Empathic Characters [VICTEC] – represent a very different position on facial expressiveness. In this project, IVAs act out short improvisational episodes on the theme of bullying, with a child user interacting with the victimised character between episodes to act as an 'invisible friend' and offer advice, influencing its behaviour in the next episode. Here, the expressiveness of the characters is required in order to engage the user in dramatic interaction of the episodes and to develop a feeling of empathy and thus of commitment towards the victim.

The approach just discussed in which FAPs are used to animate the structure of an IVA's face is clearly expensive in terms of rendering since it requires enough polygons in the facial model to support the number of parameters in use. Where an IVA is essentially a 'talking head' as in the case of GRETA, this level of detail is feasible, but once several characters must interact in a virtual world that is itself contributing substantially to the polygon count, this becomes very much less feasible. In addition, the user's viewpoint is very much further from the faces of the characters, making it harder to recognise the expressions being displayed.

The VICTEC characters are deliberately cartoon-like in design, as can be seen from the top-left component of Figure 3. In tune with this design, facial expressions have been designed as textures rather than polygonal animations, with a small set of primitives, some of which are shown in Figure 3. This is not only economical, but the expressions are dramatically obvious to users, and preliminary evaluation [Woods et

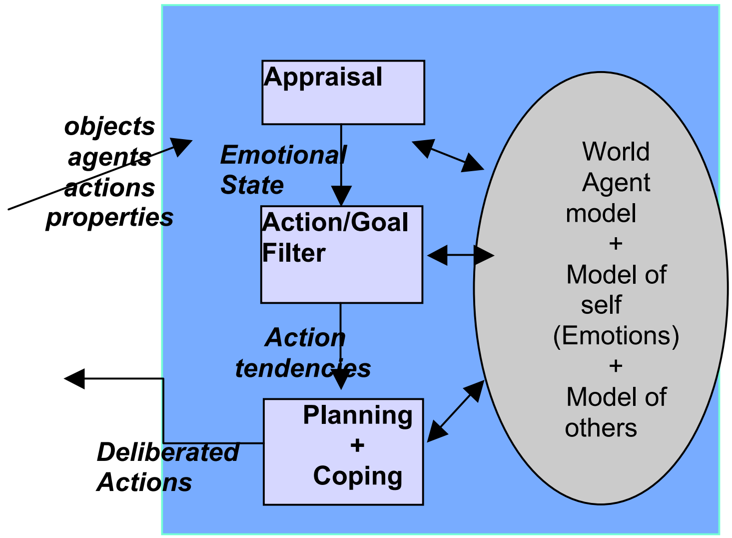

Fig. 4. The VICTEC 'Agent Mind'

al 03] has shown that believability is high even though naturalism is clearly not present.

3. Do the right thing?

Superficially, expressive behaviour may seem to belong entirely to the domain of animation: after all, to move the features of a face, make a gesture, adopt a particular posture is, for graphical characters, a graphical problem. While this is true for pre-rendered animated characters in film, once a character has to interact in real-time the problem is no longer merely one of graphically displaying an affective state, but of generating the affective state in the first place. So the XML annotation used by GRETA above to invoke facial expressions has to be generated 'on the fly' if it is to interact with a user. In other words, expressive behaviour is an instance of the generic problem of behaviour for an autonomous agent: how to select the appropriate behaviour, or in the now classic phrase 'do the right thing' [Maes 90].

Affect, however, has a special status as a behaviour – it is not only an output from a behaviour selection system, but is itself an input into that selection mechanism. As argued by Damasio [Damasio 94], affect should not be seen as the antithesis of rational decision-making but as a vital part of ensuring that decisions are timely and relevant to the situation of an agent. From this viewpoint, the expressive behaviour discussed above acts as a window for the user into the decision-making process of the IVA, giving it a leading role in supporting the decoding of motive and intention referred to above.

Where affective expressiveness is being generated by an internal architecture in which behaviour is being autonomously selected, two overall approaches can be discerned. One might be termed the 'low-level' account of affect and frequently implements models of IVA physiology, for example the endocrine system [Canamero 98, Valesquez 97]. Such architectures are very successful in handling changes in affective state since they are normally based on continuously-valued variables and reject affective labels such as the six primitive emotions already referred to in favour of emergent states resulting from the interaction of much lower-level processes.

However, difficulties emerge in applying such models successfully once IVAs must act at the higher cognitive level of responding to each other's goals and intentions. Thus in the VICTEC domain, a bullying action may cause physical pain if one character punches another, but more significant is the social humiliation of the exercise of power by the bully over the victim. Here, higher-level models taken from psychology have some advantages, since they are based on the concept of appraisal in which a person, event or situation is evaluated in relation to an IVA's own goals. Appraisal can be thought of as acting as a high-level filter on perceptual input to an agent (high-level because it uses far-from basic categories such as events, situations and persons) via an emotional state which constrains the possible behaviours to be selected for immediate execution. Figure 4 gives an example of the functional architecture that might result, as used in the VICTEC project.

Reaction Type	Definition
Attraction reactions	Which entities the agent likes or hates
Event reactions	How important are particular events to the agent?
Prospect-based reactions	How important are the agent's goals?
Attribution reactions	What are the agent' norms or standards of behaviour?

Table 1. OCC Appraisal categories

The taxonomy of Ortony, Clore and Collins [Ortony et al 88] – often abbreviated to OCC - has been used in a number of IVA systems since it supplies straightforward rules for the appraisal process. OCC divides the possible reactions of an agent into those in Table 1.

OCC emotion	Appropriate actions/behaviours
Joy	Smile, dance, laugh, wave
Happy-for	Felicitate, encourage
Sorry-for	Apologise, encourage, protect
Anger	Ignore, hit, avoid, aggress, humiliate
Distress	Cry, sit on the floor, beg

Table 2. Associating actions with OCC emotions in the bullying case

The OCC model was not in fact originally intended for the generation of affective state, but as a means of recognising it in other agents, and it has a number of weaknesses as an affective mechanism. One is that it has nothing concrete to say on how affective state produces behaviour selection. Here other work,, for example that of Lazarus [Lazarus & Folkman 84], is more useful since it introduces the concept of *coping behaviour* – behaviour that tries to equalise an imbalance between the external situation of an agent and its internal affective state. This approach has been used very successfully in the Mission Rehearsal Exercise system of USC [Gratch et al 01] as well as in Carmen' Bright IDEAS [Marsella et al 00].

An alternative is to define a behaviour repertoire indexed by the OCC taxonomic emotions, and how this might work for the type of emotions require3d for IVAs in the bullying scenarios of the VICTEC project is shown in Table 2.

4. Conclusions

In this paper we have argued that affective systems are not optional extras for IVAs but a requirement from at least two angles. The first is in the generation of the type of expressive behaviour that allows a user to understand an IVA in terms of its motivations and goals. The second is precisely in generating these motivations and goals as an input into the behaviour selection process of the IVA. In both cases, relevant work from psychology can be drawn upon, though in both cases this work undergoes modification for the purpose of IVA development.

One should note however that psychological theory is necessarily naturalistic, that is, it seeks to model and explain the natural behaviour of humans. In some sense then, the use of such theories - as against work in animation or indeed in drama, - is an implicit choice by IVA researchers in the direction of naturalism. Yet Mori's work shows that naturalism and believability are far from the same thing. It may in fact be that the aesthetic choices of the artist are at least as important as the scientific framework of the psychologist in IVAs, and that in expressive affective behaviour, narrative and drama require something that is 'larger than life' rather than the everyday norms of behaviour.

References

Badler, N. I; Reich, B. D. and Webber, B. L. (1997) Towards Personalities for Animated Agents with Reactive and Planning Behaviors, in Creating Personalities for Synthetic Actors, eds. Trappl, R. and Petta, P. Springer-Verlag pp43—57 1997

Bates, J. (1994), The Role of Emotion in Believable Agents Communications of the ACM v37, n7 pp122—125 1994

Canamero, D. (1998) Modeling Motivations and Emotions as a Basis for Intelligent Behaviour, Proceedings of the First International Conference on Autonomous Agents eds. Johnson, W. L. and Hayes-Roth, B. ACM Press pp148—155, 1998

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

Chi, D; M. Costa, L. Zhao, and N. Badler (2000) The EMOTE model for Effort and Shape, *ACM SIGGRAPH '00*, New Orleans, LA, July, 2000, pp. 173-182 2000

Dautenhahn, K. (2002) The Design Space of Life-Like Robots. Proc. 5th German Workshop on Artificial Life, Lübeck 18th - 20th March, 2002, IOS Press, Eds D. Polani, J. Kim, T. Martinetz, pp. 135-142

Damasio, A. (1994)Descartes Error. Avon Books. 1994

Duchenne (de Boulogne), G.-B. (1876). Mecanisme de la Physionomie Humaine. Atlas. Deuxieme edition. J.-B. Bailliere et Fils

Ekman,P. and Friesen, W.V (1978) Manual for the Facial Action Coding System, Consulting Psychology Press, Palo Alto, Calif., 1978

Ekman, P. (1982) Emotions on the Human Face. Cambrdige University Press

Gratch, J; Rickel, J; & Marsalla, S. (2001) Tears and Fears, *5Th International Conference on Autonomous Agents*, pp113-118 2001

INVOLVE group http://www.dsv.su.se/research/k2lab/involve_group.htm

LaFrance, M. (2002) What's in a Robot's Smile? The Many Meanings of Positive Facial Display In: eds D..Canamero, R.S.Aylett, proceedings, Aimating Expressive Characters for Social Interraction, AISB Symposia, April 2003 Imperial College London

Lazarus, R.S& Folkman, S. (1984). Stress, appraisal and coping. New York: Springer

Maes, P. (1990) How to do the right thing. Connection Science. 1990.

Marsella, S; Johnson, L.W. & LaBore, C. (2000) Interactive Pedagogical Drama Proceedings, Autonomous Agents 2000

Mori, M. (1982), *The Buddha in the Robot*, Charles E. Tuttle Co

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

Pelachaud,C; E. Magno-Caldognetto, C. Zmarich, P. Cosi, [01] Modelling an Italian Talking Head, Audio-Visual Speech Processing, Scheelsminde, Danemark, 7-9 septembre 2001

Picard, R. (1997), Affective Computing, MIT Press 1997

Sloman, A. and Croucher, M. (1981), Why Robots Will Have Emotions IJCAI81, PP197--202, Vancouver, B.C 1981

Velasquez, J.D. (1997) Modeling Emotions and Other Motivations in Synthetic Agents. Proceedings, 14th National Conference on AI, AAAI Press, pp10-16

VICTEC www.victec.org

Woods, S; Hall, L; Sobral, D, Dautenhahn, K. & Wolke, D. (2003) Animated characters in bullying intervention. In: T.Rist, R.S.Aylett, D.Ballin (eds) Proceedings, IVA2003 Springer LNAI 2003