*CHAPTER FOURTEEN*

# Semi-Autonomous Avatars: A New Direction for Expressive User Embodiment

Marco Gillies, Daniel Ballin, Xueni Pan and Neil A. Dodgson

## 1. Introduction

Computer animated characters are rapidly becoming a regular part of our lives. They are starting to take the place of actors in films and television and are now an integral part of most computer games. Perhaps most interestingly in on-line games and chat rooms they are representing the user visually in the form of avatars, becoming our on-line identities, our embodiments in a virtual world. Currently on-line environments such as "Second Life" are being taken up by people who would not traditionally have considered playing games before, largely due to a greater emphasis on social interaction. These environments require avatars that are more expressive and that can make on-line social interactions seem more like face-to-face conversations.

Computer animated characters come in many different forms. Film characters require a substantial amount of off-line animator effort to achieve high levels of quality; these techniques are not suitable for real time applications and are not the focus of this chapter. Non-player characters (typically the bad guys) in games use limited artificial intelligence to react autonomously to events in real time. However avatars are completely controlled by their users, reacting to events solely through user commands. This chapter will discuss the distinction between fully autonomous characters and completely controlled avatars and how the current differentiation may no longer be useful, given that avatar technology may need to include more autonomy to live up to the demands of mass appeal. We will firstly discuss the two categories and present reasons to combine them. We will then describe previous work in this area and finally present our own framework for semi-autonomous avatars.

## 2. Virtual Characters

This work brings together the two areas of research in virtual characters: *avatars*, which are controlled directly by the users, and *autonomous virtual characters*, whose action and behaviour are controlled by artificial intelligence.

Virtual characters that graphically represent a human user in a computer-generated environment are known as "avatars". This idea of an avatar synonymous with a user's identity in cyberspace became accepted after the science fiction novel *Snow Crash*, written by Neil Stephenson (1992). The word "avatar" comes from the ancient language of the Vedas and of Hinduism, known as Sanskrit. It traditionally meant a manifestation of a spirit in a visible form, typically as an animal or human. Examples of modern avatars can be found in virtual worlds, online computer games, and chat rooms. A lot of work has gone into developing graphically realistic avatars; this technology is now being refined and is already commercialised. However, as Ballin and Aylett (2000) point out, believable virtual characters are the summation of two key components: visual realism and behaviour. Therefore it should come as no surprise that current research is now equally focusing on behavioural attributes such as the avatar's gait and body language, and the user's individual mannerisms as captured and expressed in their avatar.

The second thread of related research has focused on virtual characters that act independently in a virtual world. These are typically referred to as autonomous virtual characters or virtual agents, and their roots stem from the area of artificial intelligence. Unfortunately for new researcher in the field, several names for these embodied entities have appeared: examples include believable and synthetic characters or virtual agents. Autonomous virtual characters have control architectures designed to make the character "do the right thing" and these usually include a sensor-reflect-act cycle. Here the character makes its decisions based on what it can sense from the environment and the task it is performing. This is compared to other virtual character applications where decisions are based on a set of predicted outcomes. This means an autonomous virtual character needs a sensory coupling with its virtual environment. Naturally, just like any autonomous agent (such as a human or dolphin), it is fallible and will make mistakes sometimes: this could be for several reasons, such as when it might base its decision on incomplete information. However in many respects this makes the character more believable, as we do not act like gods or zombies.

The designers of architectures for autonomous animated characters have taken their inspiration from the AI agent community, and they typically fall into one of two camps. At one extreme lie traditional top-down, planner-based, deliberative or symbolic architectures that typically rely on a world model for verifying sensory information and generating actions in the virtual environment. The information is used by an AI planner to produce the most appropriate sequence of actions. A good example of an autonomous character using a deliberative architecture is that of STEVE (Johnson *et al*., 1998), a virtual tutor who acts as a mentor for trainees in maintenance of gas turbines in US navy ships or the Mission Rehearsal Exercise, a training system for peacekeepers (Rickell *et al.* 2002). Both architectures are based on SOAR (Laird *et al*., 1987), a mature symbolic AI system that makes sure the sequence of actions in the world are followed correctly. At the other end of spectrum lie autonomous control architectures that are bottom-up and come from non-symbolic AI. These are referred to as *Behavioural architectures*. These are based on tightly coupled mappings between sensors and motor responses; these mappings are often competing, and are managed by a conflict resolution mechanism. It is the many interactions between the sensed signals in the environment and internal drives that produce an overall "emergent" behaviour. Examples of behavioural approaches can be seen in Terzopoulos and Tu's (1994) fish, Ballin and Aylett's (2000, 2001) 'Virtual Teletubbies', or Grand and Cliff's (1998) 'Creatures'. In the case of the Virtual Teletubbies, a robot-based architecture was modified to recreate fictional television characters for children's entertainment, and offer a level of interaction and stimulation that could not be provided by the television programme.

Of particular interest to us are autonomous characters that can interact with people using appropriate non-verbal communication skills (Vinayagamoorthy *et al.* 2006): examples include Gandalf (Thórisson, 1998), Rea (Cassell *et al*., 1999) and Greta (Pelachaud and Poggi 2002). Many characters are also programmed with models of human social relationships that enable them to interact appropriately. Examples in this volume include Rist and Schmitt's chapter, where the characters have a model of their attitude both to other characters and to concrete and abstract objects in the world. This enables them to negotiate with other characters and establish satisfactory relationships. PACEO by Hall and Oram (also this volume) is an autonomous agent that appears to display an understanding of power hierarchies in an office environment and uses this to interact appropriately with real people.

The work we have presented up to now has made a firm distinction between characters that are directly controlled by a human user (avatars and characters in animation packages) and those that are intelligently controlled by a computer (autonomous agents). This seems a logical distinction, and one that has generally divided the research into animated characters along two general directions: those where the character has no intelligence such as avatar systems or in an animation, and intelligent virtual agents, who have some degree of self-control, such as the next generation of web hosts. The idea that an avatar could have any degree of autonomy had been seen by many researchers as foreign, or even an oxymoron. However, increasingly researchers are seeing the importance of bridging this divide. Just because an avatar represents a user, does not mean that it has no independence and cannot exhibit some autonomous behaviour. The next section will firstly discuss the motivation for this sort of semi-autonomous character and then describe a number of similar, existing systems. After that we will discuss our own approach to creating semi-autonomous characters and then describe our implementation of autonomous gaze behaviour.

## 3. Semi-Autonomous Avatars and Characters

People are constantly in motion, making often very subtle gestures, posture shifts and changes of facial expression. We do not consciously notice making many of these movements and neither do we consciously notice others making them. However, they will contribute to our subconscious evaluation of a person. In particular when an animated character lacks these simple expressive motions we clearly notice their absence and judge them as lifeless and lacking personality. We would, however, often find it hard to put our finger on what it is exactly that is missing. The behaviour itself is extremely complex and subtle: LaFrance, in this volume, gives an excellent example with her discussion of vast variation and number of meanings that are possible with as seemingly simple an action as a smile. These expressive behaviours are particularly important during conversations and social interactions.

### *3.1 Avatars and chat environments*

Eye gaze and gesture play an important part in regulating the flow of conversation, determining who should speak at a given time, whereas expressive behaviours in general can display a number of intra-personal attitudes (e.g. liking, social status, emotion). These factors mean that this sort of expressive behaviour is very important for user avatars, particularly in social chat environments. Vilhjálmsson and Cassell (1998), however, note that current graphical chat systems are seriously lacking in this sort of behaviour. Interestingly they note that the problem is not that there is no expressive behaviour but that the behaviour is disconnected from the actual conversations that are going on, and so it loses most of its meaning. This is partly due to the limited range of behaviour that is currently available but they argue that the problem is in fact a more fundamental flaw with avatars that are explicitly controlled by the user. They note four main problems with this sort of system:

1. Two modes of control: at any moment the user must choose between either selecting a gesture from a menu or typing in a piece of text for the character to say. This means the subtle connections and synchronisations between speech and gestures are lost.
2. Explicit control of behaviour: the user must consciously choose which gesture to perform at a given moment. As much of our expressive behaviour is subconscious the user will simply not know what the appropriate behaviour to perform at a give time is.
3. Emotional displays: current systems mostly concentrate on displays of emotion whereas Thórisson and Cassell (1998) have shown that envelope displays* – subtle gestures and actions that regulate the flow of a dialog and establish mutual focus and attention – are more important in conversation.
4. User tracking: direct tracking of a user's face or body does not help as the user resides in a different space from that of the avatar and so features such as direction of gaze will not map over appropriately.

Vilhjálmsson and Cassell's first two points refer to the problems with simple keyboard and mouse style interfaces while point 4 shows that more sophisticated tracking type interfaces have problems of their own. Point 3 concerns the type of expressive behaviour that is not directly relevant to the discussion on semi-autonomous avatars. The major problem with the keyboard and mouse interface is that it can only input a small amount of information at a time; it is simply not possible to control speech and gesture at the same time using only two hands. Even if it were possible to create a new multimodal input device that could allow simultaneous control of both speech and gesture, it would be too great a cognitive load for the user to be constantly thinking what to do in each modality. Even if this were not so, point 2 makes it clear that we would not know which gestures to select as so many important signals are subconsciously generated. All this suggests that traditional interfaces are too impoverished to directly control an expressive avatar. Vilhjálmsson and Cassell's answer to point 4 is to add autonomous behaviours that control the avatar's expressive behaviour while leaving the user to control the avatar's speech. This creates a new type of animated character that sits between the passively controlled avatar and the autonomous agent. In the rest of this section we will develop Vilhjálmsson and Cassell's argument that this sort of semi-autonomous avatar is important for graphical chat type situations and then describe how it can be extended to other domains.

New interfaces that track the user's face and body might seem to offer an answer to this problem. They could track behaviour without the user having to explicitly think about it and could pick up subconscious cues. However, Vilhjálmsson and Cassell's point 4 argues that for desktop systems this is not possible. The position in space of the user sitting at a computer is very different from that of the avatar, and so their actions will have different meanings. For example, the user will generally look only at their computer screen while the avatar should shift its gaze between its different conversational partners. Vilhjálmsson and Cassell suggest that this sort of interface is only suitable for immersive systems.

However, even here there are problems: clearly full body tracking systems are large, expensive, and currently impractical in a domestic setting, but a worse problem is that even these complex systems are rather functionally limited. They only have a limited number of sensors and these can be noisy, thus giving only a partial view. With face tracking this is even more problematic, especially when the data must be mapped onto a graphical face that can be quite different from that of the user. These deficiencies might only introduce small errors but small errors can create a large difference in interpretation in a domain as subtle as human facial expression. There is a final problem with tracking systems; a user might want to project a different persona in the virtual world. Part of the appeal of graphical chat is to have a graphical body very different from our own. The effect of the tough action hero body would be ruined if it had the body language of the bookish suburban student controlling it.

Before leaving the subject of avatars we should briefly discuss a rather different approach suggested by Michael Mateas (1999), that he calls 'subjective avatars'. This work explores the relationship between the avatar and the user. In current narrative computer games the user tends to control a character with a strong personality and with well-defined goals in the game. However, there is little to guide the user in acting appropriately in role. Current methods tend to be crude, forcing the user down one path. Mateas' text based system uses an autonomous model of the character's attitudes to generate subjectively biased textual descriptions of events that makes the user look through the eyes of the character, instead of a more objective description that leaves the user in doubt as to how to interpret events. This is a very powerful idea potentially very important to the application of semi-autonomous avatars in games. The autonomous behaviour and interpretations of events can then give the user a stronger connection with the protagonist of the game.

### *3.2 Semi-autonomous characters in other domains*

The preceding discussion has focused on the domain of avatars for graphical chat, as this has been the field in which many of these ideas have been developed. However, those ideas are applicable to many other domains where the character does not directly represent the user. The animator generally controls animated characters directly for film but having some of the behaviour autonomously generated could greatly speed the process. This could be very useful for television where budgets are tighter than for feature films.

Moreover, computer-controlled characters do not need to be entirely autonomous. In computer games it is currently popular for the player to have allies that can be controlled indirectly through commands or requests, "Halo: Combat Evolved" is a good current example of this. Characters like these can also be classed as semi-autonomous. It might also be useful to have characters that are normally autonomous but whose behaviour can occasionally be influenced or controlled by the director of a virtual environment. This might, for instance, give a teacher the opportunity to guide a child's use of an educational Virtual Environment. Blumberg and Galyean's (1995) system is of this type.

### *3.3 Existing systems and applications*

The main problem unique to semi-autonomous avatars and characters is how to combine user input with autonomous behaviour to produce appropriate behaviour for the character. This section will discuss current solutions to this problem and applications of semiautonomous avatars and characters. The main focus of this chapter is on semi-autonomous avatars (i.e. characters that directly represent a user), however many systems described below involve other types of character. Normally the techniques used are applicable to both avatar and non-avatar characters.

There are two main approaches to combining user control with autonomous behaviour. The first is for the user to give very high-level instructions ("walk over to the door and let Jane in") and for the character to act autonomously to fulfil them. The character is normally also able to act autonomously in the world without instruction. At one extreme this type of character is manifested in graphical agents that act for the user in a virtual world where the user might not even be present. The user issues instructions or establishes a set of preferences and the agent thereafter acts autonomously to fulfil these instructions. Examples in this volume include Rist and Schmitt and also Hall and Oram. In both cases, characters act autonomously to negotiate meetings for users in an office environment.

The second approach is to leave some aspects of the characters behaviour to be controlled by the user and others to be controlled autonomously. The focus of this article is primarily on the latter, but most current work falls in the former category so we will spend rather more time discussing it. Though most systems fall into one of these two categories there is a notable exception in Mateas' subjective avatars (Mateas, 1999) described above. In that system, the character's behaviour is entirely controlled by the user but the autonomous system attempts to influence the user into acting in character.

Another important aspect of a semi-autonomous character is the type of behaviour that is produced autonomously. Expressive behaviour such as gesture, facial expression or eye gaze has been studied by researchers such as Cassell, Vilhjálmsson and Bickmore (Vilhjálmsson & Cassell, 1998; Cassell *et al*., 2001), Poggi and Pelachaud (1996), Fabri, Moore and Hobbs (this volume), Coulson (this volume), and ourselves. However, it could really be any type of behaviour that is produced currently by autonomous agent; path planning and object manipulation are popular examples.

The final factor we will consider in these systems is the method of user input. Keyboard and mouse are of course popular. Users could directly manipulate the character's body with the mouse, or they could manipulate higher-level features using menus, sliders or other GUI elements. Language-based control is also popular, whether via keyboard, or speech-based. This takes two forms. Firstly, graphical chat, as in Vilhjálmsson and Cassell, where the user enters the text to be spoken and the character autonomously generates non-verbal behaviour based on it. The other type is to give the character high-level linguistic commands, which the character then acts on. Finally, the user's face or body could be tracked and this information, rather than being directly being mapped onto the character, could be interpreted and used as input to an autonomous behaviour generation system. This approach may be promising but there has been little work on it so far, see

(Vinayagamoorthy *et al*., 2004) for an example. Barakonyi and colleagues (2002) extract MPEG-4 facial action parameters by tracking the user's face, these are used as input to an action generator for their character. This information is then used to reproduce the same emotion etc. but the character might not express it in the same way as the user would have.

Based on these categories the current work can be divided into three main types, discussed below. The first two concern high-level control of autonomous characters while the last has the user and the computer controlling different modalities in an avatar.

*Multi-layered control*

Blumberg and Galyean (1995) introduced an autonomous character that could be controlled on a number of different levels, from low-level instructions (for example, issuing commands that directly move parts of their body) to very high-level changes to the characters internal state (for example, making the character more hungry). It is a technique that is generally applied to non-avatar characters but may also be applicable to avatars. Multi-layered control architecture have been popular; for example, Caicedo and Thalmann (2000) created a character that could be controlled by issuing instructions or altering its beliefs. An interesting feature of this system is that it contains a measure of how much the character trusts the user, which influences whether it will carry out the user's instructions. Musse and colleagues (1999) have applied a multi-level system to controlling crowds. Paiva, Machado and Prada (2001) combine direct control of an autonomous character with a more reflective level of control which takes users out of the virtual world allowing users to update the internal state of their character. Carmen's Bright IDEAs (Marsella *et al*., 2000) uses high-level control of the character. Interestingly, the user influences the character's internal state but does not do so explicitly, rather they choose one of three thought bubbles which reflect different state changes. This system will be discussed further in the section on inference below.

*Linguistic commands*

An obvious way of controlling behaviour of avatars and characters is to give them commands in natural language. For example, Badler and colleagues (2000) implemented linguistic control for avatars in a multi-user VE, and for military training scenarios. Also Cavazza and colleagues (1999) used natural language to control the player character in a computer game modeled on id software's "Doom".

*Text chat*

We have already discussed this example at length. The user's only input is the text that the avatar should say. Appropriate non-verbal communication behaviour is generated autonomously based on this text. In Vilhjálmsson and Cassell's BodyChat (Vilhjálmsson & Cassell, 1998) the avatar produces suitable eye gaze and facial animation to regulate the flow of a conversation. In BEAT and Spark, their follow-up systems (Cassell *et al*., 2001; Vilhjálmsson 2005) they analyse text and determine which gestures should be produced at which particular moments in the text. Similarly, the eDrama system analyses text to extract emotional information that is used for animating avatars (Dhaliwal *et al.* 2007). Poggi and Pelachaud (1996) have done similar work for faces. Gillies and Ballin (2004) use off line customisation, real time commands and recognition of emoticons to control non-verbal behaviour. Similar methods can also be used for voice, rather than text, interaction. Vinayagamoorthy *et al.* (2002), use an autonomous model of gaze that is triggered by speech in a two part conversational setting. Cassell and Vilhjálmsson, in their evaluation work for BodyChat (Cassell & Vilhjálmsson, 1999), discovered that users find the character's behaviour more natural when it is animated autonomously as opposed to when they can control its animation. More surprising was the finding that subjects also felt more in control of the semi-autonomous character. This result is probably due to the fact that users feel overwhelmed at having to control the character's non-verbal behaviour whereas in a semi-autonomous system they can concentrate on the content, such as the speech.

### *3.4 Future developments*

In this section we will describe a number of potential research directions for semi-autonomous avatars and characters. As described earlier the central research problem for semi-autonomous avatars as opposed to other types of agent is the integration of autonomous behaviour and user control. The three areas of research above address this in one of the following ways:

### *Selective autonomy*

Multi-user virtual environments are becoming increasingly heterogeneous, with users of different skill levels accessing them through machines with different capabilities and different interaction devices. Therefore practical semi-autonomous avatar systems should be designed so each user can select which parts of the avatar's behaviour is generated autonomously and which are directly controlled, making the set of possible avatars a continuum between complete autonomy (for agents in the world) to complete user control. For example, a world might contain non-user agents which are completely autonomous; text based users whose avatars have autonomous expressive behaviour and also largely autonomous navigation behaviour; desktop graphical users whose expressive behaviour is autonomous but whose navigation behaviour is controlled with the mouse, and finally fully immersed and tracked users whose body motion is directly mapped onto the avatar.

### *Inferring avatar state*

In order to generate appropriate non-verbal behaviour for an avatar, it is useful to know certain things about the internal state of the avatar/user; for example, are they happy, do they like the person they are talking to? One approach might be to use whatever limited input comes from the user to infer what kind of internal state to project, for example, by analysing the text that the user types. This is of course a hard problem and could easily lead to very inappropriate actions due to incorrect inferences. However, it has the potential to greatly improve the experience. Existing systems such as Spark (Vilhjálmsson 2005) or eDrama (Dhaliwal 2007) use analysis of typed text to infer certain conversational or emotional states of the user. Marsella's Carmen's Bright IDEAs (Marsella *et al*., 2000) supports this type of inference in an interesting way. The user is asked to choose an appropriate thought bubble to represent what the character is thinking. These thought bubbles correspond to changes of internal state but do not expose the user directly to the internal workings of the system.

### *End-user personalisation*

Semi-autonomous avatars should reflect what the user wants them to do as closely as possible and yet with minimum of input from them. One way of trying to achieve this is to put some of the work of user control off-line by allowing the user to extensively customise the behaviour of the character before they start to use it. Users of graphical chat systems are very keen to personalize their avatar's appearance (Cheng *et al*., 2002), and there is no reason to believe that this would not be true of behaviour as well. This means not only that avatar behaviour should be very customisable but also that the tools for customizing behaviour should be easy to use for non-expert users. This second requirement is difficult as AI behaviour generation systems are complex and not very easy to understand. Our system, described below takes a few steps in the direction of building such a tool. Gillies (2006) provides a more complete tool for customising avatars. A different approach that is attracting much interest is the development of mark-up languages that can be used to design the behaviour of virtual humans. Ruttkay and colleagues provide one particularly interesting example in this volume. Their GESTYLE language provides four levels of mark up for specifying differences in style of non-verbal communication between virtual characters.

## 4. A Model for Semi-Autonomous Avatars

We propose a model of semi-autonomous avatars and characters in which the user controls different aspects of the behaviour from the autonomous system. Our model ensures that the autonomous behaviour is influenced by the actions the user performs. This is similar to systems where the user types text and the system generates non-verbal behaviour, however, we allow the user to control certain animated actions while leaving the others autonomous. We divide behaviour into *primary behaviour*, which consists of the major actions of the character, and is controlled by the user and *secondary behaviour* that is more peripheral to the action but may be vital to making the avatar seem alive.

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. In our system the primary behaviour can be *tagged* so as to provide a way of synchronising the secondary behaviour. 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 module (or set of modules) that is not directly influenced by user input and which acts to a large degree autonomously. To ensure that the secondary behaviour is appropriate to the primary behaviour it is influenced by messages sent from the primary behaviour module. These messages contain instructions for the secondary behaviour to change appropriately based on the state of the current primary behaviour. Various points in the primary behaviour are assigned tags that result in a message being sent when that point is reached. The tags contain the content of the message. 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. The tags should be probabilities of sending a message and the parameters of the message should also be expressed as probabilities. This ensures that behaviour is not entirely deterministic and so does not seem overly repetitive.

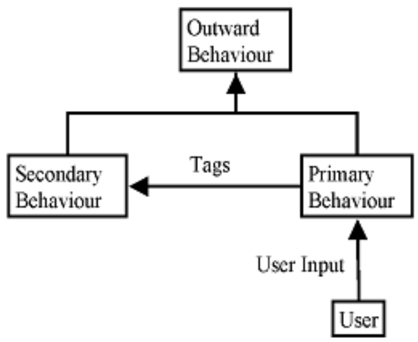



*Figure 1: The relationship between primary and secondary behaviours.*

There are two ways in which the tags could be edited. The first is when a designer of a virtual environment would want to design the behaviour traits of the characters in their environment. This would be a professional, trained in using the editing package. The end-user would also want to customise the behaviour of their particular avatar. They, however, would require easy-to-use tools and less ambitious edits. Designers could be given a tool that allows complete control of tags, allowing them to place the primary behaviour tags and edit all of their content. The end-user would be given a tool with more limited control, merely altering certain parameters of the tags, without changing their position. For example, the designer might add a tag requesting that the avatar should look at the conversational partner at the end of an utterance. The end-user might then indicate whether this should be a brief glance with just the avatar's 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.

*4.1 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. Eye gaze is a very expressive part of human behaviour and one of the most important cues we use when “reading” other people. This is of course true of gaze between people in social situations such as conversations, giving envelope cues such as that for turn-taking behaviour as well as giving information about social attitudes such as liking. There has been extensive work on simulating this use of gaze, for example (Vilhjálmsson & Cassell, 1998; Colburn *et al*., 2000; Vinayagamoorthy *et al*., 2004). However, non-social uses of gaze can also be important in interpreting people’s behaviour. What a person is looking at gives a strong indication of their intentions and what they are thinking about. Having a character look at an object before reacting to it makes clear what the reaction was to and so makes the characters behaviour easier to understand. Non-social gaze has been studied by Chopra-Khullar and Badler (1999) but they did not investigate in detail how to integrate simulation of gaze with user control of the avatar’s actions. We focus on creating a tool by which a user without programming knowledge can create both primary actions that an avatar can perform as the user requests it, and secondary gaze behaviour that will accompany these primary actions, as summarised in Figure 2.

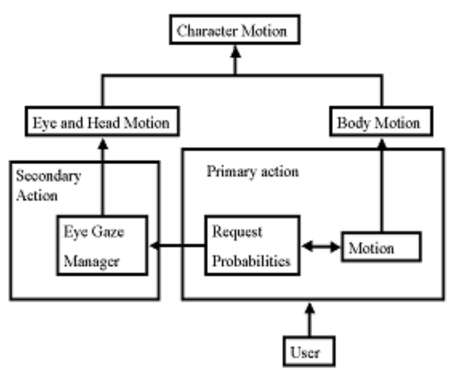



*Figure 2: Primary and secondary behaviours for the gaze example.*

Our primary behaviour consists of simple actions that an end user can invoke in real time. Each action has one or more targets, which are objects that the character interacts with during this activity. For example, a target for a drinking motion would be a cup. The user would invoke the action by clicking on a possible target. Our aim is to make it easy for the designer of a virtual environment to design a new action. The designer first chooses a piece of motion on which to base the actions and adds some mark-up information. They then designate targets for the action. When the action is invoked the motion is transformed using motion-editing techniques (see Gleicher, 2001, for an overview) to be appropriate to the new position of the target. For a more detailed description of the primary behaviour see (Gillies, 2001).

Secondary behaviour consists of gaze shifts that are controlled by an eye gaze manager described in more detail in (Gillies & Dodgson, 2002). The manager can generate eye gaze autonomously and react to events in the environment. The eye gaze can be controlled by sending requests for gaze shifts to the manager, causing the character to look at the target of the request. The gaze behaviour can be controlled by editing one of two types of parameters. Firstly there are parameters that control the character's behaviour as a whole. For example, observing people we noticed that they vary their horizontal angle of gaze but keep 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. Setting the parameters in advance allows some end-user customisation of the behaviour. The second type of parameter is attached to a request, changing the way in which the character looks at the target of the request, for example, changing the length of gaze.

As described above the primary behaviour is tagged with messages that are sent to the secondary behaviour module. In this case the messages consist of eye gaze requests. The designer of the action will add tags to various points in the original motion. These tags will contain a request to gaze at one of the targets of the action, as well as the probability of sending that request. When that point in the motion is reached the request will be sent with that probability, ensuring that eye gaze can be synchronised with the motion. Values for the parameters of the request can also be specified, allowing finer control of the gaze behaviour. The designer can also specify what the parameters of the tags including the probabilities can be edited by the end user. This allows the end user to perform a certain degree of customisation. These parameters are set with a simple interface consisting of a slider for each parameter.

*Results and evaluation*

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 it up and then at the other avatar shown in the last frame, this time just glancing and moving its eyes without turning its head. This behaviour might indicate avoiding the gaze of the other avatar, which would have a strong intra-personal meaning. 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 in the middle of the sequence it looks at a location in the distance.

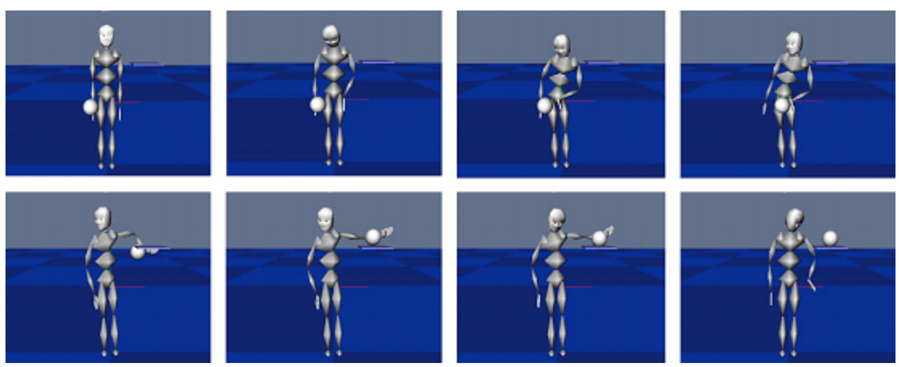


*Figure 3: An action of an avatar drinking from a can.*

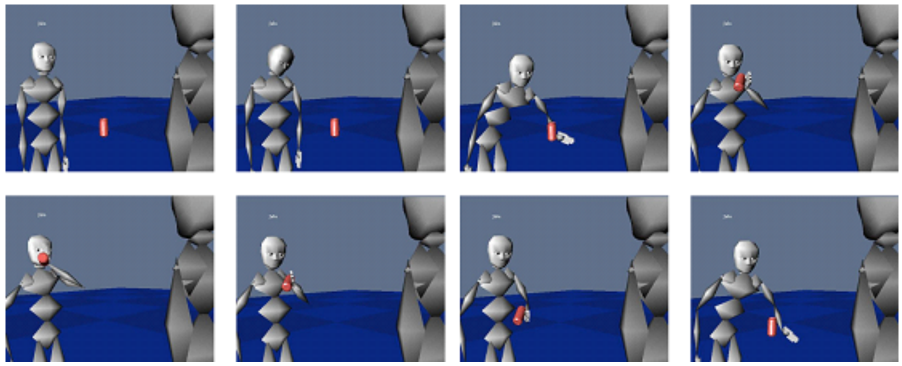
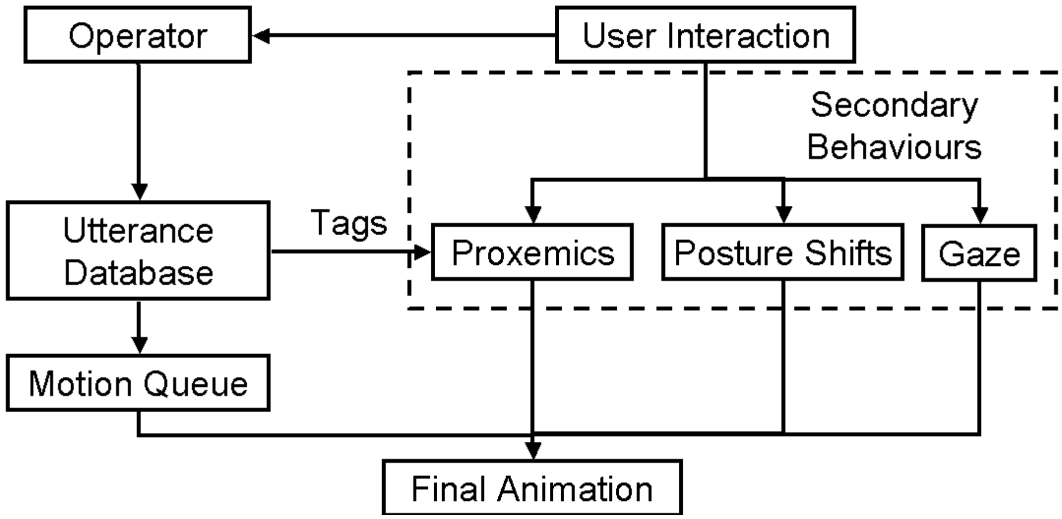


*Figure 4: An action of an avatar picking up an object and putting it down somewhere else.*

This is a first prototype of this framework, and we are not yet ready to do a formal evaluation. In our opinion the quality of the behaviour is reasonable but could be improved through more careful tagging of the primary behaviour. People viewing the system informally have reported that they consider the addition of eye gaze to add life to the characters and the connection to the primary behaviour gives a stronger sense of intentionality to the character.

Both semi-autonomous avatars in general and our particular system have a large potential for further development. As our system is a general framework, there is a potential to apply it to many different domains and different types of secondary behaviour. There are also specific improvements that could be made to our current implementation. 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. As the work focuses on animated actions rather than conversation it would be better suited to a task-based environment than a purely social one. This could form the basis of a formal evaluation of the system. An experiment could be run to compare the user’s experience with and without the use of secondary behaviour. The experiment might involve a task that consists of collaboratively manipulating the world using a repertoire of actions.

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 affect the parameters, some of which are rather counter-intuitive: 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 and we would therefore like to include more heuristics for social situations.

### *4.2 A Conversational Character*

The framework we have presented is applicable to a number of different uses of characters. This section will briefly describe another application to a character that is able to have a conversation with a real person in an immersive virtual environment. The character is designed for use in virtual reality experiments. The conversation itself is controlled in a "wizard of oz" manner. This application is closely related to the text chat avatars discussed earlier as the character is controlled by a human operator. However, the operator, rather than creating arbitrary textual responses chooses from a number of pre-recorded audio files of speech responses.

*Figure 5: The Architecture for a conversational character*

Figure 5 shows the architecture of the character. As in our previous example the characters behaviour consists of Primary Behaviour that is triggered by the operator and Secondary Behaviour that occurs largely autonomously in parallel to the Primary Behaviour. In this case the Primary Behaviour consists of a set of multi-modal utterances that the operator can choose via a graphical user interface, in response to the speech of the user that is interacting with the character. A multi-modal utterance consists of an audio clip containing speech but can also contain other animation elements such as gestures and facial expressions. The secondary behaviour consists of a number of components that respond directly, and in real time, to the behaviour of the user. The user that is interacting with the character has their position tracked and their voice recorded with a microphone. The secondary behaviours can respond in a number of ways to these inputs. The character has three secondary behaviours:

- Proxemics: the character maintains a comfortable conversational distance to the user, stepping forward if the user is too far away or backward if they come too close based on the position tracker.
- Posture Shifts: the character will shift posture occasionally. It will attempt to create a rapport with the user by synchronising its posture shifts with those of the user. This is done by triggering a shift when a large movement is detected from the position tracker.
- Gaze: the character contains a gaze model based on that of Vinayagamoorthy *et al.* (2004). This model changes the degree of gaze at the user depending on whether the character is talking or listening to the user (as detected by the microphone).

As well as directly responding to the user the secondary behaviour can also be influenced by the multi-modal utterances selected by the operator. As described in the previous example, the utterances can be tagged with information about the parameters of the secondary behaviours and how they can be changed. For example, a more intimate topic of conversation can be tagged with a closer conversational distance for the Proxemics behaviour. Similarly any significantly long speech will change the level of gaze at the user in the Gaze behaviour.

This architecture has been used for characters in a number of different experiments (figure 6 shows an example). The use of Secondary behaviours has proved very helpful in the experimental setting. Firstly, it makes it possible to have

a very rich set of behaviour without overloading the operator with excessive work. Secondly, the Secondary Behaviours can respond instantly to the actions of the users without a lag created by the operators response time. This makes it possible to create responsive effects like synchronization of posture shifts that would be otherwise impossible.

*Figure 6: A conversational character interacting with a human user.*

## 5. Conclusion

We have given an overview of the reasons why semi-autonomous avatars and characters are an important research area, described current research, and suggested possible future directions. We have also presented a framework for semi-autonomous characters, and described an application of this framework 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

Some of this work was done at Cambridge University Computer Laboratory and funded by the UK Engineering and Physical Sciences Research Council. The rest of this work is funded and carried out at BT and at University College London, funded by the UK Engineering and Physical Sciences Research Council. The authors would like to thank the members of the Cambridge University Computer Lab Rainbow research group, the Radical Multimedia Lab, UCL Virtual Environments and Computer Graphics group, Mel Slater and Tony Polichroniadis for their support and suggestions.

## Notes

[*] There is often a distinction made between envelope and emotion in expressive behaviour. We wonder if there is another type of behaviour that is less basic to conversation than envelope behaviour but more important in day-to-day conversation than emotional expressions. This is the

sort of behaviour that expresses and influences intra-personal attitudes and relationships. Whereas envelope behaviour controls the low level, moment-by-moment details of the conversation, intra-personal behaviour might control the high-level relationships between the speakers. Examples might be expression of liking or social status. There could also be more short-lived examples such as behaviour that encourages another speaker to express an opinion or behaviour involved in trying to win an argument.

† Though this point is not generally mentioned in the literature it is actually very important. If an avatar's head is made to move vertically too much it looks very wrong.

## References

Badler, N., Bindiganavale, R., Allbeck, J., Schuler, W., Zhao, L. and Palmer, M. (2000). Parameterized Action Representation for Virtual Human Agents. In J. Cassell, J. Sullivan, S. Prevost, & E. Churchill (Eds.), *Embodied Conversational Agents* (pp. 356-284). Cambridge, MA; MIT Press.

Ballin, D. & Aylett, R.S., (2000). Time for Virtual Teletubbies: The development of Interactive and Autonomous Children's Television Characters. In *Proc. Workshop on Interactive Robotics and Entertainment* (pp. 109-116). Carnegie-Mellon University, April 2000.

Ballin, D., Aylett, R.S. & Delgado, C. (2001). Towards the development of Life-Like Autonomous Characters for Interactive Media. In *Procs. BCS Conference on Intelligent Agents for Mobile and Virtual Media, National Museum of Film and Photography*. Bradford, UK, 2001.

Barakonyi, I., Chandrasiri, N. P., Descamps, S. & Ishizuka, M. (2002). Communicating Multimodal information on the WWW using a lifelike, animated 3D agent, In H. Prendinger (Ed.), *Proceedings of the PRICAI workshop on Lifelike Animated Agents* (pp. 16-21). Tokyo, Japan, August 19, 2002.

Blumberg, B. & Galyean, T. (1995). Multi-Level Direction of Autonomous Creatures for Real-Time Virtual Environments. In *Proceedings of ACM SIGGRAPH 1995* (pp. 47-54). ACM Press.

Caicedo, A. & Thalmann, D. (2000). Virtual Humanoids: Let Them be Autonomous without Losing Control. In *Proc. of the Fourth International Conference on Computer Graphics and Artificial Intelligence*. Limoges, France. May 3-4, 2000.

Cassell, J., Bickmore, T., Campbell, L., Chang, K & Vilhjálmsson, H. H. (1999). Embodiment in Conversational Interfaces: Rea. In *Proceedings of ACM SIGCHI 1999* (pp. 520-527). ACM Press.

Cassell, J. & Vilhjálmsson, H. H. (1999). Fully Embodied Conversational Avatars: Making Communicative Behaviours Autonomous. *Autonomous Agents and Mutli-Agent Systems* 2(1), 45-64.

Cassell, J., Vilhjálmsson, H. H. & Bickmore, T. (2001). BEAT: the behavior expression animation toolkit. In *Proc. of ACM SIGGRAPH* (pp. 477-486). Los Angeles California: ACM Press.

Cavazza, M., Bandi, S. & Palmer, I. (1999). Situated AI in Video Games: Integrating NLP, Path Planning and 3D Animation. In *Proceedings of the AAAI Spring Symposium on Computer Games and Artificial Intelligence* [AAAI Technical Report SS-99-02]. Menlo Park, CA: AAAI Press.

Cheng, L., Farnham, S., & Stone, L. (2002). Lessons Learned: Building and Deploying Virtual Environments. In R. Schroeder (Ed.), *The Social Life of Avatars: Presence and Interaction in Shared Virtual Worlds* (pp. 90-111). Heidelberg & Berlin: Springer.

Chopra-Khullar, S. & Badler, N. (1999). Where to look? Automating visual attending behaviors of virtual human characters. In *Proceedings of the 3rd Autonomous Agents Conference* (pp. 16-23). ACM Press. Dhaliwal, K., Gillies, M., O'Connor, J., Oldroyd, A., Robertson, D., and Zhang, L. (2007) eDrama: Facilitating online role-play using emotionally expressive avatars. in *The Proceedings of the AISB Workshop on Language, Speech and Gesture for Expressive Characters* (P. Olivier and R. Aylett eds) 2007

Gleicher, M. (2001) Comparing Constraint-Based Motion Editing Methods. *Graphical Models,* 63, 107-134

Gillies, M. (2001) Practical behavioural animation based on vision and attention Cambridge University Computer Laboratory technical report UCAM-CL-TR-522

Gillies, M. & Dodgson, N. (2002). Eye Movements and Attention for Behavioural Animation. *Journal of Visualization and Computer Animation*, 13(5), 287-300.

Gillies, M. & Ballin, D. (2004), Integrating Autonomous Behavior and User Control for Believable Agents. In *Proc. Intl. Conf. on Autonomous Agents and Multi-Agent Systems (AAMAS2004)*. New York, NY: ACM Press.

Gillies, M. (2006) Applying direct manipulation interfaces to customizing player character behaviour. in *The Proceedings of the International Conference on Entertainment Computing* 2006 Springer Lecture Notes in Computer Science

Grand, S. & Cliff, D. (1998) Creatures: Entertainment software agents with artificial life. *Autonomous Agents and Multi-agent Systems*, 1(1), 39-57

Laird, J., Newell, E. & Rosenbloom, P. (1987) Soar: An architecture for general intelligence. *Artificial Intelligence*, 33(1), 1-64.

Marsella, S. C., Johnson, W. L. & LaBore, C. (2000). Interactive Pedagogical Drama. In *Proceedings of the 4th international Conference on Autonomous Agents* (pp. 301-308). ACM Press.

Mateas, M. (1999). Not Your Grandmother's Game: AI-Based Art and Entertainment. In *Proceedings of the AAAI Spring Symposium on Computer Games and Artificial Intelligence* [AAAI Technical Report SS-99-02]. Menlo Park, CA: AAAI Press.

Musse, S. R., Garat, F. & Thalmann D. (1999), Guiding and Interacting with Virtual Crowds in Real-Time. In M. Magnenat-Thalmann & D. Thalmann (Eds.), *Proceedings of the Eurographics 1999 workshop on Computer Animation and Simulation* (pp. 23-34). Milan, Italy, September 7–8, 1999.

Paiva, A., Machado, I. & Prada, R. (2001). The child behind the character. *IEEE Transactions on systems, man and cybernetics: Part A*, 31, 361–368.

Pelachaud, C. and Poggi, I. (2002) Subtleties of Facial Expressions in Embodied Agents, *Journal of Visualization and Computer Animation*, vol. 13, 301-312, 2002.

Poggi, I & Pelachaud, C. (1996). Context Sensitive Faces. In R. Campbell & C. Benôit (Eds.), *Proceedings of the ESCA Workshop on Audio-Visual Speech Processing* (pp. 17-20). Rhodes, Greece, 24-27 September 1996.

Rickel, J. Marsella, S., Gratch, J., Hill, R., Traum, D., and Swartout, B. (2002) Towards a New Generation of Virtual Humans for Interactive Experiences, in *IEEE Intelligent Systems* July/August 2002, pp. 32-38

Stephenson, Neil (1992). *Snow Crash*. Bantam Books.

Terzopolous, D., Tu, X. & Grzeszczuk, R. (1994) Artificial fishes: Autonomous, Locomotion, Perception, Behavior, and Learning in a simulated physical world. *Artificial Life*, 1(4), 327-351.

Thórisson, K. (1998). Real-time Decision Making in Multimodal Face-To-Face Communication. In *Proc. Second Intl. Conference on Autonomous Agents* (pp. 16-23). Newy York, NY: ACM Press.

Vilhjálmsson, H. H. & Cassell, J. (1998). BodyChat: Autonomous Communicative Behaviors in Avatars. In *Proc. Second Intl. Conference on Autonomous Agents* (pp. 269 – 276). Newy York, NY: ACM Press.

Vilhjalmsson, H. (2005),Augmenting Online Conversation through Automated Discourse Tagging, *6th annual minitrack on Persistent Conversation at the 38th Hawaii International Conference on System Sciences*, January 3-6, 2005, Hilton Waikoloa Village, Big Island, Hawaii, IEEE, 2005

Vinayagamoorthy V., Garau M., Steed A., and Slater M. (2004) An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience, *Computer Graphics Forum*, 23(1), pp. 1-11, (March 2004)

Vinayagamoorthy, V., Gillies, M. , Steed, A., Tanguy, E., Pan, X., Loscos, C., and Slater, M. (2006) Building Expression into Virtual Characters. in *Eurographics Conference State of the Art Reports* 2006