Intelligent Virtual Environments

Ruth Aylett
Centre for Virtual Environments
University of Salford
Salford, M5 4WT
United Kingdom

Michael Luck
Department of Computer Science
University of Warwick
Coventry, CV4 7AL
United Kingdom

1 Introduction

In a period when niche areas of cutting-edge technological research are capturing the public imagination and moving out of the laboratory into everyday life, at least in terms of an *awareness* of the technology and the possibilities it raises if not its common application, this broader impetus can be one of the key ingredients in a recipe for dramatic progress. Though the label *dramatic* may be regarded as excessive in a quantitative sense, it is very apt in the literal sense. Indeed, the vision of exciting applications can be regarded as a driving force behind the premise of this paper that a convergence has begun to take place between branches of advanced computing and research communities which, until recently, were quite separate — namely Artificial Intelligence (AI), Artificial Life (AL) and Virtual Reality (VR), or, as it is sometimes now known, Virtual Environments (VE). This combination of intelligent techniques and tools, embodied in autonomous creatures and agents, together with effective means for their graphical representation and interaction of various kinds, has given rise to a new area at their meeting point, which we call *intelligent virtual environments*.

A number of factors are allowing the use of virtual environments in AI and AL research just at the point when the development of particular fields of exploration, including that of intelligent and autonomous agents, makes such use an obvious step to take. First, the continuing growth in the amount of computing power that can be put on a desktop not only supports a much higher degree of visual realism, but even leaves over a little processing power that can be used to add intelligence. A second factor relates to the maturing and more widespread availability of 3D graphics software, and the development of 3D graphics standards such as VRML '97 (Hartman and Wernecke, 1996). Third, AI technologies such as natural language processing have matured in parallel with this to the point where they can be used as a means of interaction with a virtual environment.

At the same time, some researchers in the field of virtual environments and advanced graphics are seeking to progress beyond visually compelling but essentially empty environments to incorporate other aspects of physical reality that require intelligent behaviour. This may involve populating urban models with crowds (Musse and Thalmann, 1997) or traffic (Wright et al., 1998), the investigation of virtual humans (Thalmann and Thalmann, 1998) or virtual actors (Shawver, 1997; Wavish and Connah, 1997), the creation of virtual non-humans (Terzopoulos et al., 1994) or, at the more abstract level, the attempt to produce more adequate representations within VE tools for modelling behaviour and intelligence (VRML Consortium, 1998).

In this paper we discuss the main research areas of this convergence, and begin in the next section with a consideration of the basic issues underlying the enabling technologies for intelligent virtual environments. The next three section focus primarily on agents, first considering broad agent issues,

and then examining the particular concerns relevant to the physical and cognitive ends of the agent spectrum in turn. Though the area of agents is overwhelmingly where most of the effort is currently being directed, and demands most attention, we also then focus on virtual worlds themselves. Finally, we conclude with a discussion of the research issues and the problems yet to to be addressed and consider possible future directions. At each point, we illustrate the issues with reference to systems and applications.

1.1 Basic Constraints

Virtual environments have at least one thing in common with robots. This is the need to respect real-time processing constraints. A VE is a system driven by a rendering cycle that ideally works at 50 or 60 Hertz so that change appears as smooth animation rather than a series of jerks. At frame rates below 10 Hertz, it becomes impossible to sustain the illusion of physical reality that is so important to the feeling of *presence* for the user of the VE.

In a given cycle, the rendering algorithm traverses a *scene graph* — usually a hierarchical structure in which all the components of a VE are represented using nodes of different types linked together. The more complex the scene graph, the longer it takes to traverse, and the harder it becomes to maintain a high frame rate. Researchers working on the development of VEs are very conscious of this problem, and much individual effort has targetted the creation of visually appealing components with as low a polygon count as possible, as well as general mechanisms, such as Level of Detail (LOD), that allow components to be represented by successively more detailed models as the user of the VE approaches them.

Just as in robotics, however, adding intelligence potentially steals processing power from the basic cycle. This is literally true where added intelligence uses the same processor, but distribution onto extra processors may have the same effect if the parallel processing is not accurately synchronised with the frame rate. A number of the systems discussed below cannot render in real-time but must instead render off-line and run subsequently as animations — an approach that precludes the normal interactive use of the VE. Thus, although the growth in processing power and the development of improved algorithms makes it possible to add intelligence to VEs, it is still currently inadequate for many of the systems that are being developed in research laboratories.

1.2 Virtual Environment Tools

1.2.1 The Level of Support

The development of intelligent VEs is even further constrained by the bias of most generally available VE toolkits towards visual realism and the graphical support of the VE, rather than towards the addition of intelligence.

At the lowest level, a system might be developed using, for example, Open GL, or some other 3D library system and C++ but, as usual, flexibility is traded-off against the time and effort required to achieve the desired functionality. One example of VE development at this lowest level is the AReVi toolkit (Reignier et al., 1998) which offers a set of C++ classes built around an *agent* programming language, *oRis* (though the sense of *agent* is in this context a weak one, and corresponds to the notion of a *programming entity* much more than to a *virtual agent* in the sense discussed below).

At the next level of abstraction, VE toolkits use the scene-graph representation already referred to. This is a convenient way of representing the graphical aspects of objects, since leaf nodes in the scene-graph normally represent graphical primitive objects as a collection of polygons. Such primitives are

then grouped into more complex graphical objects using group nodes to which the components are attached.

1.2.2 Incorporating Knowledge Representation

However, if we wish to attach knowledge to objects — and in particular to manipulate objects at a knowledge level, the scene-graph representation is much less convenient, since it is not always clear how conceptual objects can be mapped onto collections of graphical primitives. As a consequence, it has been argued that it is now time for the designers of VE toolkits to consider the incorporation of explicit knowledge representation facilities (West and Hubbold, 1998), an area in which AI has much to offer, not least in helping to avoid the reinvention of a large number of knowledge representation wheels.

The orientation of existing toolkits to graphical representations and to the visual perspective of the VE user can be seen in other ways, too. In most toolkits, the object representing the VE user has a privileged position in the system so that, for example, in VRML'97 the user is provided with automatic facilities for the detection of collisions with parts of the environments (using bounding boxes), while this must be explicitly programmed for any other object. Support for animation is widely provided, but this is oriented towards trajectories calculated by the designer in advance (automatic interpolation between way-points is usually provided), rather than to the autonomous motion of objects driven by virtual sensors.

Indeed, while many VE toolkits provide sensors, these do not correspond to a virtual representation of the type of sensing that a robot might carry out, but to facilities for detecting user interaction, such as an alarm when a wall is hit or the generation of events in response to a mouse-click. The addition of interesting behaviour in the AI sense normally requires direct programming via whatever language facility the toolkit supports (which is typically C++ in many proprietary toolkits, or Java in the case of VRML).

VRML'97 has incorporated Script nodes into its scene-graph representation in order to provide a clean interface to user-added functionality, and this is the standard method for adding behavioural complexity to VRML applications. However, criticisms of this approach include the amount of inter-node traffic that is generated for interesting behaviour, and there is at least one proposal for the packaging of a neural net into a VRML node (Nerves, 1998) in order to make behaviour more responsive without a large inter-node routing load.

1.2.3 Interaction with Complex Properties

If we wish to attach more complex properties than visual appearance to objects in a VE, a further issue arises relating to the ways in which an object and a VE interact. Visual interaction in the classical sense takes place largely between a VE object and the VE user, and is encompassed by the overall visual appearance of an object to the user, including texture, lighting effects and level of detail. In a standard VE, objects only interact visually with each other insofar as one hides another from the user's view. However, once more complex properties are introduced, the amount of interaction between VE objects and between objects and the VE itself, increases sharply. The question that needs addressing is how far such interactions should be driven by the object and how much by the environment in which the object is located.

As yet, there is no clear resolution of this problem. One approach is to embed, inside the object, the properties and the necessary knowledge of how they interact with an environment. For example, the IMPROV system (Goldberg, 1997) adopts what Goldberg calls *inverse causality*, and stores ani-

mations of the interaction between an object and a virtual actor within the object, in order to remove any learning requirement from the virtual actor. Thus, a virtual actor that points at a virtual bottle of beer is given the option of drinking from it without having to learn the necessary actions for so doing.

This seems a little counter-intuitive when considering the incorporation of aspects of physics — such as gravity (Aylett et al., 1999) — into a VE. Here, it seems preferable that all objects placed within a VE should obey whatever physical *laws* are current, whether this is falling downwards if unsupported in the case of gravity, or floating in the case of non-gravity. An example of the same type might be how a fish should behave if it is placed in a VE that is not full of water (West and Hubbold, 1998). One approach might be to provide properties — such as *force* — that allow an object to interact in a sensible way with whatever VE in which it has been embedded. It seems clear that work in AI on common ontologies may have an important application here.

In conclusion, it must be pointed out that VE toolkits were never intended to support the kind of functionality discussed in this paper, so that it would have been very surprising if the difficulties identified here had not existed. The creation of a new generation of VE tools is an enterprise in which the two communities of AI and VE might profitably collaborate, and work is already in progress within the VRML consortium (VRML Consortium, 1998) that is considering the future of VRML.

2 Autonomous Agents

The convergence between AI and AL on the one hand and VEs on the other is nowhere so obvious as in the area of agents. Autonomous agents, as a research field in AI, has burst into a frenzy of activity in the last five years or so, as indicated, for example, by the increasing number of workshops and conferences, and the large number of active research groups (Aylett et al., 1998). We distinguish here between autonomous agent research and the more general field of multi-agent systems (Luck, 1997; Luck et al., 1998). The latter area encompasses distributed problem-solving applications such as network management which do not typically involve VR or virtual environments and has a focus on inter-agent communication and negotiation that may not be required in the autonomous agents sub-field. In this paper we are focussed specifically on work using VEs as a technique for exploring agent behaviour and agent believability, or agents as a way of extending VEs into new application areas. This includes synthetic agents, virtual actors, virtual humans, as well as avatars (which are physical representations of users) in 3D multi-user web environments. We will discuss work with a starting point in the area of VEs as well as work beginning from issues in AI.

2.1 Autonomy

The notion of autonomy has become increasingly important and studied in relation to agents that must function effectively and independently in a dynamic environment. A range of work has attempted to consider many issues concerned with agent autonomy, including its nature what it entails (Luck and d'Inverno, 1995), how it may be determined by agent architecture (Castelfranchi, 1995) and the subtleties of its use by different researchers (Franklin and Graesser, 1996), for example. A question that immediately arises is whether *autonomy* is useful and appropriate for agents in virtual environments in the way it is for agents in the real world. In the real world, the environment functions independently of the agents within it — an individual agent can only perceive part of it (and may be wrong about what it does perceive) and is subject to independent processes and the activity of other agents. Under these circumstances, predictions about the world are always likely to be fallible. Autonomy is an appropriate response because leaving the agent to decide its actions allows it to take account of the

current — rather than a predicted — state of the world.

In a virtual environment, the situation is very different. The designer has a 'gods-eye' view of both the environment and the agent, and need not distinguish between them. Moreover, the whole environment is available to the agent — there need be no difference between its model of the virtual world and the virtual world itself. Autonomy might appear a needless overhead from a practical perspective, and only useful as a basis for more scientific investigations of agenthood.

However, as discussed in (Petta and Trappl, 1997), the omniscient approach to virtual agents in fact turns out to be very inefficient. The problem is that if virtual agents are to behave in a way that is convincing to the user and sustains the feeling of presence in a virtual world, they ought to appear to have the same limitations as agents in the real world. They ought to seem to collect information as it becomes available to them and to interact with objects — noticing, avoiding and manipulating — in a plausible manner. Omniscient agent management soon runs into combinatorial problems when it must keep track of what each agent is supposed to know and perceive. It is far simpler to equip each agent with virtual sensors (Thalmann et al., 1997) and use these to autonomously drive their *physical* manifestion as virtual effectors. Thus, most of the work concerned with virtual agents follows the autonomous approach.

There are other implementation-level advantages of applying the autonomous approach, especially the potential for reuse of agents in different VEs and the ability to distribute individual agents over separate processors. However these remain theoretical to a large extent with little evidence of reuse or distribution in practice, perhaps reflecting the current immaturity of the field and the diversity of the problems being tackled. It is also true that while autonomy may be a prerequisite for reusable agents, it is far from sufficient, with many representational issues of agent, environment and interaction yet to be successfully tackled.

2.2 The Spectrum of Agents

In order to divide the very large number of systems in some tractable way, we imagine a spectrum of agency. At one end of this spectrum, we place physical agents, by which we mean agents where the focus is on believable physical behaviour in a virtual environment. Topics here include realistic movement and physical interaction with the environment — for synthetic animals as well as humans — in addition to body language, gesture and facial expression. Such agents normally interact with a VE through virtual sensors working at a non-symbolic level.

At the other end of this spectrum, we place agents where the focus is on human cognitive behaviour and on cognitive interaction with the human user of the system. Many of the topics here are related to natural language and cognitive processes such as planning. Such agents often *sense* symbolic information directly from the VE and it is sometimes less obvious how far they can be said to have an autonomous perceptual apparatus.

We speak of a spectrum rather than mutually exclusive categories because more cognitive agents usually require some degree of physical interaction with a VE while more physical agents often require some kind of control at the cognitive level. Indeed, one could characterise work at the cognitive end of the spectrum as working from cognition outwards and at the physical end of the spectrum as working from the body inwards. Ideally, virtual agents should have completely realistic movement and physical interaction as well as human-like cognitive abilities. In reality, both involve solving many hard problems so that there is a tendency for groups to place their emphasis at one or other end of the spectrum. This difference of emphasis can be found in a number of specific topics within virtual agents, and is excellently illustrated by the issues involved in the key area of *emotion*, which raises problems at both ends of this spectrum.

2.3 Emotion

We should note here that work in virtual agents has given a fresh impetus to the whole field of motivation and emotion in agents, perhaps for two reasons. Firstly, an embodied virtual agent in a virtual environment provides many more external channels for the representation of emotional state — gaze, facial expression, gesture and overall body language — than was the case with disembodied intelligent agents, where language content was just about the only means of expression. Secondly, as seen below, many virtual agents domains are those in which the expression of emotional state is essential to the application. Here, the use of avatars in distributed multi-user environments has provided a driving pressure.

The emphasis of those working at the cognitive end of the agent spectrum is on emotion as a cognitive state, while for those working at the more physical end it is on emotion as a bodily state. Note that by this we mean the internal modelling of emotion, rather than its external expression. These two approaches reflect a long-standing debate within psychology itself (Picard, 1997) and can be traced back as far as the separation of body and mind by Descartes.

The more long-standing approach of cognitive modelling (Ortony et al., 1988; Frijda, 1987), has the advantage that the agent is always in an explicitly defined emotional state or states, giving a clear linkage to the external manifestation of that emotion. However, at a more behavioural level, emotion is produced by the working of lower level structures. The simplest — but rather crude — way of modelling emotion at a such a low level is to equip the agent with *meters* that are incremented or decremented according to interaction with the environment, with other virtual agents or with a human user. A more sophisticated and realistic approach at this level is to model an endocrine system, as in Creatures (Grand et al., 1997), with chemical emitters and receptors (Canamero, 1998). Emotion is then manifested as part of the overall interaction of the agent with its environment rather than being modelled as a cognitive state, and much work in this area has been done by those attempting to construct *animats* or artificial animals, such as (Schnepf, 1991) and (Donnart and Meyer, 1994), for example.

3 Physical Agents

In this section, we consider virtual agents where physical behaviour is seen as the key issue. Such agents need not be human in form: they could be abstract (Sims, 1995) or mechanical (Prophet, 1996)[1], they could be animals such as birds (Reynolds, 1987), fish (Terzopoulos et al., 1994) or dolphins (Martinho et al., 1998), or they could be fictional, such as Teletubbies (Aylett et al., 1999). Human forms (Badler et al., 1993) are, of course, also common, whether as virtual actors (Shawver, 1997; Wavish and Connah, 1997), virtual humans (Thalmann and Thalmann, 1998) or avatars (Damer, 1998) in web-based multi-user virtual environments. In all these cases, common issues have to be faced.

3.1 Physical Issues

Firstly, body movement and mobility must be handled, often raising important issues of body structure. This suggests that ideas surrounding the notion of *embodiment* may be as significant for virtual agents as they are for real agents. Once a sophisticated physical representation can be controlled, there is the opportunity to use it for non-verbal communication, including gaze, facial expression, gesture and overall body language.

[1] `http://www.technosphere.org.uk`

Secondly, once agents have mobility, they must be able to avoid undesired collisions with other objects in their environment, whether these are stationary, as in trees, buildings and furniture, or also moving, as with other agents. Indeed the introduction of other agents leads to concerns of social movement such as herding, flocking (Reynolds, 1987) or crowd motion (Musse and Thalmann, 1997), which brings still further problems to be tackled. Other kinds of interaction involving contact with the environment, beyond mere collisions, must also be considered, such as the grasping of objects, or physical interactions with other agents ranging from hugging them to eating them.

Finally, given a sophisticated physical representation and a repertoire of physical behaviours to go with it, a number of *control* issues arise. For example, the level at which control should be exercised must be determined, and includes possibilities of control at the level of individual *muscles*, or at the level of whole behavioural actions such as *walking* and *grasping*, or even at both levels. This issue is also familiar from robotics, and particularly from work on teleoperation. A second such issue is how physical behaviours should be combined, so that an agent can move and grasp at the same time.

While there are many parallels with concerns in robotics, there are, however, interesting differences of emphasis. For example, the problems of dealing with real physical forces and materials, not to mention gears and motors, make complex structures much harder to deal with in robotics than in virtual agents, as well as very much more costly. Thus virtual agents can be produced with much more interesting body structures than most robots. In the same way, virtual sensors need not suffer from the same deficiencies as real ones: virtual agents can always know their position in the world accurately, unlike robots for whom localisation is a major problem. On the other hand, gravity, friction, inertia and all the other physical properties of the real world are known for robots. In a virtual world, physics has to be added explicitly by the designer.

3.2 Bodies in Motion

Some of the impetus for physical agents has come from film animation. Hand-crafted animation is laborious and very expensive, so that it seemed an obvious idea to apply computing power to the process. Initially automation was used to produce intermediate stages between hand-drawn frames, so that far fewer drawings were necessary — this is sometimes termed *procedural animation*.

An extension of this idea can be seen in the Microsoft AGENT programming environment that has emerged from the Persona project (Ball et al., 1997) (available currently in 2D only). Here, authored chunks of animation are provided — *character-looks-left*, *character-disappears*, and *character floats upwards*, for example — and the programmer can invoke sequences that the system will merge smoothly into each other. A similar approach, but in 3D, has been taken in the IMPROV system for synthetic actors (Goldberg, 1997), which defines an action as a 'single atomic or repetitive activity' that does not require 'explicit higher-level awareness or conscious decisions'. In IMPROV, actions are divided into groups, with actions in the same group being mutually exclusive. This allows several different actions to be carried out at once so that, for example, a virtual actor can walk and chew gum at the same time.

Specifying the sequences can itself become very complex. IMPROV provides a scripting mechanism with the possibility of triggering scripts from within scripts or in response to user interface control. In the Persona project, this problem was tackled by using an off-line AI linear planner (Kurlander and Ling, 1995) which produced the sequences in a pseudo-compiled form in the spirit of Universal Plans (Schoppers, 1987). An advantage of this approach is that the agent's body need not be a complex structure but, conversely, in neither of these projects has the agent any autonomous control over its movement.

A second approach used in the animation community, and also in the manipulation of avatars (Emer-

ing et al., 1997), is to drive the agent from a real-world human, an approach known as *performance animation*. Here, a real-world human wearing special markers on important joints carries out the desired movements. These markers may be reflective if they are to be picked up by camera, or may use other modes of operation such as with magnetic markers. The movements are then reapplied to the virtual agent's body structure in order to produce movement. This technique was, for example, used very effectively to animate virtual agents on the digital model of the ship 'Titanic' in the eponymous film. Its advantage is computational tractability, but it requires a properly modelled agent body, at least as far as joint relationships are concerned. It is even less flexible than the sequencing technique described above, although the two could be combined with one being used to create the *authored segments* the other then sequences.

Once an agent has a realistic body structure, an obvious next step is to provide autonomous control over it, so that agent motion is driven by internal *direction* of the body structure rather than external animation of the body surface. Reynolds was one of the pioneers of this approach, which he described as *behavioural animation* (Reynolds, 1987). Depending on the physical accuracy of this control and the structural complexity of the body, however, this technique can be very computationally expensive. Two interesting applications of this approach can be seen in the work of Terzopoulos and his group at Toronto, which concentrates on fish and other aquatic creatures, and JACK, a family of humanoid models developed over some years by Badler and his group at the University of Pennsylvania, and now available commercially.

3.2.1 Terzopoulos' Fish

The mechanics of Terzopoulos' fish (Terzopoulos et al., 1994) are defined by a set of spring-mass models forming the skeleton, with various springs meeting together at nodes and Lagrange equations being used to determine the motion of the structure. (Such use of continuous mathematics to determine motion is relatively unusual in this field.) When a fish beats its tail, it sets in motion a volume of water, and the inertia of the displaced water produces a reaction force normal to the fish's body, proportional to the amount of water displaced per unit time. Pectoral fins on the fish are used to control its pitch and yaw. This produces realistic motion but at some computational cost. When similar work was carried out on dolphins that had to work in real time for display in the 1998 World Fair, Expo '98 at Lisbon, this approach was replaced by simple (and much cheaper) sine wave propagation through the simulated skeleton (Martinho et al., 1998). Other work continues into the use of damped spring models for body motion as, for example, [de Jong 98].

Initial versions of the fish (Terzopoulos et al., 1994) took perceptual information directly from the data structures of the world model and the geometric and photometric information available to the rendering engine, giving them perfect virtual sensors. However, subsequent work (Terzopoulos et al., 1996) incorporated a biologically based perceptual system, though taken from primate rather than piscean vision. Eyes are modelled as four coaxial cameras giving approximately the spatial nonuniform foveal/peripheral capabilities of a biological eye — that is, high resolution in the central foveal region and lower resolution peripheral vision. The 3D field of view is projected onto the 2D simulated retina with the direction of *gaze* controlled by separate horizontal and vertical *muscles*. An active vision approach is needed to allow the eye to centre an interesting object (where *interest* is determined by analysing a colour histogram of the object) and to factor out the body movement of the fish. This perceptual system is proposed for general use in artificial animals, though most other work to date uses systems of far less sophistication.

3.2.2 JACK

Where Terzopoulos' fish were conceived as wholly autonomous agents, JACK in its commercial variants (for example, Transom JACK) has been targeted at ergonomic and engineering design applications. For example, one might wish to try a JACK model in a virtual tractor cockpit to analyse the placement of the controls. Thus high-level control of JACK can be left to the user either through direct manipulation or through high-level behavioural commands such as *walk* or *reach*. At the same time, however, programming language interfaces in both C++ and Lisp allow JACK to be driven as an autonomous agent if desired.

The targeted applications do require a degree of biological realism. Transom JACK has 68 joints, giving about 120 degrees of freedom to be driven. While this is not completely biologically accurate, important joints such as the shoulder are accurately modelled and it has articulated eyeballs so that gaze can be directed. Torques can be measured at particular points on the model and account is taken of the weight of any object held. The specific JACK figure used is generated from a database containing a range of human characteristics, so that it is possible to specify a figure with height at the 40th percentile and reach at the 70th percentile for example. Separate databases are available for females (as *JILL*) and children.

Essentially, high-level behavioural commands are translated into low-level muscle commands using kinematic and inverse-kinematic models very much as in robotics. Effort has been put into making low-level behaviour accurate, so that, for example, walking includes a heel strike. In addition, a small number of autonomous low-level behaviours have been incorporated, so that, for example, if a JACK figure is tilted off balance frontways, it will automatically take a step forwards to compensate. In the same way, if JACK is instructed to track an object it will do so with its gaze alone while the object is within the field of view, but will also turn its head if the object moves outside the field of view.

This accuracy of control is computationally demanding, but since the perceptual system is modelled very much less accurately than the muscle system and high-level control is left to the user, it is possible to run JACK in real time on relatively powerful machines as long as the environment is not too complex. On top-end graphical platforms, it may be possible to run up to four JACK models at once.

A European equivalent to the JACK system has been developed in the HUMANOID project (Boulic et al., 1995). This has 75 degrees of freedom for the body and an extra 30 degrees of freedom for each hand. In order to make the physical appearance more realistic (a JACK model is not very curvy), the skeleton is covered with a second layer of *blobs* (metaballs) to represent the skin and muscle. Just as versions of JACK have been used in other research projects in the US, so the HUMANOID model has been used in a number of European projects.

Apart from individual motion, there have also been investigations into social movement of groups of agents. The seminal work of Bruce Reynolds on *boids* (Reynolds, 1987) showed that realistic flocking herding and schooling behaviour could be produced through the application of a small set of rules rather than requiring a heavy-duty analytical framework such as that offered by particle physics. These ideas have been applied to other non-human agents, such as the fish discussed above (schooling behaviour), but also to the more complex requirements of human crowds in cities (Benford et al., 1997) and to the simulation of traffic in urban environments (Wright et al., 1998). However, many interesting applications involving social movement which have been tackled in 2D simulation — such as emergency evacuation (Galea and Galparsoro, 1994), for example — remain to be tackled in a 3D virtual environments.

3.3 Non-Verbal Communication

As argued above, the development of sophisticated body structure opens up a much larger repertoire of non-verbal communication, whether through the use of glance, facial expression, gesture, stance or overall body language. Much of the work in this area has been driven by the use of avatars to represent users in distributed interactive graphical environments that have extended the originally text-based MUDs (multi-user domains or dungeons). Here the need to support complex social interaction has motivated a search for ways to expand the communications bandwidth beyond the use of typed natural language.

One approach to non-verbal communication is to concentrate on facial expression. This work need not depend on a physically realistic face: the use of *emoticons* to express mood using standard keyboard symbols in email messages has served as a precedent for fairly crude avatars in which a highly stylised 'face' may represent one of a small range of emotions (typically happiness, sorrow and anger), as in The Palace [2]. However, parallel work on generating speech has required the building of human facial models in which muscle control can allow mouths to look as if they might be speaking the actual words being generated [BT ref, Hayek friends ref]. The development of such models then also supports the use of more naturalistic human facial expressions, as for example in OnLive Traveler [3] which allows user-controlled head movement as well as expressions for happiness, sorrow, surprise and anger, together with random blinking.

One of the simplest applications of facial expression is the use of glance to indicate the attentional focus of an agent. In a system which uses virtual sensors — even very simple ones — it is interesting to note that glance is a by-product of the agent's own use of its perceptual system. Where an agent does not have a field of view determined by its perceptual system (for example where information is being extracted straight from the data structures of the VE), then glance may be more explicitly applied, as it is for example in the STEVE pedagogical system discussed below. In the same way, avatars driven wholly by their users require the user direction of glance.

At the other end of the scale of sophistication, body language has also been studied, sometimes in the context of non-human agents. Thus computer pets such as Dogz and Catz[4] are equipped with behaviours that are intended to communicate with the human user (such as begging in the case of a computer dog). The advent of more realistic human body structures has similarly led to development of, for example, the avatars in Active Worlds [5], which are full bodies that can wave, jump and dance (as well as walk) under user control.

In its use for avatars, the HUMANOID model mentioned in the last section has enabled the investigation of a much wider repertoire of gesture and body language (Guye-Vullieme et al., 1998) as well as facial expression. This includes welcoming gestures, such as bowing, and aggressive gestures through arm or hand movements. User trials with this system showed both the power of non-verbal communication and the inherent difficulty in controlling the behavioural repertoire with simple and inexpressive mechanisms such as mouse clicks and menu selection.

Ultimately, however, the most convincing representation of agents or avatars in virtual environments will arise from the integration of various such forms of non-verbal communication with speech and intonation. Though it is still early, some initial work in this respect has already been done in the development of a system for animated conversation (Cassell et al., 1994). The system developed by Cassell and colleagues automatically animates conversations between agents using synchronised

[2] `http://www.thepalace.com`
[3] `http://www.onlive.com`
[4] `http://www.pfmagic.com`
[5] `http://www.activeworlds.com`

speech, intonation, facial expressions and hand-gestures. These latter include lip and eyebrow movement, gaze from head and eyes, and a library of predefined handshapes together with wrist control and arm motion.

3.4 The Agent-World Coupling

We have not so far considered the issues that arise from allowing physical agents to perceive and have a physical effect upon the world. This is a particular case of the interaction between object and VE already discussed above. At an abstract level, perception and action define an agent's coupling to its virtual world, and we consider each in turn.

The extent to which perception in a VE is modelled on real-world perception varies widely for different agents. In the real world, perception — as everyone who works with robots soon understands — is a really difficult problem. In a virtual world, in contrast, perception need suffer no problems with ambiguity, noise or lengthy processing. The very simplest of virtual sensors might consist of projecting a line from the agent's 'eyes' and returning information about any object in the virtual world it intersects, drawing on the data structures that represent the object for its identity and properties.

In contrast, as has already been mentioned above, some work has explored a biologically plausible perceptual system for virtual agents (Terzopoulos et al., 1996), in which the agent's field of view is projected onto a simulated retina and vision algorithms are used to process the pixels into a form that can be used by the agent. In between the two extremes, there are virtual robot systems in which infra-red and ultra-sound sensors are modelled with some degree of realism (Michel, 1998) by, for example, adding noise into the simulated signal.

Irrespective of the biological plausibility of any particular system of virtual sensors, however, it is important to understand that perception is an interaction between an agent and its environment. For example, one might expect an agent in a darkened room to be able to 'see' less than one in a brightly lit room. As with other areas that balance the agent and the environment, agent perception is a pragmatic issue, once again relating to the location of knowledge. The simpler the agent's perceptual system for a given level of agent functionality, the more knowledge must be transmitted to agents from the objects they perceive. Conversely, an accurately modelled fish visual system requires much less knowledge from the environment, but a correspondingly greater amount of processing within the agent.

If perception raises questions about the extent to which an agent is embedded in a particular world, action raises the same question much more strongly. While perception is passive, action results in changes in the world that depend upon the functionality of the agent as well as the functionality and state of the world. For example, if an agent grasps an object, its ability to lift it depends on both its physique and the size and weight of the object. Moreover, the weight of the object might in turn depend on whether the world in question was on the surface of the earth, inside a space station (and thus subject to micro-gravity) or at the bottom of an ocean.

At a more detailed level, actions such as a grasp should be visually convincing — a hand should not pass through the object being grasped and the relationship between the hand and the surface of the object being grasped should look plausible. The complexity of the interaction at this level depends on the sophistication with which a hand is modelled and whether physical forces and constraints such as surface hardness are represented in the interaction or whether the grasp is merely an animation covering a grouping of primitive graphical objects.

4 Cognitive Agents

In a relatively early article, Bates very effectively makes the case for a significant effort in the area of content and structure for virtual environments through theories for agents, presentation and drama, which he argues are vital in order for virtual reality to achieve its potential (Bates, 1992). In this section, we will avoid agent research *per se*, which is reviewed extensively elsewhere in a number of survey articles, for example (Wooldridge and Jennings, 1995), Nwana (Nwana, 1996; Jennings et al., 1998), but concentrate on those particular efforts that have been concerned with agents in a virtual environment and their embodiment in that environment. We concentrate on the content and structure of virtual environments through the work that has taken up this call, and note that much of it is aimed at, and has been driven by the development of virtual environments for entertainment and drama, though other equally significant areas of application have provided impetus, too.

The issues involved in the development and construction of agents that lie at the cognitive end of the spectrum can be grouped under three key areas. First, there must typically be some traditional (agent) architectural component that is responsible for critical cognitive capabilities of reasoning, decision-making, planning, learning, and so on, regardless of whether the agent is situated in a virtual environment or not. This is exactly the same problem faced by intelligent agents in other contexts, and can be regarded as the foundation on which all else is based. As stated above, we will not consider this aspect in itself here, but only in its relation to the context of virtual environments.

The second key area relates to the realism of the agent in its environment with regard to its behaviour (in a broad sense) rather than its rendering or visualisation. In order for intelligent virtual environments to be practical, they must be *believable* (as Bates points out), both through the actions of the agents themselves, and their interaction with others. Among other things, this means that cognitive function must not be divorced from its affective influences — motivation, emotion, personality, and so on — which must find ways of expression in the virtual environment. Indeed, there has been a significant amount of research in this area of affective agent architecture and models of emotion and motivation. Examples include the work of Moffat and Frijda on developing a computational model of emotion (Moffat et al., 1993; Moffat and Frijda, 1995), and the sophisticated "computational theory of mind" developed by Sloman and colleagues over a number of years (Sloman and Croucher, 1981; Sloman, 1987; Sloman, 1997) and inspired by the seminal work of Simon thirty years ago (Simon, 1979).

The concern with *expressing* the affective influences in intelligent virtual environments brings us to the third key area of representation or visualisation. This completes the circle in that we come back to the issue of how the cognitive and affective models can be mapped onto the physical models. Thus the visualisation of the agents is not unimportant here, but can only be effective if agent models provide sufficient and appropriate detail of information.

For example, Badler *et al.* (Badler et al., 1997) describe the ways in which *personality*, which is manifested in broad characteristics such as curiosity and fatigue, for example, relates to various parameters of locomotion of an artificial agent, such as speed and anticipation. They describe a simple model in which personality can then influence future courses of action. While the model is relatively simple, and the work only an initial pointer to further directions to explore, it demonstrates the connection of the higher-level mental components to their lower-level physical expression.

4.1 Virtual Environments for Art and Entertainment

A sensible place to start in situating this aspect is with the work of Bates himself, and colleagues on the Oz project, in attempting to bring together existing technologies, and artificial intelligence techniques

in particular, for application to virtual environments. Much of the work of the Oz Project revolves around the development of a *broad* agent architecture called *Tok*, its reactive component (*Hap*), and its emotion component (*Em*) (Bates, 1994), for non-linguistic, believable agents (Bates et al., 1992; Loyall and Bates, 1993). Here, believability is all-important, and imposes certain constraints when dealing with a real-time animated environment. The reactive architecture enables speedy responses in an environment in which primitive actions last between 100 and 150 milliseconds, and fast effective responses are demanded.

Tok has been used to create several specific agents inhabiting artificial worlds. One particular early outcome was the construction of three real-time animated agents, known as *woggles*, each with different characteristics, and existing in an animated virtual world (*Edge of Intention*) (Loyall and Bates, 1993). The agents, which are stereotypic in personality, interact with each other (and sometimes with the user through a mouse-controlled fourth Woggle), and play, explore and fight. Woggles are visually very simple round shapes with expressive eyes and, while avoiding many issues that might otherwise demand addressing with more sophisticated graphical environments, some key areas of interest such as sensing within the agent architecture, and the exhibiting of emotions that vary based on internal needs, are effectively tackled.

Other work has produced Lyotard, a simulated cat that demonstrates the integration of the emotion and reaction substrates in Tok (Bates et al., 1992), and has made efforts to develop linguistic ability on top of the basic architecture (Loyall and Bates, 1997; Loyall, 1997; Bates et al., 1994). The Oz Project has managed to show how the integration of agent models that include emotion and reactivity can give rise to the beginning of effective embodied artificial agents in virtual worlds for interactive drama and storytelling.

4.2 Virtual Theatre

Drawing on the work of Bates, but with roots very firmly in the field of intelligent agents is the work being carried out by Hayes-Roth and colleagues in the Virtual Theater Project. Initially attempting to develop a new paradigm for multi-agent systems (Hayes-Roth and Brownston, 1995), based in part on the early work of the Oz project, the work is concerned with the provision of multimedia environments in which users or agents fulfill various roles including, in particular, animated actors that can be directed but can also improvise their *performances*. Much of the motivation for this work stems from the identified need in advanced HCI applications for intelligent agents that can *engage* their users (Hayes-Roth, 1995). Key to these aims is the need to develop effective agent models based on emotions, moods and personalities (Rousseau and Hayes-Roth, 1998).

This work has explored several different avenues, with one leading to the development of more or less sophisticated computational characters, and another to the development of environments that can themselves be considered to provided the intelligence the individuals within it use. The former approach is exemplified by Tigrito (Maldonado et al., 1998), an affective computational character that appears as a toy tiger, which can be used in three key modes of interaction. First, Tigrito can be used as a *virtual pet* by providing a means of interacting with it in terms of modifying a representation of the user's *mood* and choosing actions to perform. Tigrito's mood then automatically responds to the actions that it can perceive, and is expressed by Quicktime movies filmed with a stuffed animal. The second mode of interaction involves a second tiger, an avatar, whose actions can be determined by the user, and to which Tigrito reacts, rather than the user. Finally, in the third, "movie" mode, the user can only modify the moods of the tigers and watch the agents themselves select their own actions.

In contrast to focussing on the agents themselves, another strand of work aims to develop *annotated* virtual environments that provide cues that enable the agents within them to exhibit intel-

ligent behaviour. Doyle and Hayes-Roth describe work based on the human-computer interaction literature on *objects that provide knowledge in the world*, in which the virtual worlds they construct are annotated and contain explanations of, or cues for, emotional responses and other actions by agents, role-specific functionality and personality, as well as information regarding problem-solving and game-playing to suit the particular environmental circumstances (Doyle and Hayes-Roth, 1998). This appears to be an effective way of generating the overall desired behaviour of the virtual world, and gives a more direct interpretation of *intelligent virtual environments*, since much of the intelligence is part of the environment itself.

4.3 Games

Possibly the most unashamedly commercial effort is the *Creatures* game (Grand et al., 1997) has proven to be a very successful product that incorporates several of the characteristics of the previous work as well as many artificial life techniques. It comprises artificial agents that inhabit an elaborate $2\frac{1}{2}$-dimensional world containing various objects that can interact with the creatures (known as Norns) in different ways. These include automated objects such as elevators as well as the more traditional food and toys, and it is also possible to download and add other new artifacts. Users can interact through by means of the mouse which controls a disembodied hand that can reward or punish the creatures through stroking or slapping them.

Where the creatures stand out from other efforts is in their construction with a neural network *brain* of 1000 neurons and 5000 synapses, and a complex biologically plausible model. Indeed, they are very carefully modelled, with a game life-span of several hours, and different stages of creature development. They are able to sense light and sound, and though the simulation of these senses is relatively coarse, it is adequate. Simple language ability can also be learned.

Creatures is a widely-available commercial product that uses artificial intelligence and artificial life techniques in the creation of an entertaining virtual environment. Whether the combination of these techniques is sound from an academic perspective, and whether it advances the state-of-the-art are both open questions. However, as Grand and Cliff point out, it can be regarded as probably the single largest experiment in intelligent virtual environments (Grand and Cliff, 1998), and in that alone is worthy of note.

4.4 Pedagogical Agents

In contrast to the high-profile application domain for intelligent virtual environments of entertainment, Johnson and his colleagues have been developing an infrastructure for, and example applications of virtual pedagogical agents. Starting with previous work on intelligent tutoring systems and incorporating intelligent agent research, this work seeks to create animated, persona-rich agents that use eye contact and body language, for example, in order to interact effectively with students.

STEVE (Soar Training Expert for Virtual Environments) is a pedagogical agent that inhabits a virtual environment, monitoring it and periodically controlling it through virtual motor actions (Rickel and Johnson, 1997). STEVE serves as a tutor and collaborator for a student, demonstrating how to perform tasks, monitoring student performance and providing assistance. Built on the VET (Virtual Environments for Training) software that allows the environment, the pedagogical agent and the student to run as separate processes, it uses a message-passing mechanism between components to represent the occurrence of events in the world. Humans interact with the virtual world by means of a head-mounted display and a 3D mouse or interactive glove. STEVE comprises a cognitive component (implemented in Soar) for high-level processing, and a sensorimotor component for interfacing with

the virtual world and dealing with perception and motor commands. It can be rendered either as a disembodied agent (as a virtual hand that can point and grasp) or as an embodied agent (appearing as a human figure) without affecting the cognitive level.

Somewhat unlike Steve, which was designed to operate in immersive virtual environments, though with the same broad aims and amy similar capabilities, (Agent for Distance Learning Environments) is a pedagogical agent designed to operate over the World-Wide WebAdele (Johnson et al., 1998b). Adele comprises a reasoning engine and an animated persona implemented as a Java applet, and can monitor student performance and provide feedback in the prototype application domain of clinical diagnosis. Adele is much more limited than Steve because of its use of a conventional 2D graphical interface, but does, nevertheless, have the capacity to use gaze and gesture to modify emotional facial expression in order to motivate students.

5 Virtual Worlds

If much work in the field of VE is moving towards intelligent virtual environments in order to add specific pieces of functionality, while work in AI considers the use of virtual environments as a way of creating more interesting intelligent agents, researchers in the field of artificial life have altogether more ambitious aims in some cases. These include the creation of virtual worlds containing *digital life*, complete with coherent physical laws, possibly, but not necessarily, similar to those of the real world. Some see distributed interactive virtual environments such as Active Worlds as a basis for the developments of such virtual worlds, providing an on-line laboratory for the investigation of AL and of long-term autonomous interaction between AL forms and virtual worlds. Where much work in genetic algorithms stays at the level of genotype, a virtual world would allow the exploration of the phenotype consequences of a particular genetic data set and the exploration of evolution at the phenotype level in the same sort of way as it occurs in the real world.

One of the first pieces of work within this framework was that of the block creatures created by Sims, whose genetic system specifies various creatures built of collections of blocks which are linked by flexible joints (Sims, 1995). These joints are in turn activated by *muscles* in the form of circuits controlled by an evolvable network of functions. The virtual world of these block animals contains a virtual physics, and the block animals evolve locomotion abilities under these conditions. Various other tasks were explored, such as the competitive acquisition of a free block in a community of different block animals[6].

A more recent approach can be found in Technosphere, an on-line virtual world in which artificial creatures can be constructed and run (Prophet, 1996). These creatures are deliberately simple and mechanical in form, with the user able to choose the head, body, eyes and so on from a small pre-determined set, either for a herbivore or carnivore. Creatures can eat, sleep, move, fight and mate, and when they die (or are killed) their bodies gradually decompose. Statistics collected about this virtual world show a tendency for users to create carnivores rather than herbivores, producing a rather unstable ecology.

The motivation for TechnoSphere was primarily artistic rather than scientific, and much effort has gone into producing visually attractive natural scenery (landscape and vegetation) with, for example, trees built of fractals rather than of polygons. However, TechnoSphere cannot be rendered in real-time (and instead off-line animations are produced) so that, to produce some interactivity, the system mails users at intervals reporting on the progress of their creatures. Facilities are provided to check the

[6]MPEG movies of these creatures can be found at `http://www.biota.org/conf97/ksims.html`

location and activity of creatures on-line, and the system maintains a family tree for creatures which manage to breed successfully.

A final example of this approach to VEs is the *Nerve Garden* project (Damer et al., 1998), which is a client-server system that allows users to generate 3-D plant models using L-systems (Lindenmayer Systems) (Prusinkiewicz and Lindenmayer, 1990). A user can then select a particular plant and locate it in a plot in a VRML island garden VE. Multiple users were able to view and update the same island in a wall-sized display at the SIGGRAPH 97. Various viewpoints were available, including a dynamic viewpoint from the back of a flying insect animation, which circled continuously over the island. Sound effects were added, including thunder, which accompanied L-system based virtual lightning. The work is fairly primitive from an AL perspective since it does not support plant growth or interaction between plants and environment, but a Nerve Garden II project aims to meet these deficiencies by developing a simple but effective garden eco-system.

6 Directions and Issues

By way of conclusion, we discuss some of the directions and issues that appear to be emerging in the area of intelligent virtual environments. In doing so, we begin by using the issue of autonomy as a measure to which we can relate different levels of control, for example, and as a fixed point to which other issues can be connected. In this way autonomy, which runs through much of the extensive work on intelligent virtual agents as should by now be clear, can serve to focus the discussion.

6.1 Autonomy

Indeed, as some of the examples given previously indicate, virtual agents at the physical end of the spectrum exhibit everything from zero autonomy (as in performance animation or standard avatars) to complete autonomy, passing through various intermediate states. This range of autonomy contrasts with more cognitive agents which, almost by definition, are highly autonomous, though they may have a limited physical functionality in their environments.

It would be reasonable to assume that avatars as described thus far require no autonomy at all since they can be — and usually are — controlled directly by the user that they represent. With a simple agent model containing perception, reflection and activity, sensing and reflecting are directed by the user, leaving only the actuation to be carried out automatically. However, as avatars become more complex, opening up the possibility of the use of facial expression, gesture, and body posture as extra channels of communication, the standard type of control via mouse and menu becomes increasingly difficult. Just as conscious control of all these aspects of the human body might well lead to cognitive overload, so the increasing facilities offered by avatars pose overload problems for their users. Either more flexible direct manipulation interfaces are required — as in the use of a camera to capture or interpret the user's real-life facial expression and body sensors for posture and gesture — or some low-level behavioural autonomy must be incorporated into the avatars.

Agents with low-level behavioural autonomy are a good description of virtual actors (Shawver, 1997). Here, agents are incorporated into a scenario under high-level direction by a user such as with an instructor developing a training scenario. Systems of this kind have already been constructed for the practice of hostage release actions (Stansfield and Shawver, 1996), and for medical aid on the battlefield (Stansfield et al., 1998). In the first case, virtual actors are used to play a terrorist and a number of hostages, while the trainee enters the VE as an avatar, usually driven by a headset and data glove at least. The instructor controls the overall scenario via a pre-developed script, but the virtual

actors are equipped with low-level behaviours both to instantiate actions in the script and also to react to actions in real-time. Thus, if the trainee or the terrorist fire a shot, the virtual hostages will dive out of the way.

In the battlefield scenario, trainees again drive an avatar and provide medical aid to virtual actors with various wounds, determined by the script. However, the virtual casualties must respond in a realistic physical manner to the medical aid they receive: blood flows must cease, consciousness must be regained, and facial colour must change, for example. This use of the *director metaphor* seems natural for applications related to film, for example.

Other agents have a very high degree of autonomy, though this may be implemented in rather different ways. With more physical agents, autonomous action usually depends on the interaction between drives internal to the agent and stimuli from the environment. Since the agent's drives affect its behaviour and its behaviour affects the environment (especially in an environment containing multiple agents), a continuing feedback loop is produced.

For example, Terzopoulos' fish (Terzopoulos et al., 1994) have internal drives of *hunger*, *fear* and *libido*, which stimulate feeding, fleeing and mating behaviour respectively. The drives are themselves produced by the interactions between the fish and its environment, such as the time since the last feed or whether a predator is in sight. Similar approaches are taken by Virtual Teletubbies (Aylett et al., 1999), which are driven by *hunger*, *fatigue* and *curiosity*, and in Creatures (Grand and Cliff, 1998). The problem faced by all all such systems, however, is in producing a reasonable amount of persistence to avoid oscillation between competing behaviours in response to slight changes in the internal drives.

A contrasting approach may be seen in more cognitive agents. Thus, STEVE (Johnson et al., 1998a) is task driven with its autonomy based on goal expansion through a generative AI planning system. Its behaviour also changes in relation to the state of the environment — for example, if a trainee undoes an action, STEVE will plan to redo it. Similarly, if a trainee is not looking at the location of the current subtask, STEVE will ask them to do so. Clearly, it would be possible to combine this classical goal-driven approach with the drives approach just discussed, but no work has yet been found in which this is done. Apart from the conceptual issues involved in such a combination, it may be that this would pose excessive demands on currently available computational resource.

6.2 Combining the Physical and Cognitive

While the majority of systems fall into the broad categories that lie at the different ends of the agent spectrum, physical and cognitive, some efforts are combining the different emphases. One particular example, the ALIVE system, is worth a detailed examination because of the way in which it brings together these different aspects in an effective and coherent way. ALIVE allows unencumbered full-body interaction between a human and a rich graphical world inhabited by autonomous agents, by using a *mirror paradigm* in which the user sees a representation of herself in the virtual environment.

ALIVE agents are modelled in a very sophisticated way, and are constructed via a toolkit that aims to have agents do things that make the *most sense* at any time (Blumberg and Galyean, 1997). The designer of an agent specifies its virtual sensors for perception of the environment, releasing mechanisms that identify behaviourally significant stimuli (such as a human hand being extended and down for a dog), the internal needs and motivations, the motivationally-relevant behaviours that compete for control of the agent according to its internal needs and the opportunities in the environment, and the motor skills, such as *walking*, *sitting*, *wagging tail*, etc. At each time step, agent states and geometries are recomputed according to the selected behaviours and the corresponding motor activities (Blumberg and Galyean, 1995).

A special 3D agent is created to represent the user, whose position and state are computed by a vision system based on a camera image of the user (Maes et al., 1997). Interaction with the user is then accomplished through visual identification of the actual user's hand and body gestures, and auditory feedback about the agents' internal state. In this way, the artificial agents can sense the human user with the same virtual sensors used to detect other objects in the environment, and the user is rendered together with the artificial agents and environment using a live video image.

Examples of applications constructed using ALIVE include a puppet world in which a puppet would follow user directions and employed facial expressions to convey internal state, a world inhabited by hamsters and their predators and, more recently, a world containing a sophisticated virtual dog with a wide range of behaviours that used auditory as well as visual input, and that would enage in both interactive and autonomous action.

The ALIVE system thus involves important issues typically addressed by work at the physical end of the agent spectrum in its concern with the body movement of the artificial agents and provision of an adequate model for that, which includes virtual sensors, and its recognition of salient features such as hand and body gestures, for example. Conversely, it also addresses more cognitive issues in the strong model of motivation and directed behaviour, for example, in pursuit of believability. In this coming together of work from different strands, many aspects of intelligent virtual environments become better defined and understood. Arguably, the ALIVE systems is the kind of work that suggests some directions for the future of the field, and which demonstrates the important and useful cross-fertilisation of ideas from AI and VE.

6.3 Prospects

As stated at the start of this paper, resource constraints are still very evident in intelligent virtual environments. The type of physical modelling carried out in some work — the fish for example — is very expensive computationally, since it relies on a strong analytical framework that makes big demands on processing power. Modelling physics successfully may require the development of a cheaper approach based on *local* computations, possibly drawing from artificial life technology. Another possibility might be the application of qualitative physics from AI to achieve a broadly accurate model — a kind of physics *level of detail* approach. From the other perspective, real-time rendering supporting continuous interaction must be the aim for successful virtual environments, though this may not be required for many entertainment applications where the development of an off-line video is consistent with existing work.

The field of intelligent virtual environments that is introduced and reviewed in this paper, is a new and emerging area that is still relatively immature, but one which has already demonstrated a significant degree of vitality with many exciting ideas being generated and explored. At the same time, this immaturity reveals a noticeable lack of generally useful tools and architectures, however. While the development of a variety of different approaches may be inevitable as much as desirable at this stage, the lack of such tools and architectures may begin to limit the constructiveness of exploration as wheels are reinvented and consolidation slows. Some progress is underway in this respect with regard to the establishment of the VRML'97 standard, which was a major step forward but, as we have noted, it is grossly deficient for the type of work discussed here. Nevertheless, the opportunities for rapid advances are marked. Indeed, if a closer interaction and engagement between strands of research in the fields of virtual reality, artificial intelligence and artificial life can be encouraged and flourish (Fisher and Fraser, 1998), the chances of a greater appreciation of, and agreement over, the critical issues will be enhanced. We should then be able to expect a new wave of applications using the technology of intelligence virtual environments much nearer in the future.

References

Aylett, R., Brazier, F., Jennings, N., Luck, M., Preist, C., and Nwana, H. (1998). Agent systems and applications. *Knowledge Engineering Review*, 13(3):303–308.

Aylett, R., Horrobin, A., O'Hare, J. J., Osman, A., and Polyak, M. (1999). Virtual teletubbies: reapplying a robot architecture to virtual agents. In *Proceedings of the Third International Conference on Autonomous Agents*.

Badler, N. I., Reich, B. D., and Webber, B. L. (1993). *Simulating Humans: Computer Graphics Animation and Control*. Oxford University Press.

Badler, N. I., Reich, B. D., and Webber, B. L. (1997). Towards personalities for animated agents with reactive and planning behaviors. In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 43–57. Springer-Verlag.

Ball, G., Ling, D., Kurlander, D., Miller, J., Pugh, D., Skelly, T., Stankosky, A., Thiel, D., Van Dantzich, M., and Wax, T. (1997). Life-like computer characters: the persona project at microsoft research. In Bradshaw, J. M., editor, *Software Agents*, pages 191–222. AAAI Press.

Bates, J. (1992). Virtual reality, art, and entertainment. *PRESENCE: Teleoperators and Virtual Environments*, 1(1):133–138.

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

Bates, J., Loyall, A. B., and Reilly, W. S. (1992). Integrating reactivity, goals, and emotion in a broad agent. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, Bloomington, Indiana.

Bates, J., Loyall, A. B., and Reilly, W. S. (1994). An architecture for action, emotion, and social behavior. In Castelfranchi, C. and Werner, E., editors, *Artificial Social Systems: Fourth European Workshop on Modeling Autonomous Agents in a Multi-Agent World*. Springer-Verlag, Berlin.

Benford, S. D., Greenhalgh, C. M., and Lloyd, D. (1997). Crowded collaborative virtual environments. In *Proceedings of the 1997 ACM Conference on Human Factors in Computing Systems (CHI'97)*, pages 59–66. ACM Press.

Blumberg, B. and Galyean, T. (1995). Multi-level direction of autonomous creatures for real-time virtual environments. In Mair, S. and Cook, R., editors, *SIGGRAPH'95 Computer Graphics, Proceedings of the 22nd Annual ACM Conference*, pages 47–54. ACM Press.

Blumberg, B. and Galyean, T. (1997). Multi-level control for animated autonomous agents: Do the right thing ... oh, not that. In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 74–82. Springer-Verlag.

Boulic, R., Huang, Z., Shen, J., Molet, T., Capin, T., Lintermann, B., Saar, K., Thalmann, D., Magnenat-Thalmann, N., Schmitt, A., Moccozet, L., Kalra, P., and Pandzic, I. (1995). A system for the parallel integrated motion of multiple deformable human characters with collision detection. *Computer Graphics Forum*, 14(3):337–348.

Canamero, D. (1998). Modeling motivations and emotions as a basis for intelligent behaviour. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents*, pages 148–155. ACM Press.

Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., and Stone, M. (1994). Animated conversation: Rule-based generation of facial expression, gesture and spoken intonation for multiple conversational agents. In *Proceedings of the ACM SIGGRAPH Conference*, pages 413–420. ACM Press.

Castelfranchi, C. (1995). Guarantees for autonomy in cognitive agent architecture. In Wooldridge, M. and Jennings, N. R., editors, *Intelligent Agents: Theories, Architectures, and Languages, LNAI 890*, pages 56–70. Springer-Verlag.

Damer, B. (1998). *Avatars: Exploring and Building Virtual Worlds on the Internet*. Peachpit Press.

Damer, B., Marcelo, K., and Revi, F. (1998). Nerve garden: A public terrarium in cyberspace. In Heudin, J.-C., editor, *Virtual Worlds*, pages 177–185. Springer-Verlag.

Donnart, J. and Meyer, J. (1994). A hierarchical classifier system implementing a motivationally autonomous animat. In Cliff, D., editor, *From animals to animats 3, Proceedings of the Third International Conference on Simulation of Adaptive Behavior*, pages 144–153. MIT Press/Bradford Books.

Doyle, P. and Hayes-Roth, B. (1998). Agents in annotated worlds. In *AGENTS'98: Proceedings of the Second International Conference on Autonomous Agents*, pages 173–180, Minneapolis, MN.

Emering, L., Boulic, R., Balcisoy, S., and Thalmann, D. (1997). Real-time interactions with virtual agents driven by human action identification. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents*, pages 476–477. ACM Press.

Fisher, S. S. and Fraser, G. (1998). Real-time interactive graphics: Intelligent virtual worlds continue to develop. *Computer Graphics*, 32(3).

Franklin, S. and Graesser, A. (1996). Is it an agent, or just a program?: A taxonomy for autonomous agents. In Müller, J. P., Wooldridge, M. J., and Jennings, N. R., editors, *Intelligent Agents III — Proceedings of the Third International Workshop on Agent Theories, Architectures, and Languages (ATAL-96)*, Lecture Notes in Artificial Intelligence, 1193. Springer-Verlag, Heidelberg.

Frijda, N. (1987). *Emotions*. Cambridge University Press.

Galea, E. R. and Galparsoro, J. M. P. (1994). EXODUS: An evacuation model for mass transport vehicles. *Fire Safety Journal*, 22:341–366.

Goldberg, A. (1997). Improv: A system for real-time animation of behavior-based interactive synthetic actors. In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 58–73. Springer-Verlag.

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

Grand, S., Cliff, D., and Malhotra, A. (1997). Creatures: Artificial life autonomous software agents for home entertainment. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents*. ACM Press.

Guye-Vullieme, A., Capin, T. K., Pandzie, I. S., Thalmann, N. M., and Thalmann, D. (1998). Nonverbal communication interface for collaborative virtual environments. In *Proceedings, Collaborative Virtual Environments 98*, pages 105–112, Manchester.

Hartman, J. and Wernecke, J. (1996). *The VRML 2.0 Handbook: Building Moving Worlds on the Web*. Addison-Wesley.

Hayes-Roth, B. (1995). Agents on stage: Advancing the state of the art in AI. In *Proceedings of the Fourteenth International Joint Conference on Artificial Intelligence (IJCAI-95)*, pages 967–971, Montréal, Québec, Canada.

Hayes-Roth, B. and Brownston, L. (1995). Multiagent collaboration in directed improvisation. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 148–154, San Francisco, CA.

Jennings, N. R., Sycara, K., and Wooldridge, M. (1998). A roadmap of agent research and development. *Autonomous Agents and Multi-Agent Systems*, 1(1):7–38.

Johnson, W., Rickel, J., Stiles, R., and Munro, A. (to appear, 1998a). Integrating pedagogical agents into virtual environments. *Presence*, 7(6).

Johnson, W., Shaw, E., and Ganeshan, R. (1998b). Pedagogical agents on the web. In *ITS'98 Workshop on Pedagogical Agents and Workshop on Intelligent Tutoring Systems on the Web*.

Kurlander, D. and Ling, D. T. (1995). A planning-based animation controller. In *Proceedings, CHI '95*, pages 472–479.

Loyall, A. B. (1997). Some requirements and approaches for natural language in a believable agent. In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 113–119. Springer-Verlag.

Loyall, A. B. and Bates, J. (1993). Real-time control of animated broad agents. In *Proceedings of the Fifteenth Annual Conference of the Cognitive Science Society*, Boulder, CO.

Loyall, A. B. and Bates, J. (1997). Personality-rich believable agents that use language. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents*. ACM Press.

Luck, M. (1997). Foundations of multi-agent systems: Issues and directions. *Knowledge Engineering Review*, 12(3):307–308.

Luck, M. and d'Inverno, M. (1995). A formal framework for agency and autonomy. In *Proceedings of the First International Conference on Multi-Agent Systems (ICMAS-95)*, pages 254–260. AAAI Press/MIT Press.

Luck, M., d'Inverno, M., Fisher, M., and FoMAS'97 Contributors (1998). Foundations of multi-agent systems: Techniques, tools and theory. *Knowledge Engineering Review*, 13(3):297–302.

Maes, P., Darrell, T., Blumberg, B., and Pentland, A. (1997). The ALIVE system: Wireless, full-body interaction with autonomous agents. *Multimedia Systems*, 5(2):105–112.

Maldonado, H., Picard, A., Doyle, P., and Hayes-Roth, B. (1998). Tigrito: A multi-mode interactive improvisational agent. In *Proceedings of the 1998 International Conference on Intelligent User Interfaces*, San Francisco, CA.

Martinho, C., Paiva, A. M., and Gomes, M. R. (1998). Tristao and isolda: Synthetic dolphins with emotions. In *Proceedings, Workshop on AI and Entertainment, ECAI'98*, pages 16–22.

Michel, O. (1998). Webots: Synthesis between virtual and real mobile robots. In Heudin, J.-C., editor, *Virtual Worlds*, pages 245–263. Springer-Verlag.

Moffat, D. and Frijda, N. H. (1995). Where there's a will there's an agent. In Wooldridge, M. and Jennings, N. R., editors, *Intelligent Agents: Theories, Architectures, and Languages, LNAI 890*, pages 245–260. Springer-Verlag.

Moffat, D., Frijda, N. H., and Phaf, R. H. (1993). Analysis of a model of emotions. In *Prospects for Artificial Intelligence: Proceedings of AISB93*, pages 219–228, Birmingham.

Musse, S. R. and Thalmann, D. (1997). A model of human crowd behavior. In *Computer Animation and Simulation '97, Eurographics workshop*, Wien. Springer Verlag.

Nerves (1998). http://www.digitalspace.com/nerves/index.html.

Nwana, H. S. (1996). Software agents: an overview. *Knowledge Engineering Review*, 11(3):205–244.

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

Petta, P. and Trappl, R. (1997). Why to create personalities for synthetic actors. In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 1–8. Springer-Verlag.

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

Prophet, J. (1996). Technosphere. *Interpretation*, 2(1).

Prusinkiewicz, P. and Lindenmayer, A. (1990). *The Algorithmic Beauty of Plants*. Springer-Verlag.

Reignier, P., Harrouet, F., Morvan, S., Tisseau, J., and Duval, T. (1998). Arevi: A virtual reality multiagent platform. In Heudin, J.-C., editor, *Virtual Worlds*, pages 229–240. Springer-Verlag.

Reynolds, B. (1987). Flocks, herds and schools: A distributed behavioural model. *Computer Graphics*, 21(4):25–34.

Rickel, J. and Johnson, W. (1997). Integrating pedagogical capabilities in a virtual environment agent. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents*. ACM Press.

Rousseau, D. and Hayes-Roth, B. (1998). A social psychological model for synthetic actors. In *AGENTS'98: Proceedings of the Second International Conference on Autonomous Agents*, pages 165–172, Minneapolis, MN.

Schnepf, U. (1991). Robot ethology: A proposal for the research into intelligent autonomous systems. In Meyer, J. A. and Wilson, S., editors, *Proceedings of the First International Conference on Simulation of Adaptive Behaviour: From Animals to Animats*, pages 465–474. MIT Press/Bradford Books.

Schoppers, M. J. (1987). Universal plans for reactive robots in unpredictable environments. In *Proceedings of the Tenth International Joint Conference on Artificial Intelligence (IJCAI-87)*, pages 1039–1046, Milan, Italy.

Shawver, D. (1997). Virtual actors and avatars in a flexible, user-determined-scenario environment. In *Proceedings of the Virtual Reality Annual International Symposium*, pages 1–5, Albuquerque, NM.

Simon, H. A. (1979). Motivational and emotional controls of cognition. In *Models of Thought*, pages 29–38. Yale University Press.

Sims, K. (1995). Evolving 3d morphology and behaviour by competition. *Artificial Life*, 1:353–372.

Sloman, A. (1987). Motives, mechanisms, and emotions. *Cognition and Emotion*, 1(3):217–233.

Sloman, A. (1997). What sort of control system is able to have a personality? In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 166–208. Springer-Verlag.

Sloman, A. and Croucher, M. (1981). Why robots will have emotions. In *Proceedings of the Seventh International Joint Conference on Artificial Intelligence*, pages 197–202, Vancouver, B.C.

Stansfield, S. and Shawver, D. (1996). Using virtual reality to train and plan response actions to acts of terrorism. In *Proceedings of the SPIE Conference on Enabling Technologies for Law Enforcement and Security*, Boston, MA.

Stansfield, S., Shawver, D., and Sobel, A. (1998). Medisim: A prototype vr system for training medical first responders. In *Proceedings of the Virtual Reality Annual International Symposium*, Atlanta, GA.

Terzopoulos, D., Rabie, T., and Grzeszczuk, R. (1996). Perception and learning in artificial animals. In *Artificial Life V: Proceedings of the Fifth International Conference on the Synthesis and Simulation of Living Systems*.

Terzopoulos, D., Tu, X., and Grzeszczuk, R. (1994). Artificial fishes with autonomous locomotion, perception, behavior and learning, in a physical world. In Maes, P. and Brooks, R., editors, *Proceedings of the Artificial Life IV Workshop*. MIT Press.

Thalmann, D., Noser, H., and Huang, Z. (1997). Autonomous virtual actors based on virtual sensors. In Trappl, R. and Petta, P., editors, *Creating Personalities for Synthetic Actors*, pages 25–42. Springer-Verlag.

Thalmann, N. M. and Thalmann, D. (1998). The virtual humans story. *IEEE Annals of the History of Computing*, 20(2):50–51.

VRML Consortium (1998). Living worlds: Making vrml 2.0 applications interpersonal and interoperable, draft 2. `http://www.vrml.org/WorkingGroups/living-worlds/`.

Wavish, P. and Connah, D. (1997). Virtual actors that can perform scripts and improvise roles. In Johnson, W. L. and Hayes-Roth, B., editors, *Proceedings of the First International Conference on Autonomous Agents*, pages 317–322. ACM Press.

West, A. and Hubbold, R. (1998). Research challenges for systems supporting collaborative virtual environments. In *Proceedings, Collaborative Virtual Environments 98*, page 11, Manchester.

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

Wright, S., Fernando, T., Ward, N., and Cohn, A. (1998). A framework for supporting intelligent traffic within the leeds driving simulator. In *Proceedings of the ECAI'96 Workshop on Intelligent Virtual Environments*.