Motion Control of Virtual Humans

Marc Cavazza and Rae Earnshaw
University of Bradford

Nadia Magnenat-Thalmann
University of Geneva

Daniel Thalmann
Swiss Federal Institute of Technology

This article surveys virtual humans and techniques to control the face and body. It also covers higher level interfaces for direct speech input and issues of real-time control.

Computer animation technologies let users generate, control, and interact with life-like human representations in virtual worlds. Such worlds may be 2D, 3D, real-time 3D, or real-time 3D shared with other participants at remote locations. Computer graphics techniques let animators create very realistic human-like characters. Although the computational requirements remain significant (especially for real-time generation), the advances in hardware mean that animators can work mostly on the desktop and, in the near future, on the PC. Networking technologies have become more transparent, easy to use, and pervasive, allowing users to treat local and remote information in the same virtual space without considering physical location.

Producing and interacting with virtual humans requires an interface and some model of how the virtual human behaves in response to some external stimulus or the presence of another virtual human in the environment. If these virtual humans are avatars (representations of virtual environment users), then the behavior they exhibit should reflect that of their owners, especially if other users in the environment need to recognize them. Programming behavioral models with emotional responses and encapsulating them in virtual humans challenges current research.

Computer games successfully implement computer-generated characters in life-like scenarios. Such games let the characters interact with users. However, the game designers must choreograph in advance every possible action the character might make—users can't generate new movements. Thus the characters' "behaviors" remain fixed and predetermined—they can't respond in new ways to new inputs. Facilitating interesting and engaging responses that reflect the way real humans behave, in real time, remains a major challenge for the field.

Motion control

Traditionally, human animation has been separated into facial animation and body animation, mainly because of lower level considerations. Facial animation results primarily from deformations of the face. Controlling body motion generally involves animating a *skeleton*, a connected set of segments corresponding to limbs and joints. Using geometric techniques, animators control the skeleton locally and define it in terms of coordinates, angles, velocities, or accelerations. The simplest approach is motion capture. In key-frame animation, another popular technique, the animator explicitly specifies the kinematics by supplying key-frame values whose "in-between" frames the computer interpolates. In inverse kinematics,[1] a robotics technique, the end link trajectory computes the motion of a chain's links. More recently, researchers have used inverse kinetics[2] to account for the center of mass. Although efficient and easily performed in real time, geometric methods often lack realism. In physics-based methods, animators provide physical data; solving the dynamic equations obtains the motion.

Motion is controlled globally through parameter adjustment[3] methods and constraint-based methods, where animators state in terms of constraints the properties the model should have, without adjusting parameters. For example, Witkin and Kass[4] proposed the concept of space-time constraints for character animation by solving constrained optimization. Cohen[5] took this concept further and used a space-time window to control the animation. See the sidebar for more resources on facial modeling and animation.

Classifying virtual humans

Several methods exist for controlling synthetic actors' motion. For example, Zeltzer[6] classified animation systems as guiding, animator-level, or task-level systems. Magnenat-Thalmann and Thalmann[7] proposed classifying computer animation scenes involving synthetic actors according to the method of controlling motion and the kinds of actor interactions. The nature of privi-

leged information for controlling actors' motions falls into three categories: geometric, physical, and behavioral, giving rise to three corresponding motion-control method categories. More recently, Thalmann[8] proposed four new classes of synthetic actors: participatory, guided, autonomous, and interactive-perceptive.

However, all these classifications ignore facial animation. As motion control becomes increasingly high level, it's appropriate to use a unique classification for virtual humans that includes motion control of the face and body:

■ Pure avatars or clones
■ Guided actors
■ Autonomous actors
■ Interactive-perceptive actors

Pure avatars or clones

Virtual actors must have a natural-looking body and face, and the animation must correlate to the actual body and face. This technique, called the *real-time roto-scopy method*,[9] consists of recording input data from a virtual reality (VR) device in real time and concurrently applying the same data to the virtual actor onscreen. A popular way to animate the body uses sensors like the Ascension Flock of Birds or Polhemus Fastrack.

The video sequence of the user's face may be continuously texture mapped on the virtual human's face. Users must be in front of the camera, so the camera captures the head and shoulders, possibly with the rest of the body. A simple and fast image analysis algorithm finds the bounding box of the user's face within the image. The algorithm requires a head-and-shoulders view and a static (though not necessarily uniform) background. Thus the algorithm compares each image with the background's original image. Since the background is static, the user's presence effects changes in the image, so it's fairly easy to detect their position. This lets users move freely in front of the camera without losing the facial image. Instead of transmitting whole facial images as in the previous approach, this method analyzes the images and extracts a set of parameters describing the facial expression. As in the previous approach, users must stand in front of the camera that digitizes the video images of head-and-shoulder shots. Accurate recognition and analysis of facial expressions from video sequences requires detailed measurements of facial features.

Guided actors

Although driven by users, guided actors do not correspond directly to users' motions. These actors are also a type of avatar based on the concept of a real-time direct metaphor. Participants use input devices to update the virtual actor's position. These devices compute the incremental change in the actor's position (for example, they estimate the rotation and velocity of the body's center).

This approach lets users choose from a set of menu-based predefined facial expressions or movements (animations), as shown in Figure 1. The facial expression driver in this case stores a set of defined expressions and animations and feeds them to a facial representation engine as the user selects them.

Readings in Facial Modeling and Animation

Numerous research efforts have been made in the area of facial modeling and animation in the past 25 years by the following researchers:

F.I. Parke, "Parametrized Models for Facial Animation, *IEEE Computer Graphics and Applications*, Vol. 2, No. 9, 1982, pp. 61-68.

P. Kalra et al., "Simulation of Muscle Actions Using Rational FreeForm Deformations," *Proc. Eurographics 92* (Computer Graphics Forum), NCC Blackwell, Oxford, UK, Vol. 2, No. 3, 1992, pp. 59-69.

A. Pearce, B. Wyvill, and D.R. Hill, "Speech and Expression: A Computer Solution to Face Animation," *Proc. Graphics Interface 86*, Canadian Information Processing Society, Toronto, 1986, pp. 136-140.

N. Magnenat-Thalmann, E. Primeau, and D. Thalmann, "Abstract Muscle Action Procedures for Human Face Animation," *The Visual Computer*, Vol. 3, No. 5, 1988, pp. 290-297.

N. Magnenat-Thalmann and P. Kalra, "A Model for Creating and Visualizing Speech and Emotion," *Aspects of Automatic Natural Language Generation*, R. Dale et al., eds., 6th Int'l Workshop on National Language Generation, Springer-Verlag, Heidelberg, Germany, April 1992, pp. 1-12.

K. Waters, "A Muscle Model for Animating Three Dimensional Facial Expressions," *Computer Graphics* (Proc. Siggraph 87), ACM Press, New York, Vol. 21, No. 4, 1987, pp. 17-24.

S. Platt and N.I. Badler, "Animating Facial Expressions," *Computer Graphics*, (Proc. Siggraph 81), Vol. 15, No. 3, 1981, pp. 245-252.

D. Terzopoulos and K. Waters, "Physically-Based Facial Modeling, Analysis and Animation," *J. Visualization and Computer Animation*, Vol. 1, No. 2, 1990, pp. 73-80.
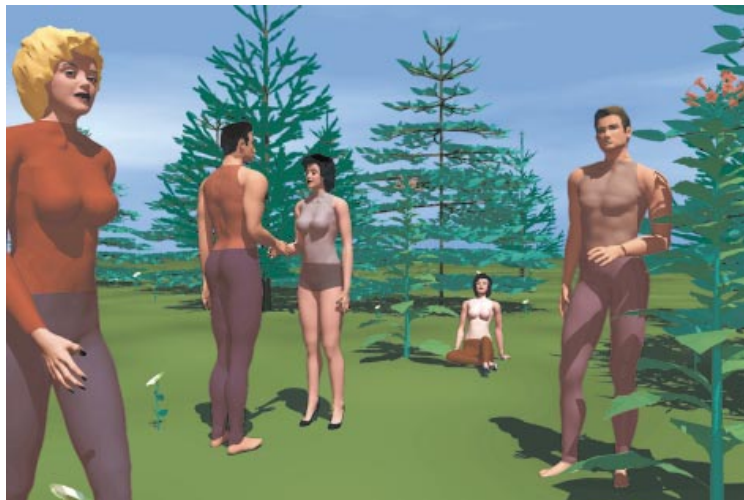
1 Predefined expressions.

Autonomous actors

Autonomous virtual humans should be able to demonstrate a behavior, which means they must have a manner of conducting themselves. Typically, the virtual human should perceive objects and other virtual humans in its environment through visual, tactile, and auditory virtual sensors.[10] Based on the perceived information, the actors' behavioral mechanism will determine the actions they perform. Actors may simply evolve in their environment, interact with this environment, or communicate with other actors. In the latter case, we consider the actor an interactive-perceptive actor.

Renault et al.[11] first introduced the concept of *virtual vision* as a main information channel between the environment and virtual actor. Virtual humans perceive their environment from a small window showing the environment rendered from their point of view. Since they can access the pixels' depth values, the pixels' color, and their own position, they can locate visible objects in their 3D environment.

2 Nonverbal intercommunication.

3 Combat between a participant and an interactive-perceptive actor.

tures and their indications of what people feel (see Figure 2). Postures provide a means of communicating, defined by specific arm and leg positions and body angles.

Enabling real people and virtual humans to communicate requires that virtual actors have some way to sense the real world. Real people become aware of virtual humans' actions through VR tools like head-mounted displays, but one major problem is making virtual actors conscious of real people's behavior. For the interaction between virtual humans and real ones, gesture and facial expression recognition proves a key issue. As an example of gesture recognition, Emering et al.[12] produced a combat engagement between a real person and an autonomous actor (see Figure 3). The real person's motion is captured using the Flock of Birds. The system recognizes the gestures and transmits the information to the virtual actor, who then reacts and decides which attitude to adopt.

The principle remains the same for facial communication, but facial recognition should be mainly video-based. Mase and Pentland[13] used optical flow and principal direction analysis for lip reading. Essa et al.[14] further refined the model. Waters and Terzopoulos[15] animated faces by estimating the muscle contraction parameters from video sequences using "snakes."[16] To obtain a more robust recognition of facial expressions and movements, some people have used external markers and lipstick on the real face.[17,18] Azarbayejani et al.[19] used a Kalman filter to retrieve motion parameters restricted to head motion and orientation. Li et al.[20] used a Candid model for 3D motion estimation for model-based image coding.

Many of these methods do not feature real-time performance. One fast method by Pandzic et al.[21] uses on a "soft mask"—a set of points on the facial image that the user adjusts interactively. Figure 4 shows a virtual actor imitating a real person with expression recognition using the soft-mask method.

As a first step to recreating a virtual audition, we must model an audio environment where virtual humans can directly access an audible sound event's positional and semantic sound source information. For virtual tactile sensors, we could use spherical multisensors attached to the articulated figure. A sensor activates whenever the actor collides with other objects. For facial animation, autonomous actors should have a way of generating spontaneous expressions depending on perception and emotional state.

Interactive-perceptive actors

We define an interactive-perceptive virtual actor as an actor aware of other actors and real people. Of course, we assume that these actors are autonomous. Moreover, they can communicate interactively with other actors and real people.

For communication between virtual actors, behavior may also depend on the actors' emotional state. Facial emotions and speech may be coordinated between virtual actors. Nonverbal communication concerns pos-

Networked real-time synthetic actors

Virtual humans also prove a key issue in networked virtual environments (VEs). For example, the VLNet[22] (Virtual Life Network) system supports a networked shared VE that lets multiple users interact with each other and their surroundings in real time. Avatars, or 3D virtual human actors, represent users so that they can interact with the environment and other avatars. In addition to guided actors, the environment can also

include autonomous and interactive-perceptive actors as a friendly user interface to different services. Virtual humans can also represent currently unavailable partners, allowing asynchronous cooperation between distant partners.
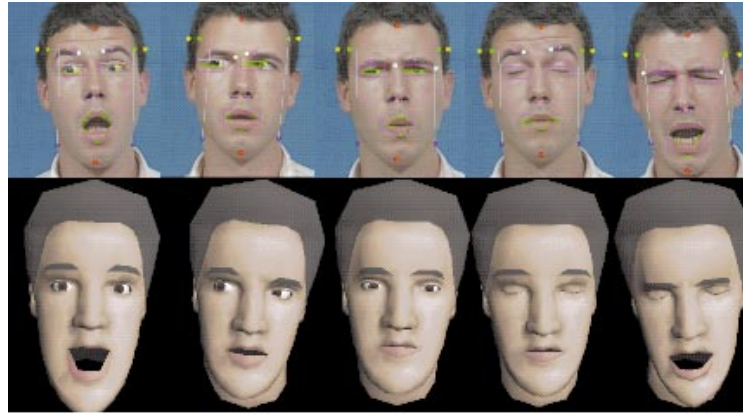
Table 1 compares the various types of actors for several types of actions.

Rationale for natural language control

As virtual humans become more sophisticated, it seems desirable to control these actors as you would direct human actors—by giving them natural language instructions during the animation. While current communication with virtual humans is limited, recent progress in speech recognition and understanding makes this a possible goal.

However, we need to discuss specific requirements for spoken interaction with animated agents, especially autonomous or guided actors. First, we must achieve "real-time" speech understanding—mandatory, for example, in controlling guided actors, but also necessary for autonomous actors.

Second, the system must include high-level representations to control agent behavior. These representations serve as the interface between the agent control mechanism and speech understanding component. Obviously, they differ for guided actors and autonomous actors. For guided actors, translating a predefined behavior into elementary steps constitutes the appropriate motion. Conversely, autonomous actors can implement more complex models that account for their interaction with the environment or other actors. In the following sections, we'll illustrate these points by discussing the possible relations between speech understanding and some behavioral modeling techniques that control autonomous actors.

Various interface paradigms can control virtual humans:

4 Facial expression recognition using a soft mask.

Table 1. Types of actors for several types of actions.

Actions	Pure Avatars	Guided Actors	Autonomous Actors	Interactive-Perceptive Actors
Walking on flat terrain	Only possible with full sensors	Requires a walking motor and procedure to input velocity and step length	Requires a walking motor	Requires a walking motor
Walking on nonflat terrain	Only possible with full sensors	No current model	Difficult to have a general walking motor; genetic algorithms are promising	Difficult to have a general walking motor; genetic algorithms are promising
Grasping	Possible with gloves, but hard to implement feedback; use of fine tuning can help	Requires inverse kinematics	Possible with an automatic procedure	Possible with an automatic procedure
Obstacle avoidance	Not possible	Could be implemented using algorithms based on graph theory	Vision-based navigation	Vision-based navigation
Gesture recognition	Not provided	Not provided	Generally not provided	Should be provided
Facial animation	Possible with video camera	Provided by a few parameters	Model-based animation	Should be provided
Communication with actors	Communication is not computer-generated	Communication is not computer-generated	Limited communication	Perception-based communication (verbal and nonverbal)
Communication with user	Corresponds to a communication between users	Corresponds to a communication between users	Generally not provided	Easy through the avatar corresponding to the user

1. classical input devices such as keyboards and pointing devices,
2. motion-capture techniques operated by a human animator, and
3. language understanding.[23]

Animation control through a keyboard mainly concerns avatars and guided actors; performing arts professionals don't accept it easily.[24] Motion capture—certainly the most natural interface—essentially controls pure avatars and reflects only the operators' actual movements, which can limit certain complex animations. Using speech understanding to control virtual actors offers the ability to convey high-level, abstract instructions that can be mapped to the actors' high-level behaviors. Plus, it's a user-friendly interface.[25] Through its relation to high-level behavioral models, speech understanding could eventually let a human director control an artificial actor without a human animator's assistance by using "artistic" concepts to match the agent emotional behavior. This interface applies equally to guided actors and autonomous or interactive-perceptive actors.

Real-time issues in the speech-based control of virtual humans

The performance of speech recognition systems has progressed considerably in recent years. Although large-vocabulary, speaker-independent speech recognition remains an active research area, speech recognition might belong among professional animation applications. Most recent high-end commercial systems (like the Nuance Speech Recognition System) offer ways to define speech input that will ensure high-recognition scores while retaining sufficient flexibility. With commercial systems providing acceptable solutions for speech recognition, most efforts in speech understanding have concentrated on developing natural language processing software. Such software connects strongly to the target representations required by the behavioral animation module.

However, not only should the speech understanding component produce accurate results, it should do so in real time. This notion depends strongly on the application requirements, and interactive animation will likely be a demanding application in this regard. In most speech-based interfaces, response time for the global sequence of operations should be kept to less than one second, but requirements for speech control of animation might be even more stringent.

Consider the case of a director instructing a guided actor's behavior on a virtual stage. The moving agent should respond as quickly as possible to any instructions requiring it to start, alter, or terminate some action sequence such as walking, jumping, or adopting a specific attitude. Interactive-perceptive actors would also have to react almost instantaneously to any directions given them, especially when interacting with human actors or their avatars. However, the need for real-time language understanding has one exception: when instructing an autonomous actor to carry out a specific task through interacting with its environment. Well-documented examples of this kind of control come from the AnimNL project[23] and the Sodajack application[26] based on the "Jack" character.

Relevant natural language processing (NLP) techniques include syntactic analysis (or parsing) and semantic interpretation. The latter comprises both lexical processing and knowledge-based inference, which converts sentence meaning into the appropriate behavioral concepts controlling the actor. Many different approaches have built on various syntactic and lexical formalisms, but the most successful systems succeed in integrating syntactic and semantic processing to produce accurate analysis in real time.

For our ongoing experiments in speech control of animation, a variant of the Tree-Adjoining Grammars (TAG) has been adopted.[27] This variant, a simplification of the original TAG formalism, leaves out some complexity while incorporating semantic processing within an elementary parsing algorithm.[28] We chose this variation because it implements a linguistic formalism that retains good descriptive power. It could also integrate semantics properly at runtime. In this regard, no separation between structural processing of a sentence and its application-oriented interpretation exists.

The most recent implementation of this NLP component achieves real time for users, processing 10- to 15-word sentences in the 20 to 200 milliseconds range on a 150-Mhz R10000 Silicon Graphics O2 workstation (depending on sentence complexity and syntactic ambiguities). The system—designed for use in conjunction with the Nuance speech recognition system—was developed specifically for speech understanding in sublanguage applications. We believe that actor directions could fit into this category. The limiting step for the whole system is thus likely to be speech recognition.

This approach to optimizing speech understanding aims to achieve real-time control of artificial actors. For animation applications, envision an additional option that would slow down the whole animation to match the speech understanding system's response time. However, this would be acceptable only insofar as it would not impair judgments on the overall agent performance, which still need investigation.

Language interpretation and behavioral models

As we have seen, natural language instructions can direct both guided actors and autonomous actors. In the former situation, the problem mainly consists in controlling animation rules, and natural language can be seen as the "ultimate scripting language." For instance, Beardon and Ye[29] developed a simple NLP interface for a set of behavioral rules controlling animation. They described their system as an evolution of scripting approaches and emphasized the system's user surveyability.

The best use of natural language directions would certainly be achieved with autonomous actors. This is further supported by the fact that natural language interfaces prove appropriate when the application creates rules for action.[25] Autonomous actors or interactive-perceptive actors can use various behavioral models for executing goal-oriented autonomous actions, inter-

acting with their environment, or perceiving and communicating a certain range of emotions.

Actually, controlling complex animations frequently involves patterns that a specific behavior module supervises. Perlin and Goldberg[30] explicitly introduced a distinction between low-level animation control and high-level behavior modules in their Improv system architecture. In this implementation, a behavior engine determines which animation to trigger, making recourse to probabilistic modeling.

The Persona project at Microsoft[31] addressed the animation of an autonomous interface agent and distinguished representations for high-level and low-level behaviors. While the main interaction loop (based on speech understanding) controls the former, the animation context can trigger the latter directly. Similarly, Magnenat-Thalmann and Thalmann[32] introduced a distinction between general motion control and more complex behavioral patterns.

The best work described on language-based animation is the AnimNL project at the University of Pennsylvania.[23] The AnimNL project focused on natural language communication with agents for human factors analysis or group training. It concentrated on understanding instructions such as "go into the kitchen and get the coffee pot," executed by the agent without further human intervention. On the other hand, strongly interactive natural language instructions—such as those issued by a human director to an artificial actor—can differ significantly from the kind of instructions described in the AnimNL system. Interactive instructions influence the course of the animation sequence, a fact that can impose stringent requirements on the response time of the speech-based interface.

Most of the above representations of actors' behavior rely on some kind of planning technology. However, because the required reactive, real-time planning algorithms are extremely difficult to implement and relate to the application properties, researchers often take a pragmatic approach to planning with emphasis on application-specific knowledge. Most of the applications described here seem to follow this approach.

In the Persona project,[31] plans are precompiled into finite-state automata. More specifically, this approach exploits coherence in the search space, which lets the system adopt a simplified approach at runtime. The Sodajack[26] architecture—part of the AnimNL project—includes a specific hierarchical planning algorithm, ITPlans. Sodajack's planning approach interleaves planning and action. It also relies on the context and task description to chose among known plans. Finally, Magnenat-Thalmann and Thalmann[32] described using displacement local automata (DLA) for the high-level control of motion from a goal-oriented perspective. Because of their relations to Scripts,[33] DLAs can also be considered a form of precompiled plans. The rationale for using plans is that they interface with both the agent world (through the actions they trigger) and the high-level behavior or natural language instructions (through the goals they process).

An even more ambitious control mechanism for agent behavior comes from the explicit representations of agent *intentions*, which Webber et al.[23] developed specifically. The most detailed work on spontaneous behavior has focused on controlling animats,[34] for which the term "intention" would not be appropriate despite its use by Beardon and Ye.[29] Note that in most cases, agent intentions are specifically "cognitive" rather than "emotive," following one of the main paradigms of cognitive science. Thus they're closer to traditional work in artificial intelligence than to work on artificial personality modeling, though these two approaches may get closer in the future when cognitive processing associates with emotional modeling in artificial actors.

Finally, because animation often involves mastering complex spatial relations, it's necessary to devise proper linguistic formulations for these relations. Clay and Wilhelms[35] described a system for controlling spatial layouts, not unlike virtual studios or virtual stages, strongly inspired by spatialized theories of language. But some of the problems they address would be more easily solved through multimodal approaches, in which spatial information is processed through pointing gestures and integrated with natural language input.[36] Though multimodal interaction has focused mostly on interface applications involving direct object manipulation, it easily extends to virtual actor supervision, where the gesture component of multimodal interaction would instruct the agent about specific directions or objects. This would thus close the loop of communication by integrating the various communications channels available to virtual humans introduced in this article.

The future of human language technologies in animation should not be restricted to user-friendly interfaces; they're appropriate for integrating control over abstract behaviors for artificial agents. This strongly relates to developing behavioral or abstract models of animation control that can also be enriched with application-oriented high-level primitives such as style.

Future research directions

In "Alan Alda meets Alan Alda 2.0" Hodgins[37] described a project to build a digital twin of actor Alan Alda and regenerate the actor's voice for the synthetic character by separating and resequencing phonemes. For other characters, Hodgins described virtual humans that move according to the laws of physics and can run, dive, bicycle, and vault. Examples of simulated human movement may be found at http://www.cc.gatech.edu/gvu/animation/. Hodgins[38] also described an algorithm for automatically adapting simulated behaviors to new characters.

Current research directions in virtual humans and their control include

- Generating and controlling virtual humans in real time
- Creating characters with individuality and personality who react to and interact with other real or virtual people (that is, manifest intelligent behavior)
- Generating appropriate and context-sensitive behaviors via higher level models
- Connecting language and animation

- Constructing a semantic representation of actions, objects, and agents that exploits natural language expression and also produces animation
- Enabling virtual humans to perform complex tasks
- More effectively linking real humans with virtual worlds for collaboration, communication, and information exchange
- Real-time speech understanding by virtual humans representing real actors
- Evaluating the effect of presence in shared VEs and the influence of virtual humans upon presence
- Exploring human factors associated with realistic looking models that move with the user but also involve subtle relationships among the user, interface, task, user's involvement in the task, and emotional responses provided by human-like objects in the virtual world

Badler[39] and others perceive the future of virtual humans as having many facets:

- Greater human centric design criteria at early stages in the design process
- Enhanced presentation of information (training, collaboration, mentoring, and coaching)
- Surrogates for medical and surgical training and telemedicine
- Personalized information spaces on the Internet

We agree. All these areas show tremendous potential, and we look forward to continuing developments in the many different approaches. ∎

References

1. N.I. Badler et al., "Positioning and Animating Figures in a Task-Oriented Environment," *The Visual Computer*, Vol.1, No.4, 1985, pp. 212-220.

2. R. Boulic, R. Mas, and D. Thalmann, "Complex Character Positioning Based on a Compatible Flow Model of Multiple Supports," *IEEE Trans. in Visualization and Computer Graphics*, Vol. 3, No. 3, 1997, pp. 245-261.

3. W.W. Armstrong, M. Green, and R. Lake, "Near Real-Time Control of Human Figure Models," *IEEE CG&A*, Vol.7, No. 6, Nov. 1987, pp. 28-38.

4. A. Witkin and M. Kass, "Space-Time Constraints," *Computer Graphics* (Proc. Siggraph 88), ACM Press, New York, 1988, pp. 159-168.

5. M.F. Cohen, "Interactive Space-Time Control for Animation," *Computer Graphics* (Proc. Siggraph 92), 1992, pp. 293-302.

6. D. Zeltzer, "Towards an Integrated View of 3D Computer Animation," *The Visual Computer*, Vol. 1, No. 4, 1985, pp. 249-259.

7. N. Magnenat-Thalmann and D. Thalmann, "Complex Models for Visualizing Synthetic Actors," *IEEE CG&A*, Vol. 11, No. 5, Sept. 1991, pp. 32-44.

8. D. Thalmann, "A New Generation of Synthetic Actors: the Interactive Perceptive Actors," *Proc. Pacific Graphics 96*, National Chiao Tung University Press, Hsinchu, Taiwan, 1996, pp. 200-219.

9. D. Thalmann, "Using Virtual Reality Techniques in the Animation Process," *Virtual Reality Systems*, Academic Press, San Diego, 1993, pp. 143-159.

10. H. Noser et al., "Navigation for Digital Actors Based on Synthetic Vision, Memory, and Learning," *Computers and Graphics*, Pergamon Press, Exeter, UK, Vol. 19, No. 1, 1995, pp. 7-19.

11. O. Renault, N. Magnenat-Thalmann, and D. Thalmann, "A Vision-Based Approach to Behavioral Animation," *J. Visualization and Computer Animation*, Vol. 1, No. 1, 1990, pp. 18-21.

12. L. Emering, R. Boulic, and D. Thalmann, "Interacting with Virtual Humans through Body Actions," *IEEE CG&A*, Vol. 18, No. 1, Jan./Feb. 1998, pp. 8-11.

13. K. Mase and A. Pentland, "Automatic Lipreading by Computer," *Trans. Inst. Elect. Information and Communication Eng*, Vol. J73-D-II, No. 6, 1990, pp. 796-803.

14. I. Essan and A. Pentland, "Facial Expression Recognition using Visually Extracted Facial Action Parameters," *Proc. Int'l Workshop on Automatic Face and Gesture Recognition*, Zurich, Switzerland, 1995.

15. K. Waters and D. Terzopoulos, "Modeling and Animating Faces using Scanned Data," *J. Visualization and Computer Animation*, Vol. 2, No. 4, 1991, pp. 123-128.

16. D. Terzopoulos and K. Waters, "Techniques for Realistic Facial Modeling and Animation," *Proc. Computer Animation 91*, Springer-Verlag, New York, 1991, pp. 59-74.

17. E.M. Caldognetto et al., "Automatic Analysis of Lips and Jaw Kinematics in VCV Sequences," *Proc. Eurospeech '89 Conf. 2*, European Speech Communication Assoc. (ESCA), Grenoble, France, 1989, pp. 453-456.

18. E.C. Patterson, P.C. Litwinowich, and N. Greene, "Facial Animation by Spacial Mapping," *Proc. Computer Animation 91*, Springer-Verlag, New York, pp. 31-44, 1991.

19. A. Azarbayejani et al., "Visually Controlled Graphics," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 6, 1993, pp. 902-905.

20. H. Li, P. Roivainen, and R. Forchheimer, "3D Motion Estimation in Model-Based Facial Image Coding," *IEEE Trans. Pattern Analysis and Machine Intelligence*, Vol. 15, No. 6, 1993, pp. 545-555.

21. I.S. Pandzic et al., "Real-Time Facial Interaction," *Displays*, Vol. 15, No. 3, 1994, pp. 157-163.

22. T.K.Capin et al., "Virtual Human Representation and Communication in the VLNET Networked Virtual Environments," *IEEE CG&A*, Vol. 17, No. 2, March/April 1997, pp. 42-53.

23. B. Webber et al., *Instructions, Intentions, and Expectations*, IRCS Report 94-01, Institute for Research in Cognitive Science, University of Pennsylvania, 1994.

24. C.R. Reeve and I.J. Palmer, "Virtual Rehearsal over Networks," to be published in *Proc. Digital Convergence*, British Computer Society, Bradford, UK.

25. P.R. Cohen and S.L. Oviatt, "The Role of Voice in Human-Machine Communication," *Voice Communication between Humans and Machines*, D. Roe and J. Wilpon, eds., National Academy of Sciences Press, Washington, DC, 1994, pp. 34-75.

26. C. Geib, L. Levison, and M.B. Moore, *Sodajack: An Architecture for Agents that Search for and Manipulate Objects*, Tech. Report MS-CIS-94-13/Linc Lab 265, Dept. of Computer and Information Science, University of Pennsylvania, 1994.

27. A. Joshi, L. Levy, and M. Takahashi, "Tree Adjunct Gram-

mars," *J. Computer and System Sciences*, Vol. 10, No. 1., 1975, pp. 136-163.

28. M. Cavazza, "An Integrated TFG Parser with Explicit Tree Typing," to be published in *Proc. Fourth TAG+ Workshop*, University of Pennsylvania, 1998.

29. C. Beardon and V. Ye, "Using Behavioral Rules in Animation," *Computer Graphics: Developments in Virtual Environments*, Academic Press, San Diego, 1995, pp. 217-234.

30. K. Perlin and A. Goldberg, "Improv: A System for Scripting Interactive Actors in Virtual Worlds," *Computer Graphics* (Proc. Siggraph 96), ACM Press, New York, 1996, pp. 205-216.

31. D. Kurlander and D.T. Ling, "Planning-Based Control of Interface Animation," *Proc. CHI 95 Conf.*, ACM Press, New York, 1995, pp. 472-479.

32. N. Magnenat-Thalmann and D. Thalmann, "Digital Actors for Interactive Television," *Proc. IEEE (*Special Issue on Digital Television, Part 2), Vol. 83, No. 7, July 1995, pp. 1022-1031.

33. R.C. Schank and R.P. Abelson, *Scripts, Plans, Goals, and Understanding: An Inquiry into Human Knowledge Structures*, Lawrence Erlbaum Associates, Hillsdale, N.J., 1977.

34. X. Tu and D. Terzopoulos, "Artificial Fishes: Physics, Locomotion, Perception, Behavior," *Proc. Siggraph 94*, ACM Press, New York, Vol. 29, No. 3, 1994.

35. S.R. Clay and J. Wilhelms, "Put: Language-Based Interactive Manipulation of Objects*," IEEE CG&A*, Vol. 16, No. 2, March 1995, pp. 31-39.

36. R.A. Bolt, "Put That There: Voice and Gesture at the Graphics Interface," *Computer Graphics* (Proc. Siggraph 80), ACM Press, New York, Vol. 14, No. 3, 1980, pp. 262-270.

37. J.K. Hodgins, "Animation Human Motion," *Scientific American*, Vol. 278, No. 3, March 1998, pp. 46-51, http://www.sciam.com/1998/0398issue/0398hodgins.html.

38. J.K. Hodgins and N.S. Pollars, "Adapting Simulated Behaviors for New Characters," *Computer Graphics*, (Proc. Siggraph 97), ACM Press, New York, 1997, pp. 153-162.

39. N.I. Badler, "Real-Time Virtual Humans," *Proc. Pacific Graphics 97*, IEEE CS Press, Los Alamitos, Calif., 1997, pp. 4-13, http://www.cis.upenn.edu/~badler/paperlist.html.

Rae Earnshaw *is professor and head of electronic imaging and media communications at the University of Bradford, UK. His research interests include imaging, graphics, visualization, animation, multimedia, virtual reality, art, design, and the convergence of computing, telephony, media, and broadcasting. He obtained his PhD in computer science from the University of Leeds. He is a member of the editorial boards of The Visual Computer, IEEE Computer Graphics and Applications, and The Journal of Visualization and Computer Animation, and as managing editor of Virtual Reality, vice-president of the Computer Graphics Society, chair of the British Computer Society Computer Graphics and Displays Group, and a fellow of the British Computer Society. He is a member of ACM, IEEE, and Eurographics.*

Nadia Magnenat-Thalmann *has researched virtual humans for more than 20 years. She studied psychology, biology, and chemistry at the University of Geneva and obtained her PhD in computer science in l977. In l989 she founded Miralab, an interdisciplinary creative research laboratory at the University of Geneva. Some recent awards for her work include the l992 Moebius Prize for the best multimedia system awarded by the European Community, "Best Paper" at the British Computer Graphics Society congress in l993, to the Brussels Film Academy for her work in virtual worlds in 1993, and election to the Swiss Academy of Technical Sciences in l997. She is president of the Computer Graphics Society and chair of the IFIP Working Group 5.10 in computer graphics and virtual worlds.*

Daniel Thalmann *researches real-time virtual humans in virtual reality, networked virtual environments, artificial life, and multimedia at the Swiss Federal Institute of Technology (Ecole Polytechnique Fédéral de Lausanne—EPFL). He received a diploma in nuclear physics in 1970, a certificate in statistics and computer science in 1972, and a PhD in computer science (cum laude) in 1977 from the University of Geneva. He is co-editor-in-chief of the Journal of Visualization and Computer Animation, member of the editorial board of the Visual Computer, CADDM Journal (China Engineering Society) and Computer Graphics (Russia). He is co-chair of the Eurographics Working Group on Computer Simulation and Animation and member of the executive board of the Computer Graphics Society.*

Marc Cavazza *is a professor of digital media in the Dept. of Electronic Imaging and Media Communications at the University of Bradford, UK. He has been working in the field of natural language understanding for more than 10 years and has taken part in developing several large fully implemented research prototypes, both for text understanding and speech understanding. His areas of expertise include both syntactic parsing in the Tree-Adjoining Grammar formalism, lexical semantics, and syntax-semantics integration in natural language understanding systems. Since 1995 he has served as an expert for the European Commission Directorate General XIII and III in the fields of natural language processing, multimedia systems, and telepresence—on programs in Advanced Communications, Technologies, and Services (ACTS); Telematics Language Engineering; and IT Long-Term Research.*

Readers may contact Earnshaw at the Dept. of Electronic Imaging and Media Communications, University of Bradford, Bradford BD7 1DP, UK, e-mail R.A.Earnshaw@bradford.ac.uk, http://www.eimc.brad.ac.uk/