

Multi-modal Authoring Tool for Populating a Database of Emotional Reactive Animations

Alejandra García-Rojas, Mario Gutiérrez, Daniel Thalmann and Frédéric Vexo

Virtual Reality Laboratory (VRlab)
École Polytechnique Fédérale de Lausanne (EPFL)
CH-1015 Lausanne, Switzerland
{alejandra.garciarojas,mario.gutierrez,daniel.thalmann,frederic.vexo}@epfl.ch

Abstract. We aim to create a model of emotional reactive virtual humans. A large set of pre-recorded animations will be used to obtain such model. We have defined a knowledge-based system to store animations of reflex movements taking into account personality and emotional state. Populating such a database is a complex task. In this paper we describe a multimodal authoring tool that provides a solution to this problem. Our multimodal tool makes use of motion capture equipment, a handheld device and a large projection screen.

1 Introduction

Our goal is to create a model to drive the behavior of autonomous Virtual Humans (VH) taking into account their personality and emotional state. We are focused on the generation of reflex movements triggered by events in a Virtual Environment and modulated by inner variables of the VH (personality, emotions). We intend to build our animation model on the basis of a large set of animation sequences described in terms of personality and emotions. In order to store, organize and exploit animation data, we need to create a knowledge-based system (animations database).

This paper focuses on the authoring tool that we designed for populating such animation database. We observe that the process of animating is inherently multi-modal because it involves many inputs such as motion capture (mocap) sensors and user control on an animation software. For simplifying the process of animating we propose to integrate the required inputs into a multi-modal interface composed of a handheld device (providing a mobile GUI), motion capture equipment and a large projection screen (to ease the interaction with virtual characters). One of the added values of our system is that it provides an immersive multimodal environment for animating characters. Our tool allows for interacting within the virtual environment as if the user were inside. Animation data produced with this tool is organized considering personality traits and emotional states.

The rest of the paper is organized as follows: next chapter presents related work on multi-modal interfaces and knowledge-based proposals for VH animation. This will be followed by our proposal of system, its architecture and implementation. Finally we present our results and plans for future work.

2 Related Work

Our authoring tool is intended to facilitate the process of producing character animation in an innovative way. Our approach is to animate a character from “inside”, having the actor (animator) immersed in a Virtual Environment and looking at the world through the eyes of the character. For this we need to address multiple interaction modalities: body motion, 3D visualization, etc. Moreover, animation data created by the user should be stored and organized in an efficient way. The following section presents a brief overview on multimodal interfaces research. The second subsection deals with models and strategies for organizing animation data in a knowledge-based system.

2.1 Multi-modal Interfaces

A multi-modal system is a system that has two or more ways of input/output data. The benefit of using multi-modal systems is to get more transparent, flexible, efficient and expressive means of human-computer interaction. A descriptive conception of multimodal interfaces can be found in Oviatt’s work[24].

Multimodal interfaces are implemented in virtual environments (VE) because they help to produce the effect of immersion. This immersion is provided through a natural interaction between the user and the environment. One pioneer multimodal application is the “Media Room” by Bolt [6], this application combines the projection in a screen and user gestures. The fact of positioning one or more users in front of a large rear-projection screen displaying the virtual world is an approach in semi-immersive VE that has given encouraging results.

Examples of systems implementing the semi-immersive approach are: “The Enigma of the sphinx” [1] and the “Magic Wand” [7]. In the same line of research, “Conducting a virtual orchestra” [25] proposes a semi-immersive VE based on a large projection screen, a handheld device and 3D sound rendering. A PDA-based GUI was implemented to conduct the orchestra and a magnetic tracker attached to it to retrieve user’s arm gestures while conducting. Handheld devices are innovate interfaces for VE. Another example of integration of handheld devices in VE can be found in [13]. These works explore the potential of using handheld devices to interact with virtual environments and VH.

Multimodal interfaces allows us to create more immersive virtual environments and they have the advantage of facilitating user interaction. It is interesting to incorporate handheld devices as an interaction device because they reinforce human-human contact in a Virtual Environment, avoiding the need to sit in front of a screen with a mouse and keyboard. We will use these ideas to build our multimodal interface. Next subsection deals with models for organizing animation, data management is essential for exploiting and reusing information.

2.2 Knowledge-based Systems for Virtual Human Animation

The animation model we are developing will require a large amount of animation sequences. We need a database for storing and organizing animation data. The

following is a brief review of research targeted at organizing animation data for Virtual Humans.

The Amoba system[11] uses a database structure to organize motion. This is a very spread and ambiguous organization structure. We require something more specialized and with richer annotations -metadata- describing the content -animation.

The work presented in [16] proposes the use of a database to retrieve the desired animation and manipulate it from a GUI. This work is interesting from the point of view of the process needed to reuse animations and models, but it does not address the problem of populating of the database. Another implementation for reusing animation data is presented in [20] this research considers the importance of a database to animate avatars in real time. But still does not deals with the process of populating the database.

A new approach towards incorporating semantics into Virtual Humans and their animation is presented in [12]. This work intends to define in a formal way, the components of a VH, including its animation, by means of an ontology.

We observe that in order to maximize the reuse and exploitation of animation data, we need to incorporate a semantic layer that enables both computer systems and human users to acquire, organize, and understand the information. For realistic animations, we need to have as much data as possible in the database. The tool we propose intends to facilitate both the data acquisition and organization.

3 Multi-modal Tool for Populating an Animation Database

This section presents the conceptual description of our authoring tool based on the multimodal interaction concepts presented in previous subsection and the requirements for associating the desired data to the animation.

3.1 Multi-modal structure

The Multimodal authoring tool we propose makes use of motion capture, a large projection screen and a handheld device. The motion capture is composed of 6-DOF magnetic sensors and is the main input data to record virtual human movements. Integrating a real time Mocap system allows the animator to acquire high quality animations and give the appropriate intention and expressiveness to each movement.

The second input modality is the handheld device that will work as a remote control. In combination with it, a large screen will provide visual feedback to the user. With the handheld device we reduce the amount of people required in the production process, and make an interface more interactive because the same person that is using the mocap can drive the authoring. Moreover, this mobile interface allows for previewing the animation and accessing the database without the need to be in front of a PC.

The handheld will act as remote control, it will provide a GUI with commands to execute and interact with the environment projected in the large screen. The user can select objects on the screen by pointing with the hand. The selection is confirmed through the GUI in the handheld device. After selecting a VH, the user can start populating the knowledge base system defining parameters in the handheld device to create a particular sequence of an animation. In the handheld device the user can start, stop, save and preview animation sequences (see figure 1). Visualization in the handheld is done with a low resolution model while on the large screen a high-resolution VH model is rendered.



Fig. 1. Previewing animation on the handheld device.

All the data from the handheld device, synchronization with mocap, projection on the screen and transactions in the knowledge base are managed by a central system. During the authoring process the tasks executed in this manager system are transparent for the user.

This section described functionality of the multi-modal implementation, next section describes the way we add an structure based on personality and emotions to organize animation in the knowledge-based system.

3.2 Knowledge-base structure

Based on the aim of creating a model for emotional reactive VH we need to associate traits of personality and emotions to the animations. By considering such inner variables we expect to increase believability of the characters.

Improving believability in computer generated characters is one of the main challenges in computer animation. A believable behavior has many aspects to consider: realism, emotions, personality and intent [9]. There are many models that approach a realistic behavior following the principle of action selection, goal-oriented animation, etc. They frequently use synthesized animations, created by means of specialized algorithms. The most realistic results are obtained with prerecorded animations performed by human actors using mocap. Several models consider that personality and emotional states are the more general traits that influence in VH behavior [21]. This is why we consider interesting to have animations influenced by these traits.

We organize animation in terms of emotions and personality because they are key components of a believable behavior. There are models of personality and emotion for VH that allow to design an emotionally personified virtual human. Among the most complete personality models for virtual humans are: [19] [18] and other less complex such as [10].

We took some of the common factors proposed by the models mentioned before to describe the metadata (attributes describing animation sequence) of our knowledge-based system. To represent personality we use the Five Factor Model (FFM) [22]. The parameters that compose this model are described in table 1. Emotional states are described by the Cognitive Structure of Emotions Model (OCC - Ortony, Clore and Collins) [23]. The OCC model categorizes several types of emotions based on the positive or negative reactions to events, actions, and objects. It defines 22 emotions. Like in the multilayered model [18] we want to integrate 6 basic expressions to represent the emotional states. This emotions are explained and categorized in the table 1.

Five Factor Model of Personality		OCC Model of Emotion	
Factor	Description	Emotion	Description
Extraversion	Preference for and behavior in social situations.	Joy	Happy-for, Gloating, Joy, Pride, Admiration, Love, Hope, Satisfaction, Relief, Gratification, Gratitude
Agreeableness	Interactions with others.	Sadness	Resentment, Pity, Distress, Shame, Remorse
Conscientiousness	Organized, persistent in achieving goals	Anger	Anger, Reproach, Hate
Neuroticism	Tendency to experience negative thoughts	Surprise	Surprise
Openness	Open minded-ness, interest in culture	Fear	Fear, Fear-confirmed
		Disgust	Disgust

Table 1. Models of personality(FFM) and emotion (OCC)

The FFM and OCC models are ideally suited to the task of creating concrete representations of personality and emotions with which to enhance the illusion of believability in computer characters [3].

The models we have defined describe an individualization of VH humans. Each VH is defined by a specific combination of attribute values corresponding to the FFM. Each VH can have many configurations of the attributes of emotional states (OCC model), and each attribute can be defined in different levels.

To describe animation we consider also events in the environment, these events are represented by objects in the VE. For example we can have a sphere that represents a ball. The ball can then produce the event “thrown ball”.

One event is associated to one animation because this event will make the VH move in reaction. One animation is performed as reaction to one event, under one configuration of emotional state for one specific personality of a VH. This conception is translated into a database diagram presented in figure 3.2.

Under this structure, when an event occurs in the environment, the animation engine will look into the knowledge base for an animation suitable to perform a reaction, taking into account the specific conditions of emotional state and personality. Next subsection gives more details on the components of the multi-

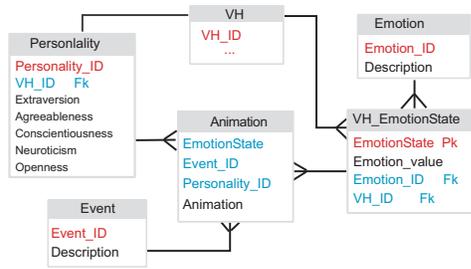


Fig. 2. Diagram of the knowledge-base system

modal interface and how they are interconnected.

3.3 Multi-modal Authoring tool

This tool contains several interactive elements to facilitate data acquisition for the knowledge-based system. Figure 3 illustrates the interaction between elements of the multi-modal tool.

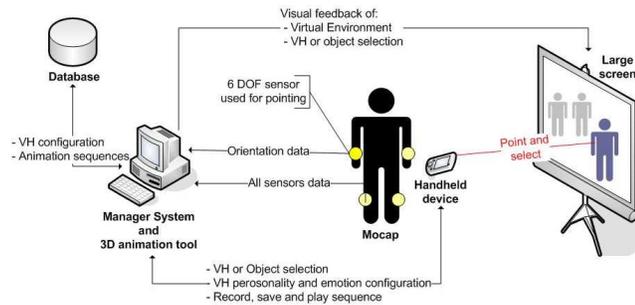


Fig. 3. Multimodal interaction between elements.

System components represented are described as follows:

- Main control of the authoring resides in the Manager system. This Manager receives commands from the handheld device and executes an action in response; it can also send information to the handheld, such as metadata (attributes) or animation data (the actual animation).
- Sensors continuously send information about the actor's position and the Manager system reads them depending of the command in process; it could be reading only the right hand orientation or all the sensors information (recording an animation).
- The Manager makes transactions in the knowledge-base system, can store or retrieve information of the metadata or animation sequences.

- The manager system communicates with the 3D viewer in which the VE is represented. A VE contains VHs and objects that trigger events, the scene is projected in the large screen.

The authoring process is driven through the handheld device. First, the user selects a VH or object by pointing at the large screen and pressing a button in the handheld device. Objects can be associated to an event. Events can be triggered from the handheld device. Depending of the event some of its parameters can be modified, for example a ball can be thrown in different directions.

Personality traits and emotion parameters of a VH can be configured for each sequence to be recorded. Many combinations of parameters of emotional states can be recorded for one event. This configurations are saved in the knowledge base system.

Animating a VH can be done in two modalities: watching the VH mimic the user movements, or viewing the VE through the VH eyes (see figure 7). The character moves in real time according to the data acquired by the mocap system.

For recording, there is a mechanism similar to a VCR with a big button to start and stop recording. To start recording we give 5 seconds for the user to get an initial posture before recording the animation. After those 5 seconds the sequence starts to be recorded until the user presses stop. When stop recording a pause of real time animation is made. When storing the recorded animation, the last three seconds are removed to avoid storing undesired motions due to the movement for pressing the stop button on the PDA. User can Save or discard a sequence. The animation is stored with the current configuration of emotional states, event, personality and VH. Finally, “Stop animating” option stops animating from the mocap and “Stop authoring”, the VH is deselected.

To reproduce animations the user can Play a sequence. It consists in reproducing the last sequence recorded in the large screen, or any of the other sequences saved. This last option will only work when the VH is neither being animated nor in pause. User can also preview recorded sequences in the handheld device (figure 1). We have explained the main components of the multi-modal tool and its work-flow. Next section provides technical details concerning the implementation.

4 Implementation

Motion capture is performed using the motion capture system from Ascension Technology [4], composed of 13 6-DOF magnetic sensors. As interface with the Mocap we used an utility developed at VRlab: Shared Input Devices (SID) [15]. The SID program gets the sensor information and puts it in a shared memory zone in the host machine. The manager program access to this shared memory to obtain the sensors data. This information is translated and sent to the 3D animation tool.

The animation tool used was Maya 5.0[8]. Maya provides a rich set of tools for rendering, modeling and animating. It is one of the leader solutions in the market

and can be considered as a defacto standard in the animation industry. In Maya we modelled VH with H-Anim [14] skeleton. H-Anim is a W3C standard for animating humanoid models. Having H-Anim skeleton in Maya we used a Plug-in presented in [2] to export the animation in MPEG-4 BAP [17]. This encoding animation is a low bit-rate representation suitable for networked applications.

Data from mocap sensors are sent to Maya using Maya’s motion capture API. We created a component that uses this API and also applies the right transformation to the raw sensors data. Each 6-DOF sensor needs to be calibrated to provide its data in the same coordinate space used by Maya. Calibration was implemented following the method described in [5]. We implemented in maya several MEL (Maya Embedded Language) scripts for: linking the sensors data with the skeleton used to animate the VH; start and stop recording movements and exporting animations. This commands are executed by the Manager system.

As handheld device we used a PDA iPAQ HP 4700. We built the GUI for the PDA in C# with controls to drive the animation process. To preview the recorded animations on the PDA we switch application and use a mobile 3D viewer based on the “Mobile Animator” [13].

The knowledge-based system was set up in MySQL. This database is free, easy to implement and provides enough performance for our needs. Data transmission to and from the database is done using ODBC for MySQL.

The Manager system was implemented in C++. The communication with the PDA is done through sockets. Interaction with maya is done through the “Command Port” interface (MEL command: commandPort).

The components diagram in figure 4 shows the interaction between the described components.

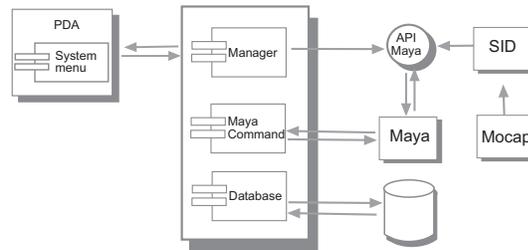


Fig. 4. Component diagram implementation.

Virtual Humans modeled in Maya are animated through their H-Anim skeleton. We used inverse kinematics to compute proper joint rotation values for the VH limbs, and applied orientation constraints to some joints (root, column and skull) with the suitable weights. We created in Maya one locator for each sensor and constrained them to the proper effector or joint. The association of the sensors in the actor with the mocap and the locators in Maya and the skeleton are shown in figure 5.

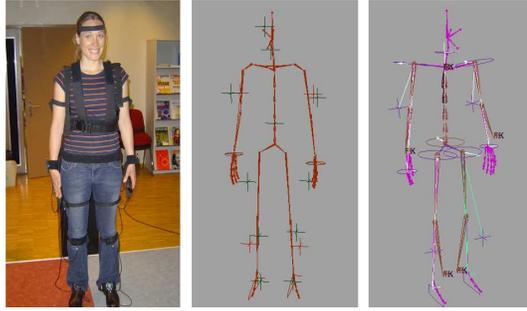


Fig. 5. Association of the mocap's sensors with the skeleton effectors/joints.

The GUI in the PDA has a “Tab menu” for selecting VH or objects, configuring personality, emotions or events and recording animation. Events can be selected from the PDA and modified and activated. The data configured in the PDA is sent as a character stream.

For selecting VHs or objects we calibrate the initial position of the right hand sensor as the center of the screen and transform sensor orientation into 2D coordinates corresponding to the computer screen. 2D screen coordinates are used to drive the mouse pointer position. This way the user can naturally interact just by pointing at the screen.

To record an animation the user has easy access to the button that starts recording, this button serves at the same time to stop recording. Screen shots of the GUI in the PDA of each menu are shown in figure 6. VH can be animated in two modalities, watching the character on the screen from a 3rd person perspective or from a first person perspective, watching the Virtual Environment through the eyes of the VH (figure 7).

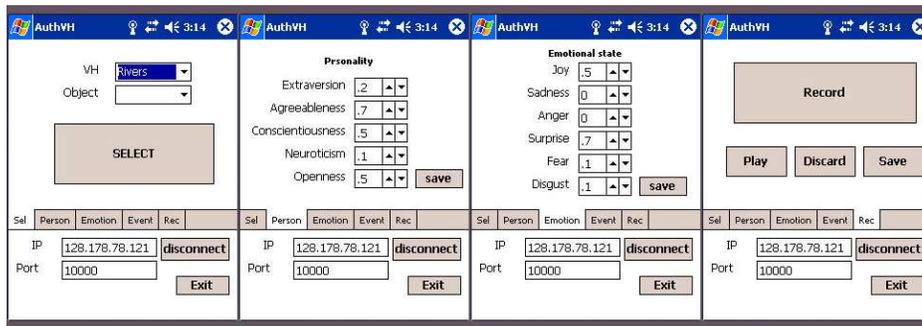


Fig. 6. Screen shots of the GUI in the PDA.

To save animations we use one a MEL script to execute a plug-in that exports the skeleton animation to MPEG-4 BAP format. The BAP file created is placed into a shared directory in the host machine and the path saved in the database. If the user wants to play the animation on the PDA the Manager sends the path and name of the animation and the Mobile animator is able to preview the animation.

5 Discussion and Results

To start populating the database, we have tested our multi-modal tool with different examples. A typical example is recording different reactions to a simple event such as a ball being thrown towards the VH. We defined the event “Ball thrown”. We recorded movements with our user in the mocap. Figure 7 shows the animation process: the user configures a VH, performs an animation, and previews the saved animation on the PDA.

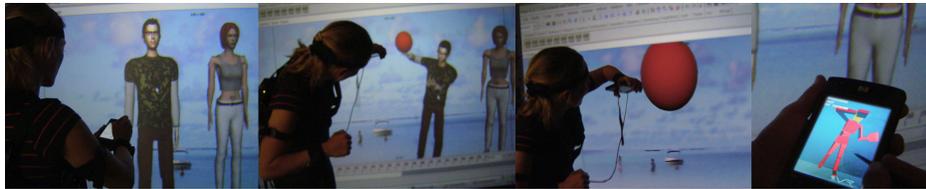


Fig. 7. Using the multimodal authoring tool.

The combination of Mocap with a large projection screen and a lightweight interface (handheld) proved to be an efficient and intuitive way to produce multiple sequences of realistic animation. The handheld device could be enhanced with speech recognition, in particular for the start/stop recording functionality. The multimodal interface provided a useful mechanism for populating the animation database that we have defined. Describing animation sequences by means of metadata introduces a semantic layer that promote the reuse and increases the productivity of animation.

6 Conclusions and Future Work

We have presented a multi-modal tool to populate a knowledge based system of VH animation. The tool involves the use of motion capture, a handheld interface, a semi immersive VE (large projection screen) and the incorporation of semantics (metadata) to the animation. Animation data is organized in a knowledge base taking into account personality traits and emotional state of the VH. The multimodal interface provides a fast and intuitive tool for populating the

animation database. However there may still be need for fine-tuning animation due to the inherent noise of the Mocap data.

Future work consists on populating the animations repository in order to provide a rich repertoire for the animation model. The model for reactive VH will be described in future publications. The current paper focused on describing the tools for acquiring and organizing the information.

Acknowledgments: This research has been partially funded by the Swiss Federal Office for Education and Science in the framework of the European Networks of Excellence IST - HUMAINE (<http://emotion-research.net>) and IST - AIM@SHAPE (<http://www.aim-at-shape.net>).

References

1. T. Abaci, R. de Bondeli, J. Ciger, M. Clavien, F. Erol, M. Gutierrez, S. Noverraz, O. Renault, F. Vexo, and D. Thalmann. Magic wand and enigma of the sphinx. *Computers & Graphics*, 28(4):477–484, August 2004.
2. Y. Amara, M. Gutierrez, F. Vexo, and D. Thalmann. A maya exporting plug-in for mpeg-4 fba human characters. In *Proceedings of the First International Workshop on Interactive Rich Media Content Production: Architectures, Technologies, Applications and Tools. (RICHMEDIA 2003)*, pages 121–130, 2003.
3. E. André;, M. Klesen, P. Gebhard, S. Allen, and T. Rist. Integrating models of personality and emotions into lifelike characters. pages 150–165, 2000.
4. Ascension Technology Corporation. <http://www.ascension-tech.com/>.
5. C. Babski and D. Thalmann. Real-time animation and motion capture in web human director (whd). In *VRML '00: Proceedings of the fifth symposium on Virtual reality modeling language (Web3D-VRML)*, pages 139–145, 2000.
6. R. A. Bolt. Put-that-there: Voice and gesture at the graphics interface. In *SIGGRAPH '80: Proceedings of the 7th annual conference on Computer graphics and interactive techniques*, pages 262–270, New York, NY, USA, 1980. ACM Press.
7. J. Ciger, M. Gutierrez, F. Vexo, and D. Thalmann. The magic wand. In *SCCG '03: Proceedings of the 19th spring conference on Computer graphics*, pages 119–124, New York, NY, USA, 2003. ACM Press.
8. A. S. Corp. Maya, 3d animation and effects software. <http://www.alias.com>.
9. P. Doyle. Believability through context using "knowledge in the world" to create intelligent characters. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 342–349, New York, NY, USA, 2002. ACM Press.
10. A. Egges, S. Kshirsagar, and N. M. Thalmann. A model for personality and emotion simulation. In *Knowledge-Based Intelligent Information and Engineering Systems (KES2003)*, 2003.
11. S. Grunvogel, J. Piesk, S. Schwichtenberg, and G. Buchel. Amoba: A database system for annotating captured human movements. In *CA '02: Proceedings of the Computer Animation*, pages 98–102. IEEE Computer Society, 2002.
12. M. Gutierrez, D. Thalmann, F. Vexo, L. Moccozet, N. Magnenat-Thalmann, M. Mortara, and M. Spagnuolo. An ontology of virtual humans: incorporating semantics into human shapes. In *Proceedings of Workshop towards Semantic Virtual Environments (SVE05), March 2005*, pages 57–67, 2005.

13. M. Gutierrez, F. Vexo, and D. Thalmann. The mobile animator: Interactive character animation in collaborative virtual environments. In *VR '04: Proceedings of the IEEE Virtual Reality 2004 (VR'04)*, page 125, Washington, DC, USA, 2004. IEEE Computer Society.
14. H-anim. The humanoid animation working group. <http://www.h-anim.org>.
15. B. Herbelin. Shared Input Device Controller, <http://vrlab.epfl.ch/bhbn/birdnet/index.html>.
16. A. Huang, Z. Huang, B. Prabhakaran, and J. C. R. Ruiz. Interactive visual method for motion and model reuse. In *GRAPHITE '03: Proceedings of the 1st international conference on Computer graphics and interactive techniques in Australasia and South East Asia*, pages 29–36. ACM Press, 2003.
17. ISO/IEC 14496-2:1999. Information Technology – Coding of Audio-Visual Objects, Part 1: Systems (MPEG-4 v.2), December 1999. ISO/IEC JTC 1/SC 29/WG 11 Document No. W2739.
18. S. Kshirsagar. A multilayer personality model. In *SMARTGRAPH '02: Proceedings of the 2nd international symposium on Smart graphics*, pages 107–115, New York, NY, USA, 2002. ACM Press.
19. S. Kshirsagar and N. Magnenat-Thalmann. Virtual humans personified. In *AA-MAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 356–357, New York, NY, USA, 2002.
20. J. Lee, J. Chai, P. S. A. Reitsma, J. K. Hodgins, and N. S. Pollard. Interactive control of avatars animated with human motion data. In *SIGGRAPH '02: Proceedings of the 29th annual conference on Computer graphics and interactive techniques*, pages 491–500, New York, NY, USA, 2002. ACM Press.
21. S. Marsella and J. Gratch. A step toward irrationality: using emotion to change belief. In *AAMAS '02: Proceedings of the first international joint conference on Autonomous agents and multiagent systems*, pages 334–341, New York, NY, USA, 2002. ACM Press.
22. R. R. McCrae and O. P. John. An introduction to the five-factor model and its applications. *Journal of Personality*, (60):175–215, 1992.
23. A. Ortony, G. L. Clore, and A. Collins. *The Cognitive Structure of Emotions*. Cambridge University Press, 1998.
24. S. Oviatt. Multimodal interfaces. pages 286–304, 2003.
25. S. Schertenleib, M. Gutierrez, F. Vexo, and D. Thalmann. Conducting a virtual orchestra. *IEEE MultiMedia*, 11(3):40–49, 2004.