The Virtual Human Interface: A Photorealistic Digital Human

Barnabás Takács
Digital Elite

Bernadette Kiss
VerAnim

Our multimodal human–computer interface employs photorealistic virtual humans that can talk, emote, and act adaptively and intelligently in response to the actions of a user in front of a computer monitor.

To mimic the quality of everyday human communication, future computer interfaces must combine the benefits of high visual fidelity with conversational intelligence and—above all—the ability to modulate the emotions of their users. From a computer graphics or visual perspective, achieving this goal requires creating synthetic digital humans that have photorealistic faces capable of expressing the finest shades of emotion. From a perceptual point of view, the same system must also read users' emotional reactions and adapt the behavior of the digital human accordingly. Our research focuses on implementing this concept to create a face-to-face system called the Virtual Human Interface (VHI).

Photorealistic virtual humans—in contrast to nonrealistic virtual characters—make a more effective communication interface between computers and humans because they can stimulate declarative and procedural memory in the human brain. Declarative memory is our memory of information and events. Procedural memory, on the other hand, is the memory of how to do something, such as riding a bike. Most computer interfaces primarily access users' declarative memory; the person learns information about a particular topic. The VHI system, however, engages the user through entertaining emotional responses that help engrave knowledge in the user's mind and even modify the user's behavior through procedural memory.

During the past several decades, researchers have conducted countless studies on agents and human animation. Early work mainly addressed low-polygon virtual humans that users could animate in real time.[1] Many of these animation systems initially addressed purposes other than the needs of human–computer interaction. However, the requirements of facial animation—and especially speech synthesis—demand a different underlying architecture that can effectively model how real faces move and change.[2-4] As a result, research began to focus on creating autonomous agents that could exhibit rich personalities and interact in virtual worlds inhabited by other characters.[5,6]

To provide the illusion of a life-like character, researchers have developed detailed emotional and personality models that can control the animation channels as a function of the virtual human's personality, mood, and emotions.[7-9] The real-time interaction with these virtual characters poses an extra set of technical challenges in terms of the speed, computational power, and visual quality required to make the user believe that he or she is interacting with a living creature. To achieve this goal, researchers eventually replaced precrafted animated actions with intelligent behavior modules that could control speech, locomotion, gaze, blinks, gestures (including various postures), and interaction with the environment. Our VHI system builds on this research by employing photorealistic virtual humans, providing users with information, learning services, and entertainment in a personalized and adaptive manner.

Interface

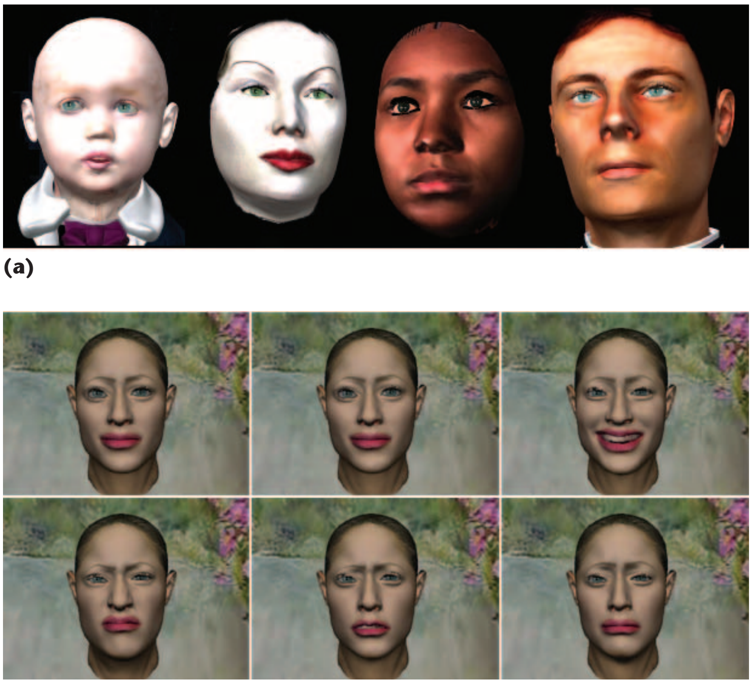
We designed our system to create emotional engagement between the virtual human and the user to increase learning efficiency. The information content broadcast by the digital human (the *what*) is independently modulated through emotion channels (the *how*) and delivered in a focused 3D environment (the *where*). This triad of *what*, *how*, and *where* forms a platform for a communication interface designed to enhance information absorption. We achieve this effect by enabling our virtual human to perceive—specifically, to see, touch, and hear. Specialized vision modules detect and recognize one or more people in front of the display and analyze their facial expressions, looking for signs of their emotional states. We model these emotional states and create a mechanism for the digital human to express its own feelings with the purpose of modulating the user's mood.

The underlying theory supported by neuropsychological evidence is that emotion is a key catalyst in the

 Published by the IEEE Computer Society

process of transforming information to knowledge.[10] Modulating emotions to support the information flow offers a critical advantage over other systems. To produce our content (the *what*), facial-tracking techniques extract and encode the speech, facial expressions, and other nonverbal communication channels of a lecturer, actor, or performer. We also record hand gestures and body motion or create them from motion captured gesture libraries. The recorded material forms a deterministic data stream specific to a given application.

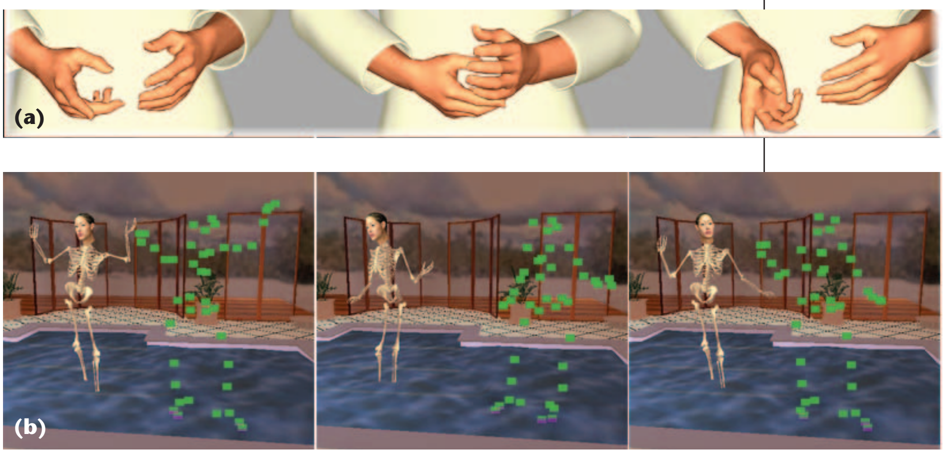
The emotional modulation channel (the *how*) focuses on changing delivered content as a function of advanced multimodal interaction with the user. Finally, the *where* channel stands for the 3D world surrounding the virtual human and the user in a real-time environment. It's responsible for mixing virtual world objects with the user's own surroundings, enhancing the perception of reality. We achieve this enhancement by the combined effect of many system components, such as controlling the main camera motion through head tracking or virtual reality glasses; combining live imagery captured from a video camera to support augmented reality; and using a 3D holographic display device that delivers a 3D experience without glasses or eyestrain.



(a)

(b)

1 **(a) Virtual human faces used in the VHI system. (b) Subtle changes of emotional display.**



(a)

(b)

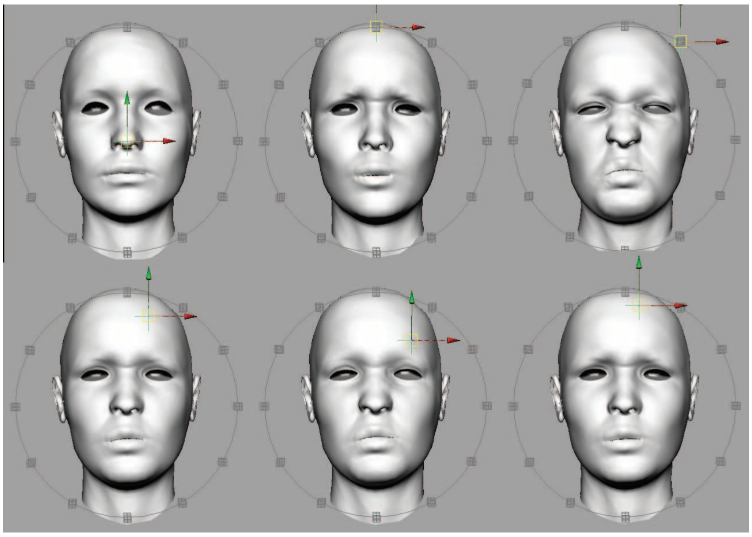
2 **Examples of virtual human (a) hand gestures and (b) body motion.**

Face and body

We built the VHI using our integrated character simulation system that lets you create virtual humans and animate them through high-level artificial intelligence methods. The VHI system builds upon facial and body modeling methods, specialized algorithms, and animation architecture suited for humanoid animation as well as intuitive interfaces to manipulate large numbers of animation channels and methods to create motion patterns through transformations of existing motion libraries.

To achieve our goals, we developed advanced modeling and deformation methods. We started with a large set of high-fidelity 3D scans of real faces and turn these models into standardized, animatable meshes and high-detail textures. For a complete model of a single person, we record up to 60 different expressions corresponding to various emotions and expressions. Figures 1a and 1b show sample 3D head models. The multitrack manager (MTM) unit holds all this information together. It acts as a tape recorder with multiple parallel tracks, each referring to different aspects of the animation process. The body modeling and deformation controlled by the MTM module comprises a fully constrained human skeleton with more than 200 dynamically configurable bones. Figures 2a and 2b demonstrate examples of how the MTM hand and the skeletal structure move the virtual body during interaction with the user.

For real-life applications, it's important to avoid having virtual humans engage in repetitive behavior because it can annoy and distract users, introducing a counter-productive element to information exchange. In the VHI framework, we record the original content—corresponding to the *what* channel—in a neutral newscaster style. The emotional modulation channel (*how*) receives this input and transforms it according to the current internal state of the digital agent, taking into consideration the facial deformation system's constraints. Our mechanism that achieves this effect uses a simple circular configurable device, called the tempo-



3 **Temporal disc controller interface configured for facial expression animation.**

4 **Creating reactive animated emotion sequences using the temporal disc controller mechanism.**

ral disc controller. The TDC (see Figure 3) is a generic interface that controls hundreds of nonlinear deformation channels and applies constraints to them in parallel. The major advantage of the TDC over other systems is that it restricts the complete animation space to a mathematically constrained subspace that encompasses only those facial expression combinations valid for that particular face.

In addition, we can configure the TDC to convey emotions consistent with psychological findings on how faces reflect emotions.[3,11] The TDC mechanism uses the same internal representation to animate an intelligent agent's emotional responses as to encode a user's internal state. This linkage helps the virtual human create an emotional bond with its natural human counterpart and even devise strategies to modulate this emotional bond during the communication process. The granularity of the resulting expression space depends on the number of example surfaces representing each emotion and the frequency of the sampling rate along the curve spanning a particular animation path defined over the disc. Figure 4 demonstrates the TDC mechanism's use in creating animated emotion sequences.

Sense and perception

The ability to perceive is a key element of a truly intelligent agent. We define VHI's input modalities corresponding to vision, touch, and hearing in the context of the communication process and the functionality of the animated character itself. In particular, the three implemented senses connect to information channels that have a direct affect on the animated character's gaze, facial expressions, locomotion, and body gestures.

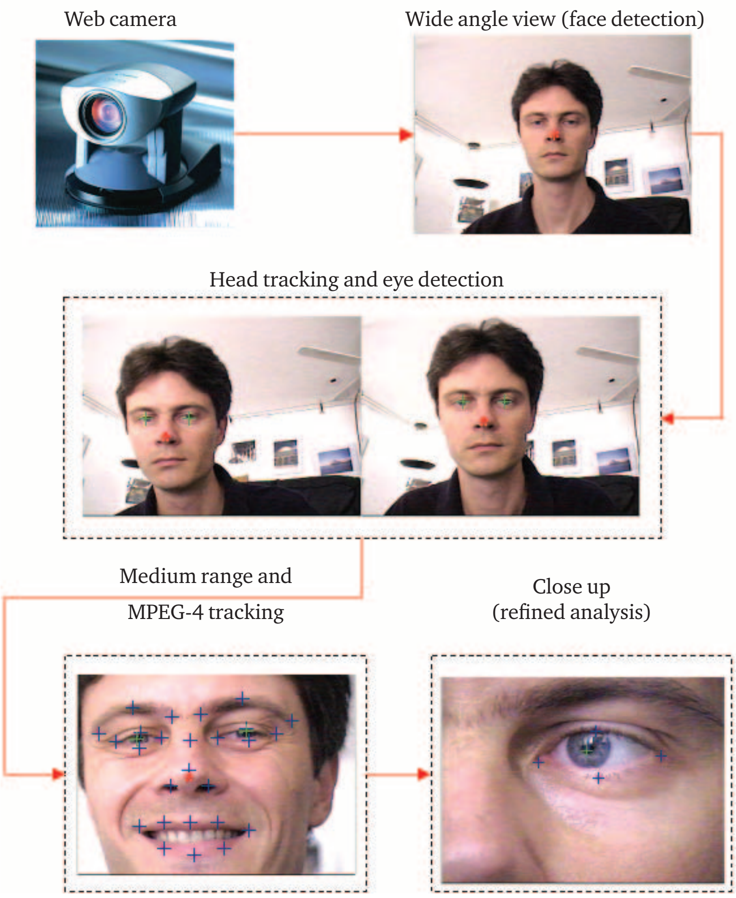
The foundation for implementing this functionality and executing these actions hinges partly on the processing of external visual, auditory, and touch-related signals. VHI connects the user to the synthetic human's 3D world by means of markers. Markers are invisible dimensionless representations of 3D positions attached to any object in the scene. They carry unique hierarchical labels and we can refer to them by their names in describing a high-level task. We could attach multiple markers to a camera, a table, the floor, or the virtual monitor—a special-purpose object in the virtual environment that is analogous to the computer screen. We map the live video stream directly onto this monitor and assign and display results of the visual processing—the locations and identities of the people in front of the terminal—here using temporary markers.

The advantage of the marker-based representation is that it defines and executes at a high level, symbolically, all tasks that the virtual human needs to carry out. We can direct the gaze, for example, by a "look at me" command where "me" is the name of the marker attached to the currently active camera. To carry out these commands, the VHI includes advanced target animation and inverse kinematics functions that take into consideration the current constraints on the virtual human's body.

Vision subsystem

VHI's visual subsystem is responsible for the virtual human's ability to see the user's surroundings as well as its own synthetic 3D environment. The most important functions of the digital human's vision system are detecting, tracking, and recognizing people in front of the screen. We used a Web camera mounted on top of the computer monitor and mapped the incoming video stream onto the virtual monitor object in the virtual human's synthetic environment. The face detection and recognition algorithms analyze each frame and mark the position of each face with a marker. These markers have 3D coordinates in the digital human's world. As a result, the user and our virtual human can visually communicate with one other.

We can program the eye and head motions of the virtual character to look at this marker, which has the effect of the virtual character maintaining eye contact with and following the movement of the userin front of the screen. To control the "look at" process, the VHI system lets us use different tracking strategies for the head and the eyes. We can set the head and eye targets to follow different markers. For example, when multiple people are present, the head could follow the moves of the second person appearing, while the eyes would still fixate on the first person. Alternately, the head could follow the motion of the camera while the eyes respond to the changing location of the user in front of the terminal.



5 **Using active and selective vision to extract high system images of the user sitting in front of the computer.**

Eye contact plays a critical role in the process of face-to-face human communication. In particular, the ability to intelligently speak to or listen to the user while looking at him or her creates—in itself—the illusion of intelligence. Furthermore, the virtual humans' ability to detect the number of people in front of the computer terminal, analyze their distance from the screen, and recognize their identity as well as their facial expressions and emotions forms a firm foundation for controlling procedural reactions in a reactive manner. The refined control and multiple available strategies of VHI's "look at" process lets the virtual human maintain natural eye contact during the entire conversation.

Recognizing user emotions

Figure 5 shows the overall architecture of the facial tracking module used in the expression analysis unit. This module first captures the user's image with a pan-tilt-zoom Web camera in a wide-angle view, letting the user move freely in a large workspace. The module detects the user's presence by finding the number of faces in the input image and tracks the face (or faces) continuously as the user moves throughout the session. The head-tracking routine measures head motion to indicate the user's restlessness. The system also detects the eyes to compute the relative changes in distance, which it can use as an indication of interest.

Once the user is settled in a relatively stable position, the camera gradually zooms in, framing the face optimally to deliver the best resolution for the feature-tracking algorithm. To remove dependencies on personal facial characteristics, VHI first normalizes the feature point's tracked motion and then encodes it in an MPEG-4 data format. If needed, the system can carry out further analysis of specific facial regions—such as the eyes or the mouth—by zooming in to the highest possible resolution. As the user slowly moves within the field of view, the camera actively follows his or her motion, trying to maintain the best possible imagery. If the user's head moves out of the frame, the camera gradually zooms out until it finds the user and repeats the process.

The emotion-recognition subsystem accepts input from the active-vision tracker, which follows the motion of specific facial features on the user's face. The algorithm takes advantage of the fact that any emotion displayed by the user in front of the VHI computer system has a unique temporal pattern, the transition of which is highly constrained by human facial musculature. In fact, psychological research in the early 1950s showed that the neutral face and the six basic human emotions—happiness, surprise, fear, sadness, anger, and disgust—can be charted in a 2D circular arrangement.[11] Transitions from one emotional state to another must take place continuously, respecting the constraints in this circular arrangement. In other words, psychology tells us that the human face cannot arbitrarily transition from happiness to anger, for example, without either passing through a neutral state or through the disgusted state—at least as far as facial muscles are concerned.

In practice, we use this psychological finding in our emotion-recognition subsystem to model and constrain the user's possible emotional states. The second important psychological aspect of the internal emotion representation model is the user's current relationship to the presented material. Specifically, we narrow the six basic expressions to four major categories on the basis of how the user's emotional state interweaves with the cognitive dynamics of the communication process. We borrowed this quadrant model from educational psychology.[12] We define movements that happen in the space in front of the computer as interaction and, consequently, information exchange occurs.

Emotional modulation

Emotion is an essential element in human communication. The presence or absence of emotion can determine the differences between lifeless, robotic, or lifelike behavior. Traditionally, animators have painstakingly created emotional behaviors for prerendered animations. However, when virtual characters interact in real time with real people who exhibit emotions, the system's ability to generate and display a range of emotions autonomously becomes critical.

Emotional processing involves many cortical areas.[10,13] As a result, researchers have developed multiple competing theories about emotion. For the purposes of animated agents, many practical applications prompted researchers not to use explicit emotion labels and categories, but to span an emotion's space along two or more dimensions. In these models, one dimension represents the *valence* or *appraisal* with positive or negative values while a second dimension represents the intensity of a felt emotional reaction. Such is the ocasionally connected computing model[14] that has recently become a frequently used system in the research community. Instead of working with basic emotional states, the OCC model provides a rule-based system for triggering some 22 emotion types.

For the purposes of the VHI system, our main focus was to provide a mechanism to link the user and the virtual human together in a manner that created an emotional bond. This emotional bonding can maximize the effectiveness of information exchange by adapting the virtual character's emotional overtone to suit the user's current mood and momentary emotion. Our system's main advantage is that it does not require us to model the internal states of the virtual character. We can instead rely on what we call mindless strategies that the users will perceive as emphatic, reactive, and intelligent.

In our artificial emotion space (AES) algorithm, we adopt the notion that emotions are generally comprised of three basic layers of behavior: personality, moods, and momentary emotions.[15] Personality is the behavior that we generally display. Motion patterns represent the character's personality in the VHI system. In other words, motion conveys the attitude we record and store. Moods are prolonged emotional states caused by the cumulative effect of the environment, including the momentary emotions we enjoy and live through. In the VHI system, we compute moods by integrating the positive and negative emotions affecting the virtual human character and express them in the same manner as the momentary emotions—with the exception of possessing different amplitudes, durations, and decay characteristics. We define momentary emotions as behaviors displayed briefly in reaction to events. Our system uses the TDC at this top level to define a link between the user's and the virtual character's emotions.

The AES algorithm is the manifestation of a simple everyday observation: When in a good mood, we are generally more susceptible to information presented in a positive fashion, and when we are sad or down, we prefer things presented in a more subdued manner. The AES defines two additional TDC configurations not used for animation, but that represent the emotional states of the virtual human and the user. The algorithm synchronizes the two TDCs to help create our goal of emotional bonding. Clearly, depending on the application domain, we could employ many strategies to link the two actors together. The digital human might exhibit layers of emotion that coincide with or even oppose the user's mood. Such behaviors could create the illusion of empathy or compassion.

With this mechanism, the system might improve the user's mood and relation to the presented material by gradually driving the user toward a happy state from a sad or neutral one. Because the AES algorithm operates directly in the TDC domain, we could record the resulting transitions through the MTM unit that controls the duration, and the decay, of these emotions. The resulting output could become a seamless deformation of facial surfaces (different emotional expressions), bone structures (tone and posture associated with moods), and even textures (skin blushing and tiredness) in real time.

Synthetic vision

The second module of the digital human's vision implementation is responsible for perception in its own virtual 3D space. To implement this functionality, we use visual fields represented as cones and modeled after human ergonomic measurements. Our system considers any markers attached to 3D objects in the synthetic environment that fall within these cones as visible by the digital human and pass them on to higher-level processing routines. The different cones can represent various states of alertness in the VHI system. For example, when the application requires the virtual human to focus intensely on a certain area, it can use a narrow field of view for the visibility test, which effectively renders objects in the wide periphery ineffective. On the other hand, when the system needs to detect events on the periphery, it can use the wide-angle field of view to encompass every marker that might bear significance to execute a specific task.

The system can also control visibility in the distance by simply adjusting cone length. As the virtual human moves in 3D space, objects and markers appear and disappear in its field of view accordingly. In addition, the synthetic vision module also encompasses low-level collision-detection routines to compute whether something is in the way or crosses the path that the virtual human is moving in while executing its task. At any time, we can have the system's higher-level behavior routines poll the list of visible objects for those potentially colliding with or obstructing the locomotion of the virtual human agent. The mechanism provides a simple means to make the virtual human aware of its 3D surroundings.

Auditory and tactile input

The sound system consists of independent speakers placed in 3D space in the virtual environment. As the camera representing the user moves around, these sound channels deliver sound in stereo or quadraphonic channels. Both the user and the virtual human have a directional speaker assigned to them that moves in 3D space like any other object in the scene. When multiple people provide input, such as in a shared collaborative environment, the VHI system maps each individual to his or her own speaker in 3D space, explicitly marking that person's position. The system can access markers attached to the speakersin high-level task descriptions much like any other object in the scene. As a result, when the user utters a command into the microphone, the virtual human perceives the command as coming from the front or behind, or from the left or the right, and can react accordingly. This simple mechanism, which does not require any complicated recognition engine, provides a layer of reactive intelligence that projects a feeling of presence to the users.

In addition to auditory input, touch is the digital agent's ability to detect collisions and accept force input from the user. Instead of predicting if and when the virtual human will collide with a given object, VHI can compute the object intersecting with the virtual human's body and send signals to the higher-level controls when collision events occur. The user can also touch the virtual human by clicking on any polygon on the agent's body (to simulate a tickle) or apply forces to the bones directly. The system stores the polygon and the body part touched and the time the user touched the agent. Users can select bones and apply forces using a keypad or joystick interface. The virtual human's architecture gives us the ability to calculate the local pain within each joint as well as a global pain measure accumulated as the result of internal and external forces. Thus the system can directly link pain, touching, and tickling values to static or procedural animation channels to provide reactions to those conditions. Although we implemented all of these mechanisms using the mouse, keyboard, or joystick, future extensions could provide force feedback and other means to further enhance the interaction process.

Real-time performance

We implemented the VHI as a high-performance, real-time rendering environment capable of running at 30 to 50 frames per second even with high-resolution geometry, textures, video, and multiple AVI files running in the background. The VHI interactive environment supports external devices to enhance the perception of reality. Such devices include a head-mounted virtual reality display, external video sources, motion capture capabilities, built-in face recognition and tracking capabilities, particle systems, chroma-key calculations, and special effects. In addition to the modeling and animation features, the VHI system supports multiple virtual cameras, cuts and moves, fast previsualization rendering to AVI video files, and compatibility with external rendering systems to achieve high fidelity. The modeling and animation systems also run separately as plug-ins for commercially available animation packages. The rendering engine supports OpenGL and DirectX, and it can run in an Internet browser as an ActiveX component.

To create this high-throughput virtual environment we devised a predictive rendering scheme to measure the most important factors in a graphics system: the speed of the graphics card, CPU, and memory. In particular, maintaining high frame rates without compromising the functionality or quality of our models became a difficult optimization problem that posed several challenges. For one, the high-resolution facial and body geometry used in the real-time animation system can exceed 65,536 polygons, which reaches the limit of 16-bit index buffers used in most graphics cards today. To overcome this limitation, our rendering process uses double buffering and slices the surface geometry in render-time to fit the data into the size supported by the applicable hardware.

Synchronizing the graphics card, CPU, and memory is also a key source of significant delay not addressed or optimized in most existing systems. The classical rendering method first requires the CPU to collect and compute all required data and then pass it on to the graphics card, which, upon receiving this data, spends additional time with internal processing before displaying it. In addition, because we use high-resolution textures and video files, the time it takes to load these data files from memory also presents a potential source of significant delay. We can use these latencies as a parameter to feed a predictive-rendering scheme that overlaps critical time periods of the CPU and the graphics hardware to produce a continuous feed. At the optimal point of operation, the speed of the CPU matches the data rate imposed by the graphics card and the memory for any given size of geometry, textures, and media.

Dynamic textures are another important feature of our system. Specifically, the VHI might have to compute a new texture map for each animation sequence frame to create subtle details and higher fidelity. Internally, we employ $1{,}024 \times 1{,}024$-pixel maps, several of which need to be blended together at higher frame rates. Currently we accomplish this task with the CPU. In the future, we might shift this burden onto the graphics card, using pixel-shader technology.

The VHI can map a new image frame captured directly from the camera or a media file 30 times a second. Decoding the AVI file and passing the information from a Web camera to the appropriate VHI module occupies the majority of the CPU's load. To evaluate the efficiency and usability of our methodology, we tested the virtual human animation system on multiple computer platforms, each exhibiting different configuration and performance characteristics. For testing purposes, we primarily focused on CPU and graphics card performance and measured the overall frame rate as a function of polygon size to establish optimal performance.

Table 1. Real-time performance comparison.

	Computer 1		Computer 2	
Polygons	DirectX (fps)	OpenGL (fps)	DirectX (fps)	OpenGL (fps)
127,709	6.15	7.92	31.89	29.25
97,305	7.73	8.65	35.72	31.33
62,859	12.86	12.07	42.84	40.27
33,017	20.44	14.95	52.07	44.75

6 External virtual-reality devices that connect to the VHI system: (a) From left to right, magnetic motion tracker, joystick, head-mounted display, , and video camera used for live input; (b) Complete portable system, and (c) user interacting with and experiencing a virtual environment.

In the performance tests, presented in Table 1, we compared two computers running DirectX and OpenGL. Computer 1 was an average computer configuration with a 3-GHz CPU speed, 1 Gbyte of memory, and a Radeon-9700 graphics card. Computer 2 was a top-of-the-line machine with an 850-MHz CPU, 256 Mbytes of memory, and a GEForce2-Go graphics card. Because of the improvements in computer manufacturing technology, the same performance will cost significantly less in the near future and will therefore be widely available on home computers.

Enhanced and augmented reality

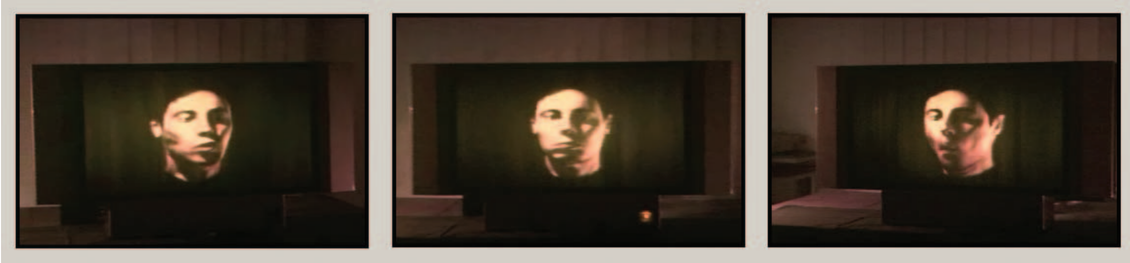
VHI supports several external devices and built-in technical systems that can immerse the user in the virtual environments. For example, in the virtual-reality module, the user wears a head-mounted display to experience visual and auditory signals delivered directly to his or her eyes and ears in 3D. A tracker attached to the head-mounted display measures the yaw, pitch, and roll of the head and updates the computer-generated imagery. The user controls the other movements in 3D space via a joystick or other external motion sensors. The system can represent sound sources and can emulate the cumulative effects of speed and motion. Figure 6 shows the key hardware elements of the VHI system.

An alternate way to increase the level of immersion is with image-based techniques to create a virtual backdrop that uses 360-degree panoramic images. (See http://www.digitalElite.net for examples.) This panoramic background might be a static image or a video stream mapped onto a sphere that surrounds the user in virtual space. As the user turns his or her head, a portion of the sphere would become visible to create the illusion of actually being at the location of the original photo. The VHI system would then mix synthetic elements and the virtual human character to create interactive functionality. The live video module offers yet another set of functionality to implement novel applications. We specifically designed a chroma-key module to let other people and real-world objects enter the user's virtual space. We implemented augmented reality support to enhance the real-time experience and allow users to combine their own true-life environment with the digitally created human and other synthetic elements.

Future work

Perhaps one of the most interesting and intriguing applications of the VHI system is building an intelligent holographic virtual human that could sense and interact with real people naturally. Although this might sound like pure fantasy, recently we managed to interface our VHI system with a novel 3D holographic display system that does not require any special glasses and is viewable simultaneously by multiple people. Figure 7 shows the television-like device from three different viewing angles while the virtual human is making faces to entertain the people around it. Such a device, when combined with the interactivity and communicative intelligence offered by the VHI system will find many application areas in human–computer interaction.

Our approach differs from classical artificial intelligence systems because our goal is to make the user believe that the virtual character is paying personalized attention during communication, conversing intelligently with the user about any particular topic. To gen-



7 Holographic virtual human.

erate this kind of low-level reactive behavior, we use strategies that ensure that the visual communication link remains unbroken. We hope the VHI system will serve as the technological foundation for many practical, real-life applications. For example, a virtual teacher implemented in this framework could deliver personalized educational material at the pace required by the student, while at the same time gathering information about the student's level of interest. Our system could also be used to create intelligent information kiosks located in stores, around airports, and in other public places where people could access the information they need quickly and easily.

Currently, we are working on medical applications of the VHI technology. Virtual reality therapy is an emerging field that has already been successfully applied in treating pain, rehabilitating stroke patients, curing fear of flying or driving, and many other areas.[7] Existing systems are not only very expensive, but often lack the visual quality required to trick the brain into believing that a given virtual situation is real. VHI's visual capabilities and functionality offers opportunities for therapeutic applications for patients dealing with issues related to interacting with other people. This kind of cyberpsychology could help treat social phobias, such as fear of public speaking, and help provide new tools for anger management and substance abuse patients. The VHI system could potentially help diagnose many brain disorders, such as Alzheimer's or Parkinson's disease. ■

References

1. N. Badler, M.S. Palmer, and R. Bindiganavale, "Animation Control for Real-Time Virtual Humans," *Comm. ACM*, vol. 42, no. 8, 1999, pp. 64-73.
2. D. Terzopoulos and K. Waters, "Techniques for Realistic Facial Modeling and Animation," *Computer Animation*, M. Thalmann and D. Thalmann, eds., Springer-Verlag, 1991, pp. 59-74.
3. S. Pasquariello and C. Pelachaud, "Greta: A Simple Facial Animation Engine," *Proc. 6th Online World Conf. Soft Computing in Industrial Applications*, Springer-Verlag, 2001.
4. S. Morishima, "Face Analysis and Synthesis," *IEEE Signal Processing*, vol. 18, no. 3, 2001, pp.26-34.
5. J. Cassell, H. Vilhjlmsson, and T. Bickmore, "BEAT: The Behavior Expression Animation Toolkit," *Proc. Siggraph*, ACM Press, 2001, pp. 477-486.
6. I. Poggi, C. Pelachaud, F. de Rosis, "Eye Communication in a Conversational 3D Synthetic Agent," *J. Artificial Intelligence*, vol. 13, no. 3, 2000, pp. 169-181.
7. A.A. Rizzo, "Virtual Reality and Disability: Emergence and Challenge," *Disability and Rehabilitation*, vol. 24, no. 11, 2002, pp. 567-569.
8. J.S. Rickel, et al., "Toward a New Generation of Virtual Humans for Interactive Experiences," *IEEE Intelligent Systems*, vol. 17, no. 4, 2002, pp. 32-38.
9. W.L. Johnson, W. Rickel, and J.C. Lester, "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments," *Int'l J. Artificial Intelligence in Education*, vol. 11, 2000, pp. 47-78.
10. D. Goleman, *Emotional Intelligence: Why It Can Matter More Than IQ*, Bantam Books, 1995.
11. H. Schlossberg, "The Description of Facial Expressions in Terms of Two Dimensions," *J. Experimental Psychology*, vol. 44, no. 4, 1952, pp. 229-237.
12. B. Kort and R. Reilly, "Analytical Models of Emotions, Learning and Relationships: Towards and Affect-Sensitive Cognitive Machine," *Proc. Intelligent Tutoring Systems Conf.*, 2002, pp. 955-962.
13. R. Adolphs, "Neural Mechanisms for Recognizing Emotion," *Current Opinion in Neurobiology*, vol. 12, 2002, pp. 169-178.
14. A. Ortony, G. Clore, and A. Collins, *The Cognitive Structure of Emotions*, Cambridge Univ. Press, 1988.
15. G. Ball and J. Breese, "Emotion and Personality in a Conversational Agent," *Embodied Conversational Agents*, MIT Press, 2000, pp. 189-219.

Barnabás Takács *is the president of Digital Elite. His research interests include photorealistic human animation, face recognition, motion tracking, and artificial-intelligence-based agents. Takács received a PhD in Computational Sciences from George Mason University.*

Bernadette Kiss *is a lead designer and developer for VerAnim's photorealistic interactive virtual humans. Her research interests include human modeling and animation, and psychology and its applications to nonverbal communication in computer interfaces. Kiss received a BS in business management from the College of Commerce, Catering, and Tourism/Budapest Business School, Hungary.*

Readers may contact Barnabás Takács at Digital Elite (http://www.digitalElite.net); btakacs@digitalelite.net.