Special Education and Rehabilitation: Teaching and Healing with Interactive Graphics

Barnabás Takács
Digital Elite

A special education and rehabilitation system employs real-time interactive computer graphics and photorealistic virtual humans to implement an emotional modulation technique that increases learning efficiency.

High-fidelity animated virtual characters that interact with students and children with special needs represent an increasingly important segment of computer graphics in domain-specific education. These pedagogical agents[1] offer important new capabilities that help improve learning efficiency by providing interactive task demonstrations, answering questions in a dialogue-based format, providing navigational guidance in game-like environments, and above all, teaching while entertaining.

In particular, emotions play a vital role in a student's ability to memorize and learn new material. Emotions act as a catalyst in the process of transforming information into knowledge and thus the effectiveness of computer-based learning can be greatly improved if we incorporate emotions into computer use as a learning tool.[2] Tutors, teachers, and professors achieve this effect—called *emotional modulation*—on a daily basis, using their charisma and other personal and motivational qualities during the time they invest in their students. Therefore, using virtual humans to mimic the natural face-to-face dialogue that normally takes place between student and tutor in real life forms the foundation of a unique and critically important enabling technology for teaching in the future.

Virtual humans, much like their real-life counterparts, use gaze and gestures as guides to focus the learner's attention, incorporating dynamic models of appropriate emotional responses while continuously providing verbal and nonverbal feedback. Furthermore, they might also monitor the student's reactions to create a closed-loop model of dialogue-based interactions. Advances in virtual human interface and animated conversational agent technology and the availability of low-cost graphics hardware that can perform cinematographic-quality real-time rendering offers a unique opportunity to adopt graphics to enhance the special education and cognitive, mental as well as physical rehabilitation toolset (see http:// www.digitalElite.net).[2]

Photoreal digital humans—as opposed to stylized humanoids or cartoon characters—provide a novel means to unlock and further use the learning capability residing in each and every one of us. The final outcome results in a channel to engrave knowledge and to even modifying behavior, both of which are critical in the applications addressed in this article.

The "Related Work" sidebar discusses other approaches to this field.

Factors of believability

The mechanisms that take place during interaction with an animated cartoon or nonhumanoid characters versus high-fidelity virtual humans are radically different. Therefore, from a practical and scientific point of view it's important to underline how the photoreal agents compare to the learning delivery through more traditional cartoon or stylized humanoids. Cartoon characters exhibit stylized and exaggerated features that help them indicate and signal emotions in a caricatured and deliberately unrealistic way. We do not expect them to behave as real human faces and thus the rules of fidelity do not apply. Therefore, as long as the viewer feels the humanness, the character will work. The lack of realism actually helps the user to identify with the characters through self-projection and empathy.

Conversely, photorealistic virtual humans take on the challenge of addressing the highly refined facial infor-

 Published by the IEEE Computer Society

Related Work

The creation of animated conversational agents and virtual humans for educational applications has been a key interest of computer graphics research during the last few decades. Pedagogical agents aim to mimic the quality of everyday human communication by combining the benefits of visual impact with conversational intelligence to teach students in a personalized manner.

Early work in the field mainly addressed the modeling and control aspects of low-polygon virtual humans[1] that can be delivered in real time, while later results focused autonomous agents that exhibit rich personalities.[2] To achieve this goal, various layers of metacommunication and nonverbal signals needed to be modeled and delivered in an automatic fashion in conjunction and synchronized with what the virtual character spoke. To generate these signals and provide an illusion of a lifelike character, detailed emotional and personality models needed to be developed that could control the animation channels as a function of personality, mood, and momentary emotions. Finally, the real-time delivery and interaction with these virtual characters posed an extra set of technical challenges in terms of the speed, computational power, and most importantly, the visual quality required to make users believe that they are interacting with a living creature. To achieve this goal, precrafted animation graphs and animated actions were replaced with intelligent behavior modules that attempted to control speech, locomotion, gaze, blinks, head and body gestures (including various postures), and object manipulations to interact with the environment.[3]

Work on the aspects of social dialogues with embodied conversational agents has recently become a distinct field of development. A good summary can be found in Cassell et al.[4] where the authors studied a virtual real-estate agent who uses vision-based gesture recognition, speech recognition, discourse planning, sentence and gesture planning, speech synthesis, and animation of a 3D body. Other major pedagogical systems developed to date all employ low-resolution human characters.[5,6]

In terms of how these virtual pedagogical representations help the learning process, studies on human interfaces with static or animated faces found that users rated them as more engaging and entertaining than functionally equivalent interfaces without a face.[7,8] Furthermore, a human face clearly commands more attention from the user than any other character. In particular, users were found more likely to be cooperative with an interface agent when it had a human face.[10] Other similar studies can be found elsewhere.[10,11]

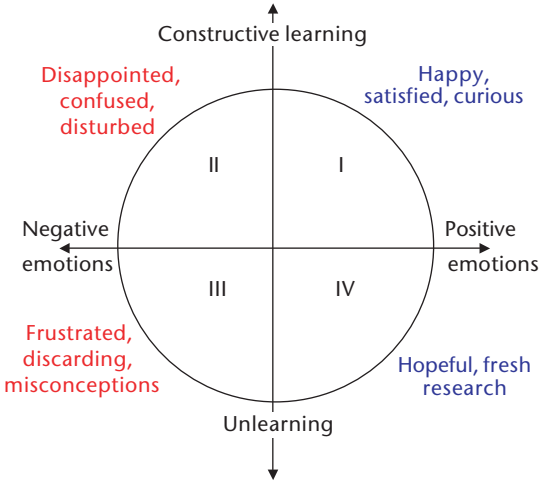
Researchers in the realm of education and rehabilitation have successfully used the approach to interact with autistic children,[12] virtual environments and multimedia systems have helped children with physical handicaps,[13] and a pedagogical agent has helped mothers of pediatric cancer patients in developing better problem solving skills.[14]

References

1. N. Badler, M.S. Palmer, and R. Bindiganavale, "Animation Control for Real-Time Virtual Humans," *Comm. ACM*, vol. 42, no. 8, 1999.
2. J. Rickel et al., "Toward a New Generation of Virtual Humans for Interactive Experiences," *IEEE Intelligent Systems*, vol. 17, no. 4, 2002, pp. 32-38.
3. B. Takács and B. Kiss, "Virtual Human Interface: A Photorealistic Digital Human," *IEEE Computer Graphics and Applications* vol. 23, no. 5, 2003, pp. 38-45.
4. J. Cassell et al., "More than just a Pretty Face: Affordances of Embodiment," *Proc. Intelligent User Interfaces* (IUI), 2000; www.isi.edu/~hannes/publications/IUI2000.pdf.
5. T. Bickmore, L. Caruso, and K. Clough-Gorr, "Acceptance and Usability of a Relational Agent Interface by Urban Older Adults," *Proc. Conf. Computer–Human Interaction* (CHI), ACM Press, 2005.
6. J. Lester, B. Stone, and G. Stelling, "Lifelike Pedagogical Agents for Mixed-Initiative Problem Solving in Constuctivist Learning Environments," *User Modeling and User-Adapted Interaction*, vol. 9, nos. 1-2, 1999, pp. 1-44.
7. T. Koda and P. Maes, "Agents with Faces: The Effect of Personification," *Proc. 5th IEEE Int'l Workshop Robot and Human Comm.* (ROMAN), 1996; ai.rightnow.com/colloquium/papers/Agents-with-faces.pdf.
8. A. Takeuchi and T. Naito, "Situated Facial Displays: Towards Social Interaction," *Proc. Conf. Computer–Human Interaction* (CHI), ACM Press, 1995.
9. S. Kiesler and L. Sproull, "'Social' Human–Computer Interaction," *Human Values and the Design of Computer Technology*, B. Friedman, ed., CSLI Publications, 1997, pp. 191-199.
10. J. Cassell, "Embodied Conversation: Integrating Face and Gesture into Automatic Spoken Dialogue Systems," *Spoken Dialogue Systems*, S. Luperfoy, ed., MIT Press, 1999.
11. D.M. Dehn and S.V. Mulken, "The Impact of Animated Interface Agents: A Review of Empirical Research," *Int'l J. Human–Computer Studies*, vol. 52, 2000, pp. 1-22.
12. J.H.G. Williams et al., "Visual-Auditory Integration During Speech Imitation in Autism," *Research in Developmental Disabilities*, vol. 25, pp. 559-575.
13. C. Sikne Lányi et al., "Developing Multimedia Software and Virtual Reality Worlds and their Use in Rehabilitation and Psychology," *Proc. 2nd Int'l Conf. E-Health in Common Europe*, IOS Press, 2004.
14. W.L. Johnson, J.W. Rickel, and J.C. Lester, "Animated Pedagogical Agents: Face-to-Face Interaction in Interactive Learning Environments," *Intl. J. Artificial Intelligence in Education*, vol. 11, 2000, pp. 47-78.

mation processing subsystem in our brain. Therefore, the closer they resemble their real-life counterparts, the greater the challenge of making them real. However, once that level of quality is attained they make a much more effective communication interface between computers and humans than stylized characters. By conveying subtle facial signals (see Figure 1, next page) of positive reinforcement, they access and activate both our *declarative* (that is, our memory of information and events) and *procedural memory* (that is, learned skills such as how to ride a bike), while engaging the student in the process of emotional responses in an entertain-

1 Subtle changes of emotional display of a photoreal virtual teacher during interaction with a student.



2 Relationship of emotions to the learning process (based on Kort, Reilly, and Picard[3]).

ing fashion. Most computer interfaces primarily access users' declarative memory; the person learns information about a particular topic.

To date, there is no virtual human-based conversational agent system operating at a comparable level of visual quality and/or nonverbal capabilities that allows a direct validation of the benefits of using photoreal agents instead of stylized characters. Evaluation studies typically employ stylized humanoid representations, and in their conclusions frequently overlook the fact that the visual representations used for their characters might hinder the users' ability to immerse themselves. On the other hand, the virtual human interface system described here is equipped with cartoon-style as well as photoreal human representations. As such, it allows comparative research and evaluation of motivational methodologies. The effectiveness of this technique in currently undergoing large-scale evaluation studies in real-life applications.

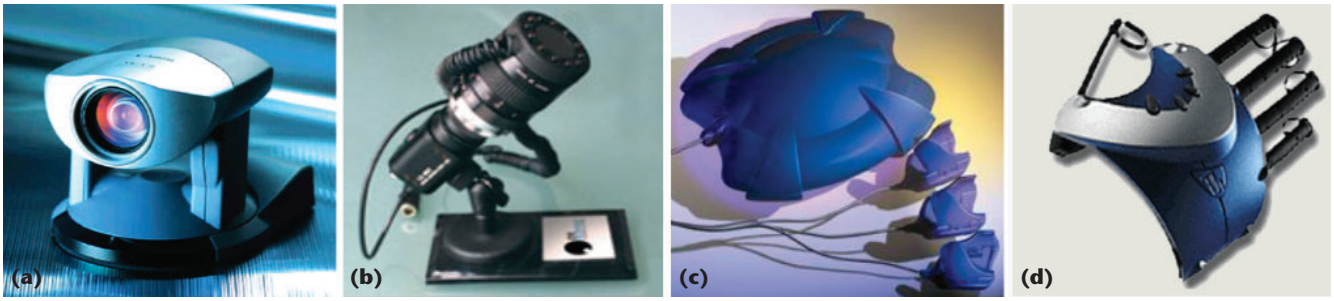
The power and sustainability of an animated conversational agent in general and virtual human in particular, lies in its ability to create the impression that it is alive. More precisely, what matters the most is how believable the character is in a specific context and how it helps suspend user disbelief. Factors of believability span the entire spectrum of content, gestures, low-level behavior, and visual realism. Designers shouldn't address believability as a single issue but need to address how it comprises many layers, infusing practically all aspects of character design and delivery.

Visual fidelity is only one of the many factors we might think of when addressing believability. From our everyday experiences it becomes clear that low-resolution digital humans or stylized, cartoon characters work perfectly well in certain applications. For some, this suggests that there is no need for increased visual accuracy. Unfortunately, that is not the case. To understand this seemingly contradicting statement we need to examine why stylization in virtual characters works. Stylization is a means to emphasize and exaggerate artistically important features at the expense of other, less expensive details. Stylization lets the system focus the user's attention and significantly reduce character complexity. As a result, the user visually processes a stylized character differently than a living face.

While some of the same basic underlying processing mechanisms are in place, when it comes to judging the believability of the digital human, stylization strongly relies on indicating events as opposed to truly expressing them. These indications of gestures and expressions let the student imagine the rest of or simply fill in the visual details otherwise missing. In other words, an underlying current of powerful self-projection mechanisms helps create an enjoyable user experience, but this approach is entirely different from the mechanisms that take place in real-life tutoring.

Learning without failure

The classical learning paradigm is based on what is often called *learning by failure*. Students' emotional states interweave with the cognitive dynamics of the learning process as shown in Figure 2. The emotional state is represented on the horizontal axis, while the vertical axis depicts the students' relationship to learning in a given moment. The students' state of mind moves continuously in this space in a spiraling motion progressing in time. Frequently students begin in quadrant I,



3 External devices used to measure a student's internal state: (a) Web camera, (b) eye tracker, (c) biofeedback, and (d) data glove).

being curious and fascinated regarding a new topic of interest, or they might be puzzled and motivated to reduce confusion and start in quadrant II. In both cases the top half of the space focuses on constructing or testing knowledge. As learning proceeds, movements in this space occur. As an example, students might get an idea about how to implement a solution or build something. However, when the result fails, the need to revise the original idea and analyze mistakes commonly moves them to quadrant III where emotions might be negative and the initial focus on learning changes to eliminating some misconceptions. Hopefully, in time the students consolidate their knowledge and gradually move to quadrant IV from where a fresh idea propels them back to the upper quadrant I.

For applications in special education as well as in rehabilitation, we need to partly revise this model as the experience of failure can often lead to such a strong negative effect that it prevents students from further participation and efficient learning. Therefore, we need a more gentle method based on positive reinforcement. My solution to this problem is to provide visual feedback in the form of encouraging facial displays, expressions, and other nonverbal signs that help students avoid the mistakes they are about to make.

The implementation of this strategy focuses on providing a mechanism and adequate strategies that link the student and virtual teacher together to create a modulated emotional bond. The purpose of this emotional bonding is to maximize the effectiveness of how information is exchanged and absorbed. This step is achieved online by adapting the emotional overtone of the virtual character to best suit the user's current mood and momentary emotions as measured by a set of sensors. This artificial emotion space algorithm[2] formalizes the notion of a simple observation, namely when in a good mood, students are generally more susceptible to information presented in a positive fashion, while on the contrary, when they are sad or down, things presented in more subdued manner are preferential. This simple notion of empathy supposes

- the recognition of the students' internal state and
- a mechanism to respond appropriately in the animation domain.

The final outcome of this process is a reactive model of communication implemented in the form of a closed-loop dialogue.

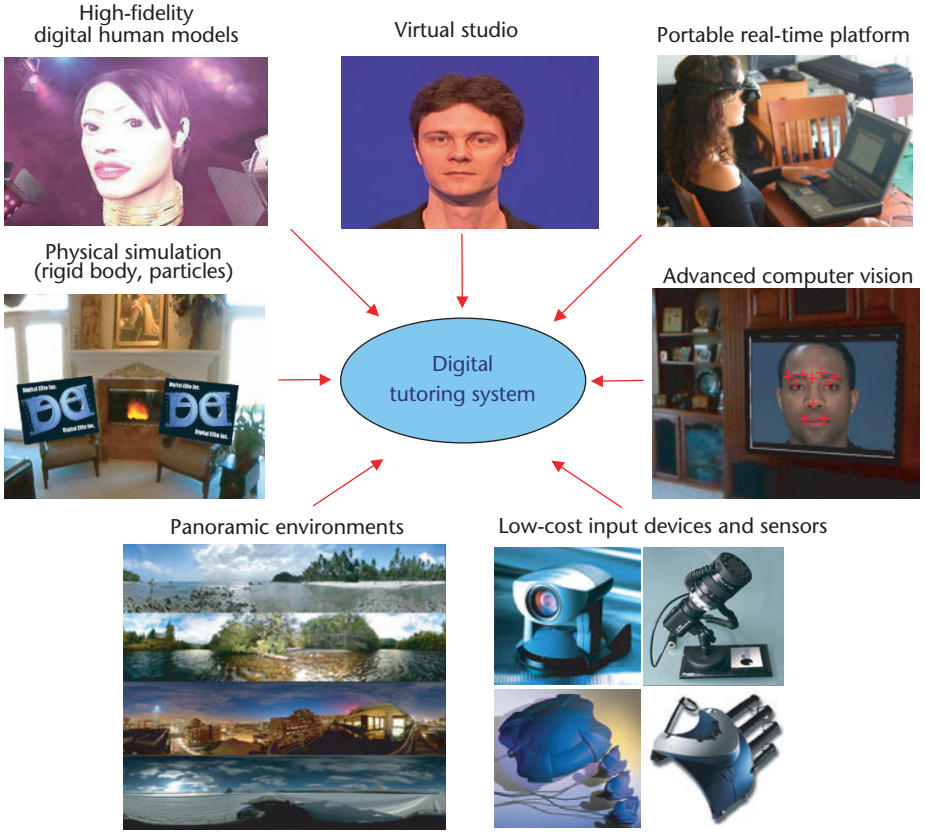
Closed-loop dialogue

The closed-loop interaction model of student–computer interaction combines the power of high-fidelity facial animation with advanced computer vision and other perception techniques to create a bidirectional user interface. In this system, the actions and reactions of a student in front of the computer monitor directly affect the behavior of the interactive content delivered. In particular, the interaction process between the student and system is defined as a closed-loop dialogue that draws on the unique characteristics of human face-to-face communication and more specifically gaze behavior following well-researched interaction patterns and rules.

The importance of gaze and facial display is paramount for social interaction and learning. Conversations are guided by a mechanism known as *turn-taking*, where the role of the listener and speaker shift periodically as indicated by the eyes. Similarly, besides gaze behavior measured with the help of an eye tracker and/or Web camera levels of interest, attention, stress, and fatigue can also be derived through other techniques such as biofeedback. Figure 3 shows some of the devices used in the system.

The mechanism linking external sensors to the virtual human—that is, how the system converts feedback from the external signals to useful and understandable input to the agent—is based on virtual markers. Markers are invisible, dimensionless representations of 3D positions attached to any object in the scene. They possess attributes (such as analog values of device measurements) and carry unique hierarchical labels, which the system can refer to by their names while describing a high-level task. As an example, you could attach several markers to a Web camera image and the monitor representation of that image would appear in the agent's 3D world. This representation is analogous to the computer screen we ourselves use while interacting with the system. The live video stream is mapped directly onto the virtual monitor and the results of the visual processing. For example, the system assigns and displays temporary markers to locations and gestures of faces in front of the terminal.

The advantage of a marker-based representation is that all tasks the virtual human needs to carry out are defined and executed in a high-level, symbolic fashion. The gaze, for example, can be directed by the command "look at 'me'" where me is the name of the marker attached to the currently active camera. Similarly, pointing and locomotion commands are also defined as "go



4 Key components of the virtual human-based digital tutoring system for applications in special education and rehabilitation.

to 'person1'" or "point at 'table corner,'" where person1 is the marker attached to the user's head and table corner refers to the part of an object.

To carry out these commands, the virtual human interfaces include advanced target animation and inverse kinematics functions that consider the current constraints on the virtual human's body. Input from all external devices is represented in this marker form. This configuration allows the tutoring system to carry information from the physical domain into an abstract symbolic representation suitable for the control of low-level reactive scripts or the high-level cognitive engine. Eventually, in the course of interaction these measurements together encode the internal state of the student. To minimize the potential for distraction and to not hinder the learning experience, the practical implementation of the digital tutoring system employs a highly optimized rendering and device architecture, eliminating latencies and providing real-time user feedback. Furthermore, when possible, nonintrusive devices, such as those that are vision-based (face recognition, eye tracking) or require minimal physical contact (biofeedback with two finger clips), are preferred to reduce the effects and inconveniences of attaching cumbersome devices to the students.

Using these measurements of internal emotional and mental states, the animated computer tutor maximizes knowledge gain by minimizing the interaction cost, thus reducing the overall demand on the student's attention. However, instead of treating these variables as simple inputs driving the animated content, the dialogue model assumes that while receiving instruction, the student plays the role of listener, and when giving active input, he or she plays the role of talker. For example, paying attention means that the guiding gestures (such as pointing to an object) that the animated virtual teacher presents are followed by an appropriate shift of gaze and attention to that screen region. Thus, by using these techniques, the tutoring system continuously gauges and monitors the student's performance while expecting certain reactions to compare the measured data with those anticipated responses.

Implementation

Based on the educational principles and emotional modulation techniques set forth in the preceding sections, a digital tutoring system (DTS), as Figure 4 shows., was implemented. The system comprises several key modules, each of which plays a vital part in the interactive learning process. These elements are

- high-fidelity digital humans,
- virtual studio technology,
- photorealistic virtual environments,
- input devices and sensors,
- computer vision, and
- a portable real-time platform.

The first key component of this solution is a virtual tutor featuring a specialized computer graphics and animation architecture to implement high-fidelity digital humans (see http://www.digitalElite.net). The virtual humans have the ability to deliver nonverbal communication signals through subtle facial gestures that support the presented verbal content. To achieve this goal high-resolution digital copies of living people were used; Figure 5 demonstrates this process. These virtual replicas can talk, act, and deliver more than 60 different facial expressions, seamlessly integrated into the interactive process.

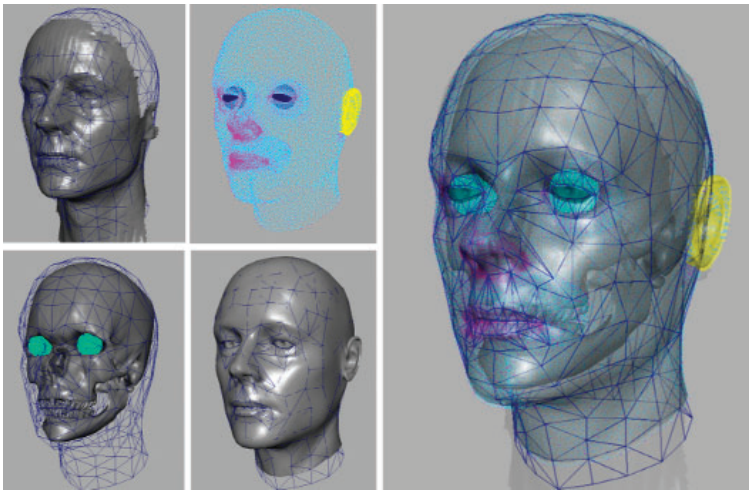
Body gestures, when required, add a further layer of complexity. Specifically, the overall direction of the virtual character's body, hand gestures and beats, transient motions, and other pointing gestures help indicate action or direct the student's attention to a particularly important piece of information. The 60 facial expressions selected for implementation are based on earlier studies of how the human face deforms and the constraints its motion involves. In particular, using the tools described here, a large 3D facial database of individuals

was created and analyzed for the most significant deformations caused by bone motion, muscle actions, and air-related deformations. As a result, the selected 60 expressions span the complete set of display possibilities a face can express.[2] Of these 60 expressions, seven hold particular importance as they refer to the basic emotions (happy, surprised, angry, disgusted, fearful, sad, and neutral) the virtual human system uses for emotional modulation.

As previously mentioned, digital human drives the student's attention and expects a response. When those expectations fail to be realized, it might indicate that the user has lost interest or could not follow the instructions that the digital human presented. In such situations, the system backtracks and adjusts its presentation strategy accordingly.

The next necessary component to implement the DTS—virtual studio technology—uses a variety of visual and auditory stimuli to cause emotions. In many practical applications for special education or rehabilitation the student is placed into a virtual space where the system constantly measures the effects of the stimuli. To adjust the pace and nature of the educational content, supervision of such a situation from a therapist, doctor, or parent is often recommended. By seating this supervising person in front of a blue or sometimes green screen, the system uses automated image-processing techniques to remove the background and composite his or her image directly inside the animated virtual world. This feature becomes particularly important when the child—for increased immersion—wears a VR helmet or sits in a physical setup where reassuring eye contact with the supervising adult is otherwise hard to maintain.

The next required element involves photorealistic virtual environments. Depending on the educational or rehabilitative application's needs, the virtual environments implemented in the DTS go beyond the state-of-the-art by implementing an animation and rendering pipeline that supports high-resolution 3D models with different material properties as well as photorealistic content in the form of image-based rendering. Specifically, the scene is divided into two distinctive regions such as the background that contains elements in weak perspective, and the foreground that holds visual objects that are close to each other and therefore need to be represented under strong perspective transformations. DTS implements the background as a spherical, 360-degree, image-based panoramic dome. The virtual camera representing the child is placed into the center of an imaginary sphere with the surroundings mapped appropriately. Foreground elements, on the other hand, are synthetic 3D objects that appear under strong perspective conditions, where an object appears very close to the viewer. Typically, the student would have these in near proximity as they represent digital props to carry out the exercises. Implemented as a collection of 3D models and particle systems, these digital objects possess physical properties, such as color, roughness, material type, weight, and resistance. Simulated forces influence them as part of the interactions. The overall result is a unique blend of foreground and the background, which creates a truly immersive environment capable of inducing emotions and reactions in the child.

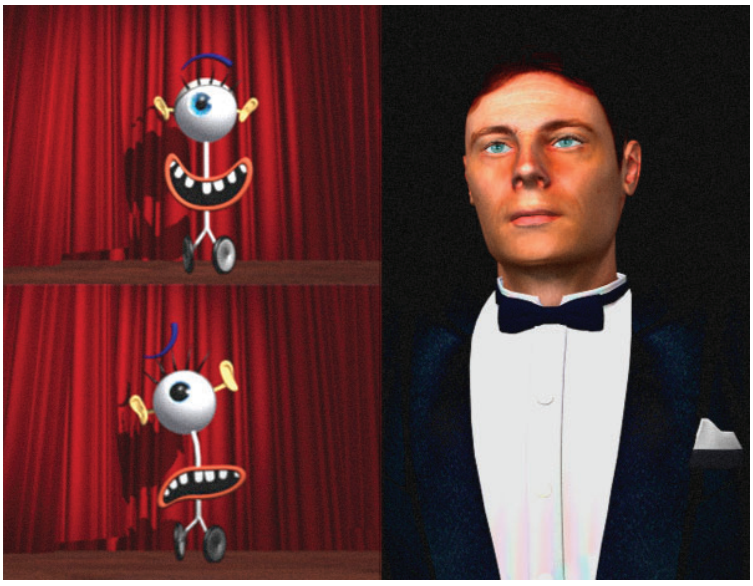


5 Digital cloning: building the virtual replica of a real person from scanned data.

Next, to implement the closed-loop dialogue model, the DTS continuously monitors the student's internal state through input devices and sensors (see Figure 3). Web cameras supply video input to analyze facial actions to measure responses in response to animated content. The eye tracker delivers a more accurate input as to where the student is looking. The system than directly translates raw screen coordinates of the fixation point to 3D scene element components. For example, the virtual teacher might know which synthetic object or even which polygon the student is observing, or if the student is looking at the virtual face, whether he or she is gazing at the eyes, nose, mouth, lips, and so on. In addition the biofeedback interface provides information regarding the student's stress levels. The system measures and monitors changes in heart rate and skin conductance and can infer whether the child is in a state of anxiety or relaxation. These regulatory cues then drive the animation system and keep the level of an exercise just at the limit the child is capable of handling. Finally, the low-cost data glove is used to track the translation and rotation of the hand (6 DOF) allowing the user to touch the virtual objects he or she is surrounded with while the finger bending sensors on the glove also serve to control animation.

The DTS's built-in computer vision module is, for most applications, the primary means of extracting information from the student. It implements a multitude of image processing and pattern recognition functions tailored specifically to analyze faces and facial expressions. Its functions range from locating the presence of a face in front of the computer (a feature used to guide the gaze of the virtual human to maintain eye contact and to look back at the child) to detecting and tracking the facial feature motion as indicators of emotions and facial gestures.

Lastly, the DTS's portable real-time platform is implemented to take advantage of the latest advances in low-cost consumer-level computer graphics. Despite the high visual fidelity virtual environments, media content, and animated virtual humans, it runs on a portable system



6 Examples of virtual teacher characters (cartoon and photorealistic).

complete with interfaces to the external sensors and VR devices described previously.

Applications

Digital interactive tutors in general and high-fidelity virtual animated faces in particular offer a novel and emerging application area for computer graphics in education. These systems are the result of many years of interdisciplinary research and provide the technological foundation for a variety of practical, real-life applications being developed by nonexpert users. While the possibilities are seemingly endless, here three case studies are presented that cover working with autistic children, cybertherapy, and cognitive rehabilitation. The virtual human system described here features cartoon characters as well as full-bodied photoreal digital human tutors as exemplified in Figure 6. The practical applications described next, employ head-only pedagogical agents and thus let the system maximize the face-to-face contact while reducing user distractions, letting them focus their attention to the task at hand.

Virtual friend for autistic children

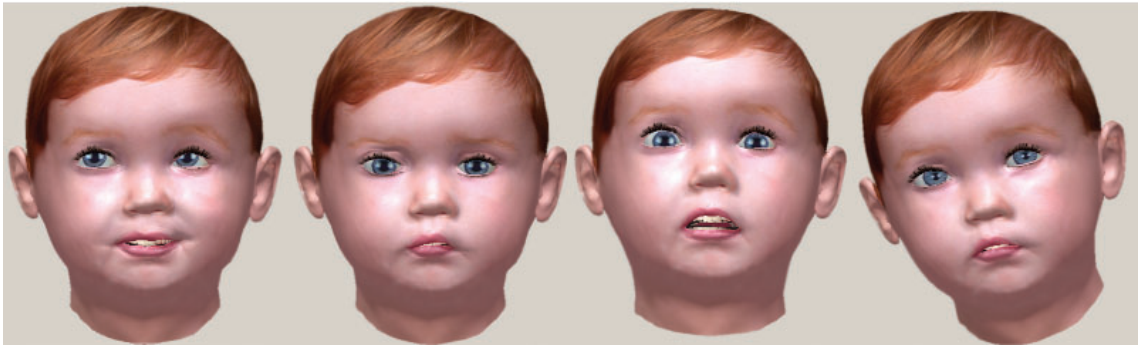
Autism is a lifelong severe disorder that robs most of the affected individuals of independence, satisfying work, and friendships. It severely stunts their social, emotional, and cognitive development. While the exact cause of autism is largely unknown, the popular theory of social gaze disruption traces many features of the disorder to a child's inability to process facial information normally. For example, children with autism have difficulty perceiving other's emotions and processing nonverbal communication. They frequently appear unaware of the information that can be gained by attending to other's facial expressions and often exhibit impaired ability to shift attention between objects and faces. These disruptions eventually result in difficulties socially interacting with others.

Facial processing, and especially eye contact, plays a critical role in the brain's development during early life. The human face exerts a strong attraction on infants' attention within minutes of birth and again compels their intense interest during later stages of early development. Therefore, photorealistic virtual faces, when used in conjunction with eye tracking and other means of measuring a child's responses and reactions, can provide healthy facial stimuli in a controlled manner. Such a system can help diagnose possible disorders at a young age—long before currently employed screening methods can—as well as help rehabilitate these children to eventually lead a better, happier life.

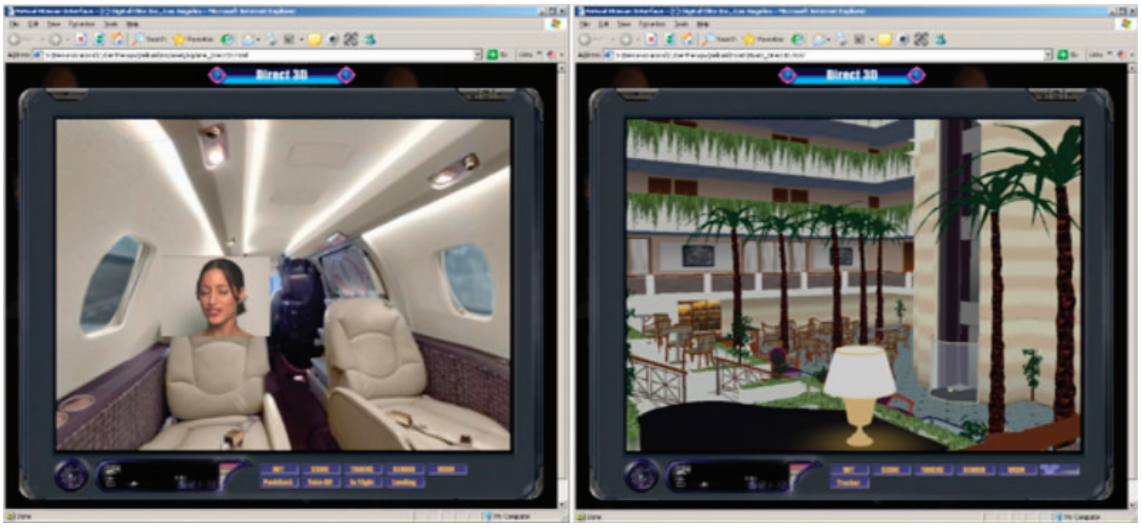
Figure 7 shows an example of such a system. The child takes a gentle, slow movement ride while "Buddy," a virtual friend (shown on the right) is displayed on the built-in monitor screen.[4] A hidden eye tracker and video camera measures and trains a child's social attention to help with his healthy development. This learning environment reinforces gaze behavior, and, in particular, it encourages the child to look at the animated face's informative area where communicative, emotional, gender, and age information can be obtained.

Figure 8 demonstrates some of the many faces of the system's virtual child. First-order expressions include facial shape changes caused by the underlying musculature and controlled via temporal disc controller interfaces.[2] Secondary effects include pupil dilation, eyes looking at targets (the user, camera, or other objects in the scene), and cornea dynamic skin tone and reflection parameter changes as a function of emotional state (blushing, slight perspiration, glinting eyes, and so on). Finally, the system can render animations to create

7 Learning environment system that includes a hidden eye tracker and "Buddy," a virtual friend for autistic children.



8 Virtual child's many faces.



9 Cybertherapy system used to help treat various phobias.

streamable video snippets or it can convert them to Macromedia Flash format to include small animated reactions for home-based care and Internet applications.

Cybertherapy

VR therapy is the emerging field of noninvasive medicine. Computer graphics researchers and clinical psychologists have successfully applied it to help relieve many physical and psychological problems during its 10-year history. VR therapy offers treatment in such broad areas as alleviating pain in burn patients, rehabilitating people after a severe stroke, or curing those with fears and phobias.[5] Furthermore, cyberpsychology can treat social phobias such as fear of public speaking, anger management, and help with the social pressure that substance abuse patients often face at a party or a bar.

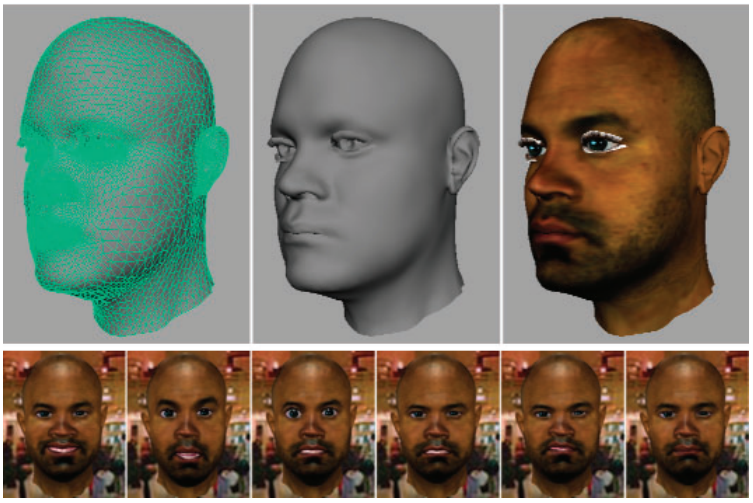
The underlying principle of VR treatment is make-believe. It works by leading patients to believe that a given virtual situation in which they learn to overcome their respective phobia is real. The important part in this process is the gradual nature of the treatment. While patients progress through increasing levels of difficulty, much like in a game, eventually their behavior changes. DTS supports this capability with its advanced visual capabilities and closed loop. Biometric measurements help adjust the stressful stimuli levels so patients remain just at the limit of tension of what they can bear. Therefore, by repeatedly allowing patients to practice the same lifelike scenarios under computer control, the system becomes a vital tool for fast rehabilitation.

Figure 9 shows two examples of the cybertherapy system using the same underlying architecture. The screen shot on the left demonstrates a private jet interior complete with the therapist sitting across from the user to treat fear of flying. This application combines panoramic video and synthetic animated 3D elements. On the right, a virtual hotel lobby depicts a fully computer-generated synthetic 3D environment that serves as practice ground for patients afraid of heights—looking down from the balcony at each floor—and fear of closed spaces—taking a ride in the elevator.

Cognitive diagnosis and rehabilitation

Photoreal animated faces created using computer graphics tools can also help diagnose psychological and brain disorders. This system employs high-fidelity 3D models of human faces to display basic emotions that can be most reliably recognized from facial expressions. The patients' ability and the reaction time it takes to recognize these basic emotional expressions (neutral, happiness, surprise, fear, anger, disgust or sadness) can help doctors determine mental health.

Recognizing the emotion shown in a face involves many different brain structures responsible for perceptual processing and recognition of facial components. Var-



10 Diagnosing psychiatric disorders using animated facial expressions of six basic emotions shown on bottom row (from left to right: happiness, surprise, fear, anger, disgust, and sadness).

ious stages of emotional processing represent different evolutionary advantages and thus they developed in the human brain at different times. This human brain's distributed architecture gives rise to the possibility of studying various brain disorders through facial image animation. Clinical psychologists have long used facial photographs to study their patients. For example, the Szondi test—to detect psychiatric disorders—is a projective technique based on a person's reaction to a series of 48 photographs of psychotic patients. Similarly, it is well documented that depression can be detected as a cognitive bias in emotion. In fact, when depressed patients are asked to select a neutral face from a database, they tend to choose faces that are approximately 15 percent biased toward sad faces. Another example comes from patients with Alzheimer's disease who appear to be significantly impaired when performing tasks related to facial emotion matching. Studies have reported similar findings with obsessive-compulsive disorder, Parkinson's disease, schizophrenia, and multiple sclerosis.[6]

Figure 10 shows the digital model of a neutral head in the 3D modeling stage (top) and the six basic emotions in the real-time rendering environment. Patients press a button when they recognize a specific expression from the animated emotional transitions, starting from the neutral expression to one of the six emotions. These experiments are repeated using different timing characteristics, 3D viewpoint, shape, and texture (for example, gray head only) to obtain a detailed characteristic profile of the subject, which a doctor can later evaluate statistically.

Conclusions

The key thesis of this work stipulates that current human–computer interfaces for education and rehabilitation can be greatly improved by recognizing the long term need for a novel communication layer that can serve as an intermediary between users and today's systems of ever increasing complexity. To fulfill this demand for such a dedicated layer is a task for interdisciplinary research involving many branches of science reaching beyond the boundaries of computer graphics and interface design. Yet this layer might be considered as the key element of a human-centered protocol that determines how users and their computers interact with one another for the decades to come. Failing to implement such a layer and the corresponding architecture will eventually hinder our ability to keep up with the amount of information our brains have to cope with on a daily basis. As such, the architecture described herein fulfills a vital need in computer graphics and education by becoming an intermediate tool that eventually will be and must be seamlessly integrated into our computer culture and become virtually transparent to students of all ages, cultures, and capabilities. ■

References

1. J. Rickel et al., "Toward a New Generation of Virtual Humans for Interactive Experiences," *IEEE Intelligent Systems*, vol. 17, no. 4, 2002, pp. 32-38.
2. B. Takács, and B. Kiss, "Virtual Human Interface: A Photorealistic Digital Human," *IEEE Computer Graphics and Applications*, vol. 23, no. 5, 2003, pp. 38-45.
3. B. Kort, R. Reilly, and R.W. Picard, "An Affective Model of Interplay between Emotions and Learning: Reengineering Educational Pedagogy—Building a Learning Companion," *Proc. Intl. Conf. Advanced Learning Technologies*, 2001; http://affect.media.mit.edu/projectpages/lc/icalt.pdf.
4. R. Ramloll et al., "A Gaze Contingent Environment for Fostering Social Attention in Autistic Children," *Proc. Eye Tracking Research and Applications Symposium on Eye Tracking Research and Applications*, 2004, pp. 19-26; http://s91149240.onlinehome.us/rampub/Pub/autistHelicop.pdf
5. C. Sikne Lányi et al., "Developing Multimedia Software and Virtual Reality Worlds and their Use in Rehabilitation and Psychology," *Proc. 2nd Int'l Conf. E-Health in Common Europe*, IOS Press, 2004.
6. G. Csukly et al., "Evaluating Psychiatric Patients Using High Fidelity Animated Faces," *Proc. CyberTherapy*, 2004.

Barnábas Takács *is the founder of Digital Elite, a research company dedicated to developing next-generation technology for real-time virtual humans and their applications. His research interests include photoreal human animation and modeling, face recognition, motion tracking, and AI-based intelligent conversational agents as well novel applications of VR in rehabilitation. Takács received a PhD in computational sciences from George Mason University. Contact him at btakacs@digitalelite.net.*