Overcoming the Uncanny Valley

Tom Geller

Stories about golems—humanlike beings molded from base clay—have ancient origins in Jewish folklore. According to traditional stories, these soulless creatures serve wise men for both household tasks and community protection. Just as often, golems turn on their masters. But mindlessness, not malice, fuels their conversions to brutality. For these automatons are humanlike but not quite human—they're like animated corpses, lacking volition and desire.

Some say the golems' nearness to ourselves makes them more frightening than other threats. Whereas a lion might tear our flesh and a fire might cause greater destruction, golems—even when peaceful—mock us with reminders of how easily we could lose the warmth that defines our humanity. Scary stories throughout history have featured such characters, varying only slightly in how their humanity fails. Vampires, zombies, aliens, devils … all are nearly human but lack that certain spark of life. To this list of golems, we can add Gollum, a recent instance captured cinematically in the *Lord of the Rings* trilogy, who lost his goodness through enslavement to the Ring.

Golems and graphics

What makes some near-human characters scary while others are merely laughable? More important, why do some human and humanlike characters fail to arouse our sympathy? Visual artists and roboticists face these questions as they seek to alternately frighten and endear. Recent attempts to create accurate human replicas have brought these questions to the fore with increased urgency.

One often-cited example is the 2004 movie *The Polar Express*, in which director Robert Zemeckis and Sony Pictures Imageworks created fully digital replicas of the actors using Vicon motion-capture technology. The movie was controversial and opened to mixed reviews that alternately commended and criticized the effect. CNN reviewer Paul Clinton said "[the] human characters in the film come across as downright … well, creepy." On the other hand, Roger Ebert at the *Chicago Sun-Times* praised the characters as having "a kind of simplified and underlined reality that makes them visually magnetic." (The market seems to have sided with Ebert, as the movie recouped its $150-million-dollar budget by its ninth week.)

Many who criticized *The Polar Express* pointed to a short 1970 essay by Japanese roboticist Masahiro Mori. The essay's title, "Bukimi no tani," is widely translated as "The Uncanny Valley." In it, Mori gave examples of several types of moving and still humanlike images. He posited that, when such characters approach realistic similarity to humans, they stop being likable and instead become eerie, frightening, repulsive—"uncanny." But if the similarity is perfected, the supposition goes, such images would become indistinguishable from humans and therefore elicit as much attraction and sympathy as an ordinary person. Figure 1 (next page) shows Mori's graphical representation of this idea.

The uncanny-valley conjecture has roots in a 1906 paper by German doctor Ernst Anton Jentsch, "On the Psychology of the Uncanny" ("Über die Psychologie des Unheimlichen"). Jentsch discussed the fear engendered by "automata" (which act as though alive) and wax figures (which appear as though alive). He asserted that the essence of this unease comes from uncertainty: is the object alive or not? Freud later focused on the aesthetic implications of this unease in his 1919 paper, "The Uncanny" ("Das Unheimlichen"), further suggesting that the fear of death was involved, referring to the uncanny effect "in relation to death and dead bodies, to the return of the dead, and to spirits and ghosts." (Mori made this connection clear in his paper, saying, "When we die, we fall into the trough of the uncanny valley." He also notably placed dead people—corpses and zombies—at its deepest point.)

And yet, anthropomorphic characters that are clearly nonhuman generally don't cause the "creepy" feeling associated with the uncanny valley. Artist and author Scott McCloud gives one possible reason for human identity with the nonhuman in his groundbreaking book, *Understanding Comics*. He says that realism shows the outside world as our

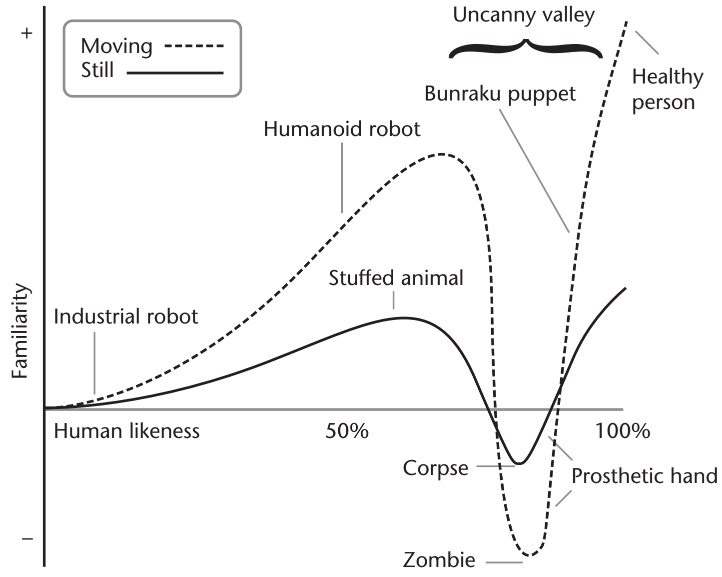

Figure 1. Masahiro Mori's graphical representation of the uncanny valley. The graph demonstrates a gap in familiarity as near-human likenesses approach realism.

eyes see it, while the vague outlines of "cartoony" characters are how we see ourselves with our inner eyes. In particular, "furry" animal-human hybrids externalize inner concepts of self by appealing to the iconic status of animal traits—a status that's cross-cultural and reaches back to the beginning of documented history.

Is there a valley or not?

The uncanny-valley conjecture is well known to anyone who creates humanlike characters or machines. But its repercussions—and its limits—extend beyond these fields, as robotic and animated characters grow ever closer to being indistinguishable from living humanity. Take psychological research, for example: If these androids can pass a Turing test, they could be used to test subjects' reactions to various human conditions. Yet unlike real people, their features and behaviors could be controlled for iterative experiments, and reactions to them could be tested in a consistent way.

So we have nonrealistic humanoid characters that are attractive, near-realistic humanlike characters that repel, and actual humans that engender familiarity. Does that prove Mori's uncanny-valley conjecture?

Many in the field say no.

Ron Fedkiw is one such naysayer. An associate professor of computer science at Stanford University and a frequent consultant to the movie industry for his work on fluid dynamics and biomechanics, he said, "The uncanny valley is usually described as [a situation where] the animations are getting better, but we have fallen into a valley where they are worse. I instead feel that we are in a multidimensional space, and that we are zigzagging all over the place." Put another way, he said, photorealistic animation isn't in a valley toward the right side of the graph in Figure 1. Rather, it's at the beginning of its own curve, separate from those for 2D and 3D nonphotorealistic animation.

Both Fedkiw and roboticist David Hanson (www.hansonrobotics.com) agree that part of the problem lies in how the uncanny valley is perceived. Hanson, whose filmmaking and sculpting background informs his work creating realistic human robots, said, "Mori put forth the uncanny valley as speculation, not as a true scientific theory. But he drew it as a graph, and that made it seem more scientific. It's not a scientific hypothesis that was tested with data, though."

Even if the uncanny-valley conjecture isn't science per se, scientists have widely tested the broader question of how human forms and behavior are perceived. However, comparatively little research has examined the uncanny-valley conjecture itself, and Hanson has decried many of those studies as "prebiased." By way of proof, he published "Expanding the Aesthetic Possibilities for Humanoid Robots," a paper describing three thought experiments that demonstrate the uncanny valley effect, its opposite, and a random distribution of "eeriness" that's unrelated to characters' perceived humanity. Figure 2 shows his graphic demonstrating this experiment. He also included a small-scale study showing that test subjects found a certain real-life human (actor Mitch Cohen as "The Toxic Avenger") to appear notably eerier than Hanson's robot replica of Philip K. Dick, as Figure 3 shows (page 14).

For animators, style triumphs over truth

Many animators are also dubious of the uncanny valley as Mori proposed it—or, at least, of how closely "familiar" and "realistic" are paired in the uncanny-valley conjecture. Sony Pictures Imageworks' Kenn McDonald, a lead character animator on *The Polar Express*, pointed to the need for moviemakers to stylize their characters away from realism to make them effective, "much like putting makeup on a flesh-and-blood actor." Speaking of his later work as animation supervisor of the 2007 film *Beowulf*, McDonald said, "Our intention's not to fool people that it's actually Anthony Hopkins on the screen—they know it's not. Instead, we want to get something that's convincing within the realm of that world we're portraying. In *Beowulf*, there was a lot of character detail, but our intent was never to hit that mark and be truly real."

He pointed to both *Beowulf* and *Lord of the Rings*' Gollum as examples. "A good way to avoid the uncanny valley is to move a character's proportions and structure outside the range of 'human.' One reason Gollum was so successful is that he has

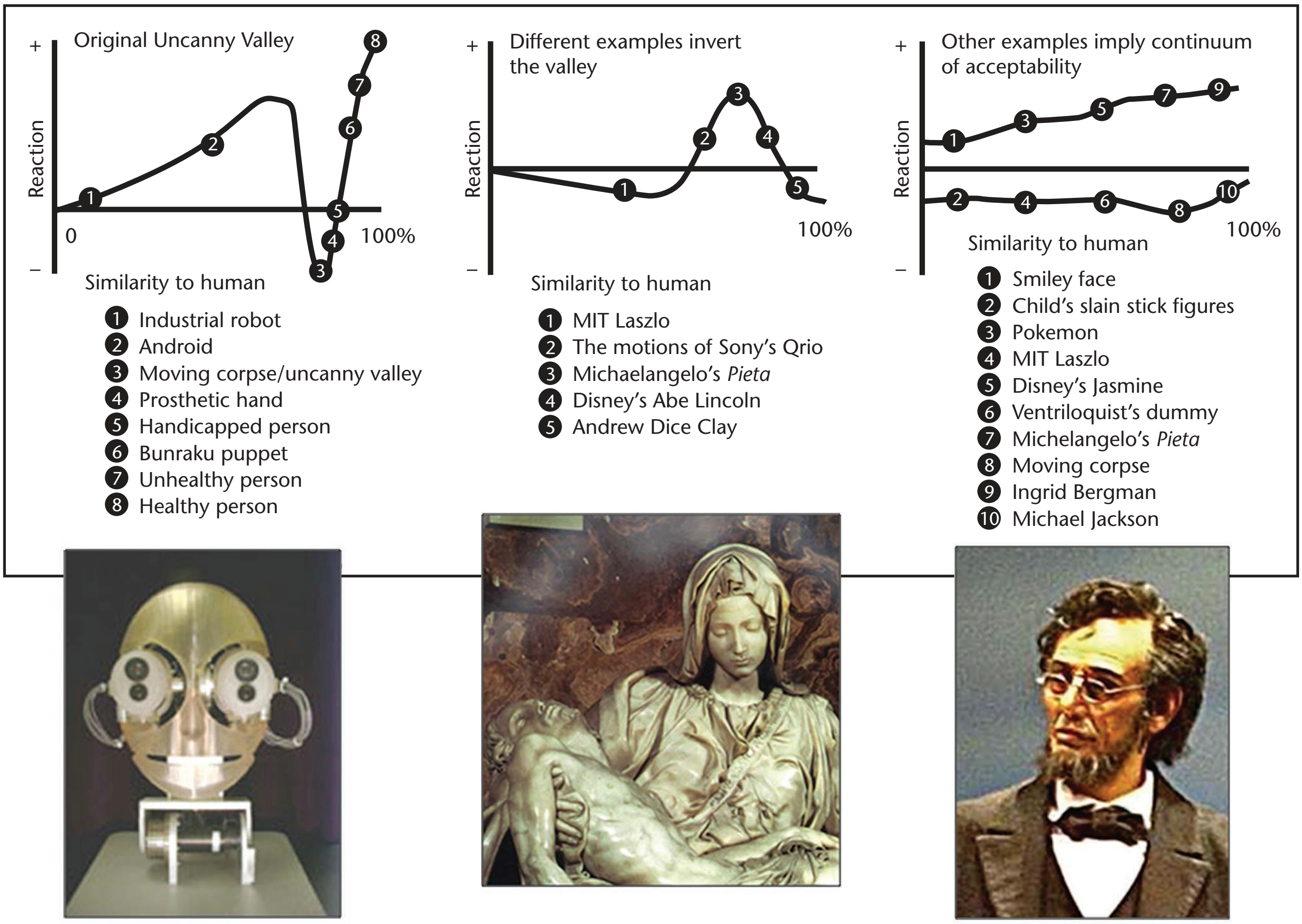

Figure 2. Three proposed models that contradict the uncanny-valley conjecture. The first graph shows the conjecture as Mori depicted it; the second, an arrangement of examples that contradict Mori directly; and the third, examples that deny either of the progressions implied in the first two. (Originally published in "Expanding the Aesthetic Possibilities for Humanoid Robots," a paper presented by David Hanson at the IEEE-RAS Int'l Conf. Humanoid Robotics, Tsukuba, Japan, 2005; used by permission of the author.)

big eyes, and the shape of his face is not quite human. Grendel in *Beowulf* is also disfigured and deformed. The audience subconsciously says, 'He's not human; I don't have to judge him by the same rules as if he were.' But when we try to portray a human, viewers notice what's missing."

Matt Aitken, a visual-effects supervisor at Weta Digital responsible for work on Peter Jackson's *Lord of the Rings* trilogy and *King Kong*, agrees that stylization is necessary in moviemaking—and helps avoid possible uncanny-valley issues. "What we're after is a digital facsimile of the human actor that's believable for that shot. That might be different from a realistic facsimile of that actor. We generally feel that if you want a natural-looking human, why not just cast a human actor? I'm not sure what would happen if a director we wanted to work with came to us and said, 'We want a completely believable digital human.' We'd rather use our skills and techniques to create nonhuman characters."

But Aitken did give one reason for using realistic digital human characters: as digital doubles, which can act as stunt performers or stand-ins. "In *Lord of the Rings*, there are shots with big, sweeping camera moves across miniature environments," Aitken says. "It would have been very difficult to film actors in that situation, so we created digital versions. In one sequence in *King Kong*, he's fighting dinosaurs and has the Ann Darrow character in his hand: That would have been difficult to do with a blue screen."

Making people the newfangled way

On the robotics side, the question is somewhat more involved. That robots exist in three dimensions is only one complicating factor: size, voice type, interactivity, material quality, and need for outboard support (such as visible wires and a power source) make robots more obviously nonhuman than movie characters. Sara Kiesler and Aaron Powers of the Project on People and Robots (www.peopleandrobots.org) have published several papers on human-robot interaction. Both said that the uncanny valley remains an uncertain quantity in

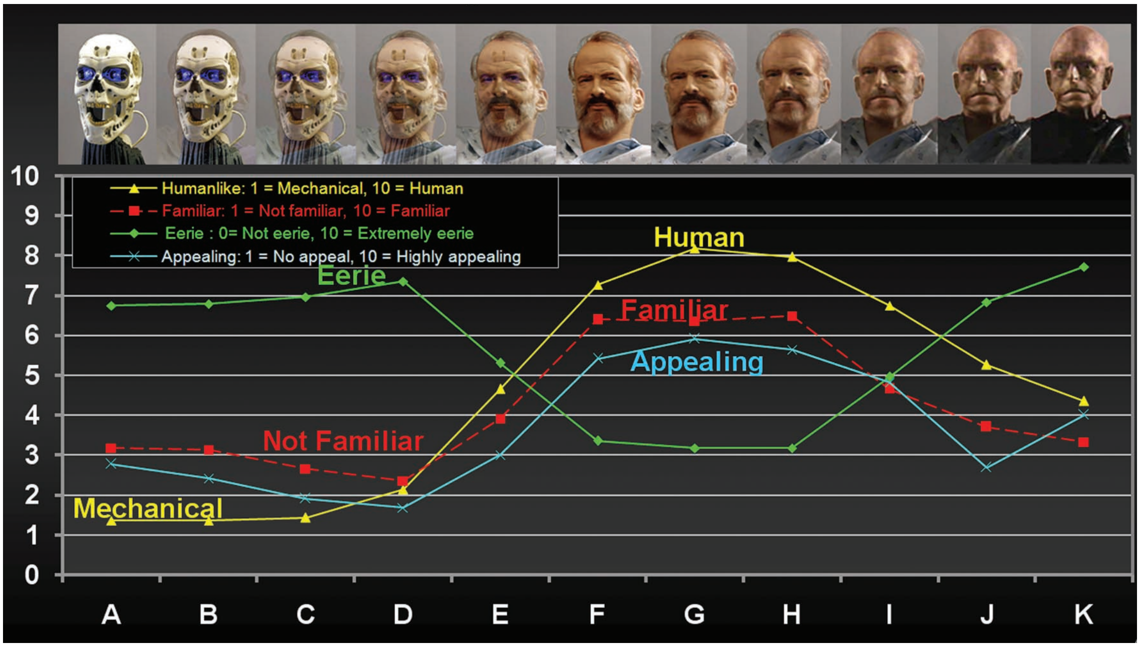

Figure 3. A clinical demonstration by David Hanson that appears to conflict with the uncanny-valley concept: Viewers tended to find the real-life human on the right more "eerie" than the robot in the middle. (Originally published in "Expanding the Aesthetic Possibilities for Humanoid Robots," a paper presented by David Hanson at the IEEE-RAS Int'l Conf. Humanoid Robotics, Tsukuba, Japan, 2005; used by permission of the author.)

such meetings. According to Powers, "We'd need a [convincing] android to really be able to do those sorts of experiments." Kiesler added, "There's some evidence that [the valley] exists, and some that it doesn't."

Realistic human likenesses in the movies, on the other hand, have successfully fooled observers for almost 20 years, helped greatly by the introduction of computer graphics (CG) technologies. Such tools were used as early as 1976, when a digitized version of a hand appeared in *Futureworld*. They were exploited further in nonrealistic contexts in *Tron* (1982) and *Young Sherlock Holmes* (1985), where a knight made up of stained-glass panes became what's considered the first narrative CG character.

But it wasn't until the 1991 movie *Terminator 2: Judgment Day* that CG humanoids started to approach realism. From that point, the game was on. CG animation allowed Jim Carrey to display a wide variety of physically impossible poses in the 1994 vehicle *The Mask*, while the nonrealistic CG-created ghosts in the 1995 *Casper* were on screen for over 40 of the film's 100 minutes. Then came attempts to portray realistic human characters—or, at least, nearly human characters, as in *The Mummy* (1999), *Hulk* (2003), and *Lemony Snicket's A Series of Unfortunate Events* (2004).

Although the outward appearance of these (and similar) humanlike characters varied considerably, their behavior was increasingly modeled through use of a motion-capture, or mocap, procedure. The essential idea of mocap is to track human performers' actual movements, usually by attaching dozens or hundreds of trackable points to their bodies, and then converting tracked data into vectors that effectively replicate their movement. Those vectors can then serve as automated or manual guides for traditional animation (as was the case for the 2000 *Sinbad: Beyond the Veil of Mists*, the first animated movie to use mocap exclusively for character animation). The vectors can also be applied computationally to a CG character's exterior "skin" or combined in several ways for hybrid solutions.

Several types of mocap systems are available, varying in tracker types and how the data is processed. The vast majority in use today are optical-marker systems, in which the actor wears a number of reflective or light-emitting "dots" at such crucial points as joints and facial high points. Some mocap systems use other types of markers—magnetic, for example—or forego markers entirely by attempting to computationally resolve a video stream into manipulable vector data. In either case, such a system's resolution is measured in two ways: by the number of reference points and by the frequency with which they're sampled.

Hal Hickel, the animation supervisor for Industrial Light & Magic who worked on the *Pirates of the Caribbean* movies, feels that mocap has overcome virtually all uncanny effects in replicating gross body movement. "Largely the problem of realistic human motion was solved with motion capture," he said. "It's in the movement of the face where you run into problems." The late-1990s development of subsurface scattering, first used for the Dobby character in the 2002 film *Harry Potter and the Chamber of Secrets*, further improved the appearance of CG skin by giving it realistic translucency. So what's left? According to Hickel, "We've gotten out of the valley with body movement—now it's all about the eyes and facial performance."

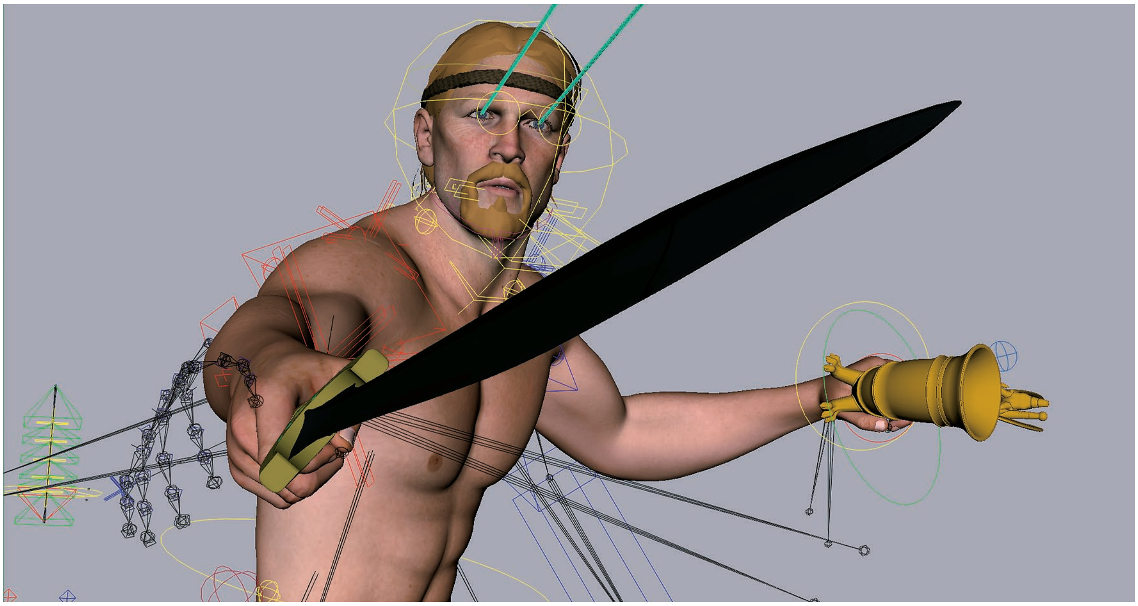

Figure 4. A realistic computer graphics character in an early production stage. Particular attention is given to the eyes, which lead the body direction. (Used by permission of Sony Imageworks.)

The eyes have it

Indeed, animators and roboticists generally agree that realistic eyes are the key to avoiding the facial "creepiness" associated with the uncanny valley. (The title of one review of *The Polar Express* was "Cold Eyes, Warm Heart.") Roboticist and toy designer George York of YFX Studio learned this firsthand by observing the reactions of people who interacted with his mass-produced, $299 robotic bust WowWee Alive Elvis at the Consumer Electronics Show. "We got mixed emotions when we presented Elvis. But as soon as we put sunglasses over his eyes, people were comfortable with it."

Kenn McDonald agreed. "We'd done a fair amount of research and work on the eyes for *The Polar Express*—both on how they move around in the socket and from a rendering point of view—but that didn't really come across. For *Beowulf*, we had experts come in and talk about eye physiology and the psychology of eye movement; we had a capture system with four electrodes around the eyes to capture the electrical impulses of the muscles. We studied how the eye floats as it moves from side to side and up and down, and how peoples' eyes move in different situations—subtle, subconscious movements. And we incorporated it all into *Beowulf*. It wasn't always perfect, but [these studies] gave us a very good basis for the characters' eye motion" (see Figure 4).

"Eyes are supercritical," said *Pirates* animation supervisor Hal Hickel. "When we were doing the scene in the first *Pirates* movie, where Geoffrey Rush turns into a skeleton, we did a trick where we preserved his eyes for a split second before he fully transformed. That's always where the audience is looking, and people are really sensitive if you get the eye line off even just a little bit. People have a lifetime of looking at others' eyes to know whether they're really listening or not."

Another unusual example of eye tracking occurred in the 2006 sequel, *Pirates of the Caribbean: Dead Man's Chest*. There, Davy Jones' face is presented as humanlike, but covered in tentacles and lacking a nose. Because of these differences, the facial animation of Davy Jones was an "interpretation" of Nighy's performance, not an exact match. While animators had to capture the same complexity and subtlety in Jones' animation as was found in Nighy's acting, Jones' bizarre appearance helped ease the character out of the uncanny valley. "We had to deal with the uncanny-valley problem with Davy Jones and his crews," Hickel said. "They were humans—hideously mutated, but still humans. It helped that [Jones] wasn't just a normal human, but really the audience is looking at his eyes, so we had probably 90 percent of the trouble that we'd have with a regular character."

After the eyes, facial movement is a close second in importance to maintaining (or failing to maintain) believability. Steve Perlman, president of mocap service bureau Mova LLC, explained some of the technical problems of doing facial mocap. "The problem [with tracking markers] is that you just can't put them too close on the face. They're expensive, they bump into each other, and the markers are uncomfortable to the actors. Very often after shooting a scene in *Beowulf*, they'd sweep a flashlight along the ground to see if they found any markers that fell off! Some people are using pen markings on the face—that's obviously easier to do, but you still have limited density, plus there are shadows and reflections that cause problems."

To solve these problems, Mova developed Contour, a markerless system that uses retrospective vertex tracking (RVT). In RVT, a phosphorescent makeup is applied to a performer's face, and the software tracks the random patterns in the

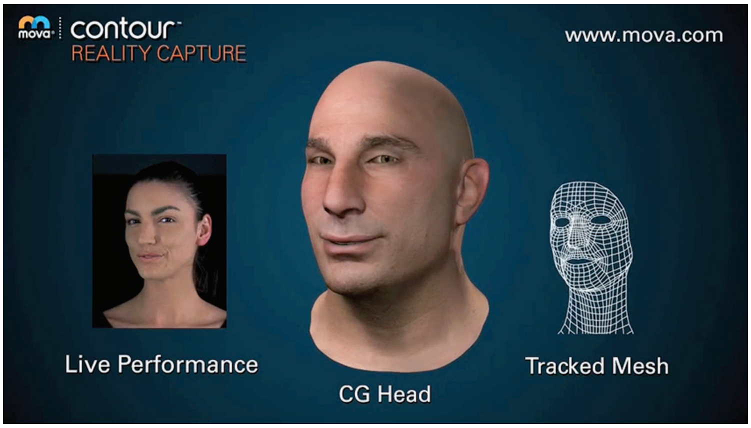

Figure 5. Mova's Contour system. A phosphorescent makeup is applied to a performer's face, which lets the software track random facial patterns. The tracked mesh can be used to retarget motion from any performer onto any computer-graphics head model. This example shows capture data from a female performer, Jessiqa, driving the photoreal head of a male character, David. (Used by permission of Mova.)

sponged-on makeup. Once the makeup has been applied, the room is darkened to pitch black, then illuminated by a light that strobes 120 times per second. The strobe frequency is above the perception threshold, so the room appears normally lit. Multiple cameras shoot simultaneously. Their shutters are open only when the strobing lights are off, thereby capturing emissive light from the phosphorescent makeup without any reflective shadows from the lights. The result is a capture of "millions of points, but we typically use between 1,000 to 100,000"—far more than with marker-based systems.

"The sponging puts on a random pattern on the face, that's the key," Perlman said. "First of all, it's easy to do—you can't *not* put a random pattern on the face. Where the standard marker system takes a random thing—a face—and makes it regular, Contour makes a face consistently random. That random pattern from frame to frame stays the same, so we can have continuous geometry." These points, shot from multiple angles, are then triangulated in postproduction to form 3D images. "We don't know which points on the face we want to track at the time of the shoot. Then, when we're animating we say, 'Well, this particular performer has this particular feature, and we can track particular points to bring that out.'" (Motion characteristics of one performer can also be mapped onto the facial structure of another in a process called retargeting, as shown in Figure 5.)

Because tracking points are determined in postproduction by essentially reducing resolution, the studio can also render characters in multiple resolutions. This allows applications that, for example, use fewer resources in low-resource situations such as video games.

As we go to press, no movies have been released using Mova's Contour system, so the market hasn't had the opportunity to judge whether it overcomes the uncanny-valley problem. But Perlman believes they've achieved success. "We knew we'd crossed over when we asked people off the street, 'What do you think of this face? What's wrong with it?' After lots of development, we finally got people saying, 'I don't know what you're talking about—it's just a video of a face, right?' The difference was a very subtle improvement to our technology. That's a sign to me that there's an [uncanny valley] threshold: You're either spot on or you're not playing ball." (Examples of animations using Contour are at www.mova.com/gallery.php?g=examples.)

The well-behaved simulacrum

So perhaps we've built something that looks and moves like a real person. But without a self-directed behavior system, it remains merely a puppet, suitable only for performance and rote interaction. Stephen Regelous, whose Massive Software was responsible for creating the crowd scenes in the *Lord of the Rings* trilogy, believes that both robots and digital actors need internal intelligence to cross the uncanny valley.

"Truly autonomous characters have a believability that's extremely hard to generate through the puppeteering method," he said. "When you watch a character that's performing on its own, with subtly reactive actions, you start to believe in something that's not possible with [digital] puppeteering." Although Massive's work in *Lord of the Rings* was limited to crowd scenes, the individual characters in those crowds were effectively thinking—that is, interacting according to algorithms, rather than explicit directions. "Using Massive, agents get a scan-line rendered image—a picture of what's in front of them—on every frame," he said. They actually see, just as you and I do. Since they can see and react, they don't run into each other. They're evaluating what they're supposed to be doing 24 times a second."

To create these self-directed characters, a user of Massive's software first breaks down the director's needs into a "motion tree" that dictates which actions can transition to others. If a character needs to go from sitting to running, for example, the motion tree would dictate that it first stand. Second, Massive's software creates a "take list" that specifies which actions must be captured with an actor and standard mocap techniques. Finally, the captured actions are imported into Massive to populate the motion tree. "So the digital actor's performance isn't directly driven by the [live] actor's performance, but rather by its reactions to what's going on around it, and largely constructed from its library of short actions," Regelous said. Put computationally, Massive's approach is similar to the switch from CISC (complex-instruction-set computing) to RISC (reduced-instruction-set computing). And just as RISC processors can combine their short instructions into practically any complex whole, "all those little pieces of motion can create any performance."

Regelous pointed to a further implication of Massive's approach: replicating functions of the human brain itself. He refers to Ray Kurzweil's concept of the "singularity," when artificial intelligence becomes greater than our own, leading to an explosion of superhuman development. He further agreed with Kurzweil's prediction, coupled with Moore's law, that this will likely happen in about 30 years. "A lot of computation goes into [some realistic animation tasks]. They're computationally expensive, but are actually using a different set of resources than the puppeteering approach. The puppeteering approach uses the 60 billion neurons in the animator's brain. If the character had the equivalent of 60 billion neurons, it could do the job [of self-animation] just as well."

In other words, research in these fields might help lead to the effective creation of new life.

Is it live, or is it eerie?

Which leads us back to Golems and our fear of the nonliving alive.

Unsurprisingly, many who confront the uncanny valley's technical challenges also concern themselves deeply with how a truly convincing "agent" would affect humanity. Yoseph Bar-Cohen, a NASA senior research scientist and coauthor (with roboticist Hanson) of the book *Humanlike Robots*, believes that humanity is preparing itself for this eventuality, pointing to movies such as *Toy Story* that portray robots as "the good guys." "For the first time we're doing pre-preparation on technology that doesn't even exist," he said. "We are starting to have discussions of robo-ethics on all levels."

Bar-Cohen also noted the influence that Eastern versus Western religion might have on our perceptions of humanlike robots. "In America, we barely have any robots that look very human. But in Asian cultures, they're more acceptable. There's a Shinto belief that God is in everything, including robots, and children there grow up with stories of a robot as a savior of the people. Meanwhile, Judaism prohibits making statues and things that look like humans, because they're an image of God. There's a concern that we can create something that will destroy us."

As if to confirm Bar-Cohen's assertion, Masahiro Mori turned to studying robotics in the context of Buddhism shortly after penning "Bukimi No Tani." Four years later, he published *The Buddha in the Robot: A Robot Engineer's Thoughts on Science and Religion*. He now leads the Mukta Institute, a think tank on the subject. In a statement presented to the IEEE Robotics and Automation Society's 2005 International Conference on Humanoid Robots, Mori said,

> Once I positioned living human beings on the highest point of the curve in the right-hand side of the uncanny valley. Recently, however, I came to think that there is something more attractive and amiable than human beings in the further right-hand side of the valley. It is the face of a Buddhist statue as the artistic expression of the human ideal. ... Those faces are full of elegance, beyond worries of life, and have aura of dignity. I think those are the very things that should be positioned on the highest point of the curve.

It's arguable whether animation technologies have reached the point where the materials have the expressive capabilities of a Buddha marble—or whether they ever will. If so, Mori's statement suggests that they could not only embody humanity, but better it.

Contact author Tom Geller at tom@tomgeller.com.

Contact Applications editor Mike Potel at potel@wildcrest.com.