Presenting in Virtual Worlds:

An Architecture for a 3D Anthropomorphic Presenter

Herwin van Welbergen, Anton Nijholt, Dennis Reidsma, and Job Zwiers,
University of Twente

To present and explain information, this 3D humanoid presenter uses output channels such as speech and animation of posture, pointing, and involuntary movements.

Meeting and lecture room technology is a burgeoning field. Such technology can provide real-time support for physically present participants, for online remote participation, or for offline access to meetings or lectures. Capturing relevant information from meetings or lectures is necessary to provide this kind of support. Multimedia presentation of this captured information requires a lot of attention.

Our previous research has looked at including in these multimedia presentations a regeneration of meeting events and interactions in virtual reality. We developed technology that translates captured meeting activities into a virtual-reality version that lets us add and manipulate information.[1]

In that research, our starting point was the human presenter or meeting participant. Here, it's a semi-autonomous *virtual presenter* that performs in a virtual-reality environment (see figure 1). The presenter's audience might consist of humans, humans represented by embodied virtual agents, and autonomous agents that are visiting the virtual lecture room or have roles in it.

In this article, we focus on models and associated algorithms that steer the virtual presenter's presentation animations. In our approach, we generate the presentations from a script describing the synchronization of speech, gestures, and movements. The script has also a channel devoted to presentation sheets (slides) and sheet changes, which we assume are an essential part of the presentation. This channel can also present material other than sheets, such as annotated paintings or movies.

The virtual presenter's architecture

Building a virtual presenter involves many different techniques, including facial and body animation and speech, emotion, and presentation style generation. The main challenge is to integrate those elements in a single virtual human.

Integration concerns

Such integration raises two major concerns.[2] The first is consistency. When an agent's internal state (for example, goals, plans, and emotions) as well as the various channels of outward behavior (such as speech, body movement, and facial expressions) are in conflict, inconsistency arises. The agent might then look clumsy or awkward, or, even worse, appear confused, conflicted, emotionally detached, repetitious, or simply fake. Because our virtual presenter currently derives its behavior from the annotated script of a real presentation, consistency conflicts arise mostly between the implemented and unimplemented channels. For example, one of our 3D models can't move its mouth. When it speaks, this looks awkward. When we extend the presenter to dynamically generate its behavior, consistency will become even more important.

The second concern, timing, is currently more crucial. The agent's different output channels should be properly synchronized. When an agent can express itself through many different channels, the question arises, what modality should primarily determine the timing of behavior? For example, BEAT (Behavior Expression Animation Toolkit), which generates nonverbal animation from typed text, schedules body movements as conforming to the time line that a text-to-speech system generates.[3] Essentially, behavior is a slave to the speech synthesis tool's timing constraints. In contrast, EMOTE (*E*xpressive *Mot*ion *E*ngine) takes a previously generated gesture and shortens it or draws it out for emotional effect.[4] Here, behavior is a slave of the constraints of emotional

1541-1672/06/$20.00 © 2006 IEEE
Published by the IEEE Computer Society

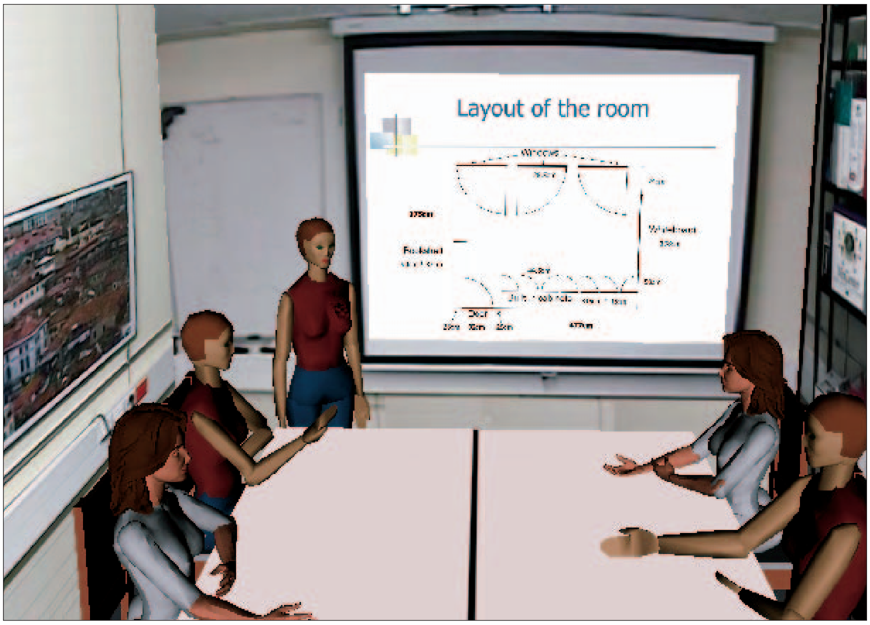


Figure 1. The virtual presenter.

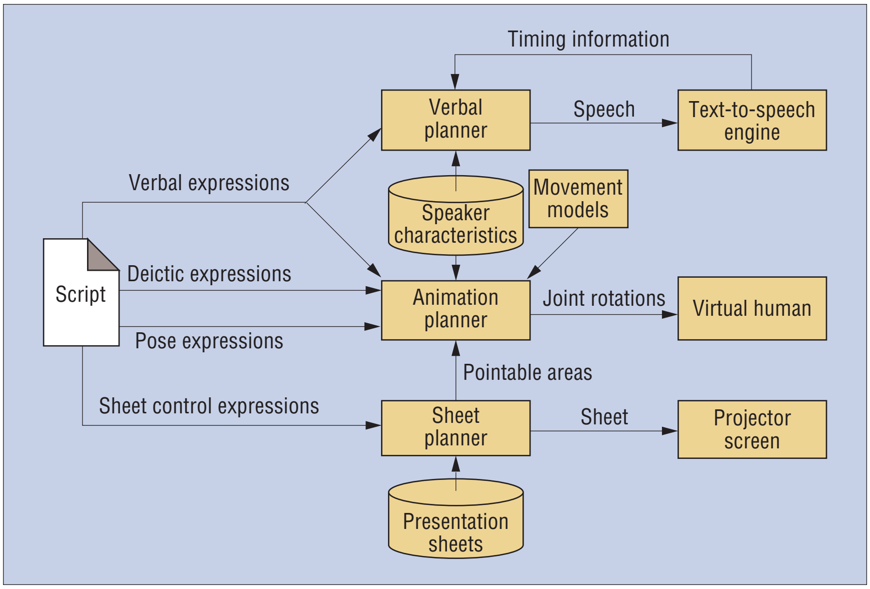


Figure 2. The virtual presenter's system architecture.

dynamics. Other systems focus on making a character highly reactive and embedded in the synthetic environment. In such a system, behavior is a slave to the environmental dynamics.

To allow for different combinations of such constraints, you need at least two things. First, the architecture should allow different leading modalities in relation to the synchronization. It should be possible to select the leading modality dynamically, in real time. Second, different components must be able to share information. For example, if BEAT had information about the timing constraints that Emote generates, it could schedule behavior better. Another option is to design an animation system that's flexible enough to handle all constraints at once. Norman Badler suggests a pipeline architecture that consists of "fat" pipes with weak uplinks.[2] Modules would send down considerably more information (and possibly multiple options) and get relevant information from a module further away in the pipeline (for example, how long it would take to point to a certain target or speak a word).

To synchronize speech and gesture, the phonological-synchrony rule[5] should be satisfied. This rule states that a gesture's peak of effort (called the stroke) precedes or ends at, but doesn't come after, the phonological peak syllable of co-occurring speech (the stressed syllable in the word that relates to the gesture).

The architecture

Our virtual presenter's architecture (see figure 2) is inspired by the pipeline architecture idea, mentioned in Jonathan Gratch's research on creating interactive virtual humans.[2] The presentation script specifies expressions on separate channels. The channels' planners determine how to execute those expressions. To do this, a planner can use information from another module (for example, it can ask the text-to-speech engine how long it takes to speak a certain sentence or ask the sheet planner which sheet is visible at what time) or from human behavior models. The planner could even decide not to execute a certain expression because doing so is physically impossible or because it wouldn't fit the presenter's style.

We choose to implement a selected set of dimensions of a presenter's behavior, using behavioral models from the literature instead of an ad hoc implementation. The presenter is extensible so that we can insert new behavior later on.

Presentation script

As figure 2 shows, the script is the starting point for visualizing a presentation. We can create a script from annotated behaviors observed in a real presentation, or we can generate it from an intention to convey certain information. (In the latter case, the script would indicate the information we want the presenter to present but not what mix of modalities the presenter will use.)

The multimodal channels that the presenter uses are scripted at different abstraction levels. We specify gestures in an abstract manner, mentioning only their type, which indicates communicative intent. We've adapted this gesture classification:[5]

- *Deictic* gestures simply point at something.
- *Beat* gestures have a visual rhythm that seems aligned with an utterance's rhythm.
- *Iconic* gestures bear a close formal relationship to the content of speech—for example, wiggling the index finger and forefinger when discussing walking.

- *Metaphoric* gestures display abstract concepts—for example, holding your nose to indicate disapproval of an opinion.

The planners determine the exact visualization (for example, which body part, hand shape, or movement path to use). We annotate speech at the word level. We annotate poses (a body's resting positions) and pose shifts by specifying the joint rotations in the presenter's skeleton for every pose. We specify sheet changes whenever they should occur.

For synchronization and timing of the presentation in its different channels, we developed the MultiModalSync language. To synchronize a presentation, we set synchronization points in one modality and use these points in another modality. We can set and use synchronization points on all modalities so that the leading modality can change over time. For example, we can set a synchronization point before a word in the verbal modality. The pointing modality can then use this point to define a pointing action that co-occurs with the spoken word. We describe MultiModalSync's constraints and synchronization definitions in greater detail and explain why we had to develop a new script language elsewhere.[6]

Speaker characteristics

Personality or style influences how a presenter presents. Different presenters will perform the same presentation differently. You can define personal characteristics that influence how the presenter performs the presentation. Zsófia Ruttkay and Catherine Pelachaud proposed using a set of static parameters—for example, gender, age, or nationality—to define and tune an embodied conversational agent's style.[7] We adapt this approach. We store individual speaker characteristics in a database. Currently, this database contains parameters about the presenter's voice and pointing movement.

Presentation planning

Each output modality has a planner that plans its execution. The input modalities for these planners aren't necessarily the same as their output modalities. For example, the animation planner takes input from the deictic, pose, and verbal channels and displays them as body movement on a virtual human. Multiple planners can plan input modalities. For example, the presentation's verbal text is planned by both the verbal planner, to generate speech, and the animation planner, to generate mouth movement.

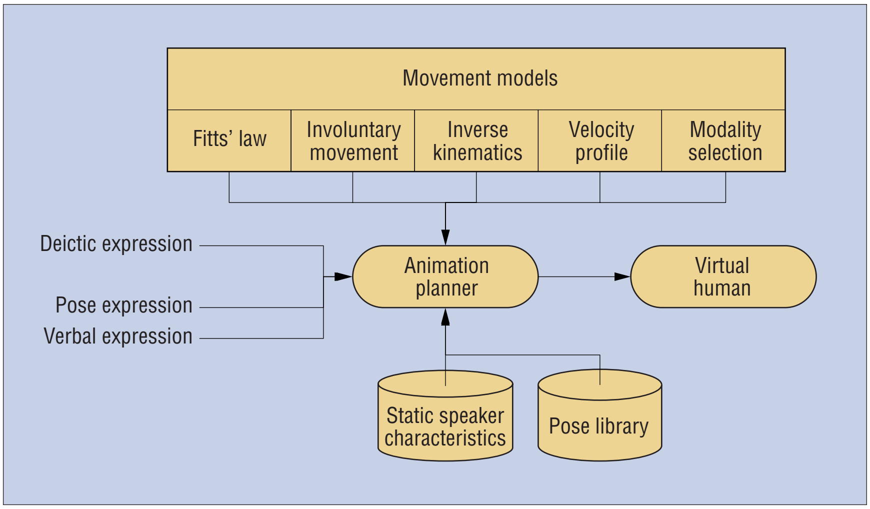


Figure 3. The animation planner.

We also use planners to play back the script. Playback has two phases. In the setup phase, the planner gathers all the necessary information that isn't specified in the script. Each modality has its own setup time, which specifies the time needed to initialize an expression on this modality before it can be played. To initialize an expression, planners can use the speaker characteristics, behavioral models, or information from other planners (such as the speech timing from the speech planner or the position and size of sheet areas from the sheet planner). The planner then stores the information needed to play an expression in the expression itself.

In the execution phase, an expression is played by its planners, which combine it with all the other expressions that they must play.

During playback of an expression, conflicts might arise. For example, two expressions could claim the right hand at the same time. The planners solve those conflicts. Some conflicts can be solved during setup and others during execution. To solve them, planners can

- cancel execution of an expression,
- combine the expressions on the same modality (for example, rhythmically move the hand while pointing, to combine a beat and a pointing gesture), or
- execute an expression on another modality (for example, if the hands are busy, point with the head).

In our current implementation, the presenter combines poses with pointing gestures by showing the pose but letting the pointing action take over the arm and head movement. It skips a new pointing gesture if another pointing gesture is active, or it overwrites the current pointing gesture with a new one if that current gesture is in its retraction phase (which we describe in more detail later).

To plan and play back body animation, the animation planner (see figure 3) uses movement models from neurophysiology and behavioral science. Currently, it can play deictic gestures, pose shifts, and speech (mouth movement) specified in the script. Static speaker characteristics influence how this behavior executes. We can easily extend the architecture to execute other gesture types.

The verbal planner regulates the text-to-speech generation, and the sheet planner regulates the sheet changes.

Speech planning

We use Loquendo's (www.loquendo.com) text-to-speech engine to generate speech, lipsyncing, and speech-timing information from the verbal text. This engine lets us obtain speech timing on the word level. To satisfy the phonological synchrony rule, we can synchronize a gesture's stroke with the start of a word. The animation planner uses a simple form of lip-syncing: the opening of the mouth is proportional to the speech's volume, averaged over a short time period.

Involuntary movement

Even while standing still, the human body moves in subtle ways: we try to maintain balance, our eyes blink, and our chest moves when we breath in and out. An avatar that doesn't perform such subtle motion will look stiff and static.

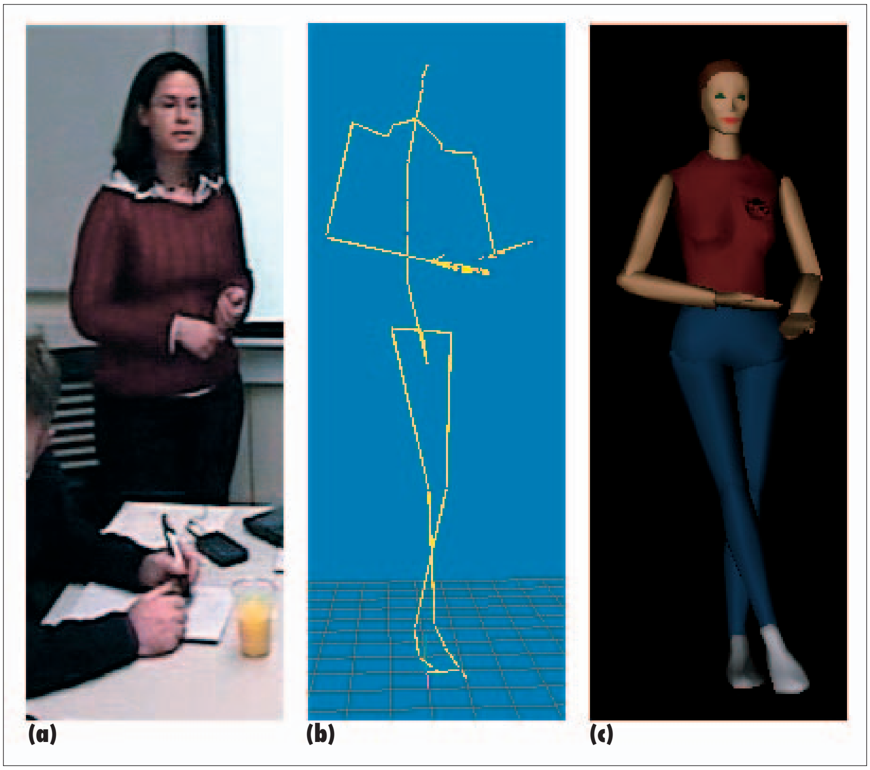


Figure 4. Annotating and simulating poses: (a) a human presenter's pose, (b) a manually created representation in the Milkshape modeling tool, and (c) the virtual presenter showing that pose.

To avoid this problem, our presenter uses an involuntary-movement method that Ken Perlin devised.[8] This method simulates involuntary movement by creating noise on some of the avatar skeleton's joints. We chose this method because it avoids the repetitiveness of predefined scripted idle animations (such as breathing or thumb twiddling) and because the presenter's model isn't detailed enough to use realistic involuntary-movement models. In our approach, we choose which joints to move in an ad hoc manner. For example, we can move the two acromioclavicular joints (between the neck and the shoulder) to simulate small shoulder movement that occurs with breathing. Small rotations of the vl1 joint (the spine's lowest joint) simulate subtle swaying of the upper body.

Posture

We use poses as start and end positions for the limbs in a gesture unit. A gesture unit is the period of time between successive rests of the limbs. It begins when a limb starts to move and ends when the limb has reached its resting position again.

Our presenter system currently specifies each pose separately in a pose library containing the joint positions, rather than using models of when and how people shift poses during real presentations. The current scripts include references to these poses based on a human presenter's pose in a real presentation (see figure 4).

Pointing

A presenter can refer to areas of interest on the sheet by using a gesture with a pointing component. Our pointing model considers several aspects of pointing movement, so that our system can generate the pointing movement given only the intention to point and a pointing target. Like Tsukasa Noma, Liwei Zhao, and Norman Badler's presenter (see the "Presentations by Embodied Agents" sidebar), ours uses its right hand to point to the right and its left hand to point to the left, to keep an open posture. When the preferred hand is occupied, the presenter will gaze at the area of interest instead of directly pointing at it.

Timing. Fitts' law, which predicts the time to move from a certain start point to a target area, is used to model rapid, aimed pointing actions. Fitts' law could thus give a minimum value for the duration of a pointing action's preparation phase. Our virtual presenter uses a 2D derivation[9] of Fitts' law:

$$T = a + b \cdot \log_2\left(\frac{D}{\min(W,H)} + 1\right)$$

where T is the time necessary to perform the pointing action, D is the distance to the object to point to, W is the object's width, and H is its height. a and b both depend on the pointing medium (in our case, the head or arm) and the pointing individual. We empirically determined a and b from a real presenter. We can set their values as static speaker characteristics to create different pointing styles.

Movement in the retraction phase. Humans execute gestures in three phases.[5] In the optional preparation phase, the limb moves away from the resting position to the position in the gesture space where the stroke begins. The obligatory stroke phase expresses the gesture's meaning. In the optional retraction phase, the hand returns to a resting position. Preparation occurs only if the gesture is at the beginning of a gesture unit, and retraction occurs only if the gesture is at the end of a gesture unit.

According to Adam Kendon, gesture movement is symmetric.[10] We analyzed videos of pointing gestures to determine whether this is true for more precise pointing actions. Figure 5 shows screen captures of such a video.

We discovered that

- pointing gestures that form a complete gesture unit by themselves are rare,
- gestures that do form a unit by themselves have symmetric-looking preparation and retraction phases, and
- as Kendon noted, it's hard to tell whether a video of such a gesture is being played forward or backward.

On the basis of these findings, we conclude that the retraction phase consists of the arm moving to the resting position in the same way it moves from the resting position to the stroke position, but in reverse.

Pointing velocity. A pointing movement's velocity profile is bell shaped.[11] This bell can be asymmetric. The relative position-time diagram is sigmoid shaped. We use the sigmoid $f(t) = 0.5(1 + \tanh(a(t^p - 0.5)))$ to define the wrist's relative position. In this function, t represents the relative movement time ($t = 0$ is the pointing movement's start time; at $t = 1$, the wrist reached the desired position). $f(t)$ describes the relative distance from the start position: $f(0) = 0$ is the start position; $f(1) = 1$ is the end position. We can use a to adjust this sigmoid's steepness and p to set the length of the acceleration and deceleration phases. We empirically determined a and p using a real presenter's movement. We add a and p to the static speaker characteristics to allow style-dependent velocity profiles.

Pointing with gaze. We implement gaze behavior during pointing movements on the basis of Donders' law.[12] This law defines the necessary movements and end orientations for the eyes and the head, given that the presenter will look at the pointing target.

Shoulder and elbow rotation. If the pointing target's location and size determine the wrist position, we can analytically calculate elbow and shoulder joint rotations using the inverse-kinematics strategy that Deepak Tolani, Ambarish Goswami, and Norman Badler describe.[13] The elbow, though, is still free to swivel on a circular arc, whose normal is parallel to the axis from the shoulder to the wrist. To create reasonably good-looking movements, the presenter always rotates the elbow downward.

Sheet planning

To display the sheets, the virtual presenter uses a virtual projector screen. These sheets have defined areas of interest at which the presenter can point. A presentation's sheets are described in an XML presentation sheets library. As we mentioned before, the sheet planner handles the display and planning of sheet changes, and it can provide other planners with planning information.

Evaluation

We informally evaluated our virtual presenter's involuntary movement to see whether it made her seem less stiff and made her movements seem more natural. We showed 20 test subjects (15 male and 5 female, ages 17–56) a virtual presentation twice. In one presentation, the presenter demonstrated involuntary movement; in the other, she didn't. We showed 10 subjects the presenter with involuntary movement first and showed the other 10 the presenter without involuntary movement first. Then we asked them which presenter moved more naturally and which one

Presentations by Embodied Agents

Here, we highlight a few projects featuring human-like presenters that use visual aids.

Helmut Prendinger, Sylvain Descamps, and Mitsuru Ishizuka worked on specifying presentations related to Web pages, which are given by Microsoft Agents, a robot, or a 3D presenter.[1] Their main focus has been development of a multimodal presentation language (MPML) with which nonexpert (average) users can build Web-based interactive presentations. Presentation actions on separate modalities (such as gestures or speech) and their synchronization are specified in the script. For every action, an emotion can be specified, which affects that action's execution.

Tsukasa Noma, Liwei Zhao, and Norman Badler simulated a professional presenter such as a TV weather reporter.[2] Their presenter usually interacts with a 2D screen. The animation model is rather simple. It implements two posture shifts—namely, those needed to look at the screen and then back at the viewer. The presenter's arm movement is determined by pointing actions specified in the animation script and by the affirmation level (neutral, warm, or enthusiastic). For hand movement, Noma, Zhao, and Badler used canned animations for grasping, indicating, pointing, and reaching.

Elisabeth André, Thomas Rist, and Jochen Müller described a cartoon-style Web agent that presents information on Web pages.[3] Most of their research focuses on planning the presentation script. The agent's character is displayed in 2D, using completely predefined animations. The agent performs pointing gestures such as direct pointing, underlining, and circling, using a pointing stick that's drawn from the hand to the target. The agent can also express emotions such as anger or tiredness. To span pauses, the agent performs idle animations such as breathing or thumb twiddling.

References

1. H. Prendinger, S. Descamps, and M. Ishizuka, "MPML: A Markup Language for Controlling the Behavior of Life-Like Characters," *J. Visual Languages and Computing*, vol. 15, no. 2, 2004, pp. 183–203.
2. T. Noma, L. Zhao, and N.I. Badler, "Design of a Virtual Human Presenter," *IEEE Computer Graphics and Applications*, vol. 20, no. 4, 2000, pp. 79–85.
3. E. André, T. Rist, and J. Müller, "Webpersona: A Life-Like Presentation Agent for the World-Wide Web," *Knowledge-Based Systems*, vol. 11, no. 1, 1998, pp. 25–36.

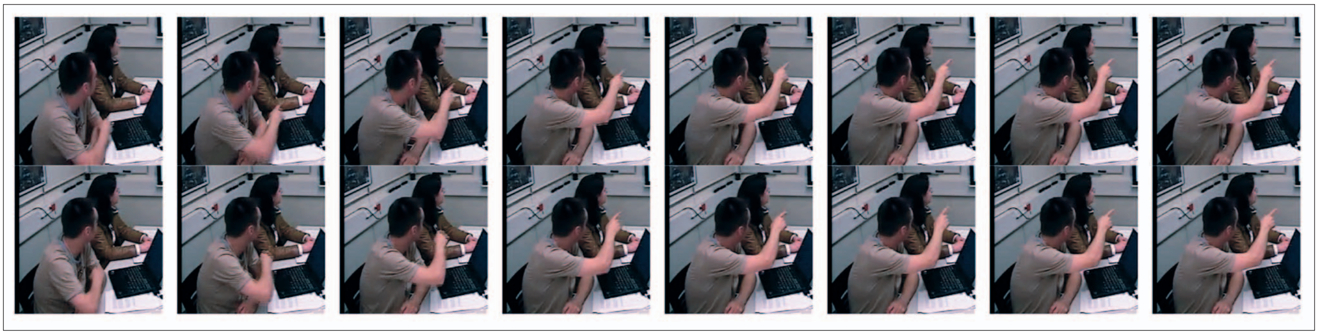


Figure 5. The preparation phase (upper half) and retraction phase (lower half, in reverse) of a pointing action, captured from video.

was less stiff. We didn't tell them about the involuntary movement beforehand.

Ten subjects noticed the involuntary movement immediately; nine had to watch the presentations twice to notice the difference. One subject didn't notice the difference between the presentations until we pointed it out. Eighteen subjects thought the presenter with involuntary movement moved more naturally. The test subjects all agreed that the presenter with involuntary movement was less stiff. Taking into account that we choose both the joints on which the noise is created and the amount of noise ad hoc, we can conclude that creating involuntary movement in such a way looks promising.

We plan to evaluate other modalities or parameters of those modalities by comparing them to real human behavior. To do this, we will directly display a recording of a real presentation on our avatar. Such a recording could consist of motion capture of the real presenter's movements and recorded audio of the presenter's voice. In this recording we would then replace the target modalities or parameters with our own models. For example, we could replace the recorded speech with synthesized speech or replace the recorded speech's rhythm with one that our speech rhythm model generated. We could also replace modalities or parameters with "null" or random models. (A null model is one that does "nothing." For example, a null model for eye blinking will not blink at all.) One person's modalities or parameters could replace those of another. However, conflicts might arise if the two persons' styles don't match.

Design of future evaluation

We would tell test subjects that the avatar's behavior could be either a recording of a real human or partly machine generated. We would then show short presentation segments and ask the subjects to judge whether the behavior was human-like or (partly) machine-like. This way, we could implicitly judge the naturalness of the avatar's behavior. The segments could show real human behavior for different humans, behavior using a theoretically sound model, behavior using a null model, behavior using a random model, or behavior using parameters from another human.

With this test, we could obtain information on how much certain modalities and parameters contribute to make behavior look natural, by comparing the real human behavior with random and null models. We could obtain information on a model's quality by comparing it with other models or real human behavior. By replacing one human's parameters with those of another, we can infer which modalities and parameters are style dependent.

Modalities to evaluate

For our first tests, it would be convenient to use modalities, or at least recorded parts of those modalities, that aren't heavily influenced by outside factors. Such factors include preceding or following behavior, or behavior on other modalities that could influence the behavior's shape (for example, the combination of an iconic gesture and a deictic gesture). This way, we could isolate a single modality's execution and observe only its own parameters. We could then later create and test separate models that combine different modalities and concatenate behaviors on a modality.

Our method of varying leading modalities is more flexible than traditional presenting systems that use speech to guide expressions on all other modalities.

As a start, we plan to execute the user test on beat gestures, because beats are the most commonly used gestures[5] and because they can be isolated more easily than other gestures. Modifiable parameters for beats include

- the beat space,
- the hand shape,
- pre- and post-stroke hold selection and duration,
- the velocity function,
- the movement path's shape and length, and
- the beat modality (left hand, right hand, or head).

A second test candidate is speech. We've already created models to replay recorded speech and extract and modify several parameters of it, including the rhythm of the phonemes, the melody, and the pitch.

Moving to a different domain: The virtual museum guide

To demonstrate the broader applicability of our virtual-presenter technology, we're investigating a different domain—a virtual museum guide. A corpus of annotated paintings such as the Rijksmuseum database used by Arnold Smeulders and his colleagues[14] shares many characteristics with the presentation sheets. The information about a painting covers general aspects as well as remarks about specific subareas of the painting (for example, relative composition and details in a corner of the painting).

Multimedia presentations using the content of the Rijksmuseum database can be generated automatically. We could easily make such a presentation interactive by using the virtual presenter as a museum guide who talks about the paintings while pointing out interesting details. Where the text only implicitly encodes the relations between the text and areas in the paintings, we could develop techniques to automatically extract those relations from the text.

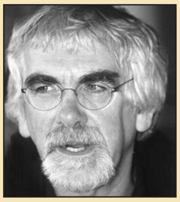
Our method of varying leading modalities is more flexible than traditional presenting systems that use speech to guide expressions on all other modalities. For example, our presenter could comment on a video she's showing by synchronizing her speech and gestures with the video's timing. If speech is the chosen leading modality, our presenter is still compatible with traditional presenting systems. Currently, we achieve synchronization of speech and gestures on word boundaries. We could achieve tighter and possibly more varying timing by identifying the phonological peak of words and using that to time gesture strokes.

Two small research projects have already adapted our architecture. One project is developing a virtual guide that gives people directions in a building. This project has added a channel for iconic gestures to the presenter. We plan to evaluate this virtual guide by comparing her multimodal way of giving directions with a written or spoken route. The other project is looking at letting the audience interrupt the presenter.

Further work could broaden the presenter's abilities to express herself. We could do this by adding additional gesture types (such as beats, iconic gestures, or metaphoric gestures). We could also raise the virtual-presenting process to a higher abstraction level. Currently, the script determines what parts of the presentation to express in speech or in gestures. The next logical abstraction step would be to implement a process that determines what to say and what gestures to make on the basis of what the presenter wants to tell. The presenter's style and emotional state could guide this selection.

The Authors

Herwin van Welbergen is a PhD student in the Human Media Interaction group of the University of Twente's Department of Computer Science. He's working on the parameterization of animation and multimodal planning. His research interests include virtual reality and graphics, animation, multimodal interaction, and virtual tutoring. He received his MSc in human media interaction from the University of Twente. This article describes his master's thesis. Contact him at the Univ. of Twente, Human Media Interaction, PO Box 217, 7500 AE Enschede, Netherlands; h.vanwelbergen@alumnus.utwente.nl.



Anton Nijholt is a full professor and the chair of the Human Media Interaction group of the University of Twente's Department of Computer Science. His main research interests are multiparty and multimodal interaction, virtual environments, and social and intelligent (embodied) agents. He received his PhD from Vrije Universiteit Amsterdam. Contact him at the Univ. of Twente, Human Media Interaction, PO Box 217, 7500 AE Enschede, Netherlands; anijholt@cs.utwente.nl.

Dennis Reidsma is a PhD student in the Human Media Interaction group of the University of Twente's Department of Computer Science. His activities focus on the problems and issues that arise in creating large, multiply annotated corpora, such as developing annotation schemes and tools and investigating annotation reliability. He received his MSc in human media interaction from the University of Twente. Contact him at the Univ. of Twente, Human Media Interaction, PO Box 217, 7500 AE Enschede, Netherlands; dennisr@ewi.utwente.nl.

Job Zwiers is an associate professor in the Human Media Interaction group of the University of Twente's Department of Computer Science. His research interests are human-computer interaction, computer graphics, and multiagent systems. He received his PhD in computer science from the University of Delft. Contact him at the Univ. of Twente, Human Media Interaction, PO Box 217, 7500 AE Enschede, Netherlands; zwiers@cs.utwente.nl.

References

1. D. Reidsma et al., "Virtual Meeting Rooms: From Observation to Simulation," *Proc. Social Intelligence Design* (SID 05), CD-ROM, Stanford Univ. Press, 2005; http://wwwhome.cs.utwente.nl/~rienks/documenten/sid2005.pdf.
2. J. Gratch et al., "Creating Interactive Virtual Humans: Some Assembly Required," *IEEE Intelligent Systems*, vol. 17, no. 4, 2002, pp. 54–63.
3. J. Cassell, H.H. Vilhjálmsson, and T. Bickmore, "BEAT: The Behavior Expression Animation Toolkit," *Proc. 28th Ann. Conf. Computer Graphics and Interactive Techniques* (SIGGRAPH 01), ACM Press, 2001, pp. 477–486.
4. D.M. Chi et al., "The EMOTE Model for Effort and Shape," *Proc. 27th Ann. Conf. Computer Graphics and Interactive Techniques* (SIGGRAPH 00), ACM Press, 2000, pp. 173–182.
5. D. McNeill, *Hand and Mind: What Gestures Reveal about Thought*, Univ. of Chicago Press, 1995.
6. A. Nijholt, H. van Welbergen, and J. Zwiers, "Introducing an Embodied Virtual Presenter Agent in a Virtual Meeting Room," *Proc. IASTED Int'l Conf. Artificial Intelligence and Applications*, IASTED/ACTA Press, 2005, pp. 579–584.
7. Z. Ruttkay and C. Pelachaud, "Excercises of Style for Virtual Humans," *Proc. Animating Expressive Characters for Social Interaction Symp.*, 2002, pp. 85–90; www.aisb.org.uk/publications/proceedings/aisb02/AISB02_ExpressiveCharacters.pdf.
8. K. Perlin, "Real Time Responsive Animation with Personality," *IEEE Trans. Visualization and Computer Graphics*, vol. 1, no. 1, 1995, pp. 5–15.
9. I.S. MacKenzie and W. Buxton, "Extending Fitts' Law to Two-Dimensional Tasks," *Proc. SIGCHI Conf. Human Factors in Computing Systems*, ACM Press, 1992, pp. 219–226.
10. A. Kendon, "An Agenda for Gesture Studies," *Semiotic Rev. of Books*, vol. 7, no. 3, 1996, pp. 8–12.
11. X. Zhang and D. Chaffin, "The Effects of Speed Variation on Joint Kinematics during Multisegment Reaching Movements," *Human Movement Science*, vol. 18, no. 6, 1999, pp. 741–757.
12. D. Tweed, "Three-Dimensional Model of the Human Eye-Head Saccadic System," *J. Neurophysiology*, vol. 77, no. 2, 1997, pp. 654–666.
13. D. Tolani, A. Goswami, and N.I. Badler, "Real-Time Inverse Kinematics Techniques for Anthropomorphic Limbs," *Graphical Models and Image Processing*, vol. 62, no. 5, 2000, pp. 353–388.
14. A. Smeulders et al., "An Integrated Multimedia Approach to Cultural Heritage E-Documents," *Proc. 4th Int'l Workshop Multimedia Information Retrieval*, CD-ROM, ACM Press, 2002.