Cursive: A novel interaction technique for controlling expressive avatar gesture

Francesca Barrientos
John Canny
Computer Science Division
University of California, Berkeley
Berkeley, CA 94720, USA
{fbarr,jfc}@cs.berkeley.edu

ABSTRACT

We are developing an interaction technique for rich nonverbal communication through an avatar. By writing a single letter on a pen tablet device, a user can express their ideas or intentions, nonverbally, using their avatar body. Our system solves the difficult problem of controlling the movements of a highly articulated, 3D avatar model using a common input device within the context of an office environment. We believe that *writing* is a richly expressive and natural means for controlling expressive avatar gesture.

KEYWORDS: Avatars, computer-mediated communication, virtual environments, gesture, novel interaction technique, pen gesture, nonverbal.

INTRODUCTION

Anything is possible in virtual worlds, yet they contain real world elements such as buildings, pavilions and human looking avatars. These elements reflect the human desire to bring to digital worlds the richness of expression possible in *real life*. In face-to-face encounters, we use gestures as well as words to communicate. Avatars, as surrogate bodies, have the *potential* to be a rich, expressive gestural communication channel. We describe a simple interface that supports discrete (symbolic) and continuous (expressive) gesture control at the same time.

A central difficulty in producing gesture is that humanoid avatars have many joints whose motions need to be coordinated all at once. Though capture (e.g. using computer vision) and reproduction of a user's own body motion in the avatar seems the most direct solution, it is not possible with standard hardware and software (yet). But we argue that even when it is, the social context in the user's physical world is radically different from the avatar's, and that this will make it inappropriate in many contexts to use this kind of input. If cell phone conversations in public spaces are annoying, imagine watching one side of a full-body conversation between recalcitrant family members.

Gesture is a complex medium. Speech has both symbolic (words) and a *prosodic* (pitch, volume, speed.) components. Gesture has symbolic and "prosodic" components too, but the continuous component is much more important. A person can wave "hello" intimately using just finger tips, or exuberantly using a whole arm that sways the body with its energetic motion. Our goal is a single interface that allows the user discrete and continuous control, and is natural enough that they can use the interface as an alternative channel while talking to another person.

FROM PEN GESTURES TO AVATAR GESTURES

We have developed an interface that we believe meets these goals. The idea is to use *gesture* as input, but pen gesture rather than body gesture. A user writes a single character on a pen tablet; the character is recognized using a standard recognition toolkit, and, at the same time, attributes such as speed, size, slant and pressure are computed. Then the gesture symbol plus these style parameters are sent across the network to any copies of the avatar which need to be animated as illustrated in figure 1.

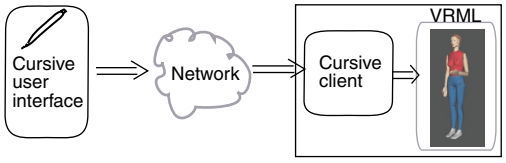

Figure 1. Cursive takes user input and sends animation data to copies of the user's avatar in virtual world browsers.

The pen gesture indexes a pre-recorded avatar gesture, which is played back and modulated using the style parameters to produce a unique, expressive body gesture. Each prerecorded body gesture is captured offline using a motion tracking system, in our case, an Ascension Flock of Birds array. In fact each body gesture is recorded many times with different styles (fast, slow, small, large etc.) and then the style parameters are used to interpolate between those reference gestures to produce the new gesture.

The advantages of this approach are:

- Discrete and continuous input derived from a single action;
- People easily write while talking and often produce expressive writing spontaneously (e.g. doodling while on the phone);
- No mode switches between gestures or, between discrete and continuous control; and
- Pen movement is minimal.

Our first testbeds will be 3D avatar worlds such as Parallel Graphics *Island Worlds* [1]. However, this technique could easily be adapted for use in other applications such as informal "sketching" of animation sequences, recording or designing choreography, or digital puppetry. A simple interface such as this may enable new ass-market uses of animation, e.g. the use of avatar animations in multimedia email.

AVATAR GESTURE PRODUCTION

When a user logs into a 3D virtual world, they view the world and any avatars in it using a browser such as *Cortona*. The browser provides an interface for navigating the avatar around the world. The user interacts with Cursive in a Java application next to the browser, using a pen tablet. Ink from the tablet is passed to a recognizer which computes the character identity and style parameters. The character identity plus the style parameters give the gesture encoding as in Figures 2.

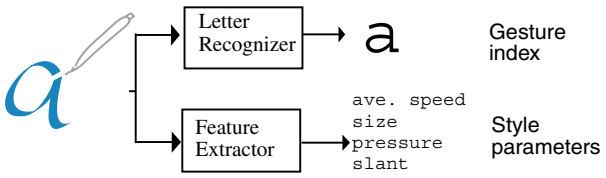

Figure 2. The character identity plus the style parameters give the gesture animation encoding.

This encoding is sent across the network directly (not via the Vworld server) to any browsers that are displaying copies of the user's avatar. An avatar is a VRML file that includes javascript nodes for animation. Code in these script nodes receives the gesture encoding via a socket connection to the Cursive application on the machine that generated the pen gesture. Because Java scripting is supported in several commercial Avatar world browsers, our approach has an elegant consequence: We can "drop" expressively animated avatars into existing avatar worlds!

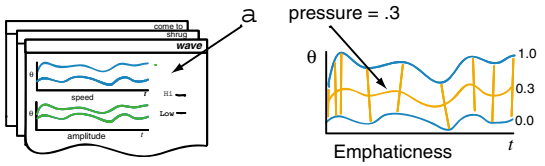

Figure 3. The character selects the gesture while the style parameters modulate the animation.

The identity of the pen gesture (which character was drawn) is used to select a particular body gesture from a (preloaded) file. A body gesture is a set of trajectories for all the joints of the avatar. We can denote such a trajectory *Jx(t)*, for the pen character *x*, where we understand that *J* is a vector with *k* components, where *k* is the number of joints of the avatar. When the gesture encoding is received, a clock starts in the avatar script that triggers playback of the gestures. This simplified story leaves out style modulation. The style parameters from the pen gesture are mapped to meaningful parameters for the body gesture, such as that shown in the table below. (This mapping need not be the same for all gestures.) The

pen gesture feature	avatar gesture parameter
bounding box area	size
average speed	speed
pressure	emphaticness
slant	direction

body gesture parameters are used to interpolate between a family of trajectories for each gesture. That is, the gesture trajectory *Jx(t)* is really a family of trajectories *Jx0000(t)* through *Jx1111(t)*. Trajectory *Jx0101(t)* has style parameters (0,1,0,1) (assuming there are four style parameters as there are now). A trajectory with parameters *(a,b,c,d)* is obtained from the 16 prototypes by multilinear interpolation. The speed parameter also requires scaling of the time parameter and is handled specially. This selection and modulation process is illustrated in figure 3.

GESTURE LIBRARY CREATION

The avatar gesture for character *x* is a family of trajectories *Jx0000(t)* through *Jx1111(t)*. These files must be created offline for all the gestures the user plans to use. Each trajectory is created by recording a person's motions while they are instrumented with position/orientation sensors on their limbs and body. For four style parameters, there are 16 combinations of extremal trajectories. Each one of these trajectories must be performed and recorded. The real meaning of each style parameter will be affected by the performer's understanding of the mapping of that parameter, e.g. pen pressure may be coded as "excitement" or "anxiety". Ideally, the performer and the online user of the avatar would be the same person. That will not always be possible, but it will be useful to have several libraries available from different performers so that a user can find one that best matches the gestures they would like to use.

RELATED WORK

ComicChat [3] is a two-dimensional chat world in which avatars appear as comic strip characters. Users select emotions using an emotion wheel. Different directions around the wheel select different expressions, and the distance from the center selects the magnitude of the expression. The non-verbal communication interface to the *VLNET* system [2] provides users with buttons to select an avatar gesture, and a separate slider to control the speed of the animation. *BodyChat* [4] treats avatars as semi-autonomous communicative agents. Users specify high level conversation goals, and the avatar agents automatically handle the gestures required to negotiate the conversation.

None of these systems provides an interface which allows a user to specify a body gesture along with multiple style performance parameters using just a single input action.

CONCLUSIONS

We have devised and implemented a novel interaction scheme for controlling avatar animation for nonverbal communication in a virtual environment. At this time we use letters because they are a natural pen gesture for most people. In the future we would like to study the use of other forms of pen gesturing to control avatar expression.

REFERENCES

1. www.parallelgraphics.com
2. Guye-Vuillème, A., Capin, T., Pandzic, I., Magnenat-Thalmann, N. and Thalmann, D. Non-Verbal Communication Interface for Collaborative Virtual Environments, in Proc. CVE 98, Manchester, 1998.
3. Kurlander, D., Skelly, T. and Salesin, D. Comic Chat. in SIGGRAPH 1996.
4. Vilhjálmsson, H.H. and Cassell, J. BodyChat: Autonomous Communicative Behaviors in Avatars. in Proc. of the 2nd Intl. Conf. on Autonomous Agents, 1998.