3 Interaction technique

We have developed an interaction technique for controlling communicative avatar gesture using a pen interface. The technique applies to avatars with articulated bodies and animated behaviors. It is designed to provide fine grained control over multiple expressive qualities of movement in gesture. In contrast, existing techniques either provide no controls for gesture modulation or offer a control over a single movement variable. We argue that the pen interface provides a natural means for controlling expressive, spontaneous avatar gesture.

3.1 Expressive movement in gesture

Part of the communicative power of gesture comes from the ability to alter the magnitude, color or emphasis of the message by changing the expressive movement of the gesture. While some features of a particular gesture remain invariant—the features that make a gesture. recognizable as, say, a wave and not a shrug—other features of movement can be changed without impairing the identifiability of a gesture. A wave of "hello" can be larger or smaller, last a fraction of a second or several minutes, oscillate quickly or sway slowly, etc. The movement variations are multi-dimensional in nature and infinite in number because the features vary continuously. We call these features the *expressive* qualities of a gesture.

Another aspect of gesture that is important to its communicative power is its capacity to be combined with speech to create a composite message. Gesture can accompany speech because it is received through a different modality—sight rather than sound—and is produced through the movement of parts of the body not involved in the production of speech.

In an avatar communication system, the ability to produce coverbal gesture requires a similar separation of the modalities of reception and production. Coverbal gesture accompanies spoken utterance and shares a concept with a spoken word or phrase. The gesture itself appears with or slightly precedes the accompanying words, so it is imperative that the avatar controls do not interfere, either mechanically or cognitively, with speech production.

Spontaneity is closely tied to the expressive aspects of gesture. It refers to the ability of the sender to intentionally vary performance of a gesture. This is different from the fact that the gesture itself can be varied. Spontaneity results from the user's ability to modulate the gestural movement at the instant it is performed.

Finally, expressive power comes from the feel of producing the expressive movement. In our own bodies, humans experience the emotions that give rise to outward expression, and the generation of the expression itself. The experience comes from the kinesthetic sense of the body's motion.

3.2 Limitations of existing techniques

This section discusses the limitations in previous techniques.

3.2.1 Narrow range of behaviors

Most virtual environments focus on only a small range of NVC behaviors, featuring affective displays and emblementaic gestures. These displays share the property that they are independent of speech. This simplifies the design of controls for these displays. Since many

virtual environments are chat applications, independence from speech is an important factor; typing and pointing (with the mouse) cannot be performed simultaneously. Affective displays—facial expressions or postures expressing emotional or mental state—are particularly tractable since they do not need to change often during a conversation.

Even when speech is available, coverbal gesture remains a problem. The gesture must coincide with or immediately precede the word or phrase it shares meaning with. Additionally, it occurs with a low level of awareness on the part of the speaker, and so appears to be inconsistent with the demands of a desktop user interface. In contrast, emblematic gesture is deemed controllable because it is produced with a high degree of intentionality. This level of awareness maps well to the selection task required by graphical user interface GUI elements such as pull-down menus and buttons.

3.2.2 Minimal movement modulation

Control of continuous expressive qualities is very limited in previous solutions if it is addressed at all. The NVC interface in [45] presents a panel of buttons, each of which maps to a different expression or gesture. Before invoking a nonverbal cue by selecting a button, the user can set the animation speed of the gesture by using a separate slider. Selection an modulation here is a two step process. In ComicChat [63], the user simultaneously chooses an avatar expression and specifies its magnitude by clicking within a circle. The invocation and modulation occurs within a single user action.

The style modulation controls in these applications do not scale well to multiple features. To add more style features, either the interface must add more controls— such as in the case of the slider—or the control must have more dimensions. These traditional GUI interfaces are fundamentally limited in their expressive power because of the two-dimensional

nature of their controls. These interfaces depend on the user visually attending to the size of the control and then selecting a point within it. Adding more controls or more complex controls would require the user to spend more time assessing and manipulating visual information. The more attention spent on the avatar control task, the less the user can attend to the overall communication task.

A possible solution to controlling movement quality is to map the position of a user interface element to the position of the limb and thereby directly control gesture movement. Most systems provide this type of control to a limited degree; they allow the user to turn the avatar's head either through the mouse or through the keyboard. In general the motion is restricted to turning the head side to side or up and down. With just these two degrees of freedom, the user is able to shake or nod the head, and change gaze direction. In [99] the user directly controls the avatar's arms using a slider for each arm joint. The avatars have single segment arms that rotate at the shoulder in a plane parallel to the sagittal plane.[1] The user can vary the speed and magnitude of the arm's rise and fall using these sliders. Avatar models with more joints and joints with multiple rotation axes would require a greater number of user interface elements.

3.2.3 Restriction to standard UI elements

The developers of the BodyChat system argue that manual control should be provided at a higher level than the level of triggering a particular gesture [110]. The user need only be aware of their communication goals and not the gestures that will communicate their goals. In the BodyChat environment, the user indicates interactional goals such as the desire to begin a conversation and with whom. A study of the efficacy of this system suggests that conversation

[1]The plane that divides left from right.

using this system is perceived as more natural than when using an interface in which users more directly control the avatar's gesture [20].

This study is limited in its applicability to the problem of avatar gesture control in general. BodyChat is a chat environment and the user cannot both type their conversation and control the avatar at the same time. In an environment in which speech carries the verbal content of the conversation, the user's hands are free to control the avatar. So, their results might not apply to a speech-based environment.

Another limitation with the study is that the researchers compared the BodyChat system to a manual interface built of pulldown menus. User's found this interface distracting. Autonomous behaviors are not the only answer to this problem. Other interfaces that require little or no visual attention may prove as natural. Autonomous behaviors may be most appropriate for a narrow range of communicative behavior, in particular those behaviors that are difficult or impossible to place under conscious control. This still leaves open the question of whether other types of manual control would seem natural if they required little attention on the part of the user.

3.2.4 Consequences

Previous solutions to the avatar gesture control problem limit their designs to the use of two-dimensional interfaces. These controls are best suited to only a narrow range of NVC behaviors and those behaviors are given only a limited range of variation. One of the consequences of relying on simple interfaces, according to [111], is that the cues for social interaction accountability, are "underdetermined." The "narrowing" of avatar gestures results in an over dependence on appearance. The rapid fire display of amusing gestures is used to impress people but is not extended to its full potential as a communicative channel.

3.3 Description of interaction

The main idea behind our interaction is to use gesture as *input*, but we use pen gesture rather than body gesture. Our technique uses the identity of the pen gesture to select a particular action, the same as in most other pen user interfaces [53][69]. In addition to extracting this symbolic information, the technique uses the writing style to extract expressive content from the pen gesture.

3.3.1 Avatar model

The technique is designed for use with a three-dimensional, multi-segmented, articulated avatar. For our testbed, we work with VRML models since they are readily available on the internet. An example of a VRML[2] avatar is shown in Figure 3-1 below. This avatar con-

Figure 3-1. "Nancy," a VRML avatar designed by Nancy Ballreich.

forms to the *h-anim*[3] specification, a standard for modeling VRML humanoid avatars. This

[2]See online reference in Appendix A.

[3]Online reference in Appendix A.

specification provides a standardized joint hierarchy and naming convention for the model's segments and joints.

3.3.2 Avatar gestures

The user can invoke a number of different gestures. By *different* we mean that the gestures are recognizably different. For example, such gestures might include a bow, a shrug and a wave. The user also specifies different expressions for each gesture. That is, the user can modulate the expressive qualities of the gesture's movement.

We define a multi-dimensional continuous space of expressive variation for the gesture movement. The dimensions include, but are not limited to, speed, size and emphaticness. Unlike previous work, in which a user specified only a single modulation parameter, our controls allow simultaneous specification of multiple parameters.

3.3.3 Avatar gesture library

The avatar gesture animation data is stored offline and consists of motion capture samples of a person performing the gesture. Several prototypes of each gesture are recorded, and each prototype exhibits an extremal variation of the gesture movement. If each of the parameters can vary from 0 to 1, then each prototype is a variation in which the parameters have the value the value 0 or 1. Taken together, the set of prototype gestures span the space of possible expressive variations. A gesture library with three different gestures varying in three dimensions is shown in Figure 3-2. Note that there are $2^3 = 8$ samples of each gesture to cover both extremities of the three dimensions.

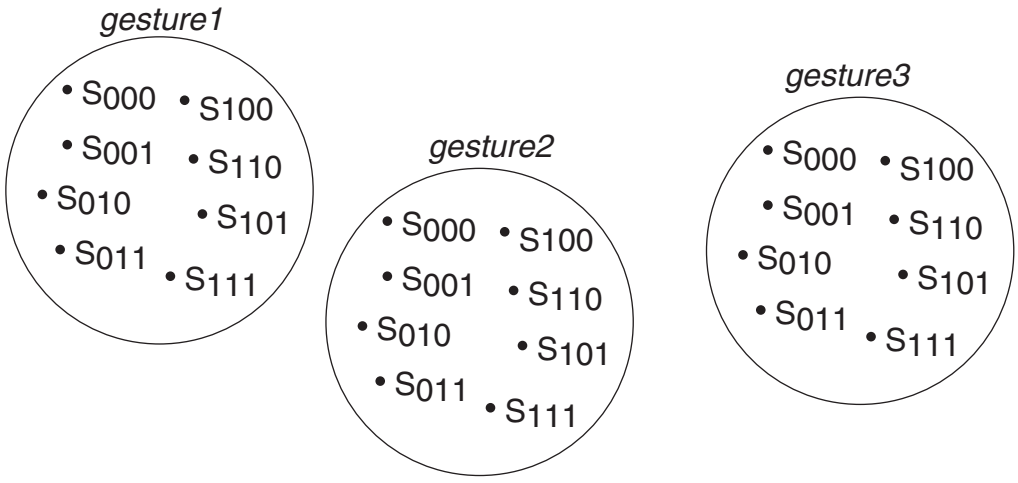

Figure 3-2. Each gesture consists of several motion capture samples.

3.3.4 User input

The user controls the avatar gesture by a writing pen gesture[4]. Unlike traditional graphical user interfaces, using pen gesture does not require visual search and selection. The gestures are written into a designated window on the desktop. This input requires pen input hardware, such as a graphics tablet. This hardware is common on handheld computing devices, and though not standard, is a common pointing input device for desktop computers.

The pen gesture set consists of letters of the alphabet. Writing a letter on the tablet invokes the gesture animation. The letter identity selects the particular avatar gesture. The way the letter is written modulates the expressive variation of the avatar gesture movement. For instance, writing the letter bigger results in a bigger avatar gesture.

[4]Because the word *gesture* can refer to both pen and avatar gesture, we will use the phrases "pen gesture" and "avatar gesture" to distinguish between the two.

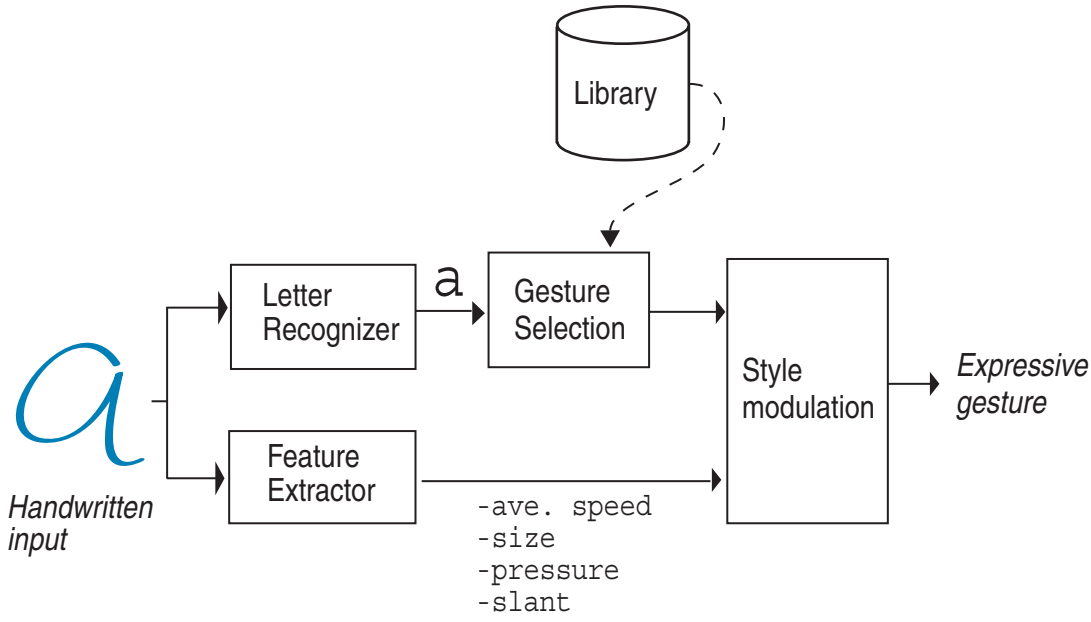

Figure 3-3. Production of avatar gesture from pen gesture input.

3.3.5 Selection and modulation of gesture

When the letter is written, the digital ink gets analyzed by two different modules. The first is a standard character recognizer. The second module computes a set of handwriting style features on the pen gesture. These continuously valued features include the speed of the writing, the size of the letter and the pressure of the pen tip on the writing surface.

The identity of the letter selects an avatar gesture type from the avatar gesture library. Values from the handwriting style features are mapped to values for the avatar gesture expression parameters. The gesture samples in the library are blended together according to the values of the parameters to produce an expressive gesture instance. This instance is sent on to the avatar to be animated. The sequence is shown in Figure 3-3. The details of this algorithm are covered in *Chapter 4, From pen gesture to expressive avatar gesture.*

3.3.6 Auxiliary features

Additional features would complement our interaction technique well. The features described in the rest of this section are not implemented, however we feel they would add greatly to the functionality and usability of a system using a pen-based interaction technique.

3.3.6.1 Direct control mode

In the basic sequence of events described above, the pen gesture identity indexes into the avatar gesture library. Other pen gestures are recognized as *escape symbols*. The escape symbols change the state of the system so that subsequent pen movements are not interpreted as pen gestures. Instead, the movements of the pen directly control the motion of the avatar's limbs.

Direct control mode accommodates production of singular gestures. These gestures are created in the moment to accompany a particular utterance or situation. For instance, an iconic gesture might be used to depict the shape of a physical object under discussion. The direct control mode is particularly suited to deictic gestures which refer to specific objects in the environment or points in space.

3.3.6.2 Reversibility

The design includes a special symbol, a back slash, for undoing a control command. This allows a user to recover when their pen gesture is misrecognized or they wish to abort the command for some other reason.

3.4 Design issues

This section addresses design issues involved in choosing a pen gesture input set and the avtar gesture vocabulary.

3.4.1 Choice of pen gesture set

Alphabetic characters are a natural starting point for our interaction technique. Unlike other artificial pen gestures, letters do not need to be specially learned. This removes one barrier to learning the interaction. Writing letters with different styles is already a natural practice for most individuals, especially for simple features such as size, speed and pressure. Letters also provide an adequate size for the pen gesture set. Future work would be to determine if other pen gestures or even words would work better for controlling avatar gesture.

3.4.2 Avatar gesture vocabulary

The library should include both a standard set of gestures and a set personalized for each user. The standard set is the default library provided with the system. Users can learn the interaction technique using these gestures. Later as they become expert users they will likely want to extend the library by adding a personalized set of gestures. The trend towards personalization is seen in other aspects of avatar design. Virtual environment visitors usually begin by using a generic avatar model. Later, as they become more sophisticated users, they design their own avatars.

The standard set can be drawn partly from emblematic gestures from a particular culture. [56] presents a survey of emblematic gesture in North American culture. Nonlinguistic forms of gesture are also culturally specific and have been characterized by communication researchers [36][39]. The simplest and most common of these might be used. Beat gestures are a good candidate since their timing and not their exact form is their important aspect. Standard sets can also be designed for a specific domain. Studies describe gestures that are particular to discussions of design [105] and math [75]. Finally, gestures can be drawn from individuals, either through their specification or through analysis of them in videotaped conversations.

3.5 Discussion

3.5.1 Pen gesture as affective input

Many interfaces define emotional states for the user to select. This requires the user to self-assess their affective state and then match it to one of the states presented in the interface. However, emotions are not always experienced categorically [1][94]. Control through handwriting specifies affective state directly without requiring a human recognition phase. This form of control maps naturally to the experience of emotions along a continuum.

Extracting emotional state from handwriting also allows users to specify affective state with less intentionality. Emotion is naturally and often unintentionally evinced in handwriting. This technique seamlessly uses this input without extra user effort.

3.5.2 Personality

Another interesting aspect of this technique is that the avatar motion can reflect the user's personality. Personality is made up of the combination of qualities that differentiate one person from another. These qualities include handwriting style. The aspects of writing that remain consistent for a particular person reflect personality. Certain features of writing style are beyond conscious control. This is the basis of forensic graphology which identifies individuals through invariants in their writing [80]. So, handwritten input may produce highly personalized avatar gesture, even when different users work with the same gesture prototypes.

The way people move is a function of the goals of the motion, the person's physiology, the environmental conditions, etc. However, there are aspects of movement that distinguish one person from another, and these might be called "expressive." Research in expressive movement has confirmed what humans intuitively discern—that people can be identified from the way they move. Through a series of experiments Allport and Vernon found that individuals

showed a high degree of internal consistency in the way they performed a variety of motions from walking to shaking hands [3]. This self-consistency extends to the way people write. There is also rhythmic coordination between speech and gesture because they are occurring in the same body [24].

The flip side of expressiveness is the ability to intentionally express one's self through motion. In order for this to work, different people have to share an understanding of the meaning of different writing characteristics. Wolff conducted an experiment in writing letters which indicates that there exists common conventions for expressing emotions through letters [114]. This result suggests that though handwriting is very individual, people are able to modify their writing in recognizable ways to convey emotion.

3.5.3 Continuous interfaces and transparency

In his book Abstracting Craft [74], Malcolm McCullough asserts that "computers fragment our thinking by substituting discrete events for continuous actions." A computer interface that divides emotions and actions into discrete categories so that they can be mapped to buttons and menu items breaks the production of personal expression. Then simple conversation requires multi-tasking between visually selecting appropriate nonverbal communication and generating verbal utterance. We developed this interaction technique with the intent of maintaining some of the feel of the continuous action of gesture.

Part of the personal experience of gesturing is the kinesthetic feel of forcing air through the larynx and moving limbs to gesture. We chose pen gesture because the pen movement is the action. The form of the movement is stored in muscle memory. By way of contrast, GUI actions require visual recognition and hand-eye coordination. Pen gestures are closer in experience to other body gestures.

This consonance with communicative gesture makes pen gesture a more natural form of interaction than the hunt and select form of GUI actions. Like other movements of the body, writing is affected by emotional state. The expressive movement in handwriting can be and often is controlled intentionally. Using a pen to doodle while conversing is a natural act.

Using the pen gestures for avatar control is natural in the sense that using any tool in a skilled manner is a natural act. Acquiring the skill takes practice. Our avatar gestures exhibit a continuous range of variation in movement style, and the richness in output is complemented by fine control of the input. This rich interface suggests that manual dexterity should be brought into play. The user must have a feeling for what the possibilities are and a sense that they can acquire the manual skills necessary to achieve any of the possibilities.

Using manual dexterity suggests that the hands need to be manipulating some kind of medium, something that can be worked, through continuous action, into an expression of the person's intent. McCullough defines a medium as something that offers a continuum of possibilities. It can be formed into different states with "continuous hand guided processes, like coaxing a metal (p. 199)". In addition, a medium also offers constraints. These limits define how the medium can be manipulated. In our case, the avatar system is the medium. The interface presents possibilities for possible outcomes, the possible gesture variations; and also limitations, the implicit rules for possible actions within the interface. McCullough suggests that it is the continuity as well as richness of a medium that supports engagement.

More than any motivation on the part of the avatar user, the design of the application is important to the learnability of the pen gesture skill. The user interface must be engaging. Laurel, likens "engagement" to the suspension of belief in theater when the audience willingly sees the actors on a stage as characters in a scene [67]. In a user interface, engagement happens

when the user willingly understands their task only within the context of the actions and representations provided by the user interface. The connection between expressive movement in the pen gesture input and in the avatar gesture output facilitates the emergence of engagement.

Engagement then permits playfulness. Playfulness is especially relevant to the skill of producing communicative gesture. Laurel describes the learning of effective gesture as follows: "...the art of finding and executing an effective gesture is learned through the…indirect means of observation, experimentation, performance, and evaluation, and it is a skill that continues to grow over time (p. 155)." Communication applications are particularly suited for play. The consequence of any gestural action is fleeting. Misinterpretations can be cleared up through verbal explanation. Repetition and correction is simple. Use of the expressive characteristics of gestures is forgiving. Creating a bigger or smaller gesture is only a matter of degree. And there is no wrong input. The movement need not be exact. An interface that allows play will enable a user to become more skilled at and expressive in communicating with their avatar through exploration of expressive possibilities. The user is able to experiment with different action and learn from observing the different outcomes.

Though the avatar range of motion will be limited, users should be able to become more facile with using them as they would with any tool. Again, using an argument in [74], as a person becomes familiar with the possibilities of the avatar body, they will begin to push use of the interface into the subconscious and be aware only of the intent of how they want the avatar to move. It is the continuous nature of the interaction that allows the use of the tool to become *transparent*.

As the interface becomes natural to the avatar user, we hope that it will become integrated into the process of communication in the way that unmediated gesture is. We know that

the synthesis of gesture along with the coding of verbal language helps the speaker to constitute thought [31][41][60], and that a person's gesture can feed back to their own perceptions [48]. Our hope is that the connection between the creation of utterance and gesture can be reestablished through this form of interaction.

3.5.4 Limitations

This method suffers from a problem inherent in pen UI's, the user interface requires learning. This technique suits expert users, but is less spontaneous for casual users. Users are required to learn the index from letters to gesture forms and the map from handwriting style to gesture movement characteristics. Controlling the expressive characteristics is a skill developed through practice and by monitoring feedback. When the indexing skills are learned they will become a part of muscle memory.

Letters as linguistic entities may have particular associations for users. The linguistic connotations may interfere with learning the index into the gesture set. Further study will determine if undesirable interactions arise. Such studies may also point to ways in which the user's associations can aid usability of the technique. For instance, some mappings between particular letters and body gestures may seem more natural than others.

Inherent in these mediated environments is the lack of proprioceptive feedback where in the user is able to sense the position of their limb in the virtual space. This technique may have less of a problem because continuous input is mapped to continuos output. Usually this problem is addressed through appropriate visual feedback, such as placing the user's viewpoint behind their avatar so they can see their own avatar movements. Visual feedback will be part of our solution as well.

3.6 Contributions

We have developed an interaction technique that transmits the expressive movement that is a part of communicative gesture. This method applies pen gesture to the avatar gesture control problem in contrast with previous work which has applied traditional graphical user interfaces.

This novel interaction technique is a natural method for specifying expressive motion. The continuous motion of the pen is mapped to continuous motions on the avatar, and continuous stylistic variations in the production of the pen gesture are mapped to continuous variations in the avatar motion. This method takes advantage of a mode that people creatively use for personal expression. People are already quite used to doodling while involved in conversation.

With this technique, users invoke and modulate avatar gesture with the same action. No mode changes are required for setting the gesture expression. Further, users can specify several expressive parameters simultaneously.

3.7 Summary

In designing an interaction method, the main goal was to provide communicative power through the control of expressive movement in avatar gesture. The use of pen gesture is a novel input mode for controlling avatar gesture. The method uses alphabetic characters as the pen gesture input set. This takes advantage of the way that humans naturally express themselves through their handwriting style. The stylistic variation in pen gesture production is mapped to the expressive modulation in avatar gesture. The design also defines how the control actions are made reversible through the use of meta control pen gestures

This interaction technique addresses some of the gaps in previous work. When controls for qualitative variation are available, expressive variation is limited to a single variable. Most interfaces focus on control of particular types of nonverbal communication such as affective display, emblematic gesture or conversation regulation. The pen-based control we have developed provides a more general method of control. This technique can be applied to a broad range of kinesic behaviors beyond gesture, such as facial displays and posture. In contrast, previous techniques are designed to control a specific and limited range of nonverbal displays.

The pen UI also results in a scalable interface because user choices are not mapped to graphical UI elements. The user does not need to commit attention to looking for items on a menu or panel, and can remain more involved in the conversation. And the number and kinds of avatar gesture can change without affecting the visual layout of the application.