7 Conclusions and Future work

Gesture is a very specific form of nonverbal communication involving the movements of the arms and hands. One of the distinguishing characteristic of gesture is that it accompanies speech during spoken verbal communication. In fact, research in psychology has shown that information is often encoded and transmitted through the combination of gesture and speech. Avatars, as visual representations of the human body, have the capacity to produce this form of multi-channel communication in a virtual environment. However, the technologies for controlling avatar gesture limit this ability, especially in desktop and other nonimmersive computing environments. This dissertation explores the use of a pen UI for controlling expressive avatar gesture in a desktop virtual environment application.

7.1 Summary of work

Virtual reality allows users to transcend the limitations of physical reality. Groups of people who are physically dispersed can meet in a virtual space. People can assume different visual forms for social, educational or other purposes. In the virtual world, their physical abilities, and indeed the physical laws of the world can be designed and built. The user's experience of the virtual world depends on the design of the world and technology that enables the user to sense and interact with the objects in that world.

One of the most important activities in social virtual worlds is interacting and communicating with other people. Graphical avatar worlds are a communication medium just like writing, telephone, or video. The limits to communication result from limitations to manipulating the medium.

This research addresses the problem of expressively gesturing with an avatar. One of the most difficult problems with controlling gesture is coordinating the control of gesture with the production of verbal communication. We assume a system in which verbal communication is transmitted through speech. Using current techniques, it is difficult to control an avatar while speaking at the same time. Further, there is a lack of expressive controls, that is, controls in which the stylistic qualities of gestural movement can be modulated.

In previous work, researchers and multi-user virtual environment (VE) application developers have developed mechanisms to handle various forms of nonverbal communication (NVC). The majority of these mechanisms focus on expressions of emotion, through facial displays or body posture, or on emblematic gesture. These techniques are inadequate for controlling more spontaneous and subtle forms of NVC such as coverbal gesture or ephemeral facial expressions.

Other research focuses on conversation regulation behaviors. These behaviors are used to control the initiation, flow and cessation of conversation mainly through eye contact and gaze. This behavior is characterized by the low level of awareness of these volitional actions on the part of the participant.

Our solution is to develop a pen user interface for directing avatar gesture. Pen gesture is a more natural form of control because it relies on manual dexterity skills rather than on hand -eye coordination. Manual dexterity combines naturally with speech since one can use one's hands while maintaining a conversation. The rhythms and dynamics of speech have their counterparts in the movements of writing.

The production of pen gesture has similarities with the production of communicative gesture. Both are particular types of bodily movement. Both can be recognized by particular movement features, yet the movements can exhibit variations that are seen as expressive.

We break down the problem of gesturing to the subproblems of selecting a particular gesture and modifying its expressive qualities. A particular gesture is a symbolic form. Its expressive qualities are those stylistic variations that are perceived as meaningful. For instance, a gesture that is performed with greater forcefulness may appear to be more emphatic.

The analysis of a user's pen gesture is broken down along these lines. The identity of the pen gesture is mapped to a particular avatar gesture while the stylistic variations of the pen gesture are mapped to stylistic variations of the avatar gesture movement. In Chapter 4 we described the mapping from pen gesture to avatar gesture. We also described an initial physical mapping in which pen gesture speed is mapped to avatar gesture speed, and pen gesture size is mapped to avatar gesture duration. In Chapter 5 we speculated about other possibly meaningful maps.

We have developed a method for synthesizing a continuous range of expressive gesture motion variations. The variations are changes in the characteristics of the movement that lend emotional color or other affect to a gesture. We begin with several motion samples. The samples consist of human performance capture data. The same human gesture is performed several times, each time with a different extremal variation on the gesture. To synthesize a new variation, we blend the motion samples using multi-linear interpolation.

Blending motion samples to create new motions or new motion variations is a technique studied extensively in computer animation. The aim of this work is to reduce the amount of manual labor required to create new animation sequences through reusing old sequences. Our work applies these techniques to interactive control of gesture animation. The use of performance capture is particularly suited to avatar animation since it allows the avatar motion to mimic faithfully the motion of the human user if the recordings are made from that same person.

The methods and techniques described above are incorporated into Cursive, an application for interactively controlling an avatar in VRML-based multi-user virtual worlds. Using this application, a user can connect to an uploaded avatar to control its gestures. The avatar itself is instrumented with locally invoked code that opens a socket connection back to the user's host.

All running instances of the avatar open a connection. Other visitors to the virtual world are able to transparently view the avatar's animation. This application will eventually enable us to test user reaction to our gesture technique.

7.2 Contributions

We have developed a technique for controlling expressive avatar gestures using pen gestures. Other desktop virtual reality applications shy away from coverbal gesture and focus mainly on affect display. Research that does consider coverbal gesture narrows their focus to conversation regulation. Our technique is designed to address a broader range of communicative gesture. We further depart from techniques that are based on traditional graphical user interface elements such as buttons and menu items by developing a pen gesture based technique. Pen user interfaces have not been previously explored as a means for avatar control.

Our research also extends control over the expressive qualities of movement. Whereas other interfaces allow the user to modulate the display by one characteristic at a time, our pen based technique permits control over several dimensions of modulation simultaneously.

Unlike other techniques, our design allows rich kinesthetic feedback. The continuous action of writing pen gestures is a performance that can be felt. One can write a letter or word without looking at it. Using graphical user interfaces also results in a kinesthetic feel, but they require the user to visually locate an element before manipulating it. We argue that the employment of manual dexterity and kinesthetics makes the pen interface more engaging and satisfying to the user.

We developed a number of algorithms for realizing the interaction technique. Among these are algorithms for extracting size and speed features from digitized pen gestures, for mapping the values of these features to movement parameter values, and for using the movement parameter values along with avatar gesture animation samples to synthesize new animations of expressive avatar gesture.

The algorithms are implemented in our working application, Cursive. Cursive allows users to control avatar gestures using our pen-based technique. Cursive is designed to work with multi-user virtual environments that require only a standard VRML browser to view the virtual world. Our application allows other virtual world users to see the Cursive animated avatar without using special software. The architecture of this application should be of interest to those who are developing behaviors for shared objects in virtual worlds.

7.3 Future work

The work presented here— the interaction technique, the scheme for breaking down the control problem, and the methods for solving the subproblems—provides a basic framework for exploring the application of pen user interfaces to avatar gesture control. There are many directions for future work to pursue. Some of this work has to do with evaluating the system we have designed and implemented so far. Other work would explore design alternatives or new features.

7.3.1 Evaluations

We would like to extend the evaluation of our system beyond our preliminary evaluations in the laboratory by conducting user studies. In using Cursive, we found that pen gesture is an effective way to index into a library of avatar gestures, and that it is straightforward to control a single movement parameter, either size of duration, by modifying a pen gesture feature. When modulating two parameters simultaneously, we found that using the system effectively requires some amount of practice. The objective of our initial studies would be to find out how well users feel they can express themselves using the basic system we have implemented. This would provide us with clues to what kinds of NVC our system is most suited for, and to what other features may be desirable in the system.

When discussing the ability to communicate nonverbally with avatars, it is tempting to break down measures of communication into the ability to send signals and the ability to receive signals. However, research into methods of measuring receiving ability suggest that receiving ability varies widely depending on the individual and the particular circumstances. Among the findings is that receiving ability is more strongly affected by interactional and contextual affects than is sending ability [16]. Any simple experimental procedure will only measure one very small aspect of a much more complex process.

User studies of avatar worlds tend to look for emergent behavior in social interactions. In [45], pairs of users were asked to meet in a virtual environment but were given no other instructions. In [102], the interaction environment was set up at a series of gallery exhibits where gallery visitor participated in spontaneous interactions. Other studies, such as [99], study structured interactions in a particular domain. (Descriptions of these studies appeared in Chapter 2.)

Our particular system is not as suited for naïve users as those previously mentioned, so we will need to design studies for different levels of evaluation. One more structured study would evaluate the control technique itself and evaluate how well users can control avatar gesture using the system. This is more a study of sender ability. Then a more unstructured study, similar to the first two mentioned, would be used to evaluate how users perform socially using this technique.

7.3.2 Additional features

A feature we would like to add is direct control mode, a feature described in Section 3.3.6. The system could be augmented to allow the user to switch from using library avatar gestures to directly controlling the avatar movements. The switch would be accomplished by writing an escape character. Direct control can be implemented a number of ways, and we would like to investigate these variations. One variation is to separate the selection of the avatar gesture from the production of the expressive movement. In such a scheme, the user selects a gesture using one pen gesture, and then subsequent pen motion "drags" the avatar's arms through the motion path of the gesture. Another possible scheme is to use the pen motion to place the position of the avatar's hand, using inverse kinematics to determine the proper joint angles.

7.3.3 Design alternatives

Our framework for building a pen-based control technique breaks down naturally into a number of subsystems and design features that are amenable to variation. Our design choices are just a starting point in exploring the design space, and each choice seems to address particular issues adequately. In future work we would like peek further into the design space to find more optimal alternatives.

One such design choice is the use of the alphabet as the pen gesture set. As mentioned earlier, letters have the advantage that writing them has already been learned by literate users. A potential problem with letters is that they are laden with psychological associations which may make it difficult to map them to particular gestural motions. We need to investigate further the design of the input gesture set. Design alternatives include more abstract pen gestures and more concrete symbols such as words.

A subproblem in our interaction technique is the design of a map from handwriting features to avatar gesture motion variations. Chapter 4 describes the physical map we implemented while Chapter 5 describes an alternative map using dimensions of emotion. An open problem we face in implementing this alternative mapping is finding the relation from handwriting features to emotion space. We described an experimental method for finding this map, and performing these experiments is the first step we would like to take in studying this alternative.

Another interesting possibility is to parameterize the avatar gesture emotions according to the Effort dimensions of Laban Movement Analysis (LMA). LMA was developed by Rudolf Laban for the analysis of dance movement, though the system has also been used in physical therapy and industrial applications [64]. The Effort portion of the system deals with the expressive portions of motion [8]. The quality of movement is described in terms of the force, tension and control exerted by parts of the body. An application of LMA to character animation is described in [21].

The limitations of our gesture synthesis method are described in Section 4.5. These include the number of sample motions required and constraints on the shape variations of the gesture. Other methods for the use and blending of existing motion samples exist in the computer graphics literature, and through further investigation we may find a method that is better suited for synthesizing gestural motion.

7.3.4 Interdisciplinary work

A suitable means for designing an avatar gesture vocabulary remains an open question. How many gestures and what range of expression is sufficient? In future work, we would like to develop a framework for creating such a vocabulary. Section 3.4.2 suggests possible starting places for designing this vocabulary.

Though literature exists on the recording and classification of gestural behavior, there are no theories on the sufficiency of nonliguistic vocabularies of gesture. Yet, there may exist patterns in the types and frequency of gestures used by people in face to face conversation, and these patterns could be used to design a vocabulary. Another way to design the vocabulary would be to select gestures appropriate to a particular domain. Finally, studies of interactions inside virtual worlds may reveal patterns in the way avatar gestures are employed in virtual worlds. Answering this question will require collaboration with communications researchers.

7.3.5 Alternative applications

Current affective computing research [83] seeks to enable computers to understand affective input and combine it intelligently with symbolic input. The goal is more powerful and natural interaction between humans and computers. Our map of handwriting features to expressive output may prove useful in an application for producing expressive synthetic speech from handwritten input. While the words in the sentence are being recognized, the handwriting style could also be analyzed and then used to control the prosody of the synthetic speech output.

7.4 Conclusions

The many users of avatar worlds have demonstrated their willingness and desire to employ virtual nonverbal communication when it made available to them. Previous techniques for controlling nonverbal communication are too cumbersome to use for producing fluid, spontaneous facial and bodily expression on an avatar. Because they carry symbolic and qualitative information, pen gestures can be employed effectively in controlling avatar gesture, adding richness and subtlety to the to the avatar's gestural movement.