Continuous Control of Avatar Gesture

Francesca Barrientos
Computer Science Division
University of California, Berkeley
Berkeley, CA 94720, USA
fbarr@cs.berkeley.edu

ABSTRACT

We are developing an application to give humans the ability to transmit nonverbal communication behaviors through an avatar: specifically *gesture*, the movements of the arms and hands that accompany speech when people speak face-to-face. In this application the user will have continuous control over the avatar animation. The avatar will be like a virtual puppet and the user will manipulate the avatar using not strings or rods but the controlled and skilled motions of their hand. The system tracks hand motions and then maps that motion to the joint motions of a three-dimensional articulated avatar. As part of this research we will try out different ways of tracking the user's hand. Eventually we plan to test the efficacy of this system by incorporating it into a networked virtual environment in which two or more people can interact through the virtual medium. Working with artists will enable us to design a system that is expressive and to better understand the expressive power of this system.

Keywords

Avatars, computer-mediated communication, virtual environments.

1. INTRODUCTION

Graphical virtual worlds are a manifestation of the desire of humans to realize, in the computing environment, interactions as rich we have in real life. Despite the fact that electronic media can create a myriad of new and strange forms of graphical, audio and other sensory experience, an intent to recreate the feel of face-to-face can be seen in the design of virtual worlds. The worlds themselves are often made up of recognizable elements such as buildings with floors and walls; and avatars, the graphical figurines meant to represent humans, often look human. Virtual reality gives people the ability to escape the bounds of having a body, and yet bodies exist!

Avatars compel us because they give humans the possibility of communicating using virtual bodies when real bodies cannot or should not be present. Yet the ability to use avatar bodies communicatively remains impoverished. Partly this has to do with existing technologies. In most virtual worlds, verbal communication takes place using chat technologies, so users are reading and typing text in addition to manipulating their avatars. Streaming media technologies make it possible to transmit our voices over the network while freeing our hands for other tasks such as controlling avatars. However, the problem of how to animate an avatar for the purpose of nonverbal communication remains unsolved.

In this project we are building an application that will allow a person to easily control the arm movements of an avatar so that the person can gesture through the avatar while in a conversation in a virtual environment. The application will consist of a computer interface and control procedures that allow a person to interactively animate the avatar's body. The idea is that this interface could be used as part of a networked desktop virtual reality system. Using the full system, users could communicate verbally using networked telephony and nonverbally using their avatar.

We want our application to be adopted by casual users, so we are designing our control interface to work with simple input devices. In one implementation a pen-and-tablet device allows a user to "doodle" with their hand. The doodling motion is tracked and is used to move the limbs of an avatar. In another implementation, we will track the user's arm motions with a single position/orientation sensor, and map that data on to the movements of the avatar.

Because we are using such simple sensing techniques, the corresponding motion of the avatar will be highly constrained. Thus, part of the problem we are solving is how to create controls for highly expressive gesture in a limited space of movement possibilities. We expect some of the expressive power to come from just the ability to control an avatar continuously while simultaneously speaking. Much of the power must come from artistic and scientific design of the avatar's movement characteristics.

We will also study how people might use a continuos avatar control system for communication. Many types of nonverbal communication exist, and some methods of control may be more appropriate than other for producing these various kinds. We expect that continuous control will give users the ability to communicate in ways that are not possible using other discrete input methods. In particular, we expect to see the emergence of the various kinds of gesture. In our studies, we want to learn how constraining motion affects the kinds of gestures that can be produced, how expressive a sender feels as they use our system and whether or not the gestures coming from an avatar make sense to a receiver. We also plan to study how the addition of continuous controls for avatars affects interaction in a virtual world.

2. RELATED WORK

Our project employs some techniques that are similar to the performance animation systems have been used to animate virtual characters on television and at trade shows. These systems are designed to take a low number of inputs, then produce complex and interesting animation by combining the control signals with algorithmic behaviors [15]. Badler and others have been developing algorithms for recreating full body postures using only four position/orientation sensors [2]. Our project looks at the more extreme problem of controlling an avatar with less sensing and more understanding of how the control technique affects human communication.

Most virtual world communication technologies combine chat with a graphical world. *ComicChat* [8] is a two-dimensional chat world in which avatars are drawn statically as comic strip characters. Users select emotions using an emotion wheel in which the different directions around the wheel select different expressions: coy, happy, scared, bored, etc. Then the avatars are drawn with the appropriate facial expressions and body postures. Users are also able to select actions for their avatars such as *waving* or *pointing*.

Other work has been done using three dimensional networked virtual environments. The non-verbal communication interface to the *VLNET* system [5] contains controls for selecting affect display, emblematic gesture, and avatar actions. Slater et. al. [17] developed a non-verbal communication interface for avatar expressions in order to study the feasibility of virtual reality for rehearsing for a live dramatic performance. Mouse movements allow the actors to generate head nods and wags. Facial expressions are created by drawing onto a generic smiley face. The bodies can be placed into standing or sitting positions, and the arms can be raised and lowered independently. *BodyChat* [22] is a virtual world chat system developed by Vilhjálmsson that treats avatars as semi-autonomous communicative agents. Users specify high level goals such as initiating, rejecting or ending conversations. The avatars automatically handle the nods, glances and other gestures required to negotiate the conversation.

Two works that study nonverbal expression but not avatars are worth mentioning. In the *digital baton* project by Marrin [10] a baton is instrumented with sensors to determine position, velocity, acceleration and orientation. A user uses the baton like a conductor's baton for gestural input to create music. Work by Snibbe and Levin [18] addresses the question of using a computer for expressive imagistic communication. Their system tracks human movement and produces complex abstract graphic animations. It has many of the features we are trying to develop: the output is more sophisticated than the initial input; the "process of interaction is also the product of interaction"; and it is used to mediate communication with another person.

Current interfaces for generating nonverbal communication on avatars do not give the user the ability to spontaneously gesture with their avatars. Typically, these applications give users the ability to display *affect*, that is, emotion or state-of-mind. Interfaces for this type of nonverbal behavior are easier to design because usable displays of emotions can be enumerated and designed ahead of time. In addition, user's do not need to change their emotional displays often while conversing, so they are easy to use.

3. SPONTANEOUS GESTURE

Gesture, as we use the word here, is a particular kind of non-verbal behavior that is always accompanied by speech. Many gestures are spontaneously created in order to participate in the efficient encoding of verbal messages. They are *sui generis* forms whose meanings are derived from their physical form and their temporal relationship with speech. Gestures are a form of *imagistic* communication, much in the way that drawings are. The movement of a gesturer's hands come to symbolize something in the way that streaks of graphite on paper, when artfully made, can mean a chair.

3.1 Types of Gesture.

The kinds of gestures we hope to recreate have been categorized by McNeill [12] as follows:

iconics. Iconics describe physical things or actions. For instance, in describing a birthday cake, one might draw a circle in the air to describe the shape of the cake.

metaphorics. Metaphorics describe abstract ideas or thought processes. These gestures often depict images derived from the metaphors embedded in the language [9]. For instance, the idea of a story as an object may be implied in a gesture that denotes a container to hold that object.

deictics. Deictics are pointing gestures. Interestingly, deictics can point to abstract things. For instance, gesturers often assign meanings to spaces around their bodies, and gesturing towards a space may emphasize that topic.

beats. Beats are simple gestures that mark particular words as important to the discourse. They serve a similar function as putting vocal emphasis on particular words in a sentence.

We are interested in these kinds of gestures specifically because they cannot be designed ahead of time. In order to create these kinds of gestures using an avatar, the user must be able to spontaneously create movement forms[1]. In addition, this categorization provide us with a way to study the kinds of gestures that users may employ using their avatars.

3.2 Communicative role of gesture

Studies show that listeners attend to gestural motion as an indication that the speaker is thinking and does not want to be interrupted. While speech is taking place, listeners rely on gestural cues when speech is ambiguous [20] or when white noise inteferes with the transmission of speech[14]. When listeners are shown narrations which are designed so that the information conveyed by gesture is distinguishable from the information conveyed by speech, listeners attend to the gestures and can relate the information that was conveyed through gesture [3]. Verbal descriptions accompanied by descriptive gestures facilitate the ability of listen-

---

[1.] Another common type of gesture is *emblematic gesture* which refers to signs made with the hands which are well understood in the culture and can be used to replace spoken words. We do not address them here because they do not depend on speech.

ers to identify objects and to recall lists of words and short stories [14].

Spontaneous gesture is often described as *unwitting* gesture because speakers are not always aware of what they are doing. However, speakers do change their gesticulative behavior depending on the situation. Whether or not a person produces gestures depends on whether or not they believe that the gestures will be available to a receiver [7]. Further, when gesture and other means of communication such as diagrams are available, the speaker will combine these different modes into integrated units of *composite* signals [4]. When mothers talk to babies, they will modify both their speech and their gestures [12]. If subjects are asked to be more persuasive, or have increased motivation for approval, they will produce more frequent gesture [13] [16].

4. DESIGN CONSIDERATIONS

4.1 Continuous and Transparent Control

We feel that the ability to synthesize gesture through continuous action is essential. Spontaneous gestures are made up of unique forms, so they cannot be designed ahead of time. Further, gesturing while generating sentences helps speakers to constitute thoughts [7][12]. Interacting with a computer while speaking and gesturing can add another layer of complexity. In [11], McCullough argues that "computers fragment our thinking by substituting discrete events for continuous actions", and that continuous interaction allows the use of a tool to become *transparent*— that is, use of the tool becomes subconscious, and the user is only aware of the intent to create an object. We believe that continuous interaction will allow our interface to become transparent so that it does not interfere with the ability to converse.

4.2 Controllability and Expressiveness

In order to make the avatar controllable using a simple input device, the user will control, in general only the direction of movement of the avatar's arm. The exact shape of the arm as it moves will be constrained. Careful design of the constrained motion will provide expressiveness. We believe that even with constrained motion, people will find a way to create interesting gestures. In face-to-face humans can send the same message using different gestures; for instance, it is possible to shrug using ones hands, eyes, head or mouth as well as with one's shoulders [Ekman personal communication].

4.3 Affordances

Hollan, et.al., suggest that trying to recreate face-to-face may blind researchers to the possibility that computer-mediated forms of communication provide affordances that in-person communication cannot [6]. And a look at the avatar's cousin, the puppet, suggests that an avatar should not be too life like.

> Puppets that attempt to imitate human movements often create a superficial sense of realism. Once this novelty has worn off, the audience usually becomes aware of the difference between puppet actions and human actions. Puppets that create the illusion of life by using movements exclusive to their con-

struction can more easily encourage an audience to accept the living existence of an otherwise inanimate object [21].

So, avatars may be at their most expressive when they perform gestures that are specifically designed for them.

4.4 Feedback and Proprioception

A drawback to controlling a virtual body is the lack of proprioceptive feedback. In some virtual reality applications, users reported losing track of the positions of their limbs [5][17]. In our application we will add visual feedback so that users can see the positions of their bodies. And we will provide a mechanism for automatically returning the avatar bodies to a "resting" position.

5. PROTOTYPE

We have developed a prototype [1] in which a single arm of an avatar is controlled by dragging a mouse. An avatar is displayed using a VRML browser, and the control interface is written as a Java applet. As the mouse is dragged is around on the plane, the avatar's shoulder, elbow and wrist joints rotate. When the mouse button is released, the avatar's arm returns to a resting position.

6. WORKING WITH ARTISTS

Besides formal user studies, we plan to have other ways of studying this technology. This section lists some possibilities.

Designing an Interactive Exhibit. We plan to set up an exhibit at a science museum or other similar venue where children and adults can use our system for communicating with one another. Our objective at the exhibit would be to make observations of novice users of our system. An artist could help us to design a compelling experience that would draw passersby.

Improvisation exercises. Actors may be able to show us types of gestures that are effective for avatars that other novice users might not have invented. Performing improvisational exercises may illicit this kind of invention. As an example, in one exercise [19], actors must use only their hands to tell a story. The acting instructor gives the following coaching:

> *Laugh with your fingers! Shrug your hands, not your shoulders! Remember, we can't see your face! Put all that energy into your fingertips!*

To transfer this exercise to our application, the actors would be able to move only the limbs that we gave them control over.

Designing stylized movement. Another way to look at the problem is to consider the function of movement constraints in stylized forms of theatrical and performance arts such as dance.The style of the dance constrains the path of the limbs as they move from one position to the next. Though the reason for a limited vocabulary is to create an aesthetic whole, it also aids the dancer in understanding out how to move. Variables in the motion, the exact attitude of the hand as well as the speed, acceleration and smoothness of the movement itself, allow each dancer to add expressive nuance to the movement.

A Puppet Play. We might write a short puppet play—perhaps a version of *Faustus* which is a classic in Western puppetry—and perform it using the avatars.

7. CONCLUSION

The invention of technologies such as writing and printing has forced humans to encode meaning into purely verbal channels. However, in face-to-face communication, language comes to each of us "embedded in a matrix of other channels" consisting of not only words but also vocal intonations and facial, hand and arm movements [23]. Given that avatars are a visual representation of the user's body, they should be exploited for their possible uses in communicating via the visual channel.

We have presented our ideas for an application for controlling avatar gesture for the purpose of communicating nonverbally in a virtual world. We have explained why we believe that continuous control is necessary for producing gesture. And we have described possible collaborations with artists that would enhance the quality of our work.

8. ACKNOWLEDGEMENTS

9. REFERENCES

[1] http://www.cs.berkeley.edu/~fbarr/Doodler

[2] Badler, N., M. Hollick, and J. P. Granieri, "Real-time control of a virtual human using minimal sensors," *Presence*, vol. 2, pp. 82-86, 1993.

[3] Cassell, J. and S. Prevost, "Distribution of Semantic Features Across Speech & Gesture by Humans and Machines," presented at Workshop on the Integration of Speech and Gesture, 1996.

[4] Engle, R. A., "Not Channels but Composite Signals: Speech, Gesture, Diagrams and Object Demonstrations Are Integrated in Multimodal Explanations," in *Proceedings of the Twentieth Annual Conference of the Cognitive Science Society*, M. A. Gernsbacher and S. J. Derry, Eds. Mahwah, NJ, Erlbaum, 1998.

[5] Guye-Vuillème, A., T. Capin, I. Pandzic, N. Magnenat-Thalmann, and D. Thalmann, "Non-Verbal Communication Interface for Collaborative Virtual Environments," in *Proc. CVE 98*. Manchester, 1998.

[6] Hollan, J. and S. Stornetta, "Beyond Being There," presented at CHI, 1992.

[7] Kendon, A., "Gesture and Speech: How They Interact," in *Nonverbal Interaction*, J. M. Wiemann and R. P. Harrison, Eds. Beverly Hills, Sage Publications, 1983.

[8] Kurlander, D., T. Skelly, and D. Salesin, "Comic Chat," in *SIGGRAPH 1996*, 1996.

[9] Lakoff, G. and M. Johnson., *Metaphors we live by*. Chicago: University of Chicago Press, 1980.

[10] Marrin, T and J. Paradiso, "The Digital Baton: a Versatile Performance Instrument," in *Proceedings of the International Computer Music Conference*. Thessaloniki, Greece, 1997.

[11] McCullough, M. *Abstracting craft: the practiced digital hand*. Cambridge, Massachusetts: The MIT Press, 1996.

[12] McNeill, D., *Hand and Mind: what gestures reveal about thought*. Chicago and London: University of Chicago Press, 1992.

[13] Mehrabian, A. and M. Williams, "Nonverbal concomitants of perceived and intended persuasiveness," *Journal of Personality and Social Psychology*, vol. 13, pp. 37-58, 1969.

[14] Riseborough, M. G., "Physiographic gestures as decoding facilitators: three experiments exploring a neglected facet of communication," *Journal of Nonverbal Behavior*, vol. 5, pp. 172-183, 1981.

[15] Robertson, B., "Feature: Cyber Acting," in *CGW Magazine*, 1996.

[16] Rosenfeld, H. M., "Approval-seeking and approval-inducing functions of verbal and nonverbal responses in the dyad," *Journal of Personality and Social Psychology*, vol. 4, pp. 597-605, 1966.

[17] Slater, M., J. Howell, A. Steed, D. P. Pertaub, M. Gaurau and S. Springel, "Acting in Virtual Reality," in *submission*, 2000.

[18] Snibbe, S. and G. Levin, "Interactive Dynamic Abstraction," *Proceedings of the Symposium on Nonphotorealistic Animation and Rendering*, 2000.

[19] Spolin, V., *Improvisation for the Theater*. Evanston, Illinois: Northwestern University Press, 1963.

[20] Thompson, L. A. and D. W. Massaro, "Evaluation and integration of speech and pointing gestures during referential understanding.," *Journal of Experimental Child Psychology*, vol. 42, pp. 144-168, 1986.

[21] Tillis, S., *Towards a New Aesthetics of Puppetry*. New York: Greenwood Press, 1974.

[22] Vilhjálmsson, H. H. and J. Cassell, "BodyChat: Autonomous Communicative Behaviors in Avatars," in *ACM Proceedings of the Second International Conference on Autonomous Agents*. Minneapolis, 1998, pp. 269-276.

[23] Wiener, M., R. Shilkret, and S. Devoe, "'Acquisition' of Communication Competence: Is Language Enough?," in *The Relationship of Verbal and Nonverbal Communication*, M. R. Key, Ed. The Hague, Mouton Publishers, 1980.