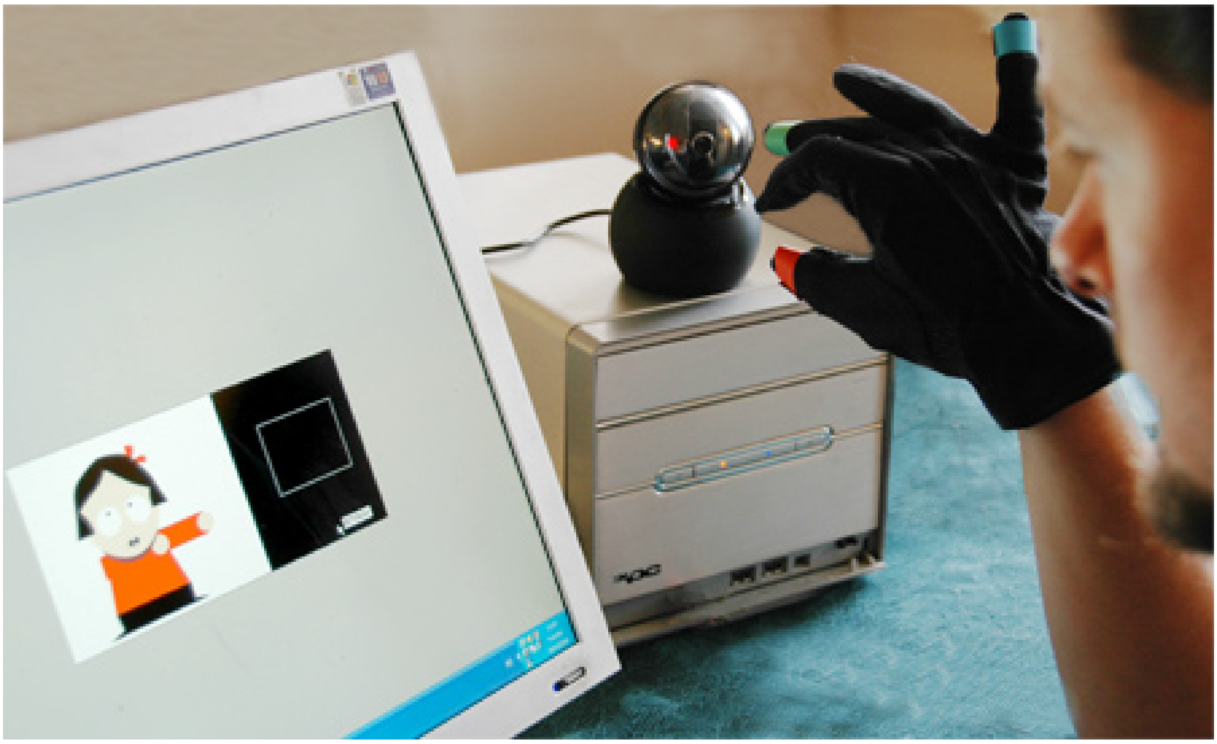

The Puppet UI

Tools for nonverbal communication in virtual environments

Authors: Fatland, Eirik & Li, Xin

Abstract

At present, a user’s control of his/her persona in an online virtual world such as “World of Warcraft” or “Second Life” is constrained to text input through a keyboard, excluding most forms of nonverbal communication. The “Puppet UI” is an ongoing exploration of alternative input methods, inspired by hand puppets, that are intended to enhance inter-personal communication and user agency in virtual environments.

1. Introduction

For most citizens in the industrialized world, and especially for the younger generations, computer-mediated communication tools such as e-mail, weblogs, social software, instant messaging and web fora have become integral parts of everyday life. With the exception of dedicated audio and video services, such as Skype or YouTube, such tools are still largely limited to text.

Reliance on text communication is especially obvious and puzzling in so-called "Virtual Worlds" that frame their communities within an imitation of the physical world and represent their users as virtual personae called "avatars". In terms of technology, text chat is the simplest solution to the needs of computer mediated communication. In terms of culture, the written word conceals its' writer, and allows for anonymity and role-play in the virtual environment. However, it is poorly suited to carry the breadth and depth of social communication – much of which relies on nonverbal signals.

To the degree that an avatar can mimic the human body, it is controlled through a mouse and the keyboard, and relies on static pre-recorded animations for movement, activated through a text command or button click. Such interfaces are cumbersome, and map poorly to user intentions. A better input method would facilitate real-time control of avatar gesture and posture as well as movement, a seamless connection between the user's nervous system and avatar locomotion.

The "Puppet UI" is a proposed input method for avatar-based communication in virtual environments. Our goal is to facilitate user expression of bodily and vocal nonverbal signals. In designing this input method, we are taking our inspiration from hand puppets - user interfaces that predate the Internet by millennia, and are well-understood and well-tested ways of animating a virtual persona. In developing the Puppet UI, we are using an iterative design process, involving benchmarking, user studies, low-fidelity and working prototypes. This paper outlines the context of this work, and presents the results of the first iteration.

An iterative design approach requires a certain degree of flexibility. We are working on an "input method", which may or may not require a specialized input device, and the "Puppet UI" may take the form of an actual puppet, worn over the user's hand, or of an interface that is only distantly inspired by puppetry but serves a similar purpose.

2. Background: Verbal and nonverbal communication

2.1 Nonverbal signals

Nonverbal signals are essential to any kind of interpersonal communication. As a "communication coding system" nonverbal signals play a key role in impression formation, complex emotional expression and conveying personality (Burgoon and Hoobler, 2002). Research has found that the majority of humans are strongly reliant on nonverbal cues, such as body movements, gestures and appearance, in order to form initial impressions of others and act upon those impressions. Burgoon and Hoobler describe seven classes of nonverbal codes in human visual cognition and sound sensation:

- **Kinesis**: bodily movements, gestures, facial expressions, posture, gaze, and gait
- **Vocalics or paralanguage**: pitch, loudness, tempo, pauses, and inflection
- **Physical appearance**: clothing, hairstyle, cosmetics, fragrances, adornments
- **Haptics**: use of touch, including frequency, intensity, and type of contact
- **Proxemics**: use of interpersonal distance and spacing relationships
- **Chronemics**: use of time as message system, punctuality, lead time, etc.
- **Artifacts**: manipulable objects and environmental features that may convey messages

In face-to-face communication, most of these nonverbal cues are rich, visible, and relatively easy to grasp. In most situations, they co-exit with linguistic cues.

2.2 Nonverbal signals in virtual worlds

Table 1 shows the degree to which these non-verbal cues are implemented in two of the most popular virtual worlds – World of Warcraft (a MMO game) and Second Life (a multi-purpose virtual world).

	Kinesis	Vocalics	Appearance	Haptics	Proxemics	Chronemics	Artifacts
Second Life	Limited, UC	No	No	No	Possible	Possible	Yes, UC
World of Warcraft	Very limited, SC	No	Partial	No	Possible	Possible	partial, SC

Table 1: *Support for nonverbal signals in "Second Life" and "World of Warcraft". UC = User Created Content, SC = System Created Content.*

Proxemics and Chronemics are possible to use as non-verbal signals in virtual worlds, but the nature of such worlds decrease the power of these signals. Since the user is not actually present in the virtual world, but observes the avatars from an external point of view (zooming in and and out at will), proxemics loose some of their power. Lacking voice, and accurate information about the user presence, chronemics – too – loose some of their communicative power. If a user delays an answer, he/she may not be sending a nonverbal signal but simply be preoccupied with something other than virtual world interaction in her real physical environment.

Haptics, the illusion of touch, pose large challenges in terms of technology and human factors, rather than design, and are therefore excluded from this study. Tools that allow the users to modify and create content, such as Second Life's 3D building tools and scripting language, are already proving effective at allowing players to use appearance and artifacts as nonverbal signals.

Our focus is on the two remaining classes of nonverbal signs: kinesis and vocalics. These are, incidentally, the most commonly used and identifiable forms of nonverbal signals.

2.3 Gesture

Colin Ware states that "gesture as linking devices" is the most natural way to link verbal content and visual imagery. He classifies gestures as belong to three distinct classes, dependent on their relationship to the speech they accompany: deictic (indicating), symbolic (illustrating) or expressive (emphasizing). According to Ware, gestures provide an additional, visual, cognition channel alongside the audible channel of speech (Ware 2004). A similar observation is made by Goldin-Meadow (1999): "Because gesture rests on different representational devices from speech, and is not dictated by standards of form as is speech, it has the potential to offer a different view into the mind of the speaker."

2.4 Voice and vocalics

Voice communication, a prerequisite for vocalic signals, require voice rather than text communication. At present, voice communication in virtual worlds is only available through third-party services such as Ventrillo and Teamspeak, but not through the user interface itself. Furthermore, speech through such services has no effect on the properties of the virtual world. When a user chats in Second Life, the avatar mimics keyboard-typing movements, but if he or she speaks through voice-chat software nothing happens to the avatar. Neither can the user's voice be heard by other users without the required software set to the appropriate channel.

2.5 The case of the emoticon

Emoticons, such as the ubiquitous smiley :-) constitute a special case of non-verbal communication. They are analogous to vocalics, in the sense that they are not linguistic but are conveyed through the same medium as language (voice or writing).

While written language is thousands of years old, emoticons appeared only after computer networks set the stage for a written culture that was informal, immediate and social. Conveying emotional intent through text traditionally belonged to the art of creative writing. The masters of this art were writers and poets, one-way communicators who could toil for years on finding the perfect stanza or paragraph to express an emotion. Amateurs practiced the same art in carefully composed letters. Handwriting style and calligraphic decoration could provide additional non-verbal cues, unavailable in the standardized fonts of digital text. Social computer-mediated communication, being sent in real-time or near real-time, could not afford the time required for careful composition. The emoticon quickly appeared as a way to convey emotional intent, to distinguish a joke from an insult.

While being analogous to vocalics, emoticons illustrate facial expressions, a subset of kinetic signals. This may partially be the result of encoding the emotion in as few characters as possible (try illustrate a hug or a high-pitched scream using text), but may also point to a user preference for facial expressions over other nonverbal signals in online environments.

Emoticons highlight the cultural context inherent in nonverbal communication. Western emoticons emphasize mouth gestures - :-) is smiling, while :-D is laughing. East Asian emoticons emphasize the eyes: ^_^ is the East Asian equivalent of :-), whereas '_' is the crying equivalent of :-(. In Chinese, Korean and Japanese cultures, individuals are expected to hide their emotional state to a higher degree than individuals in cultures of European descent. Eye movements, less consciously controlled and harder to conceal than mouth movements, thereby become the principal nonverbal clues to the emotional state of the speaker.

3. Design Process

3.1 Benchmarking

The virtual worlds that are popular at present have either no user control of avatar gestures, or implement an interface where the user can run commands – as text or button-clicks – to initiate gestures or modify posture. In Second Life, for example, a /dance command will start the avatar dancing according to a pre-defined animation. Users can replace such animations by uploading a custom animation of their own design, but not improvise a new gesture on the fly.

There are several examples of design research approaching non-verbal communication in avatar-based online environments. One of the earliest “Comic Chat” (Kurlander et al. 1996), represented users as 2D avatars on a stage composed according to the conventions of comic strips. BodyChat (Hannes and Justine 1998) and Cursive (Barrientos and Canny 2002) automate avatar gestures by estimating the user’s intentions. In BodyChat, intentions are explicitly input through user selection of events such as “greeting” and “farewell” but are modified by variables such as time and distance between avatars. Cursive, designed for pen input, allows avatar control through pen gestures, and additionally interprets the user’s handwriting to determine the avatars gesture and posture. Sentoy (Paiva et al. 2003) uses a physical doll as input device for a virtual doll, in the context of a game where the user would manipulate the doll through one of six pre-set gestures corresponding to emotions.

3.2 Target Group: Social Internet users

We are interested in designing for the Social Internet user. This group is fairly wide, but not all-inclusive. Nearly all Internet use is social in the sense that it involves communication between individuals. But much of this communication is pragmatic or utilitarian - geared towards the exchange of factual information or the maintenance of social bonds formed outside of the Internet context. Work-related e-mail and instant messenger contact with family members are examples of such communication.

Social Internet users additionally seek interaction with other users who they may never meet offline, and engage in social behaviour for the purpose of entertainment or self-expression. Examples of such behaviour include:

- “Role-play”, in its’ strict meaning of collaborative construction of fictional reality through dramatic impersonation. (Montola 2005).
- “Role-play”, in its’ looser meaning of expressing an online identity separate from the user’s offline identities. (Donath 1999)
- Social curiosity - building familiarity with (and understanding of) people the user cannot meet in the offline world
- Flirtation and virtual romance.
- Cybersex and sexually themed role-play.
- Company, the satisfaction of the need for human contact.
- Community, the satisfaction of the need to belong.
- Competitive game-play, either pitting individuals against human-controlled opponents or as teams combating AI or other human opponents.

We can draw two conclusions from this list of Social Internet behaviours: All of these practices may, to a greater or lesser degree, be aided by non-verbal communication. And they rely, to varying degrees, on hiding the actual user under the layer of the online persona.

We can assume that current social users at present find text input at least minimally acceptable, or they would not engage in social behaviour on the Internet. Our goal is thereby to enhance the experience of those users who find the text input method lacking. These fall into two groups: those who make do with text input, but would prefer a more expressive interface, and those who currently are not social users due to the lack of support for nonverbal communication.

A third group, social users who prefer the text input method, fall outside of our target group. Such users might lack interest or skill in non-verbal communication, or they might feel more comfortable and skilled in expressing themselves through text. Many of current social users, due to practice and familiarity, can be assumed to belong to the third group.

We are primarily interested in aiding the social interaction of adult users, but do not exclude potential future applications designed for children.

3.3 Design objectives

First and foremost, our ideal input method should facilitate complex non-verbal communication between users. Additionally, it should fulfil four criteria that follow from the target user group:

- Agency - the user should feel a greater sense of presence and control in the virtual environment than with the keyboard input method. (Murray 1999)
- Learnability - no prior training or explanation should be required for the user to master the basic functionality of the input method. (Norman 2003)
- Flexibility - users should be capable of learning or inventing gestures not envisioned in the original design.
- Feasibility - the input method or device should be based on off-the-shelf, inexpensive technology.

4 The Hand Puppet

4.1 Learnability of hand puppets

Puppets and their interfaces come in a wide range of complexity and required skill. Some puppet designs, such as the Balinese shadow puppet or the two-handed marionette of European puppetry, require years of practice to master. “Hand puppets”, which are worn over the puppeteer's hand, require minimal skill to manipulate. A child may pick up a hand puppet for the first time, and immediately be able to use the puppet to mimic speech and for gestural communication.

The learnability of puppet models is largely an issue of mapping and coordination. In the case of the hand puppet the fingers of the puppeteer are mapped directly to the limbs of the puppet, providing instant feedback and reducing the cognitive load of puppet operation. Single-handed puppets further reduce the need for coordination between the puppeteer's limbs. As such, hand puppets are ideal models for an avatar control method designed for end users with no particular skill in puppetry. Additionally, single-handed puppet interfaces leave the puppeteer's other hand free for mouse or keyboard interaction.

Most hand puppet designs fall into one of two categories - they are "sock puppets", where the puppeteer uses the thumb opposite the four remaining fingers to control a puppet's mouth movements, and "glove puppets" where the puppeteer manipulates the head and arms of the puppet using the thumb, middle and little fingers. The anatomy of the human hand makes it difficult to move all five digits independently of each other. Typically, a glove puppet will be manipulated through the use of three digits: the thumb, the little finger and the middle finger. The index and ring fingers may either be clutched against the palm, or (preferably) moved together with the middle finger and little finger.

4.2 User Study

We conducted an informal user study to observe how ordinary people (not professional puppeteers) express themselves through a glove puppet. As we sought to understand play behaviour, the experiments took place in a class room and a cafe, rather than in a laboratory or studio. We did two different experiments: in the first experiment, we asked users manipulating a hand puppet to have a short conversation with another person. In the second experiment, we ask users to express certain emotions, such as joy, sadness and boredom, through the hand puppet. These experiments were recorded to video and reviewed afterwards.

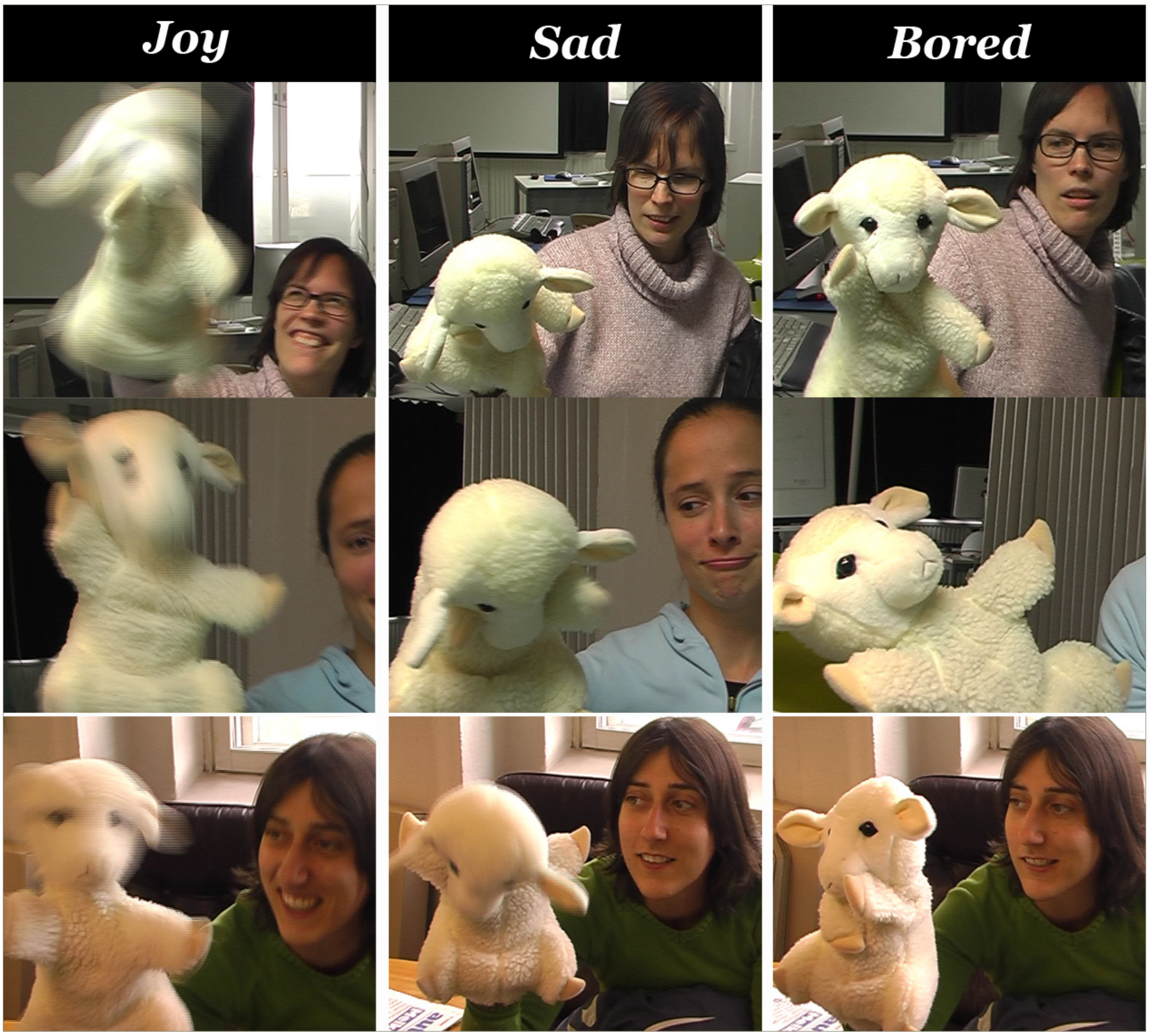

Figure 1: *User study on how people express their feeling through a hand puppet*

Findings from observation of the conversation experiments (3 experiments):

- Users made larger and more visible movements when the puppet was “speaking” than when it was “listening”. For example, users would move the puppet in a small, nodding movement to indicate that “I m listening”. (Much like two people have a conversation)
- All three of Ware’s gesture classes (symbolic, deictic and expressive) were used.

Findings from observation of the expression experiments (4 experiments):

- Joy was usually indicated by large, energetic, rapid movements with the puppets head lifted upwards.
- Sad movements were usually slow, with the puppet's head bowed down and the hands held close to the face or eyes.
- Boredom movements were less consistent than the two others. Two out of four users expressed this emotion by making puppet leaning backwards.
- Although there are general trends for both joy and sad movement, each user's movements were distinct and unique.

Findings from both experiments:

- All users employed vocalizations, such as laughs, whistles and sighs, to clarify the puppets' emotion. For example, rapidly twisting the puppet from left to right might be interpreted in several ways, but giggling sounds accompanying it were used to express joy.
- Users often twisted the puppet around. Rotation of the head or whole puppet along the vertical axis was the most common form of twisting.
- A recognizable figure might not be an ideal puppet for a generic avatar. Our puppet looked like a lamb, and so at least two users tried to express the character of a lamb rather than conveying their own emotions and intentions.

5 Input methods

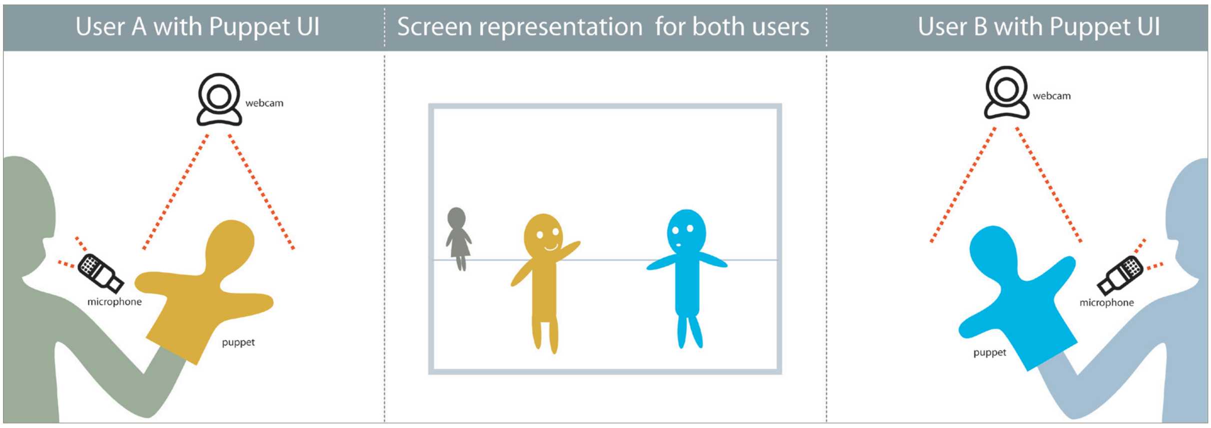

Figure 2: *Basic setting of PuppetUI*

5.1 Tracking

To emulate the interface of a glove puppet, our ideal input method needs to track three points - corresponding to the three fingers - in all three dimensions. Glove puppets only incidentally require the movement of the finger joints to operate - sufficient information can be obtained by tracking the fingertips relative to the hand.

Rotation of the fingers relative to each other is near impossible, and so tracking of finger rotation is not necessary. However, hand puppet manipulation requires rotation and movement of the whole hand. Our user interviews showed users twisting the puppet from left to right in order to communicate a gesture of refusal, or lifting it rapidly up and down in a jumping motion to communicate joy or anger.

An ideal hand-tracking device for avatar puppetry will thus have the following requirements:

- Tracking of three points, correlated with fingers, in 3 degrees of freedom (x, y and z axes) each.
- Tracking of whole hand rotation and movement in 6 degrees of freedom - x, y and z plus rotation along all three axes.

This adds up to a requirement for a 15 DOF hand-tracking device, a level of tracking offered only by expensive high-end datagloves. However, the interface method does not require precise data on hand location and rotation, as information about hand rotation along the x and z axes can be inferred by comparing fingertip location to the palm location. Additionally, the movement of fingertips is anatomically constrained, so that their position along the z axis may – at least partially - be inferred from their position along the x and y axes. sA minimal hand-tracking device for avatar puppetry therefore needs only to meet the following requirements:

- Tracking of the palm in two degrees of freedom (x,y).
- Tracking of three points correlated with fingers, each along the x and y axes, relative to the palm.
- Tracking of hand rotation along the y axis.

The minimal Puppet UI may therefore be made by tracking four points along two axes, and rotation of the whole hand along one axis. The location tracking can be accomplished through a 2D input device such as a web camera, accompanied by colour or pattern tracking software and calculation of the points' position relative to each other. Hand rotation along the y axis is not easy to infer from a 2D capture device, but early prototypes can be done without y-rotation tracking.

Additionally, certain events may need to be tracked with additional precision:

- Two fingers touching, e.g. the case of an avatar clapping.
- Fingers touching the palm, the case of a closed fist.

These events may be inferred by the proximity of tracked points – two fingertips overlapping can be interpreted as a "clap" event. Unfortunately, occlusion of a tracked point by another will yield the same data even without the fingers actually touching. Ideally, touch sensors in the device will be able to detect such events.

5.2 Feedback

Glove puppets provide tactile feedback through fingertips touching each other or the hand, and through the properties of the material itself. The material properties of the puppet are important constraints on user manipulation - by making movements easier or harder to perform, they constrain the puppet to make certain gestures easier or harder to express. In a well-designed puppet, the gestures that are easy to express will be the ones that make thse puppet more convincing. A Puppet UI may be designed along the same lines – built to constrain certain movements, and make other movements simpler. This would also provide the user with passive haptic feedback.

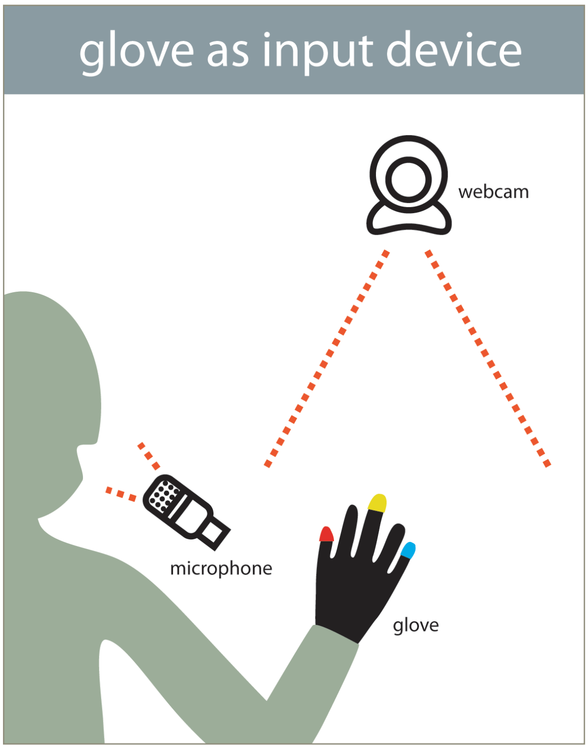

Figure 3: *Glove as main input device*

Active haptic feedback is desirable to the degree that the avatar touches other physical objects in the virtual environment. This may be the case, as with avatars shaking hands or wielding tools, but the complexity and expense of active haptic devices means that this type of feedback is not feasible for the Puppet UI project.

The primary feedback from the Puppet UI will thereby be visual. As the interface maps the user's hand to a 3D avatar, visual feedback may come from both locations:

- The user observing the avatar movements on the screen
- The user observing his/her hand as it is being tracked

This leaves us with two options for further design. If the screen is to provide all feedback, the input device does not need to resemble a puppet. But if the input device, by itself, provides visual feedback it may need to resemble the puppet to some degree. We intend to explore both options through further prototypes and user studies.

5.3 Working Prototype

To test our assumptions further, we built a rough working prototype. The prototype was programmed in Macromedia Director MX, using the TTCPro Xtra for colour tracking. Our protoype input device was a glove, with three coloured points on the fingertip, worn and held in front of a web camera. The software would track the fingertip points, and map them to the limbs of a 2D avatar. We tested two different forms of mapping. In our first experiment, points and limbs were mapped directly to each other. If the user moved her middle finger to the right, the avatars head would move to the right. The second experiment limited the avatars movements to trajectories – if the user moved her middle finger to the right, the avatar head would stay in the same location, but tilt towards the right.

The first experiment, providing direct mapping between fingertips and avatar limbs, worked better than the second experiment, where the tracked data was modified to better represent the movements of the avatar. User's reported a stronger sense of control, and viewers felt the motions of the simple avatar to be more believable. These results indicate that immediate feedback and direct mapping are necessary to sustain user agency in this form of manipulation.

The lack of precision in colour tracking led the avatar's limbs to move even when the users hand was still. The precision of colour tracking may be improved through a better camera, better tracking points, and more control over the lighting. Still, this indicates that the colour tracking method may not work well enough for an end-user product.

6. Speech

Hand puppet manipulation may be performed silently or accompanied by the puppeteer impersonating the puppet's voice. Though one might easily imagine a purely non-verbal virtual environment, where communication occurs through mimicry, the goal of our design is to enhance rather than substitute the user experience of current virtual environments. Language, whether verbal or textual, is therefore a necessary component of the puppet UI. This issue is unresolved in the 1st iteration prototype.

If the user is engaged in using one hand for avatar control, the other hand might be used for text entry through a keyboard. Keyboard text entry, while it may be done one-handed, is significantly faster when done with both hands. Furthermore, users are generally accustomed to two-handed keyboard typing. In the typical desktop computer set-up, one hand is also used for interaction with a mouse or other pointing device.

Better, then, to use voice communication – which would also add vocalics to the array of non-verbal signals. By tracking the user's voice input, especially if aided by a microphone that filters out ambient noise, the avatars mouth or head can by synchronized to the spoken word. In our initial user tests, we saw a clear difference in user behaviour depending on whether the puppet was speaking or not. A speaking puppet would be moved in large, dramatic gestures while a silent (listening) puppet might make small, almost imperceptible movements to communicate agreement, disagreement and attention. Tracking the user's voice input can thus confer the additional benefit of modifying avatar behaviour to imitate this pattern.

However, voice in virtual environments is far from unproblematic. The user's real voice risks revealing his/her gender, age, nationality, dialect or sociolect in ways that text chat does not, limiting the anonymity of the user and making role-play difficult. Practices such as gender-play, where a user of one gender plays an avatar of the opposite gender or some third gender, are well-documented and pervasive in virtual environments. The virtual world promises that "you can be whoever you wish to be" - this might be part of the core appeal for current Social Internet users.

7. Further research

The prototyping, benchmarking and user studies done in the first iteration showed that the idea itself is viable. But it also showed that further iterations are needed to refine and test this idea. Future iterations of the prototype should include audio tracking and study how this affects user behaviour and perception. The question of input device design, whether it should ideally resemble the avatar or not, needs to be resolved through comparative user studies.

The current iteration of the Puppet UI captures hand and head movements well, but it is far less precise when it comes to facial expressions. Users are limited to moving and/or tilting the avatar's head, and to underline head movements with arm movements. Given the importance of facial expression in non-verbal communication, this may prove a disadvantage. This issue can be further explored by modifying our input device to work as a facial control interface, mapping points to eyes and mouth movements rather than head and hands, or by directly tracking the user's facial expressions.

It is possible that the puppet UI would not appeal to the typical social Internet users. It would break with already established practices regarding anonymity and textual meta-communication. We hope to test whether a puppet UI might enable new kinds of social Internet practices, and whether such practices are interesting or powerful enough to replace text-based practices.

References

Barrientos, Francesca and Canny, John 2001 'Cursive: A novel interaction technique for controlling expressive avatar gesture' in *UIST 01*, November 11-14, 2001, Orlando, Florida, pp. 151-152.

Burgoon, J. & Hoobler, G. 2002 'Nonverbal Signal' in M Knapp & J Daly (eds) *Handbook of Interpersonal Communication*, Sage Publications, London, pp. 241-299.

Donath, Judith S. 1999 'Identity and Deception in the Virtual Community' in P Kollock & M Smith (eds), *Communities in Cyberspace*, Routledge, London

Kurlander, D., Skelly, T., and Salesin, D. 1996 'Comic Chat', *Proceedings of the 23rd Annual Conference on Computer Graphics and interactive Techniques SIGGRAPH '96*. ACM Press, New York, NY, pp. 225-236.

Montola, Markus 2005 'Designing Goals for Online Role-Players' in S. de Castell & J. Jenson (eds): *Changing Views: Worlds in Play, Proceedings DVD of DiGRA 2005 conference,* Vancouver, Simon Fraser University.

Paiva, A., Prada, R., Chaves, R., Vala, M., Bullock, A., Andersson, G., and Höök, K. 2003, 'Towards tangibility in gameplay: building a tangible affective interface for a computer game', *Proceedings of the 5th international Conference on Multimodal interfaces* (Vancouver, British Columbia, Canada, November 05 - 07, 2003). ICMI '03. ACM Press, New York, NY, pp. 60-67.

Goldin-Meadow, Susan 1999 'The role of gesture in communication and thinking' in *Trends in Cognitive Sciences*, Vol. 3, No. 11, November, 1999, pp. 419-429.

Vilhjálmsson, H. H. and Cassell, J. 1998 'BodyChat: autonomous communicative behaviors in avatars', *Proceedings of the Second international Conference on Autonomous Agents* (Minneapolis, Minnesota, United States, May 10 - 13, 1998). K. P. Sycara and M. Wooldridge, Eds. AGENTS '98. ACM Press, New York, NY, pp. 269-276.

Ware, Colin 2004, *Information visualization: perception for design,* 2nd edition, Morgan Kaufmann, San Francisco

Virtual Worlds:

Second Life: http://secondlife.com/

World of Warcraft: http://worldofwarcraft.com/