Body Buddies: Social Signaling through Puppeteering

Magy Seif El-Nasr[1], Katherine Isbister[2], Jeffery Ventrella, Bardia Aghabeigi[1], Chelsea Hash, Mona Erfani[1], Jacquelyn Morie[5], and Leslie Bishko[6]

[1]Simon Fraser University
[2]New York University-Poly
[5]Emily Carr University
[6]University of Southern California

magy@sfu.ca, isbister@poly.edu, Jeffrey@ventrella.com, b.aghabeigi@gmail.com, saveremreve@gmail.com, morie@ict.usc.edu, lbishko@ecuad.ca

Abstract. While virtual worlds have evolved to provide a good medium for social communication, they are very primitive in their social and affective communication design. The social communication methods within these worlds have progressed from early text-based social worlds, e.g. MUDS (multi-user dungeons) to 3D graphical interfaces with avatar control, such as Second Life. Current communication methods include triggering gestures by typed commands, and/or selecting a gesture by name through the user interface. There are no agreed-upon standards for organizing such gestures or interfaces. In this paper, we address this problem by discussing a Unity-based avatar pupeteering prototype we developed called *Body Buddies*. *Body Buddies* sits on top of the communication program Skype, and provides additional modalities for social signaling through avatar pupeteering. Additionally, we discuss results from an exploratory study we conducted to investigate how people use the interface. We also outline steps to continuously develop and evolve *Body Buddies*.

Keywords: avatar pupeteering, avatar nonverbal communication, social communication with avatars, avatar design, CVE (Collaborative Virtual Environment)

1 Introduction

Mobile devices, including phones and PDAs, are becoming the dominant method for communication and an essential part of our everyday lives. Communication and social interaction have shifted from standard face to face modality to mediated social settings, such as email, Facebook, Skype videoconferencing, and increasingly, online multi-user virtual worlds. Despite the increased technical enhancements making such synchronous and asynchronous communication modalities possible, the design of synchronous online communication systems is still limited and primitive in terms of the affordances for social interaction and the affective communication they offer.

Extensive research in areas, such as communication and social psychology, has highlighted the significance of nonverbal behaviors, such as facial expressions, turn taking signals, and body language for communication [1-3], [4]. Current synchronous communication systems use one or more of four communication modes: text, audio, video, and avatar systems. Chat or text-based interfaces are limited since users cannot communicate intricate social messages, such as turn taking, or signals of skepticism or confusion. Recently, there are systems proposed or developed to combine video and audio signals, such as Skype calls, or PortaPerson [5]. Such modalities enable users to communicate and deliver synchronous social and affective messages through non-verbal behaviors within audio and video channels. These are successful solutions for one-on-one settings; however, there are issues that constrain the use of such systems within a group mode. First, with video alone it is impossible to use gaze direction as a communicative element between more than two people. Even with a setup as the one discussed in [5], it is still hard to efficiently use gaze; though it is widely recognized as an important communicative element [6], [7]. Second, spatial placement is hard to communicate through video, especially because people are not co-located in their communication space. A better method for enabling body position and proximity is to use virtual environments or spaces with an avatar representing the user so each person in a group is co-located virtually. Several researchers explored capturing gestures and developing avatars that imitate user gestures in virtual space [8], [9]. This approach has several limitations including methods of using or transferring proximity and spatial signals. In this paper, we take a different approach.

In the past year, we formed an interdisciplinary team composed of designers, developers, artists, a graphic designer, and a communication researcher to address this issue. We developed a Unity-based avatar puppeteering system, called *Body Buddies*, that sits on top of the Skype conferencing application, allowing Skype users to socially signal messages to one another in a simple virtual environment. Avatars can be adjusted to show like/dislike, skepticism, agreement, attention, and confusion, using dynamic movement rather than static poses. The system was first demonstrated and published at the CHI 2010 workshop on social connectedness [10]. In developing *Body Buddies*, we focused on allowing users conscious control of various puppeteering parameters. This approach has various advantages and disadvantages. First, it requires the user to consciously make a signal. Second, it adds cognitive load on the user as they take the burden to communicate these signals when needed. However, using this type of interface allows users more control over their signaled behaviors. It also may alleviate video camera issues, such as users not wanting their image to be projected or feeling nervous in front of a camera [11].

In this paper, we discuss the system and its puppeteering interface. In addition, we discuss preliminary results of a study we conducted, in which we asked users to discuss and debate a particular topic using *Skype* and *Body Buddies*. We conclude the paper by discussing future research.

2 Previous Work

Previous work within this area spans multiple disciplines. We outline the following areas: *nonverbal behavior in real life*, for which we devote a section discussing models and taxonomies proposed. It should be noted that in the interest of space, we only summarize some important contributions here, highlighting what models transfer to Virtual Worlds (VWs). The second relevant area is: *avatar based online communication environments*. Although there has been little work in this area, we highlight some of the significant work here. Some of this work includes systems and/or studies of how people used avatar based nonverbal communication modalities within a virtual environment. These studies are an important corner stone to our work.

2.1 Nonverbal Behavior

The study of nonverbal behavior in the real world has received much attention, including the study of proximity, emotional expressions, and gesture. Hall and Birdwhistell [1], [12], [13] are considered the fathers of the study of *Proxemics* and *Kinesics,* respectively – two of the very important and dominant paradigms of nonverbal communication dealing with different aspects of the human body.

Hall's work on *Proxemics* discusses the notion of personal space, describing several zones of intimacy around the body. Over its 60 year history, Proxemics has been used to describe how people position themselves in space relative to each-other, and how different demographic factors alter these spatial behaviors. Recent studies in Virtual Worlds (VWs) discussed evidence found supporting the translation of real world proxemic and gaze behavior to virtual worlds [6], [14], [15]. Additionally, Yee et al. also report the presence of social norms governing the proxemic behaviors within virtual worlds resembling those of the real world [16].

Kinesics, which is the study of gesture and posture, has also received attention. In addition to the structural model developed by Birdwhistell [13], several researchers investigated a descriptive approach. Ekman and Friesen [2] present an exhaustive description of the types of non-verbal behavior that people perform. They discuss different types of acts, such as *emblems*: culture specific, learned behaviors that represent meaning, *illustrators*: socially learned behaviors that complement or contrast verbal messages, *affect displays*, *regulators*: conversational flow gestures that control the back and forth within a dyad, and *adaptors*: learned actions based on satisfying bodily needs, based on child-hood experience. This model has been used by several researchers within the HCI field [8].

Additionally, there has been much work on the use of gesture in speech and communication. An important work in this area is the work of McNeill and Cassell [4], [17], [18], who explored the use of communicative gestures by observing and analyzing people talking about specific subjects, such as real estate, etc.

2.2 Avatar-based nonverbal communication within online meeting environments

Several researchers empirically investigated the communicative power of nonverbal behaviors within virtual environments. In a study conducted by Allmendinger, they compared conditions with video, audio, inferred-gaze avatar, and random-gaze avatar. They found that video was most favored followed by inferred-gaze avatar system [11]. This confirms the role of gaze in nonverbal communication as discussed in previous work [6], [7]. Additionally, automated gaze within avatar groups were explored and implemented in the socially-focused virtual world There.com. Through in-house user testing, such use of gaze was found to significantly increase users' sense of social engagement [3]. In addition to gaze, Allmendinger argued that avatars can provide cues to support (a) *group awareness*, such as focus attention and position in an argument, as well as (b) *communicational gestures*, such as signals to identify who is talking [11].

Empirical work exploring the design of such avatars is sparse, although there are some. Anderson et al. presented a study combining testing and participatory design to infer the usability and presence of avatar systems. Their experiments showed that users needed a level of control on avatar animation to show who is talking and support control and turn-taking [19]. Similarly, Guye-Vuilleme et al. [20] stressed the use of agreement and space in avatar design; they deduced these results through a qualitative experiment with a CVE (Collaborative Virtual Environments). This is important as turn-taking in distributed synchronous environments is seen as a problem area [21], [22]. Allmendinger et al.'s study confirmed these results by concluding that important signals for avatars in CVEs were: *thumb up*, *gestures highlighting information on slides*, and *turn taking signals* [23].

The work on developing virtual meeting spaces can be grouped into two groups: sensor-based intelligent environments, where gestures are entered through devices, camera or other sensors and are transferred to an avatar model [24], [25], and lightweight interactive virtual environments, such Lucia et al.'s *SLMeeting* [26] which supports collaborative and management activities within a web based interface, and *Porta-Person* [5], which enhances the sense of social presence through a controlled display of a video image of the participant. Another example of lightweight interactive virtual meeting system is Shami et al.'s Olympus [27], which is a flash-based light-weight virtual meeting place that allows chat based input, and can link specific commands to gestures. For example, the character can shrug based on the text '?'. Similar, to WOW interface, users can type specific gestures by typing '/' then the name of the gesture animation. They tested this system in three meetings to assess its effectiveness. In terms of the use of gesture, they found that in meetings users used a combination of gesture and chat in general. The most popular three gestures, confirming previous research, were: clap, agree, and wave. They, however, concluded that users did not move their avatars. Our work extends the work discussed here to present a new avatar puppeteering system and test its interface.

3 Body Buddies

The architecture of the Body Buddies system is shown in figure 1. The system consists of two components. The avatars along with the virtual environment were developed in Unity. The Unity-based system is developed to augment Skype voice interaction with avatar based controls. It is activated once the user logs into the Skype program. The Skype – Unity interaction is implemented using a middleware protocol, which acts as a mediator between the Unity client avatar system and the Skype program, transferring messages from Skype to Unity and vice versa. As shown in the diagram, the middleware application communicates with Unity through TCP/IP protocol and with the Skype client through an ActiveX component representing Skype API as objects called Skype4Com. The middleware first establishes a connection between Unity and Skype through a handshaking routine. This is done for all users within a Skype call. The middleware application then uses the application-to-application protocol command in the Skype API, *AP2AP*, to send Unity messages from one Skype client to other peers in the conference call. For example, when a user activates a gesture command for an avatar in a client, the Unity side sends a message to the middleware using TCP/IP protocol, and the mediator uses the *AP2AP* commands for sending this animation change to other peers in the Skype conversation. On the other side of the communication, the Skype client communicates with the middleware side using a Skype4Com registered call back, and then the application sends the animation command to the other Unity part by TCP messaging system. Finally, the Unity side parses the animation command, and executes the related avatar changes.

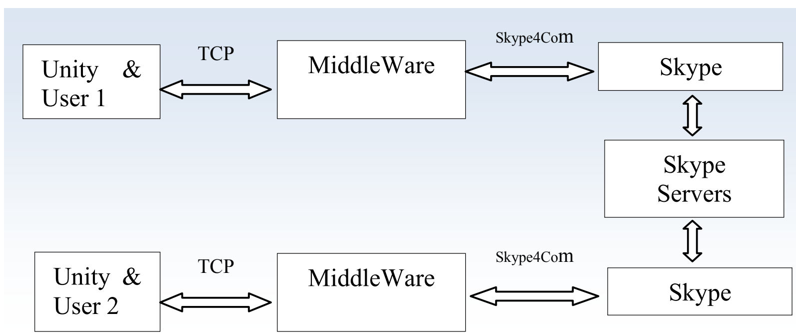


Fig. 1. Body Buddies Architecture.

Once the user logs in, the Unity-based interface shown in figure 2 appears showing the user the other avatars with the Skype name for each avatar displayed above its head. The user can interface with the avatar through the buttons shown in figure 2, where he/she can move, rotate, lean the avatar forward or backward, as well as execute the social gestures ‘skeptical’ and ‘my turn’.

The avatars in the *Body Buddies* system use a hybrid set of techniques. The 3D representations were modeled and rigged in Maya, and then imported into the Unity game engine, along with accompanying short-lived full-body gestural and postural animations. The UI controls for moving the avatar (forward, backward left and right) as well as *skeptical* and *My Turn!*, shown in figure 2, were linked to the Maya animations developed for the avatars, thus allowing users to trigger animations in real-time. In addition to these triggered animations, controls were implemented allowing the user’s avatar root position and heading to be adjusted – permitting a rudimentary form of navigation. This allowed users to shift the positions of the avatars in relation to each other and also to face towards or away from each other – for the purpose of social signaling.

Fig. 2. Body Buddies interface developed in Unity. For a full video of a demo see: http://www.sfu.ca/~baa17/SkypeBuddy/DemoLeslie.avi

In addition, we also added modifiers to the avatar joint rotations allowing the user to adjust parameters such as *Arch forward* and *Arch backward* [28]. These were procedurally-generated postural stances involving several joints. These procedural modifications were layered on top of the avatar joint array such that they could be smoothly blended with any imported animation playing simultaneously. The blending of postural and gestural movement created a palette of body signals that the user could combine in a variety of ways.

4 Study

To investigate how users interacted with the avatar system, we ran a study with 9 groups of 2-3 participants. Unfortunately, due to technical difficulty and problems with videos we had to disregard data from 3 groups; thus, we analyzed only 6 groups for a total of 11 participants.

4.1 Procedure

Participants were invited for a debate session in the lab in pairs. Once they arrived, they were asked to sign a consent form and then asked to complete a survey designed to measure their social connectedness. We then took each participant to a different room equipped with a laptop or desktop computer running *Skype* and the *Body Buddies* system. We then asked each participant to discuss a given topic (social network and Facebook) using the Skype and Body Buddies interface. We did not enforce any specific interface use during the session, leaving them to chat freely using the given tools. We video taped their interaction session for later analysis. We also logged all their actions, including button presses, the amount of time the Unity window was active, their button-pushing frequency, etc. This interaction session lasted between 15-18 minutes. After the session, we asked participants to fill out a questionnaire and the social connectedness survey again.

4.2 Results

Figure 3 shows the total time vs. the time spent using the avatar interface. Our results show that the users employed the avatar interface considerably more than any of the other interfaces. This result is statistically significant.

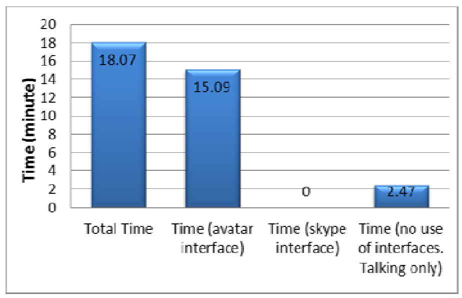
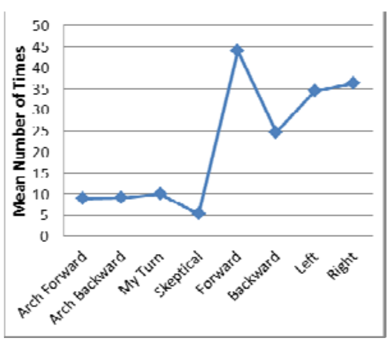


Fig. 3. Average time (error bars and bar chart)

Results from the before and after social connectedness tests show that IOS Scale upgraded on average from mean IOS Scale (before) = 3.53 to mean IOS Scale (after) = 4.4. While on average there is a difference, it was not significant — an expected result as 15 minutes is too short to cause a major improvement on social connectedness. However, interacting with avatars may improve social connectedness in the long run.



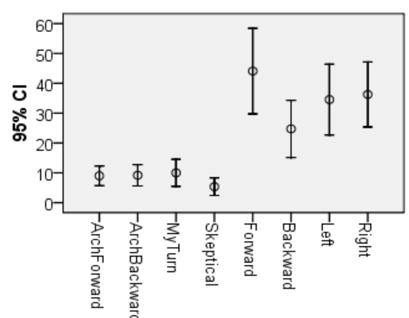


Fig. 4. Button press analysis. The figure on the right shows averages and the figure on the right shows error bars calculated from Standard Error given the sample.

The investigation of the interface use led to interesting results. The interface has 8 different buttons (Forward, Backward, Left, Right, Arch Forward, Arch Backward, My Turn and Skeptical). Analysis of the times these buttons were pushed, shown in figure 5, showed that participants used mostly the movement keys (Forward, Backward, Left and Right). It should be noted that the movement buttons were counted differently, as we only counted the first button press of a sequence of button presses. This is because users will probably press several times to move the avatar, to a specific place. We counted this as one event so as not to skew the data. We divide the eight buttons into two groups: movement buttons (Forward, Backward, Left and Right) and other buttons (Arch Forward, Arch Backward, My Turn and Skeptical). We ran Mann-Whitney U Test to determine any significance between the use of these two groups of buttons. The results show significance ($p<.05$). Participants pushed the movement buttons four times as much as the other buttons, which is an interesting result as it is in conflict with the other results from the literature discussed above, where Shami et al. [27] deduced that participants did not move avatars at all within meetings. Figure 4 shows the error bars on the average number of times of button presses and the average number of times of button presses over different actions.

In addition to this quantitative analysis, we also looked at the qualitative feedback given by users. First, some participants were enthusiastic with the addition of another layer of expressiveness, as one said, “being able to visually interact with other Skype users through emotions is great, when you do not wish to use your camera or do not have access to one.” But some expressed concerns, such as “it is difficult to concentrate on both moving the avatar around and talking at the same time.”

5 Conclusion and Future Work

The goal of this project was to investigate types of interfaces that could support better communication within computer-mediated meetings. We found, similar to previous work, that some affordances for avatar puppeteering were used more than others. Unlike previous work, we found that users used movement the most. However, the study is limited in several ways. The study was conducted in a lab setting with undergraduates. We think this limits the result as the behavior of participants in a real meeting versus a made up scenario will be different. Overall we saw participants were more interested in ‘playing’ with the system than communicating, perhaps due to its novelty or due to the setting itself. Therefore, we believe as we move on to a different setting for testing, the use of real meeting environments will be necessary to understand and investigate the use of nonverbal communication mediated by avatars.

Understanding the communicative affordances for the system is an interesting and complex problem. We suggest several future directions towards achieving this goal, including adding other social signals such as expressive emotions: ‘happiness’, ‘sadness’, etc., expressive confusion, greetings, thumb up and thumb down which previous literature has noted as important. We also hope to engage in additional investigations with the use of body animations, gestures, and postures as techniques for expressing these variables, and use of other devices beyond keyboard and mouse.

7 Acknowledgements

We would like to thank GRAND (Graphics Animation and New Media) – NCE (Network of Centers of Excellence) for making this work possible.

8 References

[1] E. Hall, "Proxemics," *Current Anthropology*, vol. 9, Jan. 1968, p. 83.
[2] P. Ekman and W. Friesen, "The repertoire of nonverbal behavior: Categories, origins, usage, and coding," *Semiotica*, vol. 1, 1969, pp. 49-98.
[3] J. Ventrella, *Virtual Body Language*, Eyebrain Books, 2011.
[4] D. McNeill, *Gesture and thought*, University of Chicago Press, 2005.
[5] N. Yankelovich, N. Simpson, J. Kaplan, and J. Provino, "Porta-person," *CHI '07 extended abstracts on Human factors in computing systems - CHI '07*, San Jose, CA, USA: 2007, p. 2789.
[6] J.N. Bailenson, A.C. Beall, and J. Blascovich, "Gaze and task performance in shared virtual environments," *The Journal of Visualization and Computer Animation*, vol. 13, 2002, pp. 313-320.
[7] A.R. Colburn, M. Cohen, and S.M. Drucker, *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*, Microsoft Research, 2000.
[8] A. Vinciarelli, M. Pantic, H. Bourlard, and A. Pentland, "Social signal processing: state-of-the-art and future perspectives of an emerging domain," *Proceeding of the 16th ACM international conference on Multimedia*, New York, NY, USA: ACM, 2008, pp. 1061–1070.
[9] A. Vinciarelli, M. Pantic, and H. Bourlard, "Social signal processing: Survey of an emerging domain," *Image and Vision Computing*, vol. 27, Nov. 2009, pp. 1743-1759.
[10] K. Isbister, M. Seif El-Nasr, and J. Ventrella, "Avatars with Improved Social Signaling.," *CHI 2010 Workshop on Designing and Evaluating Affective Aspects of Sociable Media to Support Social Connectedness*, 2010.
[11] K. Allmendinger, "Social Presence in Synchronous Virtual Learning Situations: The Role of Nonverbal Signals Displayed by Avatars," *Education Pyshcology Review*, vol. 22, 2010, pp. 41-56.
[12] E.T. Hall, *The Hidden Dimension*, Anchor, 1990.
[13] R.L. Birdwhistell, *Introduction to kinesics: An annotation system for analysis of body motion and gesture*, University of Michigan Library, 1952.
[14] J. Bailenson, A. Beall, J. Blascovich, M. Raimundo, and M. Weisbuch, "Intelligent Agents Who Wear Your Face: Users' Reactions to the Virtual Self," *Intelligent Virtual Agents: Third International Workshop, IVA 2001, Madrid, Spain, September 10-11, 2001, Proceedings*, 2001, p. 86.
[15] D. Friedman, A. Steed, and M. Slater, "Spatial Social Behavior in Second Life," *Intelligent Virtual Agents*, Paris, France: Springer-Verlag, 2007, pp. 252-263.
[16] N. Yee, J. Bailenson, M. Urbanek, F. Chang, and D. Merget, "The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments," *CyberPsychology & Behavior*, vol. 10, 2007, pp. 115-121.
[17] J. Cassell, C. Pelachaud, N. Badler, M. Steedman, B. Achorn, T. Becket, B. Douville, S. Prevost, and M. Stone, "Animated conversation: rule-based generation of facial expression, gesture \& spoken intonation for multiple conversational agents," *Computer Graphics*, vol. 28, 1994, pp. 413-420.
[18] J. Cassell, "A Framework For Gesture Generation And Interpretation," Computer Vision

in Human-Machine Interaction, pp. 191--215.

[19] J. Anderson, N. Ashraf, C. Douther, and M.A. Jack, “Presence and Usability in Shared Space Virtual Conferencing: A Participatory Design Study,” *CyberPsychology <html_ent glyph="@amp;" ascii="&"/> Behavior*, vol. 4, 2001, pp. 287-305.

[20] A. Guye-Vuilleme, T.K. Capin, S. Pandzic, N.M. Thalmann, and D. Thalmann, “Nonverbal communication interface for collaborative virtual environments,” *Virtual Reality*, vol. 4, 1999, pp. 49-59.

[21] J. Bowers, J. Pycock, and J. O'Brien, “Talk and embodiment in collaborative virtual environments,” *Proceedings of the SIGCHI conference on Human factors in computing systems common ground - CHI '96*, Vancouver, British Columbia, Canada: 1996, pp. 58-65.

[22] A. Lantz, “Meetings in a distributed group of experts: comparing face-to-face, chat and collaborative virtual environments,” *Behaviour & Information Technology*, vol. 20, 2001, pp. 111-117.

[23] K. Allmendinger, “Social Presence in Synchronous Virtual Learning Situations: The Role of Nonverbal Signals Displayed by Avatars,” *Education Pyshcology Review*, vol. 22, pp. 41-56.

[24] E. Frecon and A.A. Nou, “Building distributed virtual environments to support collaborative work,” *Proc. VRST 1998*, 1998, pp. 105-113.

[25] C. Greenhalgh and S. Benford, “MASSIVE: a collaborative virtual environment for teleconferencing.,” *ACM Trans. Comput.-Hum. Interact.*, vol. 2, 1995, pp. 239-261.

[26] A.D. Lucia, R. Francese, I. Passero, and G. Tortora, “SLMeeting: supporting collaborative work in Second Life,” *AVI 2008*, 2008, pp. 301-304.

[27] N.S. Shami, L. Cheng, S. Rohall, A. Sempere, and J. Patterson, “Avatars Meet Meetings: Design Issues in Integrating Avatars in Distributed Corporate Meetings,” *Group*, 2010.

[28] C. Moore, “Movement And Making Decisions: The Body-mind Connection,” Dance & Movement Press, 2005.