Considering Talk and Emotion when Creating and Deploying Realistic 3D Avatars

Steve Jones
University of Illinois at Chicago
Gordon Carlson
Fort Hays State University

Author Note

This research was funded by grants from the National Science Foundation, Funding: #CNS-0703916 and #CNS-0420477. The authors wish to thank anonymous reviewers for their helpful suggestions for revision.

Abstract

We outline the setup and results of three studies conducted to assess Project LifeLike, a trans-disciplinary collaboration that investigates, develops and evaluates lifelike, natural computer interfaces that incorporate dialogue and nonverbal communication. The goal of Project LifeLike is to provide a natural interface that supports transferring knowledge over time using realistic spoken dialogue and nonverbal cues. The first study, Significant Aspects of Realism and Utility in 3D Avatars, builds on a previous pilot experiment and compares the relative importance of four key dimensions of nonverbal communication. It also compares a first, baseline avatar with a more recent one created as part of LifeLike. The second study, Intelligent User System Avatars Field Experiment, is a field test of the avatar as implemented by LifeLike. Members of the target audience at a National Science Foundation conference used the system and answered questions about the avatar's utility and compared it to their experiences with real people at the NSF. The third study, Comparative Emotive Capabilities of 3D Avatars, assesses the ability of the LifeLike avatar to accurately display and indicate emotional states. Results showed that the avatar is capable of successfully indicating two emotions, happiness and sadness, but not necessarily anger, disgust, surprise, or fear. Based on the subjects' comments it appears that shoulders, arms, and hands are more important, in terms of movement, than an avatar's head. Subjects did not show a clear preference for computer-generated voices or recorded voices was not identified.

Keywords: Avatars, lifelike interfaces, nonverbal communication, emotion display

Considering Talk and Emotion when Creating and Deploying Realistic 3D Avatars

Project LifeLike is a trans-disciplinary collaboration that investigates, develops, and evaluates lifelike, natural computer interfaces as portals to intelligent programs in the context of Decision Support Systems (DSS). Funded by the National Science Foundation (NSF)[1], the goal of the project is to provide a natural interface that supports realistic spoken dialogue and nonverbal cues in transferring knowledge over time. Research objectives focus on the development of an avatar-based interface with which a DSS user can interact. Communication with the avatar occurs in spoken natural language combined with gestural expressions or pointing on a screen. A database driven information system responds intelligently to the questions asked by DSS users with spoken responses by the avatar using realistic inflection and visual expressions.

This article outlines the setup and results of three studies conducted to assess Project LifeLike specifically and lifelike avatars more generally. The first study, *Significant Aspects of Realism and Utility in 3D Avatars*, builds on a pilot experiment and compares the relative importance of four key dimensions of nonverbal communication. It also compares a baseline avatar with the most recent one created as part of Project LifeLike. The second study, *Intelligent User System Avatars Field Experiment*, is a field test of the avatar as implemented by Project LifeLike. Members of the target audience at the National Science Foundation Industry-University Cooperative Research Annual Conference used the system and answered questions about the avatar's utility and compared it to their experiences with real people at the Foundation. The third study, *Comparative Emotive Capabilities of 3D Avatars*, assesses the ability of the LifeLike avatar to accurately display and indicate emotional states. Based on the work of Paul Ekman, the avatar's facial characteristics were manipulated to recreate emotive faces of the human who served as the avatar's inspiration.

The recent proliferation of inexpensive computer technologies has created an insatiable demand, by consumers, for quick and easy access to information. While software and websites have offered users the ability to query textual information for many years, a more gratifying approach is sought in many sectors of technological development. A current research interest in

many fields is the use of avatars, graphical representations of human beings designed to simulate the oral-aural and visual interaction between human beings in conversation (Kshirsagar, Magnenat-Thalmann, Guye-Vuillème, Thalmann, Kamyab, & Mamdani, 2002). An important role of research is to determine those qualities that make an avatar more believable or more useful (Tanikawa, Suzuki, Hirota, & Hirose, 2005; Garau, Slater, Bee, & Sasse, 2002; Lee, Chai, Reitsma, Hodgins, & Pollard, 2002). This study addresses several aspects of nonverbal communication that could be valuable in the pursuit of realistic and useful avatars.

In order to simulate a human and provide a means of simulated social interaction, an avatar must engage in many of the same nonverbal cues as people. While simulating a conversation with a human, an avatar must choose appropriate cues and indicators because it is difficult to separate verbal communication from its nonverbal counterparts (Kendon, 1983). When humans communicate, eye contact can be used to convey emotion and rapport (Tiemens, 1978). An avatar's amount of eye contact can be controlled and the amount of gaze an avatar provides will likely impact the rapport a user experiences when interacting with it (Garau, Slater, Vinayagamoorthy, Brogni, Steed, Sasse, 2003). Facial expressions, much the same as eye contact, can convey emotion, indicators of honesty or sincerity, and context for verbal communications (Ekman, 1982; Knapp & Hall, 2002). Body movement is recognized as a powerful communicative tool and vocal behavior (*how* something is said rather than *what* is said) can directly impact understanding and comfort for persons engaged in conversation. Voice style including paralinguistic and extralinguistic acoustic elements are important as well (Scherer, 1982; Hall, Roter, Rand, 1981), have been seriously studied for a number of years (Poyatis, 1993; Trager, 1958), and can be studied in useful units (Scherer, 1986) applicable to avatar research.

Literature Review

Interpersonal Communication.

Ancient Greeks and Romans who investigated communication realized that it was not just what a person said, but also how a person said it and what that person might look like (Aristotle, 1984; Plato, 1952; Plato, 1956) that constituted communication. Modern scholars have further

---

[1] Funding: #CNS-0703916 and #CNS-0420477

recognized the visual component is innate to the needs of producers and consumers of communications (Burke, 1969). Many psychologists and communication scholars even believe that when body language and verbal messages conflict, people are apt to believe the nonverbal elements rather than the verbal ones (Conniff, 2004; Beebe & Masterson, 2000). The emphasis placed on nonverbal communication has led to the creation and splintering of several fields dedicated to the study of how people communicate visually and physically.

Contemporary approaches to nonverbal communication typically focus on several types of characteristics: facial movements including the eyes, body movement with special emphasis on posture and the hands, and the various sounds employed by the speaker (rate, pitch, volume, etc.). The usefulness and overall utility of nonverbal components are constrained by social norms within a society, whether codified or not. These norms typically define appropriate and inappropriate behavior (Goffman, 1967) which, in turn, affects the quality of communication: nonverbal communication can be the determining factor in the quality and success of a given communication act.

A specific form of nonverbal communication that attempts to encompass many of these ideas holistically is known as mirroring. A common practice (Bavelas, Black, Chovil, Lemery, & Mullett, 1988), mirroring is the act of matching posture and body language with other members of a communicative dyad (LaFrance, 1979, 1982; LaFrance & Broadbent, 1976). This is not only common but often unconscious (Chartrand & Bargh, 1999; Chen, Chartrand, Chai, & Bargh, 1998). Van Swol (2003) conducted a study that resulted in quantitative support for the hypothesis that "people in a group will judge interactional partners who mirror their nonverbal behavior as more persuasive than partners who do not mirror their nonverbal behavior" (p. 464). Mirroring, as a method for creating and maintaining rapport and competent communication, is a specific application of a more general idea: nonverbal communication and paralanguage have direct and important impacts on persuasion and information conveyance.

Avatars as Interfaces.

With the development of Internet-based technologies and the increasingly common demand for instant gratification, new approaches to software interfaces are being developed. One approach is the use of avatars as front ends for computer software (Kandogan, Krishnamurthy, Raghavan, Vaithyanathan, & Zhu, 2006). Avatars began as digital

representations of people used in computing environments: “Our Avatars are 3D models that can represent us to communicate with others and objects in the environment” (Yarning, Juner, & Bin, 2004, p. 116). As avatars become more than simple shapes and move toward metaphorical people they begin to represent *real* people in *virtual* space and “take the user from the real world to the virtual world” (Yin, Yang, Wen, Lai, & Shen, 2006, p. 2). Avatars provide the ability to merge reality with the computer to the point that communication between a human and a computer shifts from metaphor to actual communication, from simple interaction to social interaction. The ultimate goal of avatar creators is to create one which would pass a Turing test (Turing, 1950), a subjective test whose goal is to determine whether a user can tell the difference between the avatar and a real person.

Avatars Mimic Real Persons.

As more avatars are created and demands for realism increase, the specific elements important to creating a realistic person in the digital realm are of prime importance.

> Especially intriguing and elaborate studies on norms in digital visual spaces have uncovered systematic similarities between offline and online space norms. For example, a study by Yee et al. (2007) found that offline personal-space norms apply among avatars in the virtual world Second Life. By calculating the head placement and placement of avatar dyads, they found that female pairs tend to stand closer to others and maintain more eye contact, whereas male pairs tend to stand farther away with less eye contact. In addition, they found that men tend to stand farther away from each other in outdoor versus indoor visual settings online. In similar research, Becker and Mark (2002) found that offline norms, such as proximity and use of private space, are stronger when settings and avatars are more realistic in online settings. These findings emphasize the importance of both avatars and visual settings to social interaction online. (Marty, 2007, p. 316)

The pervasiveness of avatars only amplifies the importance of improving all aspects of communication with them; with contemporary text-to-speech synthesis performing very well, nonverbal elements demand the most scrutiny. As Cassell, *et. al.*, have noted, the use of “embodied interface agents can provide a qualitative advantage over non-embodied interfaces,

if the bodies are used in ways that leverage knowledge of human communicative behavior" (Cassell, *et. al*, 2001).

Avatar facial expression, its significance and implementations, has been addressed by some researchers (Zhan, Li, Safaei, & Ogunbona, 2007; Salem & Earle, 2002) but these studies have dealt mostly with technical implementation (Kalra, Gobetti, Magnenat-Thalmann, & Thalmann, 1993) and not communication theory. One approach has been to implement the Facial Action Coding System (Ekman & Friesen, 1978) that attempts to code natural human facial motions and map them to the faces of avatars. Another approach has been to distinguish between fidelity (how much an avatar looks like a real person) and immersiveness (whether a user is engrossed by the system) (Ducheneaut, Wen, Yee, Wadley, 2009).

Most producers tend to bias towards fidelity; they create avatars that look real, but do not necessarily act real. This is an admirable yet daunting task. It may be better to select specific elements to support fidelity and specific elements to support immersion. The only way to select these elements is to determine the relative importance of each in constructing dyadic relationships between avatars and users. Implementations commonly use articulation, where a human is used as a model to create the avatar and a computer augments the model with movement information (Baddler, Phillips, & Webber, 1999). With this emphasis on creating realistic people, the most common aspect that has been studied is eye contact, or gaze (Garau, Slater, Bee, & Sasse, 2001; Salem, Earle, Argyle, & Cook, 1976; Bowers, Pycock, & O'Brien, 1996; Kendon, 1967). With only sporadic attention paid to individual elements of nonverbal communication and paralanguage, the field of avatar development lacks an important piece of information: Which elements of nonverbal communication and paralanguage are most important in constructing believable avatars in digital settings?

Conceptual Definitions.

Avatars have traditionally been any graphical representation of a person. Contemporary use of the term avatar in scholarly contexts typically refers to front-end user interfaces for computer software (Kandogan, Krishnamurthy, Raghavan, Vaithyanathan, & Zhu, 2006) and are often three dimensional recreations of actual persons that can be highly detailed (Yarning, Juner, & Bin, 2004). Current avatar technology allows software designers to represent real people in

virtual space and "take the user from the real world to the virtual world" (Yin, Yang, Wen, Lai, & Shen, 2006, p. 2).

A major concern for developers of avatar systems is realism. The basic approach to creating realistic avatars is to employ various physical measures of interpersonal communication to enhance a sense of social relations. To this end there are four specific elements that must be defined in the context of computer driven avatars: gaze, voice quality, head motion, and body motion.

Eye contact, or gaze, is the eye movement we make in the general direction of another's face (Knapp & Hall, 2002) and is important because humans use eye movements to read emotion (Ekman & Friesen, 1975) and gain context. Gaze is also important because humans look at places that reflect their cognitive processing (Kaur, Tremaine, Huang, Wilder, Gacovski, Flippo, et al). Gaze, therefore, represents what a person is focusing on both physically and mentally.

Voice quality is the combination of rate, loudness, variety, and pitch in the verbal presentation of information by a person (Osborn, Osborn, & Osborn, 2009). In terms of computer generated speech (most commonly referred to as text-to-speech or TTS) each of these are variables that can be controlled. TTS voices are difficult and expensive to create so there are not many of them, but by manipulating each variable it is possible to modify a voice to sound more or less like a person. An alternative to TTS systems is to pre-record a real person reading the information a designer would like an avatar to say.

Realism is difficult to assess objectively and there is little in the literature regarding a *definition* of what is realistic. Certainly it incorporates looking like a real person, but more specifically it must incorporate the identifiable nuances of a person or persons such as movement and appearance (Salem & Earle, 2000). For the purpose of our studies, realism will be measured in terms of how effective an avatar's articulation is in terms of movement and appearance as judged by participating subjects.

Because the literature makes so plainly clear the significance of eye contact, physical movement, and voice in the interactions of humans, and because avatars seek to create human-like communication scenarios, it seems that body motion, head motion, and voices will play vital roles in the quality of interactive avatars.

User Studies

Study One: Significant Aspects of Realism and Utility in 3D Avatars.

Goals.

To study the user acceptance of an avatar in this environment, we conducted a study to determine which communication elements are most important in creating realistic and useful avatars as interfaces for question and answer systems by focusing on specific paralanguage variables: gaze (eye contact), head motion, body motion, voice, and a baseline comparison of the current version of the avatar and a version from a year prior.

Procedure.

The approach employs a classic experimental design model and is based on Koon's (2006) and Garau's work with user interface and avatar design: within each specific experimental trial (called a pairing) an independent variable is manipulated in hopes of causing a direct result in a dependant variable.

Subjects sit in a chair at a conference table. The wall in front of them displays video clips on a large projection screen (over 70" in size). The avatar in each video clip is specifically scaled to appear life-sized on the wall. A subject is shown two videos which keep all attributes constant except for one key variable. There are five pairs in order to track each independent variable: head motion, body motion, voice, gaze, and the baseline comparison. The order of the pairings and the order of the videos within the pairings are randomized. The videos (5 pairs, 10 total) are each approximately 30 seconds long. They are developed in such a way that all the elements of the video not being tested do not change (same text spoken, same appearance, same environment, etc). The total time of engagement for each subject ranged from 28-45 minutes depending on how much time they spent responding to the instruments.

Subjects did not know what they were looking for before each video was shown and were given numerical assessment Likert scales ranging from 1-7. This is a useful tool for quantifying and analyzing the subjects' responses with regard to realism and satisfaction in computer interfaces (Epps and Close, 2007). After each pair of videos was shown, an additional instrument was provided which asked open-ended questions to compare and contrast the videos with one another with regard to the isolated variable for that pair. This allows researchers to

identify biases and extra results useful in later refining the avatar (Wacker, Stoev, Keckeisen, Straßer, 2003). Each assessment was identical so that the independent variable being isolated was not evident to the subjects before watching each video pair. The results indicate what impact each independent variable has on avatar realism and utility. By revealing the isolated variable before asking the open-ended questions the subject is allowed to reflect on the experience and critically assess each video in terms of the isolated variables providing rich qualitative data.

The sample (n=30) consisted of students representing diverse backgrounds (9 females, 13 ESL, ages 18-46 with 17 subjects between 20-25). Recruitment materials announced that one person would be randomly selected to receive an iPod valued at 100.00 US dollars.

We hypothesized that: (**H_1**) Subjects would report greater realism using the latest avatar generated by the NSF project than the initial baseline avatar; (**H_2**) Subjects would report a greater significance in poor body motion than they will in poor head motion; (**H_3**) Subjects would place greater emphasis, with regard to realism, on body movement than on voice. These hypotheses support a more general research question: *which elements are most important when creating realistic and useful avatars as front-ends for database driven software applications?*

Results.

The quantitative results were analyzed primarily through paired T-Tests and by cross tabulating the data to determine if extra factors played a role. Cross tabulation highlights relationships between subjects and their response (e.g., sex, age, experience with avatars). There are seven questions within each pair of videos. The open ended questions were analyzed through a non-rigorous content analysis.

The two most significant variables were body motion and the baseline test. Subjects clearly preferred the movement of the motion-captured avatar compared to an avatar with little body movement. All seven questions showed a preference for the motion; paired t-tests showed significant shifts in the mean responses on the Likert questions (ranges from 0.633 to 1.533) with all seven having two tail statistical significance values better than ($p < .011$).

Even stronger results were found within the baseline comparison pairing; all seven means shifted strongly in favor of the newer avatar (ranges from 1.53 to 2.33) with two tail statistical significance scores of ($p < 0.01$) or better. The baseline test compares two versions of the avatar. The first was created roughly a year before the second revision. The early one lacks structured or

purposeful motions while the second incorporates motion capture data from the real life basis for the character. The Text-To-Speech (TTS) system of the first avatar is several years old while the newer avatar incorporates the latest TTS voice technology and runs on the most recent version of Microsoft SAPI, a voice synthesis engine. Finally, the textures, model, and background of the newer avatar are more detailed and precise creating a more compelling and realistic looking character. These changes represent significant improvements in image, movement, and sound. Subjects prefer body motion to a still avatar and there is strong evidence that our work over the last year has yielded a substantially improved avatar.

Intriguing but inconclusive, there was a statistical tie in preference between a TTS and pre-recorded content. Results indicated that 13 preferred TTS and 15 preferred a pre-recorded voice (1 reported an exact tie); 14 thought TTS was more realistic and 13 thought pre-recorded was more realistic (2 reported exact ties, 1 failed to report). Reasons for each preference were varied and complex indicating a strong need for further research into this area, see tables 1 and 2.

Sex of the subject * Which Voice Video is More Realistic?
Crosstabulation

Count		Which Voice Video is More Realistic?		
		TTS	Recorded	Total
Sex of the subject	Male	10	8	18
	Female	4	5	9
	Total	14	13	27

Table 1. There was an even split among subjects in terms of *perceived realism* of TTS and recorded voice. Sex was not a factor.

English is their first language * Which Voice Video Did You Prefer?
Crosstabulation

Count		Which Voice Video Did You Prefer?		
		TTS	Recorded	Total
English is their first language	English	8	8	16
	ESL	5	7	12
	Total	13	15	28

Table 2. There was an even split among subjects in terms of *preference* for TTS versus recorded voice. Sex was not a factor.

The qualitative responses indicate high interest in the project and the avatar used in the research. Further, most subjects reported using avatars in their everyday lives indicating the pervasiveness of the technology. **H_1** was clearly confirmed: subjects reported greater realism using the latest avatar generated by the NSF project than the initial baseline avatar. The qualitative data indicates that higher resolution images and bump maps may be important in creating realistic avatars. **H_2** was also upheld though not as strongly. Subjects did report a greater dissatisfaction in poor body motion than head motion. **H_3** was indicated but not conclusive. While subjects did indicate that body motion was important there was no clear preference for recorded versus TTS voice. This makes the hypothesis hard to prove but yields numerous questions ripe for further investigation.

Study Two: Intelligent User System Avatars Field Experiment.

Goals.

The purpose of this study was to determine which characteristics of conversational communication are most important when interacting with avatars in the context of a NSF program with special emphasis on how it is received by the organization's employees. The study tests facial expressions, voice, wording, and technical implementation. Essentially our first field test, the case study sought to improve the LifeLike avatar based on feedback from members of the intended user base.

Procedure.

The population is men and women who attended a NSF conference in January 2009. They were all employees of the federal government, a private or public university, or a business partner. They were chosen because of their affiliation with the NSF and because they are the target audience of the software being tested. Ages ranged from 25-60 years.

The procedure employed a field test experience and survey method. It is not an experiment as there is no control group or controlled independent variables. Participants spent roughly four minutes engaging the avatar in a conversation about the National Science Foundation. Each participant was asked to fill out a short survey instrument after the experience.

Results.

Based on feedback from the Likert scale questions two things become readily apparent as illustrated in figures 2 and 3. First, subjects did not believe the avatar recreated the experience of interacting with a human; subjects did not believe the avatar was realistic.

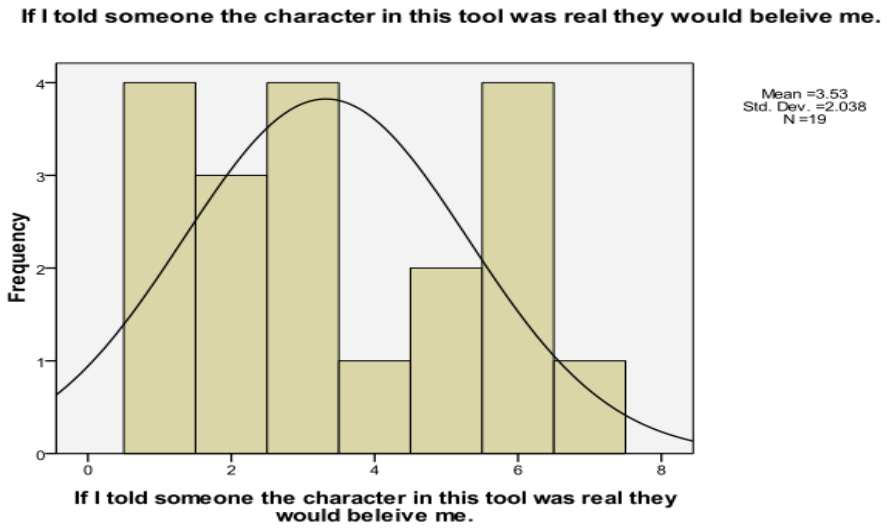

Figure 1. Answers ranged from 1 (not realistic) to 7 (very realistic). Subjects did not believe the avatar was a realistic representation of a human being.

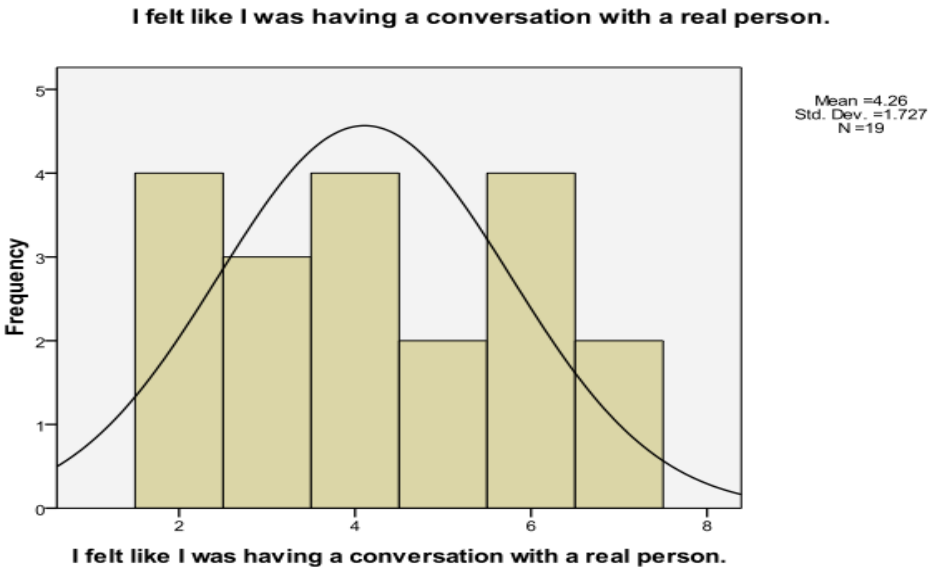

Figure 2. Answers ranged from 1 (not realistic) to 7 (very realistic). Mixed results indicate subjects did not believe the avatar engaged in a realistic conversation or acted like a human being.

Second, it appears subjects saw potential utility in a tool such as LifeLike. Though not a resounding endorsement, subjects indicated more often than not the avatar presented a *potentially* useful way of accessing information about the NSF program. The strongest responses were to the statements "I would be more productive if I had this system in my place of work" (mean=4.11, std. dev.=1.56, n=19) and "The character on the screen seemed smart" (mean=4.89, std. dev.=1.72, n=19).

Qualitative results were very useful. Subjects indicated that the largest drawback to the system was the inability to interrupt the avatar while it was speaking. When a subject was done listening to the avatar, either because they had obtained enough information or the avatar was providing an incorrect response, they were unable to stop the avatar or change topics. Interestingly, while most comments on this topic indicated it hurt the utility of the avatar, several indicated it was an issue with realism. Because they could not interrupt the avatar they felt it was not humanlike.

It seems that subjects like the idea of a multimodal approach to information access that includes non-keyboard interfaces. One subject wrote that "the ability to use speech to interact" was one of the best parts of the experience. Another responded "I like talking instead of constantly typing." Subjects indicated "There is a future potential [but it's in the] very early stages" and "potentially easier than search for info in the traditional way." However very few subjects indicated the avatar was realistic commenting on its conversational style ("It felt like I was talking to a machine" and "[The avatar] got confused when I said thank you") and failure to replicate the real person upon which the avatar is based saying "the avatar does not move naturally" (though one subject indicated the tone of the avatar's voice was similar to the real person's). Another subject contended that the realism of the avatar's appearance wasn't as important as the movement and ease of information access in terms of providing users satisfaction and a sense of successful interaction.

The numerical assessment drew upon a small sample size and so it is difficult to generalize results. However, the open-ended questions reinforce the numerical results indicating a dominant conclusion: the system appears to hold promise as an interface for a DSS but subjects still treat it as a computer system and are not engaging the avatar as they would a human because the system is not yet realistic enough for them to suspend disbelief.

Study Three: Comparative Emotive Capabilities of 3D Avatars.

Goals.

The goals of this study were twofold. First, determine whether the specific avatar being developed is capable of conveying emotional states. Second, determine more generally whether realistic avatars are good vehicles for conveying emotional states accompanying spoken information. In this study we use still renderings of an avatar and photos of the human the avatar is based on to determine whether subjects identified the emotional states comparably between the two. The rendered images used in this study are the intermediate result of graphical enhancement while we continuously improve our visual representation techniques meaning the avatar images are not the most current version we have generated.

Procedure.

The human model for the avatar, Alex, was chosen for our study. Our work here is based on Ekman's (Ekman, 1972; Ekman and Davidson, 1994; Scherer and Ekmna, 1982) approach to expressing emotions. The method draws from two sources (Euison and Massaro, 1997; Mendolia, 2007) and merges them together to focus specifically on the avatar, and to incorporate a larger pool of research subjects available online. Photographs were taken of Alex exhibiting six classic emotional states: anger, fear, disgust, happiness, sadness, and surprise. Three photos of each emotional state were selected based on how well they corresponded to the elements of Ekman's emotional characteristics (18 total human images). Images of the avatar were rendered to mimic the photos of Alex as closely as possible by manipulating key facial variables (18 total avatar images). Eyeglasses were removed to avoid interfering with facial features. The avatar renderings used were not photorealistic but had the prominent facial features necessary (see figure 3). In some cases, we modified phoneme shape, head orientation, and eye gaze in addition to the avatar mesh shape to obtain the best match.

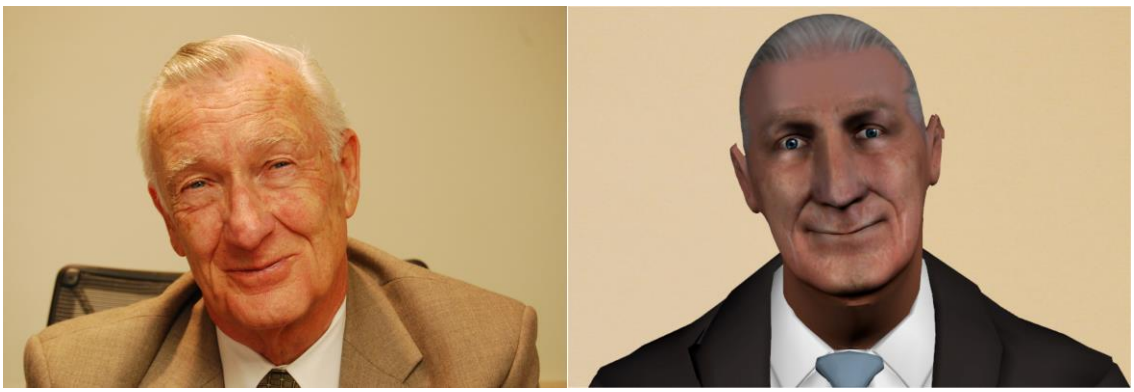

Figure 3. A sample happiness emotion. Avatar emotion is rendered with weight parameters as follows: Smile (1.0), Blink Left (0.2), Blink Right (0.1), Brow Up Left (0.2), Brow Up Right (0.1), and Phoneme B (0.65)

Subjects were directed to an online survey tool where they are shown the 36 images, randomly ordered, and asked to identify which of the six emotional states the face portrays. Subjects are only allowed to pick from the six emotional states and there was no "other" or "none" option. Recruitment incentives were used to create a sample (n=1744) taken from across the undergraduate and graduate student population of a major research university with

approximately 25,000 students. Gender was split almost evenly: 864 males and 867 females; ages ranged from 18 to 64 (mean=23.5, median=22, mode=20).

Results.

We sought two measures: (1) did the subjects correctly identify the emotion displayed and (2) did the subjects match the emotion for each human/avatar pair? Subjects did not identify anger in either the human or avatar to a useful degree. In four of the six anger images the most common response was anger but it was never the majority answer. Disgust did not fare well either, though in one pair subjects did correctly identify the emotion as disgust even if less overwhelmingly in the avatar image. Subjects could not correctly identify the human or avatar images with regard to fear, indicating that perhaps the human images were not sufficiently prototypical. The images indicating surprise were also met with mixed results.

The largest successes were the emotions happiness and sadness. In all six happiness images (three human, three avatar) the results were overwhelmingly correct and sadness was also identified with a high degree of accuracy. It appears that happiness and sadness are the easiest emotions to artificially indicate on the human face and the easiest to accurately replicate on the avatar.

This study provided useful feedback for our work and informs the decisions we will make in the next phases. It appears the current avatar is capable of successfully indicating happiness and sadness. Our avatar indicates happiness to roughly the same degree as the human upon which it was based; the same is true of sadness. The other four emotions - anger, fear, surprise, and disgust - are not currently indicated by our avatar to any useful degree.

Table 3 illustrates the successful pairs, happiness and sadness, including three pairs of images (each pair made up of one avatar rendering and one photo of the real human) for each emotion tested. The left column indicates what emotion was intended to be depicted. Columns represent what percentage of subjects identified each emotion in the image. Percentages in bold represent the most popular selection made by subjects regarding that pair. The far right column represents the number of valid responses from subjects (n). Highlighted (grey) pairings are the most successful based on Paired Samples T Test for each human/avatar pairing (threshold is $p < 0.05$). Thus, the table illustrates that subjects seem to recognize happiness and sadness between the human picture and avatar rendering.

Emotions	Anger	Disgust	Fear	Happiness	Sadness	Surprise	*n*
Happiness	0.0 / 0.7	0.1 / 0.8	0.1 / 0.3	**98.7** / **93.9**	0.2 / 0.9	0.9 / 3.3	1599
	0.2 / 1.1	0.2 / 0.5	0.4 / 0.5	**93.5** / **89.1**	0.2 / 0.5	5.5 / 8.2	1685
	0.3 / 1.2	0.2 / 0.4	0.1 / 0.2	**98.6** / **94.9**	0.1 / 0.4	0.7 / 2.9	1600
Sadness	0.7 / 20.9	20.2 / 13.5	1.8 / 5.8	0.8 / 9.5	**74.7** / **46.7**	1.7 / 3.6	1595
	0.9 / 1.9	2.7 / 4.7	2.2 / 6.3	0.2 / 0.6	**93.6** / **85.5**	0.4 / 1.1	1610
	1.1 / 1.6	7.7 / 4.1	3.2 / 3.7	0.2 / 1.8	**85.6** / **87.6**	2.2 / 1.2	1586

Table 3. Percentage of valid responses identifying the emotion displayed. This is a subset of a larger table showing all results. These subsets illustrate the successfulness of sadness and happiness in both the human model and the avatar.

While the avatar did not successfully display the other four emotions, the human photos did not achieve reliable levels of emotional indication either. In fact, there were several pairs where the avatar and human photos were identified in the same incorrect way (e.g., confusing sadness and disgust). One interpretation is that the avatar's emotional state was sometimes being interpreted the same as the human but the human image was not a good prototypical indication of the given emotion. Therefore, it is possible that the avatar was fundamentally correct in recreating the human expression but that we chose the wrong human face on which to base the avatar. Further research needs to be done to determine whether the remaining four emotions can be better indicated by the human model and, if not, whether we may want to choose a new human to serve as the basis for the avatar.

Project and Study Conclusions

In this paper we have outlined work done in the development of the Lifelike Responsive Avatar Framework and presented three studies designed to assess, critique, and improve the success of the avatar. The first study looked at those elements of paralinguistic and extralinguistic elements important to creating realistic representations of people. The second field-tested our avatar and framework on subjects drawn from the intended audience at the

National Science Foundation. The third study evaluated the ability of our avatar to accurately recreate the emotional states of a human.

Results.

Our avatar is capable of successfully indicating two emotions, happiness and sadness, but not necessarily anger, disgust, surprise, or fear. Because the human photos showed the same mixed results it is possible we need to reassess the human model for our project. Based on the subjects' comments it appears that shoulders, arms, and hands are more important, in terms of movement, than an avatar's head. This is important because articulation of the joints and extremities is more difficult to recreate than motion of the head as a single object (excluding facial elements and hair).

A clear preference for TTS or recorded voices was not identified. This is important because much time and energy is spent on TTS systems and if they are not significant to utility or perceived realism that may be time better spent elsewhere.

Members of the target audience of Project LifeLike, employees of the National Science Foundation, indicate the avatar is not realistic but they still see potential in the system as a fruitful way of accessing information stored in a database. This indicates there are instances where avatars can serve as interfaces for a DSS.

Limitations and Further Research.

One aspect we did not consider in our studies is that a human recognizes emotions within a context accompanying temporal changes. Further investigation of how those factors affect on avatar's ability to emote is necessary. Dynamic face features such as wrinkle generation will also be considered in later study. In future research it may be important to determine the specific body motions that are most important to realism (e.g., hands versus shoulders) because of the cost and effort necessary to render them in a virtual world.

Our assessments serve as a starting point to launch further research into the elements of avatars that are most important for realism and utility. In further work a comprehension test would help determine whether these variables are linked to understanding.

Our experiments only asks the subjects whether they *felt* the avatar would be useful for learning, but there was no objective measure of whether learning had actually occurred. Similarly, there was no test of how long new information was retained. While survey questions might be interpreted differently by different subjects, care was taken to craft questions that were as narrow as possible. There is ample room here for a future study to determine the utility of similar avatars.

The text the avatar uses for speech might be hard to contextualize. Out of necessity for the project driving the research, the avatar spoke about the National Science Foundation. If a subject is not familiar with the organization (as was in the case in studies 1 and 3) they may become confused. Subjects may also compare this avatar to the ones they are familiar with in video games (a comment was made by at least four of the subjects in study 1 making this direct comparison). Games are generally hyperactive and often hypersexual. These elements are not applicable to the work of this avatar because it serves to provide information, not entertainment. But, based on subjects' responses we may need to account for this issue in future research by more constructively evaluating background exposure to avatars.

References

Argyle, M., Cook, M. (1976.) Gaze and Mutual Gaze. Cambridge: Cambridge University Press.

Aristotle. (1984). The Rhetoric and the Poetics. Trans. W. Rhys Roberts. New York: Random House.

Badler, N. I., Phillips, C. B., Webber, B. L. (1993). Simulating Humans: Computer Graphics Animation and Control. New York: Oxford University Press.

Bavelas, J. B., Black, A., Chovil, N., Lemery, C. R. & Mullett, J. (1988). Form and function in motor mimicry: Topographic evidence that the primary function is communicative. Human Communication Research, 14, 275-299.

Becker, B., & Mark, G. (2002). Social conventions in computer-mediated communication: A comparison of three online shared virtual environments. In R. Schroeder (Ed.), The social life of avatars, 19-40. New York: Springer.

Beebe, S. A., & Masterson, J. T. (2000). Communicating in small groups. 6th ed. New York: Addison Wesley Longman.

Bowers, J., Pycock, J., O'Brien, J. (1996). Talk and embodiment in collaborative virtual environments. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, Canada, 58-65. ACM Press.

Burke, K. (1969). A Rhetoric of Motives. Berkeley, California: University of California Press.

Cassell, J., Bickmore, T., Campbell, L., Vilhjalmsson, H., Yan, H. (2001). More than just a pretty face: conversational proto- cols and the affordances of embodiment. Knowledge-Based Systems, 14: 52–59.

Chartrand, T. L. & Bargh, J. A. (1999). The chameleon effect: The perception behavior link and social interaction. Journal of Personality and Social Psychology, 76, 893-910.

Chen, M., Chartrand, T. L., Lee Chai, A. & Bargh, J. A. (1998). Priming primates: Human and otherwise. Behavioral and Brain Sciences, 21, 685-686.

Conniff, R. (2004). Reading Faces. Smithsonian, 34(January 2004), 49.

Ducheneaut, N., Wen, M., Yee, N., Wadley, G. (2009). Body and mind: a study of avatar personalization in three virtual worlds. Proceedings of the 27th international conference on Human factors in computing systems. Boston, MA. 1151-1160.

Ekman, P. (1972). Emotion in the human face. New York, New York: Pergamon Press.

Ekman, P. (1982). Emotion in the human face. 2nd ed. New York: Cambridge University Press.

Ekman, P., Davidson, R. J. (1994). The nature of emotion. New York, New York: Oxford Press.

Ekman, P & Friesen, W. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Palo Alto, California: Consulting Psychologists.

Ekman, P & Friesen, W. (1993). Unmasking the face. Eaglewood-Cliffs, NJ: Prentice-Hall.

Epps and Close. (2007). A study of co-worker awareness in remote collaboration over a shared application. CHI '07 extended abstracts on Human factors in computing systems. 2363 – 2368.

Euison, J. W., Massaro, D. W. (1997) Featural evaluation, integration, and judgement of facial affect. Journal of Experimental Psychology: Human Perception and Performance (23). 213-226.

Garau, M., Slater, M., Bee, S., Sasse, M. A. (2001). The impact of eye gaze on communication using humanoid avatars. CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems, March 2001. ACM Press.

Garau, M., Slater, M., Bee, S., & Sasse, M. A. (2002). The impact of eye gaze on communication using humanoid avatars. Proceedings of the SIGCHI conference on Human factors in computing systems, Seattle, Washington, 2001, pp. 309-16. New York: ACM.

Garau, M., Slater, M., Vinayagamoorthy, V., Brogni, A., Steed, A., Sasse, M. A. (2003). The impact of avatar realism and eye gaze control on perceived quality of communication in a shared immersive virtual environment. Proceedings of the SIGCHI conference on Human factors in computing systems, Ft. Lauderdale, Florida, 2003, pp. 529-536. New York: ACM.

Goffman, E. (1967). Interaction ritual: Essays on face-to-face behavior. New York: Anchor.

Hall, J. A., Roter, D. L., Rand, C. S. (1981). Communication of affect between patient and physician. Journal of Health and Social Behavior (22). 18-30.

Kalra, P., Gobetti, E., Magnenat-Thalmann, N., & Thalmann, D. (1993). A multimedia testbed for facial animation control. International Conference of Multimedia Modeling, November 1993, 59-72.

Kandogan, E., Krishnamurthy, R., Raghavan, S., Vaithyanathan, S., Zhu, H. (2006). Avatar semantic search: a database approach to information retrieval. SIGMOD '06: Proceedings of the 2006 ACM SIGMOD international conference on Management of data, June 2006. ACM Press.

Kaur, M., Tremaine, M., Huang, N., Wilder, J., Gacovski, Z., Flippo, F., et al. (2003). Where is "it"? Event Synchronization in Gaze-Speech Input Systems. Proceedings of the 5th international conference on Multimodal interfaces, Vancouver, BC, Canada. 151-158. New York: ACM.

Kendon, A. (1967). Some functions of gaze-direction in social interaction. Acta Psychologica, 26, 22-63.

Kendon, A. (1983). Gesture and Speech: How they interact. In J.M. Wiemann & R.P. Harrison (Eds.), Nonverbal interaction. Beverly Hills, California: Sage.

Knapp, M. L., & Hall, J. A. (2002). Nonverbal communication in human interaction. 5th ed. Crawfordsville, IN: Thompson Learning.

Koon, K. (2006). A case study of icon-scenario based animated menu's concept development. Proceedings of the 8th conference on Human-computer interaction with mobile devices and services. 177-180.

Kshirsagar, S., Magnenat-Thalmann, N., Guye-Vuillème, A., Thalmann, D., Kamyab, K., & Mamdani, E. (2002). Avatar Markup Language. Proceedings of the workshop on Virtual environments 2002, Barcelona, Spain, 2002, pp. 169-77. Aire-la-Ville, Switzerland: Eurographics Association.

LaFrance, M. (1979). Nonverbal synchrony and rapport: Analysis by the cross-lag panel technique. Social Psychology Quarterly, 42, 66-70.

LaFrance, M. (1982). Posture mirroring and rapport. In M. Davis (Ed.) Interaction rhythms: Periodicity in communicative behavior, 279-298. New York: Human Sciences Press.

LaFrance, M. & Broadbent, M. (1976). Group rapport: Posture sharing as a nonverbal indicator. Group and Organization Studies, 1, 328-333.

Lee, J., Chai, J., Reitsma, P. S. A., Hodgins, J. K., Pollard, N. S. (2002). Interactive control of avatars animated with human motion data. ACM Transactions on Graphics, 21(3). ACM Press.

Martey, R. M., Stromer-Galley, J. (2007). The Digital Dollhouse: Context and Social Norms in The Sims Online. Games and Culture, 2(4), 314-34.

Mendolia, M. (2007). Explicit use of categorical and dimensional strategies to decode facial expression of emotion. Journal of Nonverbal Behavior (31). 57-75.

Osborn, M., Osborn, S., Osborn, R. (2009). Public Speaking. 8th ed. Boston, MA: Pearson Education.

Plato. (1952). Gorgias. Trans. W. C. Helmbold. New York: Bobbs-Merrill.

Plato. (1956). Phaedrus. Trans. W. C. Helmbold & W. G. Rabinowitz. NewYork: Macmillan Publishing.

Poyatis, F. (1993). Paralanguage: A linguistic and interdisciplinary approach to interactive speech and sound. Amsterdam: John Benjamins.

Salem, B., Earle, N. (2000). Designing a non-verbal language for expressive avatars. CVE '00: Proceedings of the third international conference on Collaborative virtual environments, September 2000. ACM Press.

Scherer, K. R. (1982). Methods of research on vocal communication: Paradigms and parameters. In K. R. Scherer & P. Ekman, Eds. Handbook of methods in nonverbal behavior research. Cambridge, United Kingdom: Cambridge University.

Scherer, K. R. (1986). Vocal affect expression: A review and a model for future research. Psychological Bulletin (99). 134-165.

Scherer, K. R., Ekman, P. (1982). Handbook of methods in nonverbal behavior research. New York, New York: Cambridge University Press.

Tanikawa, T., Suzuki, Y., Hirota, K., & Hirose, M. (2005). Real world video avatar: real-time and real-size transmission and presentation of human figure. Proceedings of the 2005 international conference on Augmented tele-existence, Christchurch, New Zealand, 2005, pp. 112-18. New York: ACM.

Tiemens, R. K. (1978). Television's portrayal of the 1976 presidential debates: an analysis of visual content. Communication Monographs, 45, 362-370.

Trager, G. L. (1958). Paralanguage: A first approximation. Studies in Linguistics (13). 1-12.

Turing, A. (1950). Computing machinery and intelligence. *Mind* LIX(236). 433-460.

Van Swol, L. M. The Effects of Nonverbal Mirroring on Perceived Persuasiveness, Agreement with an Imitator, and Reciprocity in a Group Discussion. Communication Research, 30(4), 461-480.

Wacker, Stoev, Keckeisen, Straßer. (2003). A comparative study on user performance in the Virtual Dressmaker application. Proceedings of the ACM symposium on Virtual reality software and technology. 73-80.

Yanning, X., Juner, L., Bin, W. (2004). Research on intelligent avatar in VRML worlds. Proceedings: The 8th International Conference on Computer Supported Cooperative Work in Design, 1(May 2004), 26-28.

Yee, N., Bailenson, J. N., Urbanek, M., Chang, F. & Merget, D. (2007). The unbearable likeness of being digital: The persistence of nonverbal social norms in online virtual environments. CyberPsychology and Behavior, 10, 115-121.

Yin, G., Yang, D., Wen, Q., Lai, C., Shen, J. (2006). Sincerity and User Avatar Research Based on Binocular Vision in Virtual Reality. 2006 IEEE Conference on Cybernetics and Intelligent Systems, June 2006, 1-5.

Zhan, C., Li, W., Safaei, F., Ogunbona, P. (2007). Emotional states control for on-line game avatars. NetGames '07: Proceedings of the 6th ACM SIGCOMM workshop on Network and system support for games, September 2007. ACM Press.