In Proceedings of the First Workshop on Embodied Conversational Characters (WECC '98), October 12-15, 1998, Lake Tahoe, California.

Intelligent, Expressive Avatars

[α]Jimena Olveres, [α]Mark Billinghurst, [β]Jesus Savage, [α]Alistair Holden

[α]Human Interface Technology Laboratory
University of Washington
Box 352-142
Seattle, WA 98195. USA

[β]Laboratorio de Interfaces Inteligentes HITL-South
Universidad Nacional Autonoma de Mexico
UNAM
Mexico

{jimena, grof}@hitl.washington.edu, savage@servidor.unam.mx, holden@cs.washington.edu

ABSTRACT

Facial displays are key for communicating emotions in face to face conversation and can be made simultaneously with speech. However most collaborative virtual environments force the user to explicitly set avatar emotions after they have entered text or voice input. In this paper we present an intelligent system that will infer different emotions from textual input, parsing emotive expressions so that these emotions can be automatically displayed on the corresponding virtual avatars appearance. Although our intelligent avatars have their emotions driven by text input, our technique could also be applied to fully autonomous agents.

1 INTRODUCTION

On-line collaborative virtual environments have become popular with Internet users. In these 3D chat environments users can pick virtual avatars to represent them and then move around the space talking to each other. In most cases communication is via text input, although some more recent systems such as Onlive's *Traveler* [1] also support audio communication. Unlike face to face conversation non-verbal communication is very limited in these settings. The interfaces typically require users to input avatar body language and facial expression via key presses which means is it almost impossible for users to chat and emote at the same time. In face-to-face conversation people do not have to think about their expressions before making them and can display non-verbal cues at the same time as speaking. In fact non-verbal communication is a very important aspect of face to face communication.

It is the goal of our research to develop intelligent avatars which use expert system software that can infer user emotions from text input and set a facial representation to the appropriate expression. This will reduce the need for users to switch between typing messages and controlling their avatar representation. As a consequence of this, the user will be able to maintain more natural communication with the other people in the collaborative virtual environment.

In this paper we present preliminary results from this work. In next sections we describe the role of the face and human emotion displays in communication and review previous research on emotional display in avatar environments. We then describe our approach for emotion-extraction from text-only input and some results from informal user studies with the system. Finally we summarize the lessons we have learnt thus far and outline the future directions we intend to take.

2 FACIAL DISPLAY FOR COMMUNICATION

From a communications viewpoint the human face can be seen as an independent channel that conveys emotional and conversational signals encoded as facial displays. Facial displays have been the subject of scientific study for a long time involving a number of different disciplines, including psychology, ethnology, anthropology and communications. This research has identified a number of characteristics that could be useful for designing human-computer interfaces, including:

- communication redundancy.
- a social interface.
- communicative/emotional display.

2.1 Communication Redundancy

One of the major features of face-to-face communication is the multiplicity of communication channels and modalities used. Conversation is supported by multiple coordinated activities of various cognitive levels. For instance, in conversational understanding syntactic and semantic processing are coupled, and other processes are executed in parallel as part of these coordinated activities. As a result communication becomes highly flexible and robust, so that failure of one channel is recovered by another channel, and a message in one channel can be explained by the other channel. Among humans, the face is used extensively to

display redundant or complimentary information. For example visual cues from the lips are used to help speech understanding, while gaze plays an important role in conversational turn taking [2].

To realize a true multimedia/multimodal user interface with the same robustness, it is necessary to study how humans perceive emotional information, which information humans are sensitive to, and how they display this information.
Studying facial displays in discourse implies that they are tightly integrated with verbal and nonverbal acts. Although verbal and non-verbal acts have most often been studied separately a number of researchers (see Chovil [3] for a review) have advocated an integrated or multichannel approach to the study of communication. Pike [4] also states that a unified theory of human behavior must allow for the integration of various behaviors and not just the summation of them.

The studies of Chovil [3] mention that facial displays are regarded as linguistic elements of a message rather than outputs or outputs or "spillovers" of emotion processes. Chovil's main purpose is to uncover the ways in which facial displays contribute to the production of messages in conversation. He analyzes facial displays and classifies them according to the kind if information conveyed and whether they the information was redundant with the verbal context. These categories were syntactic, speaker redundant, speaker not redundant, listener comment and adapters. The most frequent displays were the syntactic and semantic speaker redundant. Syntactic displays are those which appeared to mark stress on particular words or clauses, associated with syntactic aspects of an utterance or organizational structure of the talk, like emphasizers, underliners, and question markers. The most common facial actions observed in this category are the raising or lowering of brows, and widening or tightening of the eyes. The second largest linguistic category is semantic speaker redundant in which facial displays convey information that is also conveyed in speech. These displays emphasis a part of the idea being conveyed verbally and often resemble many of the speech-related hand gestures that occur with words. For example, a person might have said, "I think liver is disgusting" and wrinkled her nose simultaneously.

2.2 A Social Interface

Another use for facial display is to provide a human-like social interface. Humans are social beings so communication technologies should be designed to maximize the potential for establishing and maintaining social relationships [5]. Facial displays are usually directed not at oneself, but at others and so help develop better social relations. Similar results can be found with computers. Recent advances in human-computer interfaces have added speech and vision input to the desktop. As computers have more human-like sensory capabilities users will increasingly treat them as humans [6]. Providing computers with a facial display is a way of taking advantage of this and displaying interface information in a form that naive users will understand.

Experiments have shown that facial displays are helpful especially upon first contact with a computer, and early interaction with a facial displays improves successive interaction, even when there is no facial display [5]. These results suggest that interfaces with facial displays reduce the mental barrier between the users and the computing systems.

2.3 Communicative/Emotional Display

Nagao and Takeuchi [7] assert that facial expressions serve two functions; as expressions of emotional states, or as communicative signals. Of these, Chovil asserts that the primary function of facial displays is to communicate messages to others [8]. This means that the introduction of facial displays into computer-human interfaces can make the interaction more efficient by lessening the amount of emotional information that must be explicitly stated.

There have been several attempts to categorize facial displays according to their communicative roles. Ekman and Friesen [9] group expressions according to six types of emotion they can display; surprise, fear, disgust, anger, happiness and sadness. They mention that emotional facial displays are independent of the situation, that is their meanings are the same wherever and whenever they appear.

In contrast, Izard [10] provides a complete definition for emotion that includes three components: (a) the experience or conscious feeling of emotion, (b) the processes that occur in the brain and nervous system, and (c) the observable expressive patterns of emotion, particularly those on the face. He claims that ten fundamental emotions have been identified and defined empirically; joy, surprise, distress-anguish, anger-rage, disgust-revulsion, contempt-scorn, fear-terror, shame/shyness-humiliation and guilt. Although these are the fundamental emotions, other factors can affect emotional display and generate new emotions like the love of hate.

3 PREVIOUS WORK

As can be seen from the previous section facial display has a number of important roles for face-to-face communication. However in many collaborative virtual environments simple avatar representations are used so this facial information is

lost. In this section we review work which has been developed to address this limitation.

In the most advanced collaborative virtual worlds users can talk to each other using text or voice input and are represented by two or three-dimensional virtual avatars. These avatars can also demonstrate non-verbal behavior such as lip-movements, facial expressions, gesture and body posture. However this non-verbal behavior is typically controlled by explicit key press or mouse input. For example, Comic Chat [11] implements the display of facial expressions by generating default gestures and facial expressions for its characters. Simple key word spotting is also used to set different avatar appearances, although there is no underlying emotional model, Participants can also select the desired expression through an emotion wheel. Clicking and sliding with a mouse around the wheel changes the avatar's emotional representation.

Nagao and Takeuchi [7] have developed an intelligent agent that can conduct a speech dialogue while displaying facial expressions. They change the facial display with movements of eyebrows, lips, closing eyes, and add behavioral features like avoiding eye contact. This is in response to speech input and depends on the syntactic displays like exclamation marks, question marks, speaker displays like "hello" phrases and listener comments like the beginning of a dialogue. Their agent recognizes conversational structure rather then emotional content.

The closest work to our own is the BODYCHAT project of Cassell and Vilhjalmsson [12]. BODYCHAT automates low level communication behaviors between avatars, such as gaze direction. Cassell et. al. have also developed a system which automatically generates a conversation between human-like agents with appropriate facial expression and synchronized speech and gesture[13]. They focus on gaze and communicative facial expressions to establish and maintain a link between participants in a conversation.

In contrast to this work we focus on generating emotional facial expression from the user's textual input. This involves the development of an expert system that can predict emotional state from minimal cues and display the appropriate facial expressions. In the rest of the paper we provide complete details of our approach and some initial user experiences with the system.

4 EMOTIONAL SYSTEM

In face-to-face conversation humans use a wide range of non-verbal cues to recognize emotions, including gesture, gaze, fixation or prosodic information (like voice intonation). It would be possible to detect these non-verbal emotional cues using additional hardware such as cameras, pressure sensitive keyboards, or microphones. However we are interested in exploring the accuracy of a system that uses no extra hardware and relies on text-only input. In this situation there may be other cues that can be used to measure emotional states, including explicit emotional display cues such as:

- Words that define emotional states
- Use of emoticons, such as :-) or :-D or :-(to show a happy or sad emotional state [22],

or implicit emotional cues such as:

- Use of upper case letters to denote that someone is shouting; i.e. YOU ARE ALWAYS GETTING ME ANGRY! [21]
- Length of phrases typed.
- Rate of typing and number of mistakes. This last two items are in reference to certain states already shown like for example if someone is already angry it can give short answers and if extremely angry, maybe the rate of mistakes is bigger than the usual, which is difficult to measure.

In our initial prototype we are just concerned with the explicit cues and simple implicit cues. These can be captured using natural language processing and inferences about their emotional content made by an expert system. To do this our system has the following components, shown in figure 1:

- A simple Natural Language Parser for key word spotting, phrase length measurement and emoticon identification.
- A rule based expert system that constructs emotion scores based on the parsed text and contextual information.
- An animation engine which displays the facial expression using a graphics model of the face.
- A speech synthesis system that speaks the typed text with emotional intonation.

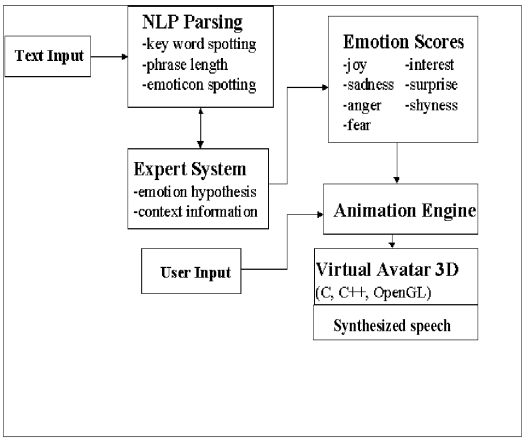

Figure 1. Intelligent Agent Block Diagram

In producing emotional display, the first step is Natural Language Parsing (NLP) of the user's text input. NLP of conversational text is traditionally a difficult problem, however in this case we are just searching for keywords, modifiers like adverbs, and emoticons that could give cues about a user's emotional state. This parsing is accomplished using a rule-based expert system developed with CLIPS, an expert system shell language [14]. CLIPS uses *if-then* production rules to match user input against emotional inference conditions.

Once parsing is complete a second CLIPS-based expert system proceeds to formulate an emotion hypothesis using a fuzzy logic process described later in the paper. This module generates emotion scores within a set of predefined emotional categories. These are sent directly to the Animation Engine controlling the facial display of the user's virtual avatar.

One of the important aspects of the interface is visual feedback to the user so they can see what their avatar looks like. To compensate for errors in the emotion extraction, the user can also change the avatar expression by hand using a set of on-screen slider bars, one for each emotion. This helps the user's avatar to have the expression that they want to portray, however the use of the intelligent avatar modules mean that ideally the user will rarely have to do this.

The text output is also spoken using a synthesized speech module with affective expression. The addition of affect to synthesized speech is useful in any application in which expressiveness is appropriate and desirable for reasons of naturalness [15]. This module can currently simulate a voice of a happy, sad and normal person, using the genders: male, female or boy. The emotional expert system also coordinates the synthesized speech.

4.1 Emotions and Keyword Recognition

To use emotions in a communication interface a classification system is required, with associated measures of probability and degree (or strength). We follow the work of Ekman [9] and Izaard [10] and assume that strong emotions, such as passion, hatred, fear and greed are universally experienced.

For our system we use seven emotional states from the classification scheme proposed by Izard; interest, joy-happiness, surprise, sadness, anger, fear, and contempt. The NLP parser tries to classify text input into one of these categories with a measure of the emotion's strength. The main words recognized are adjectives (like happy, sad, angry, and excited), that give an clearer idea of emotion. For example, consider two people talking at each other trying to express their feelings:

John: "I am very angry at you."
Jim: "Well, I am resentful and afraid."

One can extract keywords such as *angry* from the preceding input, to characterize John's emotional state. Similarly, Jim's emotional state is characterized with the words *resentful* (or angry) and *afraid*. These emotional states are then modified depending on the frequency in which these keywords are found. Once the emotional strength of a particular category passes a certain threshold then the user avatar's representation can be changed to show the appropriate emotion.

We distinguish between emotional state and intensity. Each of the emotional states has intensity rating associated with it on a scale of one to four, as shown in Table 1. The avatar initially begins with neutral emotion for all states (except for interest, which is assumed as an initial state), but these intensities change as the text is parsed for emotional content. Various stimuli may perturb an emotional measure by different degrees, which may vary for extreme emotional intensities. For example, a certain comment of anger may perturb the anger emotion from neutral (degree=1) to irritated (2) but may perturb irritated (2) to enraged (4). Conversely a compliment may perturb anger positively e.g. from angry to neutral.

Emotion	1	2	3	4
Joy	Neutral	Happy	Joyful	Elated
Anger	Neutral	Irritated	Angry	Enraged
Sadness	Neutral	Unhappy	sad	Depressed
Fear	Neutral	Apprehensive	afraid	Terrified

Table 1. Intensity of Emotions

Other modifier cues such as emoticons are also used. Some of the common emoticons that can be used in simple text to express deeper emotions were taken from Damer [16] and include :-), :-(, and :-D. In addition punctuation such as ! and ? are used as emotional modifiers, also words such as *not, very, so, more, less, much, lot, little, extremely* and *highly*.

The way this simple Natural Language Parser works is that the system recognizes the keyword or emoticon and identifies it as one of the basic emotions, then it identifies if there exists any modifier next to the keyword. Once this is done the NLP is complete and a correspondent emotional intensity found using the method described in the next section.

4.2 Emotional Model with Fuzzy Logic

After natural language processing, emotional state and intensity values are assigned to each text input string. These are then processed by an emotional expert system module to find the final intensity values that are assigned to each of the emotional states. Although the intensity of the emotional states is represented in discrete form, changes in emotion are represented using a continuous transition function so the avatar expression varies smoothly within emotional states. Two methods are used to obtain the final intensity of the avatar's present emotion; an exponential transition function, and a fuzzy logic function.

Several researchers model the change of emotion over time with an exponential function [17] [18]. We follow this in our work, so when the user changes from one emotion to another (e.g. from happy to angry) the initial transition has the biggest effect. But if the user continues reinforcing the same emotion (which could give a transition from happy to joyful or to elated), then the emotion increases as the exponential function,

$$em\_n = \Sigma_j \, em\_j * e^{-j} \quad j \geq 0,$$

where *em_n* is the actual emotional intensity and *j* is the number of times that the same emotion has been expressed in the current context. The *em_j* values are the past emotional intensities for each of these occasions. As this equation shows, for a given emotion the current emotional intensity is a function of the past emotional intensities and the newly expressed intensity, so another way to represent it is by the function,

$$em\_current = em\_old + em\_j * e^{-j}.$$

As soon as the system changes to a new emotional state the value of j is reset to zero and the quantification method is started again. The *em_n* emotion intensity value is used by the fuzzy logic function to determine transitions between emotional states. In the case of a negative modifier the *em_j* exponential term is subtracted instead of added, further reducing the emotional intensity. This emotional value is calculated for each emotion expressed by the user, and the emotion with the highest intensity is used to set the avatar expression.

For a given emotional intensity and state we must decide the corresponding emotional level. For example if the user is in the *Joy* state and have a current intensity of 2.5 does this correspond to *Happy* or *Joyful*? We cannot just give an emotional intensity to a certain emotion and decide if falls in only one emotional state because there could exist overlapping between states. To address this problem we define a fuzzy logic function that determines the probability that a given emotional intensity is a member of a particular emotional state. Fuzzy logic is a method of easily representing analog processes on a digital computer [19]. It is widely used in areas where it provides decision-support and expert systems with powerful reasoning capabilities bound by a minimum of rules. Figure 2 shows the fuzzy membership functions corresponding to the emotional states of the Joy emotion. In this case an intensity of 2.5 has a 0.5 probability of being in the state *Happy* and a 0.5 probability of being in the state *Joyful*.

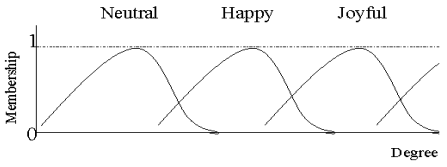

Figure 2. Fuzzy Membership Functions for Joy

Once the keywords are recognized and grouped into a set corresponding to a particular emotion, the system checks if the agent is in a different emotional intensity in the same state (say happy). The program gives priority to the last emotion showed, and uses the *em_n* value coming from the exponential function in the fuzzy subsystem to set the transitions within the same emotion. On the other hand, if the system recognizes an emotion corresponding to a different state, say Sadness while the user was previously in Joy, it immediately raises the level of Sadness. At the same time the intensity of Joy is diminished and compared to the Sadness level. Whichever is greater will be used to set the final facial expression.

5 RESULTS

The intelligent avatar runs as a client server architecture using a single DEC Alpha computer to run the expert system modules and a graphical client running on SGI computers for each user. The animation and visualization module consists of a three-dimensional computer graphics face rendered using openGL on the SGI computer. The model used is the one developed for the DECface talking head project [20] and consists of a 500 polygon skin surface overlaying 16 virtual muscles. Facial expressions are caused by changing the muscle values, producing local deformations of the face. As users type text chat to each other the text is also sent to the Expert System server and processed. The output of this is sent back to the local and remote graphics clients to set the avatar face expressions.

One of the important aspects of the interface is visual feedback to the local as well as remote users. The local user can see what their avatar looks like and, if necessary, change the avatar body expression by hand using expression slider bars. This will help the user's avatar to ideally have the expression that they want to portray. Figure 3 show some typical facial expressions produced and the emotion slider bar interface.

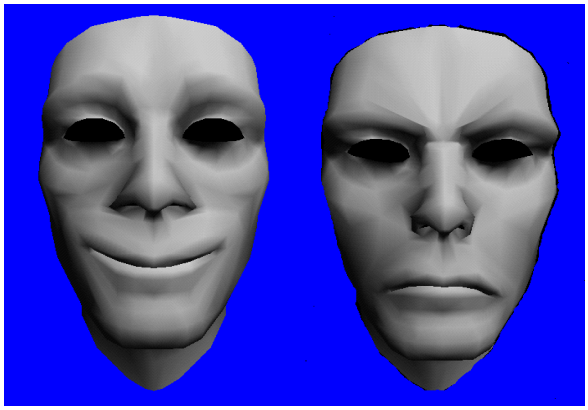

Figure 3a. Emotional Displays - happy and angry.

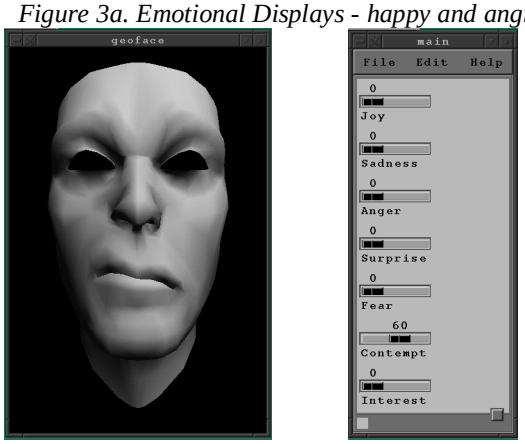

Figure 3b. The Slider Bar Producing Contempt.

6 USER TESTING

In order to evaluate the effect of adding intelligent emotional display to a chat interface we conducted a simple user study. Twenty subjects participated forming 10 conversational pairs. Their ages ranged from approximately 18 to 34 years, and all had prior computer experience although 11 persons had used a computer based chat system all of them had had access to the email.

6.1 Method

The study was conducted at the Human Interface Technology Laboratory. Upon arrival the participants were informed about the general purpose of the study and videotaping of the conversation. In these tests we compare communication under three different conditions:

- Text only.
- Text with Static Avatar Representation.
- Text with Emotive Avatar Representation.

In the text only condition users could just see a text chat window containing both their comments. The second condition added the picture of an avatar with no movement, while the last condition used our emotional expert system to automatically set the avatar expression. These conditions were applied in a random and different order with every dyad.

Before subjects started chatting to each other, they were asked what words they commonly used in email to portray an emotional state. This was used to gather additional keywords that could be added to the system in the future. After the experiment, the content of the subject's text chat was also analyzed for additional emotive terms. Subjects were also given a set of pictures showing the avatar with different expressions and asked to associate the seven emotional states with the corresponding facial expression. This was to evaluate the accuracy of our emotional representation.

Before the participants started with the discussion topic, they were asked to take a few minutes to get acquainted with each other and with the chat system. Users generally gave their name and occupancy and answered questions from the other subject. Also users were asked to tell a fantastic story or adventure of their own, and to ask questions about what their partner was saying. This helped to ensure several of the emotional states were seen before the experiment started.

Subjects were then given a discussion topic to talk about, these topics included: a time when they got a depression, movies, a time they were sick, cars, work, a time when they felt in love, the last great party they had, the last trip, a wonderful birthday party, an scared situation, pets and there were cases in which some of the participants decided to continue with the fantastic story cause they prefer to do it in

that way. The first two chat conditions lasted approximately 15 minutes each, while when using the Text with Emotive Avatar Representation it was approximately 20 minutes. After experiencing each condition a survey was given to measure the condition and to notice some of the important aspects for us, and after the three conditions they were given another to measure their preferences. This is shown in Appendix A.

6.2 User Testing Results

From the pre-experiment survey of common emotional words used in electronic communication, we were able to identify 37 additional keywords not already in our system. For six of the seven facial expressions subjects were able to correctly label them with 95% accuracy. However the interest expression was only correctly identified by 10% of the subjects. However during the text chat the conversational content enabled subjects to easily disambiguate the face expressions.

In the post-experiment survey results 90% of the users did not consider there was a great different between text only and text with static avatar. This is because they noticed the avatar just at the beginning of the second condition and once they discovered it wasn't changing didn't give it anymore attention. However this wasn't the case when using the text chat with emotional avatar.

After trying the chat interfaces, 75% of the subjects reported that they enjoyed using the intelligent avatar more than the text only chat or the text with static avatar, 15% though that it was too complicated, and 10% said they did not like it and would not use it again. The subjects that enjoyed the system when the avatars were present also tended to type more emotive expressions to cause the facial expressions to change. This indicates that the presence of the avatars will impact the conversation.

The most of the users commented that because of the combination of face and text, the emotions of the other person are more understood however it was a little difficult to handle all together: text, scrollbars and faces at the same time, but anyway it is more fun. The slide bars were helpful though too distracting some times, they would prefer not to use them and instead use some of the key functions. Some said: "in conversation we want a continuous talking, not stopping to move the mouse". This corroborates once more the need of an intelligent system in the chats.

Also users ask for the possibility to combine the emotional faces cause they tried to do it most of the times at first when they were acknowledging with the system.

In order to make this system friendlier, there was a very common suggestion of merging the answers of the written text with the face to avoid distraction and turn the attention from the face to the text or from the text to the face.

Although users were able to chat easily with one another, some users commented the delays between the text input and the avatar's expressions changes. Facial expressions were only set at the end of every typed sentence, rather than phrase or word level which made it seem a little disjoint.

Subjects also commented on the avatar appearance. About the facial display the general idea users gave was that they reinforce the conversations but also some distortions added on the hair, eyes or ears could make it look even more exaggerated. They would enable the displays to be more expressive, as well as texturing the face to make it look more natural. They also suggested the use of a quizzical expression for when a question is being asked. We have now added a quizzical expression which is displayed every time the expert system recognizes a question mark and words like how, what, where or when.

7 CONCLUSIONS

Traditionally, computer graphics characters have been entirely autonomous, or entirely human controlled. However there are a third class of avatars which have some level of communicative autonomy while allowing the user to exert control over the content of the conversation.

In this paper we have described a prototype intelligent avatar system which infers emotion from the user's text input to automatically set the avatar's facial expressions. We use a simple emotional model for possible emotional states and exponential and fuzzy logic transition functions to change emotional intensity within an emotional state.

Our preliminary user studies have shown that users respond positively to emotive avatars in a chat environment, although the nature of the conversation changes. However more work needs to be done on the interface to enable more natural communication, including reducing the delay between input and changing the facial expression, and providing a wider range of expressions.

In the future, we intend to use this interface to identify which textual cues are accurate predictors of emotion and how accurately emotional states can be predicted. We will also develop a larger knowledge base of adverbs or modifiers of

emotions, and explore the use of conversational context to increase recognition accuracy and more complex emotional phrase spotting. We would also like to start inferring from higher level conversational structure. That is meanings from the phrase, sentence or paragraph level. Finally, it would be good to implement automatic low level conversational behaviors such as those used in BODYCHAT so we have a complete communicative avatar.

A longer term goal would be to add another expert system module that learns from when the user sets avatar expression by hand and remembers the textual context in which that occurred so that over time the user will have to set expression by hand less and less.

8 REFERENCES

[1] www.onlive.com

[2] Kendon, A. Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press, 1992.

[3] Chovil, Nichole. Discourse-Oriented Facial Displays in Conversation. University of British Columbia. Research on Language and Social Interaction Vol. 25, 1991/1992:163-194.

[4] Pike, K.L. Towards a theory of the structure of human behavior. General systems yearbook, 2, 135-141

[5] Takeuchi, A., Nagao, K. Communicative facial displays as a new conversational modality. Human Factors in Computing Systems. INTERCHI '93. Amsterdam, Netherlands. April 1993, pp. 187-193.

[6] Reeves, B., Nass, C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places. New York: Cambridge University Press, 1996.

[7] Nagao, K., Takeuchi, A., Speech Dialogue with Facial Displays: Multimodal Human-Computer Conversation. Proceedings of the 32nd Annual Meeting of the Association for Computational Linguistics. Las Cruces, NM, June 1994.

[8] Chovil, N. Social Determinants of Facial Displays. Journal of Nonverbal Behaviour. 1991 vol.15(3).

[9] Ekman, P., Friesen, W. Unmasking the Face. Prentice Hall, New Jersey. 1997.

[10] Izard, Carrol E. Human Emotions. Plenum Press, NewYork 1977.

[11] Kurlander, D., Skelly, T. Salesin, D. Comic Chat. In Proceedings of SIGGRAPH 96. ACM Press, July 1996.

[12] Vilhjalmsson, H., Cassell, J. BodyChat: Autonomous Communicative Behaviors in Avatars. MIT Media Laboratory. In Proceedings of the 2nd annual ACM International Conference on Autonomous Agents, Minneapolis, 1998.

[13] Cassell, J., Pelachaud, C., Badler, N., Steedman, M., Achorn, B., Becket, T., Douville, B., Prevost, S., Stone, M. Animated Conversation: Rule-Based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. In Proceedings of SIGGRAPH '94, ACM Press, July 1994.

[14] Giarratano, J., Riley, G. Expert Systems. Principles and Programming. PWS Publishing Company, 1994.

[15] Chan, J. The Generation of affect in Synthesized Speech. In the Journal of the American Voice I/O Society. Vol. 8, July 1990, pp. 1-19.

[16] Damer, B. Avatars! Peachpit Press. 1998.

[17] Schmalz-M-S. Dankel-D-D-II. HEART-a model of emotion for natural language interpretation. 1. Background and model requirements. Proceedings of the Eighth Florida Artificial Intelligence Research Symposium. FLAIRS-95. Melbourne, FL, USA. April 1995.

[18] Piesk-J. Trogemann-G. Animated Interactive Fiction: Storytelling by a Conversational Virtual Actor. In Proceedings of the International Conference on Virtual Systems and MultiMedia, VSMM '97. Geneva, Switzerland. September 1997.

[19] Cox, E. The Fuzzy Systems Handbook. AP Professional Press. 1994.

[20] Waters, K., Levergood, T. DECFace: An Automatic Lip-Synchronization Algorithm for Synthetic Faces. Technical Report CRL 93/4, Digital Equipment Coporation, Cambridge Research Lab, September 1993.

[21]Kuilboer, Michelle. A Rationale for Synchronous Computer-Mediated Communication in an Adult Virtual Second Language Classroom. Department of Modern Languages, James Cook University, Townsville, Australia. http://sacwwww.jcu.edu.au/dm1staff/mlmak/Rationale.htm

[22] Werry, C. (1996). Linguistic and Interactional Features of Internet Relay Chat. Computer-Mediated Communication: Linguistic, Social and Cross-Cultural Perspectives. S. C. Herring. Amsterdam, John Benjamins: 47-63

8 APPENDIX A

Survey applied after each condition.

Chat ____ Condition_____

Answer the next questions crossing in the degree that corresponds. .And try to explain the reason of your answers.

1.Answer this question only if you had a previous chat condition, did you feel a great difference on the communication from the last time you tried without the avatar?

1 2 3 4 5
Not Very Much Very Much
Why?

2. In what degree do you think is good to look at a representation on the other person in the computer, like you just did?

1 2 3 4 5
Why?

3.How well do you understand how the other person was feeling:

1 2 3 4 5
Not Very Well Very Well
Why?

This next question was not included in the text only condition.

4.How much did the avatar of the other person help with the conversation:

1 2 3 4 5
Not Very Much Very Much
Why?

5.Please mark the degree in which you think you establish a real communication

1 2 3 4 5
Not established Well established

Some questions added after the chat with emotive avatar.

8. Do you consider the slide bars were helpful? Why?
9. How often did you use the slide bars?

Survey after the three conditions had been tested.

1.In a scale from 1 (very difficult) to 5 (very easy), could you rank each condition in order of how difficult was to communicate your emotional state?

1 Text Only
1 2 3 4 5
2 Text with Static Avatar Representation
1 2 3 4 5
3 Text with Emotive Avatar Representation (and sinthetized voice)
1 2 3 4 5

2.Try to explain your general impression about the system. How did you like it and what would you like to improve.?

3. What improvement can you suggest for the chat application?