Multi-Media Communication with Emotion-driven "Believable Agents"*

Clark Elliott
Institute for Applied Artificial Intelligence
DePaul University
243 South Wabash Avenue
Chicago, IL 60604
and
School of Education and Social Policy
Northwestern University
Evanston, IL
email: elliott@ils.nwu.edu
tel: (708) 467-2813
fax: (708) 491-8999

Abstract

The Affective Reasoner is a real-time simulation platform which supports multiple agents. These agents are capable of simple "emotional" reactions to situations that arise in the simulation, including those generated by interaction with the user. The platform has developed over time and currently includes, besides the underlying emotion engine, a speech recognition package which is able to discriminate at least some broad categories of emotion content; a music indexing and playback mechanism allowing virtually instant access to hundreds of hours of midi format music to aid in the expression of emotion; a schematic representation of approximately 70 emotion faces; and a text-to-speech module for expressing dynamically constructed text including a minimal amount of emotion inflection. In this paper we discuss the multi-media environment being constructed for the Affective Reasoner, followed by some brief comments on personality, interestingness of automated agents, and the merits of a "believable agent" approach. Lastly we discuss the design of three proposed applications: affective user modeling, a sympathetic computer, and an automated poker player with "emotions."

Introduction

A central assumption of our work is that simulations of social interactions between agents, including users-as-agents, should incorporate models of individual affect and personality. Most human interaction revolves around people's individual needs and goals. These lead to idiosyncratic, internally motivated behavior, and to emotional responses to situations that arise. If these aspects of mental life are not captured in simulations of interpersonal interactions, then simulated agents will be at best bland, lifeless, and unrealistic. To create "believable" agents we must not only have some internal model of emotion, but also a way to communicate with the user based on these models.

Our current work addresses these issues in several ways. First, we are developing what we hope will become a comprehensive model of certain cognitive components of emotion. This model includes twenty-four distinct emotion types (see figure 1), a rudimentary mapping of emotion to action with approximately 1000 different action paths (e.g., *shouting* is partially characterized as a *verbal, communicative, expressive, non-goal oriented* action expressing *anger*), and intensity reasoning based on a pool of approximately twenty different emotion intensity variables (e.g., *importance* and *degree of manifestation* of a goal) (cf, [Ortony *et al.*, 1988; Elliott and Siegle, 1993; Elliott, 1993]).

Second, we are currently attempting to increase the user's ability to suspend disbelief when interacting with the automated agents by using various multi-media channels to increase the emotional feedback from the agents to the users, and to give the agents the ability to understand simple spoken dialog containing affective content.

In this paper we discuss the multi-media environment being constructed for the Affective Reasoner (hereafter AR) emotion simulation program, followed by some brief comments on personality, interestingness of automated agents, and the merits of a "believable agent" approach. Lastly we discuss the design of three proposed applications: affective user modeling, a sympathetic computer, and an automated poker player with "emotions."

*Preparation of this article was supported in part by Andersen Consulting through Northwestern University's Institute for the Learning Sciences. Citation appears in bib.

Figure 1: Emotion types (Table based on [O'Rorke and Ortony, 1992] and [Elliott, 1992])

Group	Specification	**Name** and Emotion Type
Well-Being	appraisal of a situation as an *event*	**joy**: pleased about an *event* **distress**: displeased about an *event*
Fortunes of Others	presumed value of a a situation as an *event* affecting another	**happy-for**: pleased about an *event* desirable for another **gloating**: pleased about an *event* undesirable for another **resentment**: displeased about an *event* desirable for another **sorry-for**: displeased about an *event* undesirable for another
Prospect based	appraisal of a situation as a prospective *event*	**hope**: pleased about a prospective desirable *event* **fear**: displeased about a prospective undesirable *event*
Confirmation	appraisal of a situation as confirming or disconfirming an expectation	**satisfaction**: pleased about a confirmed desirable *event* **relief**: pleased about a disconfirmed undesirable *event* **fears-confirmed**: displeased about a confirmed undesirable *event* **disappointment**: displeased about a disconfirmed desirable *event*
Attribution	appraisal of a situation as an accountable *act* of some agent	**pride**: approving of one's own *act* **admiration**: approving of another's *act* **shame**: disapproving of one's own *act* **reproach**: disapproving of another's *act*
Attraction	appraisal of a situation as containing an attractive or unattractive *object*	**liking**: finding an *object* appealing **disliking**: finding an *object* unappealing
Well-being / Attribution	compound emotions	**gratitude**: admiration + joy **anger**: reproach + distress **gratification**: pride + joy **remorse**: shame + distress
Attraction / Attribution	compound emotion extensions	**love**: admiration + liking **hate**: reproach + disliking

Multi-Media extensions to the Affective Reasoner

Faces

We currently use a set of approximately 70 schematic "emotion faces" based on different intensities of the twenty-four different emotion categories. These faces are based on those currently in use in clinical settings for iconically representing different human emotions (e.g., as an aid for hospitalized children to express their emotions). Currently the drivers for the face displays are attached to the output of the emotion engine, circumventing the action component, so that agents' faces are "honest representations" of their internal states. In the future the face drivers will be incorporated in the regular action generation component of the AR so that they may, for example, deceptively smile when they are angry. Additionally, moods generate default face states after the emotions decay.

Rudimentary run-time morphing of faces is also under development, based on line drawings. This allows smooth flow from any one of the 70 faces to any other, a daunting task using static maps. A simple grammar allows us to specify the decay rate of the face back to its current default state.

Music

A wide variety of midi-format music files are used as data for the expression of mood and emotion states. The agents currently have the ability to play music files, and to change from one file to another rapidly, as a way of expressing rudimentary affective "life." Thus *triumph* might be associated with the beginning of the Fourth movement of Beethoven's Fifth Symphony, *joy* with the last movement of Beethoven's Ninth, and so forth. Preliminary work has lead us to believe that this, when used in concert with the faces, is one of the more powerful expressive mechanisms of the automated agents. In other words, rather than music being used to directly express emotions, it is best used to significantly enhance the expressive impact of emotional expression already identifiable through other means.

Working on the project is a concert pianist who is able to transcribe any form of western music, by "ear", into midi format, so there are few restrictions on the range of music available to us. This approach allows us to avoid bothersome details in acquiring the rights to performances, and of course allows us to avoid the difficult space problems associated with actual recordings.

Text-to-Speech

Agents have text-to-speech capability so that they can verbally express themselves as well. This is used sparingly, and explicitly to express emotion states and moods (and explanations of such states based on their cognitive antecedents). The goal of "believability" calls for this conservative approach, because any hint that the computer will engage in actual dialog will of course ultimately disappoint the user, leading to disillusionment.

For example, on the one hand, the following computerized utterances can be accepted without invitation for language interaction: "Jane, I am only a computer, but in my very simple way, I believe that I *like* you very much." And (if explanation is called for), "This *liking* is based on a simple preference. You are an instance of the class *people*, with the property that you are *female*, and I have interacted with you more than 10 times. Thus I have come to *like* you." On the other hand, the utterance, "Jane, I know how you really feel about me. I wish that I were alive instead of only being a computer" does express emotion, but not those actually represented in the computer.

Speech Recognition

Agents have simple speech recognition capability. The lexicon is extendable to many hundreds of words, and the system has been informally shown to recognize several different emotional inflections in speech, in constrained cases (see *Affective User Modeling* below). It has also been shown that agents are able to communicate with one another through sound waves (i.e., one agent on the computer speaks using the text-to-speech code and a speaker, while another listens using voice recognition code and a microphone). Implicit is that agents on multiple computers can speak to each other as well, without an electronic link, as part of a community of human and automated agents.

Interestingness versus accessibility

Even with the conservative approach, however, it is our feeling that the robustness of the emotion model is sutiable to give enough complexity to the agents, without sacrificing internal consistency, to make them interesting, and surprising. Here are two examples to consider: (1) Agents are able to reason about each other, about the user to a limited degree, and to have emotions based on the concerns of these others (cf, [Elliott and Ortony, 1992]). Thus, "Jane, I am *somewhat sorry for you* because you are my *good friend* and you are *distressed* that *the test did not go well*, and it was *important* to you," and, "I am *anxious* because you are my *friend* and you are *fearful* of *not playing well at the concert*," are both valid machine-based rudimentary "emotion sequences" because there is (an admittedly simple) internal semantic content for them. (See *A Sympathetic Computer* below.) (2) Agents can ask the user for English utterances that are meaningful to the user, which express the agent's "emotional" state. Thus the user might tell the automated agent to say it is *annoyed* when it is mildly reproachful over the user's violation of a principle that lead to the blocking of a minor goal. So, in the future when the automated agent says it is *annoyed*, this can be presumed to be understandable to the user.

Rudimentary Personality Models

Agents have unique personalities, represented in two parts. On the input side, agents make appraisals of situations that arise in the simulation based on their distinct dispositions – how they see the world. On the output side, agents express the "emotions" that do arise in them as a function of their distinct temperaments. That agents are "believably" different from one another is enhanced when, for example, one agent plays the music of Beethoven when *happy* (e.g., the finale of the *Ninth Symphony*) whereas another plays the music of Sonny Rollins (jazz) (e.g., the Rollins interpretation of *St. Thomas*).

The usefulness of loosely-mapped communication channels in broad agents

It is clear that sophisticated dynamic three-dimensional models of faces, such as that shown by Pelechaud at IJCAI-93, have the power to delight audiences and convey expressive information[Pelachaud *et al.*, 1993]. Nonetheless it may also be argued that much emotional content can be delivered in schematic format as well. Cartoon faces, for example, convey much about the current appraisals of the cartoon character, and schematic faces have been used in clinical settings to help children identify their emotions. Such facial representations are lacking much in the fine detail of the 3-D models, but they are much less computationally expensive to manipulate. Begging the question of which representation is more powerful, a more interesting question would appear to be regarding the effectiveness of the emotion representation on which the dynamic face depends. Our approach is that, at present, low-level personality and emotion representations are too underdeveloped to simulate complex social interaction. Since we use an *architectural* approach, independent of any domain, we feel that schematic faces are sufficiently expressive of raw emotion categories, and have the benefit of allowing for real-time manipulation on commonly used platforms.

The use of music to express mood and emotion enters into an arena in which music psychologists and others have argued for some time.[1] For example, it would be a difficult argument to say that for any particular piece of music, X, that X is *happy* music, because just as surely as we claim that Beethoven's previously mentioned "Ode to Joy" is *happy* someone will claim that it is merely *invigorating*. (Interestingly it is much easier to argue the converse, i.e., that some particular piece of music is *not sad*.) The structure of music cognition is a wide-open question (cf. [Gaver and Mandler, 1987]). That this is so, however, does not mean that music cannot be used as a tool for the coarse expression of emotion, and in particular, to enhance the emotional impact of scenes already carrying emotional content (e.g., through the use of schematic faces, and spoken text).

Music clearly has a high place in the movie and television industry because of its mood-induction capabilities, and is an accepted mood-induction technique in psychological studies. We have no plans to express each of the approximately 70 emotion/intensity states through distinct categories of music, but we feel that a loose mapping between these and perhaps eight meta-categories might be appropriate. There is also the case to be made that communication of *some* emotional presence, although not clearly understood by the witness, is a process present in human interaction as well.

[1]For example, at the Music Educators National Conference in 1981, where Roger Brown, among others, spoke.

Believablitity and the "Better Mechanical Man" Issue

One might ask, will the result of this endeavor be any better than a 90's version of Wizard of Oz illusion? There are two approaches we can take in answering this. First, one could argue that these pursuits are *not* any more than illusion. Still, to the extent that the illusion is successful, might we not use it to make significant gains in, for example, computerized education, through the enhanced *engagement* of the user? The second approach is one we prefer: Assuming that one is straightforward about what is actually meant by a "computerized agent's emotional response," has it ever actually been shown that users will not accept this as a simple, but valid, form of simulated life? A colleague once said that the only thing interesting about stupid people is that they are stupid. I would argue that quite the contrary is true: even stupid people are interesting because they have their own desires, goals, principles, and so forth, leading to interesting, if simple, emotional lives (see *A Sympathetic Computer* below).

Three example applications

Affective user modeling.

One hard, and divisive, problem facing the AI community is that of building user models. Rather than more traditional models which focus on the mental processes of the user in problem solving situations [Van Lehn, 1988], we propose an alternative wherein only the *affective* state of the user is modeled. This is a much smaller problem, but one which should provide useful leverage. It might be considered akin to the feedback a responsive speaker might make use of when "playing" to her audience.

To implement affective user modeling, several components are required: (1) A structure which allows us to capture an agent's (in this case, the user's) outlook on situations that arise. This structure must include some concept of the role, personality, and current state of the user (within the context of shallow emotion reasoning), which together comprise the basis for the user's idiosyncratic way of construing the world. (2) A lexicon through which the user expresses his or her emotions to the computer. (3) A comprehensive set of emotion categories which allow for the mapping of emotion eliciting situations to the emotion expression lexicon, and vice versa. In our current work, as discussed above, we have implemented a broad, albeit shallow, representation of the first component, and a comprehensive, descriptive representation of the third.

The weakest link in such a system is in the lexicon. How does a computer, which cannot understand faces[2], and which presumably has no plausible inference mechanism for understanding enough about the world to determine which emotions are likely to have arisen, know what emotion a user is expressing? In addressing this question we consider several hooks which show promise in allowing us to work around this problem, at least to some degree. First, and most importantly, it might well prove to be true that users are *motivated* to express their emotions to the computer, provided that there is at least the illusion that the computer understands how they are feeling. Should this be so, then some latitude is afforded us in requiring that the user, who is adaptable in communication, conform to the protocol of the computer, which is not. Second, the comprehensive emotion model allows us to develop a large lexicon categorized by both emotion category and intensity. Third, speech recognition packages are advanced enough to capture some of what is of interest to us with respect to the lexicon.

For example, in simple preliminary runs of our speech recognition package the computer was able to recognize the seven emotion categories, anger, hatred, sadness, love, joy, fear, and neutral, which we did our best to communicate to it, when speaking the sentence, "Hello Sam, I want to talk to you." In this small exercise we broke the sentence up into three parts, identifying each part as a "word" to the speech recognition system. We then trained each phrase for the seven different inflections. With practice we were able to get very accurate recognition of the intended emotional state.[3] To achieve this we had to be slightly theatrical, but not overly so, and there was a flavor of speaking with someone who was hard of hearing, but again, not overly so.[4]

In addition, while we have not yet pursued the development of the emotion lexicon, the package is capable of hundreds of word nodes in its grammar, sufficient to capture a relatively rich channel for communication from the user to the computer.

Once the lexicon is established, tokens can either be interpreted directly as situations in themselves, or as a set of features indexing into a case-base to retrieve similar cases indicating the presence of an instance of an emotion category. For example, in the first case, from user input of "I am satisfied" we might yield the situation: "the user is *satisfied* now in response to the comparison of her answer with the one just provided by the computer"). In the second case, user input of "I am happy now" (spoken with a hateful inflection), yields the set of features: *user expresses happiness*, and *user's inflection expresses hate*, which in turn retrieves the case of *hatred* masked by a contrasting verbal expression.

Finally, given that the computer is able to determine at least some minimum set of emotion inflections in user speech, we have found that even a simple list of spoken menu choice identifiers (e.g., *"Choice A" – "Choice F"*), can convey much about the user's state which was previously inaccessible.

A sympathetic computer

Sidestepping the issue of whether a a user's *engagement* with a computer system will be increased by dealing with

[2] Although reportedly the work of Eckman may address this.

[3] And this, even though the goal of the developers of the package was to *eliminate* inflection and simply recognize the words.

[4] Interestingly, when not *imagining* a situation consistent with the emotion we were attempting to express by voice, we were consistently less accurate. In addition, it appeared that the computer was biased somewhat in its errors toward the category the user was imagining, for certain pairs of categories, which suggests an interesting study.

an emotionally believable agent, we might first ask what is required to achieve believability. How much does it matter if the computer is transparent in its capabilities and motivations? What level of sophistication is required? How intelligent does the system have to be to achieve a minimal level of believability?

Consider the following: A computer system detects, or is told by the user (as discussed above), that the user is fearful (anxious, worried, scared, – any number of words from the lexicon might be suitable). The system is able to respond by asking what might be considered reasonable questions: "What is the name of the EVENT over which you are fearful?," "How far is this event in the future?" (According to the underlying theory, as embodied in the Affective Reasoner, *fear* is a prospect-based emotion resulting from the likely blocking of a future *event*.) "What is the likelihood of this coming about?" (*Likelihood* is an intensity variable). It is then able to comment on relationships between the various *intensity variables* and then to recall cases from a library, where either the automated agent itself ("I was afraid of getting turned off!" [Frijda and Swagerman, 1987]), or some other agent ("I was previously told the following story by another user...") was *fearful* with similar intensities. Lastly, after verifying that the retrieved case involves an emotion similar to that of the user, the computer agent responds by saying, "I am only a stupid computer program. Nonetheless, in my own simple way, the following is true: I consider you my friend (*friendship* is one of the relationships modeled in the AR). I *pity* you for your *fearful* state. I *hope* to become aware in the future that you are *relieved* about the outcome of situation <*input-situation*>, and that your preservation goals with respect to this are not blocked after all."

Such a system is within the range of our current technology, and representational capabilities. At what level of sophistication are users willing to accept that, in its own (remarkably) simple way, the computer does nonetheless feel pity for them? Given the tendency of users to anthropomorphize even the simplest video game characters, this would seem to be an important question to answer.

Games

Consider the game of poker, and how an emotion model can enhance the believability, and possibly interest, of a computerized version of the game. Using a preliminary model, we have discussed (and built partial toy versions of) the following system:

The computer simulates one, or more, agents who play five card draw poker against the user. Each agent knows the basic rules of poker, including betting. Agents have their own *goals*, *principles*, and *preferences* so that they respond differently to different situations that arise. Bluffing, required if one is to truly simulate poker, leads to fear for some agents, hope for others. Some agents show their emotions (through facial expression, music selection and verbal communication of physical manifestations), others suppress them (as in real life). An agent who has been bluffing and wins, might *gloat*, or if it loses it might feel *remorseful*. An increase in the amount of the "pot" increases the *importance* of winning and thus the intensity of many of the emotions. Likewise *surprise* (generally hard to represent) can be derived from losing even though one has a good hand, or vice versa, thus intensifying the respective emotion. Additionally agents can feel reproach over, for example, a player (such as the user) who is too cautious, or takes to long to decide what to do (based on a violation of the principled customary way of playing). Many variations are possible, such as allowing the computer to attempt to cheat without being caught by the user, and so forth.

One claim that might be made is that to approach a realistic computer model of games such as poker, some model of personality and emotion, minimally a broad and shallow model as described here, is essential. Whether or not this makes the computer a better playing companion is an open question, but again, one that seems worth answering.

References

[Elliott and Ortony, 1992] Clark Elliott and Andrew Ortony. Point of view: Reasoning about the concerns of others. In *Proceedings of the Fourteenth Annual Conference of the Cognitive Science Society*, pages 809–814, Bloomington, IN, August 1992. Cognitive Science Society.

[Elliott and Siegle, 1993] Clark Elliott and Greg Siegle. Variables influencing the intensity of simulated affective states. In *AAAI technical report for the Spring Symposium on Reasoning about Mental States: Formal Theories and Applications*, pages 58–67. American Association for Artificial Intelligence, 1993. Stanford University, March 23-25, Palo Alto, CA.

[Elliott, 1993] Clark Elliott. Using the affective reasoner to support social simulations. In *Proceedings of the Thirteenth International Joint Conference on Artificial Intelligence*, pages 194–200, Chambery, France, August 1993. Morgan Kaufmann.

[Elliott, 1994] Clark Elliott. Multi-media communication with emotion-driven 'believable agents'. In *AAAI Technical Report for the Spring Symposium on Believable Agents*, pages 16–20, Stanford University, March 1994. AAAI, American Association for Artificial Intelligence. THIS PAPER.

[Frijda and Swagerman, 1987] Nico Frijda and Jaap Swagerman. Can computers feel? theory and design of an emotional system. *Cognition & Emotion*, 1(3):235–257, 1987.

[Gaver and Mandler, 1987] William W. Gaver and George Mandler. Play it again sam: On liking music. *Cognition & Emotion*, 1(3):299–322, 1987.

[O'Rorke and Ortony, 1994] Paul O'Rorke and Andrew Ortony. Explaining emotions. *Cognitive Science*, 18:283–323, 1994.

[Ortony *et al.*, 1988] Andrew Ortony, Gerald L. Clore, and Allan Collins. *The Cognitive Structure of Emotions.* Cambridge University Press, 1988.

[Pelachaud *et al.*, 1993] Catherine Pelachaud, Marie-Luce Viaud, and Hussein Yahia. Rule-structured facial animation system. In *Proceedings of the Thir-*

teenth Annual Joint Conference on Artificial Intelligence, Chambery, France, August 1993. Morgan Kaufmann.

[Van Lehn, 1988] Kurt Van Lehn. Student modeling. In Martha C. Polson and J. Jeffery Richardson, editors, *Foundations of Intelligent Tutoring Systems.* Lawrence Erlbaum Associates, 1988.