Tears and Fears: Modeling emotions and emotional behaviors in synthetic agents

Jonathan Gratch
University of Southern California
Institute for Creative Technologies
13274 Fiji Way Marina del Rey, CA 90292
gratch@ict.usc.edu

Stacy Marsella
University of Southern California
Information Sciences Institute
4676 Admiralty, Marina del Rey, CA 90292
marsella@isi.edu

ABSTRACT

Emotions play a critical role in creating engaging and believable characters to populate virtual worlds. Our goal is to create general computational models to support characters that act in virtual environments, make decisions, but whose behavior also suggests an underlying emotional current. In service of this goal, we integrate two complementary approaches to emotional modeling into a single unified system. Gratch's Émile system focuses on the problem of emotional appraisal: how emotions arise from an evaluation of how environmental events relate to an agent's plans and goals. Marsella et al.'s IPD system focuses more on the impact of emotions on behavior, including the impact on the physical expressions of emotional state through suitable choice of gestures and body language. This integrated model is layered atop Steve, a pedagogical agent architecture, and exercised within the context of the Mission Rehearsal Exercise, a prototype system designed to teach decision-making skills in highly evocative situations.

1 INTRODUCTION

A person's emotional state interacts with numerous aspects of mental and physical behavior. Decision-making, actions, memory, attention, voluntary muscles, etc. may all be impacted, which in turn may impact emotional state (e.g., See Berkowitz, 2000). This pervasive impact is reflected in the fact that a person will exhibit a wide repertoire of nonverbal behaviors consistent with emotional state, behaviors that can serve a variety of functions both for the person exhibiting them as well as for people observing them. For example, shaking a fist at someone plays an intended role in communicating information to another person. On the other hand, behaviors such as rubbing one's thigh, averting gaze or a facial expression of fear may have no explicitly intended role in communication. Nevertheless, they may suggest considerable information about them, their emotional arousal, their attitudes and what they are attending to.

This paper will attempt to show how some of this daunting subtlety in human behavior can be modeled by intelligent agents, from the perception of events in the world, to the appraisal of their emotional significance, through to their outward impact on an agent's behavior. The focus for our work is on general software agents that model human performance in rich simulated worlds. In particular, we focus on virtual training environments where intelligent agents interact with a human participant to facilitate the training objectives. Emotions play an important role in such environments by enhancing believability and realism, increasing ones sense of empathy and attachment to synthetic characters, and adding to the suspense of the simulation. Rather than creating carefully crafted models tuned to a specific scenario, we put forth a domain-independent solution that addresses modestly the problem of modeling "task-oriented" emotions – emotions that arise from the performance of a concrete task.

We describe an integration of two research efforts focused on creating engaging and believable characters to populate virtual worlds. Gratch's Émile system focuses on the problem of emotional appraisal: how emotions arise from an evaluation of how environmental events relate to an agent's plans and goals (Gratch, 2000). Marsella's IPD system addresses different, complementary aspects of the complex interplay of emotion, cognition and behavior. In this paper, we will be concerned with how IPD models the impact of emotions on behavior, in particular the impact on the physical expressions of emotional state through suitable choice of gestures and body language (Marsella et al. 2000). This integrated model is layered atop Steve, a pedagogical agent architecture designed to support plan-based reasoning and flexible interactions with a human student. (Rickel and Johnson, 1999)

A secondary goal is to illustrate the workings of this unified approach within the context of a rich virtual environment. We describe how our emotional models contributed to the development of the Mission Rehearsal Exercise (MRE) system, a prototype training environment designed to teach decision-making skills in highly evocative situations. The MRE system provides an immersive learning environment where participants can experience the sights, sounds and circumstances they will encounter in real-world scenarios while performing mission-oriented training (Figure 1). Intelligent agents control characters in the virtual environment with which the participants must interact in the course of their training, and our emotional models attempt to augment the believability, realism and suspense of these interactions.

2 FROM COGNITION TO EMOTION

Many psychological theories of emotion emphasize the relationship between emotions and cognition. How one responds to some external events seems closely tied to their implications for ones plans and goals (Ortony et al, 1988; Lazarus, 1991). Even purely mental "events" can evoke strong emotions: most of us have ex-

Figure 1: Virtual Bosnian village

perienced a flash of insight in our research that leaves us with intense feelings of joy, only to be crestfallen seconds later by the realization of some crucial flaw. Emotions clearly have a strong influence over our decision-making abilities as well (Damasio 1994; Sloman, 1987).

Gratch (2000) has argued that artificial intelligence planning techniques provide a powerful and general mechanism for modeling a key aspect of the interplay between cognition and emotion, namely "task-oriented" emotions (those emotions that arise from the performance of some concrete task). Adopting a plan-based approach has some key advantages. By maintaining an explicit representation of an agent's plans one can easily reason about future possible outcomes – essential for modeling emotions like hope and fear that involve future expectations. Explicit representations allow one to recognize how the plans or actions of an agent facilitate or hinder the goals of others – essential for modeling emotions like anger or reproach which typically involve multiple actors. A plan-based approach also models some of the dynamics of emotional state by tying appraisals to the current state of plans in memory which changes via the information processing of the planner. Finally, by providing an explicit and rich reasoning infrastructure, plan-based approaches facilitate models of how emotions impact decision-making.

Émile (Gratch, 2000) provides a rich plan-based model of emotional appraisal, the task of assessing the relationship between external events and an agent's internal beliefs, plans, desires, social norms and so forth. Émile does not explicitly address the problem of how this assessed emotional state impacts behavior, or how to effectively convey this state to a human participant. Building on Elliott's (1992) construal theory, Émile characterizes the emotional impact of external events through a set of knowledge structures called *construal frames*. These frames are created whenever certain syntactic features are recognized in the agent's internal state. For example, whenever the agent adopts a new goal (or is informed of a goal of some other agent), frames are created to track the status of that goal. Each frame describes the appraised situation in terms of a number of specific features, including the point of view from which the appraisal is formed, the desirability of the situation, whether the situation has come to pass or is only a possibility and whether the situation merits praise or blame. These features are derived from domain-independent rules that examine the state of plan memory, an advance over prior approaches that utilize large numbers of domain specific rules to form the same assessment. Some examples of these domain independent rules (there are about thirty) include:

> If an agent has a goal and no known action achieves this effect, this is undesirable
>
> If an agent intends to use an action to achieve a goal and a subsequent action defeats the effect of this action, this is undesirable
>
> If an agent intends to perform an action that achieves a goal for another agent, this is praiseworthy

Émile also draws heavily on the explicit plan representation to derive the intensity of emotional response, incorporating the view of Oatley and Johnson-Laird (1987) and Neal Reilly (1996) that emotions are related to changes in the perceived probability of goal attainment. Intensity relates to the probability of the event in question (e.g. the probability of goal achievement or the probability of a threat) and the utility of the impacted goals, both of which are derived from the current plan structure. The importance of subgoals is related to how they further intrinsic goals. As intensity is based on the current plans, the assessment is a reflection of their current state and changes with further planning.

Each appraisal frame corresponds to an emotion instance. These instances are aggregated into "buckets" corresponding to emotions of the same type, and instances decay in intensity over time. Thus, threats to multiple goals will be aggregated into an overall level of fear. The aggregate buckets roughly correspond to the overall assessment of the agent's emotional state and are used to drive emotional expression as discussed next.

3 FROM EMOTION TO BEHAVIOR

People exhibit a wide repertoire of nonverbal behaviors consistent with their emotional state, through facial expressions, gestures, body posture, etc. Whether these behaviors are intentionally communicative or not, they often suggest considerable information about a person, their emotional arousal, their attitudes and what they are attending to. Indeed, observers can reliably infer a person's emotions and attitudes from their nonverbal behaviors (Ekman et al. 1969) and therefore potentially respond in a variety of ways. Thus, when creating virtual humans that maintain and convey an internal emotional state, we must ensure that the agent's performance suggests a corresponding emotional state to the observer, or run the risk of creating confusion or disbelief.

For our purposes, we need a model of agent behavior that appropriately suggests an emotional undercurrent. Such a model must address particular concerns. Of particular concern for the agent characters we design is that they provide convincing portrayals of humans facing difficult, dangerous problems. To that end, they must have emotionally revealing nonverbal behaviors and expressions consistent with deeply evocative/disturbing situations. These behaviors must also change in concert with the emotional state of the characters; obviously people express themselves differently when sad, happy or angry. Further, they must have behaviors unique to the individual since not everyone exhibits the same behaviors, in the same way.

Another key concern here is that the agent's mix of nonverbal behavior at any time appear emotionally consistent. Consider severe depression. There are many ways to convey severe depression; it may be effective for an agent to appear withdrawn, inattentive, or perhaps hugging themselves. However, if a supposedly depressed agent used various open, communicative gestures such as beats (McNeill, 1992) while expressing something to another agent, then the performance may not "read" correctly. The behavior may not appear consistent with depression. This is especially so if the agent had previously been exhibiting behaviors more consistent with depression. In fact, the mix of gestures used by an agent must be coherent and avoid unintended interpretations. For example, people don't tend to nonchalantly use deictic gesture while simultaneously averting their gaze due to mild feelings of anger or guilt. Such behavior may look unnatural, inconsistent, or may convey a different shade of meaning depending on context. Which is not to say that the overall mix of behaviors should always be monolithic. People do say one thing while expressing another. At the least, the mix of nonverbal behaviors often shade the meaning of what is said or communicated nonverbally. Returning to the previous example, if an agent does combine deictic gesture with gaze aversion, it may shade the interpretation dramatically, towards an expression of extreme emotion and a desire to control that emotion. For example, the agent is so disgusted with the "listener" they can't bear to look at them.

Implicit in these various concerns is that the agent has what amounts to a resource allocation problem. The agent has limited physical assets, e.g., two hands, one body, etc. At any point in time, the agent must allocate these assets according to a variety of demands, such as performing a task, communicating, or emotionally soothing themselves. For instance, the agent's dialog may be suggestive of a specific gesture for the agent's arms and hands while the emotional state is suggestive of another. The agent must mediate between these alternative demands in a fashion consistent with their goals and their emotional state

3.1 Physical Focus

To address these concerns, the emotional behavior component of this agent architecture relies on the Physical Focus model that was part of the IPD system (Marsella et al. 2000). The IPD work was in turn heavily influenced by work on non-communicative but emotionally revealing nonverbal behavior (Freedman 1972) as well as Lazarus's (1991) delineation of emotion-directed versus problem-directed strategies for coping with stress.

The Physical Focus model bases an agent's physical behavior in terms of what the character attends to, how they relate to themselves and the world around them, specifically whether they are focusing on themselves and thereby withdrawing from the world or whether they are focusing on the world, engaging it. The intent of the model is to refine down all the variegated ways in which emotional state impacts the agent's nonverbal behavior into distinct modes of relating to the world that provide a consistent resolution of the resource allocation problem.

The choice of nonverbal behaviors is determined by the agent's Physical Focus mode, which characterizes the mix of behaviors exhibited by an agent. At any point in time, the agent will be in a specific mode based on emotional state that predisposes it to use particular nonverbal behavior in a particular fashion. Each behavior available to an agent is categorized according to which subset of these modes it is consistent with. Any specific nonverbal behavior, such as a particular nod of the head, may exist in more than one mode and conversely a type of behavior, such as head nods in general, may be realized differently in different modes. Transitions between modes are based on emotional state.

By grouping behaviors into modes, the physical focus mode attempts to mediate competing communicative and non-communicative demands on an agent's physical resources, especially gesturing and gaze, in a fashion consistent with emotional state. This grouping model is designed with the intent that it be general across agents. However, realism also requires that specific behaviors within each mode incorporate individual differences, as in human behavior. For example, we would not expect a mother's repertoire of gestures to be identical to that of an army sergeant.

Marsella et al. (2000) discuss five distinct focus modes. Here we discuss the three modes that are most relevant to the current application: body-focus, transitional and communicative. Body focus is marked by a self-focused attention, away from the conversation and the problem-solving behavior. Emotionally, it is associated with considerable depression or guilt. Physically, it is associated with the tendencies of gaze aversion, paused or inhibited verbal activity and hand to body stimulation that is either soothing (e.g., rhythmic stroking of forearm) or self-punitive (e.g., squeezing or scratching of forearm). The agent exhibits minimal communicative gestures such as deictic or beat gestures (McNeil 1992, Cassell & Stone 1999) when in this mode. Transitional indicates an even less divided attention, less depression, a burgeoning willingness to take part in the conversation, milder conflicts with the problem solving and a closer relation to the listener. Physically, it is marked by hand to hand gestures (such as rubbing hands or hand fidgetiness) and hand to object gestures, such as playing

defPlan handle-accident

tasks: {accident, lt-arrives, evaluate, implore, evacuate, move-out, reassure, treat}

causal constraints:

{accident	{disables	*child-healthy*}		end-handle-accident}
{evaluate	{disables	*facilities-ok*}		end-handle-accident}
{move-out	{disables	*troops-helping*}		evacuate}
{treat	{enables	*child-healthy*	0.4}	end-handle-accident}
{implore	{enables	*help-requested*}		reassure}
{lt-arrives	{enables	*authority-present*	0.7}	evacuate}
{lt-arrives	{enables	*authority-present*	0.7}	implore}
{lt-arrives	{enables	*authority-present*	0.7}	treat}
{evacuate	{enables	*facilities-ok*	0.65}	treat}
{reassure	{enables	*troops-helping*	0.5}	evacuate}

ordering constraints: accident $\succ$ move-out; lt-arrives $\succ$ reassure

role assignments: mother {implore}; lt {accident move-out reassure evacuate}; medic {treat evaluate}

defGoal *child-healthy* {boy-health good} :probability 0.2 :location victim :concerns {{mother 80.0} {lt 40.0}}

Figure 2: A portion of the mother's domain knowledge

with a pen. There are more communicative gestures in this mode but they are still muted or stilted. Finally, communicative indicates a full willingness to engage in the dialog and problem solving. Physically, it is marked by the agent's full range of communicative gestures, use of gaze in turn taking, etc.

Transitions between modes are based on emotional state derived from the appraisal model. Rules map the current aggregate emotional state into a specific mode. High levels of guilt or sadness, both in absolute terms and relative to other emotion levels, induces transitions towards Body Focus. Increased hope or anger induces transitions towards Communicative. Transitional Focus lies between these extremes. Transitions are designed with hysteresis so that the agent does not readily pop into and then out of a mode.

4 FROM BEHAVIOR TO COGNITION

The agent's Physical Focus mode does more than convey an impression, via their behavior, of whether they are inwardly or outwardly directed. The focus mode also impacts the agent's awareness of, and attention to, external stimuli. This in turn impacts their decision-making and subsequent behavior as related to these stimuli in a fashion consistent with their physical focus.

Specifically, the focus mode influences an agent's sensitivity to external stimuli. Currently this is realized in a simple fashion. Rather than modeling the full complex interplay of how people can focus their perception and attention (Wells & Matthews, 1994), we provide a domain specific mechanism for ranking stimuli by their intensity. Certain stimuli are then filtered depending on if the focus mode is inner (Body Focus) or outer directed (Communicative).

5 MISSION REHEARSAL EXERCISE

We have unified ideas from Émile and IPD for modeling emotional characters within the Mission Rehearsal Exercise (MRE) system, a real-time virtual training environment. The goal of the MRE system is to provide an immersive learning environment where the participants experience the sights, sounds and circumstances they will encounter in real-world scenarios while performing mission-oriented training. The MRE system pushes the state-of-the-art in simulation technology through the integration of high-fidelity real-time graphics, intelligent agents, immersive audio and interactive story. An initial prototype of the system now exists and its improvement is a subject of ongoing research (Swartout *et al*, 2000).

Intelligent agents control characters (virtual humans) in the virtual environment, playing the roles of locals, friendly and hostile forces, and other mission team members. The goal is to support realistic face-to-face interactions, requiring an emphasis on creating "broad agents" that integrate motor skills, problem solving, emotion, gestures, facial expressions, and language.

MRE creates a heightened sense of realism through the use of immersive audio synchronized to the events occurring in the virtual world. This involves simulating the characteristics of the human ear to create immersive acoustics, canceling cross talk in real-time for rendering over loudspeakers, and correcting local acoustical environment using psychoacoustic principles.

MRE's training scenarios are created with the input of professional storywriters in an attempt to engage the learners as they are achieving pedagogical goals related to the mission. A training scenario is essentially an interactive story whose outcome depends on the decisions and actions that participants take during the simulation. The ultimate goal is to prepare decision-makers who must think on their feet under realistically bewildering circumstances.

The initial prototype contains a mixture of three interactive and about forty pre-scripted virtual humans that play the parts of characters in a military peacekeeping exercise. In the prototype scenario, a human participant is in charge of a platoon of soldiers that have become involved in an automobile accident while driving to meet another platoon in need of reinforcement. The student must decide how best to allocate his forces between the conflicting goals of assisting an injured civilian and completing his mission, all under the watchful eyes of a "ZNN" cameraman.

5.1 Steve

The three interactive agents in the scenario are modeled using the Steve system of Rickel and Johnson (1999), and have been integrated with greatly improved body and motion models developed commercially by Boston Dynamics. Steve is a plan-based pedagogical agent architecture designed to interact with human participants in well-structured environments. Students can interact with Steve agents via speech recognition, asking questions or giving commands as they relate to some concrete task that must be performed in the virtual world.

We have augmented one of these interactive Steve agents, the mother of the injured civilian, with our emotional models. This allows her to add emotional color to her actions as well as to respond in an emotionally appropriate way to the student's actions or events in the world. Steve's design facilitates this integration. Both Steve and Émile are implemented in Soar (Newell, 1990) and share quite similar plan representations. This allowes us to integrate Émile's machinery for inferring emotional state into Steve with very little modification. Furthermore, Soar makes it easy to integrate additional knowledge into an existing system. Marsella's IPD model of body focus and gesturing were straightforwardly implemented as additional procedures that can be interleaved with Steve's decision-making.

Figure 2 illustrates a (slightly paraphrased) portion of the mother's domain knowledge. Steve's representation language allows one to specify a space of possible plans that is compared with the current world state to decide the best current course of action. The figure illustrates a task decomposition schema for the "handle-accident" task. This task is broken down into several subtasks (accident, lieutenant-arrives, etc.). The schema also specifies ordering and causal relationships between tasks (the lieutenant arriving enables the condition that authority is present with 70% probability, which is a precondition of treating the victim). Finally, the schema specifies which agents are responsible for executing which tasks (the medic is responsible for evaluating and treating the child). The figure also illustrates how one defines conditions used as preconditions or effects of plan steps. "Child-healthy" is a proposition that is true if the perceptual state indicates that the "boy-health" attribute has a value of "good". The system *a priori* expects with twenty percent likelihood that this goal, should it be unsatisfied, can be attained. The location attribute tells Steve where to look or gesture when referring to this condition. Finally one can specify a set of agents who are concerned with the truth value of this condition and the utility they place on it being satisfied (the mother cares a lot about the boy being healthy). This information is used to infer the intrinsic and extrinsic utility of goals and subgoals.

5.2 Expressive Characters

The Physical Focus routines interface with human avatars modeled in Boston Dynamics, Inc.'s PeopleShop run-time environment. PeopleShop provides body models that can be either prescripted or controlled in real-time through an API. Character animation is based on motion capture: an actor wearing special sensors is recorded performing certain actions and this data is carved into segments and played back on demand. Boston Dynamics worked with us to provide a number of custom features and behaviors including procedural control of gaze and the integration of their software with face models provided by another corporation, Haptek, that provides procedural control over facial expressions.

Motion capture is good for creating natural body movements but it is rather awkward to use in conjunction with our reasoning and emotional models. Motion capture is inflexible and you have to anticipate in advance all of the actions and gestures that you will require for the scenario. This inflexibility is especially problematic for our emotional models. A character's motions and gestures should change noticeably as a function of the current emotional state. Ideally, we could procedurally adjust the behavior in realtime. In fact, some research has begun to explored how to alter motion-capture in just such a fashion (Chi et al., 2000). Until such technology is available, our solution has been to carefully organize the motion capture segments to get the desired flexibility and range of emotional expression.

Figure 3 illustrates the representation of motion capture segments. They are organized into a finite-state machine, loosely structured as a hub-and-spoke. The hubs are a set of stationary body poses that correspond to the three Physical Focus modes: body, transitional, and communicative. The spokes are various behavior segments that transition from a hub, through a sequence of movements, then back to the hub. Behaviors are further sub-divided into task related behaviors (such as imploring the lieutenant) and idle-time behaviors (such as rocking back and forth). Behaviors generate call-backs to the agent, informing it when the behavior is complete and what state the body is in.

When selecting a behavior, the agent compares the current body state, emotional state, physical focus, and whether a behavior is currently executing. Some behaviors, such as task related behaviors and reactions to perceptual events (e.g., look at an explosion) have precedence and interrupt other ongoing behaviors. If neither of these behaviors are pending, the system simply chooses some behavior that is consistent. Responses to external events are further modulated by Physical Focus (the mother doesn't respond to low intensity perceptual events when in body focus). In some cases multiple behaviors may apply (a resource conflict). Soar provides a general arbitration scheme that resolves such conflicts.

5.3 Integration Issues

Steve is designed to model team behavior; however, in this scenario the mother and the soldiers, while sharing some similar goals, would hardly be described as being on the same team. In particular, they have expectations about the desired course of events. We chose to model this by providing different domain knowledge to the mother and the soldier agents. The models are similar and refer to many of the same tasks and perceptual events, but this allows the mother to have a different understanding of the flow of events. For example, the mother understands the soldiers plans in much less detail and in one case mis-interprets the intent of one of the soldier's actions (when the lieutenant sends some squads forward to reinforce the other platoon – "move-out" – the mother infers that the troops are no longer helping her child – {disables troops-helping}).

Some software modifications were necessary to integrate Steve with Émile. Steve's representation language had to be extended to represent the probabilities and utilities needed for Émile to calculate the intensity of certain emotional responses. Steve also had to be extended to infer that certain tasks could disable conditions needed by other tasks (after the medic evaluates the child it is clear that the facilities are inadequate to treat the child without evacuating him to another location). This is necessary for reasoning about the undesirability of certain events. We also slightly changed how Steve processes information, essentially slowing its reaction time to draw out the dynamics of changes in mental state. Finally, we incorporated some knowledge from Émile's planning system to allow Steve to detect un-planned for perceptual events and express an appropriate startle reflex.

Some changes were also needed to integrate IPD's Physical Focus into the current system. The original body models in IPD were two dimensional, composed of many roughly orthogonal parts (hands, arms, etc) that could be separately animated. In MRE, the animation is three dimensional, far more realistic looking, though much more constrained as it is based on motion capture. This led to several simplifications. Most notably, because of the reduced flexibility of motion capture, and consequently the reduced need to manage the agent's behavior, we only implemented the three Physical Modes discussed above. These modes then served to drive our specification of what behavior to capture.

In the MRE system, Physical Focus uniformly impacts the agent's deliberative (task-related) behaviors, idle behaviors as well as their attention. To incorporate the impact on deliberative behavior, we modified the underlying Steve system so that when performing a task the selection of the specific behavior could be determined by physical focus mode. As an example, the mother will implore the Lieutenant to help her child differently when in communicative mode as opposed to transitional mode. Physical focus also makes behavioral choices when the agent is not explicitly engaged in a task (idle behaviors). Finally, we added to Steve the ability to react or not to react to unexpected events in the environment based on physical focus. For instance, when in body focus mode the mother is less attentive to minor events that occur in the environment. As of yet, certain capabilities that were part of the Physical Focus model as realized in IPD have not been realized in MRE. In particular, IPD considered both deliberative emotional expression (those consciously added to convey a certain meaning) as well as non-deliberative emotional expression (those arising from emotional appraisal). In MRE we have focused exclusively on non-deliberative emotions.

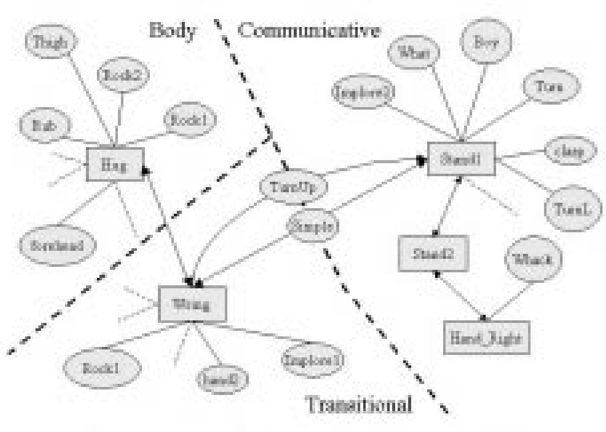

Figure 3: Organization of task-related and idle behaviors

Physical Focus also requires an appraisal of anxiety, which Émile did not support. According to most psychological theories, anxiety is treated as a non-specific threat to a goal in contrast to fear, which is treated as a specific threat. Émile previously only considered specific threats in its models (i.e., one task has an effect that disables a precondition of some other task). In the current implementation, we use the probability model to infer non-specific threats. If a task achieves predicate P with some probability less that 1.0, there is a non-specific threat to the achievement of P. It is non-specific in the sense that the goal may not be achieved, with probability 1-Pr(P), yet there is no explicit reason why not (as opposed to a goal which has a low probability of achievement because an anticipated task disables it with high probability). This covers anxiety arising from non-specific threats to goal achievement, but does not account for other sources of anxiety, for example non-specific threats to already achieved goals. A more complete model of anxiety is the subject of future work.

5.4 Illustration

We now walk through some of the key points of the scenario as they relate to the mother to illustrate how the emotional model influences her behavior. In the opening scene, the mother is waiting for the lieutenant to arrive, which she views as a precondition for her child to be treated. She is somewhat angry at the lieutenant as she perceives him as responsible for the accident (as the lieutenant is assigned the role of executing the "accident" task). Initially she believes the facilities-ok is satisfied, meaning she has the simple plan in memory that the lieutenant should arrive and her child will be treated, neither task being under her control. Since her child is hurt she has high levels of distress. Since the lieutenant arriving and the treatment tasks have low probability effects (non-specific threats), she is also extremely anxious, though also somewhat hopeful. The high distress and anxiety leads her to have an inner-directed Physical Focus. Her body gestures are directed inward and she will not attend to most stimuli.

When the lieutenant arrives in his jeep the mother perceives that "authority-present" is now satisfied in the current state. As this subgoal is now attained, the non-specific threat associated with its attainment disappears, the probability that the child will be treated increases somewhat, and the mother's anxiety and distress diminish somewhat. This is enough to transition her into transitional focus, her gestures become more outward directed and she attends to more perceptual stimuli and her child.

The lieutenant asks for a report of the child's health. The mother attends to this exchange and essentially eavesdrops on the medic's statement that the facilities are inadequate. Steve's reasoning mechanism infers the current plan is invalid and that the child must now be evacuated. This change in plans leads to a change in evaluation of her goals and thus a change in emotional state. She lowers her estimate that the child will be successfully treated and the evacuation introduces several new sources of dis-

Figure 4: Subdued and angry variants of imploring the lieutenant

tress and anxiety. She transitions back to body focus, which is articulated physically through visible and audible weeping.

Later in the scenario, the lieutenant orders one or two squads forward ("move-out") to reinforce the platoon downtown. The mother interprets this as disabling her subgoal that the troops are helping her child. The strength of this interpretation is influenced by the number of squads that move forward (implemented by domain-specific rules that infer conclusions from the agent's perceptual input). The emotional model treats this as a blameworthy event, causing the mother to become angrier at the troops. This anger is sufficient to transition her into communicative mode. The mother also updates her plans, deciding that the troops will return to helping her child if she implores them to stay (via the "implore" task). Her body language in performing this action is colored by her body focus and anger level, either remaining seated and gesturing mildly or raising to a standing position and gesturing strongly (Figure 4).

6 DISCUSSION

This project is still in its early stages (the initial prototype was completed at the end of September 2000). From a research perspective the biggest limitation is the lack of evaluation. Is MRE a viable learning environment? Does the addition of emotional models increase the realism of the scenario? Do people find the character's reactions plausible? How do emotional models impact the learning experience? Our plan is to begin formal evaluations in the coming year in conjunction with other research groups in the psychology and communications departments at the University of Southern California. Our anecdotal feedback has been encouraging. We have demonstrated the system to a number of military personal and those who served in Bosnia or Kosova seemed strongly affected by the experience. One U.S. Army Colonel began relating a related incident after seeing the demo, became quite emotional, and concluded by saying, "this system makes people feel, and we need that." In another anecdote, someone playing the role of the lieutenant became agitated when the mother character began yelling at him and when she wouldn't respond to his reassurances (she cannot be mollified when her anger exceeds some threshold).

While this is encouraging, a number of problems must be addressed before we can exercise the MRE system's potential as a learning environment and evaluate its effectiveness. The prototype is not very interactive. Although the system uses speech recognition, the recognition grammar is quite limited. Furthermore, while there is some variability in the order events can occur, the scenario is essentially a linear narrative with one branch point (based on how many squads the lieutenant sends to reinforce the other platoon). As such, the scenario does not exercise the flexibility of our emotional models, and provides little evidence that the emotional responses would appear appropriate over a wider range of interactions. Before performing any rigorous evaluation we need allow the student to exercise more flexibility by adding domain knowledge to cover other possible decisions. Steve's reasoning capabilities will also have to be augmented as Steve has been designed to teach a single correct procedure (e.g. how to repair an engine) rather than a range of possible alternatives. This lack of alternatives also makes it difficult to model the impact of emotional state on decision-making, which is most naturally encoded as some preference over alternative courses of action.

Another limitation is our current reliance on motion-capture data for the motions and gestures of the animated characters. Motion capture generates fluid and realistic motion but it is not well suited for real-time interactions. Our solution – a hub and spoke model with short motion-capture segments – allowed us to express some of the dynamics of the mother's emotional state, but there is no substitute for procedural control. As a solution we propose to integrate our work with Badler's EMOTE system (Chi et al., 2000). Emote can procedurally "morph" motion capture date along a number of dimensions, making a gesture seem to have more or less energy and gestures to be directed more inward or outward, much as is advocated by the Physical Focus model.

Finally, there are a number of limitations in how the system infers emotional state that need adjustment or re-thinking in light of this application. One key issue is the notion of responsibility. For example, whom should the mother blame for the accident? The troops? Herself? Our sense is she should have a shared sense of responsibility and that this sense should change dynamically, influenced by her emotional state and subsequent actions of the troops. Currently, we simply use Steve's responsibility constraints to assign blame. Our treatment of anger is also too simplistic. Anger seems influenced by the extent to which we decide someone intended the offending action and the extent to which they show remorse or attempt to redress the offence. We suspect the explicit use of plans can assist in forming such assessments, but we still sorting out how.

These limitations not withstanding, the integration of plan-based appraisal of emotional state with the model of Physical Focus provides a great deal of architectural support for emotional modeling. Furthermore, anecdotal evidence suggests that people not only find the agent's emotions to be plausible, but in fact, people occasionally responded emotionally to our agents.

Acknowledgments

This research was funded by the Army Research Institute under contract TAPC-ARI-BR and by the U.S. Army under contract to the Institute for Creative Technologies. The content of this article does not necessarily reflect the position or the policy of the Government, and no official endorsement should be inferred.

References

Beaudoin, L. 1995. *Goal Processing in Autonomous Agents*. Ph.D Thesis (CSRP-95-2), Univ. of Birmingham.

Berkowitz, L. 2000. *Causes and Consequences of Feelings*. Cambridge University Press, 2000.

Cassell, J. & Stone, M. 1999. Living Hand to Mouth: Psychological Theories about Speech and Gesture in Interactive Dialogue Systems. AAAI Fall Symposium on Narrative Intelligence.

D. Chi, M. Costa, L. Zhao, and N. Badler: "The EMOTE model for Effort and Shape," ACM SIGGRAPH '00, New Orleans, LA, July, 2000, pp. 173-182,

Damasio, A. 1994. *Descartes' Error*. Avon Books, NY, NY.

Elliott C. D. 1992. *The Affective Reasoner: A Process Model of Emotions in a Multi-agent System*. Ph.D Thesis (TR#32), Northwestern University.

Ekman, P. and Friesen, W.V. The Repertoire of NonVerbal Behavior: Categories, Origins, Usage and Coding. *Semiotica* 1:49-97, 1969.

Freedman, N. The analysis of movement behavior during clinical interview. In *Studies in Dyadic Communcation*, 1972, 153-175.

Gratch, J., 2000. Émile: marshalling passions in thraining and education. *Proceedings of the Fourth International Conferenced on Intelligent Agents*, Barcelona, SPAIN.

Lazarus, R.S. 1991. *Emotion and Adaptation*. Oxford Press.

Lester, J., Towns, S.G.,FitzGerald, P.J. 1999 Achieving Affective Impact: Visual Emotive Communication in Lifelike Pedagogical Agents. *International Journal of AI in Education, 10 (3-4)* pp. 278-291.

Marsella, S. Johnson, W.L. & LaBore, C. 2000. Interactive Pedagogical Drama. Proceedings of the Fourth International Conference on Autonomous Agents. Pp 301-308.

McNeil, D. *Hand and Mind*. University of Chicago Press, Chicago IL, 1992.

Neal Reilly, W.S., 1996. *Believable Social and Emotional Agents*. Ph.D Thesis CMU-CS-96-138. Carnegie Mellon Univ.

Newell, A. 1990. *Unified Theories of Cognition*. Harvard Press.

Oatley, K. & Johnson-Laird, P.N. 1987. Towards a Cognitive Theory of Emotions. *Cognition and Emotion*, 1 (1).

Ortony A., Clore, G. L., & Collins, A. 1988. *The Cognitive Structure of Emotions*. Cambridge University Press.

Rickel, J. & Johnson, L. 1999. Animated agents for procedural training in virtual reality: perception, cognition, and motor control. *Applied Artificial Intelligence, v13*:343-382.

Sloman, A. 1987. Motives, mechanisms and emotions. *Cognition and Emotion*, 1, pp 217-234.

Swartout, W., Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., Labore, K., Lindheim, R., Marsella, D., Moore, B.,Morie, J., Rickel, J., Thiebaux, M., Tuch, L., Whitney, R. 2001. Towards the Holodeck: Integrating Graphics, Sound, Character and Story, in *Proceedings of the Fifth International Conference on Autonomous Agents*, Montreal, CANADA.

Wells, A., and Matthews, G. 1994. *Attention and emotion: a clinical perspective.* Lawrence Erlbaum Associates.