Emotion and Dialogue in the MRE Virtual Humans

David Traum, Stacy Marsella, and Jonathan Gratch

Institute for Creative Technologies
University of Southern California
Marina del Rey, CA, USA

Abstract. We describe the emotion and dialogue aspects of the virtual agents used in the MRE project at USC. The models of emotion and dialogue started independently, though each makes crucial use of a central task model. In this paper we describe the task model, dialogue model, and emotion model, and the interactions between them.

1 Introduction

In the Mission rehearsal exercise project at USC, we have implemented virtual humans [Rickel et al., 2002], with a large range of functionality, who can interact with people and other virtual humans in a range of roles, for simulation-based training. The virtual humans use the core architecture inherited from the Steve agents [Rickel and Johnson, 2002], but add significant new capabilities, including new bodies and extended non-verbal behaviors, a new more extensive spoken language dialogue capability [Traum and Rickel, 2002,Traum et al., 2003a,Traum, 2003], and emotional modelling [Gratch and Marsella, 2001,Marsella and Gratch, 2002], [Marsella and Gratch, 2003]. In this paper we outline some of the dialogue and emotion capabilities of these agents, and describe how each influences the other, concluding with current integration work.

In Section 2, we describe the team task model, used as a basic object of analysis by both the emotion and dialogue models. In Section 3, we summarize the dialogue model. In section 4, we describe the emotion model. In section 5 we describe the ways in which we have started using each of these models to inform the processing of the other. Finally, we conclude in section 6 with our current integration work in these areas.

2 The Virtual Human Task Model

The ability of our agents to collaborate with human and agent teammates on tasks in the virtual world stems from their understanding of those tasks. Agents must understand task goals and how to assess whether they are satisfied, the actions that can achieve those goals, the selection and execution of those actions, and how to adapt execution to unexpected events. To provide this understanding, our agents use domain-independent reasoning algorithms operating over a

general, declarative representation of team tasks, and this representation is used to encode their domain-specific task knowledge for a given training scenario (or class of scenarios). The task model also serves as a basic source of reasoning both for dialogue (talking about tasks, negotiating on tasks [Traum et al., 2003b]) and for reasoning about emotion (how an agent feels, given its plans, goals, and beliefs).

Our task representation extends earlier work on virtual humans for team training [Rickel and Johnson, 2002], augmenting this relatively standard plan representation so that it supports conditional, decision-theoretic and intentional reasoning. Each task description includes of a set of steps, each of which is either a primitive action (e.g., a physical or sensing action in the virtual world) or an abstract action (i.e., itself a task). Abstract actions give tasks a hierarchical structure. There may be ordering constraints among the steps, which define a partial order. Interdependencies among steps are represented as a set of causal links and threat relations [McAllester and Rosenblitt, 1991]. Each causal link specifies that an effect of a step in the task achieves a particular goal that is a precondition for another step in the task (or for termination of the task). Threat relations specify that an effect of a step threatens a causal link by unachieving the goal before it is needed.

To support decision-theoretic reasoning, preferences over action effects are represented by numeric utility values that are either primitive (indicating that the effect has intrinsic worth for some entity) or derived (indicating that the effect has worth as a means towards some end, as in a subgoal) and a probability calculus to indicate the likelihood of certain actions and effects.

To support intentional reasoning, action effects and other state predicates are tagged with a belief, indicating if the virtual human has committed to a particular truth value concerning this predicate. The probability calculus represents the virtual human's measure of belief. Actions and action effects are also tagged with an intention, indicating the virtual human's commitment to bringing about the indicated action or effect.

Finally, to support negotiation, the representation encodes multiple, exclusive ways to achieve goals. These alternatives may differ in terms of their effects, likelihood, and utility, entities involved, etc.

An agent's task model represents its understanding of the task in general, independent of the current scenario conditions. To guide execution of the task and robustly handle unexpected events that require adaptive execution or replanning, agents use a partial-order planning algorithm over the task model. Agents continually monitor the state of the virtual world via messages from the simulator [Rickel and Johnson, 1999] that are filtered to reflect perceptual limitations [Rickel et al., 2002]. The result of this planning algorithm specifies how the agent privately believes that the team can collectively complete the task, with some causal links specifying the interdependencies among team members' actions. Agents continually revise this private plan as the scenario unfolds.

3 Dialogue Model

Our agents use a rich model of dialogue that is closely linked with the task model both for interpretation of utterances as well as for decisions about when the agent should speak and what to say. Our dialogue model supports multiple simultaneous conversations among potentially overlapping groups of interlocutors in a shared virtual world [Traum and Rickel, 2002].

We follow the Trindi project approach to dialogue management [Larsson and Traum, 2000]. The part of the context deemed relevant for dialogue modelling, termed *information state*, is maintained as a snapshot of the dialogue state. This state is then updated by dialogue moves, seen as abstract input and output descriptions for the dialogue modeling component. A complex environment such as the MRE situation obviously requires a fairly elaborate information state to achieve fairly general performance within such a domain. We try to manage this complexity by partitioning the information state and dialogue moves into a set of *layers,* each dealing with a coherent aspect of dialogue that is somewhat distinct from other aspects.

- contact
- attention
- conversation
  - participants
  - turn
  - initiative
  - grounding
  - topic
  - rhetorical
- social commitments (obligations)
- negotiation

Fig. 1. Multi-party, Multi-conversation Dialogue Layers

The layers used in the current system are summarized in Figure 1. The *contact* layer [Allwood et al., 1992,Clark, 1996,Dillenbourg et al., 1996] concerns whether and how other individuals can be accessible for communication. Modalities include visual, voice (shout, normal, whisper), and radio. The *attention* layer concerns the object or process that agents attend to [Novick, 1988]. Contact is a prerequisite for attention. The *Conversation* layer models the separate dialogue episodes that go on during an interaction. Each conversation consists of a number of sub-layers, each of which may have a different information content for conversations happening at the same time. The *participants* may be active speakers, addressees, or overhearers [Clark, 1996]. The *turn* indicates the (active) participant with the right to communicate (using the primary channel) [Novick, 1988,Traum and Hinkelman, 1992]. The *initiative* indicates the participant who is controlling the direction of the conversation [Walker and Whittaker, 1990]. The *grounding* component of a conversation tracks how information is added to the common ground of the participants [Traum, 1994]. The conversation structure also includes a *topic* that governs relevance, and *rhetorical* connections

between individual content units. Once material is grounded, even as it still relates to the topic and rhetorical structure of an ongoing conversation, it is also added to the social fabric linking agents, which is not part of any individual conversation. This includes *social commitments* — both obligations to act or restrictions on action, as well as commitments to factual information [Traum and Allen, 1994,Matheson et al., 2000]. There is also a *negotiation* layer, modeling how agents come to agree on these commitments [Baker, 1994,Sidner, 1994]. More details on these layers, with a focus on how the acts can be realized using verbal and non-verbal means, can be found in [Traum and Rickel, 2002]. We focus here on the level of social commitments, as this has the most direct connection to the task model and the emotion model.

3.1 Obligations and Social Commitments

Core speech acts have functions related to influencing the topic under discussion and establishing and resolving the commitments and obligations of speakers and other conversational participants towards states and actions. Core speech acts have a content which is either a state, an action description or a question about one of these.

Each of the states and actions in the task model is annotated with semantic information that can be used to describe and recognize description of those states in natural language (and our speech-act based agent communication language). For example, the action of the sergeant securing the assembly area (which can be accomplished by having the squad leaders each secure a quadrant) is represented as shown in (1). The resulting state of the assembly area being secure is represented as shown in (2).

(1) **agent** sgt
event secure
patient assembly-area
type act

(2) **object-id** assembly-area
attribute safety
value secure
polarity positive
type state

Speech recognition and natural language interpretation produces similar contents from spoken utterances. Dialogue processing then compares the NL representation to the relevant task model representations, and, if a sufficiently close match can be found with a task model state or action, that is seen as the referent.

The core speech acts that are currently modelled include **assert**, **info-request**, **order**, **request** and **suggest**. Unlike many accounts of the effects of these speech

acts (e.g. [Cohen and Perrault, 1979,Allen, 1983,Cohen and Levesque, 1990,FIPA, 1997]), there are no direct effects on the beliefs, desires or intentions of the conversational participants. This allows for the possibility that participants are insincere in their utterances. Following [Traum and Allen, 1994], the direct effects involve social commitments, and one may then infer from these commitments the beliefs or intentions commonly associated with these utterance types, given additional assumptions.

Assertions will have the effect of establishing a commitment by the speaker that the state holds, or that action happened, is happening, will happen, or should happen, depending on the tense and aspect of the utterance. **Info-requests** have a question as their contents. Questions are (possibly partial) propositions together with a designated *q-slot* indicating the part of the proposition asked about. For example, (3) shows an info-request by the LT to the Sgt with the content being a question about whether the assembly area is secure. Info-requests have as their effect an obligation to address the question. **Requests** have an action as content, and the effect is an obligation to address the request, e.g., to consider and give feedback on the request. **Orders**, which can only be performed by a superior to a subordinate in the social structure, have as their effect an obligation to perform the action that is its content. **Suggestions** do not impose obligations, but do focus the topic on the action.

(3) **action** info-req
actor lt
addressee sgt
type csa
content **q-slot** polarity
type question
prop **object-id** assembly-area
attribute safety
value secure
time present
type state

In addition to these *forward-looking* acts [Discourse Resource Initiative, 1997], there are also backward-looking acts, that point back toward previous dialogue acts or aspects of conversational structure. These will tend to relieve obligations e.g., by performing obliged actions or addressing other utterances. These include acceptances of requests (which will create an obligation to the requested act itself) as well as rejections and other moves that won't.

3.2 Dialogue Processing

Language processing occurs in two distinct and interleavable "cycles", one for understanding language and updating the information state, and a second for producing language. This separation of input and output processing cycles allows

the agent to have an arbitrary interleaving of contributions by itself and others rather than enforcing a rigid turn-alternation. Each communicative contribution is simultaneously interpreted at each layer, and may correspond to a number of acts at different layers. Generation usually starts from an intention to perform a main act, however any realized utterance will also correspond to a number of acts, some of which (e.g., turn-taking) may be as much a result of the timing of the performance with respect to other events as to the planned behavior.

4 Emotion Model

The computational model of emotion in our virtual humans is called EMA (**EM**otion and **A**daptation) [Gratch and Marsella, 2004]. Like many computational models of emotion, EMA is informed by by the group of psychological theories of emotion collectively termed appraisal theory [Lazarus, 1991]. Unlike most computational models, however, EMA provides a deep process model of the mechanisms underlying emotion, including the cognitive assessments that precede emotion, their consequences on cognition and behavior, and the way these consequences impact subsequent assessments. In terms of the antecedents of emotion, it broadens the goal-based reasoning underlying prior process models to cover other cognitive assessments implicated by appraisal theories. It is also the first process model of coping, a process associated with emotion in appraisal theories, and one which is implicated in emotional biases on cognition , but typically neglected by computational models. By modeling these deep processes explicitly, we are able to facilitate a tight integration of emotion with dialogue processing.

Appraisal theories argue that emotion arises from a person's assessment of their relationship with their environment, including not only their current condition but past events that led to this state as well as future prospects. Appraisal theory argues that people possess many distributed processes for interpreting this relationship (e.g., planning, explanation, perception, memory, linguistic processes) but that appraisal maps characteristics of these disparate processes into a common set of intermediate terms called appraisal variables. These variables characterize the significance of events from the individual's perspective. Events do not have significance in of themselves, but only by virtue of their interpretation in the context of an individual's beliefs, desires and intention, and past events. For example, the outcome of the latest presidential election might inspire joy, anger or indifference, depending on which candidate one desires and one's anger towards an individual may be mitigated by whether one believes they intended an offensive act. As in many appraisal models, events are characterized in terms of a number of appraisal variables:

- Perspective: from whose viewpoint is the event judged
- Desirability: what is the utility of the event if it comes to pass, from the perspective taken (e.g., does it causally advance or inhibit a state of some utility)

- Likelihood: how probable is the outcome of the event
- Causal attribution: who deserves credit or blame
- Temporal status: is this past, present, or future
- Controllability: can the outcome be altered by actions under control of the agent whose perspective is taken
- Changeability: can the outcome be altered by some other causal agent

EMA differs from many models by modeling the processes that derive these assessments, and in particular, how subsequent coping responses can bias these assessments. We treat appraisal as a set of feature detectors that characterize the current state of the agent's mental processes. For example, an event is interpreted through the lens of the agents current plans to assess its potential to impact current goals. Each appraised event is mapped into an emotion instance of some type and intensity, following the scheme proposed by Ortony et al [Ortony et al., 1988].

EMA supports multiple appraisals of the same event and can simultaneously appraise multiple events, forcing the issue of how the agent focuses on some aspect of its relationship with the environment. An activation-based focus of attention model computes a current emotional state based on most-recently accessed emotion instances. So, for example, the act of talking about an event can bring associated emotions into focus.

Coping determines how one responds to the appraised significance of events. Emotions can motivate people to act on the world (problem-focused coping) but they can also motivate us to re-interpret our circumstances, thereby influencing our prior beliefs, desires and expectations (emotion-focused coping). In EMA, coping strategies essentially work in the reverse direction of appraisal, identifying the precursors of emotion that should be maintained or altered (e.g., beliefs, desires, intentions, plans, expectations, etc.). Strategies include:

- Action: select an action for execution
- Planning: form an intention to perform some act (the planner uses intentions to drive its plan generation)
- Seek instrumental support: ask someone that is in control of an outcome for help
- Procrastination: wait for an external event to change the current circumstances
- Positive reinterpretation: increase utility of positive side-effect of an act with a negative outcome
- Acceptance: drop a threatened intention
- Denial: lower the probability of a pending undesirable outcome
- Mental disengagement: lower utility of desired state
- Shift blame: shift responsibility for an action toward some other agent
- Seek/suppress information: form a positive or negative intention to monitor some pending or unknown state

Strategies change behavior, but they also change the agent's interpretation of its circumstances, leading to re-appraisal. For example, simply intending to

perform an act can improve the agent's appraised sense of control and generate positive emotions. In terms of behavior, coping strategies provide the input to the behavioral, task and language processes that actually execute these directives. For example, plan related coping will generate an intention to perform some action that will make an undesirable situation better which in turn leads to the planning system to generate and execute a valid plan to accomplish this act. Alternatively, coping strategies might abandon the goal, lower the goal's importance, or reassess who is to blame. This close connection between appraisal, coping and cognition provides the processes "hooks" that facilitate the influences between emotion and dialogue.

5 Emotion and Dialogue: reinforcing each other

There are several aspects of dialogue interaction that have already begun to make use of the emotion model described in section 4. These include: reference resolution, initiative, content selection, and content realization.

Reference resolution and providing appropriate answers for general questions can be fairly difficult. For example, if a conversation is started with a question like "What happened here?", there can be many true answers indicating events that occurred in the present location. Some of these may be filtered based on common ground or assumed knowledge of the hearer, but otherwise it may be difficult to prefer one to another and provide natural answers without a lot of fairly involved commonsense reasoning. Using the "concerns" of the agent, calculated by the emotion reasoning, the agent can report on the one that is causing itself the strongest emotion. While this may not be ideal, in a Gricean cooperative sense, it does often mimic human behavior.

Likewise, as part of the coping mechanisms, 'the emotion module can signal to the dialogue manager that there is an important issue to discuss, even when not prompted by a specific question. In this case, the agent can take the initiative, when appropriate, and bring up the new topic.

Emotion reasoning is also used in the agent's language generation procedures. First, the concerns and coping strategies provide information that can be used to assess a suggestion or order, leading to choice of negotiation move [Traum et al., 2003b]. Next, emotion reasoning can signal which aspects of a situation to focus on. For example, considering a head-on crash between two vehicles, one could focus on the event itself, or each of the contributing sub-actions of the individual vehicles hitting the other. Finally, emotion reasoning is used to select the words that are used to describe the situation, trying to find a closest match in affect between the words used and the desired coping strategy. For example, as part of a coping strategy of shifting blame, the agent can inform the content realization to bias the way it phrases dialogue[Fleischman and Hovy, 2002]. For example, instead of neutrally describing a car accident, "We collided", a shift-blame strategy could bias the phrasing to be "They rammed into us."

The dialogue model also influences the emotion reasoning. In general, dialogue brings events into focus, which in turn brings into focus the associated

appraisals, altering emotional state and potentially invoking a coping response. Moreover, how the agent copes with a stressful depends on the dialogue context. Recall, EMA can appraise events from different perspectives. So, if a superior asks a subordinate agent about an event that not only brings into focus the event and the agent's appraisals of it but also how the subordinate agent feels the superior might feel about the event. Coping strategy selection takes this into account and therefore the subordinate agent is, for example, more likely to shift-blame in response to a question of "What happened here?" if the event in question is something that the superior is unhappy about.

6 Current and future integration plans

There are several further areas in which we are starting to integrate further the emotion and dialogue reasoning to lead to better affective dialogue. First, we have implemented a system for agents to talk about their emotional state, using the same dialogue model used for talking about tasks [Muller et al., 2004]. The basic mechanism of describing emotions can also be used for other coping strategies, intended to induce particular effects other than just expressing a true inner state. For example, an agent can evoke sympathy by appearing distressed, or can perhaps cast blame aside by acting angry at another. Also, the coping strategies can be used to influence the type of dialogue behavior. E.g., when initiative is taken, how much attention to pay to current focus vs possible other interpretations, etc. Likewise, more of the dialogue state can be used for emotional assessment. For instance, some agents may feel distress over unaccomplished obligations or cases in which their actual beliefs differ from their commitments.

Acknowledgements

The work described in this paper was supported by the Department of the Army under contract number DAAD 19-99-D-0046. Any opinions, findings and conclusions or recommendations expressed in this paper are those of the authors and do not necessarily reflect the views of the Department of the Army.

References

[Allen, 1983] Allen, J. F. (1983). Recognizing intentions from natural language utterances. In Brady, M. and Berwick, R. C., editors, *Computational Models of Discourse*. MIT Press.

[Allwood et al., 1992] Allwood, J., Nivre, J., and Ahlsen, E. (1992). On the semantics and pragmatics of linguistic feedback. *Journal of Semantics*, 9.

[Baker, 1994] Baker, M. (1994). A model for negotiation in teaching-learning dialogues. *Journal of Artificial Intelligence in Education*, 5(2):199–254.

[Clark, 1996] Clark, H. H. (1996). *Using Language*. Cambridge University Press, Cambridge, England.

[Cohen and Levesque, 1990] Cohen, P. R. and Levesque, H. J. (1990). Rational interaction as the basis for communication. In Cohen, P. R., Morgan, J., and Pollack, M. E., editors, *Intentions in Communication*. MIT Press.

[Cohen and Perrault, 1979] Cohen, P. R. and Perrault, C. R. (1979). Elements of a plan-based theory of speech acts. *Cognitive Science*, 3(3):177–212.

[Dillenbourg et al., 1996] Dillenbourg, P., Traum, D., and Schneider, D. (1996). Grounding in multi-modal task-oriented collaboration. In *Proceedings of the European Conference on AI in Education*.

[Discourse Resource Initiative, 1997] Discourse Resource Initiative (1997). Standards for dialogue coding in natural language processing. Report no. 167, Dagstuhl-Seminar.

[FIPA, 1997] FIPA (1997). Fipa 97 specification part 2: Agent communication language. working paper available at http://drogo.cselt.stet.it/fipa/spec/fipa97/f8a21.zip.

[Fleischman and Hovy, 2002] Fleischman, M. and Hovy, E. (2002). Emotional variation in speech-based natural language generation. In *Proceedings of The Second International Natural Language Generation Conference (INLG'02)*.

[Gratch and Marsella, 2001] Gratch, J. and Marsella, S. (2001). Tears and fears: Modeling emotions and emotional behaviors in synthetic agents. In *Proceedings of the Fifth International Conference on Autonomous Agents*.

[Gratch and Marsella, 2004] Gratch, J. and Marsella, S. (2004). A domain-independent framework for modeling emotion. *Journal of Cognitive Systems Research*. to appear.

[Larsson and Traum, 2000] Larsson, S. and Traum, D. (2000). Information state and dialogue management in the TRINDI dialogue move engine toolkit. *Natural Language Engineering*, 6:323–340. Special Issue on Spoken Language Dialogue System Engineering.

[Lazarus, 1991] Lazarus, R. (1991). *Emotion and Adaptation*. Oxford Press.

[Marsella and Gratch, 2002] Marsella, S. and Gratch, J. (2002). A step towards irrationality: using emotion to change belief. In *In proceedings of AAMAS 2002: First International Joint Conference on Autonomous Agents and Multi-Agent Systems*.

[Marsella and Gratch, 2003] Marsella, S. and Gratch, J. (2003). Modeling coping behavior in virtual humans: Don't worry, be happy. In *In proceedings of AAMAS 2003: Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*.

[Matheson et al., 2000] Matheson, C., Poesio, M., and Traum, D. (2000). Modelling grounding and discourse obligations using update rules. In *Proceedings of the First Conference of the North American Chapter of the Association for Computational Linguistics*.

[McAllester and Rosenblitt, 1991] McAllester, D. and Rosenblitt, D. (1991). Systematic nonlinear planning. In *Proceedings of the Ninth National Conference on Artificial Intelligence (AAAI-91)*, pages 634–639, Menlo Park, CA. AAAI Press.

[Muller et al., 2004] Muller, T. J., Hartholt, A., Marsella, S., Gratch, J., and Traum, D. (2004). Do you want to talk about it? a first step towards emotion integrated dialogue. unpublished manuscript.

[Novick, 1988] Novick, D. (1988). *Control of Mixed-Initiative Discourse Through Meta-Locutionary Acts: A Computational Model*. PhD thesis, University of Oregon. also available as U. Oregon Computer and Information Science Tech Report CIS-TR-88-18.

[Ortony et al., 1988] Ortony, A., Clore, G., and Collins, A. (1988). *The Cognitive Structure of Emotions*. Cambridge University Press.

[Rickel and Johnson, 1999] Rickel, J. and Johnson, W. L. (1999). Animated agents for procedural training in virtual reality: Perception, cognition, and motor control. *Applied Artificial Intelligence*, 13:343–382.

[Rickel and Johnson, 2002] Rickel, J. and Johnson, W. L. (2002). Extending virtual humans to support team training in virtual reality. In Lakemayer, G. and Nebel, B., editors, *Exploring Artificial Intelligence in the New Millenium*, pages 217–238. Morgan Kaufmann, San Francisco.

[Rickel et al., 2002] Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., and Swartout, W. (2002). Toward a new generation of virtual humans for interactive experiences. *IEEE Intelligent Systems*, 17.

[Sidner, 1994] Sidner, C. L. (1994). An artificial discourse language for collaborative negotiation. In *Proceedings of the Fourteenth National Conference of the American Association for Artificial Intelligence (AAAI-94)*, pages 814–819.

[Traum, 2003] Traum, D. (2003). Semantics and pragmatics of questions and answers for dialogue agents. In *proceedings of the International Workshop on Computational Semantics*, pages 380–394.

[Traum et al., 2003a] Traum, D., Fleischman, M., and Hovy, E. (2003a). Nl generation for virtual humans in a complex social environment. In *Working Notes AAAI Spring Symposium on Natural Language Generation in Spoken and Written Dialogue*.

[Traum et al., 2003b] Traum, D., Rickel, J., Marsella, S., and Gratch, J. (2003b). Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *In proceedings of AAMAS 2003: Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 441–448.

[Traum, 1994] Traum, D. R. (1994). *A Computational Theory of Grounding in Natural Language Conversation*. PhD thesis, Department of Computer Science, University of Rochester. Also available as TR 545, Department of Computer Science, University of Rochester.

[Traum and Allen, 1994] Traum, D. R. and Allen, J. F. (1994). Discourse obligations in dialogue processing. In *Proceedings of the 32^{nd} Annual Meeting of the Association for Computational Linguistics*, pages 1–8.

[Traum and Hinkelman, 1992] Traum, D. R. and Hinkelman, E. A. (1992). Conversation acts in task-oriented spoken dialogue. *Computational Intelligence*, 8(3):575–599. Special Issue on Non-literal language.

[Traum and Rickel, 2002] Traum, D. R. and Rickel, J. (2002). Embodied agents for multi-party dialogue in immersive virtual worlds. In *Proceedings of the first International Joint conference on Autonomous Agents and Multiagent systems*, pages 766–773.

[Walker and Whittaker, 1990] Walker, M. A. and Whittaker, S. (1990). Mixed initiative in dialogue: An investigation into discourse segmentation. In *Proceedings ACL-90*, pages 70–78.