A Model of Compliance and Emotion for Potentially Adversarial Dialogue Agents

Antonio Roque and **David Traum**
USC Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292
{roque,traum}@ict.usc.edu

Abstract

We present a model of compliance, for domains in which a dialogue agent may become adversarial. This model includes a set of emotions and a set of levels of compliance, and strategies for changing these.

1 Overview

We present an information-state based model of compliance for an agent who is questioned. The agent tracks several emotional and interpersonal variables, which can be updated depending on the dialogue act, content, and other features of utterances. A compliance level is computed based on the values of these variables. This work is in the tradition of research in building affective dialogue systems (André et al., 2004a) embodied as virtual humans (Rickel et al., 2002), with emotional components for training or tutoring purposes (Gratch and Marsella, 2005).

A model of emotion in an affective dialogue system may, among other things, influence that system's cognitive behavior (Becker et al., 2004), model the effects of social language (Cassell and Bickmore, 2003), or control behavior such as its level of politeness (André et al., 2004b). Our study is closer in spirit to (Traum et al., 2005), in which a virtual human decides on a negotiation strategy based on its emotional appraisal of the topic, of its negotiation options, and of the human speaker. Our study also overlaps somewhat in topic with (de Rosis et al., 2003), in which a computer decides whether or not to deceive.

In this work we build a model of compliance - how helpful the agent will be - in a domain in which the agent may become reticent or adversarial, along with the emotional components that direct that agent's decision.

2 Testbed Domain

Our testbed application is in the domain of *Tactical Questioning*, in which small-unit military personnel hold conversations with individuals to produce information of military value (Army, 2006). We are specifically interested in this domain when applied to civilians, when the process becomes more conversational and additional goals involve building rapport with the population and gathering general information about the area of operations.

We have developed an application for training individuals in conducting Tactical Questioning sessions with civilians. The scenario takes place in contemporary Iraq, where the trainee must talk to Hassan, a local government functionary. If the trainee convinces Hassan to help him, the trainee will confirm suspicions about an illegal tax being levied on a new marketplace; if exceptionally successful, the trainee may even discover that the tax has been placed by Hassan's employer. But if Hassan becomes adversarial, he may lie or become insulting.

Figure 1 shows the beginning of a typical dialogue with Hassan. Rather than working to determine what the human user wants and then providing it, in turns 6 and 8 Hassan provides replies that are off-topic or of low information value. The trainee's goal is to increase the value of Hassan's responses by appealing to Hassan's emotions and making him more compli-

Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue, pages 35–38,
Antwerp, September 2007.

ant. Section 4 describes how this can happen.

1	Trainee	Hello Hassan
2	Hassan	Hello
3	Trainee	How are you doing?
4	Hassan	Well, under the circumstances we are fine
5	Trainee	I'd like to talk about the marketplace
6	Hassan	I hope you do not expect me to tell you anything
7	Trainee	I just want to know why people aren't using the marketplace
8	Hassan	I don't feel like answering that question

Figure 1: Scenario Dialogue

3 System Implementation

As a training application, Hassan incorporates "human-in-the-loop" interactivity, and logs utterances, language features, and emotional states at every turn, with the aim of producing a summary for *after-action review*, at which time a human trainer and trainee may discuss the session. For this reason, Hassan may react realistically to a trainee's bribes or threats of force, even though such actions are against policy for Tactical Questioning of noncombatants (Army, 2006): these behaviors would be reviewed by a human trainer during or after the training session.

The natural language components of our dialogue agent include a set of statistical classifiers working together with a rule-based dialogue manager. The Automated Speech Recognition output is sent to the classifiers, three of which detect language features, and three of which suggest possible replies. The Dialogue Manager uses its model of emotions and compliance to determine which of the suggested replies, if any, are to be made back to the user, as described in the next section. Further system implementation details are given in (Traum et al., 2007).

4 Model of Dialogue, Emotions, and Compliance

In our training scenario, trainees have a specific set of information that they want to learn from Hassan. In the general Tactical Questioning domain, a questioner seeks **compliance**: that the interviewee at least answers any questions truthfully, and ideally that the interviewee takes the initiative in offering information. Note that this is different from *cooperation* as in (Allwood, 2001), as it does not make any assumptions about cognitive consideration, joint purpose, ethical consideration, or trust; compliant behavior might or might not be cooperative. The components of our model were developed based on a study of Tactical Questioning domain documents such as (Army, 2006) and (Paul, 2006).

More details about our model of compliance are given in section 4.3. The following sections describe how the human speaker's utterances indirectly update the agent's level of compliance by means of a model of emotion.

4.1 Dialogue Features

A human trainee's utterance is analyzed by statistical classifiers to detect its principal dialogue move, topic, and degree of politeness.

We define several dialogue moves relevant to the domain of tactical questioning. *Opening* moves are general greetings and introductions. *Complimentary* moves are those in which the trainee compliments or flatters the person being questioned. *General Conversation* includes talk meant to build a sense of social bonding between the agent and the trainee, as well as expressions of goodwill and off-topic statements. *Task Conversation* is talk related to information the trainee is interested in: in the case of this scenario, questions about the marketplace and taxation, about the agent and his business, and so on. *Threatening* moves are those that include a threat against the agent, and *Offering* moves offer to provide something. Finally, *Closing* dialogue moves are those that end the conversation.

The topic of the utterance will be a topic from one of three sets, or 'other'. The Information Request topics allow the agent to identify what the trainee is referring to in Task Conversation dialogue moves: the marketplace, taxation, and so on. The set of Threat-related and the set of Offer-related topics refer to the kinds of threats and offers that a trainee may make in the course of a conversation.

Finally, the third language feature to analyzed is the utterance's level of politeness. This will be identified as either polite, impolite, or neutral.

4.2 Emotional and Social variables

We identify four emotional and social variables (emotions, for short) applicable to the domain. They have been named to be intuitive to a trainer over-

seeing a session. *Respects Trainee* represents the degree of trust and respect the agent feels for the trainee. *Feels Respected* represents the extent to which the agent feels honored and respected. *Social Bonding* represents how much of a social relationship the agent feels for the trainee, and *Fear* represents how afraid the agent feels.

These emotions are represented as integer value components in an Information State dialogue manager (Traum and Larsson, 2003). They are updated by rules based on the state of the information state components and the language features identified in the trainee's utterance. For example, a Complimentary dialogue move would increase the agent's Feels Respected and Social Bonding values and decrease its Fear. A Threatening dialogue move would increase the agent's level of Fear but decrease its Feels Respected and Social Bonding values. A General Conversation dialogue move that was Polite would increase the Social Bonding value.

4.3 Compliance

For this study, we focused on the effect of compliance on the agent's verbal responses in terms of how much information the agent provides in response to the trainee's questions, whether the information is useful, to what extent the information is true, and whether the reply includes polite, neutral, or rude words.

Our model of compliance consists of three levels, which have the following effects.

At the *Compliant* level, the agent will answer the trainee's direct questions truthfully, and will try to provide useful information. The agent will be friendly and polite.

At the *Reticent* level, the agent will not provide any useful information. The agent may express that they do not wish to comply, may reply with off-topic remarks, or may make other low-information responses. The agent will generally be neither rude nor polite, but may be dismissive.

At the *Adversarial* level, the agent again will not provide any useful information, and may reply with off-topic or low-information responses. However, the agent may also be rude or insulting. Furthermore, the agent may reply deceptively: offering, in a neutral or polite way, high-information statements that are not true.

The agent's level of compliance may not be immediately apparent to the human speaker: for example, an agent replying in a neutral way with no information may be at the Reticent or Adversarial level, or it may be at the Compliant level and simply not have any useful information to provide. Similarly, answers with expected responses, such as greetings or farewells, may be answered the same at many compliance levels. Finally, if an agent is providing high-information responses, the human participant may not know if those are useful truths or plausible lies.

4.4 Compliance and Emotions

In the course of a dialogue, the agent's level of compliance may vary. After every utterance, the agent's emotions are checked to see if they change the agent's level of compliance. The goal of the trainee is to make the agent compliant by producing utterances that will update the agent's emotions in ways that will make the agent compliant. There are three basic strategies that the trainee can pursue, which are defined by the ways in which emotions affect compliance.

In the *Empathic* strategy, the trainee attempts to make the agent sympathetic to the trainee, and therefore to the trainee's goals. This is modeled by having the agent's compliance level become Compliant when the agent's Respects Trainee, Feels Respected, and Social Bonding scores all rise above a certain threshold. However, if those three emotions are below a given threshold, the agent's compliance level becomes Adversarial.

In the *Offering* strategy, the agent becomes compliant after the trainee makes an Offering dialogue move whose Topic is from the set of Offers that the agent is defined as being receptive to.

In the *Threatening* strategy, the trainee uses a Threat dialogue move to raise the agent's Fear above a certain threshold. If the trainee then makes a Threat that the agent is vulnerable to, the agent will become Compliant.

5 Future Directions

An evaluation of the entire system is described in (Traum et al., 2007). We hope to perform an evaluation of the compliance and emotion components

separately. One possibility is to do a semi-Wizard of Oz evaluation in which the ASR and language analysis tasks are performed by a human, to factor out errors in those components. Another possibility is to compare the system's performance in updating its information state with the performance of human coders in updating the information state, as was done in (Roque et al., 2006). Alternately, we could focus on how plausible the model of emotions and compliance is in terms of human processes by comparing it to data from human surveys, as was done in (Mao and Gratch, 2006).

The model of emotion and compliance that we have presented is motivated by the domain of Tactical Questioning, and the features and policies that we have implemented have been guided by that domain. As we continue to develop Hassan and other Tactical Questioning agents, we plan to add capabilities that will allow us to build more general and sophisticated models of emotion and compliance.

6 Acknowledgments

This work has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred. We would like to thank the other members of the TACQ team at ICT for work on the system in which these ideas are implemented and discussions regarding the compliance model, especially Anton Leuski, Susan Robinson, and Bilyana Martinovski. We would also like to thank the anonymous reviewers for their comments.

References

Jens Allwood. 2001. Cooperation and flexibility in multimodal communication. In Harry Bunt and Robbert-Jan Beun, editors, *Cooperative Multimodal Communication*, volume 2155 of *Lecture Notes in Computer Science*. Springer Verlag, Berlin/Heidelbert.

Elisabeth André, Laila Dybkjær, Wolfgang Minker, and Paul Heisterkamp, editors. 2004a. *Affective Dialogue Systems, Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004, Proceedings*, volume 3068 of *Lecture Notes in Computer Science*. Springer.

Elisabeth André, Matthias Rehm, Wolfgang Minker, and Dirk Bühler. 2004b. Endowing spoken language dialogue systems with emotional intelligence. In *Affective Dialogue Systems, Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004, Proceedings*, pages 178–187.

Department of the Army. 2006. Police intelligence operations. Technical Report FM 3-19.50. Appendix D: Tactical Questioning.

Christian Becker, Stefan Kopp, and Ipke Wachsmuth. 2004. Simulating the emotion dynamics of a multimodal conversational agent. In *Affective Dialogue Systems, Tutorial and Research Workshop, ADS 2004, Kloster Irsee, Germany, June 14-16, 2004, Proceedings*, pages 154–165.

Justine Cassell and Timothy Bickmore. 2003. Negotiated collusion: Modeling social language and its relationship effects in intelligent agents. *User Modeling and User-Adapted Interaction*, 13:89–132.

Fiorella de Rosis, Cristiano Castelfranchi, Valeria Carofiglio, and Giuseppe Grassano. 2003. Can computers deliberately deceive? a simulation tool and its application to turings imitation game. *Computational Intelligence*, 19(3).

Jonathan Gratch and Stacy Marsella. 2005. Some lessons for emotion psychology for the design of lifelike characters. *Journal of Applied Artificial Intelligence*, 19(3-4):215–233. Special issue on Educational Agents - Beyond Virtual Tutors.

Wenji Mao and Jonathan Gratch. 2006. Evaluating a computational mode of social causality and responsibility. In *5th International Joint Conference on Autonomous Agents and Multiagent Systems*, Hakodate, Japan.

Matthew C. Paul. 2006. Tactical questioning: human intelligence key to counterinsurgency campaigns. *Infantry Magazine*, Jan-Feb.

Jeff Rickel, Stacy Marsella, Jonathan Gratch, Randall Hill, David Traum, and Bill Swartout. 2002. Towards a new generation of virtual humans for interactive experiences. *IEEE Intelligent Systems*, pages 32–38, July/August.

Antonio Roque, Hua Ai, and David Traum. 2006. Evaluation of an information state-based dialogue manager. In *Brandial 2006: The 10th Workshop on the Semantics and Pragmatics of Dialogue*, University of Potsdam, Germany, September 11-13.

David Traum and Staffan Larsson. 2003. The information state approach to dialogue management. In R. Smith and J. van Kuppevelt, editors, *Current and New Directions in Discourse and Dialogue*, pages 325–353. Kluwer, Dordrecht.

David Traum, William Swartout, Stacy Marsella, and Jonathan Gratch. 2005. Fight, flight, or negotiate: Believable strategies for conversing under crisis. In *5th International Conference on Interactive Virtual Agents*. Kos, Greece.

David Traum, Antonio Roque, Anton Leuski, Panayiotis Georgiou, Jillian Gerten, Bilyana Martinovski, Shrikanth Narayanan, Susan Robinson, and Ashish Vaswani. 2007. Hassan: A virtual human for tactical questioning. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*.