Hassan: A Virtual Human for Tactical Questioning

David Traum and **Antonio Roque** and **Anton Leuski**
Panayiotis Georgiou and **Jillian Gerten** and **Bilyana Martinovski**
Shrikanth Narayanan and **Susan Robinson** and **Ashish Vaswani**
University of Southern California
Los Angeles, CA USA
`traum@ict.usc.edu`

Abstract

We present Hassan, a virtual human who engages in Tactical Questioning dialogues. We describe the tactical questioning domain, the motivation for this character, the specific architecture and present brief examples and an evaluation.

1 Introduction

Virtual Humans can be useful for tutoring or training in a variety of interactive situations in which experiential learning can be beneficial, such as in (Traum et al., 2005a) and (Rickel et al., 2002). Virtual humans contain a number of components, including a virtual body, usually embedded in a virtual world, actions that the agent can perform, including movements and sound, cognitive capabilities to decide on which actions to do and updating internal state, and perceptual abilities for recognizing the actions of users and other things in the world.

In this paper we present Hassan, a virtual human for training in Tactical Questioning dialogues. We focus on the spoken dialogue components. A companion paper (Roque and Traum, 2007) describes the dialogue manager and emotion model more fully.

Currently there is no single "best practice" model for building virtual humans or especially their spoken dialogue components. While generally there are separate modules for speech recognition, natural language understanding, dialogue management, and output (e.g., Generation and Synthesis, or text selection and audio clip playing), there is no consensus on the best ways of engineering these modules.

Part of the reason for this is that we are still fairly early in the search space, considering all of the possible techniques applied to the various domains that require spoken dialogue capability. Another issue is that there are several different goals for dialogue systems, and optimizing on one may lead to suboptimality for other goals. Some of these goals include: task success & efficiency, correct understanding & output, user satisfaction, believability/realism, authorability, reusability, revisability, and short development time.

Given the different relative importance of these goals and the specific features of the domain can lead to different choices for the spoken language technology components. For example, the virtual humans in (Rickel et al., 2002; Traum et al., 2005b) put a premium on depth of understanding within complex domains (teamwork, negotiation), but were somewhat narrow in the scope of what the virtual humans could talk about, and had a heavy authoring burden, requiring experts to create new domains. On the other hand, question-answering characters (Leuski et al., 2006) have a lower burden for depth, but must handle a broader range of questions and maintain believability and user satisfaction.

For our current endeavor, tactical questioning (see Section 2), we require capabilities between these two extremes. We need the authorability and general robustness of characters like SGT Blackwell (Leuski et al., 2006) but with more of the emotional and cognitive modeling of the situation from agents like Dr Perez (Traum et al., 2005b).

In this paper, we present Hassan, a Virtual Human for Tactical Questioning implemented using this in-

termediate architecture. In section 2, we describe the Tactical Questioning Domain and the Hassan scenario. In section 3, we describe the components of the system. In section 4, we describe the preliminary evaluation, and we conclude with future directions in Section 5.

2 Domain: Tactical Questioning

Tactical Questioning dialogues are those in which small-unit military personnel, usually on patrol, 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. Hassan is a virtual human designed to act as a roleplayer and allow trainees to practice tactical questioning and get feedback from experienced instructors on their performance on several learning goals.

The scenario for Hassan takes place in contemporary Iraq. In a fictional storyline, the US authorities have built a marketplace as part of the reconstruction effort, but the local population continues to use the old, broken-down marketplace instead. It is the goal of the trainee to discover why. To do this, the trainee talks to Hassan, a local politician. If the trainee convinces Hassan to help him, the trainee will confirm that a tax has been levied on the new marketplace, and that the tax has been placed by Hassan's employer; if exceptionally successful, the trainee may even learn where that employer lives. If Hassan becomes adversarial, he may lie and tell the trainee that an American soldier is collecting the tax. Figure 1 shows the beginning of a typical dialogue with Hassan.

3 Virtual Human Implementation

Figure 2 shows several components of Hassan during a session. The virtual environment includes the embodied character, which is the only component the trainee usually sees. Above that is a speech capture component showing the Automated Speech Recognition (ASR) results of an utterance. Also visible is a GUI showing the state of various of Hassan's emotional components. Although the system

| Trainee	Hello Hassan
Hassan	Hello
Trainee	How are you doing?
Hassan	Well, under the circumstances we are fine
Trainee	I'd like to talk about the marketplace
Hassan	I hope you do not expect me to tell you anything
Trainee	I just want to know why people aren't using the marketplace
Hassan	I don't feel like answering that question
Trainee	I think you know something about a tax
Hassan	I am simply doing business. It is rude of you to imply otherwise

Figure 1: Scenario Dialogue
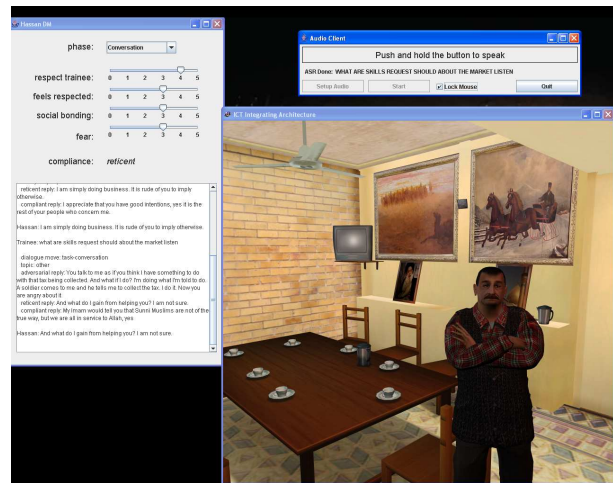
Figure 2: Hassan, a Virtual Human for Tactical Questioning, with some other components

can run autonomously, its emotional state can also be modified at run-time by an instructor. The virtual environment is set in the Unreal Tournament game engine, similar to the agent in (Traum et al., 2005b). It also uses the *Smartbody* character controller (Thiebaux et al., 2007) to control the movements of the character, including lipsync and nonverbal communicative behaviors, and the Nonverbal Behavior generator (Lee and Marsella, 2006) to select and synchronize non-verbal behaviors with the output text.

The language components include a speech recognizer, a set of statistical classifiers to recognize dialogue features and suggest responses, and a dialogue manager, to maintain a current cognitive and emotional model and chose the appropriate response. Our initial version of Hassan used the same architecture as SGT Blackwell, with a single clas-

sifier to pick the answer, and rudimentary dialogue manager to avoid repetition where possible and be able to answer further on the same topic. Our initial tests showed that this was inadequate for the tactical questioning domain, where one needs not just local coherence between questions and answers, but also an emotional progression of the character in which the kinds of questions and behavior early on in the conversation will effect the kinds of answers given later on. E.g., a trainee can increase or reduce fear.

In order to address this issue, we added a more sophisticated information-state based dialogue manager which can track several states that are important to deciding how compliant an agent should be. We also introduced a number of statistical classifiers (built using our NPCEditor software) to pick out important dialogue features as well as the best answer given a particular compliance level. Figure 3, shows the natural language components of our dialogue agent, including a set of NPCEditors working together with a rule-based Dialogue Manager. We discuss each of these components briefly below.
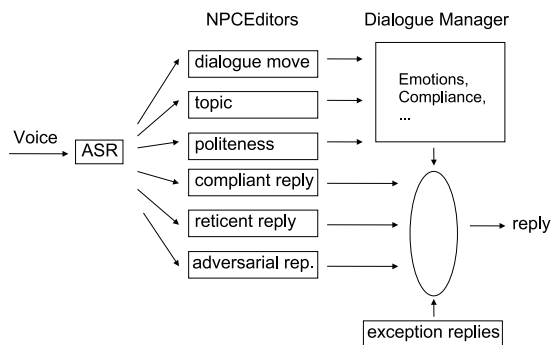

Figure 3: Architecture of Language Components

3.1 Automated Speech Recognition

The trainee talks to Hassan using a headset microphone and a push-to-talk button. The ASR component uses the Sonic statistical speech recognition engine (Pellom, 2001), with custom acoustic and language models (Sethy et al., 2005).

3.2 NPCEditor: Statistical Classification

Our NPCEditor tool allows one to build statistical classifiers for "non-player characters". It allows several output modes including email, chat, and several interprocess communication protocols. The classification can be between input and output text (e.g., the answer to a question), or between input text and output features (NLU) or input features and output text (NLG). It has been used in a variety of ways in our Virtual Human agents. The NPCEditor allows inputting and annotation of training data, training a classifier, and run-time performance all within the same software platform. The classification techniques and their use to select answers is described in (Leuski et al., 2006).

3.3 Dialogue Features

The NPCEditor statistical classifiers identify three utterance features of the user utterance: a dialogue move, a main topic and a level of politeness. The set of dialogue moves for the Tactical Questioning Domain are shown in Figure 4. The main topic is an aspect of significance for the domain and character. There are different topics for requests (e.g. marketplace, taxation), threats (e.g. loss of status) and offers (e.g. security, recognition, or secrecy). Politeness is one of *polite, neutral* or *impolite*. These three features work together to inform the decisions made by the dialogue manager.

Opening	greetings, introductions, ...
Complimentary	compliments, flattery, ...
General Conversation	non-task-related talk
Task Conversation	task-related talk
Threatening	threats
Offering	offers to provide something
Closing	moving to end the conversation

Figure 4: Dialogue Moves

3.4 Dialogue Manager

The dialogue manager of the system is based on the information-state approach (Traum and Larsson, 2003). It tracks a set of four information state variables relating to respect, bonding and fear, and calculates from these a current *compliance level* for the character. The utterance features from the classifiers are used to update these variables, which may result in a change in compliance level. A response is selected by choosing the response given by the classifier for that compliance level (or an exception reply for special circumstances). More about the dialogue manager and compliance computation can be found in (Roque and Traum, 2007).

4 Evaluation

A preliminary evaluation of the first version of this agent was held to produce data for analysis and to measure user satisfaction. Eight sessions were held with an equal combination of college-level military trainees, and information professionals in our research facility. Post-questionnaires allowed the trainees the opportunity to rate their experience.

Preliminary results indicate the users felt the system was off-topic too often to adequately judge the effects of the emotional components. In reply to ranking from 1 to 7 how satisfied they were with their questioning of the agent, the mean value given was 3.4. In reply to ranking from 1 to 7 how they rated Hassan as an interviewee, the mean value was also 3.4. A partial review of the logs indicates that these low scores may have been due to discrepancies in the reply authoring, which did not properly handle the generation of off-topic replies when confidence in an on-topic reply was low.

5 Future Work

While the current version of Hassan, with several information state variables, dialogue features, and 3 compliance levels is definitely an improvement in consistency over the previous version with one NPCEditor and no emotion-based information state, there is still much room for improvement. We are currently investigating techniques to track longer segments than the question-answer pair, as well as more sophisticated discourse processing on both the NLU and NLG side, while keeping the authoring relatively simple.

Acknowledgments

References

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

Jina Lee and Stacy Marsella. 2006. Nonverbal behavior generator for embodied conversational agents. In Jonathan Gratch, Michael Young, Ruth Aylett, Daniel Ballin, and Patrick Olivier, editors, *IVA*, volume 4133 of *Lecture Notes in Computer Science*, pages 243–255. Springer.

Anton Leuski, Ronakkumar Patel, David Traum, and Brandon Kennedy. 2006. Building effective question answering characters. In *Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, Sydney, Australia, July.

Bryan Pellom. 2001. Sonic: The university of colorado continuous speech recognizer. Technical Report TR-CSLR-2001-01, University of Colorado.

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 and David Traum. 2007. A model of compliance and emotion for potentially adversarial dialogue agents. In *Proceedings of the 8th SIGdial Workshop on Discourse and Dialogue*, september. this volume.

Abhinav Sethy, Panayiotis Georgiou, and Shrikanth Narayanan. 2005. Building topic specific language models from web-data using competitive models. In *Proceedings of Eurospeech*, Lisbon, Portugal.

Marcus Thiebaux, Andrew N. Marshall, Stacy Marsella, Edward Fast, Aaron Hill, Marcelo Kallmann, Patrick Kenny, and Jina Lee. 2007. Smartbody: Behavior realization for embodied conversational agents. In *7th International Conference on Intelligent Virtual Agents (IVA)*.

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, Jonathan Gratch, Stacy Marsella, Patrick Kenny, Eduard Hovy, Shri Narayanan, Ed Fast, Bilyana Martinovski, Rahul Baghat, Susan Robinson, Andrew Marshall, Dagen Wang, Sudeep Gandhe, and Anton Leuski. 2005a. Dealing with doctors: A virtual human for non-team interaction. In *Proceedings of the 6th SIGdial Workshop on Discourse and Dialogue*, September 2-3.

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