SemDial 2011: Proceedings of the 15th Workshop on the Semantics and Pragmatics of Dialogue, pages 63–72.
Los Angeles, California, 21–23 September 2011.

Toward Rapid Development of Multi-Party Virtual Human Negotiation Scenarios

Brian Plüss
Centre for Research in Computing
The Open University
Milton Keynes, UK
b.pluss@open.ac.uk

David DeVault and **David Traum**
Institute for Creative Technologies
University of Southern California
12015 Waterfront Drive, Playa Vista, CA 90094
{devault,traum}@ict.usc.edu

Abstract

This paper reports on an ongoing effort to enable the rapid development of multi-party virtual human negotiation scenarios. We present a case study in which a new scenario supporting negotiation between two human role players and two virtual humans was developed over a period of 12 weeks. We discuss the methodology and development process that were employed, from storyline design through role play and iterative development of the virtual humans' semantic and task representations and natural language processing capabilities. We analyze the effort, expertise, and time required for each development step, and discuss opportunities to further streamline the development process.

1 Introduction

This paper reports on an ongoing effort to enable the rapid development of multi-party virtual human negotiation scenarios. This work is part of a research effort that has been underway at USC's Institute for Creative Technologies for a number of years, which has been developing methodologies and tools that can support the rapid development of virtual human dialogue systems. Virtual humans (Rickel and Johnson, 1999; Swartout et al., 2006) are implemented virtual characters that are designed to participate in face to face natural language dialogue interactions with human users.

The methodologies and tools that have been developed have been tailored to support several types (or genres) of interaction with virtual humans. The genres that have been explored range in complexity from straightforward question-answering characters (Leuski et al., 2006; Leuski and Traum, 2010) to more strategic tactical questioning systems (Traum et al., 2007; Gandhe et al., 2009) and full negotiation scenarios (Traum et al., 2003; Hartholt et al., 2008; Traum et al., 2008).

In many ways, negotiation scenarios are the most complex genre of virtual human dialogue interaction that has been implemented to date. These scenarios are designed to allow a trainee to practice their negotiation skills by engaging in face-to-face negotiation with one or more virtual humans. To understand and respond to user utterances – such as assertions, proposals, and offers – virtual humans make use of natural language processing capabilities including automatic speech recognition (ASR), natural language understanding (NLU), and natural language generation (NLG). To reason about their negotiation, they draw on formal ontologies and task models for their negotiation domain, multi-party negotiation strategies that range from team-based to adversarial negotiation and incorporate factors like trust and emotions, and an ability to simultaneously discuss multiple potential courses of action (Hartholt et al., 2008; Traum et al., 2008). Previous negotiation scenarios have included a Mission Rehearsal Exercise in which a lieutenant talks with a virtual platoon sergeant about how to respond to a car accident (Traum et al., 2003; Swartout et al., 2006), and a negotiation with either one (Traum et al., 2005; Core et al., 2006) or two (Hartholt et al., 2008; Traum et al., 2008) virtual humans to find a way to relocate a virtual doctor's medical clinic out of an unsafe market area.

One of the goals of this ongoing research is the development of methodologies, authoring tools, and natural language processing techniques that en-

able new negotiation scenarios to be developed more rapidly, yielding new possibilities for more widespread practice and training of negotiation skills. However, rapid development has been limited by the variety and complexity of the knowledge and resources that are required to build these systems, and to date, developing a new negotiation scenario has typically required months of effort by a team of researchers and developers.

In this paper, we assess our progress in streamlining and simplifying this effort using a new case study in which a four-party negotiation scenario was designed and implemented to a prototype stage by a single researcher who had no previous experience with this technology. We present the design and authoring process that was used, starting from role play dialogues, proceeding through various development steps, and concluding in the production of an implemented prototype over a span of 12 weeks. We quantify the development effort that was needed, and conclude with a discussion of the remaining challenges and opportunities in enabling the rapid development of new virtual human negotiation scenarios.

2 Case Study Negotiation Scenario

We developed the storyline for our target scenario through an iterative design process involving brainstorming, discussion of technical details, and role play sessions. We provide here a high-level description of the resulting scenario:

> An American Old West town has been freed from a dangerous outlaw, defeated by a U.S. Ranger with the help of Utah, the local bartender. The Ranger and his Deputy must now leave town to pursue their mission elsewhere. But before leaving, they need to recruit a town sheriff, so they offer the job to Utah. He will need resources – e.g., money to buy guns and to hire men – guaranteed before considering the offer. As owner of the saloon, Harmony is an influential woman in town. She will be present in the discussions, pushing forward her own agenda of demands, part of which she cannot discuss in front of Utah and must be dealt with in private by one of the officers. The Ranger and the Deputy have very limited resources, so they must negotiate to reach an agreement by committing as little as possible.

Figure 1: Utah and Harmony

In the implemented scenario, the roles of Utah and Harmony are always played by virtual humans, which we picture in Figure 1. The art assets needed to depict these characters were borrowed from the existing Gunslinger system (Hartholt et al., 2009). The roles of the Ranger and Deputy are to be played by human negotiation trainees.

The storyline was designed to be somewhat more complex than previous implemented negotiation scenarios (Traum et al., 2003; Swartout et al., 2006; Traum et al., 2008). The new complexities included: the presence of two simultaneous human participants, and the possibility of a 4-party dialogue splitting into simultaneous 2-party dialogues; a greater number of possible solutions to the negotiation problem; and the presence of a hidden agenda in a virtual human, necessitating a private discussion away from the other virtual human. However, most of the development infrastructure from the previous SASO-EN scenario (Traum et al., 2008) was reused, as we detail in Section 3.

As we developed the details of the storyline, we held several human role play sessions. These sessions were video recorded and transcribed, both for analysis and also to serve as a source of linguistic data. We provide an excerpt from a role play in Figure 2 and describe more specific aspects of these sessions in Section 3.1.

Role plays were crucial to our development process, as they identified several gaps and implausible elements in early versions of the storyline, and also provided several creative elements that ended up serving as natural storyline extensions.

Further, role play sessions are valuable as a source of concrete examples of utterances and subdialogues, which translate into demands on the virtual humans' natural language processing capabilities, and can be used to anticipate technical chal-

Ranger	Um we can help you. What do you need?
Utah	We need some additional resources uh you know and the sheriff's not ... a sheriff can't keep the town safe alone. You need a good set of deputies, like you have your deputy here and uh
Ranger	You can hire deputies.
Utah	Well with the you know it takes some some uh some money to to do that uh um and so if you if you have enough money to to help us out I think we can probably reach some kind of arrangement.
Ranger	Um we can help you. Yeah. We can we we we will be able to help you my get some money for the that to support your deputies.
	(. . .)
Utah	Well it sounds very good.
Ranger	So I think we've got a deal here.
Harmony	No no there's no deal. There's no deal here. This this isn't this isn't right. This works for you guys but this doesn't this doesn't work for uh Utah.

Figure 2: Dialogue excerpt from one of the role plays

lenges. For example, the complexity of the speaker turns in Figure 2 suggests several implementation challenges, including the presence of multiple utterances and speech acts in a turn, complex rhetorical structure, and numerous speech disfluencies. We discuss how we addressed these challenges using a dialogue simplification procedure in Section 3.2.

3 System Development

In this section, we describe the development and implementation of our case study scenario. Authoring negotiation scenarios involves a heterogeneous set of technologies, which need to be developed in a coordinated manner, while keeping a storyline that allows for believable and engaging interaction.

Conceptually, we can characterize the overall development process as involving five main tasks. The first task is creating the *storyline*, i.e., deciding on the creative elements of the scenario. The storyline was described in Section 2.

The second task is defining a *task model*, i.e., formalizing states, actions, beliefs, goals, and plans to formally model the storyline. The task model is a core representation that both constrains and motivates the behaviour of the virtual characters, and is necessary for them to reason about the negotiation, as well as generate and understand language related to the negotiation.

The third task is creating the *language resources*, i.e., collecting the linguistic data, crafting the semantic representations, and building the models that

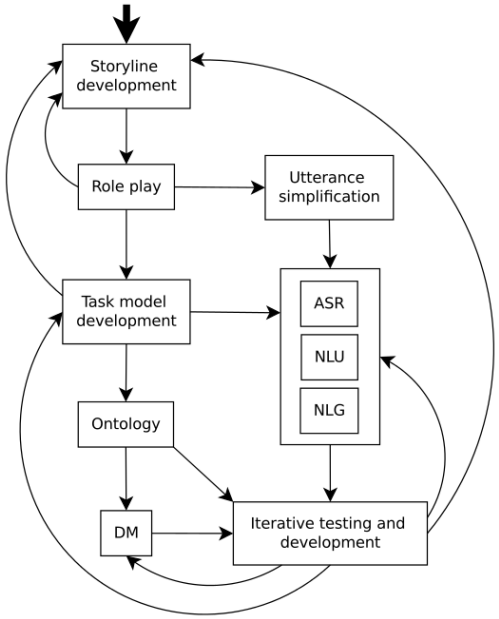

Figure 3: Development steps workflow

allow the system to communicate using natural language following the task model. This includes: an automatic speech recognition (ASR) model, semantic frames for natural language understanding (NLU) and natural language generation (NLG), and lexical elements for the dialogue manager (DM).

The fourth task is *implementing* the scenario, i.e., creating the instances of the virtual humans, with the linguistic resources and the task model in place. An ontology is used to centralize knowledge representation (Hartholt et al., 2008).

The fifth task is *iterative testing and development*, i.e., interacting with the virtual humans, assessing their behaviour, and extending coverage of the target scenario.

In practice, these tasks are not carried out as a sequence, but are highly interdependent, which necessitates a spiral design process, as shown in Figure 3. The level of formalization required to produce a task model, for instance, can help shape details of the storyline. The data collection for generating some of the language resources can expose gaps in the task model or point out unrealistic assumptions about an expected dialogue interaction. Iterative testing and development will inform all the other tasks, prompting revisions – e.g., fine-tuning the task model, extending NLG capabilities, improving NLU coverage

Phase	Description	Time
1	Storyline and task model design	4 weeks
2	Skeletal scenario implementation	3 weeks
3	Iterative testing and development	5 weeks

Table 1: Scenario development time (main phases)

of user utterances – and motivating further implementation, testing, and debugging.

To cope with this interdependence, we took an incremental approach for implementing our case study scenario. We began with an initial phase of story development and task model design. In a second phase, we started with a small skeletal subset of the elements in the scenario and implemented this subset as a running system. At the completion of this second phase, we were able to perform interactive testing with the virtual humans. We then proceeded in a third phase to iteratively extend the virtual humans' capabilities through testing, development, and debugging. In this phase, we iteratively added small increments to the functionality, such as extensions to the natural language resources or fine-tuning of the task model.

In Table 1, we summarize the development time that was required for each of the three phases in our case study scenario. As stated above, this development effort was carried out by one researcher who had no previous experience with the virtual human technology.[1] In presenting and analyzing this development process, our aim is to assess the strengths and weaknesses of the process we used to construct these dialogue systems, and to identify opportunities to streamline future development. In the rest of this section, we continue by discussing and analyzing the activities in each phase in more detail. These detailed activities and their associated development times are summarized in Table 2.

3.1 Storyline and task model design

This first phase of the scenario development took 4 weeks in total, starting from discussions and brainstorming of preliminary ideas, and leading to a well rounded, realistic, believable, and feasible storyline along with a design for a supporting task model.

[1]More specifically, this researcher was a PhD student with a background in computational linguistics and theoretical dialogue modeling, but who had no previous practical experience in dialogue system development.

Phase	Development Step	Time
1	Initial storyline development	2 weeks
	Role play	1 week
	Task model design	1 week
2	Simplified task model development	1 week
	Utterance simplification	1 day
	ASR	1 day
	NLU	1 day
	NLG	2 days
	Ontology & DM editing	1 week
3	Running tests	1 week
	Component interaction diagnosis	2 days
	Task model fine-tuning	2 weeks
	ASR/NLU/NLG extension	2 days
	Ontology & DM fine-tuning	2 days
	Debugging	2 days
	Consulting & collaboration	2 days

Table 2: Scenario development time (detail)

Figure 4: Role play session

Initial storyline development. We began with initial storyline development, which consumed 2 weeks. This step involved deciding on the creative elements of the scenario, such as time, place, characters, underlying story, current conflict, individual and shared goals, available resources for negotiation, possible outcomes, etc.

We started with brainstorming sessions. To assess the feasibility of some of the creative ideas, we consulted with several researchers who have extensive knowledge of the technical capabilities and limitations of the system modules. These consultations were followed by creative writing and more discussion. We went through two versions of the storyline, to make it both realistic and feasible, before we proceeded to the role play. The resulting storyline was presented in Section 2.

Role play. Throughout a week, we held three role play sessions. A role play session is pictured in Figure 4, and a role play excerpt is provided in Figure 2.

The role play served two purposes. First, to test

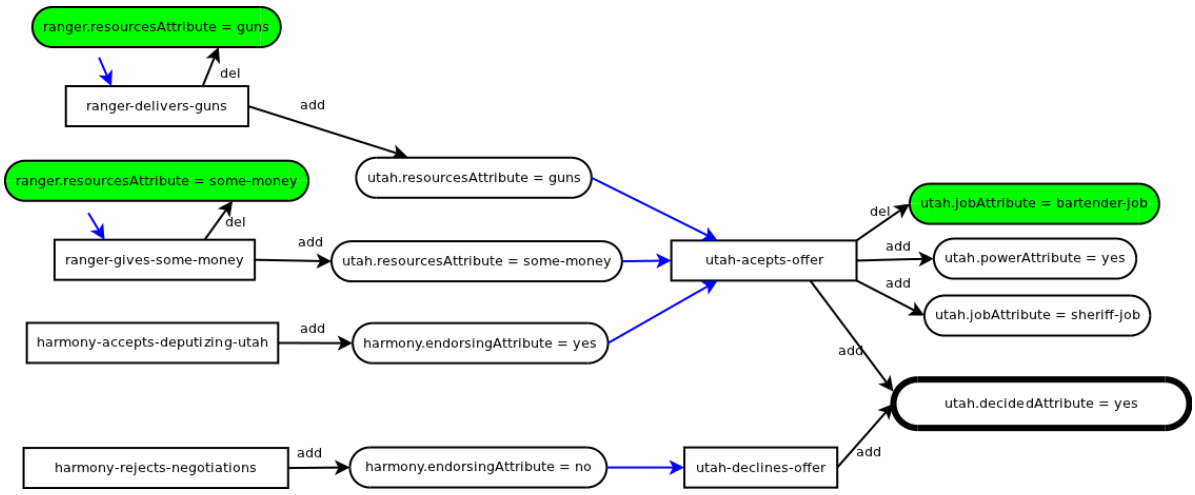

Figure 5: Simplified task model for Utah

the initial storyline and provide insight on the creative elements: likely (and unlikely) topics of discussion in the dialogues, credibility and naturalness of the overall story, etc. Second, to provide initial data for the natural language resources needed to implement the system. All sessions were video recorded and transcribed, and transcribed utterances were subsequently used to train ASR and NLU models and build an NLG corpus, as described below.

Before each session, role players received written instructions for the role play. The instructions included general details about the scenario (such as time, place, characters, and underlying story) as well as specific details about their role (such as available resources, individual goals, and desired outcomes). After each role play, the participants held discussions and brainstorming for improvements to the storyline.

Except for one player who took part in two sessions, we used a different set of players for each session. About two thirds of the role players were experienced researchers familiar with the system's underlying technology and had some basic knowledge of the storyline. However, we preferred naive participants for the human player roles. In latter sessions, once the characters were better defined, the players in the roles of the virtual humans were given increasingly more detailed instructions. This modality, somewhat closer to Wizard-of-Oz simulation, improved the suitability of the resulting natural language data and also narrowed the possible courses of interaction, bringing them closer to the task model being designed.

Task model design. The task model is a core representation used by the virtual humans to participate in their negotiation (Hartholt et al., 2008). In concert with storyline design, we therefore designed a provisional task model as a way of confirming the technical feasibility of the storyline. The task model formalizes the storyline using a collection of world states, actions, tasks (courses of action), and goals. For example, Utah's task model formalizes his option to become the town sheriff as a possible course of action, which relates to various other states and goals in a STRIPS-like formalism (Hartholt et al., 2008). The task model is used by a task planner to represent and decide between courses of actions for achieving desired goals, based on the current state of the world, possible actions (with preconditions and effects), and associated utilities (Traum et al., 2003).

Currently, the task model is authored partially using a Protégé GUI (Knublauch et al., 2004) and partially in TCL code. The designed task model for each of the virtual humans in the new scenario had two alternative courses of action, with approximately 25 world states and 20 actions. Designing these models required substantial collaboration with the designers of previous task models, and consumed 1 week of development effort. A visualization of a subsequent, simplified task model for this scenario is provided in Figure 5.

Utah	We need some additional resources uh you know and the sheriff's not a sheriff can't keep the town safe alone. You need a good set of deputies, like you have your deputy here and uh	**Utah** [16]	We need a good set of deputies.
Ranger	You can hire deputies.	**Ranger** [17]	You can hire deputies.
Utah	Well with the you know it takes some some uh some money to to do that uh um and so if you if you have enough money to to help us out I think we can probably reach some kind of arrangement.	**Utah** [18]	It takes some money to do that.
		[19]	If you have enough money to help us out we can probably reach an arrangement.

Figure 6: Example of a step in the utterance simplification process

3.2 Skeletal scenario implementation

Once the first phase was completed, we had a full storyline and the design for a full task model. We then proceeded in a second phase to implement a runnable skeletal version of the scenario. In this section, we describe this 3-week implementation effort.

Simplified task model development. We began by selecting a core subset of the storyline and task model to implement first. This involved eliminating certain aspects of the storyline. In particular, we removed some hidden agenda details that we had explored in the role plays, but which would have necessitated splitting the 4-party dialogue into two simultaneous 2-party discussions. In the simplified storyline, Utah will only accept the Ranger's offer if he is given guns and money to hire deputies, and if Harmony supports his designation. The Ranger has money and guns, so he can satisfy Utah's demands. In order to endorse Utah's designation as sheriff, Harmony needs a promise from the Deputy that they will keep protecting the town for some time, in support of Utah; otherwise, she would block the negotiations. The whole conversation plays out with all 4 parties present.

The simplified task model for Utah is shown in Figure 5. It still has two courses of action, corresponding to Utah accepting or declining the offer, but only 10 states, and 6 actions. Three of the states (shown as colored ovals in the diagram) are true when the interaction starts. The other 7 are false (shown as white ovals). Four of the actions (shown as white boxes) are enabled. This means they have no preconditions (states connected to the action with an unlabeled blue arrow) or that all their preconditions are true. If performed, enabled actions can have two possible effects (black arrows connecting an action with a state): making a state true (arrow labeled with *add*) or making it false (arrow labeled with *del*). One of the states in the task model is a goal (shown with a thick border).

Implementing the initial version of the simplified task model took one week of work. This required creating all the world states, actions and tasks in the ontology, and generating the TCL code that is used by the virtual human's task reasoner.[2]

Utterance simplification. To overcome the technical challenges discussed in Section 2, in about one day of work, we simplified the utterances in the role play transcripts. This step-wise process, also referred to as *dialogue distillation* (Jönsson and Dahlbäck, 2000), consisted of: segmenting each turn into single speech act utterances; selecting those relevant to the pragmatics of negotiation dialogues; and re-writing these into progressively simpler forms – e.g., by removing speech disfluencies and simplifying rhetorical structures – while preserving as much as possible of the semantic and pragmatic meaning. In this way, the overall flow of the conversation remains close to the original, but the utterances become suitable sources of data for the tools supporting the development of the virtual human's natural language resources. We give an example of this process in Figure 6, using a fragment of the role play dialogue presented in Section 2.

ASR. We defined a set of user utterances that we anticipated might appear in a typical dialogue, and used this to train an N-gram language model for automatic speech recognition (ASR). Using the results from the utterance simplification stage, defining this

[2] Apart from the information in the diagram, implemented task models have notions of authority associated to actions and utilities associated to states, which are used by the task reasoner and by the emotions module to guide the virtual human's negotiation behaviour.

corpus and training the language model was a quick task, taking about a day to complete.

NLU. The virtual human's NLU module converts text utterances into meaning representations (called *frames*) used for calculating the semantic and pragmatic effects of communication (Traum, 2003). The NLU consists of two parts, a context-independent part, classifying word sequences into initial meaning representations, and a context-dependent part, that uses the agent's information state to do reference resolution and compute speech and dialogue acts. The semantic components are derived from the task model representations, using Protégé. The context-independent NLU uses a framebank (pairings of word sequences to frames) to train a classifier that can recognize frames for novel utterances. It took 1 day to prepare the NLU framebank and train the NLU module (Sagae et al., 2009). Creating the framebank required us to pair the simplified utterances from the role play dialogues to their corresponding semantic representations.

NLG. The NLG module (DeVault et al., 2008) uses a similar semantic frame representation to that used by the NLU, the difference being that the frames contain more context-dependent and pragmatic information than the NLU frames. The NLG module converts semantic frames chosen by the DM into text. To support this translation, the NLG needs a training corpus of examples – the NLG framebank – linking frames to their natural language realizations. We crafted a corpus of semantic frames and simplified example utterances for the NLG model in two days of work. This process was somewhat slower than for the NLU framebank, mostly because the set of possible NLG frames produced by the DM was somewhat large at this stage of development.

Ontology and DM Editing. Most of the knowledge representation for these virtual humans, such as the elements in the task model and the components of the semantic frames, is centralized in an ontology (Hartholt et al., 2008), which can be edited by using custom extensions of Stanford's Protégé GUI (Knublauch et al., 2004). To extend the ontology requires interactive editing using this GUI. Additionally, to enable the DM to participate in a new scenario, at the time this effort was carried out, it was necessary to create a separate lexicon of domain-specific concepts.[3] The lexicon connects elements in the task model, such as people, places, objects and actions, with their counterparts in the semantic frame representation. Editing the ontology to include all domain-specific concepts, and creating the lexicon for the DM, took one week.

3.3 Iterative testing and development

Once the second phase was complete, the virtual humans could be run interactively for testing and further development. In this section, we describe the third phase, in which we used an iterative testing and development cycle to extend the system's capabilities over a period of 5 weeks.

Running tests. We spent a total of about a week in running tests of the system. These tests were spread over many small iterations of development. Each test run could take anything from a few seconds to several minutes, depending on the occurrence of errors or of unexpected behaviour.

Component interaction diagnosis. With the exception of ASR, all the modules were tested together and by interacting with the virtual humans. This required diagnosing problems in the interaction between the system components. For example, whether a semantic representation given by the NLU matched the one in the DM's lexicon, whether the entries in the lexicon were consistent with the elements in the task model, etc. This diagnosis, also spread over several iterations, took about 2 days.

Task model fine-tuning. Fine-tuning the task model to make the negotiation work as desired was the biggest task within this phase, taking about 2 weeks of work. It involved, for instance, improving the way actions depended on and affected states through preconditions and effects; adding or removing elements (such as attributes and possible values) to the set of world states; and adjusting the utility values a character associates with certain world states becoming true or false (which affects the virtual humans' negotiation decisions).

ASR/NLU/NLG extension. Around 2 days were used in extending the initial ASR, NLU and NLG re-

[3] This step has since been automated.

sources, as the possible interactions became longer and more complex. This included, for instance, authoring desired character utterances for new NLG frames, and extending the ASR and NLU resources to improve performance and coverage.

Ontology & DM fine-tuning. Similarly, a total of an extra 2 days were spent making small changes to the ontology and to the lexicon in the DM.

Debugging. Code debugging took about 2 days.

Consulting & collaboration. Finally, as the prototype grew larger, it was necessary to consult with experts on the different system modules. This took another 2 days in total, and involved discussions with researchers familiar with the emotions model, the task planner and the rules that implemented the functionality of the DM, among others.

4 Implemented Prototype

Through the steps discussed in Section 3, we arrived at an implemented prototype. The prototype uses the simplified task model shown in Figure 5. We show an example dialogue illustrating the capabilities of the prototype system in Figure 7. Some perspective on the implemented system can be achieved by contrasting this dialogue with the human-human role play in Figure 2. In comparison, the prototype is able to participate in simplified negotiation dialogues (especially in the limitation of requiring shorter turns and simplified utterances, as described above), but it does succeed in exhibiting many of the creative elements from the storyline and role play.

5 Results, Limitations, and Discussion

Through the process described above, we were able to develop a new virtual human negotiation scenario to an implemented prototype stage over a period of 12 weeks. While authoring previous negotiation scenarios had been a more orchestrated undertaking, requiring months of effort by a team of researchers and developers, this experience has produced promising results through the full-time effort of just one researcher, who had no previous experience in virtual human scenario development. The involvement of other researchers with extensive knowledge of the technical capabilities of the system modules was limited to brainstorming and consultations.

Ranger	Hello.
Utah	Howdy, Ranger.
Harmony	Hi, Ranger.
Ranger	Utah do you want the job of town sheriff?
Utah	Well that's your job, we don't need another sheriff.
Ranger	But we need you to become sheriff.
Utah	Okay, Ranger, let's discuss your offer.
	I would need money to hire deputies.
Ranger	We can give you money.
Utah	Harmony, I need your support to become sheriff.
Harmony	It's better if you remain bartender.
	The Ranger and the Deputy should protect the town.
Deputy	We will stay around patrolling the area.
Utah	Ranger, I should have guns to become sheriff.
Ranger	We can deliver a crate of guns.
Utah	Well it sounds very good.
Ranger	Harmony, do you agree?
Harmony	Sure, Utah can be the sheriff.
Ranger	So I think we've got a deal here.
Utah	Okay, I'll pick up the badge and be the sheriff.
Harmony	Okay Ranger, let's make further arrangements later.
Ranger	Thank you, Utah.
	Thank you, Harmony.
	Goodbye.
Harmony	Goodbye.
Utah	Goodbye, Ranger.

Figure 7: A dialogue with the implemented prototype.

In addition to the use of the tools that have been developed to streamline development of these characters, an important factor in this rapid development was the decision to initiate role playing sessions at a very early stage, even while the scenario was still being defined. This enabled a rapid process of extending and improving the developing storyline and task model design, and also provided seed linguistic resources for the prototype implementation.

In the implementation phases, authoring the task model and including all the elements in the ontology required most of the effort, and we observed several opportunities to improve this process. The ontology is designed to keep representations consistent across different modules. However, these elements are dispersed throughout a complex Protégé GUI, and formalizing and representing the model proved very time consuming. New tools could support editing the main elements of a task model – objects, attributes, values, world states, etc. – in a single view, either as a diagram (as in Figure 5) or using an equivalent textual representation, saving considerable time. Further, the process of fine-tuning the task model during testing could be further streamlined. The approach to testing was holistic, i.e., by interacting through conversation with the

implemented virtual humans. This meant that any change to the task model would need to be tested through re-engaging the virtual humans in another testing dialogue. The development of a new tool to automatically identify the effects that changes in the task models would have on the virtual humans' negotiation and dialogue decisions could provide substantial reductions in development time.

Even after 12 weeks of development, the implemented prototype provides only limited coverage for dialogue that could occur naturally with users in the target scenario.

The prototype does include elements that were not present in previous scenarios – most noticeably, the ability to interact with two human players simultaneously – while the complexity of the implemented task models remains comparable. On the other hand, limitations include a reduced robustness in ASR/NLU and a relatively small set of utterances produced by the NLG, when compared to the desired NLG capability for the new scenario. Also, further instances of negotiation are missing: e.g., Harmony's hidden agenda which causes the 4-party conversation to split into two simultaneous 2-party negotiation dialogues. To give a better idea of what has yet to be implemented, the task model for Utah shown in Figure 5 and included in the prototype has 10 states and 6 actions, whereas the target task model has approximately 25 states and 20 actions.

Partly as a result of the limited coverage discussed above, we have not yet evaluated the prototype with live users, and are deferring a user evaluation until we make further extensions to the system through our ongoing iterative development process. However, in this short effort, we have managed to quantify the development effort and difficulties associated with various system building steps, and to identify several opportunities for improvement.

6 Conclusion

We have presented a case study in which a new multi-party virtual human negotiation scenario was implemented over a period of 12 weeks. We have analyzed the effort, expertise, and difficulties encountered at each development step, and identified several opportunities to further streamline the development process. In future work, we intend to use these insights to further lower the development costs and barriers to rapid development of virtual human negotiation scenarios.

Acknowledgments

The project or effort described here 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.

References

M. Core, David R. Traum, H. C. Lane, W. Swartout, S. Marsella, J. Gratch, and M. van Lent. 2006. Teaching negotiation skills through practice and reflection with virtual humans. *Simulation: Transactions of the Society for Modeling and Simulation, 82:685–701, 2006.*, Volume 82, November 2006.

David DeVault, David Traum, and Ron Artstein. 2008. Making grammar-based generation easier to deploy in dialogue systems. In *Ninth SIGdial Workshop on Discourse and Dialogue (SIGdial)*.

Sudeep Gandhe, Nicolle Whitman, David R. Traum, and Ron Artstein. 2009. An integrated authoring tool for tactical questioning dialogue systems. In *6th Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, Pasadena, California, July.

A. Hartholt, T. Russ, D. Traum, E. Hovy, and S. Robinson. 2008. A common ground for virtual humans: Using an ontology in a natural language oriented virtual human architecture. In *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may. European Language Resources Association (ELRA). http://www.lrec-conf.org/proceedings/lrec2008/.

Arno Hartholt, Jonathan Gratch, Lori Weiss, and The Gunslinger Team. 2009. At the virtual frontier: Introducing gunslinger, a multi-character, mixed-reality, story-driven experience. In *Intelligent Virtual Agents*, volume 5773 of *Lecture Notes in Computer Science*, pages 500–501. Springer Berlin / Heidelberg.

A. Jönsson and N. Dahlbäck. 2000. Distilling dialogues: a method using natural dialogue corpora for dialogue systems development. In *Proceedings of the sixth conference on Applied natural language processing*, pages 44–51. Association for Computational Linguistics.

H. Knublauch, R.W. Fergerson, N.F. Noy, and M.A. Musen. 2004. The protégé owl plugin: An open de-

velopment environment for semantic web applications. *The Semantic Web–ISWC 2004*, pages 229–243.

Anton Leuski and David R. Traum. 2010. NPCEditor: A tool for building question-answering characters. In *The 7th International Conference on Language Resources and Evaluation (LREC)*, Valletta, Malta.

Anton Leuski, Ronakkumar Patel, and David Traum. 2006. Building effective question answering characters. In *In Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue*, pages 18–27.

Jeff Rickel and W. Lewis Johnson. 1999. Virtual humans for team training in virtual reality. In *the Ninth World Conference on AI in Education*, pages 578–585.

Kenji Sagae, Gwen Christian, David DeVault, and David R. Traum. 2009. Towards natural language understanding of partial speech recognition results in dialogue systems. In Short Paper Proceedings of NAACL HLT.

William R. Swartout, Jonathan Gratch, Randall W. Hill Jr., Eduard H. Hovy, Stacy Marsella, Jeff Rickel, and David R. Traum. 2006. Toward virtual humans. *AI Magazine*, 27(2):96–108.

David Traum, Jeff Rickel, Stacy Marsella, and Jonathan Gratch. 2003. Negotiation over tasks in hybrid human-agent teams for simulation-based training. In *Proceedings of AAMAS 2003: Second International Joint Conference on Autonomous Agents and Multi-Agent Systems*, pages 441–448, July.

David Traum, William Swartout, Stacy Marsella, and Jonathan Gratch. 2005. Fight, flight, or negotiate: Believable strategies for conversing under crisis. In *proceedings of the Intelligent Virtual Agents Conference (IVA)*, pages 52–64. Springer-Verlag Lecture Notes in Computer Science, September.

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 *The 8th SIGdial Workshop on Discourse and Dialogue*.

D. Traum, S. Marsella, J. Gratch, J. Lee, and A. Hartholt. 2008. Multi-party, multi-issue, multi-strategy negotiation for multi-modal virtual agents. In *Intelligent Virtual Agents*, pages 117–130. Springer.

David Traum. 2003. Semantics and pragmatics of questions and answers for dialogue agents. In *proceedings of the International Workshop on Computational Semantics*, pages 380–394, January.