Talking to Virtual Humans: Dialogue Models and Methodologies for Embodied Conversational Agents

David Traum

Institute for Creative Technologies, University of Southern California
traum@ict.usc.edu

Abstract. Virtual Humans are artificial characters who look and act like humans, but inhabit a simulated environment. One important aspect of many virtual humans is their communicative dialogue ability. In this paper we outline a methodology for study of dialogue behavior and construction of virtual humans. We also consider three architectures for different types of virtual humans that have been built at the Institute for Creative Technologies.

Keywords: Spoken dialogue, methodology, virtual humans, embodied conversational agents.

1 Introduction

Virtual Humans are artificial characters who look and act like humans but inhabit a simulated environment [1]. Virtual Humans can be built for a variety of purposes, including serving as role players in training applications, and non-player characters or artificial players in games. Since Virtual humans are built with explicit computational models of behavior, they can also be used to study how well these models work as models of human behavior. As with other aspects of artificial intelligence, one might focus on just the performance in a task (engineering approach) or the fidelity in modelling human behavior (cognitive science approach). These two goals converge more than for most AI applications, however, because for many purposes one wants a virtual human with human-like behavior rather than efficient behavior which may not be human-like.

We focus here on the dialogue aspect of virtual humans, though the same remarks could also be applied to other aspects. While it is still beyond the state of the art to build virtual humans with all the same capabilities as real people, there are a range of applications for which virtual humans can be useful. Many advantages can also be made by tailoring the virtual human to a specific domain and task rather than trying to attempt general coverage. First, some aspects of human behavior can be elided, as they are not relevant to the given domain. Secondly, one may also be able to take short-cuts in terms of how behavior in that domain is understood and generated, given a smaller set of relevant options. One must be careful, though, to not cut too deep, depending on the purposes. E.g., a more general theory can make it easier to extend the capabilities or move to a new domain.

In the next section, we outline a methodology for the study of interactive dialogue behavior and construction of virtual humans. Key is the use of several different types of scientific activities, applied in a spiral approach to increasing knowledge and virtual

I. Wachsmuth and G. Knoblich (Eds.): Modeling Communication, LNAI 4930, pp. 296–309, 2008.

human capabilities. In section 3, we describe the approach to dialogue modelling and three different virtual human dialogue architectures we have used at the Institute for Creative Technologies.

2 General Methodology

At the current state of practice in building virtual human characters, each one is a different, given it's specific domain and personalized knowledge, but also characteristics of the domain and genre of dialogue it is to engage in. While a lot can be re-used from character to character, we also use fairly different architectures and components for different classes of characters. However, we follow the same broad methodology for development of these architectures and characters. We endorse a cyclical approach, in which multiple passes are made at improving the virtual human, including building a full system fairly early in the process. Figure 1 shows several aspects of the complete process. One may start anywhere for which there are sufficient resources. This process combines a number of different scientific and engineering skills, including observation of behavior, annotation and analysis, theory formation and formalization, and computational modelling and implementation. In many cases, one may start from prior work on some of these aspects and not go through all the steps directly in one project. In other cases, however, the requisite data and understanding of the domain does not exist, and one must spend time developing a corpus of relevant dialogue and/theoretical or notions suitable for formalization of the domain.

On the empirical side, one may start with observation of the communicative behavior of the type of people that virtual humans are to emulate. The kind of behavior performed will depend on a number of factors, including some internal to the people involved, and some based on external aspects of the situation in which they find themselves. There are several issues with respect to which kind of data to collect. First, one needs to collect data from the same sort of activity. Allwood defines social activities as having the following parameters: [2]

1. type, purpose, function: procedures
2. Roles: competence/obligations/rights
3. Instruments: machines/media
4. Other physical environment

We will see very different kinds of language interaction depending on the number and nature of the participants and the activities, e.g., between a formal presentation, a travel agency booking, a courtroom trial, an auction, a press conference, and an informal negotiation. While broad investigation is still needed to be able to recognize the commonalities and differences such activities have on interaction, there have been a few efforts to try to explicitly capture some of this range of interaction. These include the Swedish Spoken Language Corpus [3] and the Dialogue Diversity Corpus collected by William Mann[1]. We may also distinguish activities as to whether they are fully natural (happening on their own, for their own purposes, without regard to experimenter collection), or

[1] http://www-rcf.usc.edu/ billmann/diversity/DDivers-site.htm

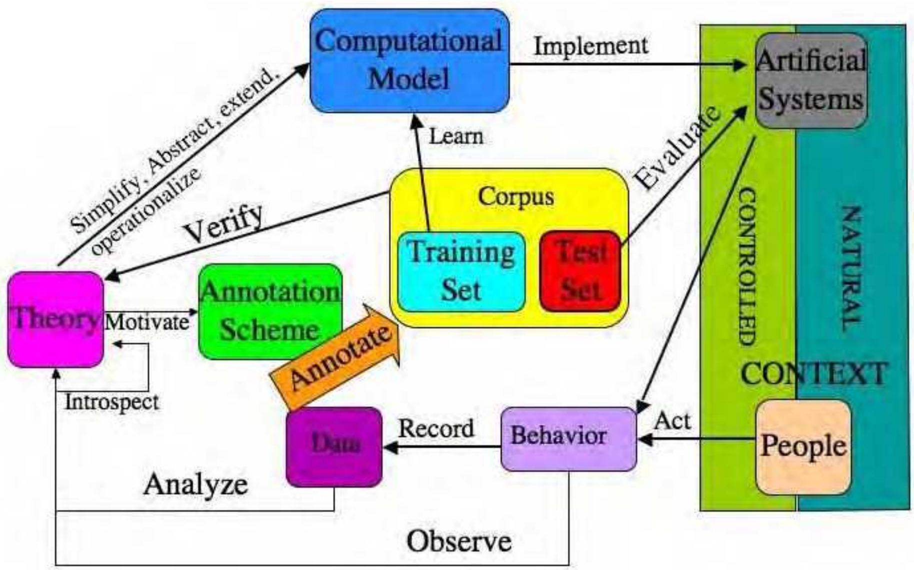

Fig. 1. Methodology for Virtual Human Creation

controlled in some way. The nature of control can also vary quite a bit. At one extreme are situations in which experimenters bring subjects in to participate in laboratory interactions which the participants would never engage in on their own. These include wholly artificial tasks meant to test specific theories, such as minimal pair differences based on different conditions. There might also be naturalistic tasks which participants are asked to role play at for the purposes of the experiment rather than engaging in for their own reasons. Or the tasks may be completely natural except for the presence of experimental observers and/or recording devices. All of these artificial interventions can change the nature of the activity and the interactions which take place. This data can still be of much interest however, as many of the most important characteristics of dialogue will remain, despite the artificial elements.

Observation and analysis of this data can provide insight on some of the common and important aspects, including theories about what kinds of behaviors are produced and the patterns and relationships of behaviors to other behaviors and other aspects of the context. In some cases these can be quite elaborate accounts of specific mechanisms and types of behaviors. In other cases, these might be accounts of possibly significant features that might predict certain types of behaviors.

Theories can also be used to calculate the effects of behavior on the participants, context and future interaction. Some theories will have very broad applicability across a range of participant types, activities, cultures, and specific contexts. Others will be more limited to the specific situation under observation. Without broad observation or detailed generative accounts of the mechanisms causing the behaviors, it can be difficult to tell how widely applicable a theory is. At the current state of the art, both broad and narrow theories are very important: narrow theories can more easily lead to empirical validation and computational models. On the other hand, broad theories will generally be more useful for adapting to new domains with (slightly) changed activities and context.

There are generally two routes to computational models. One is to formalize the theory using human-constructed rules. Often this is not a straightforward process, as the theory is constructed at a very different level than is directly suitable for computational modelling. In some cases, one may need to extend the theory since it depends on commonsense concepts which are not amenable to formalization or have no good extant theory. In other cases, one may need to simplify some aspects that are important for the theory but inaccessible to the computational model. The question then arises as to how much to simplify. Here are several, guidelines for which phenomena in a theory to represent in a computational model for a virtual human in a specific domain. They are ordered from least to most stringent.

1. Represent phenomenon if there is general evidence of its presence in a cognitive model in some domains.
2. Represent phenomenon only if there is evidence from data that it occurs in this domain.
3. Represent phenomenon only if it leads to a functional consequence in agent behavior.
4. Represent phenomenon only if it is the simplest (not necessarily most faithful to the theory) way to achieve the consequence.
5. Represent phenomenon only if it leads to a necessary function for the domain tasks the character must perform.

Each of these guidelines may be appropriate for some modelling tasks, yet inappropriate for others, depending on whether one is most focused on getting a specific character built quickly, or on more extensible and generalizable principles that could be used to model more general behavior or re-use across characters and domains.

In the second approach to forming computational models, one uses theory only to pick out some of the most relevant types of features for analysis rather than a complete algorithmic process for recognizing, processing and producing behaviors. Here one also relies on a corpus of collected interactions with annotations of both the relevant features and the behaviors of interest, and uses machine learning techniques to learn decision procedures. These learning techniques could be of two types. One type includes explicit rules that can be inspected and compared to theoretical constructs, the other type has numerical representations that can be used to compute recognition of categories and behaviors, but does not directly lead to comparison with theories.

It is also possible to combine both the theoretical/algorithmic and machine learning approaches, so that theoretically derived models are used for some aspects and machine learning for others. For example, one might use a data-driven classifier to recognize some aspects of inputs, and a logical or rule-based system to calculate the effects in context, as is done in the MRE and SASO systems described below. One can also apply both types of processing methods to the same phenomena and arbitrate the results when they differ. These hybrid models hold great promise for allowing both robustness to noisy or unseen input while still having broad capability and generative capacity across various content topics.

However the computational model is derived, it can be used as a foundation for an implementation of a virtual human, or component of a virtual human. The implementation is not the end point, however, as the system can now be used to interact with people (and/or other systems) to generate behavior to study. One can evaluate the system from multiple perspectives, including:

– Is it a valid implementation of the computational model?
– Does it faithfully encode a theory?
– Does it have acceptable performance on a "test set"?
– Can it behave appropriately in interaction with people?

Unless virtual humans behave exactly like people, there may also be some reciprocal differences in the way people interact with the virtual human. Human-virtual human interaction thus represents a new type of context that must be analyzed. Purely human data can be used both as a starting point for analysis and implementation of virtual humans, and also as an ultimate performance goal, however it may not be the most adequate direct data, especially for machine learning. For this reason a cyclical approach allows study of how people react to the (previous generation) system, and produces more and more relevant data. In the case where data is needed before it is feasible to build a system, a Wizard of Oz approach [4] is often used, in which a person plays the role of the system, and is limited in some respects to the kinds of interaction that the system will have, while using human-level cognition for other tasks.

Thus, building a virtual human potentially requires all of the following skills:

– Minimally invasive observation and recording of natural human interaction
– experimental design, for controlled data collection
– data recording and organization
– behavior annotation
– theory-formation
– computational modelling
– machine learning algorithms
– programming and system design
– role-playing for specific domains
– wizard abilities
– dialogue evaluation

Not every project will include all of these tasks. Sometimes one can make do with prior work in some areas (e.g., a large extant corpus of recorded and annotated behavior, or a well-developed theory of human behavior). A number of these tasks are required, however, for proper spiral methodology.

Iterations of this process can be used to produce better and better virtual humans. There are several scales on which performance can be increased, including: accuracy of the phenomenon model, complexity of behavior modelled, robustness with respect to types of user, and complexity of tasks that are engaged in. We will look at some of these in more detail.

In terms of complexity of behavior, probably the simplest type of virtual human would be one that focuses on just some aspect of behavior, such as gaze, or backchannels. Here the system is not really engaging in a task with the user, but just displaying this behavior rather than all the other behaviors that would be needed for task performance. This kind of system can be very useful for exploring in detail how that phenomenon works, but does not address the interactions of multiple phenomena within a task. One can also build virtual humans for artificial "toy" domains, which illustrate multiple phenomena interacting in full task behavior, but are not tasks that anyone would naturally do. Examples include games, such as simple matching tasks. These kinds of domains allow progress on some very important phenomena and their integration, while abstracting from the complexities of more realistic tasks. There are also some real-world tasks that are relatively simple, such as information-seeking and direction-giving. More complex tasks, such as negotiation, tutoring, and collaborative construction often involve more complex reasoning, longer interaction, and multiple phases. Finally, one could design a virtual human for integration in a long-range virtual interaction, spanning many interactions with different people and engaging in different tasks.

Robustness of the system interaction can be measured in several ways. One important factor is the type of user involved. Many have remarked on the difference in performance of naive vs expert users of complex systems and user interfaces (e.g., [5]). In many cases this can be overcome by training a user to a system. In other cases, however, it is not practical to train users before interaction. There are, however more degrees of differences in the user population. The easiest type of user to achieve robust behavior with is a demonstrator. This user knows how to follow a "script" to show off the high points of the system while avoiding the weak points. Showing that at least a single reasonable interaction path works can be important in both verifying the integration of the system and fidelity to expected behavior. However, if one needs to do something other than the demonstration, it is not clear that the system will be as robust. The next level of user is a trained expert user. These people will know what works and what doesn't work and how to perform a range of useful tasks even with a system that has some serious flaws. Even novice users fall into multiple categories with respect to robustness. A *motivated* user, who really wants to get the task done with the system, will be willing to try multiple approaches until something works. This user is thus easier to achieve robustness with than a more general population user, who may be using the system only because they are told to (e.g., as part of an experiment) or because it is available (e.g., in a museum) without a specific need. These unmotivated users may quickly give up or move on to other items if the system does not quickly produce desired or interesting results. Finally, there is the *malicious user*, whose main goal is to "break" the system. Here the system must be much more robust to achieve the same levels of performance as with easier users.

This methodology is broadly similar to that employed by other designers of dialogue models for virtual humans, disembodied dialogue systems, and robots. For instance, in the TRAINS project (University of Rochester, 1990-1996), there were several cycles of data collection, theory formation, system building, and evaluation [6,7]. The current exposition is strongly influenced by the methodology used by Cassell and colleagues at MIT and Northwestern [8]. One difference in presentation, at least, is that the model presented in this chapter does not require an initial starting point of collection of human-human data, and the model can be influenced directly by human-computer testing, without explicit re-collection of human data. Li also discusses the use of multiple

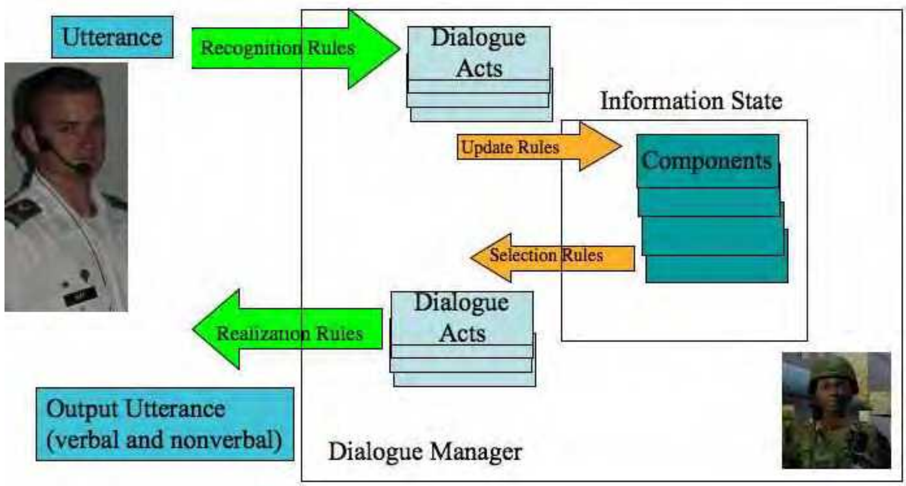

Fig. 2. Information State Approach to Dialogue

implementation-evaluation cycles as the method for design of the dialogue manager of a robot companion [9].

3 Aspects of Dialogue Theory for Virtual Humans

Dialogue interaction, whether in virtual humans or disembodied dialogue systems, can be built using many computational paradigms, from stimulus-response pairs, to finite state machines, to full agents including attitudes such as beliefs, desires, intentions, and complex reasoning. The information state approach [10,11] allows more direct comparisons between these mechanisms and different theories of dialogue. Following this approach, we conceptualize dialogue as a static part, consisting of a set of information state components and current values, and a dynamic part, consisting of dialogue acts that change the information state. These dialogue acts are abstractions of communicative behaviors (including speech non-verbal communicative behaviors) that would achieve the same effect. A dialogue manager for a virtual human consists of at least four processes, as shown in Figure 2. For each dialogue agent architecture (and perhaps even for each domain), there will be a different set of dialogue acts, and different processes. Different architectures may also assign different sets of these functions to different software components.

These processes mediate between observations, internal state, and actions that the agents perform. *Interpretation* is the process of recognizing important actions as having communicative function. From each observation, the interpretation process produces hypotheses about a set of dialogue acts that have been performed. The interpretation process could be formalized as a set of *recognition rules* in a rule-based system. *Update* changes the information state to be in accord with the performance of dialogue acts, given the previous context. This can include adding, deleting, or modifying some aspects of the information state. A theory of information state update can also be organized as a rule-based system, with specific effects for the performance of dialogue acts as well as other update rules. *Selection* is the process of deciding what to do given the current information state. It can be formalized as a choice of dialogue acts to perform, and could be implemented as a set of selection rules in a rule-based system. Finally, *realization* is in some sense the inverse of interpretation, deciding on an ordered set of physical behaviors that can be used to perform the selected dialogue acts give the current context.

Dialogue managers can differ in terms of several features, including the nature of components and dialogue acts, processing mechanisms for each of these processes, and how these processes are apportioned into multiple software modules. In the rest of this chapter, we outline three different architectures for virtual humans that have been built at ICT and their information states.

3.1 Question Answering Characters

Question answering characters have a set of knowledge they can impart when asked and goals for the presentation of this information subject to appropriate conditions. Question-answering characters must remain in character when deciding how to react to questions. Unlike question-answering systems [12] (which slavishly try to find the desired answer), question-answering characters should react to questions the way a person in that situation would, which may include lying, misleading, or finding excuses or other ways to avoid answering questions that they don't want to or are unable to answer. Question answering characters can be used for training, education, and entertainment. At the Institute for Creative Technologies we have recently built several question answering characters, including Sgt Blackwell – a simulated Army soldier who can be interviewed about ICT, the army and virtual human technology, a set of characters a reporter can interview to piece together a news story, and more recently characters who can be interviewed for training tactical questioning. These characters have a limited dialogue model of the character and focus on retrieval of appropriate answers given a question.

Sgt Blackwell, shown in Figure 3, is described more fully in [13,14]. Sgt Blackwell was designed as a technology demo exhibit for a conference. His speech model was designed for limited domain and three specific demonstrators. Sgt Blackwell's dialogue model includes a set of answers constructed ahead of time. These answers are in three categories, (1) "in domain answers", which are simple answers to questions, (2) "off-topic" answers, which are a set of responses to give when there is no appropriate in-domain answer, such as "I don't know" or "why don't you ask someone else", and (3) "prompts", to direct questioners back to the proper domain. The information state is very simple, and consists only of the local history of the last few utterances, and two thresholds: one for avoiding duplication of in domain answers (when possible), and a second threshold for avoiding repetition of off-topic answers. There is also a translation model mapping a language model for questions to a language model for answers, use to score each answer as to how well it addresses a new input question. This allows both high confidence on known questions as well as robustness to speech recognition errors and other small differences in asking the question.

Fig. 3. Sergeant Blackwell Question Answering Character

As described in [14], Sgt Blackwell is indeed robust to speech recognition errors. For known questions, accuracy does not decrease significantly until the word error rate is more than 50%. For novel questions, speech recognition does not significantly impact performance even at higher levels. While further study is required to fully understand the relationship, our hypothesis for at least part of the explanation is that this is so because the same language model is used to train the speech recognizer and the question answer classifier.

3.2 Group Conversation Characters

We have also been working on Group conversation characters, to serve as background characters for larger virtual simulations. These characters are not meant for direct interaction with users, but to serve as a middle level of detail [15]. Their behavior should be natural for a crowd, engaged in conversational interactions, and allow for natural variation for extended durations. We based this work on the conversation simulation of [16].

We have built several models of group conversation, with some examples shown in Figure 4. In [17], we extended the simulation of [16], and used this to animate bodies to drive the minor characters in the Mission Rehearsal Exercise (the left of Figure 4, also seen in the upper right corner of Figure 5). This model was extended in [18], with a new animation system, and including the ability to have members enter and

Fig. 4. Examples of Simulated Group Conversation

leave conversations and have conversation groups separate into subgroups. In [19] we extended the model to include locomotion and positioning.

For these characters, the information state consists of the set of characters and conversations. Each conversation has a set of participants, a turn-holder, a (forcasted) transition relevance place (TRP), and sequences of utterances, consisting of speaker, addressee, and whether it is main content or feedback.

Agents can perform a number of actions, including two types: Speech - which is not directly observable by humans, and non-verbal actions, which are. Speech actions include: beginning to speak, ending speech, TRP signals (that signal a possible end of turn), Pre-TRP signals (that signal that a possible end of turn is coming soon), Addressee selection, and positive and negative feedback. Non-verbal acts include position shifts (movement), orientation shifts, posture shifts, nodding, speaking gestures, and gaze.

Agents also have a set of adjustable parameters that govern their behavior in a probabilistic way. The main parameters are:

talkativeness: the likelihood of wanting to talk
transparency: the likelihood of producing explicit positive and negative feedback, and turn-claiming signals
confidence: the likelihood of interrupting and continuing to speak during simultaneous talk
interactivity: the mean length of turn segments between TRPs
verbosity: the likelihood of continuing the turn after a TRP at which no one is self selected
proxemic distance: the ideal distance between speakers of different familiarity
gaze distribution: the amount of time spent looking at different types of conversational participant (e.g., speaker, addressee, bystander)
overlap offset: the average point at which one will tend to start speaking at an oncoming TRP (before, at, or after) - leading to either small overlaps in speech, exact transitions, or pauses between turns.

These values of these parameters are used to influence behavior according to a probabilistic algorithm that will test against parameter values given configurations of the information state.

Fig. 5. An interactive peacekeeping scenario featuring (left to right in foreground) a sergeant, a mother, and a medic

So far we have evaluated these characters with respect to believability, fidelity of inferences from observed behavior to guiding parameter.

3.3 Advanced Virtual Humans

For deep interaction with humans, we need a richer model of information state. We have developed dialogue models for virtual humans that need to engage in multiparty teamwork and non-team negotiation. In the mission rehearsal exercise project [20] a human user (Army lieutenant) cohabits a 3D graphical virtual environment with animated virtual humans (a sergeant, a medic, a squad of soldiers, and some civilians) and interacts with them through face-to-face spoken dialogue to deal with an unanticipated dilemma (Figure 5) involving a traffic accident causing potentially serious injuries, and a weapons inspection where another unit may require urgent assistance.

Aspects of the information state and dialogue moves are described in [21], and the teamwork model is described in [22]. Figure 6 shows some of the conversational layers. We have evaluated several aspects of the mission rehearsal system, including a number of components of the language understanding capabilities, the system responsiveness and initiative, task success, and user satisfaction. This work is summarized in [23]. The original version of the system was one that was suitable for demonstrators but performed poorly for other classes of users. The final version had suitable performance for motivated users who were familiar with military protocol, but who were not necessarily familiar with interacting with virtual humans.

In the SASO-ST project [24,25], we go beyond team collaboration and negotiation to look at negotiation in a context where collaboration must be achieved rather than taken as a given. The virtual human model was thus extended to include representations of trust and explicit negotiation strategies in addition to the other aspects of information state.

For our first testbed domain, we developed a training scenario in which a local military commander (who has the rank of captain) must negotiate with a medical relief organization. A virtual human plays the role of a doctor running a clinic. A human trainee plays the role of the captain, and is supposed to negotiate with the doctor to

- contact – are individuals available accessible for interaction
- attention – what are individuals attending to
- conversation – what conversations are currently active
  - participants – who are the participants in the conversation
  - turn – who has the right to currently speak in the conversation
  - initiative – who is leading the progression of the conversation
  - grounding – how is information added to the common ground
  - topic – what is the conversation about
  - rhetorical – how is content in the conversation related
- social commitments (obligations)
- social roles – how are individuals related to each other
- negotiation – how do groups converge on shared plans
- individual model (beliefs, desires, intentions)

Fig. 6. Multi-party, Multi-conversation Dialogue Layers

Fig. 7. SASO-ST VR clinic and virtual human doctor

get him to move the clinic, which could be damaged by a planned military operation. Ideally, the captain will convince the doctor without resorting to force or threats and without revealing information about the planned operation. Figure 7 shows the trainee's view of the doctor in his office inside the clinic. The success of the negotiation will depend on the trainee's ability to follow good negotiating techniques, when confronted with different types of behavior from the virtual doctor.

As in the MRE project, we started with a simple version of the character that was suitable for demo users. The initial version was built very quickly, reusing over 80% of the programming of the MRE characters. By using this version to collect data with test subjects, as well as conducting additional role-play and wizard of oz data, we were able to more than double performance of the recognition components and reach a level where users have satisfactory experiences in which their success or failure has more to do with their negotiating tactics than ability to use the system.

4 Conclusions

In this article we have discussed general methodologies for building dialogue components for virtual humans, as well as several examples of different types of such dialogue models. For each of the architectures and domains, a spiral methodology involving all of study of human dialogue behavior, building computational models, implementation of systems, and evaluation of human interaction with systems has led to improved performance along multiple dimensions. This included both allowing a broader class of users to robustly interact with the systems as well as covering more aspects of the phenomena of multi-party multi-modal dialogue.

Acknowledgments

The author would like to thank the many others who have worked on the virtual humans described here and helped form the authors ideas, including Anton Leuski, Bilyana Martinovski, Susan Robinson, Jens Stephan, Ashish Vaswani, Sudeep Gandhe, Dusan Jan, Ronak Patel, Ed Hovy, Shri Narayanan, Rahul Bhagat, Dagen Wang, Jigish Patel, Michael Fleischman,Yosuke Matsusaka, Jeff Rickel, Jon Gratch, Stacy Marsella, Bill Swartout, Lewis Johnson, Patrick Kenny, Jarrell Pair, Pete McNerney, Ed Fast, Arno Hartholt, Andrew Marshall Marcus Thiebaux, Diane Piepol, Ernie Eastland, Justine Cassell, Matthew Stone and Staffan Larsson.

The 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

1. Gratch, J., Rickel, J., Andre, E., Cassell, J., Petajan, E., Badler, N.: Creating interactive virtual humans: Some assembly required. IEEE Intelligent Systems, 54–63 (2002)
2. Allwood, J.: An activity based approach to pragmatics. Technical Report (GPTL) 75, Gothenburg Papers in Theoretical Linguistics, University of Göteborg (1995)
3. Allwood, J.: The Swedish Spoken Language Corpus at Göteborg University. In: Andersson, R., Abelin, Allwood, J., Lindblad, P. (eds.) Fonetik 1999: Proceedings from the Twelfth Swedish Phonetics Conference. Gothenburg Papers in Theoretical Linguistics, Department of Linguistics, Göteborg University, vol. 81, pp. 5–9 (1999)
4. Dahlbäck, N., Jönsson, A., Ahrenberg, L.: Wizard of oz studies – why and how. Knowledge-Based Systems 6(4), 258–266 (1993)
5. Paris, C.: Description strategies for naive and expert users. In: Proceedings of the 1985 Annual Meeting of the Association for Computational Linguistics (ACL-1986), pp. 238–245 (1985)
6. Allen, J.F., Schubert, L.K., Ferguson, G., Heeman, P., Hwang, C.H., Kato, T., Light, M., Martin, N., Miller, B., Poesio, M., Traum, D.R.: The TRAINS project: a case study in building a conversational planning agent. Journal of Experimental and Theoretical AI (to appear, 1995)

7. Allen, J.F., Miller, B.W., Ringger, E.K., Sikorski, T.: A robust system for natural spoken dialogue. In: Proceedings of the 1996 Annual Meeting of the Association for Computational Linguistics (ACL-1996), pp. 62–70 (1996)
8. Cassell, J.: Sistine gap: Essays in the history and philosophy of artificial life. In: Riskin, J. (ed.) Body Language: Lessons from the Near-Human, University of Chicago Press, Chicago (2007)
9. Li, S.: Multi-modal Interaction Management for a Robot Companion. PhD thesis, Bielefeld University (2007)
10. Larsson, S., Traum, D.: Information state and dialogue management in the TRINDI dialogue move engine toolkit. Natural Language Engineering. Special Issue on Spoken Language Dialogue System Engineering 6, 323–340 (2000)
11. Traum, D., Larsson, S.: The information state approach to dialogue management. In: van Kuppevelt, J., Smith, R. (eds.) Current and New Directions in Discourse and Dialogue, pp. 325–353. Kluwer, Dordrecht (2003)
12. Voorhees, E.M.: Overview of the trec 2003 question answering track. In: Proceedings of The Twelfth Text Retrieval Conference, pp. 54–69 (2003)
13. Leuski, A., Pair, J., Traum, D., McNerney, P.J., Georgiou, P., Patel, R.: How to talk to a hologram. In Edmonds. In: Edmonds, E., Riecken, D., Paris, C.L., Sidner, C.L. (eds.) Proceedings of the 11th international conference on Intelligent user interfaces (IUI 2006), Sydney, Australia, pp. 360–362. ACM Press, New York (2006)
14. Leuski, A., Patel, R., Traum, D., Kennedy, B.: Building effective question answering characters. In: Proceedings of the 7th SIGdial Workshop on Discourse and Dialogue, pp. 18–27 (2006)
15. O'Sullivan, C., Cassell, J., Vilhjalmsson, H., Dingaliana, J., Dobbyn, S., McNamee, B., Peters, C., Giang, T.: Levels of detail for crowds and groups. Computer Graphics Forum 21(4) (2002)
16. Padilha, E., Carletta, J.: A simulation of small group discussion. In: Proceedings of EDILOG 2002: Sixth Workshop on the Semantics and Pragmatics of Dialogue, pp. 117–124 (2002)
17. Patel, J., Parker, R., Traum, D.: Simulation of small group discussions for middle level of detail crowds. In: Proceedings of the Army Science Conference (2004)
18. Jan, D., Traum, D.R.: Dialog simulation for background characters. LNCS, pp. 65–74 (2005)
19. Jan, D., Traum, D.: Dynamic movement and positioning of embodied agents in multiparty conversations. In: proceedings of AAMAS 2007: Sixth International Joint Conference on Autonomous Agents and Multi-Agent Systems, pp. 59–66 (2007)
20. Rickel, J., Marsella, S., Gratch, J., Hill, R., Traum, D., Swartout, W.: Toward a new generation of virtual humans for interactive experiences. IEEE Intelligent Systems 17, 32–38 (2002)
21. Traum, D.R., Rickel, J.: Embodied agents for multi-party dialogue in immersive virtual worlds. In: Proceedings of the first International Joint conference on Autonomous Agents and Multiagent systems, pp. 766–773 (2002)
22. Traum, D., Rickel, J., Marsella, S., Gratch, J.: 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, pp. 441–448 (July 2003)
23. Traum, D.R., Robinson, S., Stephan, J.: Evaluation of multi-party virtual reality dialogue interaction. In: Proceedings of Fourth International Conference on Language Resources and Evaluation (LREC 2004), pp. 1699–1702 (2004)
24. Traum, D., Swartout, W., Marsella, S., Gratch, J.: Virtual humans for non-team interaction training. In: proceedings of the AAMAS Workshop on Creating Bonds with Embodied Conversational Agents (July 2005)
25. Traum, D., et al.: Fight, flight, or negotiate: Believable strategies for conversing under crisis. In: Panayiotopoulos, T., et al. (eds.) IVA 2005. LNCS (LNAI), vol. 3661, pp. 52–64. Springer, Heidelberg (2005)