Dialog Simulation for Background Characters

Dušan Jan and David R. Traum

USC Institute for Creative Technologies, 13274 Fiji Way, Marina del Rey, CA 90292

Abstract. Background characters in virtual environments do not require the same amount of processing that is usually required by main characters, however we still want simulation that is more believable than random behavior. We describe an algorithm that generates behavior for background characters involved in conversation that supports dynamic changes to conversation group structure. We present an evaluation of this algorithm and make suggestions on how to further improve believability of the simulation.

1 Introduction

When we are dealing with virtual environments with a large number of virtual characters we encounter the problem when it is no longer feasible to simulate each character as a fully animated conversational agent. It is acceptable and desirable to provide main characters with speech recognition, speech synthesis, high-fidelity gestures and lipsynch, analysis of input and an AI agent capable of making informed decisions. On the other hand we would have to spend too many resources to support this for many characters.

While we can ignore characters that are too far away from action, we need at least some form of scripting to control the behavior of background characters in view, to maintain the illusion [6]. Such scripting in form of hand animating all motions is labor intensive and, if the simulation goes on for longer than the amount of hand-animated material, usually provides repetitive behavior which detracts from realism. One solution to this problem is use of simulation algorithms to generate dynamic behavior for background characters. In our case [2] we were mainly interested in behavior of agents involved in conversation. While [8] bases its multimodal conversation model on information structure, a model based on visual perception of the scene rather than on speech is more appropriate for our domain. Therefore we decided to use the algorithm [1] proposed by Padilha and Carletta as a starting point for our simulation.

We have extended the work in [2], adapting to characters in the unreal tournament game engine, and allowing more dynamic starting, ending, and joining of conversation flow. One of the limitations of the simulation algorithms in [1] and [2] was the fact that it only supported one dialog going on at a time, meaning that all characters participated in the same conversation. While we could run multiple conversation simulations and explicitly assign different characters to different conversations, this is still not realistic for many situations in which characters move around and join or leave conversations. Likewise, even when

T. Panayiotopoulos et al. (Eds.): IVA 2005, LNAI 3661, pp. 65–74, 2005.

people stay in the same position (e.g., at a meal or meeting), there are often dynamic splits and realignments into sub-conversations. Adapting a simulation that can handle dynamic creation and entry to conversations will allow more realism as well as scaling to situations with larger numbers of characters than would be supported by a single conversation.

2 Background

Since the conversation simulation is meant for background characters, who are too far away from the main action to hear the content, we focus on the appearance of conversation and the patterns of interaction, rather than actual information exchange or communication of internal state. To achieve realistic behavior we use behavior observed in real human conversations, as synthesized in [1]. Participants in conversation take turns at talk. During the turn they monitor others to see if the others follow the conversations and to react to feedback given by other participants. When the speaker is speaking there are natural points where others can begin their turn. These are called transition relevance points (TRPs). If the speaker addresses a particular participant with a question then that person will usually take a turn at the TRP. On the other hand if the speaker leaves a free TRP anyone can select to speak or the speaker may choose to continue to talk.

At a free TRP we can have more than one participant deciding to start to talk. In such cases we may have overlapped speech and there are various factors that influence who keeps speaking. Another case that involves simultaneous speech are interruptions. These can have several outcomes. The interrupter may stop after a false start, the original speaker may be cut off or the original speaker could decide to ignore the interrupter if he perceives the interruption as side talk and is not bothered by it.

Most transitions however will happen at TRPs with only a small gap or no gap at all. This is possible since the participants can anticipate the time TRP will occur based on speech characteristics and also other non-verbal behavior. Since we do not generate actual content of conversation speakers have to provide explicit pre-TRP cues to give participants the level of information required to behave realistically such as changing postures and similar non-verbal behavior that indicates the intention of taking the turn at the next TRP.

3 Aspects of Conversational Simulation

In our test scenario we connected our algorithm to virtual characters within the Unreal Tournament game engine. These characters had a small set of animations available to indicate different modalities of conversations we were simulating. These animations could be triggered by calling Unreal Tournament commands from an external character controller. Besides the outputs that trigger animations we also have messages between characters (such as TRP signals, selection

speech
- begin speaking
- end speaking
- pre-TRP signal
- TRP signal
- selection of addressee
- positive or negative feedback

non-verbal
- nodding
- gestures
- posture shifts
- gaze

Fig. 1. Conversation Agent Message types

of addressee) that do not result in any direct visible outcome in the simulation. Message types are shown in Figure 1.

The implementation in [2] used a blackboard where all participants of the conversation would exchange their messages and also had a fixed conversation cycle synchronized between all participants. We found this setting limiting in making the algorithm suitable for large number of agents with several ongoing conversations so we implemented a distributed solution in which each character implements its own decisions independently from other characters. To facilitate this we have each character controller running in a separate thread, communicating with other characters using messages. When a character receives a new message it can react to it immediately or just update its internal state and make a decision during normally scheduled processing.

In the algorithms in [1] and [2] every character was in conversation all the time. Our extensions, however, allow situations when a character is not involved in a conversation at all. From this arises the need to have some higher order planning involved which decides when the character should join an existing conversation, when it should start a new conversation and also when to leave conversation because of the reasons external to the dialog simulation itself. In a real virtual environment simulation this would also include planning for walking around and performing other activities, but in this simple scenario we started with conversation activities only.

Behavior of characters is controlled by a set of properties in a probabilistic manner as in the previous algorithms. Each of these properties has a value from 0 to 1. Whenever one of these properties is tested, a random number is selected and compared to the property value (possibly scaled based on contingent factors of the conversation). The properties currently used are shown in Figure 2.

Each character also keeps track of information about other characters. Each character tracks the gaze of each other, and whether they are speaking, and how long since that character has interacted in the conversation group of the tracker. Characters also track the composition of their conversation group — conversation groups are not defined externally but interpreted on the basis of perceived actions. Characters can also mis-interpret the actions of others, and

talkativeness	likelihood of wanting to talk
transparency	likelihood of producing explicit positive and negative feedback, and turn-claiming signals
confidence	likelihood of interrupting and continuing to speak during simultaneous talk
interactivity	the mean length of turn segments between TRPs
verbosity	likelihood of continuing the turn after a TRP at which no one is self selected

Fig. 2. Conversational Agent Properties

can have different ideas about the composition of a conversation group. In future work, we will use more realistic approaches to perception (e.g., [7]), so that agents will only observe gaze and speech within their focus of attention.

4 Conversational Participation Algorithm

Each character runs a separate instance of the algorithm in its own thread, with its own setting for the attributes, and its own internal representation of the behaviors of others and group composition. The algorithm mainly consists of a series of event handlers. We briefly describe the major events and behaviors.

4.1 High-Level Planning

This part of the code is external to the main conversation algorithm. It represents the high order planning of the character and in our case makes characters join or leave conversation.

```
every planning cycle (approx. every 5 sec)
    if in conversation
        test to leave conversation
    else if talkativeness test successful
        decide to join existing conversation
        or start a new conversation
```

4.2 Claiming a Turn

Characters decide (using the talkativeness parameter) whether or not to take a turn when they receive pre-TRP signal. If they decide they will speak, they will also decide (using the transparency parameter) whether to signal their intention to speak with turn claiming signals if appropriate.

```
when receiving pre-TRP signal
    test talkativeness to decide to speak
    if so, test transparency to make turn claiming signal
```

4.3 Starting to Speak

Whenever the character starts to speak it determines the timing of its turn, including when to send a pre-TRP signal.

```
when starting to speak
    if at TRP and someone already started speaking
        test confidence to continue speaking
    select segment length based on interactivity
```

4.4 Continuing Speaking

Sometimes when one finishes a segment, no one else takes over. In this case the agent has the option to continue his own speech beyond what was initially planned.

```
when you end segment and no one takes turn
    test verbosity to continue speaking
```

4.5 Tracking Others Participation

Whenever an agent speaks or gives feedback to someone in a conversation group, they will be an active participant as well. This section maintains the conversational group and activity of its members.

```
when receiving input from other characters
    if they are signalling to someone in my group
         then add them to group (if not already there)
    if they are in my group and addressing someone in my group
          update last time they were active
```

4.6 Responding to Others

This section calculates how an agent should respond to the initiation of speech by another. Reaction will depend on whether the agent is also speaking and who started first, whether the agent is part of the same conversation as the speaker, and parameters of confidence (whether to continue speaking or not), talkativeness (whether to join a conversation), and transparency (whether to show feedback behavior). Decisions about leaving one conversation for another are also made if a character is addressed by someone who is not in the same conversation.

```
when someone starts to speak
    if in conversation with me
        if at TRP and I already started speaking
            test confidence to continue speaking
        if not speaking
            test transparency to gaze at speaker
    if I am not in conversation and they are speaking to me
        test talkativeness to join conversation
        test transparency to give signals of joining
```

4.7 Main Loop

Below is the main loop that agents go through, as modified by the above events.

```
every conversation cycle (approx. every 0.5 sec)
    remove characters that were inactive for too long
    if no one is speaking
        test talkativeness to start to speak
        if so, start with random interval
            select addressee
            test transparency to shift posture
        if no one was speaking for some time
            if talkativeness test fails leave conversation
    if listening to someone
        if there is more than one speaker for some time
            group was split into two or more conversations
                keep speaker that I am listening to
                remove participants that are  attending to others
         test talkativeness and confidence to interrupt
    if speaking simultaneously
        if there is only one additional speaker
            and their addressee attends to them
                then treat this as a side talk
                   remove both from conversation
        otherwise test confidence to continue speaking
    if speaking alone in a turn
        decide when to gesture and gaze away
        if no one is paying attention to me
            if confidence test fails stop speaking
```

5 Evaluation

There are many possible ways to evaluate the simulation. One can try to fit the model to observed conversations, as suggested by [1]. One could also test the differences in simulation that would result from different sets of characters with different sets of parameter values, e.g., whether it leads to domination of the conversation by a single character or small set of characters. As suggested in [2], we decided to test if the simulation "looks like a conversation" to the viewer.

In our test scenario we used for our characters 6 Iraqi civilians that initially are not involved in conversation. We recorded several simulations with different character attributes and stored videos and internal logs of each agent to later analyze and compare their internal states with responses from the viewers. A snapshot from a conversation simulation is shown in Figure 3. We balanced selection of attributes with the physical bodies to control for surface characteristics of the bodies and the effect of positioning. We also made one simulation where characters decided randomly when to start speaking and who to gaze to in order to have a baseline for comparison with our algorithm.

Fig. 3. Iraqi civilians engaged in conversation

We created 3 different tests for the viewers. In the first part they were asked to view several 30 second clips of simulations and decide how believable they think each simulation was on a 7-point Likert scale. We also asked them to provide any information about what factors they thought made the conversation less believable. In the instructions we also made clear to viewers that when judging believability of the simulation they were to pay most attention the appropriateness of behavior, particularly gaze and dialogue rather than animation quality of the characters.

In the second part we asked viewers to view multiple 2 minute clips of simulations. We instructed them to pay attention to only one of the characters (different characters for different clips) and analyze their behavior. Since the attributes used in the algorithm are not all very visible in such a short dialog we decided to ask viewers about the perceived properties of the characters rather than about underlying attributes. We asked viewers to judge the following properties on scale from 1 to 7:

talkative how often is he talking:
- **1** – almost never talks
- **4** – talks about as much as everyone else
- **7** – talks almost all the time

predictive does he give any signals before speaking:
  1 – never gives any hints that he is about to speak
  7 – always indicates that he wants to speak
transparent is he giving any signals that he is attending to the speaker:
  1 – seems oblivious to others
  7 – always signals understanding of others
interruptive is he interrupting when others are speaking:
  1 – always waits for others to finish
  7 – jumps into conversation all the time
confident is he likely to keep talking if others speak at the same time:
  1 – gives up his turn if someone else starts to speak,
  7 – never shuts up when others speak

How talkative a character is is influenced by talkativeness attribute, predictive and transparent are both influenced by transparency. Confident characters have high confidence attribute and interruptiveness is determined by combination of both talkativeness and confidence. We have not asked about verbosity or interactivity because that would require observation of longer segments to get significant results.

In the last part we asked viewers to track who they think is speaking with whom, again for clips of 2 minutes in length. We used this data to compare how the internal state of each character correlates to what is perceived by the viewer.

6 Results and Future Work

Eight people of various ages and cultural background anonymously participated in our web-based evaluation. The average believability score for our algorithm was 5.3 compared to score of 3.3 for random behavior. The difference is statistically significant which indicates that most viewers were able to identify the random behavior. We found that the highest scores were received by simulations where either all characters participate in the same conversation or where the conversation groups correspond to positioning of the characters in the setting. Since our algorithm does not take positioning of characters into effect when deciding about creating new conversations and allowing conversations to split it is not able to prevent this kind of undesirable behavior from happening. We propose to make modifications to the algorithm that will take positioning into account and will also control character movement to achieve positioning where characters in the same conversation group separate themselves from other characters. We plan to achieve this by tracking noise level for each character. Each speaking character that is speaking, but not in the conversation group of this character, would contribute to the noise level based on their distance. If the noise level would get too high characters would either decide to break their conversations or move away from characters that bother them in their conversation.

Part 2 proved to be a lot more difficult than we expected. Not only were there differences between the values predicted by underlying attributes and results from viewers, but also the values varied widely between viewers. This would suggest that it is hard for humans to judge what the personality of a virtual character is, probably because of the lack of expressiveness when we compare virtual characters to real humans. We guess that it would be hard to grasp the personality of a background character anyway. However, we still think that having parameterized algorithm has its benefits since the structure of dialog changes with different attribute settings. Since it is hard to evaluate personality of a single character we propose to evaluate how different personality compositions influence believability of simulation.

Results from part 3 showed that what viewers perceived roughly agreed with the internal state of the characters. When a certain group composition was held for a longer time most of the characters and viewers agreed with what the current group composition was. Most of them correctly differentiated between normal transitions, interruptions and side conversations. However when the side conversations do not last long the results vary between characters and also between viewers.

We have not yet tested the algorithm with large numbers of virtual characters, but as Ulicny reports in [9] the limiting part in large scale crowd simulations is usually in rendering and not in the behavior generation. From our work so far we can see that it is beneficial to dynamically create behavior for background characters as it both removes labor intensive work of creating scripts and also improves believability of the simulations. However, we have seen from the evaluation results that we have a lot of room for improvements, especially in incorporating character movement in the simulations.

Acknowledgements

We would like to thank Patrick Kenny, Richard Almodovar, and Kurosh Valenejad for help with unreal tournament animations. We would also like to thank the anonymous reviewers for helpful comments on this paper. The project 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. Padilha, E., Carletta, J.: A simulation of small group discussion. Proceedings of EDILOG 2002: Sixth Workshop on the Semantics and Pragmatics of Dialogue. (2002) 117–124
2. Patel, J., Parker, R., Traum, D.R.: Simulation of Small Group Discussions for Middle Level of Detail Crowds. Army Science Conference (Orlando, November 2004)
3. Padilha, E., Carletta, J.: Nonverbal behaviours improving a simulation of small group discussion. Proc. 1st Nordic Symp. on Multimodal Comm. (2003) 93–105
4. O'Sullivan, C., Cassell, J., Vilhjalmsson, H., Dingliana, J., Dobbyn, S., McNamee, B., Peters, C., Giang, T.: Levels of detail for crowds and groups. Computer Graphics Forum. 21(4) 2002

5. 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** (2002)
6. Swartout, W., Hill, R., Gratch, J., Johnson, W.L., Kyriakakis, C., Labore, K., Lindheim, R., Marsella, S., Miraglia, D., Moore, B., Morie, J., Rickel, J., Thiebaux, M., Tuch, L., Whitney, R., Douglas, J.: Toward the holodeck: Integrating graphics, sound, character and story. Proceedings of 5th International Conference on Autonomous Agents. (2001)
7. Hill, R. Perceptual Attention in Virtual Humans: Toward Realistic and Believable Gaze Behaviors. In Proceedings of the AAAI Fall Symposium on Simulating Human Agents, pp.46-52, AAAI Press, Menlo Park, Calif., 2000.
8. Cassell, J., Torres, O., Prevost, S.: Turn taking vs. Discourse Structure: How Best to Model Multimodal Conversation
9. Ulicny, B., Thalmann, D.: Crowd simulation for interactive virtual environments and VR training systems