Controlling Three Agents in a Quarrel: Lessons Learnt

Cyril Brom, Petr Babor, Markéta Popelová, Michal Bída, Jakub Tomek, Jakub Gemrot

Charles University in Prague, Faculty of Mathematics and Physics,
Malostranské nám. 2/25, Prague, Czech Republic

Abstract. Steering behaviours can be used to position 3D embodied agents in small groups engaged in relatively simple social interactions such as in group conversation or walking while talking. Less is known about scaling these mechanisms for situations with complex dynamics requiring agents to perform actions beyond walking, turning, talking and gesturing. Here, we present a model for controlling three agents in an example of such a situation: a vigorous quarrel. The model combines a general steering behaviour for keeping the three agents in a triangular formation with a probabilistic two-level hierarchical state machine (hFSM) for unfolding the quarrel by means of changing parameters of the steering behaviour and issuing actions to the agents. The model has been implemented using UnrealEngine2Runtime on the example of a boy dating two girls at the same time who do not know about each other. The user can influence the course of the quarrel by changing attitudes among the agents. To create a list of the agents' actions and the hFSM, we video-taped about 40 episodes in which three actors improvised on the topic of the quarrel, and we manually annotated the videos. The evaluation with 67 human participants indicates that the model produces outcomes comprehensible and believable even for persons with limited previous experience with 3D graphics. On a more general level, this paper suggests that augmenting steering behaviours by a non-trivial higher-level controller is a feasible approach to modelling behaviour of 3D agents interacting in small groups in a complex way and presents a possible workflow for developing scenes featuring such agents.

1 Introduction

Populating virtual worlds with human-like agents is becoming a norm in many applications, yet controlling these agents during complex social interactions in small groups is largely a terra-incognita. As examples of "small group social interactions," consider a couple of friends walking in a shopping mall, looking at shop windows and discussing the wares, or one of these friends buying something from a sales clerk while the other friend starts arguing with the clerk about quality of the goods. The trouble is that we need to control not only positioning of these agents, but also their bodily actions, such as pointing at a shop window at the right time.

In today's commercial computer games, when such interactions should be depicted, designers traditionally employ a scripted cut-scene. Years of research have produced a body of work on modelling gazing behaviour, facial expressions and

adfa, p. 1, 2011.

gesturing of embodied conversational agents, and also on setting distances between these agents when they speak; all of this considering gender and cultural differences, e.g. [2]. However, to our knowledge, this research body is almost silent on the topic of sequencing complex actions while (at the same time) dynamically re-positioning the agents during the conversation. At the same time, in the field of crowd simulations, e.g. [9, 18], the attention is predominantly devoted to movement of agents in large groups and formation/disintegration of these groups, but less on complex interaction in small groups. In a nutshell, these two notable research directions are *complementary* to attempts at modelling agents engaged in *complex* interactions in small groups.

Walking of pedestrians in groups of two or three was already modelled [5], including reshaping their formation in a narrow corridor or when passing through a larger crowd. Popelova et al. [14] modelled behaviour of two friends walking together, including one of them waiting for the other. Jan and Traum [4] modelled agents conversing in small groups, including their turning, re/positioning on a circle, and joining and leaving the group. The same seems to be the case with BierGarten simulations, though the BierGarten team's report is less detailed [2]. Others [12, 17] used more general models to model various small group conversation types. However, to our knowledge, none of these works feature agents performing more complex actions such as pushing each other or giving/taking an object, leading to constant reshaping of the agents' formation. Finally, Mateas'es Façade [8] featured agents performing complex social actions. That interactive drama project, however, was (intentionally) not full 3D and it used two computer-controlled agents plus a human avatar.

The goal of this paper is to present a model for controlling three 3D agents in a quarrel that goes beyond previous models for controlling agents in a small group conversation in two ways. First, the agents are not positioned in a simple circular (or, in our case, triangular) formation, but the formation dynamically reshapes as the quarrel unfolds and the attitudes among agents change. Second, the agents perform complex actions such as pushing one another, walking backwards, slapping, and conversing using emoticons (Fig. 1 Left). Emoticons are used to substitute verbal conversation, not just to express the agents' emotions.

We have chosen the following situation for modelling: *A boy dates two girls at the same time, but the girls do not know about each other. One day, the trio meets and things happen...* Our motivation for picking this particular situation was: a) three agents interact, b) complex actions and repositioning of the agents happen naturally, c) the situation is expressive enough for a human observer to understand it, while the agents use body language and emoticons. The model is implemented in Unreal-Engine2Runtime (UE2) using our toolkit Pogamut [3].

The structure of the paper is as follows: Sec. 2 introduces architecture of the model and design considerations. Sec. 3 details how data for constructing and parameterising the model were acquired. Sec. 4 details the model. Sec. 5 presents evaluation of the model we conducted with 67 human participants. The purpose of the evaluation was to investigate whether the outcome is understandable to a human observer familiar with 3D graphics. Sec. 6 discusses the outcomes, emphasising lessons learnt.

2 Architecture and Design Considerations

The technical starting point is a 3D virtual environment with mocapped animations. In our work, we abstract from generating gestures procedurally, head movements and facial expressions, and their blending with underlying mocap animations (that can be in fact conceived as an additional complexity layer). Instead, we focus on production of the overall quarrel, which should last around 1 – 3 minutes and should be engaging and understandable for a human observer from the beginning to the end.

The general idea of our approach is to organise tiny pieces of adaptable scripts on the top of a triangular steering behaviour, which repositions the agents. By “adaptable scripts” we mean a couple seconds-long sequences of several animations and emoticons triggered by actual context: an “adaptable script” may not always generate exactly the same outcome. By “organise” we mean that when a script finishes, a new one is chosen reactively in real time according to the context, an input from a user who can change emotional attitudes among the agents, and a random element. This “organisation” is controlled by a probabilistic hierarchical state machine (hFSM).

Our idea vaguely corresponds to the notion of sequencing “units of drama” as used in interactive storytelling, e.g. [8, 13]. The hFSM fulfils the role of a reactive drama manager. Based on the results of the video annotations (Sec. 3), we decided to use a two-layered hFSM (Sec. 4). Inspired by the terminology of Mateas [8], we call the higher-level states *beats* and the lower-level states *minibeats*.

From the point of view of the three-layered architecture for controlling motion of virtual agents by Reynolds [16] our model looks as follows (see Fig. 1 Right):

1. At the action selection layer, the drama manager determines the active steering behaviours and their parameters, active animations and emoticons (all this is defined in the current minibeat). This layer is common to all three agents.

2. At the steering layer, the velocity of the steered agent is computed according to the active steering behaviours. Besides the traditional steering behaviours (Leader Following, Target Approaching, Obstacle Avoidance, etc.), our model uses a new Triangular Steering Behaviour: to keep three agents in a specific triangular formation (Sec. 4). This layer is autonomous for every agent.

3. The locomotion layer moves the agent according to the given velocity. This layer is autonomous for every agent.

Note that in most of the prototypes of small group interactions mentioned above, the action selection layer is missing or relatively trivial, e.g. [4, 17].

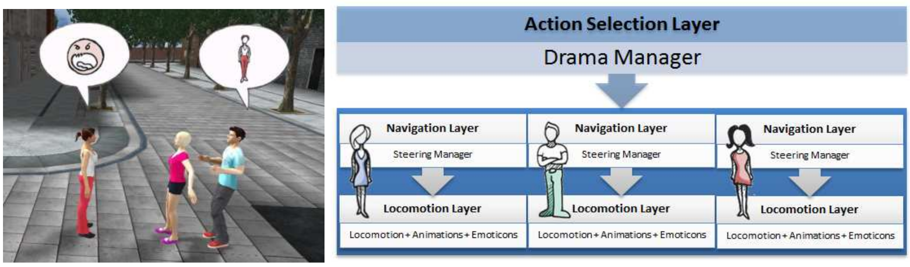

Fig. 1. Left: A quarrel example. Right: Architecture of the model, inspired by Reynolds [16].

3 Acquiring and Annotating Background Video-materials

It would be hard if not impossible to model a quarrel without knowing how quarrels unfold. How many quarrel types exist among a boy dating two girls, when the two girls suddenly meet the boy and realize they are being cheated on? How long do quarrels last and how complex they are? What are the possible outcomes? Crucially, this type of questions is rising when modelling almost any complex social interaction. To answer such questions, it is a good idea to run a small exploratory study in which we acquire data on how real humans behave in respective situations, identify common patterns and, consequently, develop a formalism capturing these patterns.

Recording the background video-materials. We hired two groups of three actors each and a theatre director to improvise on our topic. This approach is inspired by the work on Magerko, e.g. [10], on "digitalising" improvisational theatre, and Kendon [6] on analysing formations during multi-party conversations. Our aim was more focused than those of Magerko and Kendon, we aimed at identifying a reasonable number of behavioural patterns during our type of quarrel. The patterns were operationally defined as a) common actions that last approx. 1 – 3 seconds, correspond to a verb and can be mocapped, b) common action pairs between two agents, c) common groups of about 3 - 6 actions and/or action pairs that visually constitute a unit and can be described by a sentence or two, d) large units composed of (a) - (c) that constitute phases of the quarrel and can be described by a paragraph. Because our study was an exploratory one, we in fact did not know in advance if we would detect any (c) or (d).

We had two recording sessions, one with each group and the same director, two months apart. At the beginning of each session, the trio and the director were introduced to the general setting. Then, the actors and the director were demonstrated our 3D virtual world with some agents and their animations and were explained the purpose of the work. Then, the director had to figure out as many variants on the topic (i.e. quarrel instances) as possible, verbalise every variant using one or few sentences and let the actors to improvise it. The following constraints were given:

1. Every variant should have an understandable beginning and an ending phase, having a clear "narrative arc" is an advantage.
2. At the beginning and in the end, the actors may not stay together, but they should stay together at least for a while during the acting the situation variant.
3. The situation variant should last up to 5 minutes.
4. There should be no objects involved.
5. The actors can speak, but they should pay attention to the fact that the virtual agents will express themselves only using bodily gestures, motion, and emoticons.
6. Behaviours involving extreme contact, such as fighting on the ground, should be avoided.
7. The variants should differ but there can be overlaps between their parts.

An example of the director's description is: "The boy walks with the girl A while the girl B runs up from behind. They start arguing, the boy jilts both the girls and leaves, the girls leave together." Note that the point (7) is crucial because we wanted parts of the virtual quarrel to be reused in multiple quarrel instances.

The session ended when the director was unable to come up with a new variant. Each session lasted over an hour and we recorded about 20 variants, some actually very similar. After the 2nd session, we felt that the topic is nearly "depleted," signaling us that a third session would not bring much new data.[1]

We used two cameras to avoid occlusion.

Annotating the video-materials and results. Several situations violated some of the constraints, mainly (5) and (6), and some situations (mainly across the sessions) were very similar. In the end, we manually annotated 13 different situations in detail, identified behavioural patterns (a) – (d) in them, and created a list of animations and emoticons, which we do not have, but can mocap/create. We focused also on actors' emotional state during situations and used it later in the emotional parameterization of the model.

In addition, we scripted several machinimas in our freely available toolkit StoryFactory [1] based on the annotated data to verify that it is possible to mirror the "real" quarrels in UE2 – that turned out to be possible.

Table 1. Examples of minibeats from the middle phase of the quarrel.

Name	Description
handshaking	the girls handshake
speaking (4 variants)	regular conversation, different triangular formation or conversational style
the boy in the middle	the girls arguing angrily, the boy goes between them to calm them down
triangular pointing	everyone argues at the same time and are pointing at one another
the boy takes one girl away	the boy tries to go away with one girl
a girl waving	the girl is waving at the other two
a girl repelling	one girl is repelling the other girl away from the boy
a girl protecting	the boy is behind a girl, who protects him against the other girl (who has arms outstretched)
remote talking	one girl is close to the boy, the other girl is 5 meters apart talking to the boy
the boy taking both	the boy attempts to take both girls by hands and leave
the boy calming both	both girls standing arms akimbo, the boy calming them, in a triangle
the boy going for a hug	the boy goes from one girl to hug the other one
kissing/jumping back	the boy tries to kiss a girl, she jumps back
successful hugging / caressing (4 variants)	a person hugs/caresses a different person
attempting to push	a person pushes a different person, who remains still
the girls fighting (2 var.)	the girls fight, the boy either applauds ironically or leaves
a girl winning a fight	the girl wins the fight
all fighting	the trio is fighting
girls beating the boy	the girls beat the boy, each from one side
slapping	a person slaps another person
pushing each other	the trio is pushing each other chaotically
a girl kicking the boy	the girl kicks the boy, the other girl may applaud
the boy decoying	the boy decoys a girl away pretending showing her something interesting

[1] Note we would have needed much more data would have we wanted to *learn* a model using the data, cf. [11].

We identified over 70 different actions of type (a), from which about 30 were newly mocapped for the purpose of this project (and about 20 impossible to mocap). In the prototype described in this paper, we finally employed over 50 different actions, i.e., individual animations, and over 30 emoticons. We identified 10 common action pairs of type (b) such as hysterical action – calming down, over 50 units of type (c), and 13 units of type (d).

For the purpose of creating the formalism for behavioural patterns, we merged action pairs (i.e., (b)) with units of type (c), giving us three levels of abstraction: (a), (b + c), (d). The reason was that branching between levels (b) and (c) was small. The (b + c) layer corresponds to minibeats (see Tab. 1) and (a) layer to beats.

In addition, we had enough data to identify roughly common transitions among minibeats and beats.[1] At the end, it appeared that the structure of the quarrel could be modelled using a two-level hFSM with conditional and, to some extent, probabilistic transitions. We will return to the idea whether this was a good choice in Sec. 6.

4 The Model

We first introduce the hFSM for unfolding the quarrel and then detail the Triangular Steering Behaviour.

Action Selection Layer: Unfolding the Quarrel. Based on the outcome of our analysis described in Sec. 3, the state machine has two layers: the top layer comprises beats while the bottom layer minibeats. Beats model larger units of the scenario lasting about 30 - 60 seconds, minibeats are usually shorter than 10 seconds. Specification of the state machine can be given in an *xml* file. Some beats are specific, such as “Boy is coming with Girl-1 while Girl-2 is waiting,” others are generic, having role-slots instantiated in real time, such as “X calms Y while Z is watching.”

We will first zoom to minibeats. In a minibeat, the designer can specify:

1. type of the steering behaviour used and its parameters (i.e., Triangular Steering Behaviour, Leader Following or Target Approaching);
2. the list of animations and emoticons to be triggered (see below);
3. changes of attitude among agents (attitude of an agent towards each of the other two agents is represented by a number between < -1, 1 >);
4. a timeout (the maximal duration of the minibeat).

A minibeat uses an animation and emoticon streams. In the future, it would be advantageous to blend motion animation, e.g., walking, with a torso/head animation, giving us three or four expression streams, but for the sake of present prototype, we use only animation and emoticon streams. Note that it is not possible to just sequence animations within a minibeat like in a cut-scene, because the agents’ initial positions may differ in different minibeat’s runs, which means it may be needed to schedule actions differently from run to run. Therefore, animations and emoticons have specific constraints such as that the action must start during or after the end of other actions (of any agent). Actions can also be interruptible or uninterruptible.

Knowing the structure of minibeats, we can look at beats. Beats represent various types of beginnings, middle phases and ends of the quarrel and their main purpose is to simplify design. A beat consists of several minibeats with specified transitions between them. A minibeat can be reused in several beats. The transitions are probabilistic and depend on current relations between agents. E.g., if Thomas still likes Barbara, there is a higher probability that the next minibeat will be “Thomas kisses Barbara and Nataly is angry” than “Thomas kisses Nataly and Barbara is angry”. The relations between agents can be changed in the minibeat, or by a user, or – in theory – by any higher layer of the drama manager.

Triangular Steering Behaviour. The Triangular Steering Behaviour (TS behaviour) was designed to steer an agent during a conversation with the other two agents. The steered agent (agent X) should keep a specific position and heading depending on location of the other agents (agents A, B). The TS behaviour has these parameters:

1. *Agent A* – the name of the second agent;
2. *Agent B* – the name of the third agent;
3. <*minA*, *maxA*> – the interval specifying required distance from the agent A;
4. <*minB*, *maxB*> – the interval specifying required distance from the agent B;
5. <*minγ*, *maxγ*> – the angular interval specifying the required AXB angle: the angle between vector X to A and X to B should be between *minγ* and *maxγ*;
6. *Heading* – preferred heading of the steered agent with regard to the agents A, B.

Regardless the initial positions, TS behaviours steer the agent X to be *minA* to *maxA* far from the agent A, *minB* to *maxB* far from the agent B, form with them an angle between *minγ* and *maxγ* and be correctly headed. The TS behaviour is autonomous and steers just one agent, without any need of communication with the other agents (it just needs to know their current locations). To achieve a good positioning of all the three agents, they all need to be steered by the TS behaviour.

The resulting force of the TS steering behaviour attracts the steered agent X to a target location, which fulfils all conditions given by parameters, if such a location exists. If not, supporting forces are used to steer the agent X to a satisfying location.

The TS parameters define a region R, where the agent X can be located. This region must fulfil three conditions: the distance of the agent X from the agent A must be in the interval <*minA*, *maxA*>, the distance of the agent X from the agent B must be in the interval <*minB*, *maxB*> and the angle between vector X to A and X to B must be between *minγ* and *maxγ*. The first two conditions have the shape of an annulus. The third condition has the shape of two crescents defined by two circles with the circumferential angle *minγ* and *maxγ* with two intersections: the locations of the agents A and B (see Fig. 2 Left).

Fig. 2 Right shows four possible configurations of all three conditions:

1. The region R is a coherent area.
2. The region R has two separated coherent areas with centroids lying inside them.
3. The region R has two separated coherent areas with centroids outside them.
4. No location fulfils all three conditions.

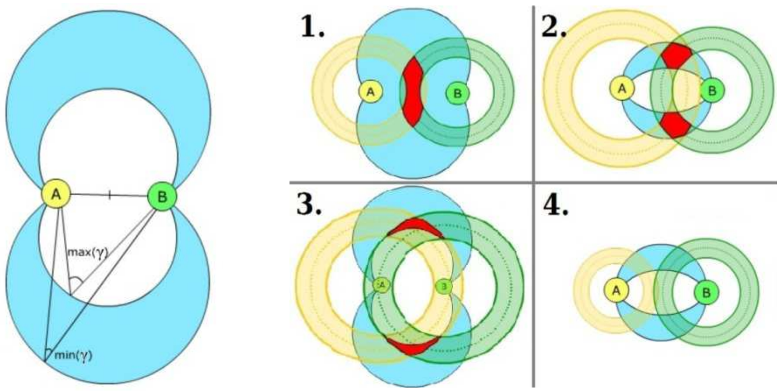

Fig. 2. Left: The shape of the region defined by the third condition (permitted angles). Right: Four possible configurations of region R (dark red colour).

Instead of solving 8 quadratic inequalities (which would be potentially slow), our algorithm computes all intersections of defining circles (56 points) and tests which of them fulfil all three conditions. The fulfilling points will be called border points. If there are some border points (which means that region R is not empty), the following algorithm is used to determine the target location:

1. If the centroid of the border points fulfils all three conditions, it will be the target location.
2. Else if the vector between location of the agent A and location of the agent B divides the region R in two separated regions R1 and R2 and their centroids fulfil all three conditions, the target location will be the centroid nearer to the agent X.
3. Else the target location will be the border point nearest to the agent X.

The resulting force of the TS behaviour consists of attractive force to the computed target location (if exists) and small attractive forces to/repulsive forces from the two other agents. These small forces lead to higher robustness and smoothness and help to solve situations when region R is empty. Apart from locomotion, the TA behaviour handles the heading of the agent X according to the parameter Heading.

One of the main advantages of this algorithm is its speed, as with others steering behaviours. According to our observations, it appears that the TS behaviour is able to navigate agents to their desired positions and generates good behaviour, even if the region R is empty. The TS behaviour also belongs to the group of steering behaviours with a certain social purpose (as well as the Walk Along steering behaviour [14]), since it can be used to express relations between agents.

5 Evaluation

We now present results of the initial exploratory evaluation of our system. The main evaluation question is whether the model's outcome is understandable by human observers. Because there is a solid evidence from the field of multimedia learning [7] that focusing on visual representations plus reading a text is cognitively demanding, our second question is how cognitively demanding is comprehension of emoticons. Our hypotheses are: 1) modelled virtual quarrels are comprehensible in general; 2) modelled virtual quarrels are more comprehensible to participants with 3D graphics experience than participants without such an experience; 3) interpreting emoticons is harder for participants without 3D graphics experience.

For the study's purpose, three different model outcomes were recorded, plus a fourth control video that was intentionally nonsensical.[2] An on-line questionnaire with 35 questions was constructed. Of present interest are three questions for each video asking a participant: a) to judge if the situation depicted seems natural or artificial; b) to what extent the quarrel looks believable; c) if he/she thinks he/she understands what has happened on the video. These were Likert items with 7 point Likert scale, with "1" or "7", respectively, meaning a) "very natural, like in a good silent movie" – "very artificial," b) "very believable; that could really happen" – "very unbelievable" c) "definitely yes, like right next to me" – "definitely no." The purpose of the control video actually was to ground the upper end of the scale. An additional Likert 7 question asked "how much of your mental capacity took you focusing on emoticons," with "very little" – "very much" scale. Four or five (14 in total) multiple-choice questions for each of the three experimental videos asked participants what had really happened to verify whether they indeed understood the situation. Of these, a total *knowledge score* was computed (0 – 14 points). Finally, participants were asked about frequency of their playing 3D games and professional usage of 3D graphics (scale: "less than a year or never," "at least once a year," "at least once a month," "at least once a week"). The sum produced a *3D graphics experience score* (0 – 6 points).

We recruited 67 anonymous Czech and Slovak participants (m=44; f=23) with average age 26.1 (SD=6.64). The average 3D graphics experience score was 1.66 (SD=1.62). Participants were instructed to focus on how the quarrels unfold, abstracting from the fact that the virtual city is devoid of objects and other people.

The rating of the videos is given by Table 2. We see that all the experimental videos score better than the control video and that there is a significant or nearly significant correlation between a knowledge score for each video and the subjective comprehensibility questions (i.e., (c) questions). The total knowledge score is 10.1 (SD=1.7), suggesting that participants understood the videos well. We ran nine planned paired comparisons among the control video and other videos for all three questions (a), (b), and (c) (paired t-tests). All three videos scored highly significantly better along all the three axes (every $p < 10^{-10}$). Thus, we consider the Hypothesis 1 supported: the videos are indeed comprehensible to the audience, and also relatively believable and natural.

The average mental load due to emoticons is 4.06 (SD=1.63) and it is not significantly different from the middle point 4 (single sample t-test; p=0.75). Thus, it cannot be concluded that following emoticons is easy for the participants, which

---

[2] The videos are available at:
http://pogamut.cuni.cz/pogamut-devel/doku.php?id=subprojects:emohawk_virtual_argument
Note some emoticons used are meaningful in Czech and Slovak cultural context only.

accords with the theory of learning from multimedia [7]. Yet there is no significant correlation between the 3D graphics experience score and the values of the question on emoticons' mental load (Pearson's rho=0.13; p=0.29). There is also only a mild trend concerning correlation between the 3D graphics experience score and the knowledge score of video content understanding (Pearson's rho=0.21; p=0.08). Thus, Hypothesis 2 is not supported by the data and Hypothesis 3 is only weakly supported. This is a surprising, though positive outcome.

We are now extending the evaluation along two axes: first, we investigate whether user participation is engaging compared to just watching generated videos, second, we investigate if explicit symbolic depiction of the agents' emotional states increase comprehensibility of the outcomes. Preliminary results indicates that user participation increases engagement and depiction of emotions weakly increases comprehensibility, but these data will be presented elsewhere in future.

Table 2. Average video scores and their standard deviations. The last row presents correlations (Pearson's rho) between subjective comprehensibility of videos and total knowledge scores from multiple-choice questions concerning each video.

	Video 1	**Video 2**	**Video 3**	**Contr. video**
natural?	3.9 (1.0)	2.7 (1.6)	3.1 (1.5)	5.6 (1.3)
believable?	3.0 (1.2)	2.2 (1.7)	3.2 (1.7)	5.2 (1.2)
comprehensible?	2.1 (1.0)	1.2 (.73)	2.8 (1.1)	4.6 (1.2)
normalized kn. score	.73 (.19)	.75 (.21)	.68 (.20)	
kn. score vs. compreh.?	rho=.28, p=.02	rho=.31, p=.01	rho=.23, p=.06	

6 Discussion and Conclusion

In this paper, we have presented a new hybrid model for controlling three agents engaged in a complex social interaction, during which they dynamically change their positions and perform actions beyond walking, turning, talking and gesturing. The model was implemented on UE2 agents participating in a vigorous quarrel. The key idea of the model was layering a hierarchical finite-state machine controller on the top of steering behaviours where the hFSM was specified based on annotations of video-recordings of actors improvising on the topic of the situation being modelled.

In general, we are relatively satisfied with the model we produced and the way we produced it but improvements are certainly possible. On the positive side, the method of constructing the behaviour model by modelling the situation using improvising actors, then manually annotating the resulting videos and, again manually, detecting common behavioural patterns turned out to be productive (the recording, annotating and StoryFactory scripting took less than four days). We believe the method can be used for other similar projects, though an open question is how well it would scale for larger groups. We are also pleased that the idea behind the model's architecture of controlling the steering layer by the action selection layer, including switching/altering steering behaviours in real-time, resulted in a prototype that swiftly and smoothly generates behaviour understandable by the target audience (as demonstrated by our evaluation). However, note that one minibeat, by definition, employs a single steering behaviour, which is set at the minibeat's start. It turned out that, occasionally, it would be an advantage to postpone turning on of a steering behaviour for one of the agents for a while (e.g., two characters go closer together and the third character starts to follow them half a second later). At the same time, explicit support for transition animations, i.e., animations between minibeats, would likely contribute to increased believability. Presently, transitions between minibeats are sometimes visible as interruptions to the quarrel's flow. Finally, we are pleased that the triangular steering behaviour leads agents to the goal positions most of the time.

On the more negative side, the triangular steering tends to be fragile in some situations: the agent's trajectories from their starting positions to the target positions are not always as one would wish (see, e.g., the boy on Video 3, at 0:15). Furthermore, the mechanism has 18 parameters and while it offers autonomy to individual agents, it is quite possible that having a steering behaviour that would control the agents centrally would not only be simpler from the design perspective, but would also ameliorate some unnatural twists in agents' trajectories.

Because this work is a prototype, there are also some minor technical limitations: due to the virtual environment constraints, we cannot use blending of walking animations with torso/head animations, which would increase believability of the scenario and also give the designer more freedom. The control mechanism also works on 4 Hz due to technical reasons: that also produce some artificialities in the resulting behaviour, mostly in exact positioning of the agents. The emoticons sometimes occlude each other and some situations might be expressed with better images: that might decrease relatively high cognitive load on processing emoticons reported by many participants. Also controlling a camera was out of our focus. In the prototype, the user can "fly" over the scene freely.

Concerning scaling, the most pressing issue, and our future work, is testing the quarrel model in a setting featuring agents passing by. While the triangular steering behaviour can be combined with, e.g., obstacle avoidance behaviour, it is less clear how to specify complex actions in relation to nearby agents, including transitions into/out of minibeats in case the quarrel should be interrupted.

Finally, the somewhat naive Markov model approach to representing the possible unfolding of the quarrel leads inevitably to some strange quarrel instances, e.g., with repetitive behaviours. An option is to plan the course of the quarrel in advance based on "aesthetic" constraints and then possibly re-plan in real-time (with minibeats corresponding to planning operators). State-of-the-art narrative generators employing planning can generate stories of the complexity of the quarrel, e.g., [15] but it is not clear to us whether they can do it rapidly enough, say, below 100 ms.

Acknowledgments. This research was supported by the project P103/10/1287 (GACR), by a student grant GA UK No. 0449/2010/A-INF/MFF, by student grant GA UK No. 655012 and partially supported by SVV project number 265 314.

References

1. Bida, M., Brom, C., Popelova, M., Kadlec, R.: StoryFactory–A Tool for Scripting Machinimas in Unreal Engine 2 and UDK. In: Proceeding of ICIDS 2011, LNCS, vol. 7069, pp. 334–337, Springer, Heidelberg (2011).
2. Damian, I., Endrass, B., Huber, P., Bee, N., André, E.: Individualized Agent Interactions. In: Allbeck J., Faloutsos P.: Motion in Games, LNCS, vol. 7060, pp. 15–26, Springer, Heidelberg (2011).
3. Gemrot, J., Kadlec, R., Bida, M., Burkert, O., Pibil, R., Havlicek, J., Zemcak, L., Simlovic, J., Vansa, R., Stolba, M., Plch, T., Brom C.: Pogamut 3 Can Assist Developers in Building AI (Not Only) for Their Videogame Agents. In: Agents for Games and Simulations, LNCS 5920, pp. 1–15, Springer, Heidelberg (2009). http://pogamut.cuni.cz (6.7.2012).
4. Jan, D., Traum, D. R.: Dynamic movement and positioning of embodied agents in multiparty conversations. In: Proceedings of the Workshop on Embodied Language Processing, pp. 59–66 (2007).
5. Karamouzas, I., Overmars, M.: Simulating the Local Behaviour of Small Pedestrian Groups. In Proc. Of the 17th VRST, pp. 183–190 (2011).
6. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press (1990).
7. Mayer, R. E.: Multimedia learning. New York: Cambridge University Press (2001).
8. Mateas, M.: Interactive Drama, Art and Artificial Intelligence. PhD thesis. Department of Computer Science, Carnegie Mellon University (2002).
9. Narain, R., Golas, A.,Curtis, S, Lin, M. C.: Aggregate dynamics for dense crowd simulation. In ACM SIGGRAPH Asia 2009 papers, pp. 122:1–122:8 (2009).
10. O'Neill, B., Piplica, A., Fuller, D., Magerko, B.: A Knowledge-Based Framework for the Collaborative Improvisation of Scene Introductions. In: Proc. ICIDS 2011, LNCS, vol. 7069, pp. 85–96 (2011).
11. Orkin, J, Roy, D.: Automatic learning and generation of social behavior from collective human gameplay. In Proceedings of The 8th International Conference on Autonomous Agents and Multiagent Systems, vol. 1, pp. 385–392 (2009).
12. Pedica, C., Vilhjálmsson, H.: Spontaneous avatar behavior for human territoriality. In: Applied Artificial Intelligence, vol. 24, pp. 575–593, (2010).
13. Peinado, F., Cavazza, M., Pizzi, D.: Revisiting Character-Based Affective Storytelling under a Narrative BDI Framework. In: Proc. ICIDS 2008, LNCS, vol. 5334, pp. 83–88 (2008).
14. Popelová, M., Bída, M., Brom, C., Gemrot, J., Tomek, J.: When a Couple Goes Together: Walk along Steering. In: Allbeck J., Faloutsos P.: Motion in Games, LNCS, vol. 7060, pp. 278–289, Springer, Heidelberg (2011).
15. Porteous, J., Cavazza, M., Charles, F.: Applying Planning to Interactive Storytelling: Narrative Control using State Constraints. In: ACM TIST, 1(2), pp. 1–21 (2010).
16. Reynolds, C.: Steering behaviors for autonomous characters. In: GDC, pp. 763–782 (1999).
17. Ricks, B. C., Egbert, P. K.: More realistic, flexible, and expressive social crowds using transactional analysis. In: The Visual Computer, vol. 28, issue 6-8, pp. 889–898, Springer-Verlag (2012).
18. Thalmann, D.: Crowd Simulation. In: John Wiley & Sons, Inc.: Wiley Encyclopedia of Computer Science and Engineering, (2007).