In Albeck et al. (Eds.): Proceedings of the 10th International Conference on Intelligent Virtual Agents, September 20-22, Philadelphia, PA, 2010

Avatars in Conversation: The Importance of Simulating Territorial Behavior

Claudio Pedica, Hannes Högni Vilhjálmsson, and Marta Lárusdóttir

Center for Analysis and Design of Intelligent Agents, School of Computer Science, Reykjavik University, Menntavegur 1, IS-101 Reykjavik, Iceland

Abstract. Is it important to model human social territoriality in simulated conversations? Here we address this question by evaluating the believability of avatars' virtual conversations powered by our simulation of territorial behaviors. Participants were asked to watch two videos and answer survey questions to compare them. The videos showed the same scene with and without our technology. The results support the hypothesis that simulating territorial behaviors can increase believability. Furthermore, there is evidence that our simulation of small scale group dynamics for conversational territories is a step in the right direction, even though there is still a large margin for improvement.

1 Introduction and Background

This paper reports the results of a survey we conducted to evaluate the believability of our simulated avatar behavior for conversations. Our approach is strongly influenced by the theories of Human Territories [1] and F-formations [2], and it has been described in two previous publications [3, 4]. It models the group dynamics of positions and orientations as a result of a special class of behaviors, conventionally called territorial behaviors, that we believe are essential for a complete simulation of a believable conversation.

Most of the work on automating avatar conversations focuses on the generation of communicative behaviors after the conversation has already started. The assumption is that the avatars are already in the right location and correctly oriented for engaging each other. However, we developed a method to let the avatars autonomously cluster and arrange themselves when a conversation takes place. Our approach generates an emergent dynamic arrangement as the result of the behavioral constraints suggested by the territorial field of the conversation. As a result, the avatars dynamically react to each other's position and orientation in a given social context. The purpose of the study is to investigate whether our approach, based on human territoriality, improves the believability of few avatars having a conversation, compared to the state of the art in avatar animation, that still relies on user control for moving them or at most arranges them into fixed formations such as a circle. A study by Jan and Traum [5] reports how the wrong positioning of virtual characters in a conversation significantly reduced the believability of the simulated social interaction. The finding lead to their model of small scale group dynamics for conversations [6], the first to keep proper positioning but not

proper orientations. So far, an evaluation of believability of that model has not been conducted.

Believability is hard to define. It is a construct and a hypothetical variable that cannot be measured directly. Therefore, we have chosen four variables we believe relate to believability when we evaluate simulated social interactions. Having better scores on these variables will increase the overall believability of the scene. The four variables we identified are: *artificiality*, *appeal*, *social richness* and *avatar awareness*. To evaluate our technology we built a questionnaire around these measurable quantities. Our dynamic avatars' conversations competed against a static version of them where avatars were chatting but not rearranging according to important events. Notice that while a static arrangement would be plausible in certain social situations, a random collective movement of the participants would not be. As confirmed by other studies on gaze behavior, an algorithmically generated gaze shift has proven more effective than just a randomized motion [7, 8]. Synchronization and timing are important for simulating social behaviors and testing against random motion would not have been informative.

2 The Survey

We conducted the survey by means of a web questionnaire[1]. The survey was open to anyone interested in it and we sent an invitation email to students, teachers and staff members of Reykjavik University, reaching about 5000 potential target respondents. Anonymity and confidentiality were assured. 171 people responded.

The only information about the participants we were interested in, apart from typical demographic information, was how often they played video games in the last three months. We wanted to discriminate between those accustomed to state-of-the-art character animation and those who were not.

The respondents were asked to answer a questionnaire organized into four test cases. Each case was focused on one of the following important situations: joining a conversation, moving within a conversation, avoiding a conversation and passing-by a conversation. For each case, there were two videos to watch and some questions to answer. At the end, we asked the subjects to answer some extra questions about the survey itself to learn more about the quality of the survey procedure.

2.1 Subjects

The mean age was 31 years with the majority of the subjects between 20 and 40 years old. 35% of the subjects hadn't played video games in the last three months, 32% played once a week (*casual players*), 21% played four times a week (*habitual players*) and 12% played every day (*hardcore players*). We classified the people who don't play and the casual players as *non-gamers* (113) while habitual and hardcore players were classified as *gamers* (57).

[1] http://populus.cadia.ru.is/survey

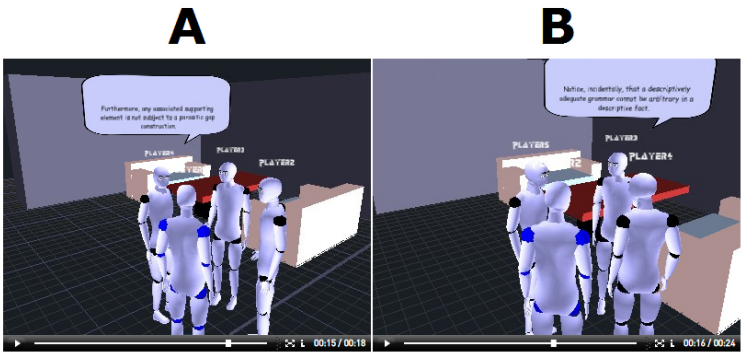

Fig. 1. Screenshot from the *joining* test case (Here A is dynamic and B static).

TEST CASE	QUESTIONS
Joining	1. Which scene do you find more artificial? 2. Which scene do you find more appealing? 3. Which scene do you find more socially rich? 4. In which scene does the group appear more disconnected from the situation? 5. In which scene does the group seem more aware of the new person joining?
Moving within	1. Which scene do you find more artificial? 2. Which scene do you find more appealing? 3. Which scene do you find more socially rich? 4. In which scene do the others appear more disconnected from the situation? 5. In which scene do the others seem more aware of the blue person moving around?
Avoiding	1. Which scene do you find more artificial? 2. Which scene do you find more appealing? 3. In which scene does the blue person's behaviour seem more appropriate?
Passing-by	1. Which scene do you find more artificial? 2. Which scene do you find more appealing? 3. Which scene do you find more socially rich? 4. In which scene does the group seem more aware of the blue person's existence? 5. In which scene does the group appear less involved in their conversation?

Table 1. The set of questions for each of the four test cases.

2.2 Questionnaire

For each of the four test cases, the subjects were asked to watch two videos of the same scene, with and without our technology (Fig. 1), and answer a set of simple questions (Table 1) about the artificiality or appeal of those scenes. The questions made the subjects compare video A and B by evaluating a statement about what they saw and associating it with one value of the following rating scale: A - much more, A - more, A - slightly more, the same, B - slightly more, B - more and B - much more.

To avoid any trivial differentiation, each pair of videos was recorded from the same system, with the same camera angle, running the same animated simulation of turn taking to let the conversation appear "alive". The difference was that in one instance our group dynamics were disabled and enabled in the other. The order of the two scenes was alternated to avoid a systematic order bias.

3 Results

The results of the survey are shown in Fig. 2. The graphs are vertically grouped by test case and horizontally by question. The horizontal axis within each graph denotes the preference for either the *Static* version (numbers lower than 4) or the *Dynamic* version (numbers higher than 4).

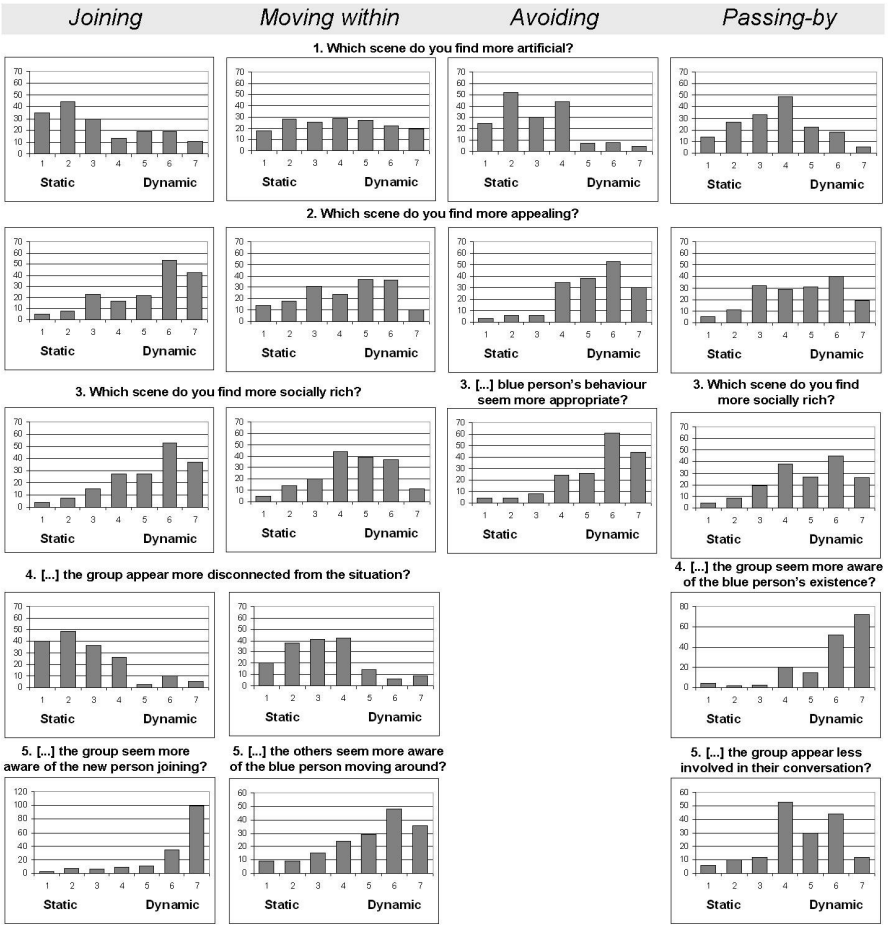

Fig. 2. Results of the survey.

Table 2 shows mean value and standard deviation of the answers for each question. The numbers in bold indicate where our technology performed effectively better. For those mean values the effect size was "medium" or "large" (Cohens's $|d| > 0.45$).

	Joining					*Moving within*					*Avoiding*			*Passing-by*				
Question:	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**4**	**5**	**1**	**2**	**3**	**1**	**2**	**3**	**4**	**5**
Mean:	**3.22**	**5.18**	**5.19**	**2.72**	**6.05**	3.95	4.18	4.49	**3.27**	**5.02**	**3.00**	**5.21**	**5.47**	3.67	4.59	**4.87**	**5.88**	**4.62**
Std dev:	1.90	1.68	1.55	1.55	1.51	1.87	1.71	1.47	1.54	1.72	1.49	1.39	1.44	1.52	1.60	1.56	1.39	1.46

Table 2. Mean values and standard deviations. Bold numbers indicate where our technology performed better (Cohen's estimated effect size $|d| > 0.45$)

Our results are statistically significant against our testing hypothesis, for which we had close to zero p-values for all the questions, except for question n. 1 ($p = 0.37$) and n. 2 ($p = 0.08$) in the second test case. Some of the negative comments on those questions were about how both scenes were unlikely to happen in real life. One of the subjects commented as follow:

> "Yes, in the second test, the blue person moved to the side [in] a way I haven´t seen anyone move. It look the most fake of all off the test, both A and B."

3.1 Quality of the Survey

On the last page of the questionnaire, the respondents answered some questions about the survey itself so that we could verify its quality. The results are shown in Fig. 3.

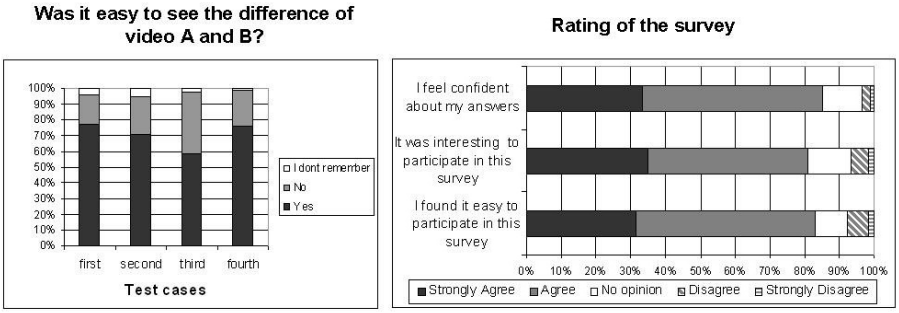

Fig. 3. Results from questions about the survey itself indicate adequate quality.

4 Discussion

The results confirm that simulating human territories significantly increases believability, by reducing artificiality and increasing appeal, social richness and apparent avatar awareness in simulated conversations.

For two out of four tests, the scene powered by our technology was judged significantly less artificial, more appealing, more socially rich and with a higher avatar awareness. Social richness was also slightly higher in the second and fourth tests, although the effective sizes are "medium/small", or 0.33 and 0.37 respectively.

In the second test case both the static and dynamic videos were judged equally artificial and appealing. This is a clear limitation of our small group dynamics model that needs to be improved. The group was too responsive to the avatar's movement and some subjects felt this was unnatural. Another reason was that the group went along with the avatar moving, giving the feeling he was followed by the others:

> "For the second scene, it is clear that the moving group is too artificial: if a group is speaking and someone joins, it is the group that rules over the single person; this means that the person can move but the group should stands still."

The *gamers* were generally tougher on our technology than the *non-gamers* although the difference is only significant for the second test case (Cohen's effect size $0.46 < |d| < 0.49$). This further suggests we need to improve group dynamics for conversations.

4.1 Dynamics of Orientation

Avatar awareness was judged significantly higher with our technology in all test cases probably due to our simulation of gaze and body orientation. This indicates the importance of controlling orientation of body parts for correctly simulating conversation. Introducing orientational motivation to our model of group dynamics improved the believability of the overall scene, making the avatars look more aware of their surroundings and more connected to the social context. This is not surprising, considering that the orientation of some bodily regions normally express temporary membership to a group or a subgroup, or more generally our claim of territory as argued by Scheflen [1].

We believe that a proper simulation of body part orientation is essential for simulating, not only conversations, but social interactions in general. This provides clues and signals that retain their expressiveness when simulated. To correctly realize these orientations requires them to be incorporated into the model of group dynamics, but also to have a behavioral architecture where they can be easily controlled in a reliable way. Our technology provides both such a model and architecture.

4.2 Conclusions

Our model of group dynamics can be improved. This is not surprising considering that it was originally inspired by models of the dynamics in a flock of birds. Two fundamental concepts of the theory of Human Territories are still not included in our model, which may impact the simulation quality. They are the *transactional segment* [2] and *locations* [1]. The transactional segment is an amount of space necessary to carry on an interaction. A participant in a conversation will adjust his position and orientation

to keep his transactional segment intersected with the segment of the other participants. Locations are placeholder amounts of space that contribute to shaping the territory of an interaction. Each participant in a conversation holds a location but is allowed to temporarily leave it and then get back to it. In the meantime the territory will probably keep the same shape, without requiring a rearrangement of all the participants.

Territorial behaviors are important for the overall believability of a simulated avatar's conversation. They show participation in a social interaction and awareness of the context and therefore they are necessary for socially intelligent agents. Simulating this class of behaviors by reactive responses appears to be an approach worthy of pursuit.

Further investigation is necessary and we plan to evaluate our technology in a real game environment with high quality character models and animations. The higher visual quality may increase expectations of the character behavior and result in lower ratings. However, we may still find evidence of the validity of our approach and may even match it with a brand new model of social territorial dynamics.

Acknowledgments. We are grateful to Jina Lee, Brent Lance and Stacy Marsella for web survey ideas. Big thanks to the CADIA team, but special thanks to Ankell Logi for web work. This was supported by the Humanoid Agents in Social Game Environments Grant of Excellence from The Icelandic Research Fund.

References

1. Scheflen, A.E.: Human Territories: how we behave in space and time. Prentice-Hall, New York, NY, USA (1976)
2. Kendon, A.: Conducting Interaction: Patterns of behavior in focused encounters. Cambridge University Press, Cambridge (1990) Main Area (multimodal commnication).
3. Pedica, C., Vilhjálmsson, H.: Social perception and steering for online avatars. In: IVA '08: Proceedings of the 8th international conference on Intelligent Virtual Agents, Berlin, Heidelberg, Springer-Verlag (2008) 104–116
4. Pedica, C., Vilhjálmsson, H.H.: Spontaneous avatar behavior for human territoriality. In: IVA '09: Proceedings of the 9th International Conference on Intelligent Virtual Agents, Berlin, Heidelberg, Springer-Verlag (2009) 344–357
5. Jan, D., Traum, D.R.: Dialog simulation for background characters. Lecture Notes in Computer Science (2005) 65–74
6. Jan, D., Traum, D.: Dynamic movement and positioning of embodied agents in multiparty conversation. In: Proc. of the ACL Workshop on Embodied Language Processing. (June 2007 2007) 59–66
7. Garau, M., Slater, M., Bee, S., Sasse, M.A.: The impact of eye gaze on communication using humanoid avatars. In: CHI '01: Proceedings of the SIGCHI conference on Human factors in computing systems, New York, NY, USA, ACM (2001) 309–316
8. Vertegaal, R., Ding, Y.: Explaining effects of eye gaze on mediated group conversations:: amount or synchronization? In: CSCW '02: Proceedings of the 2002 ACM conference on Computer supported cooperative work, New York, NY, USA, ACM (2002) 41–48
9. Cassell, J., Vilhjalmsson, H.: Fully embodied conversational avatars: Making communicative behaviors autonomous. Autonomous Agents and Multi-Agent Systems **2**(1) (1999) 45–64