Spontaneous Avatar Behavior for Human Territoriality

Claudio Pedica[1,2] and Hannes Högni Vilhjálmsson[1]

[1] Center for Analysis and Design of Intelligent Agents, School of Computer Science, Reykjavik University, Iceland
[2] School of Computer Science, Camerino University, Italy

Abstract. The challenge of making a virtual world believable includes a requirement for AI entities which autonomously react to a dynamic environment. After the breakthroughs in believability introduced by modern lightning and physics techniques, the focus is shifting to better AI behavior sophistication. Avatars and agents in a realistic virtual environment must exhibit a certain degree of presence and awareness of the surroundings, reacting consistently to unexpected contingencies and social situations. Unconscious reactions serve as evidence of life, and can also signal social availability and spatial awareness to others. These behaviors get lost when avatar motion requires explicit user control. This paper presents a new approach for generating believable social behavior in avatars. The focus is on human territorial behaviors during social interactions, such as during conversations and gatherings. Driven by theories on human territoriality, we define a reactive framework which allows avatars group dynamics during social interaction. This approach gives us enough flexibility to model the territorial dynamics of social interactions as a set of social norms which constrain the avatar's reactivity by running a set of behaviors which blend together. The resulting social group behavior appears relatively robust, but perhaps more importantly, it starts to bring a new sense of relevance and continuity to virtual bodies that often get separated from the simulated social situation.

1 Introduction

Most of Multiplayer Massively Online Games (MMO) available nowadays portray their players as animated characters or avatars, under the user control. One of the open challenges for the state of the art in interactive character animation is to measure up to the standard of visual quality that the game industry reached with their environments. There are numerous examples of 3D rendered characters in films and digital media that look quite realistic when you see a screenshot or a still frame, but once they start moving, they look totally non-lifelike giving to the viewer a slight sense of discomfort. This feeling is explained in robotics as the "uncanny valley", a terminology first introduced by Masahiro Mori in 1970 [1]. Applying the "uncanny valley" hypothesis to an animated character, the conclusion is that the more a virtual creature looks realistic the more we will expect a realistic behavior when it is moving. For those companies which plan to create close to photorealistic characters, it must be ensured that their behavior matches the quality of the visual rendering. This is particularly true when they need to simulate human communicative behavior in face-to-face interactions, such as conversations.

Zs. Ruttkay et al. (Eds.): IVA 2009, LNAI 5773, pp. 344–357, 2009.

In most commercial avatar-based systems, the expression of communicative intent and social behavior relies on explicit user input [2]. For example, in both Second Life[1] and World of Warcraft[2] users can make their avatars emote by entering special emote commands into the chat window. This approach is fine for deliberate acts, but as was argued in [3], requiring the users to think about how to coordinate their virtual body every time they communicate or enter a conversation places on them the burden of too much micro-management. When people walk through a room full of other people, they are not used to thinking explicitly about their leg movements, body orientation, gaze direction, posture or gesture, because these are things that typically happen spontaneously without much conscious effort [4]. Some of these behaviors are continuous and would require very frequent input from the user to maintain, which may be difficult, especially when the user is engaged in other input activities such as typing a chat message. In the same way that avatars automatically animate walk cycles so that users won't have to worry about where to place their virtual feet, avatars should also provide the basic behavioral foundation for socialization.

Interestingly, even though users of online social environments like Second Life appear sensitive to proximity by choosing certain initial distances from each other, they rarely move when approached or interacted with, but rely instead on the chat channel for engagement [5]. Since locomotion, positioning and social orientation is not being naturally integrated into the interaction when relying on explicit control, it is worth exploring its automation. For the particular case of a conversation, some have suggested that once an avatar engages another in such a face-to-face interaction, a fixed circular formation should be assumed [6]. Even though the idea matches with our common sense and daily experience, it is in fact a simplification. A conversation is indeed a formation but a more dynamic one. The circle we often see is merely an emergent property of a complex space negotiation process, and therefore the reliance on a fixed structure could prevent the avatars from arranging themselves in more organic and natural ways, with the net result of breaking the illusion of believability.

2 Related Works

2.1 Automating Avatar Control

As we summarized in [7] automating the generation of communicative behaviors in avatars was first proposed in BodyChat [3] and then further explored in the Spark system, that incorporates the BEAT engine [8] to automate a range of discourse related co-verbal cues in addition to cues for multi-party interaction management. With a special interest for postural shifts, the Demeanor system [9] generates avatar posture based on affinity between conversation partners. A study [2] showed how automating avatars' communicative behaviors provides some level of support to their users.

In this current work we focus on the yet "unconquered land" of simulating small scale group dynamics, while keeping in mind that other layers of behavior control introduced in previous work will need to be added for fully supporting the social interaction process.

[1] http://secondlife.com/

[2] http://www.worldofwarcraft.com/

2.2 Simulating Group Dynamics

Simulating group dynamics concerns with modelling and imitating the kinetic evolution over time of a group of individuals. This is different from another plausible definition of group dynamics which concerns how the social relationships evolves amongst the members of a group, like for example when two persons become friends. In the domain of our definition we can talk about large scale group dynamics and small scale group dynamics, and the difference between them is just in the order of magnitude of the number of individuals we consider. For our purposes we are more interested in the second kind of dynamics even though the scientific community has been much more prolific in dealing with large groups. Of course, simulating large scale groups is different from simulating small scale groups but the approaches used for modelling the former can be adapted for the latter. Numerous works have been published in the area of large scale group dynamics. Most of them simulate natural systems like crowds of people or formations of animals such as flocks of birds or schools of fish. These sort of global collective phenomena have been modeled with different approaches but the most interesting and successful of them define the group dynamics as an emergent behavior. In this direction, there are two main approaches to the problem:

- The particle-based system approach, where particles are animated in real time by application of forces.
- The agent-based systems approach, in which each agents are managed in real time by rules of behavior.

The main difference between them is how much sophistication we want for the behavior of each single individual. The first approach focuses more on the collective group behavior as a whole whereas the second focuses of the richness of behavior of each single individual.

Most of the Crowd Simulators use a particle-based approach because it is well suited for modeling global collective phenomena (such as group displacement and collective events) where the number of individuals is huge and they are all quasi-similar objects. Usually each individual is not doing more than just moving toward a destination, therefore its motion is easily modeled as a particle. One of the classical work on particle-based systems is the one of Helbing and Molnár [10] which clearly describes the concept of a social force model for simulating dynamics of walking pedestrians. Social forces are defined as a psychological tension toward acting in a certain way. Quoting the authors, a social force is a "[...] quantity that describes the concrete motivation to act". An interesting extension to the basic social force model has been introduced by Couzin et al. [11]. They define three concentric proximity zones around a particle where each zone exerts a different prioritized force on the particle's constant velocity. In the work of Pelechano et al. [12] the social force model is taken a step further with the introduction of line formation and psychological factors which induce a pushing behavior in panicking situations.

A social force model is not the only way of modelling crowds of people. Heigas et al. [13] use a model of fluid dynamics that incorporates two elementary repulsive forces to simulate jamming and flowing found in real life situations. Besides the specificity of the solution, as argued in [14] fluid dynamics not always can correctly model individual interactions which can be better described with a gaskinetic model. Another different approach is the one of Treuille et al. [15] which presents a model for crowd dynamics continuously driven by a potential field. The integration of global navigation planning and local collision avoidance into one framework produces very good video results. Furthermore, their model seems to integrate well with several agent-based models promoting interesting future integrations. Yet another approach consists of recording a crowd and then directly replicating the phenomenon using the recorded data. This is for example the approach taken by Lee et al. [16]. In this work the authors have been recording a crowd of people from a top down view and then used computer vision to extract motion trajectories out of it. Afterwards the data are fed into an agent model which can learn how to replicate the crowd motion driving each individual. Interestingly, the work also addresses some small group management but individuals are not aware of their social context and do not react to unexpected contingencies unless the centralized agent model has been trained specifically for that.

Thalmann et al. [17] use complex finite automata to determine the behavior of actors. The purpose of the model is still to simulate human crowds but this time the introduction of structured behavior of groups and individuals is remarkable. A hierarchical model describes the behavior of each part, but still the set of norms of social interactions such as conversations are not taken into account. Very comprehensive is also the work of Shao and Terzopoulos [18], which presents fully autonomous agents interacting in a virtually reconstructed Pennsylvania Station. The integration of motor, perceptual, behavioral, and cognitive components within a single model is particularly notable. Apart from the outstanding video realized with this technology, it is also very interesting to see how they use perceptual information to drive low-level reactive behaviors in a social environment. What is missing is how to constrain an agent's reactivity outside the special situation of walking down the hall of a train station. Rehm et al. [19] use a more fine-grained approach for conversations, recognizing the value of social proxemics and formation theories. They use them to inform their models of dynamic distance and orientation between pairs of humanoid agents based on their interpersonal relationship. While interpersonal relationships are necessary to fully simulate small scale group dynamics, they are not sufficient as is evident from Kendon's work [4].

In a pure agent-based approach the action generation loop produces a discrete sequence of behavior, that turns to be a downside. Human behavior is not discrete but rather continuous. One of the main advantage of the particle-based approach is in the continuity of the simulated behavior, which looks quite believable once animated. A sequential generation of behaviors is a discretization of the continuous process of perceiving and acting which take place at the lower levels of the artificial intelligence. Abstract high level behaviors can be decomposed in lower level, more fine grained, behaviors until they eventually merge into a continuum of behavioral control. For this reason we believe that the best approach for simulating social interaction territorial dynamics is to use a combination of the agent-based and particle-based approaches, where a set of reactive behaviors generates motivational forces which eventually are merged together in a continuous input control for the agent's motion generation layer. From this perspective, the work of Reynolds on the so called Steering Behaviors has been very helpful.

2.3 Reynolds Steering Behaviors

In 1999 Craig Reynolds [20] presented a simple and elegant solution, to allow simulated characters to navigate around their world in a life-like manner. These behaviors continuously generate steering forces to drive the underlying model of locomotion. So to speak they are slightly higher level than motor controllers and lower level than the actions selected by a planner for example. The advantage of using steering forces is that they can be easily combined before being applied to the agent's body. The composable nature of the steering behaviors allows building complex commands like 'go from here to there following a given path, while avoiding obstacles along the way without getting too close to the walls and maintaining a minimal distance from your neighbors'. When multiple commands are called at the same time, they generate steering forces which have to be blended. One of the simplest blending schema consists of linearly combining the steering vectors but, unfortunately, suffers of a common problem called cancellation. Basically, it could happen that the combination results in a null steering force. Amor et al. [21] recognize how sometimes a combination of behaviors result in a suboptimal steering force with the net result of having the agent behaving not in such an intelligent way, for example trying to avoid an incoming obstacle and ending up bumping into a wall. Instead of steering the agent toward a desired velocity, they invert the process by assigning a cost to a set of sampled directions. Afterwards they perform a heuristic to select the cheapest direction to follow. So far it is not clear how many sampled directions are necessary to create good behaviors while keeping the computational costs low. Some behaviors may require a large span of directions but with a very low sample resolution, while others require exactly the converse.

2.4 Small Scale Group Dynamics

The pioneering work of Jan et al. [22], for the first time exploits some of the techniques used in the field of Crowd Simulators to replicate small scale group dynamics with a special interest in conversations. In their work, the authors recognize the importance of managing the correct positioning and orientation of agents in conversation to avoid breaking the fragile illusion that makes a social virtual environment believable. They report an evaluation made in one of their previous works [23] where agents could group together and move from a group to another one but always maintaining a fixed position, and quoting the authors "[...], this significantly decreased believability when conversation groups did not coincide with positioning of the agents". To solve this problem, they took the social force field model idea from the Crowd Simulators literature and applied it to dynamically rearranging a group of agents engaged in a situated conversation inside a virtual training environment. While the approach looks promising, the main problem is that motivational forces applied to agent orientation are not taken into consideration. Reorienting is also part of the behavior expressed by people during social interactions. Moreover as we know from Scheflen [24] the orientation of some bodily regions normally express temporary membership to a group or a subgroup, or more generally our claim of territory. Such claims should be maintained as long as a member attends to that social interaction. Therefore it is important to extend the social force field model in such a way that reorientations can be also motivated. Furthermore, we should remember that a conversation is a unit at interactional level and has a territorial domain [4] and therefore we can think of it as an abstract social context but also as situated region of space. The conversation's spatial domain delimits a region which casts a behavioral influence not only on the participants but also on external individuals who stop close or just pass by [4].

Since this behavioral influence is common to other forms of territory [24] as well, we could use the name *social place* to refer to the spatial domain of a social interaction and the name *social situation* to refer to its abstract social context. Thus, a complete model of small scale group dynamics should also take into account the behavioral influence that any social situation casts in the environment, concretely bounded by its social place.

3 Conversations and Human Territories

A conversation is an example of human territorial organization. The space around it is structured in a certain fashion and specific behaviors take place demonstrating that, not only the same idea about the social context is shared amongst the participants, but also the same idea about the territorial domain of such context. These kinds of behaviors have been classified by Scheflen [24] as territorial behaviors. They don't have to be considered in isolation but rather as a particular way of looking at the behavioral relationship amongst the participants in a social interaction and as such, they are influenced by affiliation, involvement and social status. An automation of human communicative intent must take into account such a special class of behaviors in order to properly simulate individuals engaged in a social interaction, especially when we account not only for conversations but also for a variety of different social situations each of which could have is own territorial form. Territorial behaviors are intrinsically unconscious and reactive, therefore they as well are ill suited for explicit user control and must be automated.

Like some of the more interesting previous works, our primary inspiration has been Kendon's research on face-to-face interaction [4]. Individuals in a conversation tend to arrange in a way that gives all of them equal, direct and exclusive access to a common space. This positional and orientational arrangement is called an F-formation and the set of the behavioral relationships among the participants defines a behavioral system called the F-formation system (Fig. 1). From the definition of the F-formation comes an explanation of the usual circular arrangement of people in conversation with more than two participants: it is simply the best way to give everybody equal access to a common focused space. Since participants in a focused social interaction share a common space with equal access rights to it, a series of compensatory movements has to be simulated before one can hope to completely model social group dynamics.

Kendon explains a connection between the F-formation system and Goffman's concept of a frame [25]. A frame is a set of rules and social norms that all the participants in an interaction silently accept. The frame comes from the experience of the individual and states what behaviors are meaningful and what conduct is expected in that particular face-to-face interaction. The process of frame-attunement is tightly linked to the F-formation system. By actively maintaining a formation, participants inform each other that they share the same frame for the situation. This further reinforces the definition of an F-formation as a system and moreover describes the system as a unit of behavior at the interactional level of organization, not at the individual level (Fig. 1).

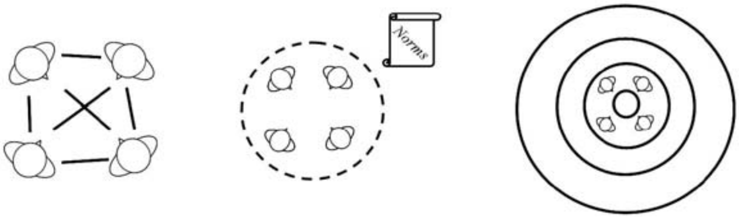

Fig. 1. A schematic representation of the three main concepts of a conversation as an instance of F-formation. The first picture shows how the participants are involved in a behavioral relationship which sustains a stable formation. The second picture shows a conversation as a unit at interactional level which states a set of norms silently accepted by its participants. The third picture shows how these set of norms organize a conversational human territory.

Scheflen [24] further proposes a general paradigm of human territorial organization (Fig. 2). He proposes a central space called *nucleus*, which comprises a common orientational space and a space of the participants, surrounded by a *region* which is commonly used as a buffer area for potential newcomers or as a passageway for passersby. Such general structure of space is applicable to all levels of territorial organization and frames the whole area in six concentric zones. The first three are called o, p and q spaces and they belong to the nucleus of the territory, while the remaining zones are called r, s and t spaces and they belong to the region or court. For small territories such as conversations or gatherings in a living room, the s and t spaces are irrelevant and the r and q spaces can be merged. Therefore, for simple and small formations the whole region around the nucleus can be unstructured. Notice that the concentric spaces define zones of progressive growing status, starting from the outermost and moving toward the innermost. In fact the region is meant for passersby, spectators and associated people while the nucleus is for the participants that get direct access to the o-space and have a claim on the territory.

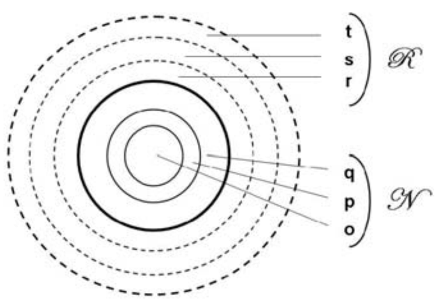

Fig. 2. The paradigm of territorial organization proposed by Scheflen and applicable to all levels of territorial organization. The o, p and q spaces belong to the nucleus N whereas the r, s and t spaces belong to the region R.

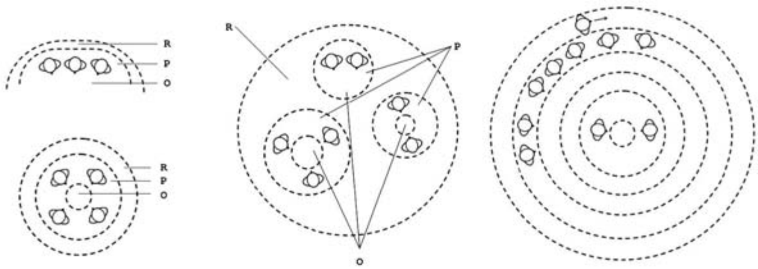

Fig. 3. From left to right, element, F-formation, gathering and hub. These are the main territorial fields of Scheflen's classification. Each of them can be seen as increment of the previous one for required space and complexity. In addition, more complex territory can contain simpler ones. The O and P in the picture stand for o and p spaces whereas the R stands for an unstructured region.

The classification of the territorial organization of small groups goes from the *element* to the *hub* in order of size and complexity, where the latter can get quite big in some situations (Fig. 3). An *element* is an array of people sufficiently close to each other and commonly oriented. Usually people arrange in adjacent locations but sometimes they also crowd in a single location. Participants in an element show a certain degree of affiliation because they are engaged in a common activity or involved in a special relationship. Examples of elements include a couple in a normal conversation, a queue of people waiting in line or a row of persons walking down a street. The next kind of simple and small territorial organization is the face formation, or *F-formation*, which has been extensively covered above and elsewhere. Here we want to point out that the region of an F-formation comprises the q and the r space in the paradigm of territorial organization (Fig. 2). The q-space is meant for salutations when joining and leaving the conversation or just as a buffer area for people who may want to join. The r-space could be used as a passageway by the participants themselves or for less close salutations. What in Kendon's work is called the c-space of a conversation is in fact an s-space in the Scheflen's territorial paradigm meant mainly for spectators and outsiders, conversely to the innermost q and r spaces where, whoever stops by, gets the higher status of being associated with the conversation. When elements and F-formations combine we have a more complex territorial organization called the gathering.

A *gathering* generalizes social situations like a group of people chilling out in a living room or at a party. Participants in a gathering do not share the same orientational spaces but rather there are many o-spaces sustained by several small groups. Indeed, a gathering is a collection of elements and F-formations which share a common region (Fig. 3). Another way of looking at it is considering the gathering as an increment of the F-formation. As such, we can have gatherings that naturally evolve from an F-formation which splits into more subgroups due, for example, to a higher number of participants. Notice that a gathering can also be just a collection of individuals clustered together a in closed space but not affiliated in any way. An example would be a waiting room where people do not interact. Usually a gathering consists of less than a dozen of people and takes the space of a room. Thus, in a bar situation we could find multiple gatherings, one for each room or floor, or multiple gatherings in the same room when several social contexts take place. The largest territorial field in terms of size and complexity is called *hub*. The hub is an increment of the gathering where the inner and outer zones are differentiated by role and status of the individuals. Unlike a gathering, the hub has a nucleus of people which perform for the spectators in the region. Thus, people in the region orient toward the nucleus while the performers can orient everywhere. The nuclear formation can be a single individual, an element, an F-formation or a gathering. Examples of hubs are a crowded theater or a cluster of persons watching at a street performer. The region of the hub, which is called surround, is usually structured in two levels of spectators plus an extra zone which is used as a passageway by passersby or people who simply want to join and attend the performance (Fig. 3).

4 Reaction Generation Paradigm

In our approach, the group dynamics of a simulated social interaction emerges from the avatars' territorial behavior. We have chosen to simulate such class of behavior as an avatar's reactive response to the environment and social context. The term reactive response should be clearly distinguished from other agent-based solutions where the agent goes through a higher level cognition process which involves some reasoning about its internal state and the state of the environment, to come up with a plan or a strategy to reach a given goal. There are fewer reasoning steps involved in our avatar's reactivity, which by definition should provide a quick response to changes in the environment, and therefore we can think of it as the simulation of a low level mental process much closer to perception than higher levels of reasoning. Thus in our reaction generation paradigm, low level perceptual information is analyzed by a set of reactive behaviors which motivate an immediate motion to accommodate contingent changes in the perceived surroundings (Fig. 4).

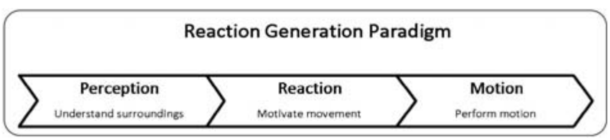

Fig. 4. Outline of our Reaction Generation Paradigm. The agent understands its surroundings through its senses. Successively, perceptual stimuli are used for generating motivations to move. Finally, motivations are transformed into physical forces which steer the agents motion.

The reaction paradigm is in effect a loop of continuous control of the avatar's motion. The surrounding environment stimulates the avatar's perceptual apparatus producing information which is later used by the reactive behaviors to generate motivations to move. Movement changes the state of the environment and therefore the set of perceptual information the avatar will perceive in the next round of the control loop.

At first, the avatar perceives its surroundings through a sense of vision and proximity both of which can be tailored to a specific individual (Fig. 5). The sense of proximity is a simple way of simulating the human awareness over the four distances of the Proxemics Theory [26]. A sensor structured in four concentric areas continuously informs the avatar about who or what is in the range of its intimate, personal, social or public zone. The public and the social zones cover a larger area of space which is also more distant from the avatar than the intimate and personal zones. Therefore we have two blinded cones for the public and social zones which extend from the avatar's back. For the sense of vision we have a peripheral and central visual area, where the former is larger and shorter whilst the latter is the converse. At the moment vision is at its early stage of design and we are planning to extend it with a more accurate model where one can specify a specialized perceptual ability for each area. For example, the peripheral vision should perceive movement better whereas central vision should be better at shapes and detail. These two senses continuously gather information about the avatar's surroundings, producing perceptual data which can be analyzed and used for generating reactions.

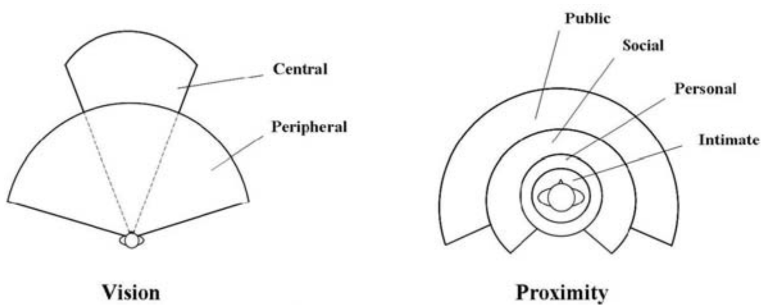

Fig. 5. Two diagrams showing the spatial structure of the sense of vision and proximity. Notice that proportions have not been respected to make the drawing more readable.

A set of reactive behaviors computes motivations for performing movement. For example, a *keep-personal-distance* behavior lets an avatar react by generating a motivation for moving away when other individuals come too close. Motivations are vectorial quantities that represent a psychological stimulation in performing a linear or rotational motion. The reactive framework permits applying motivations virtually to every part or joint of the avatar's body or to a higher level motor controllers such as a gaze controller. However, in our current implementation we limit our control to motivating the linear motion of the whole body and the rotational motion of eyes, head and rest of the body (torso and legs). When all the motivations have been generated, they are grouped per type and then blended into a composition which results in a net final motivation of each type. The combination of motivations allows multiple behaviors to stimulate the body at the same time. Several composition algorithms can be chosen for this step. In fact, motivations for linear motion usually need to be composed in a different way than motivation for rotational motion. After computing the set of final motivations, each of them is sent to the actuation module which performs a movement that respects the constraints imposed by the physical motion model.

5 Social Situation and Territorial Awareness

Now we are going to explain how, in our system, an avatar can show a certain degree of context awareness when engaged in a social interaction. To seem aware of the social context, a person has to show its acceptance of the norms that regulate the social interaction as we saw in [4] and [24]. Such norms state territorial rights and responsibilities of participants and outsiders and the acceptance of them makes the interaction possible. Thus, the attunement to such norms declares an intent of interaction and therefore the awareness of the social situation and its territorial organization. The spatial boundaries of a social situation, that we call social place, determine when the context should start influencing an avatar's behavior. In our system, such behavioral influence is realized by the activation of a set of reactive behaviors, each of which realizes a norm underlying the social situation that the avatar is participating in (Fig. 6). The activation of this set of behaviors produces the reactive dynamics expected from a group of people that has accepted the same social context.

In order to provide an example of how the behavioral influence works, we are going to succinctly describe how an avatar joins a conversation. As soon as an individual gets close enough to an ongoing conversation, it steps inside the conversation's territory. If it keeps moving closer to the nucleus, the individual receives an *associated* status. An associated person will be allowed to join a conversation if certain requirements are met. Since we assume that the conversation takes place amongst friends, the requirements are very loose. In fact it is sufficient to have the body oriented toward the o-space

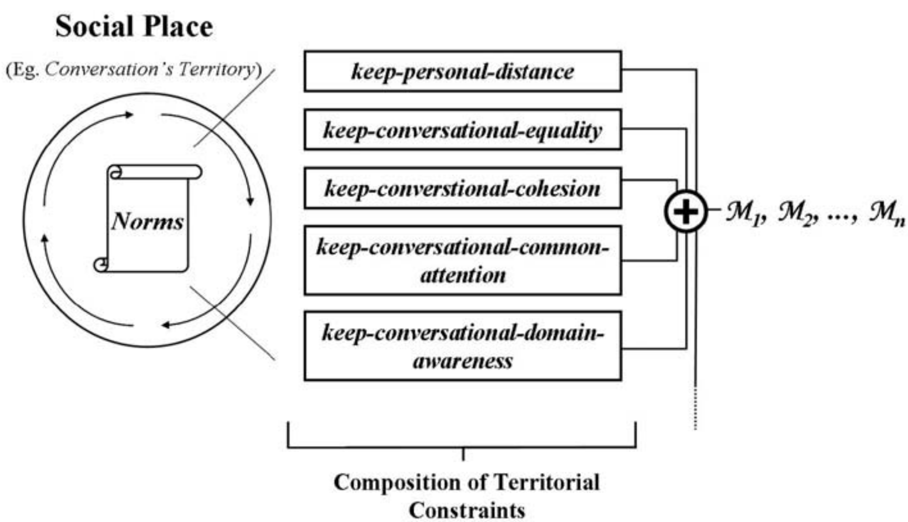

Fig. 6. A diagram to explain how social situation awareness is realized. Inside the border of the social place, the territorial organization states a set of norms that constraints the avatar's reactivity. Thus, the set of norms maps into a set of reactive behaviors that implements a composition of territorial constraints. The composition blends with other behaviors leading to a behavioral influence marked on the resulting final motivations $M_1, M_2, ..., M_n$.

and stop in front of it claiming access rights on the common space. Once an avatar is allowed to join, it is considered inside the conversation social situation, and therefore it is necessary to activate the proper set of territorial constraints in order to adapt the agent's behavior to the ongoing social situation and smoothly blend in. Conversely, an avatar can leave a conversation simply by going away from the nucleus. Moving out of the territory will stop the behavioral influence, releasing the avatar from its territorial constrains. This example is not meant to explain how an avatar should generally join or leave a conversation, but how the avatar's behavior is immediately and smoothly influenced by the simple fact that it enters or leaves a social place.

6 Conclusions and Future Work

The approach described here has been implemented in the CADIA Populus social simulation platform for virtual environments [7]. An important contribution is in the field of graphical avatars for games. With the advent of even more powerful 3D engines, faster and more robust physics engines and affordable dedicated hardware, games are rapidly growing as complex, fully dynamic and photorealistic worlds. Such rich and complex environments are now much more effective in providing the feeling of presence, immersion or intense atmosphere, than only few years ago. This generates higher expectation for behavioral AI, which has to perform at the same level of believability as the graphics and physics in order to not break the illusion. Today we have many commercial AI middleware software packages that address the need for better game AI with elegant and effective solutions, but all of them deal primarily with combat or explorative behaviors which are not suitable for social environments.

While evaluation of the effectiveness of our approach for multi-player games has not been performed yet, the visual results[3] and informal interaction tests support our direction. Having agents and avatars powered by this technology will ensure that they will immediately show a certain degree of social presence when placed in a virtual situation. Moreover, this technology will demonstrate the validity of some of the theories of Kendon on face-to-face interaction and Scheflen on human territories. These theories cannot be formally proven because of the intrinsic nature of dealing with human behavior. However, an application of their principles demonstrates their consistency as behavioral models and proposes possible extensions to clarify some of their ambiguities.

A current system limitation that we are working on is that motion generation is currently restricted to a simple point mass model plus head and eyes rotations that, while good in its simplicity, is really far from producing believable animations. Since the framework is totally independent from the implementation we are using to realize the avatar's motion, it sounds natural to plug an animation engine into it. Thus a Motion Generation module will provide an interface to control the avatar's embodiment while translating motivations into requests for the animation system. For example, a strong motivation to moving to the left would be translated into a walking cycle toward the a destination point while a weaker motivation would result just in a posture shift. Plugging

[3] See companion video at: http://cadia.ru.is/projects/cadiapopulus/pedica2009.avi

an animation engine into our framework will allow the generation of far more convincing movement. At that point, we will have the possibility of realistically evaluate our technology to its practical contribution to believability.

Acknowledgments. We are grateful to Dr. Adam Kendon for discussions and sending the unobtainable Scheflen's book. Also big thanks to the CADIA and CCP team. This work is supported by the Humanoid Agents in Social Game Environments Grant of Excellence from The Icelandic Research Fund.

References

1. Mori, M.: The uncanny valley. Energy 7(4) (1970)
2. Cassell, J., Vilhjalmsson, H.: Fully embodied conversational avatars: Making communicative behaviors autonomous. Autonomous Agents and Multi-Agent Systems 2(1), 45–64 (1999)
3. Vilhjalmsson, H., Cassell, J.: Bodychat: Autonomous communicative behaviors in avatars. In: Autonomous Agents, pp. 477–486. ACM Press, New York (1998)
4. Kendon, A.: Conducting Interaction: Patterns of behavior in focused encounters. Cambridge University Press, Cambridge (1990); Main Area (multimodal commnication)
5. Friedman, D., Steed, A., Slater, M.: Spatial social behavior in second life. In: Pelachaud, C., Martin, J.-C., André, E., Chollet, G., Karpouzis, K., Pelé, D. (eds.) IVA 2007. LNCS (LNAI), vol. 4722, pp. 252–263. Springer, Heidelberg (2007)
6. Salem, B., Earle, N.: Designing a non-verbal language for expressive avatars. In: Collaborative Virtual Environments, pp. 93–101. ACM, New York (2000)
7. Pedica, C., Vilhjálmsson, H.H.: Social perception and steering for online avatars. In: Prendinger, H., Lester, J.C., Ishizuka, M. (eds.) IVA 2008. LNCS (LNAI), vol. 5208, pp. 104–116. Springer, Heidelberg (2008)
8. Cassell, J., Vilhjalmsson, H., Bickmore, T.: Beat: the behavior expression animation toolkit. In: SIGGRAPH 2001, August 12-17, pp. 477–486. ACM Press, New York (2001)
9. Gillies, M., Ballin, D.: Integrating autonomous behavior and user control for believable agents. In: Autonomous Agents and Multi-Agent Systems, July 19-23, pp. 336–343. ACM Press, New York (2004)
10. Helbing, D., Molnár, P.: Social force model for pedestrian dynamics. Physical Review E 51(5), 4282 (1995)
11. Couzin, I., Krause, J., James, R., Ruzton, G., Franks, N.: Collective memory and spatial sorting in animal groups. Journal of Theoretical Biology, 1–11 (2002)
12. Pelechano, N., Allbeck, J.M., Badler, N.I.: Controlling individual agents in high-density crowd simulation, pp. 99–108 (2007)
13. Heigeas, L., Luciani, A., Thollot, J., Castagne, N.: A physically-based particle model of emergent crowd behaviors. In: Proc. of GraphiCon, September 5-10 (2003)
14. Helbing, D., Molnar, P., Schweitzer, F.: Computer simulations of pedestrian dynamics and trail formation (1994)
15. Treuille, A., Cooper, S., Popovic, Z.: Continuum crowds. In: SIGGRAPH 2006 Papers, pp. 1160–1168. ACM, New York (2006)
16. Lee, K.H., Choi, M.G., Hong, Q., Lee, J.: Group behavior from video: a data-driven approach to crowd simulation, pp. 109–118 (2007)
17. Musse, S.R., Thalmann, D.: Hierarchical model for real time simulation of virtual human crowds. IEEE Transactions on Visualization and Computer Graphics 7(2), 152–164 (2001)
18. Shao, W., Terzopoulos, D.: Autonomous pedestrians. Graph. Models 69(5-6), 246–274 (2007)

19. Rehm, M., André, E., Nischt, M.: Let's come together — social navigation behaviors of virtual and real humans. In: Maybury, M., Stock, O., Wahlster, W. (eds.) INTETAIN 2005. LNCS, vol. 3814, pp. 124–133. Springer, Heidelberg (2005)
20. Reynolds, C.W.: Steering behaviors for autonomous characters. In: Proc. of the Game Developers Conference, pp. 763–782. Miller Freeman Game Group, San Francisco (1999)
21. Amor, H.B., Obst, O., Murray, J.: Fast, neat and under control: Inverse steering behaviors for physical autonomous agents
22. 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, pp. 59–66 (2007)
23. Jan, D., Traum, D.R.: Dialog simulation for background characters, pp. 65–74 (2005)
24. Scheflen, A.E.: Human Territories: how we behave in space and time. Prentice-Hall, New York (1976)
25. Goffman, E.: Frame Analyses: An Essay on the Organization of Experience. Harvard University Press, Cambridge (1974)
26. Hall, E.T.: The Hidden Dimension. Doubleday, New York (1966)
27. Vilhjalmsson, H.: Animating conversation in online games. In: Rauterberg, M. (ed.) ICEC 2004. LNCS, vol. 3166, pp. 139–150. Springer, Heidelberg (2004)