Applied Artificial Intelligence, Volume 24 Issue 6, July 2010, 575-593
Taylor & Francis, Inc. Bristol, PA, USA

# Spontaneous Avatar Behavior for Human Territoriality

Claudio Pedica and Hannes Högni Vilhjálmsson

Center for Analysis and Design of Intelligent Agents
School of Computer Science, Reykjavik University
Menntavegur 1, IS-101 Reykjavik, Iceland
{claudio07, hannes}@ru.is

**Abstract**

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 avatar group dynamics during social interaction. We model the territorial dynamics of social interactions as a set of social norms which constrain the avatar's reactive motion 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 left behind when social situations are simulated. We carried out an evaluation of the technology and the result confirms the validity of our approach.

## 1 Introduction

Most Massively Multiplayer Online Games (MMO) available nowadays portray their players as animated characters or avatars, under the user's 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 still frame but, once they start moving, they don't look life-like anymore giving to the viewer a slight sense of discomfort. This feeling is explained in robotics as the "uncanny valley" (Mori 1970). When we apply the "uncanny valley" hypothesis to animated characters, the conclusion is that the more realistic a virtual creature looks the more we will expect a realistic behavior from it. 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.

In most commercial avatar-based systems, the expression of communicative intent and social behavior relies on explicit user input (Cassell & Vilhjalmsson 1999). 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 (Vilhjalmsson & Cassell 1998), 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. 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 social engagement (Friedman, Steed & Slater 2007). 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 (Salem & Earle 2000). 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

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

avatars from arranging themselves in more organic and natural ways, with the net result of breaking the illusion of believability.

Therefore we decided on an approach that uses a set of composable reactive behaviors which have as input the knowledge of the territorial structure of a social interaction (e.g., the territory of conversations), and as output generate a social motivational force field inspired by work on dynamic group behavior in embodied agents. We focus on continuous social avatar positioning and orientation in this current work, while keeping in mind that other layers of behavior control introduced in previous work will need to be added to fully support the social interaction process.

## 2 Related Work

### 2.1 Automating Avatar Control

Automating the generation of communicative behaviours in avatars was first proposed in BodyChat where avatars were not just waiting for their own users to issue behavior commands or social clues, but were also reacting to events in the online world according to preprogrammed rules based on a model of human face-to-face behavior (Vilhjalmsson & Cassell 1998). The focus was on gaze cues associated with establishing, maintaining and ending conversations. A study showed that the automation did not make the users feel any less in control over their social interactions, compared to using menu driven avatars. In fact, they reported they felt even more in control, suggesting that the automated avatars were providing some level of support (Cassell & Vilhjalmsson 1999). The Spark system took this approach further by incorporating the BEAT engine (Cassell, Vilhjalmsson & Bickmore 2001) to automate a range of discourse related co-verbal cues in addition to cues for multi-party interaction management, and was able to demonstrate significant benefits over standard chat interaction in online group collaboration (Vilhjalmsson 2004). Focused more on postural

shifts, the Demeanor system (Gillies & Ballin 2004) blends user control at several different levels of specification with autonomous reactive behavior to generate avatar posture based on affinity between conversation partners.

However, both Spark and Demeanor assume that the user will bring their avatar to the right location and orient correctly for engaging other users in conversation. We believe that the user should not have to control such sublte bodily cues and that they should be part of the autonomous behavior of the avatar itself. The user should put conscious effort only in deciding to engage in social interaction. At that point, his avatar will autonomously perform all the silent communicative behaviors expected in that particular instance of interaction, avoiding asking for user intervention when it is not needed. Thus, to achieve this goal we need to start exploring the "unconquered land" of simulating small scale group dynamics in avatars.

### 2.2 Simulating Group Dynamics

Simulating group dynamics concerns with modelling and imitating the kinetic evolution over time of a group of individuals. We can talk about large scale group dynamics and small scale group dynamics, and the first clear difference between them is 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 agent is managed in real time by rules of behavior.

The main difference between them is how sophisticated we expect the behavior of each single individual to be. The first approach focuses more on the group behavior as a whole whereas the second focuses of the richness of behavior of the single entities.

Most 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 towards a destination, therefore its motion is easily modeled as a particle. One of the classical works on particle-based systems is the one of Helbing and Molnár (Helbing & Molnár 1995) 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. (Couzin, Krause, James, Ruzton & Franks 2002). 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. (Pelechano, Allbeck & Badler 2007) 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 different approach is the one of Treuille et al. (Treuille, Cooper & Popovic 2006) 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. (Lee, Choi, Hong & Lee 2007). In this work the authors recorded a crowd of people from a top down view and then used computer vision to extract motion trajectories out of it. Afterwards the data is fed into an agent model which learns 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. (Musse & Thalmann 2001) 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. The work of Shao and Terzopoulos (Shao & Terzopoulos 2007) is also very comprehensive, and 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. (Rehm, Andre & Nisch 2005) 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 (Kendon 1990).

In a pure agent-based approach the action generation loop typically produces a discrete sequence of behavior, which has a downside. Human behavior is not discrete but rather continuous. One of the main advantages 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 takes place at the lower levels of the artificial intelligence. Abstract high level behaviors can be decomposed into 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 (Reynolds 1999) has been of great inspiration.

### 2.3 Small Scale Group Dynamics

The pioneering work of Jan et al. (Jan & Traum 2007), 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 (Jan & Traum 2005) where agents could group together and move from one group to another, 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 that affect agent orientation are not taken into consideration. Reorienting is an important part of the behavior expressed by people during social interactions. Moreover as we know from Scheflen (Scheflen 1976) 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 also be motivated. Furthermore, we should remember that a conversation is a unit at the interactional level and has a territorial domain (Kendon 1990) 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 (Kendon 1990).

Since this behavioral influence is common to other forms of territory as well (Scheflen 1976), 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 as territorial behaviors (Scheflen 1976). 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 (Kendon 1990). 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 (Goffman 1974). 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).

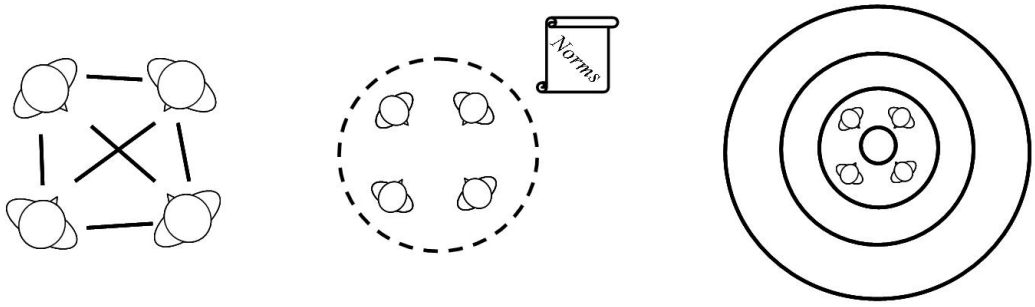


**Figure 1:** *A schematic representation of the three main concepts of a conversation as an instance of F-formation. The first picture shows how the participants are involved in a behavioral relationship which sustains a stable formation. The second picture shows a conversation as a unit at interactional 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 (Scheflen 1976) 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. Notice that the concentric spaces define zones of progressively 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 shared center and have a claim on the territory.

The classification of the territorial organization of small groups goes from an *element* to a *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

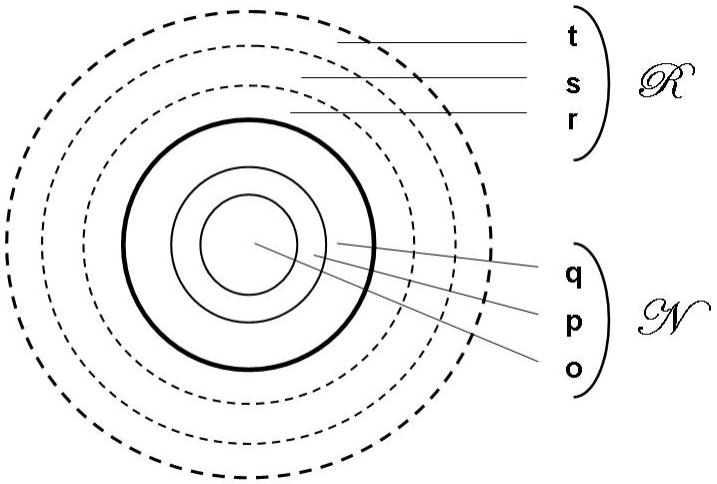


**Figure 2:** *The paradigm of territorial organization proposed by Scheflen and applicable to all levels of territorial organization. The o, p and q spaces belong to the nucleus N whereas the r, s and t spaces belong to the region R.*

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. 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 nuclei 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

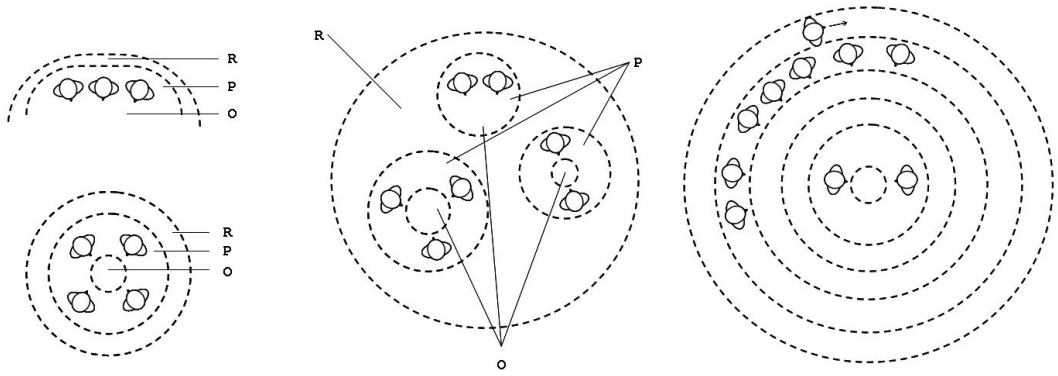


**Figure 3:** *From left to right, element, F-formation, gathering and hub. These are the main territorial fields of Scheflen's classification. Each of them can be seen as increment of the previous one for required space and complexity. In addition, more complex territory can contain simpler ones. The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

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 Outline of the Approach

Simulating human territoriality requires that agents and avatars are aware of the social interaction they are engaged in and display a behavior influenced by its constraints. The approach combines the knowledge on the territorial organization of a social situation (e.g.

the territory of a conversation) with low level reactive behaviors which generate "social forces" in a way similar to other works on dynamic group behavior in embodied agents, such as (Jan & Traum 2007). The focus of our approach is on automating continuous social avatar positioning and orientation as a function of the territorial field and the set of norms of a given social context.

Many approaches propose interesting solutions for generating the stream of actions that an agent, or in our case, an automated avatar needs to perform in order to believably simulate the behavior of a human engaged in interaction. Each action usually triggers some motor function directly at locomotion level in order to animate the agent. We believe that the sequence of actions needs to pass through an intermediate layer in order to achieve the desired fluidity of movements and reactions (Fig. 4). This extra level between action planning and motion generation, is responsible for smoothing the agent's overall behavior by applying motivational forces directly to the underlying motion model. Therefore a reactive middle layer provides a suitable solution for filling the gap between two consecutive actions, generating a net continuous fluid behavior. This approach is particularly well suited for modelling unconscious reactions and motion dynamics in social interactions, where the contextual territoriality places behavioral constraints on an avatar's reactivity, building an illusion of continuous awareness.

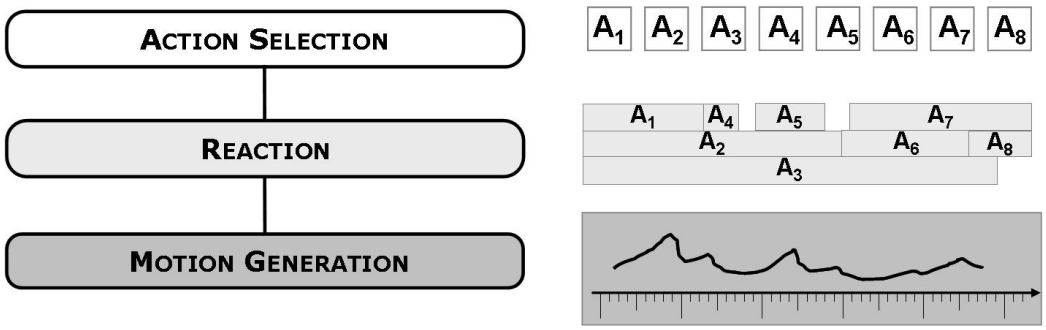


**Figure 4:** *An abstract diagram of the framework's architecture. The Reaction middle layer provides an interface between Action Selection and Motion Generation supporting the transformation of the sequence of discrete actions $A_1$, $A_2$, ..., $A_8$ into continuous motion. Each action activates a reactive behavior which generates a motivation for a certain time interval. Afterwards, the whole set of generated motivations are combined and submitted to the lower Motion Generation layer.*

We should keep in mind that this approach was chosen to simulate territorial behavior and that other layers of behavior control introduced in many related "Embodied Conversational Agent" works, such as conversational gesturing and communicative facial expression, will need to be added for full support of the social interaction process. Nevertheless, our architecture proved to be flexible and extensible enough to fully integrate an external module for simulating multi-party turn taking in conversation. The turn-taking module implements an extended version of (Thorisson 2002) and was developed at CADIA research lab following the CDM approach (Thorisson, Benko, Arnold, Abramov, Maskey & Vaseekaran 2004). Its integration into our architecture was straight forward. It was enough to encapsulate the external module into a reactive behavior to run when the avatar joins a conversation.

## 5 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 a 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 raw 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. 5).

The reaction paradigm is in effect a loop of continuous control of the avatar's motion. The

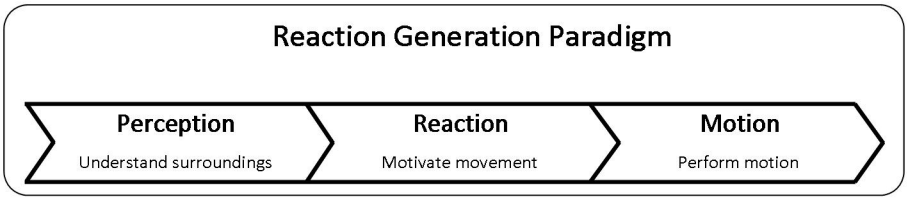


**Figure 5:** *Outline of our Reaction Generation Paradigm. The agent understands its surroundings through its senses. Successively, perceptual stimuli are used for generating motivations to move. Finally, motivations are transformed into physical forces which steer the agent's motion.*

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. 6). The sense of proximity is a simple way of simulating the human awareness over the four distances of the Proxemics Theory (Hall 1966). 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. These two senses continuously gather information about the avatar's surroundings, producing perceptual data which can be analyzed and used for generating reactions.

Reactions are generated by a set of reactive behaviors that compute 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 to perform a linear or rotational motion. The reactive framework permits applying motivations to certain part of the

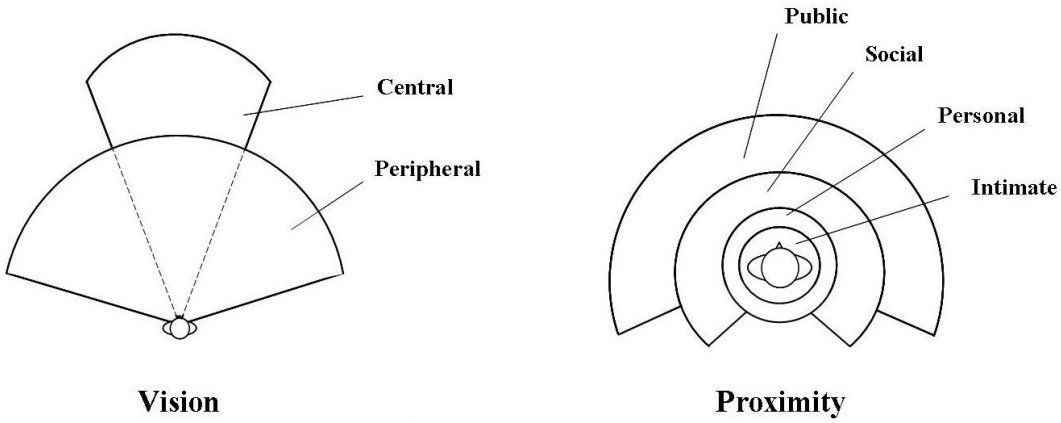


**Figure 6:** *Two diagrams showing the spatial structure of the sense of vision and proximity. Notice that proportions have not been respected to make the drawing more readable.*

avatar's body. Assuming we have a collection of motor controllers controlling those parts, then each motivation is nothing but a *command* to a motor controller. For example, a motivation to look at a point in space will be a command to the gaze motor controller that then moves the torso, head and eyes. Even though motivations may look like raw inputs for motor controllers, they are actually much more and, in fact, they have a couple of nice properties. They can be inhibited by other modules of the agent's mind and can be mixed into a final result. 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, where each type commands a specific motor controller. The combination of motivations allows multiple behaviors to stimulate the body at the same time. Several composition algorithms can be chosen for this step. For example, motivations for linear motion could be linearly combined whereas motivation for rotational motion could be simply selected based on their weight. After computing the set of final motivations, each of them will command a motor controller which performs a movement that respects the constraints imposed by the physical motion model.

## 6 Implementing Awareness of Social Context

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 (Kendon 1990) and (Scheflen 1976). 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. 7). The activation of this set of behaviors produces the reactive dynamics expected from a group of people that has accepted the same social context.

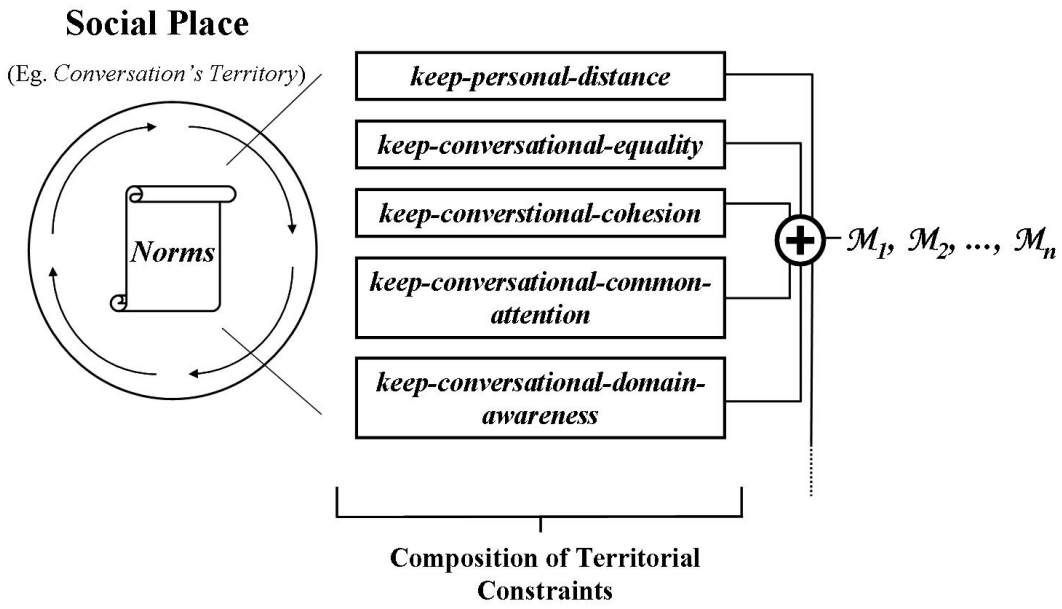


**Figure 7:** *A diagram to explain how social situation awareness is realized. Inside the border of the social place, the territorial organization states a set of norms that constrain the avatar's reactivity. Thus, the set of norms maps into a set of reactive behaviors that implements a composition of territorial constraints. The composition blends with other behaviors leading to a behavioral influence marked on the resulting final motivations* $M_1, M_2, ..., M_n$.

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 towards the nucleus 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's 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.

## 7 Resulting Behavior

Let's take a closer look at the resulting behavior with a concrete example of the joining example introduced above. Figure 8 describes the social occasion with a sequence of screen shots that will be used to describe how our system works in this situation.

At the beginning (Fig. 8a) the avatar is far away from the group of his friends while they are engaged in a conversation as clearly shown by their formation. The user issues a command to get closer to the conversation, telling the avatar to move and reach the spot indicated by the red marker. A *move-to-destination* and an *avoid-obstacles* behaviors will be activated in the avatar's mind. The first behavior will see the given destination spot as an attractor point that generates a sort of a psychological attraction force towards it. This force will be a

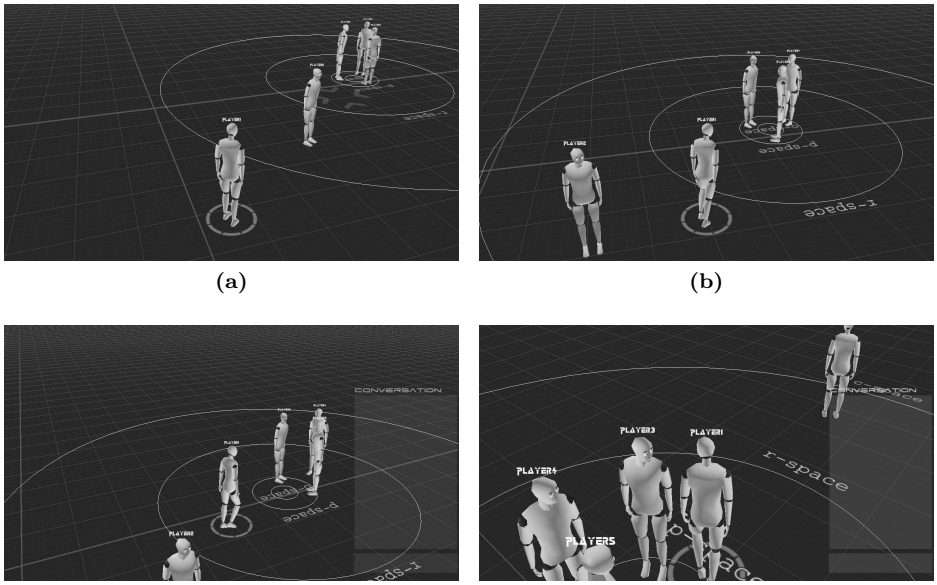


**Figure 8:** *A scenario described by a sequence of four figures: (a) the avatar starts moving to reach the destination issued by the user and marked on the ground; (b) the avatar approaches the conversation and the members of it glance at the newcomer; (c) the avatar joins the conversation while the other members move away to make space; and, (d) the avatar leaves the conversation while the rest of the group will soon rearrange.*

candidate motivation generated by the reactive behavior. This motivation does not have any effect on the body yet, it is only a candidate. First it has to go through an arbitration phase where it will be blended with other candidate motivations that also claim control over the body velocity and therefore the avatar's ability to move across the environment. It should be clear, that the candidate motivation generated by the *move-to-destination* behavior will have a rather high priority. The reason is that the behavior is executing an order that came straight from the user. The user expects his avatar to be exactly on the marked spot when it stops moving. Therefore the candidate motivations of this reactive behavior have to be strong enough to overcome other candidates who might steer the avatar away from the user-expected destination spot. Having a higher priority will let our candidate motivation overwhelm other lower priority candidates, assuring a good advantage in the competition for controlling the body. In case of other candidates with equally high priority, all of them will be blended together and result in a final motivation. For example, that is the case

when our avatar approaches the fellow who blocks his path to the destination point (Fig. 8a). The *avoid-obstacles* behavior, recognizing the potential threat, will generate a candidate motivation to move away from the fellow. At this point the attraction toward the destination and the repulsion from the obstacle will blend using a chosen blending schema, each of them with different weights so that the avatar will strongly avoid bumping into an obstacle while still trying to get closer to the destination.

In the next picture (Fig. 8b) our avatar gets closer to the conversation but still has to reach the destination point. He will be close enough to attract the attention of at least one of the members in the conversation. There is a behavior active inside the mind of the members which controls the avatar gaze and motivates a quick glance towards those who step into the territorial domain of the social situation. Quickly glancing towards an approaching fellow makes the avatar look aware of the territorial boundaries of the conversation. This behavior will generate a candidate motivation to turn the avatar's gaze toward the target. We want this candidate to be of high priority because other latent candidates may keep control over the gaze. We expect the avatar to glance and then go back watching where he was looking before. Although only one of the members of the conversation glances at the passerby at first, from the picture (Fig. 8b) we see that the other members glance as well. This is the effect of another behavior activated in those having a conversation. This behavior checks where the other members of the conversation are looking and motivates a quick glance to check if anything noteworthy is going on there. Thus it might happen that only one member will recognize the approach of a new individual but the rest of group glances at him, reacting to the gaze of the first member.

In the third figure (Fig. 8c) we see our avatar finally at the destination spot. Moreover he is now close enough to the conversation to have the rights to join the other friends. Since the conversation is assumed to be open and everybody is welcome to join, it is enough for him to have his whole body oriented towards the nucleus of the conversation to become part

of the activity. This will activate a set of behaviors inside his mind, meant to influence his position and orientation so to keep a proper formation with the other members. In fact, such behaviors can influence the avatar's dynamics by generating a motivational force field similar to the one described in (Pedica & Vilhjálmsson 2008), but with a couple of major changes to the model of dynamics. The first is the adding of an extra and stronger motivational repulsive force to keep the avatar's intimate zone safe. The second is that we compute forces using the future locations of the others. Notice that positions are predicted only few seconds ahead and in such a short amount of time, we can safely assume them to change linearly.

Once all the candidate motivations have been generated, they will blend eventually into a final motivation that will control the body and gently steer the avatar towards a position and an orientation that functions well for the ongoing social interaction. The other members will do the same thanks to the same set of behaviors that was already activated in their minds. They will recognize the attendance of a new member and move away to accommodate his presence and leave him some space.

In the fourth and last picture (Fig. 8d) we see our avatar leaving the conversation and moving towards a new destination point issued by the user. In moving away from the nucleus the avatar will not be a member anymore and consequently all the behaviors related to the social situation will be deactivated. Therefore the behavioral influence of the social context will end and the avatar will behave like someone external to the conversation. The rest of the members will recognize the event and will rearrange accordingly, aware of having one less person in the group.

## 8 Evaluation

An evaluation was conducted to test the validity of our approach and the advantage of simulating group dynamics in conversation. We decided to let people judge our technology

in four different test cases, each of them focused on one of the following important situations:

1. a person joining a conversation;
2. a participant moving around within the conversation;
3. a person trying to avoid a conversation;
4. a person passing by a conversation.

For each test case we asked subjects to watch two videos of the same scene with and without our technology and, afterwards, to answer a set of five simple questions about the artificiality or appeal of them. To avoid any trivial differentiation, each pair of videos had been recorded from the same system running a simulation of turn taking to let the conversation appear 'alive'. The difference was that in one instance the group dynamics were disabled and in the other enabled. To have as many participants as possible, we conducted the survey using a web page[3]. We had 171 people participating, 66% classified as *non-gamers* and 33% as *gamers*. The results confirm our hypothesis that simulating group dynamics in social interactions significantly improves believability, appeal and the feeling of avatar awarennes. For three out of four tests, the scene powered by our technology was judged significantly less artificial, more appealing and more socially rich. Moreover the avatars looked more aware of the context and their surroundings. Only in test number 2, the users found the two videos equally artificial and appealing, even though our avatars looked more socially rich and connected with the situation. In conclusion, the result of the evaluation shows margins of improvement for our group dynamics in conversation but confirms the promising direction and effectivness of our approach. More details about the study and its results appear in a future publication.

[3] http://hasge.cs.ru.is/survey

## 9 Conclusions and Future Work

The approach described here has been implemented in the CADIA Populus social simulation platform for virtual environments (Pedica & Vilhjálmsson 2008) and is an important contribution in the field of graphical avatars for games. 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.

An evaluation of our results confirms the effectiveness of the approach and a video of group dynamics in conversation is available at our platform's website[4]. 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 independent from the implementation of the avatar's motion, it sounds natural to plug an animation engine into it to produce more believable gaze movement and legged locomotion. Another improvement we plan to work on, would be a new model of group dynamics built on top of two of the most basic concepts of the theories of Kendon and Scheflen, namely the *transactional segment* and *locations*.

[4] http://populus.cadia.ru.is/

## 10 Acknowledgments

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

## References

Amor, H. B., Obst, O. & Murray, J. (n.d.), 'Fast, neat and under control: Inverse steering behaviors for physical autonomous agents'.

Cassell, J. & Vilhjalmsson, H. (1999), 'Fully embodied conversational avatars: Making communicative behaviors autonomous', *Autonomous Agents and Multi-Agent Systems* **2**(1), 45–64.

Cassell, J., Vilhjalmsson, H. & Bickmore, T. (2001), Beat: the behavior expression animation toolkit, *in* 'SIGGRAPH01', ACM Press, New York, NY, pp. 477–486.

Couzin, I., Krause, J., James, R., Ruzton, G. & Franks, N. (2002), 'Collective memory and spatial sorting in animal groups', *Journal of Theoretical Biology* pp. 1–11.

Friedman, D., Steed, A. & Slater, M. (2007), Spatial social behavior in second life, *in* '7th International Conference on Intelligent Virtual Agents', Vol. 4722, Springer-Verlag, Berlin, pp. 252–263.

Gillies, M. & Ballin, D. (2004), Integrating autonomous behavior and user control for believable agents, *in* 'Autonomous Agents and Multi-Agent Systems', ACM Press, pp. 336–343.

Goffman, E. (1974), *Frame Analyses: An Essay on the Organization of Experience*, Harvard University Press, Cambridge, MA.

Hall, E. T. (1966), *The Hidden Dimension*, Doubleday, New York, NY.

Helbing, D. & Molnár, P. (1995), 'Social force model for pedestrian dynamics', *Physical Review E* **51**(5), 4282.

Helbing, D., Molnar, P. & Schweitzer, F. (1994), 'Computer simulations of pedestrian dynamics and trail formation'.

Jan, D. & Traum, D. (2007), Dynamic movement and positioning of embodied agents in multiparty conversation, *in* 'Proc. of the ACL Workshop on Embodied Language Processing', pp. 59–66.

Jan, D. & Traum, D. R. (2005), 'Dialog simulation for background characters', pp. 65–74.

Kendon, A. (1990), *Conducting Interaction: Patterns of behavior in focused encounters*, Cambridge University Press, Cambridge. Main Area (multimodal commnication).

Lee, K. H., Choi, M. G., Hong, Q. & Lee, J. (2007), 'Group behavior from video: a data-driven approach to crowd simulation', pp. 109–118.

Mori, M. (1970), 'The uncanny valley', *Energy* **7**(4).

Musse, S. R. & Thalmann, D. (2001), 'Hierarchical model for real time simulation of virtual human crowds', *IEEE Transactions on Visualization and Computer Graphics* **7**(2), 152–164.

Pedica, C. & Vilhjálmsson, H. (2008), Social perception and steering for online avatars, *in* 'IVA '08: Proceedings of the 8th international conference on Intelligent Virtual Agents', Springer-Verlag, Berlin, Heidelberg, pp. 104–116.

Pelechano, N., Allbeck, J. M. & Badler, N. I. (2007), 'Controlling individual agents in high-density crowd simulation', pp. 99–108.

Rehm, M., Andre, E. & Nisch, M. (2005), Let's come together - social navigation behaviors of virtual and real humans, *in* 'Intelligent Technologies for Interactive Entertainment', Vol. 3814, Springer-Verlag, Berlin, pp. 124–133.

Reynolds, C. W. (1999), Steering behaviors for autonomous characters, *in* 'Proc. of the Game Developers Conference', Miller Freeman Game Group, San Francisco, CA, pp. 763–782.

Salem, B. & Earle, N. (2000), Designing a non-verbal language for expressive avatars, *in* 'Collaborative Virtual Environments', ACM, pp. 93–101.

Scheflen, A. E. (1976), *Human Territories: how we behave in space and time*, Prentice-Hall, New York, NY, USA.

Shao, W. & Terzopoulos, D. (2007), 'Autonomous pedestrians', *Graph.Models* **69**(5-6), 246–274.

Thorisson, K., Benko, H., Arnold, A., Abramov, D., Maskey, S. & Vaseekaran, A. (2004), 'Constructionist design methodology for interactive intelligences', *A.I. Magazine* **25**(4), 77–90. ID: 515.

Thorisson, K. R. (2002), *Natural Turn-Taking Needs No Manual: Computational Theory and Model, from Perception to Action*, Vol. 19 of *Multimodality in Language and Speech Systems*, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 173–207. ID: 482.

Treuille, A., Cooper, S. & Popovic, Z. (2006), Continuum crowds, *in* 'SIGGRAPH 2006 Papers', ACM, New York, NY, USA, pp. 1160–1168.

Vilhjalmsson, H. (2004), 'Animating conversation in online games', *Lecture Notes in Computer Science* **3166**(International Conference on Entertainment Computing), 139–150.

Vilhjalmsson, H. & Cassell, J. (1998), Bodychat: Autonomous communicative behaviors in avatars, *in* 'Autonomous Agents', ACM Press, New York, NY, pp. 477–486.