REYKJAVÍK UNIVERSITY
HÁSKÓLINN Í REYKJAVÍK

Spontaneous Avatar Behaviour for Social Territoriality

Claudio Pedica

Master of Science
May 2009

Reykjavík University - School of Computer Science

M.Sc. Thesis

Spontaneous Avatar Behaviour for Social Territoriality

by

Claudio Pedica

Thesis submitted to the School of Computer Science
at Reykjavík University in partial fulfillment
of the requirements for the degree of
Master of Science

May 2009

Thesis Committee:

Dr. Hannes Högni Vilhjálmsson, supervisor
Assistant Professor, Reykjavík University

Dr. Adam Kendon
Visiting Scholar, University of Pennsylvania, USA

Dr. Emanuela Merelli
Associate Professor, University of Camerino, Italy

Spontaneous Avatar Behaviour for Social Territoriality

by

Claudio Pedica

May 2009

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 behaviour sophistication. Avatars and agents in a realistic virtual environment must exhibit a certain degree of presence and awareness of the surrounding, reacting consistently to unexpected contingencies and contextual social situations. Unconscious reactions serve as evidence of life, and can also signal social availability and spatial awareness to others. These behaviours get lost when avatar motion requires explicit user control. This thesis presents new AI technology for generating believable social behaviour in avatars. The focus is on human territorial behaviours during social interactions, such as during conversations, gatherings and when standing in line. Driven by theories on human face-to-face interaction and territoriality, we combine principles from the field of *crowd simulators* with a *steering behaviours architecture* to define a reactive framework which supports avatar group dynamics during social interaction. This framework manages a set of prioritized behaviours updated at various frequencies, which can be combined in different ways. This approach provides enough flexibility to model the territorial dynamics of social interactions as a set of social norms that constrain the avatar's reactivity. The resulting social group behaviour appears relatively robust, but perhaps more importantly, it starts to bring a new sense of relevance and continuity to the virtual bodies that often get separated from the simulated social situation.

Ósjálfráð félagsleg svæðishegðun fyrir avatara

eftir

Claudio Pedica

Maí 2009

Útdráttur

Trúverðugt sýndarumhverfi krefst þess að allar vitrænar verur bregðist sjálfvirkt við síbreytilegu umhverfi. Eftir þann mikla árangur sem hefur náðst við að ná fram eðlilegu útliti og réttri hegðun dauðra hluta, er gervigreindinni nú gefin aukinn gaumur. Stafrænir holdgervingar, eða avatarar, og vitverur í trúverðugu sýndarumhverfi verða að sýna að þau séu hluti af umhverfinu og að þau skynji það sem fram fer í kringum þau með því að bregðast stöðugt við óvæntu félagslegu áreiti. Ósjálfrátt viðbragð er eitt merki lífs og getur einnig gefið til kynna hvort viðkomandi hefur athyglina í lagi og gefur kost á samskiptum. Slíku viðbragði verður ekki við komið ef notandi stýrir avatar sínum handvirkt. Þessi ritgerð kynnir nýja gervigreindartækni sem framleiðir trúverðuga félagslega hegðun sjálfvirkt fyrir avatara. Sjónum er beint að svæðishegðun fólks við félagsleg samskipti, til dæmis við samræður, hópamyndun og þegar beðið er í biðröð. Byggt er á kenningum um tjáskipti fólks augliti-til-auglitis og svæðishegðun, ásamt því að hugmyndum frá sviði *hóphermunar* og sviði *stýrihegðunar* er blandað saman. Þannig er skilgreint viðbragðskerfi sem stýrir kvikri hóphegðun avatara við félagslegar aðstæður. Þetta kerfi heldur utanum forgangsraðað mengi hegðana sem hver er uppfærð með mismunandi tíðni. Hegðununum má blanda saman á mismunandi hátt. Þessi aðferð er nógu sveigjanleg til að útfæra líkanið af kvikri svæðishegðun sem sérstakt hegðanamengi félagslegra venja sem þvinga avatarinn til að bregðast við á ákveðinn hátt. Útkoman er félagsleg hóphegðun sem virkar sannfærandi og virðist gæða sýndarlíkamana, sem hingað til hafa verið einangraðir frá félagslegu umhverfi sínu, samfelldu og markvissu lífi.

Dedicated to Marialina, Michele and Lorenzo.

Acknowledgements

This thesis is supported by the Humanoid Agents in Social Game Environments Grant of Excellence from The Icelandic Research Fund and the Italian Ministry of Education, University and Research.

First of all, my deepest thanks to my supervisor and teacher Dr. Hannes Vilhjálmsson. He has been a guidance along my first experience with research and AI but, most of everything, he has been a friend. He gave me numerous directions but also enough freedom to explore my ideas and to believe in them. Upon many others, this is for me a real sign of great intelligence. This work would never been done without him.

I am are very grateful to the other project team members, in particular Marta Larusdottir, Dr. Kristinn R. Thorisson and Eng. Eric Nivel, for numerous discussions and suggestions. If I have learned something on how to carry on a real project, it is especially thanks to our meetings.

I would also like to thank Torfi Olafsson and the rest of the team at CCP Games for their productive collaboration.

Special thanks go to Professor Adam Kendon for his precious material on face-to-face interaction and valuable personal communication. He suggested me to have a look on the Scheflen's work, sending to me his unobtainable book on human territories that turns to be a great contribution for this thesis.

I would like to thank Dr. Luca Aceto for his 'Italian' welcome at Reykjavik University and for looking after me during my permanence, especially at the beginning. He certainly is an excellent representer abroad for my country as well as a great teacher. Apparently, he is also the only one able to organize a soccer match at the department of Computer Science.

I am grateful to my teacher at Camerino University, Dr. Emanuela Merelli. She encouraged me to take one of the most critical steps in every new adventure; the first. Hope she will continue supporting the idea of going abroad as an essential stage in the training of every student. The international collaboration with Reykjavik University has been very fruitful for me and I am sure it will be for many others in the future.

Thanks to my friends Agnes, Anu, Ari, Fids, Jan and Paul. Their are my family in Iceland and without them my permanence would have been much colder, darker and lonely. Thanks also to my flatmates Mathilde, Brandon, Richard and Tom to have been the most unexpected and beautiful surprise of 2008.

I would like to thank: my ex-flatmate Lisa for all the fascinating conversations we had about theater, acting and personal space; all the guys at CADIA Research Lab. for being the main contribution of such a nice and compelling atmosphere; Angelo, Enrico and Raffaele for bringing the colors of the beautiful south of Italy to CADIA Research Lab. and for extending the Italian lunch menu with new dishes; Matteo that just arrived in September; Marco for composing the songs that I have been listening almost every morning in the last three months; Paolo to continuously remember everybody that the life has to be faced with courage; Raphael for sharing almost my same misadventures and for letting me talk about them for hours.

Finally, thanks also to all the people which tried to discourage me in going to Iceland. I achieved certain results also to show how wrong you were.

Publications

Pedica, C. and Vilhjálmsson, H. (2008). Social perception and steering for online avatars. In IVA '08: Proceedings of the 8th international conference on Intelligent Virtual Agents, pages 104-Ű116, Berlin, Heidelberg.Springer-Verlag.

Contents

List of Figures

Chapter 1

Introduction

Life is what happens to you while you're busy making other plans.
John Lennon

In June 2003, my teacher and supervisor professor Hannes Vilhjálmsson defended a thesis on "Avatar Augmented Online Conversation" (H. H. Vilhjalmsson, 2003) at the Massachussets Institute of Technology which pushed a step further the state-of-art on automated conversational avatars. In his essay, he showed how to automate the behaviour of virtual personal conversational representative in such a way to improve the communicative power of a computer mediated conversation. The present work fits into the larger picture of his automated avatars approach, that had a first seminal incarnation in 1996. Since research is also a history of people, before taking up the baton and start moving into the specific subject of the present, I would like to quote him on the subtleties of mediating human conversation, helping people communicating better through machines.

> *One of the most important roles played by technology is connecting people and mediating their communication with one another. [...] Building technology that mediates conversation presents a number of challenging research and design questions. Apart from the fundamental issue of what exactly gets mediated, two of the more crucial questions are how the person being mediated interacts with the mediating layer and how the receiving person experiences the mediation. [...] Establishing and maintaining a channels of communication with other human beings face-to-face is an ability that has evolved since the dawn of humanity. The coordination of a conversation is not merely a personŠs spoken transmission of thought, but rather it is a dynamic process involving exchanges of gesture, gaze, facial expression and body posture, carefully coordinating the mutual understanding about what is being shared and how to proceed with the conduct. [...] When communicating, we expect everyone to adhere to a shared protocol. The protocol allows us to interpret everyoneŠs words and actions, in the current context. While ensuring that the conversation unfolds in an efficient and orderly fashion, this elaborate process does not ask for much conscious effort beyond what is required to reflect on the topic at hand. Conversation, and the processes*

> *that contribute to its successful execution, have been studied extensively in the fields of discourse analysis, conversation analysis, sociolinguistics, social psychology and computational linguistics.* (H. H. Vilhjalmsson, 2003, pages 19–20)

Technologies that mediate conversation between two or more people have of course continued to evolve since the telegraph and today networked computers play an essential role. As we know from our experience a number of limitations make the medium unsuitable for many of the tasks traditionally solved face-to-face. Nevertheless, there are application domains in which a computer mediated communication successfully responds to the user needs. Due to their intrinsic casual nature, such domains can take advantage of every technological improvement in mediated conversation. Amongst many others, one particular domain has been of special interest for the present work.

Massively Multiplayer Online Role Playing Games (MMORPGs) are a rapidly growing form of mass entertainment delivered over the Internet in the form of live game worlds that persist and evolve over time. Players connect to these worlds using client software that renders the world from their perspective as they move about and meet fellow players. The community is the cornerstone of these games. Therefore, any effort spent on supporting communication and social interaction between players has to be considered valuable for the application. When games wish to use avatars to represent players in environments where they can virtually meet face-to-face, all the animated behaviors that normally support and exhibit social interaction become important. Since players cannot be tasked with micro-management of behaviour, the avatars themselves have to exhibit a certain level of social intelligence (Cassell & Vilhjalmsson, 1999 ; H. Vilhjalmsson, 2004). The purpose of such avatar AI is in fact twofold: to give players helpful cues about the social situation and to ensure that the whole scene appears believable and consistent with the level of game world realism.

1.1 Avatars in Social Virtual Environments

Avatars in social virtual environments are used as two dimensional or three dimensional human or fantastic representations of a person's self. A virtual avatar is under the control of the user that can issue commands as a way of interacting with the virtual world. Technically, an avatar can be seen as a sophisticated user interface toward the affordance of the virtual environment. In fact, such representations can be commanded to explore the universe they are in, interact with objects and facilities or conduct conversations with avatars under other users' control. Usually, the purpose and appeal of such virtual representations is to enhance online conversation capabilities or to allow a greater usability of the virtual universe, seen as a way of providing services to the users or engaging them in some way. This brings us to the primary motivation behind this research.

Since an avatar is a user interface toward the affordance of the virtual environment, it is also an interface toward the conversations taking place inside such a virtual universe. A poor interface leads to poorer communication. Therefore, an avatar must be powered with a certain degree of social intelligence and context awareness in order to look believable improving the users' social experience and their ability to communicate.

1.2 Problem Statement

In most commercial avatar-based systems, the expression of communicative intent and social behaviour 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 (H. 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. 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 (Kendon, 1990). Some of these behaviours 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 behavioural 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 (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, this is in fact a simplification. Most of the communicative content shared in a conversation is not spoken and is involuntary. This is something we are not aware of most of the time. A conversation is indeed a formation but a very 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.

A conversation is much more than a circular disposition of people. It defines a positional and orientational relationship amongst its participants. This arrangement has been described by Kendon (Kendon, 1990) as an instance of an F-formation system. Moreover, since a conversation is not a fully rigid formation, external events from the environment or individual behaviour of single participants may produce fluctuations inside this system

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

that lead to compensational rearrangement that avatars can automate without requiring input from their human users. In fact, this is the kind of reactive behaviour which is ill suited for explicit control (Cassell & Vilhjalmsson, 1999). More generally a conversation is also an example of human territorial organization. The space around it is structured in a certain fashion and specific behaviours take place within it. The behaviours not only indicate a shared idea about the social context, but also a shared idea about the territorial domain of that context. These kinds of behaviours have been classified by Scheflen (Scheflen, 1976) as territorial behaviours. They don't have to be considered in isolation but rather as a particular way of looking at the behavioural 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 this special class of behaviours in order to properly simulate individuals engaged in a social interaction, especially since we don't want to limit ourselves to simple conversations, but also deal with a variety of different social situations each of which could have is own territorial field. Territorial behaviours are intrinsically unconscious and reactive, therefore they are especially ill suited for explicit user control and must be automated.

1.3 Humanoid Agents in Social Game Environments

This thesis is a part of the Humanoind Agents in Social Game Environments Project (HASGE), a larger collaborative effort between CCP Games Inc.[3], a major Icelandic MMORPG developer, and the CADIA Research Lab[4] at Reykjavik University where appreciable effort has been spent in the field of embodied conversational agents specializing in multi-modal behaviour generation based on social and cognitive models. The work presented by this thesis is one piece of the project which deals with modelling the relatively low-level motion dynamics in conversational interaction, which despite several ongoing efforts, is largely an unsolved problem.

The overall goal of HASGE is to develop new methods to create believable human behaviour in animated characters for massively multiplayer games. The focus is on natural motion and believable social interaction. The characters are either fully controlled by the game AI or they are avatars under the direction of human players. A particular emphasis is placed on the generation of spontaneous behaviour that supports communication. In general the approach involves scientific, engineering and artistic components. The analysis and modeling of communicative human behavior is drawn from an extensive scientific knowledge base constructed over the course of CADIA's prior research as well as new research. The engineering component turns scientific models into computational modules that receive a description of characters, goals and environment and return a detailed description of appropriate behaviour. Engineering and art meet in the design of an animation engine that turns the description of behavior into a smooth continuous performance of articulated human figures on the screen, in an aesthetic and engaging presentation.

[3] http://www.ccpgames.com/
[4] http://cadia.ru.is/

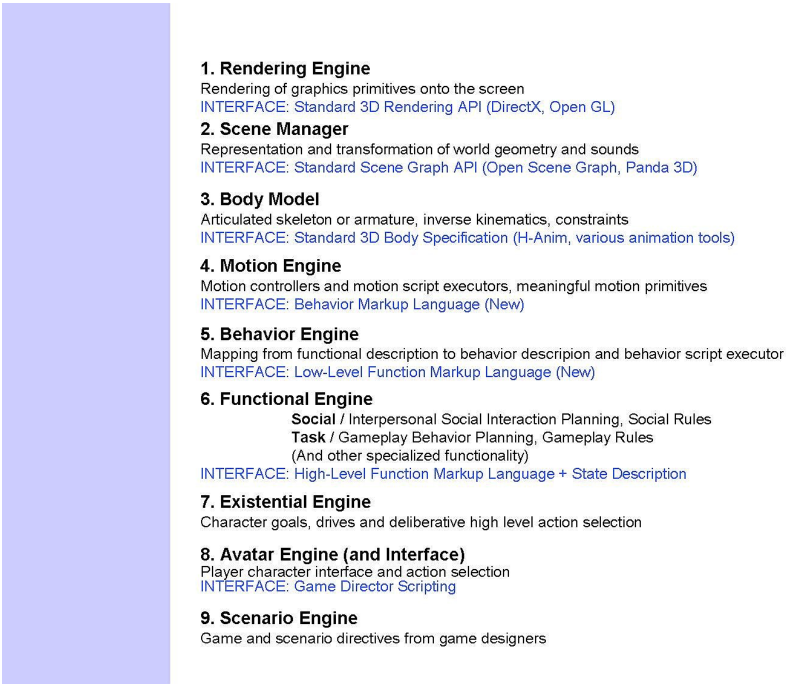

Figure 1.1: The layered architecture for intelligent animated characters as proposed in the HASGE project.

The realization of an intelligent animated character can be accomplished with a layered architecture where the first layers are responsible for getting moving images on the screen and the later layers are responsible for making those images do something sensible (Fig. 1.1). In a way, one could imagine the intelligence spread across all the layers, with a higher concentration at the later stages.

This architecture is partially based on a framework developed with an international working group on multimodal behavior generation (Kopp et al., 2006). In particular, there is a fundamental distinction between a functional description of what a character intends to accomplish, such as 'greet' and a behavioral description of actual physical behavior that accomplishes the function, such as 'wave hand'. Therefore these two layers of abstraction are clearly distinguished in the architecture, with the evolving Function Markup Language (FML) and Behaviour Markup Language (BML) serving as their interfaces respectively.

1.4 Case Study: EVE Online Walking in Stations

CCP Games was founded in the summer of 1997 with its first department located in Iceland. With the launch of its space themed MMORPG EVE Online[5] in May 2003, CCP has established itself as one of the leading companies in the field of massively multiplayer online game developers. EVE Online has about 300,000 active subscribers spread

[5] http://www.eveonline.com/

over 80 countries. Its universe runs on a single server cluster with daily peaks of 50,000 simultaneously connected users.

While many MMORPGs portray their players as animated characters, the players of EVE never leave their space ships and therefore never see each other in person, save for their static portraits. CCP Games has been holding off on using animated characters in part because the state-of-the-art in interactive character animation has not measured up to the standard of visual quality that CCP Games has reached with their environments (Fig. 1.2 and 1.3). Another factor is that many of the characters in the world of EVE are autonomous entities and the level of AI needed to exhibit believable animated behaviour in face-to-face interactions with players has been considered much to high. 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 un-lifelike giving to the viewer a slight sense of discomfort. This feeling is called in robotics the "uncanny valley" (Fig. 1.4), a terminology first introduced by Masahiro Mori in the 1970 (Mori, 1970). 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 behaviour when it is moving.

Figure 1.2: The picture shows an early character test from the CCP "Walking-in-Stations" Project. This is an example of the quality reached by CCP in real-time rendering and shading of characters.

With a ground breaking "Walking-in-Stations" expansion planned in late 2009 or in 2010, CCP is ready to tackle the "uncanny valley" problem head on. They will introduce space stations throughout the online universe where players can dock their ships, exit and mingle with other players and AI agents that hang out in lounges or mill around the commons. The stations are realized as 3D virtual environments and both players and agents are represented by fully articulated animated characters. The avatars are expected to exhibit a level of autonomy that makes them appear fully reactive to the environment they are in. Various fully autonomous agents interact with players in order to further the EVE storyline, filling well defined roles given to them by the CCP team. Other autonomous agents may interact with players to a lesser extent and some may merely appear in the background, going about their business, to bring the environments to life. A large portion of the audience, the EVE player-base, is familiar with the back-story and has particular

Figure 1.3: Two pictures show the level of details of the environments in "Walking-in-Stations". A first demonstration was performed by CCP in November of 2008 and showcased an incredible level of visual realism. The challenge then becomes to make the characters behave correctly.

expectations about what should happen and how the environments should feel. Rooting all behaviour in an already existing game-world provides an additional challenge that needs to be approached with fullest respect for the history and community of EVE Online.

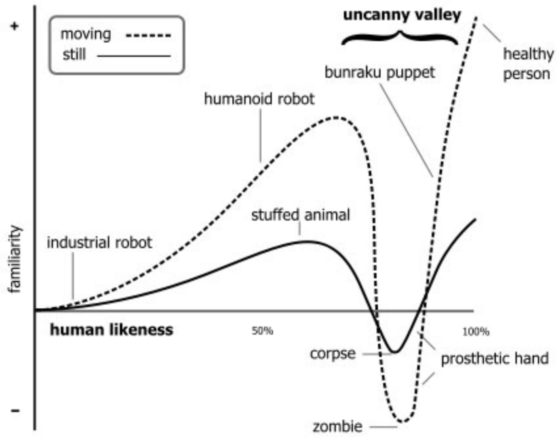

Figure 1.4: The "uncanny valley" hypothesis explained with a graph. The more human-like a synthetic creature becomes, the more deeply unfamiliar it becomes when behavioral expectations are not met.

While the world of EVE serves as an excellent target for believable social characters, it is clear that the research undertaken in the HASGE Project is about much more than populating this particular game universe. The research is approached in general terms and all the results are being shared with the rest of the research community.

1.5 Outline of the Proposed 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 a higher level territorial organization of a social situation (e.g. conversations) with low level reactive behaviours, affected by a "social force field" inspired by work on dynamic group behaviour in embodied agents such as (Jan & Traum, 2007). The focus of this work is on automating continuous social avatar positioning and orientation as a function of the territorial field and set of norms of a given social context, while keeping in mind that other layers of behaviour control introduced in related "Embodied Conversational Agent" works, such as gesturing and posture shifting, will need to be added for full support of the social interaction process.

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 behaviour of a human engaged in interaction. Each action usually triggers some motor function directly at locomotion level in order to animate the agent. The sequence of actions usually needs to pass through an intermediate layer in order to achieve the desired fluidity of movements and reactions (Fig. 1.5). This extra level between action planning and motion generation, is responsible for smoothing the agent's overall behaviour by applying motivational forces directly to the underling physical model. Therefore a reactive middle layer provides a suitable solution for filling the gap between two consecutive actions, generating a net continuous fluid behaviour. This approach is particularly well suited for modelling unconscious reactions and motion dynamics in social interactions, where the contextual territoriality places behavioural constraints on an avatar reactivity building an illusion of awareness.

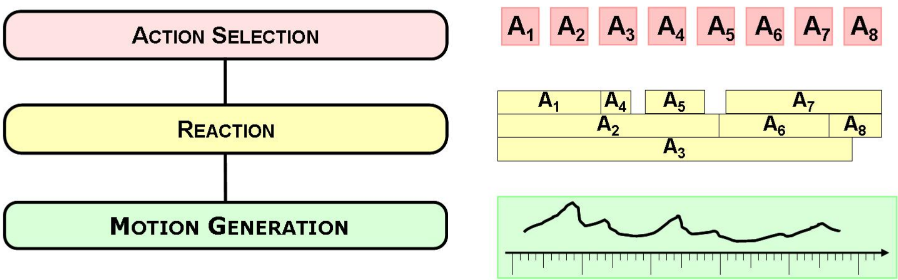

Figure 1.5: An abstract diagram of the framework's architecture. The Reaction middle layer provides an interface between Action Selection and Motion Generation supporting the transformation of the sequence of discrete actions A_1, A_2, ..., A_8 into continuous motion. Each action activates a reactive behaviour 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.

Chapter 2

Related Works

To know the road ahead, ask those coming back.
Chinese Proverb

When we first approached the problem at hand we followed a classical rule of thumb which worked well in many other occasions: to start from a particular case, understand its deep implications and then move toward a general solution. Originally our goal was to come up with a prototype to simulate the dynamics of a group of individuals engaged in a conversation, without considering other forms of social interaction such as queuing in front of an ATM or chilling out at a bar table. We assumed conversations taking place in an open setting, such as a square or a park, and we assumed all the individuals being equally friends or at least acquaintances, to avoid situations in which a potential newcomer was not welcome to join or there was a different level of affiliation amongst the members of the interaction. Simulating a conversation between acquaintances in a open space is simpler because its territory is not influenced by objects in the surrounding environment or by interpersonal relationships. In a way is a pure form of territory and therefore suitable for investigating the properties of a first prototype simulation. Surprisingly we found very few works about simulating small scale group dynamics, especially in conversations.

Another reason why we chosen to simulate conversations first, comes from the strong research tradition of CADIA Research Lab. Some members have been doing research for more than a decade in the field of conversational agents producing relevant contributions for the scientific community. Despite the numerous results achieved in modelling and simulating turn taking or production of communicative content, the problem of how individuals in conversation react to each other from a spatial perspective was never been faced before. At this point is important to remember that human territoriality is just one aspect of the complex picture which describes group dynamics in conversation. Many other aspects should be taken into account to fully simulate people engaged in a face-to-face group interaction such as personality, emotions, goals, affiliation, social status, involvement, context awareness, facial expressions, gesturing, posture shifts, comprehension of spoken content and even the state of the surrounding environment. Each aspect directly or indirectly influences the actuation and meaning of a communicative behaviours and could provide a relevant contribution to the overall group dynamics. Nevertheless in the vast world of conversational agents, human territoriality is still an undiscovered land which can shed some light on how we interact with each other or, quoting Scheflen (Scheflen, 1976), on "How We Behave in Space-Time".

What follows is a overview of the scientific works most related to the present. We will start from describing what it has been done so far in Automating Avatar Control in conversations. Then will move toward simulation of group dynamics talking about two of the main approaches in this field and in particular about Reynolds Steering Behaviours approach which has been of great help for our purposes. Finally we will give credits to the work of Jan and Traum in Dynamic Movement and Positioning of Embodied Agents in Multiparty Conversations (Jan & Traum, 2007) which has been so far the closest work to ours and one of the very first sources of inspiration.

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 behaviour 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 behaviour (H. 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 (H. 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 behaviour 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. 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 (Friedman et al., 2007). Since locomotion and positioning is not being naturally integrated into the interaction when relying on explicit control, it is worth exploring its automation as well. Some have suggested that once an avatar engages another in a conversation, a fixed circular formation should be assumed (Salem & Earle, 2000). This is a simplification, as 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. Therefore we decided on an approach that combines a higher level organizational structure (e.g., that of conversations) with lower level reactive behaviours, affected by a social motivational force field inspired by work on dynamic group behaviour in embodied agents. We focus on continuous social avatar positioning in this current work, while keeping in mind that other layers of behaviour control introduced in previous work will need to be added for fully supporting the social interaction process.

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 with 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 of group dynamics 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 behaviour. 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 behaviour.

The main difference between them is on how sophisticated we want the behaviour of each single individual. The first approach focuses more on the collective group behaviour as a whole whereas the second focuses of the richness of behaviour of each single individual. As we will see in the next chapters, the approach we chosen for simulating group dynamics in conversation is a combination of both.

2.2.1 Large Scale Group Dynamics in Crowd Simulators

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 in Social force model for pedestrian dynamics (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". In their model of dynamics motivations are generated and influenced by external perceived stimuli and personal aims of an individual. If a pedestrian wants to reach a location in the environment then the destination generates a motivation in keeping a certain desired walking direction. External stimuli like other individuals getting closer or the approaching of walls and obstacles, generates motivations as well which the model sum up together to result in a final social force which is interpreted as a desired velocity. Finally, the desired velocity is translated into a physical force acting on the particle acceleration which realizes a motivated action of moving.

An interesting extension to the basic social force model has been introduced by Couzin et al. in Collective Memory and Spatial Sorting in Animal Groups (Couzin, Krause, James, Ruzton, & Franks, 2002), which uses a slightly different approach to model schools of fish. They define three concentric proximity zones around a particle where each zone exerts a different force on the particle's constant velocity, modelling a particular behaviour. These forces are prioritized, that is an higher priority motivation suppresses any lower priority ones. For example, keeping a minimum distance from the neighbors should be more important then keeping the same orientation with them. Avoiding collisions should be the most tightening motivation and therefore it will have the highest priority. What is very interesting in this approach is the relationship between priorities the individual's proxemics, defining essentially a more sophisticated way of mixing up motivations than just a simple sum. 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 behaviour in panicking situations. The model is meant to simulate an evacuation of people from a building when safety is in danger. The line formation is achieved by creating an area of influence in front of each individual used by a waiting behaviour. In a normal situation without stress factors, each individual will slow down and eventually stops if somebody steps into the influence area. The influence area disappears when the individual goes in panic. In that case it will start pushing slower individuals to get its space in front of the group.

A social force model is not the only way of modelling crowds of people. Heigas et al. (Heigeas, Luciani, Thollot, & Castagne, 2003) use a more sophisticated approach to simulate some specific emergent human crowd behaviour. They 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 (Helbing, Molnar, & Schweitzer, 1994) not always fluid dynamics can correctly models individuals interactions which can be better described with a gaskinetic model. Another different approach is the one of Treuille et al. in Continuum crowds (Treuille, Cooper, & Popovic, 2006) which presents a model for crowd dynamics continuously driven by a potential field. Potential field techniques are very similar to the social force model with the difference that only one global force field is computed, considering the group of individual and the characteristics of the environment. 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 different approach consists in recording a

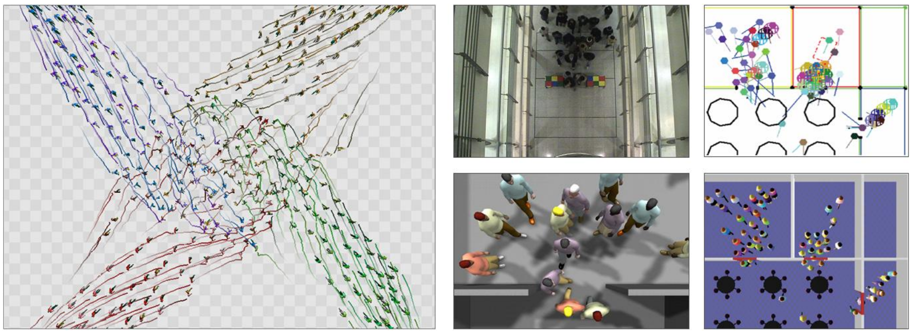

Figure 2.1: Some shots from the Crowd Simulators cited in this section. On the left, Continuum Crowd and the emergence of a vortex when four lanes of pedestrians intercross. This emergent behaviour is a good evidence of the accuracy of the model. In the middle, couple of pictures from the data driven approach of Lee et al. Above is the recorded real crowd and below its simulation. On the right, last set of pictures for the work of Pelechano et al. showing a 2D and 3D view of an high density crowd. All images are courtesy of the authors.

crowds and then directly replicate the phenomenon using the recorded data. This is for example the approach taken by Lee et al. in Group behavior from video: a data-driven approach to crowd simulation (Lee, Choi, Hong, & Lee, 2007). 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 the single individuals. Even though the approach is data driven, it is still classifiable as particle-based. The individuals are controlled by a central agent model which steers their low-level motion changing their position dynamically. Interestingly, the work addresses also 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.

In conclusion, particle-based systems simulates well the low-level behaviours which continuously induce a reaction and produce a smooth movement of the particle used to model the behaviour of individuals in large groups. They also offer good performance and scalability but they put more emphasis on the large-scale collective phenomena then on modelling the micro-dynamic of interpersonal interactions. Their goal is to simulate the crowd as a whole and therefore the single individuals usually lack in intelligence and sophistication. For example, they rarely express context awareness and usually the social orientation of the body is not considered by the model of dynamics. Forces are acting on particles changing their velocity but not directly controlling their orientation. In the dynamics of small scale groups the body orientations, for example the orientation of head or torso, are quite important to determine the membership of an individual to a groups rather than another. Scheflen (Scheflen, 1976) recognizes four regions of the human body a person may use to state his or her membership to an on going social interaction. Therefore in order to simulate small group dynamics such as people in conversation, also the evolution in the orientation of these four body regions must be simulated.

2.2.2 Agent-based Approach to Group Dynamics

As we saw after the last paragraph, a particle-based approach is not enough to simulate small scale group dynamics. If we think about people in conversation is not hard to imagine behaviours which cannot be simulated if we simply model an individual as a particle. For example, participants during a conversation are used to gaze each other, change their posture, gesticulate and show a rich repertoire of facial expressions for example. Even though our goal concerns the simulation of human territoriality and not the whole spectrum of human behaviours in social interactions, our solution have to fit naturally into an architecture which integrates several modules responsible to simulate different classes of behaviour. This general architecture is the one we showed in the first chapter (Fig. 1.1) and it concretely realizes an agent-based approach in simulating avatars in a social virtual environment. The advantage resides in the possibility of selecting actions based on behavioural rules. In general these rules can be of any kind but some of them will model the norms of the social interaction in which the agent is engaged, and this is important if we are aiming toward agents which have to show a certain degree of social context awareness. Controlling the agent through behavioural rules is definitely what we need, however a pure agent-based approach to simulate group dynamics is not completely satisfactory for our purposes. To understand why, we can describe some works which exploits an agent-based approach to simulate group dynamics.

Figure 2.2: Some pictures taken from the works cited in this section. On the left, Autonomous Pedestrians with its reconstructed old Pennsylvania Station, where a population of autonomous agents walks around and interacts with the environment. In the middle, couple of shots from the simulation of Rehm et al. which shows interacting agents. The couple above is having a lovable conversation whereas the one below is arguing. Notice the different arrangements based on the agents interpersonal relationship. On the right, two pictures from the work of Thalmann et al. which show a synchronized group reaction and a park situation. All images are courtesy of the authors.

Thalmann et al. in Hierarchical Model for Real Time Simulation of Virtual Human Crowds (Musse & Thalmann, 2001) use complex finite automata to determine the behavior of actors. These automata represent intelligent autonomous behaviors defined by a set of clever rules. The purpose of the model is still to simulate human crowds but this time the introduction of structured behaviour of groups and individuals is remarkable. A hierarchical model describes the behaviour of each part, but still the set of norms of social interactions such as conversations are not taken into account. Each agent is aware of belonging to a group which belongs to a crowd, but there is no mention of conversations or other forms of social interaction. Very comprehensive is also the work of Shao and Terzopoulos in Autonomous Pedestrians (Shao & Terzopoulos, 2007), which presents fully autonomous agents interacting in a virtually reconstructed Pennsylvania Station. Worthful is the integration of motor, perceptual, behavioural, and cognitive components within a single model. Apart 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 environments. Their work is a good combination of particle and agent based approach, where a sequence of basic reactive behaviours are executed in a certain optimal order, generating a resulting motivation to apply to the physical model of each pedestrian. The basic set of reactive behaviours relies on Reynolds Steering Behaviours which we will describe in the next paragraph. What is not considered in Autonomous Pedestrians is how to constraint an agent reactivity beside the special situation of walking down the hall of a train station. In fact, motion dynamics for conversations and social interactions in general are not taken into account. For example, is not clear what would happen when a group of agents engaged in a conversation must rearrange due to external events like the approaching of a newcomer. In such situations we expect that the formation of agents will break for a while to reform right afterwards. This special case of dynamics needs a different set of reactive behaviour to be simulated as well as the agent awareness of its social context. For example, Rehm et al. in Let's Come Together - Social Navigation Behaviors of Virtual and Real Humans (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. The focus of their work is on modeling how the relationship changes over time based on simulated personality and a theory driven classification of small group interactions. Their underlying model provides a representation of the interpersonal relationships in focused social interactions and how they influence distances and orientations in small groups of two participants. 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). Same for the work of Anthony Guye-Vuillème in "Simulation of Nonverbal Social Interaction and Small Groups Dynamics in Virtual Environments" (Guye-Vuillème, 2004) that again focuses on the group dynamics of the interpersonal relationship and social identities of focused encounters. The simulation of the basic dyadic formations described in (Ciolek & Kendon, 1980) is of particular interest, especially the outlining of an intelligent approaching behavior.

An agent-based approach can combine behaviour generation at several levels of abstraction. Starting from perception, an agent goes through some reasoning process which will end up selecting some actuator behaviour. These action generation loop produces in fact a sequence of behaviours, which turns to be a possible downside of a pure agent-based approach. Human behaviour is not discrete but rather continuous. One of the main advantage of the particle-based approach is just in the continuity of the simulated behaviour, which looks quite believable once animated. A sequential generation of behaviours 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 behaviours can be decomposed in lower level, more fine grained, behaviours until they will eventually merge into a continuum of behavioural control. For this reason we believe that the best approach for simulating social interaction territorial dynamics is to use a combination of agent-based and particle-based, where a set of reactive behaviours 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 Behaviours has been very helpful.

2.3 Reynolds Steering Behaviours

In 1999 Craig Reynolds presented a simple and elegant solution in Steering Behaviors For Autonomous Characters (Reynolds, 1999), to allow simulated characters to navigate around their world in a life-like manner. If necessary we could classify his approach as agent-based since a set of controlling behaviours as well as an abstract agent architecture are proposed, however these behaviours continuously generates control signals in a similar way a system controller does. What makes the difference from a classical set of motor controllers is that steering behaviours are not explicitly acting on the character's mean of locomotion. Instead, they generates 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 together before being applied to the body. The composable nature of the steering behaviours 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'. The schematic architecture for motion behaviours proposed by Reynolds defines three layers, where the upper layer decides goals and creates plans to accomplish them, a middle layers drives the character's embodiment by application of steering forces and the lower layer models the character's mean of locomotion. Since the steering behaviours layer is in between action selection and motion generation, it is by all means a concrete interface over the character's model of locomotion. No matter in what way we model the character's movement, we will still control it through the same set of commands like *turn left*, *turn right*, *go slower*, *reach the destination* and so on. The application of forces resembles in a way the particle-based approach and indeed Reynolds' idea has been used to simulate crowds as well. For example in Big Fast Crowds on PS3 (Reynolds, 2006) the same Reynolds shows how to simulate a school of fish on the parallel architecture of a PS3 using a variant of his *boids* flock behaviour model (Reynolds, 1987). Instead, Erra et al. in Massive Simulation using GPU of a distributed behavioral model of a flock with obstacle avoidance (Erra, Chiara, Scarano, & Tatafiore, 2004) show how to run the same flock behaviour on a GPU, proving how steering behaviours can considerably run faster taking advantage of the modern graphics hardware. The Reynolds' approach is halfway between particle-based and agent-based. In fact, forces are indeed applied to the character's model of locomotion but such model is interchangeable and can be different from a simple point-mass. Moreover his approach present a whole behavioural architecture in which runs multiple behaviours and mix them together.

When multiple commands are called at the same time, they generates steering forces which have to be blended. The simplest blending schema consists in just summing up the steering vectors. A possible and quite recommendable extension is to linear combine the steering forces by multiplying them for a weight. The weight is meant to be a measure of the importance of a certain behaviour in a blended composition. For example, if we combine *reach the destination* and *avoid obstacles along the way*, we expect the latter to have an higher weight then the former. Even using weights to linearly combine steering forces, the blending schema suffers of a common problem called cancellation. Basically, it just can happen that the combination results in a null steering force because for example a behaviour wants to steer to the right whilst another wants to steer to the left. To avoid this problem, Amor et al. suggest a more sophisticated blending schema in Fast Neat and Under Control: Inverse Steering Behaviors for Physical Autonomous Agents (Amor, Obst, & Murray, 2006). The authors recognize that the problem of blending behaviours together resides in their intrinsic definition. They are simple and specialized behaviours which, even though possible to merge together, they still make decisions independently from the others. Sometimes happen that a combination of behaviours leads to a suboptimal resulting 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 a in wall. Amor et al. propose what they called Inverse Steering Behaviours, a variant of the one proposed by Reynolds. Basically, 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.

From their experimental tests, Inverse Steering Behaviours perform better in reaching their goal than Reynolds Steering Behaviours. However, we have implemented Inverse Steering Behaviours in our system and they do produce better results, especially with obstacle avoidance, but at the price of 15% more of computational costs. Of course this is just a speculation since we don't have data to present here, but running the same behaviour for every sampled direction and then perform a heuristic to choose the cheapest one, in principle sounds more computational expensive. Moreover it is not clear how many sampled directions are necessary to create good behaviours while keeping low the computational costs. Some behaviours may require a large span of directions but with a very low sample resolution, whereas other exactly the converse. Finding a compromise seems to be difficult at the moment. Despite these considerations, we will continue using and testing Inverse Steering Behaviours in our system because the idea sounds much promising than many others and the combination of behaviours is the most critical part of the Reynolds' proposed architecture and need to be treated in best way possible.

The original work of Reynolds was meant for character navigational behaviours which does not directly concern the simulation of group dynamics in conversation. Despite this consideration, we can adapt Reynolds' approach to our purposes leveraging on the flexibility of his architecture. From this point of view, our approach can be seen as an application of Steering Behaviours in simulating small scale group dynamics proposing an extension to the Reynolds' seminal idea.

2.4 Group Movement Dynamics in Conversation

At the time we are writing this essay, only two works have been published on simulating movement dynamics of a group of agents engaged in a conversation. The most recent of them is a publication I made in collaboration with my supervisor at CADIA Research Lab. professor Hannes Vilhjálmsson (Pedica & Vilhjálmsson, 2008), to which this thesis largely represents an extension of. The other was published before and was one of the very first sources of inspiration, together with Kendon's opus on Conducting Interaction: Patterns of behavior in focused encounters (Kendon, 1990). Our precursor in this field was the pioneeristic work of Dusan Jan and David Traum in Dynamic Movement and Positioning of Embodied Agents in Multiparty Conversation (Jan & Traum, 2007), which 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 for conversations.

In their work, Jan and Traum for first 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 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 rearrange a group of agents engaged in a situated conversation inside a virtual trainer. Based on its personal believes about who is participating at the conversation, each agent is subjected to a set of social forces which can motivate a variation in positioning. A set of four motivational forces model the tendency of people of moving closer to the speaker, moving away from sources of loud background noise, keeping a minimum distance from other participants and trying to keep a roughly circular formation during conversation. As it has been done for other models, here as well forces are prioritized by rules. For example, if a noise is too loud then the agent will move away from it while keeping a minimum distance from others, but if the voice of the speaker goes too low then it will forget about the noise and start moving closer to the speaker. Once been evaluated, the final motivation is used to compute a new destination point to command the underlying Unreal Tournament Engine[1] to perform a movement.

While the approach looks promising, the main problem is that motivational forces applied to agent orientation are not taken into consideration. Also reorienting is part of the behaviour 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 in general our claim of territory. Such claim should be maintained as long as a member attends to that social interaction. Therefore is important to extend the social force field model in a way that also reorientations can be motivated. Furthermore, we should remember that a conversation is a unit at 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

[1] http://www.unrealtechnology.com/

Figure 2.3: Two pictures from the simulation environment of Jan and Traum. On the left, we can see an agent who wants to approach a conversation. The arrows show the motivation acting on each agent. On the right, a conversation splits in two and the agents rearrange accordingly. All images are courtesy of the authors.

spatial domain delimits a region which casts a behavioural influence not only to the participants but also to external individuals who stop close or just pass by (Kendon, 1990). Since this behavioural influence is common also to other forms of territory (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 take into account also the behavioural influence that any social situation casts in the environment, concretely bounded by its social place.

Chapter 3

Social Theory

The greatest step is out the door.
German Proverb

This chapter covers some of the fundamental concepts of two important theories that shaped this research. We are talking about the works of Kendon and Scheflen, who deserve credit for the valuable contribution to shaping the system requirements of our approach. At the very beginning, Kendon's work on "Conducting Interaction: Patterns of behavior in focused encounters" (Kendon, 1990) was the main influence. His book focuses on the study of face-to-face interactions with a special chapter on conversations. From Kendon's own suggestion, we started looking into the work of Scheflen and his book "Human Territories: how we behave in space and time" (Scheflen, 1976). If Kendon describes the spatial organization of people in conversation, Scheflen goes even deeper by proposing a general paradigm of human territorial organization.

What follows is a quick overview of some of the ideas from the two sociologists mentioned above, Adam Kendon and Albert E. Scheflen. We will start with the concept of F-formation as a way of explaining the arrangement of people in conversation, which Kendon relates to the concept of frame as described by Goffman (Goffman, 1974). Later, we will see how the F-formation fits perfectly into a more general paradigm of territorial organization and classification proposed by Scheflen.

3.1 Face to Face Interaction in Conversation

To define a formal model that approximates the physical group dynamics of a real conversation, we start from the basic concepts of social theory. Like some of the more interesting previous works, our primary inspiration has been Kendon's research on social interaction. In his book "Conducting Interaction: Patterns of behavior in focused encounters" (Kendon, 1990) Kendon presents the concept of F-formation system to explain why people tend to spatially arrange in a particular way when engaged in a social interaction, in which all the participants have equal access and participation rights. Successively, the

F-formation theory has been further explored and extended in (Ciolek & Kendon, 1980), where observational data on how people actually were observed to behave in a public space were presented and discussed. When people pursue a certain line of activity, they do so by claiming an amount of space related to the activity itself they are currently engaged in. This space tends to be respected by other people and Kendon referred to it as the individual's *transactional segment*. Others are able to judge a person's transactional segment from his orientation in the environment, his line of activity, what appears to be the 'feedback cycle' between the person's actions and the things in the environment to which these actions relate. Consequently, its shape and extension tends to vary with the person's body size, posture, position and bodily parts orientation during the interaction. The transactional segment sustains the line of activity as a necessary component of the interaction itself. When outsiders do not respect this segment and intrude into it, the activity may be interrupted and the person who claimed such space reacts immediately to the invasion. Usually people tends to act accordingly to what they think to be the transactional segment of another, continuously adjusting what they are doing so to do not interfere with it.

However, people may establish a joint or shared transactional segment and this is a space that those sharing it have privileged access to. They will act to protect it and that others tend to respect, just as they do with individual transactional segments. It is the system of spatial-orientational arrangement that people enter into when they set up and maintain and preserve a shared transactional segment, called the *o-space*, that constitutes the F-formation system (Fig. 3.2). 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.

The o-space, as a common shared transactional segment (Fig. 3.1), is necessary for the social interaction. All the participants perform a series of behaviours that aim to create and defend the o-space. The set of these behaviours defines a system of positional and orientational reactionary behaviours for maintaining an equilibrium in the system. Typically the o-space is sustained through the appropriate orientation of the lower body. People are not allowed to step into the o-space and participants stand on a ring of space around it called the p-space. In addiction, also a region of space outside the p-space is part of the conversation's domain. Such region defines two concentric zones called r- and c- spaces (Ciolek & Kendon, 1980). Everything inside the c-space is of pertinence to the conversation and participants will cast their gaze to other individuals stopping or just passing by. Whereas, the r-space is smaller and gives higher status to the people inside than just being an outsider. People in the r-space are considered associated with the conversation and this title gives the right to eventually joining the conversation. Interestingly the r-space, because of its proximity to the o-space and the participants, casts a strong behavioural influence on people who stop or pass close by. For example, it is not unusual to see passersby that dip their head and eyes when walking close to a conversation.

Any external or internal perturbation to the system leads to a series of compensational movements to again reach a stable formation. Such movements are meant to sustain the integrity of the o-space which otherwise will break, giving a clear message to the

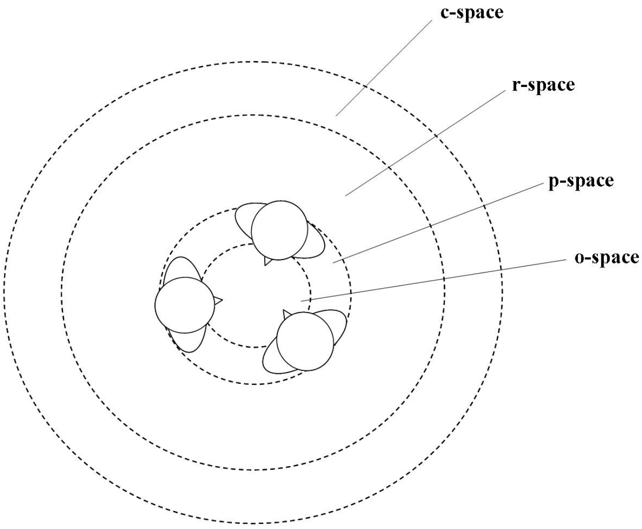

Figure 3.1: A diagram of an F-formation and its space organization. The o-space is the innermost and right in front of the participants, which lies in the p-space. Outside there is a region divided in an r-space and an outermost c-space.

participants about lack of involvement in the same activity. From this point of view then, conversation formation is more similar to a bubble than a rigid circle. However, the shape of the formation strongly depends on the number of people involved, the interpersonal relationship among them, the group attentional focus and on environmental constraints (e.g. the position of furniture inside a room).

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 behaviours are relevant and what conduct is expected in that particular face-to-face interaction. 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 behaviour at the interactional level of organization, not at the individual level (Fig. 3.2).

Since an F-formation system defines a positional relationship among the participants, it is quite natural to take into account the work of Hall and his Proxemics Theory (Hall, 1966). Basically, in his work Hall claims the existence of four important distances between people as they interact. From shorter to longer, these are: intimate, personal, social and public distances. According to the *interaction project* the participants are entering into together, so they will assume a certain spacing. Usually, normal conversations between acquaintances take place in the social area of the participants, while the personal area is reserved for interaction between close friends.

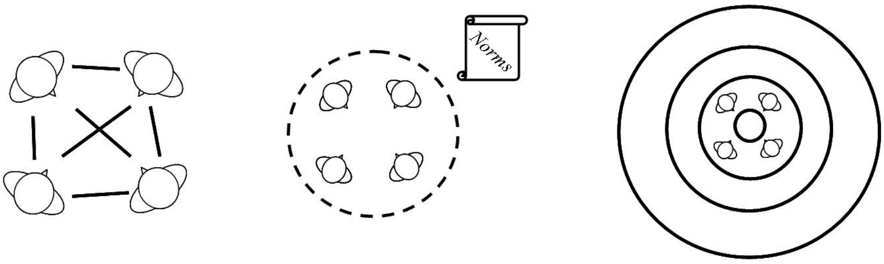

Figure 3.2: A schematic representation of the three main concepts of a conversation as an instance of F-formation. The first picture shows how the participants are involved in a behavioral relationship which sustains a stable formation. The second picture shows a conversation as a unit at interactional 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.

3.2 Human Territories

As a unit at the interactional level where a set of norms defines behavioural relationships amongst members, a conversation delimits its space and defines how it is structured (Fig. 3.2). Scheflen (Scheflen, 1976) describes such a spatial organization as a special case of the more general concept of human territory. Scheflen argues that, when people cluster together, they define a territory which organizes the space around them in a particular structured fashion. The structured space influences and is sustained by a class of specific behaviours called territorial behaviours. This circular relationship between structure of space and territorial behaviours is evident also in the special case of a conversation, where we have compensatory movements for rearranging a formation as mentioned earlier. In general, such a relationship determines a series of movements which concern the position and orientation of a territorial member. Scheflen proposes a classification of territories which ranges from the simplest cases of elements and F-formations, to the more complex structures of gatherings and hubs. This classification is hierarchically organized in a way that simpler territories can be nested in more complex ones. For example, a hub could contain a gathering which is composed of many elements and F-formations which could be further subdivided into smaller elements. This human territorial organization can be marked somehow in the environments where we are accustomed to live or interact, suggesting a certain non-casual arrangement of furniture in rooms. A table surrounded by chairs can be seen as the footprint of an F-formation whilst the seats in a stadium or in a theater are arranged in a classical hub structure (Fig. 3.3).

Scheflen splits the human body in four regions, each of which has an orientation which is communicationally active. People spontaneously use the orientation of the bodily regions are to state their temporary claim on a territory. Usually the lower part of the body is used to state temporary but longer claims, for example the orientation of pelvis and legs during a conversation. Whereas the upper part is generally used to state shorter and highly transitory membership to sub interactions, for example when a participant in a conversation asks another to light a cigarette and both turn head and torso to look at the lighter. Scheflen reintroduced an old English unit of measurement called *cubit* to quantify the

Figure 3.3: Some interesting examples of human territorial organization and spontaneous reaction to it. The first picture shows a typical situation in a subway. In this case, the arrangement and orientation of the seats mark a territory which would suggest a certain degree of affiliation between the girl and the man to her right. The girl's posture, as much as the bag lying there almost as a protective wall, are clear clues of voluntary denying the territorial implicit involvement. The second picture shows a Greek outdoor theater, an example of a general territorial classification which Scheflen named hub. The central region, in this case the stage, represents a common focus point for an audience which is arranged in concentric circular zones. Newcomers are suppose to join through a passageway which typically is an external ring, far away from the central region of the hub.

space required by the human body. A cubit amounts to eighteen inches or about forty five centimeters, and it roughly covers the distance from elbow to elbow approximating the width of an average-sized Nordic body. An amount of four cubic cubits is enough to cover the whole body, where each cubic cubit comprises a bodily region. The space of four cubic cubits of volume around the human body is called k-space and it is a minimal space allocation for a person in stationary posture (Fig. 3.4). Surprisingly, many appliances and small domestic spaces are increments of the cubit and therefore it can be used as unit of space for human territorial organization. The k-space described so far is the minimal space occupancy of a person and as such is usually the space claimed in overcrowded settings. In more normal situations though, each person claims a larger area of space which amounts to four times the k-space, that is, sixteen cubic cubits or four square cubits of ground (Fig. 3.4). This larger area is called *location* and allows enough space to contain the body in several postures without colliding with the neighbors. A location is then four square cubits of ground which is approximately a squared meter of space, therefore covering the intimate zone declared by the Proxemics Theory (Hall, 1966).

Scheflen further proposes a general paradigm of human territorial organization (Fig. 3.5). 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 space and they belong to the nucleus of the territory, while the remaining zones are called r, s and t space and they belong to the region. 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. The re-

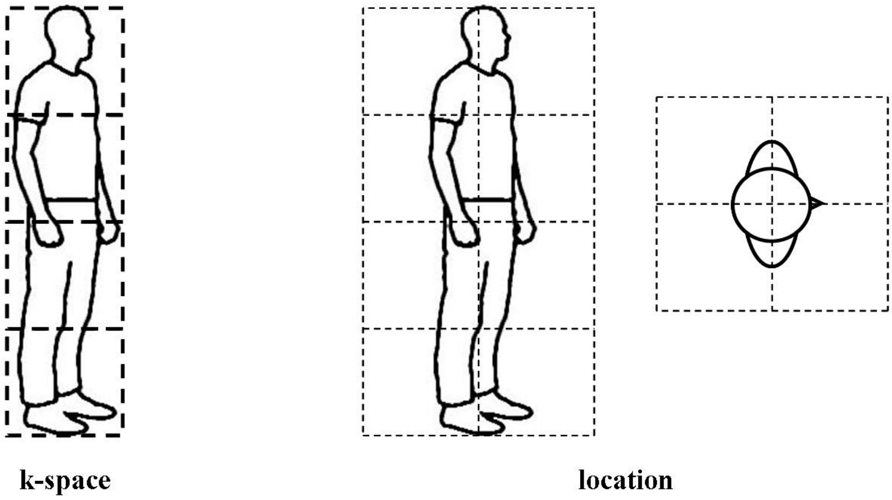

Figure 3.4: On the left, the total amount of space required by the human body called k-space, which amounts to roughly 4 cubic cubits. Each cubit comprises a communicationally active bodily region. On the right, the space of a location from side and top projection. The location is the unit of space usually claimed by people in daily situations.

gion 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.

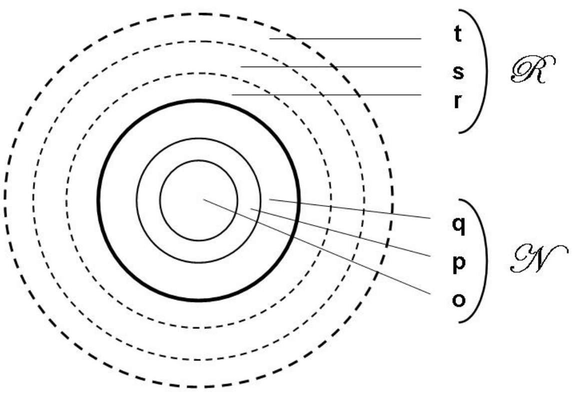

Figure 3.5: The paradigm of territorial organization proposed by Scheflen and applicable to all levels of territorial organization. The o, p and q spaces belong to the nucleus N whereas the r, s and t spaces belong to the region R.

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. 3.6). 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

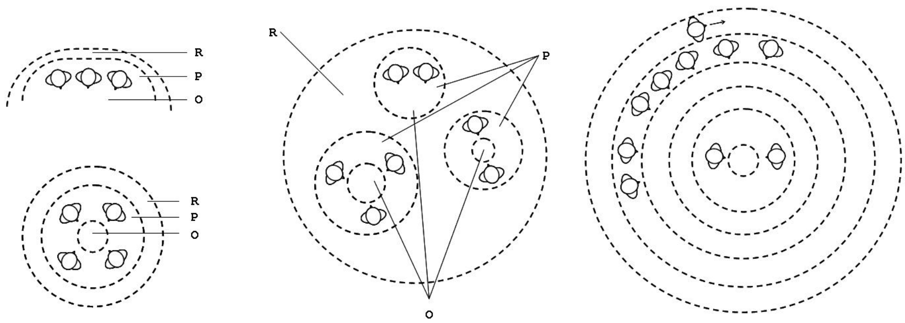

Figure 3.6: From left to right, element, F-formation, gathering and hub. These are the main territorial fields of Scheflen's classification. Each of them can be seen as increment of the previous one for required space and complexity. In addition, more complex territory can contain simpler ones. The O and P in the picture stand for o and p spaces whereas the R stand for an unstructured region.

territorial organization is the face formation, or F-formation, which has been extensively covered above. 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. 3.5). 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. 3.6). 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 crowded of people 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 join and attend the performance (Fig. 3.6).

Chapter 4

Approach

No matter how high the mountain,
to reach the peak you only need to care about your next step.
Anonymous

This chapter will describe our approach in more details. Our goal is to improve the start-of-the-art in the simulation of human social interaction. At the moment, we are not dealing with interpersonal relationships, emotions, affiliation and other important social aspects of social interaction because they are extensively addressed in many other works. What we want to do is to take a step further in the almost untouched ground of simulating human territoriality. As we saw in the previous chapter, when people interact they spontaneously claim an amount of space organized in a certain fashion, which influences the behaviour of participants and outsiders. This behavioural influence is a clear sign of context awareness and we want to simulate it.

Our solution is based on three main pillars which together provide a theoretical and technological background to it. The theories of Kendon, Goffman and Scheflen help in a better understanding of human behaviour and sketch a set of requirements for our technology that we have to meet, for properly simulating human territoriality. Furthermore, the literature on Crowd Simulation suggests a "social force field" for modelling the motivations acting on human reactivity when they have to cope with a dynamic environment. Human territorial behaviour is mostly involuntary and provides a reactive response to a given context. Thus a "social force field" model seems to be a good path to pursue for the simulation of such behaviour. Finally the steering architecture proposed by Reynolds, with an important middle layer between action selection and locomotion, helps to safely incorporate a "social force field" model into an existing agent-based solution for behaviour generation of human multimodal communication.

This theoretical and pragmatic background leads to an AI architecture for our avatars which allows territorial organization and social situation awareness. Considering face-to-face interaction as a social situation, we can delimit its territory determining the border of a social place where the avatars reactive behaviour is influenced by the context. The acceptance of such context involves a set of commonly understood behaviour norms that need to be observed by the participants in the particular face-to-face interaction. Each norm is then a constraint on the avatar's reactive movement when it steps into that social place. The set of constraints is therefore directly mapped into a set of reactive behaviours that generate motivations for the avatar's motion. Finally, this contextually constrained reactivity leads to the net result of an emerging continuous group dynamics as an evidence of the avatar's territorial awareness.

The next sections will cover several aspects of our approach in simulating human territoriality with a special attention to conversations. Even though our architecture proposes a framework which could deal with territoriality in a variety of social situations, at this time only conversations have been simulated and tested. Other arrangements can be added and they will be in the future. This chapter will cover the fundamental contribution of the approach without going into too much implementation detail, which will be reserved for a separate technical report.

4.1 Reaction Generation Paradigm

In our approach, the group dynamics of a simulated social interaction emerges from the avatars' territorial behaviour. We have chosen to simulate class of behaviour 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 an 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 behaviours which motivate an immediate motion to accommodate contingent changes in the perceived surroundings (Fig. 4.1).

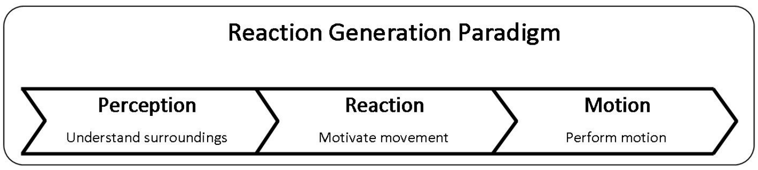

Figure 4.1: Outline of our Reaction Generation Paradigm. The agent understands its surroundings through vision and a sense of proximity. Successively, perceptual stimuli are used for generating motivations to move. Finally, motivations are transformed into physical forces which steer the agentŠs motion.

The reaction paradigm is in effect a loop of continuous control on the avatar's motion. The surrounding environment stimulates the avatar's perceptual apparatus producing information which is later used by the reactive behaviours 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 (Fig. 4.2). 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 on 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.

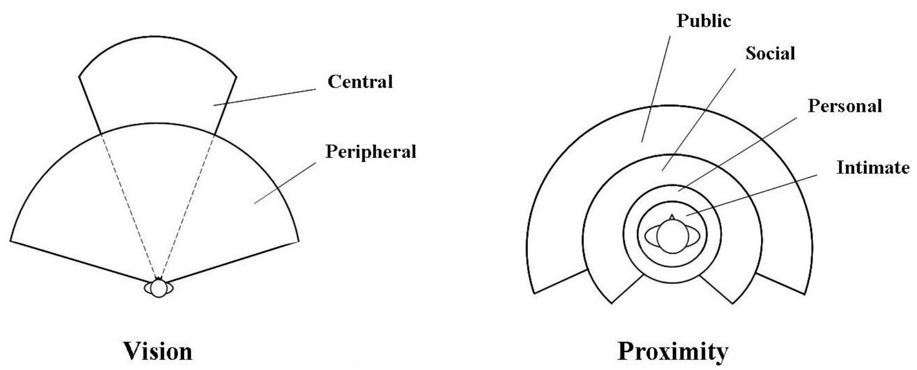

Figure 4.2: Two diagrams showing the spatial structure of the sense of vision and proximity. Notice that proportions have not been respected to make the drawing more readable.

A set of reactive behaviours computes motivations for performing movement. For example, a *keep-personal-distance* behaviour 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 behaviour to stimulate the body at the same time. Several composition algorithms can be chosen for this step. In fact, usually motivations for linear motion 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. So far we have used a point-mass as a model of locomotion, with the added possibility of controlling the orientation of eyes, head and rest of the body. Of course, a point-mass model cannot be accurate enough to simulate human locomotion. The reason for using it in the beginning is for its simplicity and stability, two properties for a physical model which make things much easier during the design and development of a new system. Thus, when considering this motion model, every final motivation is translated into a physical force applied to the body. The translation process strictly depends on the properties of the physical model. In our system, motivations to perform linear motion are translated into steering forces whereas motivations to perform rotational motion are translated into rotation angles. Finally, once the forces have been computed in light of the motion model requirements, they are applied to the body and the motion performed.

4.2 The Social Simulation and Visualization Platform

For conducting this research, we felt we needed a special virtual environment that supported and highlighted the type of information and behavior we wanted to develop. We created a tool called CADIA Populus that combines a fully capable simulation environment with clear visual annotation of the social situations in terms of the theoretical models being used to inform avatar behaviour (Fig. 4.3).

Figure 4.3: Some pictures from the CADIA Populus social simulation and visualization tool.

The avatars themselves are human 'clay' figures with articulated necks and movable eyes, drawn to scale in the environment for accurately reflecting distances. Users can drive their avatars around and strike up conversations with each other using the built-in text chat. A single user can generate any number of avatars and switch between them at will, which is perfect for manipulating the social situation. Our framework for automating the avatars' and agents' reactive behaviour is provided with the tool. Behaviours can be started, stopped or scheduled and multiple instances of the same behaviour are allowed. Running behaviours can be investigated to get a snapshot of their internal status or the status of their profile.

Visual annotations can be inserted through the behaviour code or created on the fly to visually debug the avatar reaction to the environment. Annotations support drawing of lines and shapes to mark the scene with meaningful visual information. Common uses of annotations in CADIA Populus include the highlighting of sensory apparatus, the motivations driving the avatars' behaviours as well as the territorial organization of conversations (Fig. 4.4).

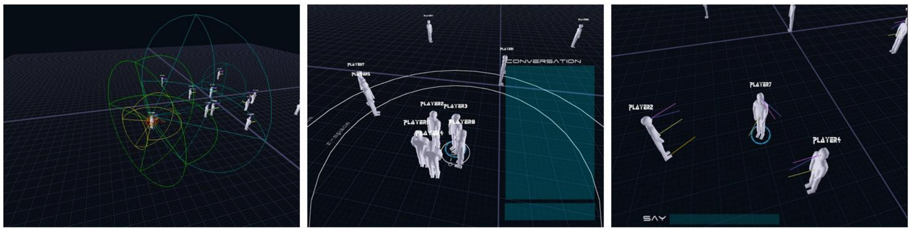

Figure 4.4: Examples of annotations in CADIA Populus. Relevant information can be visually marked into the environment using lines, shapes and labels.

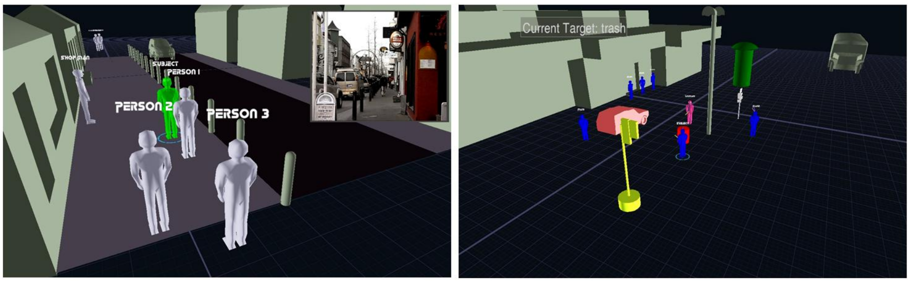

Figure 4.5: Two scenarios created with CADIA Populus to simulate people gaze behaviour in special situations. On the left, a group of people walking down a sidewalk of a main street while, on the right, another group waiting at the bus station. Both examples are simulated reconstructions of recorded real life situations in downtown Reykjavik.

CADIA Populus supports the authoring of various social environments to simulate special situations or debug new avatar behaviours (Fig. 4.5). The tool is written in Python which assures flexibility, fast prototyping and deployment of new scenarios and behaviours. Panda 3D[1] game development library from Canigie Mellon University and Disney is used as the 3D engine powering the graphical component of our simulation environment. Moreover, our tool embeds the NVIDIA PhysX[2] engine to create a fully and reliable physical simulation of the scenario. PhysX allows rigid body dynamics, fast spatial queries and high performance physical simulation.

4.3 Reactive Behaviour Framework

In order to generate reactive avatar response to the environment and to a social place in particular, we designed a framework to manage reactive behaviour and influence it by the knowledge of the social context. Our reactive behaviour framework is based on the architecture proposed by Reynolds in (Reynolds, 1999). Even though the framework has considerably evolved from a Steering Behaviours framework, the original influence of Reynolds is still recognizable in the way our framework manages reactive behaviours. In chapter one we introduced the outline of our approach, showing a diagram of the framework's architecture that we would like to present again (Fig. 4.6). As

[1] http://www.panda3d.org
[2] http://www.nvidia.com/object/nvidia_physx.html

in Reynolds (Reynolds, 1999), the Action Selection in our architecture chooses the next action to perform in order to reach a given goal, whereas Reaction and Motion Generation generalize what Reynolds called Steering layer and Locomotion respectively. The generalization lies in the fact that our reactive behaviours are meant for the simulation of human multimodal communication, whereas Reynolds was concerned only with navigational behaviours. Therefore it should be clear why we used a different terminology. Locomotion in Reynolds deals with modelling and simulating the agent's means of locomotion whereas for us Motion Generation deals also with posture shifts, gaze control, gesturing and facial expression for example. Thus, what Reynolds calls Steering Behaviours layer in ours architecture generalizes into Reaction because we want to address the range of behaviour mentioned above.

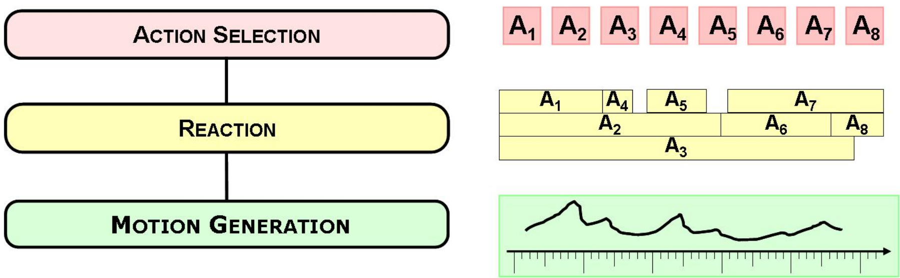

Figure 4.6: An abstract diagram of the framework's architecture. The Reaction middle layer provides an interface between Action Selection and Motion Generation supporting the transformation of the sequence of discrete actions A_1, A_2, ..., A_8 into continuous motion. Each action activates a reactive behaviour 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.

When we look at the framework, the main concept to focus on is Reaction as the middle layer between Action Selection and Motion Generation. As a consequence, the set of reactive behaviours is indeed a control interface on the avatar's embodiment and means of locomotion. Therefore the whole architecture is independent of the specific implementation of Motion Generation and theoretically allows the use of any implementation of it, provided that we reasonably tune the system again. But there is one more advantage of this approach that we would like to point out. The Action Selection layer abstracts all the reasoning and planning processes inside an agent's brain, producing a *discrete* sequence of actions to perform. Without a middle layer, these actions would directly take control of the agent's embodiment, asking it to execute a sequence of discrete movements which would look pretty unnatural. Thus a possible solution is to have a middle layer of extra behavioural control where the actions blend into a composition, generating a continuous stream of motivational control on the underlying Motion Generation layer. The compositional nature of the motivations leads to a society of behaviours where a competition is open to take control over the embodiment. Some reactive behaviours can motivate a movement with greater strength than others and some of them could get so strong that they totally override the effect of a group of behaviours. To sum, the core idea of this architecture is to transform the rigid nature of a collection of behaviours into a more elastic, continuous and fine grained flow of signal control on the underlying Motion Generation.

The framework supports the creation of a library of reactive behaviours which can be executed by other modules or by other reactive behaviours. For each behaviour we can set a priority, a weight, a running frequency, a profile, and a task. Priority and weight are used by the blending schema and priority has a stronger effect than weight. If we consider a set of behaviours having the same priority, then the weight is a factor used by the blending schema to compute the final motivation and basically states how much a behaviour should influence the final result. On the other hand, the priority is used to completely override motivations. Motivations generated by a lower priority behaviour are canceled by the motivations generated by a higher priority one. In other words, the blending schema is applied only to the motivations generated by the highest priority behaviours. A behaviour update frequency can be specified as an extra level of control on the behaviour execution. Sometimes, tuning the frequency can result in a much better simulation and there is evidence that continuous behaviour control in humans is performed at different frequencies, this is especially apparent in the case of impairments such as drugs, alcohol, fatigue and medical conditions (Allen, Marcotte, Rosenthal, & Aponso, 2005). The frequency is specified as a constant time step after which the framework will again call the behaviour's task. The task is a function that realizes the behaviour. There are no specific requirements on the implementation of the behaviour's task but a certain structure emerged naturally. Usually a task starts with some precondition that must be true in order to generate motivations. Then motivations are computed and possibly post processed afterwards. Finally, the motivations are submitted to the blending step. Notice that tasks do not necessarily have to generate motivations and some of them could even start or stop another reactive behaviour. For example, our *keep-personal-distance* behaviour can request a quick glance to the person who enters the avatar's personal space. When the task is called for an update, the framework passes the behavior's profile as a parameter. The profile is a set of variables used internally by the task's code and it is used to define default settings at creation time or to control the behaviour's functionality at runtime.

When a behaviour has been created, it is possible to start it when required. Starting a behaviour creates a request to the framework that will create a behaviour environment to keep track of its state. The behaviour environment includes many attributes and amongst other the behaviour's priority, the weight, the frequency and the profile. Every behaviour's task gets access to the environment of the intended behaviour. Since the behaviour environment is accessible from the task's code, its attributes can be changed along the behaviour's execution allowing, eventually, for the creation of adaptive behaviour which can autonomously change their internal variables to accommodate external contingencies. An example is our *idle-gaze* behaviour, which simulates the behaviour of people casually looking around. The behaviour checks the avatar's states and increases its update frequency when the avatar is walking rather than standing. Notice that the framework is allowed to call the same task multiple times to satisfy different behaviour requests. For example, this is the case for our gaze behaviours. There are three different ways of looking at something in our system: glancing, focusing and gazing. The three behaviours have different settings even though they use the same gaze controller task. The framework allows for multiple requests of the same behaviour or to schedule them setting a starting and ending time, which can be used to generate scripted-like animations.

When there are multiple running instances of a behaviour, sometimes it can be useful to keep a reference to a specific instance in order to explicitly control it. The framework supports tagging for every behaviour instance with a unique name. When the same behaviour is requested again using that name, the framework does not start a new behaviour instance but instead takes control over the one we tagged before. The framework also provides a higher level interface to command the execution of reactive behaviours. For example, the command *move to* lets the avatar move toward a given destination. These higher level interfaces are exposed to external and internal modules and they represent the usual way of executing behaviours. The higher level interfaces also provide an extra functionality called context propagation. Basically, whenever a behaviour is called, it is possible to specify which reactive behaviour called it. In this way, the new behaviour will get the same priority and weight of the caller behaviour allowing different results from the same behaviour called from different behavioural contexts.

4.4 Social Situation and Territorial Awareness

Now it is time to move a step further and describe how 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 chapter three. 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, tell when the context should start influencing an avatar's behaviour. In our system, such behavioural influence is realized by the activation of a set of reactive behaviours, each of which realizes a norm underlying the social situation that the avatar is participating in. 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 behavioural 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 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 behaviour to the ongoing social situation and smoothly blend in. Conversely, an avatar can leave a conversation just simply going away from the nucleus. Moving out of the territory will stop the behavioural influence releasing the avatar from its territorial constrains. This

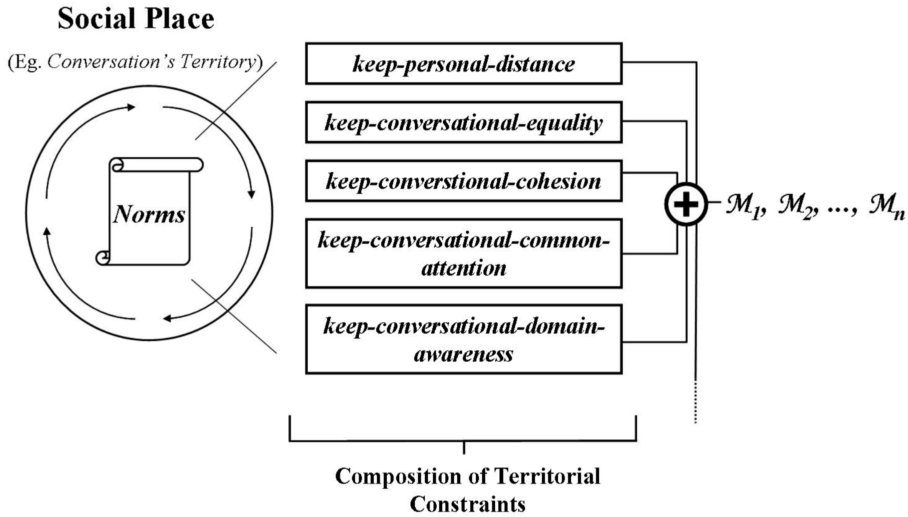

Figure 4.7: A diagram to explain how social situation awareness is realized. Inside the border of the social place, the territorial organization states a set of norms that constraints the avatar's reactivity. Thus, the set of norms maps into a set of reactive behaviours that implements a composition of territorial constraints. The composition blends with other behaviours leading to a behavioural influence marked on the resulting final motivations $M_1, M_2, ..., M_n$.

example is not meant to explain how an avatar should join or leave a conversation, which can be improved, but how the avatar's behaviour is immediately and smoothly influenced by the simple fact of entering or leaving a social place.

4.5 Formation Dynamics in Conversation

As mentioned earlier, as soon as an avatar enters a new conversation, a particular set of reactive behaviours is activated and run. Our purpose is to simulate the compensational movements of the formation when the system is subjected to some fluctuations and for doing so we constraint the avatar's reactive motion using the following set of behaviours: *keep-personal-distance*, *keep-conversational-equality* and *keep-conversational-cohesion*. These behaviours have the same priority but different weights and they generate motivational forces at a given frequency using perceptual information.

The behaviour keep-personal-distance prevents avatars from crowding in a small area, pulling them apart to maintain a minimum distance between each other. When someone steps into the avatar's personal space a repulsive force starts increasing, motivating the avatar to move away from the other. The behaviour keep-conversational-equality constraints an avatar to keep roughly the same distance from the nucleus as the others, stating an equal claim on the shared space. In addition, the behaviour generates an orientation of the body toward the rest of group. Motivational forces increase only if other participants are in the range of the avatar's social distance. Finally, the behaviour keep-conversational-cohesion motivates an avatar to stay close to the other participants in the conversation. Furthermore, the behaviour generates a motivation to rotate the body toward those participants, that for some reason, are further away from the conversation. The motivations computed by these three behaviours are blended together, with keep-personal-distance having the highest weight, followed by keep-conversational-cohesion and then keep-conversational-equality. What follows is a more detailed discussion on how to compute such motivations for a conversation.

To motivate an avatar to keep a certain minimum distance between individuals, we calculate a *repulsion force* as follows (Fig. 4.8).

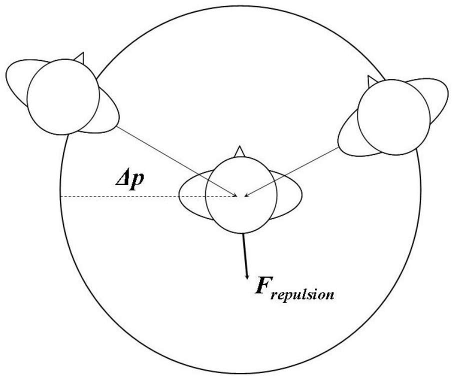

Figure 4.8: A diagram schematizing how to compute a repulsive force to use as a motivation to maintain personal distance. In this case the avatar will back off some steps.

Let N_p be the number of individuals inside the personal area of the avatar, $r \in \Re^3$ the position of the avatar and $r_i \in \Re^3$ the position individual i inside the personal area and Δ_p the personal distance:

$$F_{repulsion} = -\left(\Delta_p - d_{min}\right)^2 \frac{R}{\|R\|} \tag{4.1}$$

where $R = \sum_i^{N_p} (r_i - r)$ and d_{min} is the distance of the closest individual.

To motivate an avatar to sustain the o-space along with the other participant inside its social area, keep-conversational-equality calculates an *equality force*, $F_{equality}$, and a *body orientation*, $D_{sustain}$, as follows (Fig. 4.9). This equality force acts like a repulsion or attraction force, depending on the difference between the distance of the avatar from the center of the group in its social area and the mean of every group member's distance from the same center. Whereas the body orientation keeps the avatar facing the o-space.

Let N_s be the number of individuals in the avatar's social area, and m the mean distance of the members from the group center c. We can calculate the centroid, the mean distance, the equality force and the body orientation as follows:

$$c = \frac{1}{N_s+1}\left(r + \sum_i^{N_s} r_i\right) \tag{4.2}$$

$$m = \frac{\|r-c\|}{N_s+1} \sum_i^{N_s} \|r_i - c\| \tag{4.3}$$

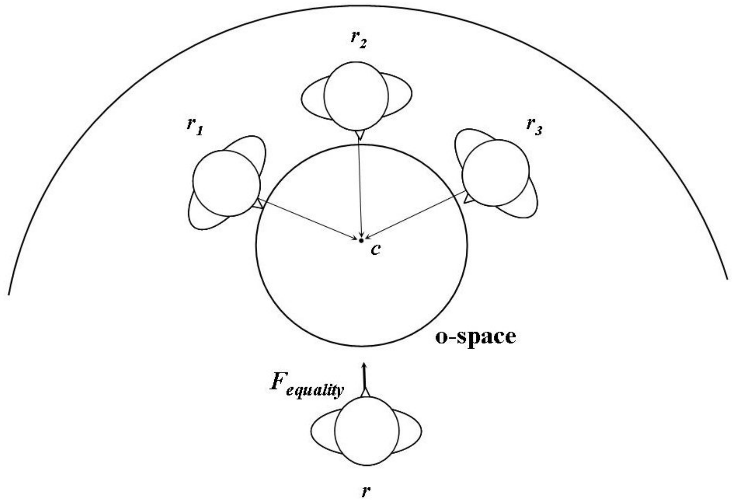

Figure 4.9: A diagram schematizing how to compute an equality force to use as a motivation to maintain an equal distance from the o-space. In this case the avatar will move forward few steps to be at the same distance from the shared space as the other participants.

$$F_{equality} = \left(1 - \frac{m}{|c-r|}\right)(c-r) \quad (4.4)$$

$$D_{sustain} = \sum_i^{N_s} (r_i - r) \quad (4.5)$$

To avoid being isolated, individuals are attracted toward the o-space by means of a *cohesion force*, $F_{cohesion}$, generated by keep-conversational-cohesion (Fig. 4.10). In addition, this behaviour calculates a *cohesive orientation*, $D_{cohesive}$, to face those participants moving further away from the nucleus.

Let N_a be the number of individuals in the avatar's public area, $o \in \Re^3$ the center of such group and s the size of a cubit. Then we can calculate the cohesion force and cohesive orientation as follows:

$$\alpha = \frac{N_a}{(N_s + 1)} \quad (4.6)$$

$$F_{cohesion} = \alpha \left(1 - \frac{s}{\|o-r\|}\right)(o-r) \quad (4.7)$$

$$D_{cohesive} = \sum_i^{N_a} (r_i - r) \quad (4.8)$$

the scaling factor α for the cohesion force is used to reduce the magnitude of this force if the avatar is surrounded by individuals in its social area. The cohesion force is intended to be stronger for those participants far away and isolated from the conversation.

At this point, it must be noticed that a set of three behaviours is not enough to properly simulate group dynamics in conversation. In fact, our model should be seen as an example of how to map the norms of a social situation into a set of reactive behaviours and not

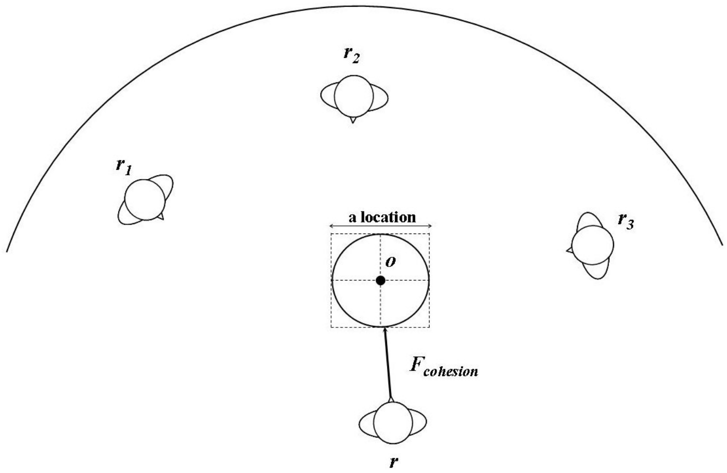

Figure 4.10: A diagram schematizing how to compute a cohesion force to use as a motivation to keep the group of participants united. In this case the avatar will move some steps forward to get closer to an area of the size of a location in the middle of the group.

has the ultimate model for small group dynamics. Differently from the approach of Jan and Traum (Jan & Traum, 2007), where the agents get closer to the speaker if they cannot hear it well, our model defines a social force field based only on the avatar's proximity perception. Of course getting closer to the speaker or moving away from sources of loud background noise are important motivated movements which will serve to fully simulate small group dynamics, but not necessarily they can suffice for the scope. Conversations are dynamic formations even when nobody is speaking. People are repositioning and re-orienting themselves also to show their membership to a specific social interaction, which states an higher level of involvement amongst the participants. People communicate such a higher involvement through a complex and refined orientation of bodily parts and space management where proximity, or distance amongst participants, plays an important role. From this perspective, our model shares many similarities with the flocking behaviour of Reynolds (Reynolds, 1987), in which three different behaviours are acting on three progressive larger distances. The novelty introduced by ours is in postulating a relationship between the reactive behaviours and the four zones of the Proxemics Theory (Hall, 1966). A relationship that has never been proposed before but that sounds pretty reasonable considering the behavioural categorization based on the different human proxemics zones.

Chapter 5

Evaluation

The disciple is not above his master;
but everyone, after he has been fully trained, will be like his master.
Bible, Luke 6:40

Evaluating a software system is always a sensitive task. Depending on the point of view when we look at the system, we find different properties to evaluate. The problem is what set of properties best describes the contribution of the system when used in practice. There is also the issue of what can be realistically evaluated of the technology at this point in time. Of course the technological contribution is not mature enough to be fully evaluated with respect of to its full potential. Nevertheless we know that believability is the ultimate quality we want to evaluate.

Believability is not a system property that describes the quality of a system design or an architecture. Believability depends on people, on their sensations when using the system. As a consequence believability is evaluated on what the users can experience with the system, that is on the output generated by it. Do our avatars look more believable when compared to the state of the art? This is the question we want to answer when evaluating believability. The answer will speak to the practical contribution of our technology. Remembering what was said in chapter one about addressing the "uncanny valley" problem for game environments, evaluating the believability means to understand whether the contribution succeeds in merging high quality graphics and simulation standards with the generated continuous avatar's motion. We cannot yet evaluate believability under these terms however, because the continuous avatar motion has not been linked to a graphically rich environment yet. The collaboration with CCP will cumulate in such a demonstration further down the road, which will make evaluation from this point of view possible. For this reason believability will not be formally evaluated at this time, but a plan for the evaluation will be presented. Some informal description of the visual results produces by the social simulation platform powered by our technology will also be presented.

Even though believability is the ultimate quality to evaluate, it is also important to focus on the technical evaluation of the contribution. The architectural qualities of the technology are mainly a heritage of the Reynolds Steering Behaviours architecture that was used a source of inspiration. In particular, several specific properties of the proposed system architecture for reactive behaviours need to be evaluated: scalability, extensibility, reusability and flexibility.

5.1 Considerations on Believability

As mentioned earlier, the contribution needs to be evaluated in terms of believability of the output, that is how well the user recognizes that an avatar is aware of the social context. This definition of believability is hard to evaluate precisely and, in our case, strongly depends on the quality of the generated avatar or agent motion, which is a problem of low level body animation and besides the scope of this research. To address this, a future evaluation should examine whether the new technology increases the believability when compared to the state of the art. Comparisons must be done using two instances of the same overall software system in order to avoid biased results due to irrelevant technological disparities. In this case, state of the art techniques should be implemented as part of the baseline system for the comparison. Independent human subjects should be used to rate the believability of the outcome.

While this evaluation of the effectiveness of the approach for multi-player games has not been performed yet, the visual results and informal interaction tests support the direction chosen for the approach. When observing a number of avatars in the system, one immediately gets the feeling that conversation groups are robust elements, that organically maintain sensible formations as players bring their avatars in and out of contact with other players. Additionally, social perception and reactive maneuvering seems to give the avatars a heightened level of behavioural continuity and a greater relevance to the ongoing social interaction, potentially avoiding the dreaded disconnect between the visual environment and the classic text box for chatting. Of course, we are still far from simulating a realistic conversation dynamics due, for example, to lack in animation sophistication. Avatars are sliding around instead of walking, they do not change posture during the conversation and their shoulders are unrealistically stiff resulting in a limited gaze excursion compensated by unexpected full body rotations. Moreover, the dynamics appear to get unstable when more than six or seven individuals join a conversation. We believe that a proper tuning of the territorial behaviours can stabilize the dynamics but also being aware that a higher number of participants may split the conversation into multiple sub-conversations, and our model should account for such possibility in the future.

5.2 Scalability

Scalability is often a desirable quality for AI architectures which in our system relates to the number of avatars and agents than can populate a scene. In that sense, we are interested in the size of the simulated situation. As a minimum requirement, our architecture should be scalable enough to easily manage scenarios with tens of individuals in continuous interaction. Imagine a simulated situation like a crowded bar or a conference room during a meeting. In the long term, our solution should eventually scale up to, or at least easily integrate with, a crowd simulator. The idea would be to manage both large and small scale group dynamics by extending, or composing, an existing crowd simulator with our technology supporting the simulation of a massively crowded environments where social interactions take place. We know that our architecture is scalable because it more or less represents an instance of a Steering Behaviours architecture, that proved to scale up to thousands of individuals (Reynolds, 2006) and which can potentially take advantage of a GPU to achieve an even better performance (Erra et al., 2004). The integration of Steering Behaviours with other crowd simulation techniques has been shown in (Treuille et al., 2006), where Reynolds agents were mixed with a different sort of agents driven by a dynamic potential field.

What has the greatest adversarial effect on the scalability of our implementation is the avatar's perception of the surroundings and the number of running behaviours per avatar. Perception can be considerably improved using a spatial sorting data structure as shown in (Reynolds, 2006) and (Shao & Terzopoulos, 2007). For example, in our social simulation platform we chose to integrate a physics engine and to use its fast collision detection to implement the physical interaction between a sensory apparatus and the world. Our perception system is based on spatial filtering to spatially reduce the amount of information about the surroundings an avatar can access at every point in time. This process of spatial filtering based on the avatar's perceptual abilities can be considerably sped up taking advantage of the spatial sorting data structures usually implemented in a physics engine to perform fast collision detection. Regarding the number of running behaviours it is important to realize that not all of them need to be running at the same time. If a conversation takes place in a location far away from the camera then we are not really interested in producing facial expressions, posture shifts or glances to passersby for example. A level-of-details (LOD) technique could be use to reduce the number of running behaviours when they are not essential, leveraging on the composability of our reactive behaviours framework. The behaviours of the greatest importance are the ones which define the territorial field of a social interaction, because they will determine the position of the whole body in the scene. Of course, all other social behaviours will considerably improve believability, but only provided that the camera will be close enough to the group to let the user detect every subtle movement of the body.

5.3 Extensibility and Reusability

The most natural way of extending our design is to add new reactive behaviours and social situations. The reactive framework supports the creation of behaviours as new system components that can be shared with the research community. We envision a library of behaviors that keeps growing.

The creation and deployment of new behaviours and their various combinations needs to be straight-forward to assure extensibility and reusability in the architecture. This needs to be evaluated with respect to the use of possible third parties, and here the close collaboration with a group of actual game developers will be helpful. So far other students at CADIA have been working on extending the set of behaviours of the reactive framework. For example, Angelo Cafaro and Raffaele Gaito worked on an *idle gaze behaviour* that simulates people's gaze shifts when no particular task has to be accomplished. They worked on their implementation independently with very little supervision. The framework provided a solid base for the creation of the idle gaze behaviour while placing few constraints on its implementation. Only few one-time changes were necessary in the framework but all of them fell naturally into the boundaries of the original design. Also the tuning of the idle gaze has been quite easy but it is important to realize that our library contains about twenty behaviours which is a fairly small quantity. Moreover the idle gaze generates orientational motivations and these are much easier to tune than motivations for linear motion. In fact, the tuning of a new behaviour is probably the most critical step when using our system. It could turn out to be a time consuming task which rapidly grows in complexity with the number of available behaviours in the library. To fully address extensibility the framework should provide some level of support for tuning (e.g. a tool based on reinforcement learning) so to ease or even automate such a potentially complex task.

Choosing the right tunning is essential to achieve a stable dynamics but sometimes this is not enough. In fact badly designed behaviours will hardly produce a stable dynamics. We noticed that a good style in designing behaviours starts to emerge but it is to early to derive conclusions about it. To foster a good style in designing new behaviours for our framework we could provide a set of standard building blocks for the implementation, while keeping in mind that a programmer should be constrained as little as possible. Such a set of building blocks could help the behavioral designer to implement good behaviours and promote reusability in the architecture. This could be supported by a graphical user interface similar to those used for the designers of visual effects. At the moment, the framework provides a set of *steering functions* to easily compute motivations inside a reactive behaviour. Steering functions are mapped onto the set of Reynolds Steering Behaviours (Reynolds, 1999), which are designed with a "building blocks philosophy" and, therefore, meant to be reusable. For improving reusability even more, our architecture support attaching the same behaviour task to multiple behaviours. For example, we have three different type of gazes in our system. Each of them realizes a different way of looking at people, but all of them use the same gaze motor controller just with different parameters and different weight or priority. The final mark in terms of reusability supports the specification of new reactive behaviours in separate files than can be exported from one system and imported into another. But for that, we have to find an efficient and effective solution to the tuning problem.

5.4 Flexibility

The reactive framework supports the creation of new social situations that avatars and agents can become a part of or be under an influence from. These social situations and reactive behaviours are expected to depend on the application. Therefore an important part of the technical evaluation is to see how flexibly the system copes with different application domains. Good flexibility will contribute to greater longevity of the implemented solution. Since our architecture presents a certain degree of extensibility, as a consequence it is also flexible enough to stretch and accommodate possibly new required features.

The architecture has prove to successfully manage both territorial and navigational behaviours within the same framework. This is a heritage of the Steering Behaviours architecture, that was originally meant for navigational behaviour but we extended for general reactivity. Also the motivational middle layer improves the flexibility of the whole design since we can apply motivations to an avatar's embodiment at different levels of control. For example, we could have a motivation for controlling the gaze direction or a set of motivations for controlling the orientation of shoulders, head and eyes. We could go from motivating the whole body to motivating the movement of a single finger, presenting to the higher layers of the system an interface with a spectrum granularity. Although we would prefer the use of an animation engine for the finer grained control of the low level body movements, having such a broad definition for motivations, should in principle make the integration with an animation system easier in the near future.

Thanks to its technical kinship with the Reynolds architecture, our reactive framework can power several typologies of agents and avatars. Even though we focused on human-like territorial behaviours, the framework supports plugging in different ways of generating motion for agents that are powered with it. Different behavioural controls for the same embodiment are even possible. That means we can use the same technology to simulate the reactivity of vehicles and creatures in general that may have distinct means of locomotion. We could also start experimenting with this feature, mixing up a certain reactive intelligence with uncommon models of embodiment. That aside, what concretely remains is the possibility of creating different typologies of agents within the same simulation platform, which suggests a highly flexible nature of our framework. Imagine a scenario in which a pedestrian wants to cross the road and first looks for passing cars. Our framework could flexibly simulate both kinds of agents, the pedestrian and the car, within the same system architecture. Also the basic design concept of territory is adaptive in nature. Think about the soccer field in a soccer simulation for example. We could model the field as a territory with its peculiar spatial structure and therefore its way of influencing the reactive behaviours of the players. At least in principle, our framework can flexibly adapt to a variety of simulation domains.

Chapter 6

Conclusions and Future Work

God is really only another artist. He invented the giraffe, the elephant, and the cat.
He has no real style. He just goes on trying other things.
Pablo Picasso

6.1 Main Contributions

The primary contribution of this thesis work is in the field of graphical avatars in social virtual environments, considering the MMORPG as a special case. With the advent of even more powerful 3D engines, faster and more robust physics engines and affordable dedicated hardware, virtual environments are rapidly growing into 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 a higher expectation for behavioural AI, which has to perform at the same level of believability as the graphics and physics in order not to break the illusion. It is a problem quite similar to the so called "uncanny valley" (Mori, 1970) in robotics: when visual realism grows, so does our expectation that the virtual world should match our daily experience or imagination. Perfectly human looking robots may therefore look uncanny or disturbing when their behaviour fails to be as perfect as their looks. 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 behaviours which are not suitable for social environments.

This thesis contributes a technological tool for artists and designers of interactive virtual environments, which ultimately allows them to devise social situations in the same way an art director does on stage. Once a particular social situation has been created, all the characters in it will contribute to making the whole scene lively and dynamic with continuous reaction based on context awareness. 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. We believe that context bounded reactivity is one of the key elements for realizing a strong virtual social situation that immediately brings to the viewer an instinctual feeling and knowledge about the intended social activity. Helping the viewer acquiring this knowledge in a natural and non invasive way will improve the social affordance of the virtual environment, enriching in principle the user's experience.

Moreover, this technology is an application of some of the theories of Kendon on face-to-face interaction and Scheflen on human territories. These theories cannot be formally proven correct because of the intrinsic complexity of dealing with human behaviour and because a formal language to explain and reason about it does not exist. We cannot proof theorems on human behaviour. However, an application of their principles using an agent-based simulation approach is a way to realize these behavioural models in practice and consequently to investigate some of their ambiguities, promoting discussions to clarify them. Agent-based simulations can be used as a way of encouraging a closer dialogue between Social Science and AI. In our research, the former helped providing a theoretical framework to explain the human behaviours we wanted to reproduce on a machine, yielding in fact to a set of requirements for the design of our software system. Forcing these requirements into practice, through the phases of design and coding, reveals some obscurities in those theories that deserve some further clarification. For example, we are all familiar with the concept of personal distance. But how do we measure it in practice? Is it a distance between arbitrary points on the body surface of two individuals, or the distance between the center of masses of two bodies? And what if one will lean forward so to have his or her face closer to the other. Should we consider the personal distance to be between the faces or still be the distance between the whole bodies? These are questions we need to answer in geometrical terms in order to build a simulation, but we need a collaboration with the Social Science to get some insights for a possible answer.

6.2 Known Limitations

An obvious current limitation that we plan to address 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 results. Since the framework is totally independent from the implementation we are using to realize the avatar's motion, as we showed in chapter four, a very natural extension would be to plug a whole animation engine into it. Thus Motion Generation 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 move to the left would be translated into a walking cycle toward the destination point while a weaker motivation would result just in a posture shift. Plugging 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 evaluating our technology in terms of its practical contribution to believability, disclosing what we believe to be a technology with good potentiality.

In addition, we will investigate better techniques for combining motivations to avoid jitter and achieve real-time performance with large number of agents and avatars. The blending of motivations is a critical part in our framework. Inverse Steering Behaviours present a good alternative to the more common blending schemes out there, but they were designed especially to avoid the problem of cancellation that turns out to be a desirable property when simulating a group of avatars trying to reach a stable formation. To achieve stability we expect that repulsive and attractive social forces are going to cancel each other after a while. Using Inverse Steering Behaviours to avoid cancellation leads more naturally to the definition of an unstable group dynamics.

In this first attempt of modelling small scale group dynamics and in particular the one of people engaged in conversations, we deliberately omitted a concept of clear importance: the *transactional segment*. This concept, as introduced in chapter three, could be used to define a more general model of dynamics since the spatial-orientational arrangement of participants in a conversation is the result of preserving a shared transactional segment. The problem here is that it is hard to geometrically define the amount of space a person claims when engaged in an interaction and moreover, this amount of space changes over time. Since our model is based on a social force field, we usually start from a geometrical description of the territorial field to implement the territorial constraints which realize the group dynamics. If the conversation takes place in an open space then we know that the shape of the shared transactional segment will be roughly circular, but in general this assumption might be false. Considering the geometric variability of the transactional segment, a good idea seems to be to model it as a gradient to sum up to a contextual overall potential field. Group dynamics driven by a potential field is not a novel as shown in (Treuille et al., 2006). The problem here, is that potential fields are well suited for particle-based simulation so an adaptation toward our agent-based approach would be necessary.

We also want to further investigate the benefits of explicitly representing social situations. We believe we have the right underlying model here to simulate a range of different social situations within the same complex environment. Each situation will have its social place, that is a territory, and there are still many other forms of them beside the F-formation that we would like to simulate. Moreover, we want to investigate the "chines boxes" relationship pointed out by Scheflen about human territories. Social contexts seem to be nested inside each other, a relationship which is not hard to simulate and that will allow situations like standing in line while having a conversation for example. Of course, realizing such a territorial relationship will open up a window on a new problem concerning dependencies amongst behaviours. In our future behaviour society, some of them may be in opposition to each other, such that only one can be running at the same time. If we admit nested territories then it is possible that opposing behaviors are forced together, leading to an erroneous avatar net behaviour. The framework should be improved in a way to automatically dealing with such dependencies. Taking inspiration from the extensive literature on behaviour networks can help us solve this problem.

Finally, as we move up from the lower levels of purely reactive behaviors we continue to add higher levels of intelligent control, incorporating some of the behaviors that we have ready from previous work on Embodied Conversational Agents, while also adding new ones that address social scenarios that the specific MMORPG setting calls for. The user interface will also need to grow and adapt to handling the increased complexity without risking the seamless user experience. This is something we believe automation makes easier, but that is an empirical question we look forward to answering.

Bibliography

Allen, W. R., Marcotte, T. D., Rosenthal, T. J., & Aponso, B. L. (2005). Driver assessment with measures of continuous control behavior. In *Paper presented at the 3rd international driving symposium on human factors in driver assessment, training, and vehicle design.*

Amor, H. B., Obst, O., & Murray, J. (2006). Ai game programming wisdom 3. In S. Rabin (Ed.), (pp. 221–232). Boston, MA : Charles River Media. (ISBN 1-58450-457-9)

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, August 12-17). Beat: the behavior expression animation toolkit. In *Siggraph01* (p. 477-486). New York, NY : ACM Press.

Ciolek, T. M., & Kendon, A. (1980). Environment and the spatial arrangement of conversational encounters. *Sociological Inquiry Nos 3-4*, *50*, 237-271.

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

Erra, U., Chiara, R. D., Scarano, V., & Tatafiore, M. (2004, November). Massive simulation using gpu of a distributed behavioral model of a flock with obstacle avoidance. In *Proceedings of vision, modeling and visualization 2004 (vmv).*

Friedman, D., Steed, A., & Slater, M. (2007, September 17-19). Spatial social behavior in second life. In *7th international conference on intelligent virtual agents* (Vol. 4722, p. 252-263). Berlin : Springer-Verlag.

Gillies, M., & Ballin, D. (2004, July 19-23). Integrating autonomous behavior and user control for believable agents. In *Autonomous agents and multi-agent systems* (p. 336-343). ACM Press.

Goffman, E. (1974). *Frame analyses: An essay on the organization of experience*. Cambridge, MA : Harvard University Press.

Guye-Vuillème, A. (2004). *Simulation of nonverbal social interaction and small groups dynamics in virtual environments*. Thèse de doctorat non publiée, Ècole Polytechnique Fèdèrale De Lausanne.

Hall, E. T. (1966). *The hidden dimension*. New York, NY : Doubleday.

Heigeas, L., Luciani, A., Thollot, J., & Castagne, N. (2003, September 5-10). A physically-based particle model of emergent crowd behaviors. In *Proc. of graphicon.*

Helbing, D., & Molnár, P. (1995, May). 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. *Evolution of Natural Science*, *230*, 229-234.

Jan, D., & Traum, D. (2007, June 2007). Dynamic movement and positioning of embodied agents in multiparty conversation. In *Proc. of the acl workshop on embodied language processing* (p. 59-66).

Jan, D., & Traum, D. R. (2005). Dialog simulation for background characters. *Lecture Notes in Computer Science*, 65-74.

Kendon, A. (1990). *Conducting interaction: Patterns of behavior in focused encounters.* Cambridge : Cambridge University Press.

Kopp, S., Krenn, B., Marsella, S., Marshall, A. N., Pelachaud, C., Pirker, H., et al. (2006, August 21-23). Towards a common framework for multimodal generation in ecas: The behavior markup language. In *Aamas '08: Proceedings of the 7th international joint conference on autonomous agents and multiagent systems* (Vol. 4133). Richland, SC : International Foundation for Autonomous Agents and Multiagent Systems.

Lee, K. H., Choi, M. G., Hong, Q., & Lee, J. (2007). Group behavior from video: a data-driven approach to crowd simulation. In *Sca '07: Proceedings of the 2007 acm siggraph/eurographics symposium on computer animation* (p. 109-118). Eurographics Association.

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* (p. 104-116). Berlin, Heidelberg : Springer-Verlag.

Pelechano, N., Allbeck, J. M., & Badler, N. I. (2007). Controlling individual agents in high-density crowd simulation. In *Sca '07: Proceedings of the 2007 acm siggraph/eurographics symposium on computer animation* (p. 99-108). Aire-la-Ville, Switzerland, Switzerland : Eurographics Association.

Rehm, M., Andre, E., & Nisch, M. (2005, November 30 - December 2). Let's come together - social navigation behaviors of virtual and real humans. In *Intelligent technologies for interactive entertainment* (Vol. 3814, p. 124-133). Berlin : Springer-Verlag.

Reynolds, C. W. (1987). Flocks, herds and schools: A distributed behavioral model. *Computer Graphics*, *21*, 25–34.

Reynolds, C. W. (1999). Steering behaviors for autonomous characters. In *Proc. of the game developers conference* (p. 763-782). San Francisco, CA : Miller Freeman Game Group.

Reynolds, C. W. (2006). Big fast crowds on ps3. In *Sandbox '06: Proceedings of the 2006 acm siggraph symposium on videogames* (p. 113-121). New York, NY, USA : ACM.

Salem, B., & Earle, N. (2000). Designing a non-verbal language for expressive avatars. In *Collaborative virtual environments* (p. 93-101). ACM.

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

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

Treuille, A., Cooper, S., & Popovic, Z. (2006). Continuum crowds. In *Siggraph 2006 papers* (p. 1160-1168). New York, NY, USA : ACM.

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* (p. 477-486). New York, NY : ACM Press.

Vilhjalmsson, H. H. (2003). *Avatar-augmented online conversation*. Thèse de doctorat non publiée, Massachusetts Institute of Technology. (Supervisor-Justine Cassell)

REYKJAVÍK UNIVERSITY
HÁSKÓLINN Í REYKJAVÍK

School of Computer Science
Reykjavík University
Kringlan 1, IS-103 Reykjavík, Iceland
Tel: +354 599 6200
Fax: +354 599 6301
http://www.ru.is