Dynamic Movement and Positioning of Embodied Agents in Multiparty Conversations

Dušan Jan
USC Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292
jan@ict.usc.edu

David R. Traum
USC Institute for Creative Technologies
13274 Fiji Way
Marina del Rey, CA 90292
traum@ict.usc.edu

Abstract

For embodied agents to engage in realistic multiparty conversation, they must stand in appropriate places with respect to other agents and the environment. When these factors change, for example when an agent joins a conversation, the agents must dynamically move to a new location and/or orientation to accommodate. This paper presents an algorithm for simulating the movement of agents based on observed human behavior using techniques developed for pedestrian movement in crowd simulations. We extend a previous group conversation simulation to include an agent motion algorithm. We examine several test cases and show how the simulation generates results that mirror real-life conversation settings.

1 Introduction

When we look at human conversation in a casual, open setting, such as a party or marketplace, one of the first things we notice is a tendency for people to cluster into sub-groups involved in different conversations. These groupings are not fixed, however, people will often join and leave groups and often move from one group to another. Groups themselves may fragment into subgroups, and smaller groups sometimes merge into one larger group. Participants in these groups adapt their positions and orientations to account for these circumstances, often without missing a beat or otherwise disrupting their conversations.

In order to create believable social environments for games or training simulations we need agents that can perform these same kinds of behaviors in a realistic way. There are a number of crowd simulations (Sung et al., 2004; Shao and Terzopoulos, 2005; Still, 2000; Helbing and Molnár, 1995), but most of these place an emphasis on large-scale movement of agents and do not model the low-level aspects of conversational interaction in a realistic way — movement of agents in multiparty conversation is more about positioning and repositioning on a local scale. There is also a large body of work on embodied conversational agents (Cassell et al., 2000), which attempt to model realistic conversational non-verbal behaviors. Most of this work focuses on aspects such as gaze, facial expressions, and hand and arm gestures, rather than positioning and orientation in a group. There is some important work on authored presentation agents and avatars for human participants which take account of position in the modelling (Vilhjalmsson and Cassell, 1998; Rehm et al., 2005), but none of this work presents fully explicit algorithms for controlling the positioning and movement behavior of autonomous agents in dynamic conversations.

In previous work, it has been shown that incorrect positioning of animated agents has a negative effect on the believability of dynamic group conversation (Jan and Traum, 2005). Research from anthropologists and social psychologists such as the classic work on proxemics by Hall (1968) and positioning by Kendon (1990) provide social reasons to explain how people position themselves in different situations. It is also important to know that people expect

Proceedings of the Workshop on Embodied Language Processing, pages 59–66,
Prague, Czech Republic, June 28, 2007.

similar behavior in virtual environments as in real life as shown by Bailenson et al. (2003). This gives us basic principles on which to base the simulation and provides some qualitative expectations, but is not suitable to directly convert into algorithms. The social force model (Helbing and Molnár, 1995) developed for crowd simulations gives a good framework for movement simulation. While the basic model shows how to handle pedestrian motion we apply the model to the problem of movement in conversation setting.

Our implementation of conversational movement and positioning is an extension of prior work in group conversation simulation using autonomous agents. Carletta and Padilha (2002) presented a simulation of the external view of a group conversation, in which the group members take turns speaking and listening to others. Previous work on turn-taking is used to form a probabilistic algorithm in which agents can perform basic behaviors such as speaking and listening, beginning, continuing or concluding a speaking turn, giving positive and negative feedback, head nods, gestures, posture shifts, and gaze. Behaviors are generated using a stochastic algorithm that compares randomly generated numbers against parameters that can take on values between 0 and 1.

This work was further extended by (Jan and Traum, 2005), who used new bodies in the Unreal Tournament game engine, and added support for dynamic creation of conversation groups. This simulation allowed dynamic creation, splitting, joining, entry and exit of sub-conversations. However, the characters were located in fixed positions. As indicated in their subject evaluations, this significantly decreased believability when conversation groups did not coincide with positioning of the agents. Adding support for movement of characters is a natural step to counter these less believable situations. We augment this work by adding a movement and positioning component that allows agents to monitor "forces" that make it more desirable to move to one place or another, iteratively select new destinations and move while remaining engaged in conversations.

The rest of the paper is organized as follows. Section 2 describes the main motivations that agents have for moving from their current position in conversation. Section 3 presents the social force model, which specifies a set of forces that pressure an agent to move in one direction or another, and a decision algorithm for deciding which forces to act on in different situations. Section 4 presents a series of test cases for the algorithm, demonstrating that the model behaves as desired for some benchmark problems in this space. We conclude in section 5 with a description of future work in this area.

2 Reasons for Movement

There are several reasons why someone engaged in conversation would want to shift position. Some of these include:

- one is listening to a speaker who is too far and or not loud enough to hear,
- there is too much noise from other nearby sound sources,
- the background noise is louder than the speaker,
- one is too close to others to feel comfortable,
- one has an occluded view or is occluding the view of others.

Any of these factors (or a combination of several) could motivate a participant to move to a more comfortable location. During the simulation the speakers can change, other noise sources can start and stop, and other agents can move around as well. These factors can cause a variety of motion throughout the course of interactions with others. In the rest of this section we describe these factors in more detail. In the next section we will develop a formal model of reactions to these factors.

The first reason we consider for repositioning of conversation participants is audibility of the speaker. The deciding factor can be either the absolute volume of the speaker, or the relative volume compared to other "noise". Noise here describes all audio input that is not speech by someone in the current conversation group. This includes the speech of agents engaged in other conversations as well as non-speech sounds. When we are comparing the loudness of different sources we take into account that intensity of the perceived signal decreases with the square of the distance and also that the loudness of several sources is additive.

Even when the speaker can be heard over a noise source, if outside disruptions are loud enough, the group might want to move to a more remote area where they can interact without interruptions. Each of the participants may decide to shift away from a noise source, even without an explicit group decision. Of course this may not always be possible if the area is very crowded.

Another reason for movement is proxemics. Hall (1968) writes that individuals generally divide their personal space into four distinct zones. The intimate zone is used for embracing or whispering, the personal zone is used for conversation among good friends, the social zone is used for conversation among acquaintances and the public zone for public speaking. The actual distances the zones span are different for each culture and its interpretation may vary based on an individual's personality. If the speaker is outside the participant's preferred zone, the participant will move toward the speaker. Similarly if someone invades the personal zone of a participant, the participant will move away.

The final reason for movement is specific to multiparty conversations. When there are several people in conversation they will tend to form a circular formation. This gives the sense of inclusion to participants and gives them a better view of one another (Kendon, 1990).

3 Social Force Model

We present our movement simulation in the context of a social force model. Similar to movement in crowds, the movement of people engaged in conversation is to a large extent reactionary. The reaction is usually automatic and determined by person's experience, rather than planned for. It is possible to assign a vectorial quantity for each person in conversation, that describes the desired movement direction. This quantity can be interpreted as a social force. This force represents the influence of the environment on the behavior of conversation participant. It is important to note however that this force does not directly cause the body to move, but rather provides a motivation to move. We illustrate these forces with figures such as Figure 1, where each circle

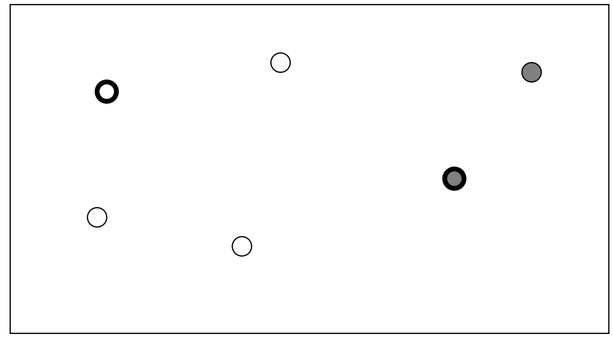

Figure 1: A sample group positioning. Each circle represents an agent. A thick border represents that the agent is talking, filled or empty shading indicates conversation group membership.

represents an agent, the different shadings represent members of different conversation groups, thicker circles represent speakers in that group, and arrows represent forces on an agent of interest.

We associate a force with each reason for movement:

$\vec{F}_{speaker}$: attractive force toward a speaker

$\vec{F}_{noise}$: repelling force from outside noise

$\vec{F}_{proximity}$: repelling force from agents that are too close

$\vec{F}_{circle}$: force toward circular formation of all conversation participants

$\vec{F}_{speaker}$ is a force that is activated when the speaker is too far from the listener. This can happen for one of two reasons. Either the speaker is not loud enough and the listener has to move closer in order to understand him, or he is outside the desired zone for communication. When the agent decides to join conversation this is the main influence that guides the agent to his conversation group as shown in Figure 2. $\vec{F}_{speaker}$ is computed according to the following equation, where $\vec{r}_{speaker}$ is location of the speaker, $\vec{r}$ is location of the agent and k is a scaling factor (we are currently using $k = 1$):

$$\vec{F}_{speaker} = k(\vec{r}_{speaker} - \vec{r})$$

$\vec{F}_{noise}$ is a sum of forces away from each source of noise. Each component force is directed away from

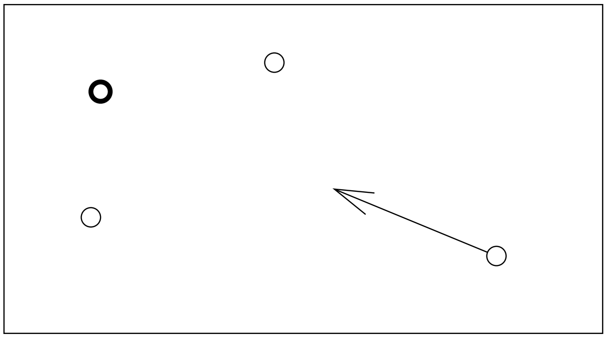

Figure 2: Attractive force toward speaker $\vec{F}_{speaker}$.

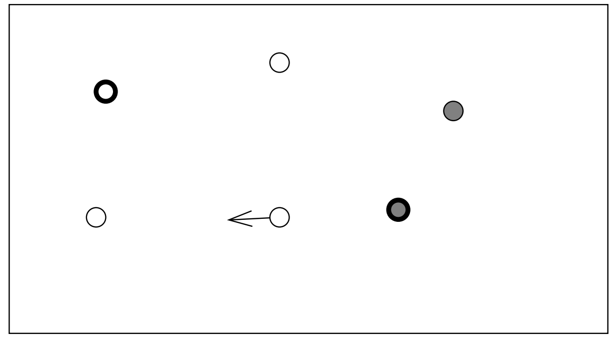

Figure 3: Repelling force away from other speakers $\vec{F}_{noise}$.

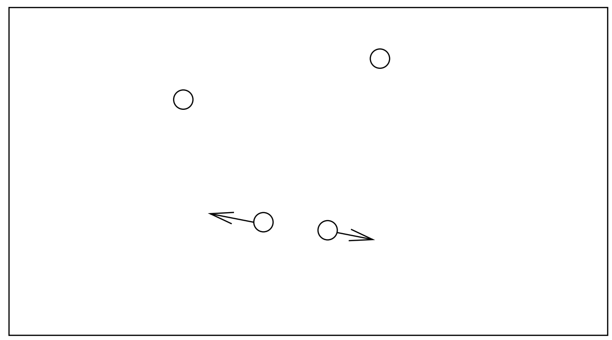

Figure 4: Repelling force away from agents that are too close $\vec{F}_{proximity}$.

that particular source and its size is inversely proportional to square of the distance. This means that only sources relatively close to the agent will have a significant influence. Not all noise is a large enough motivation for the agent to act upon. The force is only active when the noise level exceeds a threshold or when its relative value compared to speaker level in the group exceeds a threshold. Figure 3 shows an example of the latter. The following equation is used to compute $\vec{F}_{noise}$:

$$\vec{F}_{noise} = -\sum_i \frac{\vec{r}_i - \vec{r}}{\|\vec{r}_i - \vec{r}\|^3}$$

$\vec{F}_{proximity}$ is also a cumulative force. It is a sum of forces away from each agent that is too close. The force gets stronger the closer the invading agent is. This takes effect for both agents in the conversation group and other agents. This is the second force that is modeling proxemics. While $\vec{F}_{speaker}$ is activated when the agent is farther than the desired social zone, $\vec{F}_{proximity}$ is activated when the agent moves to a closer zone. Based on how well the agents know each other this can be either when the agent enters the intimate zone or the personal zone. Figure 4 shows an example when two agents get too close to each other. The following equation is used to compute values for $\vec{F}_{proximity}$:

$$\vec{F}_{proximity} = -\sum_{\|\vec{r}_i - \vec{r}\| < distance_{zone}} \frac{\vec{r}_i - \vec{r}}{\|\vec{r}_i - \vec{r}\|^2}$$

$\vec{F}_{circle}$ is responsible for forming the conversational group into a convex, roughly circular formation. Each agent has a belief about who is currently participating in the conversation. An agent will compute the center of mass of all these assumed participants and the average distance from the center. If an agent's position deviates too much from the average, the $\vec{F}_{circle}$ gets activated either toward or away from center of mass. Notice that $\vec{F}_{proximity}$ takes care of spreading out around the circle. The situation in Figure 5 is an example where an agent decides that he has to adapt his positioning. Notice that if this agent was not aware of the agent to his left, the force would not get triggered. This can be a cause for many interesting situations when agents have different beliefs about who is part of the conversation.

$$\vec{r}_m = \frac{1}{N} \sum_i \vec{r}_i$$

$$\vec{F}_{circle} = \lambda \left(\frac{1}{N} \sum_i \|\vec{r}_i - \vec{r}_m\| \frac{\vec{r} - \vec{r}_m}{\|\vec{r} - \vec{r}_m\|} - \vec{r} \right)$$

As described above, each force has some conditions that determine whether the force plays an ac-

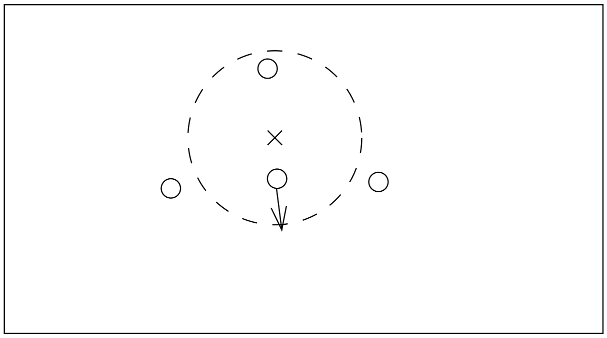

Figure 5: Agent's deviation from circular formation exceeds threshold and triggers force $\vec{F}_{circle}$.

tive role in motivating movement. Since the forces are not actually physically acting on agent's bodies, it is not unreasonable for agents to suppress a certain force. All the possible causes for movement are always present, but the agents selectively decide which ones they will act upon in a given situation. This is unlike a kinematics calculation with physical forces where all forces are always active. Combining all the conditions we can define which forces are active according to a simple decision procedure. We can view this as priorities the agent has that decide which conditions are more important to react to.

In our implementation we use the following priorities:

if speaker is too low $\vec{F} = \vec{F}_{speaker} + \vec{F}_{proximity}$

else if noise is louder than speaker $\vec{F} = \vec{F}_{speaker} + \vec{F}_{noise} + \vec{F}_{proximity}$

else if noise is too loud $\vec{F} = \vec{F}_{noise} + \vec{F}_{proximity}$

else if too close to someone $\vec{F} = \vec{F}_{proximity}$

otherwise $\vec{F} = \vec{F}_{circle}$

Using the above priorities we have a force defined at each point in space where an agent could be located. We do not use this for the continuous computation of movement, but rather use it to compute destination points. In each planning cycle the agents will consider whether they should move. To do this an agent considers his position in the force field and computes a destination in the direction of the force field. This process is performed iteratively a constant *bound* times (unless there is no movement in an earlier iteration). This is described in the following equations, where $\vec{r}$ is the initial position, α is a scaling factor, and $\vec{P}_{bound}$ is the destination for the movement of this planning cycle:

$$\vec{P}_0 = \vec{r}$$

$$\vec{P}_{i+1} = \vec{P}_i + \alpha \vec{F}(\vec{P}_i)$$

$$Desti\vec{n}ation = \vec{P}_{bound}$$

Once we have computed the destination, we use it as a destination point for the character movement algorithms in the Unreal Tournament game engine. These will manage character animation and collision avoidance.

Figure 6 shows an example with two separate conversation groups, where one agent decides to leave the shaded group and join the unshaded conversation. The figure shows the iterations he is performing in his planning cycle and the resulting final destination.

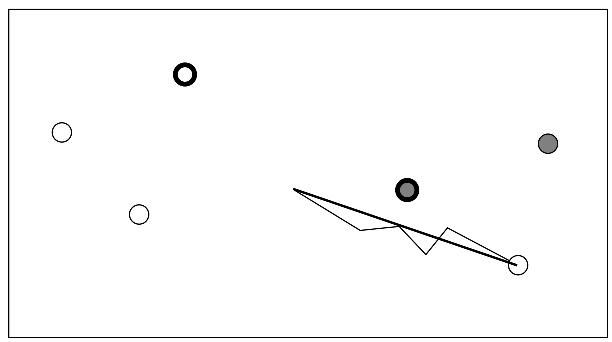

Figure 6: Example of motion computation: The lower right agent decided to join the unshaded conversation. He iteratively applies movement in the direction of local forces. In each iteration the effects of different component forces may take effect. The thick line indicates the final destination and path the agent chooses for this planning cycle.

4 Test Case Analysis

A full evaluation of the social-force based positioning algorithm presented in the previous section would involve analysis of simulations to see if they improve believability over static simulations such as simulation of Jan and Traum (2005), or other algorithms. While this remains future work for the moment, we did evaluate the algorithms against a series of test cases where we know what behavior to expect from known forces. In this section we present three such cases, showing that the algorithm does have the power to represent several aspects of conversational positioning.

In the simulations we describe here we did not change the conversational attributes of agents, but we did constrain the grouping dynamics. In a normal situation the agents would randomly form conversation groups, based on their stochastic decisions. Here we wanted to examine particular scenarios and how the movement algorithm would react to specific changes in conversation group structure. For this reason we disabled conversational grouping decisions in the algorithm and triggered the group structure changes manually from the user interface.

The only variable input to the movement algorithms for different agents is the preferences for proxemics. Each agent has defined values for all zones, but we set all agents to use social zone for communicating. The other parameters such as thresholds for hearing a speaker and noise and circular formations were fixed for these experiments.

4.1 Joining conversation

In this test case we have 4 agents. In the initial condition three agents are engaged in conversation while the fourth one is away from the scene. We let the simulation run and at some point we give a command to the fourth agent to join the group of three. At first the agent will move toward the group until he is in a comfortable range as shown in Figure 7.

At the point in which the fourth agent decides to join the other three, he is the only one who knows he wants to join the conversation. The other agents know of the presence of the fourth agent, but they have no idea that he would like to join them. The fourth agent is listening for a while and when he gives a feedback signal the other agents interpret that as a signal that he wants to join the conversation. As a result the agents reevaluate their positioning and one agent decides it would be appropriate to move a step back to give more space to the new agent. Given more space the new agent is able to move in circular formation with the rest of the group without intruding on the personal zones of other agents. The stable point of simulation is shown in Figure 8.

Figure 7: The agent on the left is approaching a conversation. Arrows indicate where the agents will move from now until the simulation stabilizes.

Figure 8: Stable point after the fourth agent joins the conversation.

4.2 Conversation splitting into two separate conversations

In this test case, we have 6 agents. After initial placement of the agents we issue a command for all the agents to form one conversation group. As a result they form a circular formation as can be seen in Figure 9.

We let the agents talk for a while and then give a command to the two agents on the right side of the group to start a side conversation. After this a complex sequence of events takes place. Initially the remaining agents still think that those two agents are part of their conversation group. They have to disambiguate the speech of those two agents and decide whether this is just an interruption or a split in the

Figure 9: Agents form in a circle to engage in a single conversation.

conversation. After a while they realize that those agents are having a separate conversation.

Deciding that the agents on the right have left the conversation leads to a change in the force field. The agents that were closest to the split are bothered by the noise and start adjusting by moving away. By doing this they change the shape of formation which causes the farther agents to also adapt back into circular formation. At the same time the agents who split also move away from the others until they get to a point where all are satisfied. The point where the simulation stabilized is shown in Figure 10.

Figure 10: After two agents leave the conversation the agents adapt to it by repositioning.

4.3 Effect of proxemics

In this test case, we examine the effects when the social zones of the agents are not compatible. This frequently happens when we have people from different cultures with a large difference in distances for social zones. An example would be North Americans compared to Arabs. Americans prefer a much greater inter-personal distance than Arabs. Empirical data shows that in many such situations there is a sort of dance with one agent moving in while another moves away (Scheflen, 1975).

Figure 11: Incompatible social zones.

Figure 11 shows an example of agents with incompatible social zones. The markings on the ground indicate the minimum and maximum acceptable distance for social zone for each agent. We can see that the agent on the left has a much smaller comfortable distance than the one on the right. In the current position the left agent feels that the other one is too far, while the right agent thinks everything is fine. This causes the left agent to make a step forward. Consequently by doing so he steps into personal zone of the right agent. Now the left agent is satisfied with the situation but the right agent feels uncomfortable and decides to take a step back to keep the other agent out of his personal zone. If nothing else intervenes, this process can continue, as the agent on the left "chases" the one on the right out of the marketplace.

5 Conclusions

In the previous section, we have shown examples of how the movement algorithm can mirror many effects we see in real conversations. The examples however were very constrained and could not show all the possible combinations that could result from random choices the agents can make. Given the fact that each agent maintains his own belief about who is currently in their conversation we can see many interesting effects when those beliefs become unsynchronized.

As seen in the third test case, we can get some very interesting results when we simulate agents of different cultures. We think that this simulation approach can be fruitful for modeling cultural differences in conversational behavior, and could be used for inter-cultural and cross-cultural awareness and training. We are currently exploring whether we can model different cultural norms for conversational behaviors in ways such that the resulting agent interaction can be recognized as appropriate to one culture or another.

There are still several improvements possible for the conversation simulation. On the presentation side we are planning to make some improvements to the bodies and number and types of conversational gestures they can display. We also plan to improve the algorithm so that it will be able to generate different conversation styles. Currently all conversations take the same form where all the agents have the same goals, their only goal is to engage in conversation with other agents. We plan to introduce the notion of tasks so that we can better simulate different kinds of activities such as asking for directions, a political debate, or casual conversation.

Acknowledgments

The project described here has been sponsored by the U.S. Army Research, Development, and Engineering Command (RDECOM). Statements and opinions expressed do not necessarily reflect the position or the policy of the United States Government, and no official endorsement should be inferred.

References

Jeremy N. Bailenson, Jim Blascovich, Andrew C. Beall, and Jack M. Loomis. 2003. Interpersonal distance in immersive virtual environments. *Personality and Social Psychology Bulletin*, 29:819–833.

Justine Cassell, Joseph Sullivan, Scott Prevost, and Elizabeth Churchill, editors. 2000. *Embodied Conversational Agents*. MIT Press, Cambridge, MA.

Edward T. Hall. 1968. Proxemics. *Current Anthropology*, 9(2/3):83–108, apr.

Dirk Helbing and Péter Molnár. 1995. Social force model for pedestrian dynamics. *Phys. Rev. E*, 51(5):4282–4286, May.

Dusan Jan and David R. Traum. 2005. Dialog simulation for background characters. *Lecture Notes in Computer Science*, pages 65–74.

Adam Kendon, 1990. *Spatial Organization in Social Encounters: the F-formation System*, pages 209–237. Cambridge University Press.

E. Padilha and J. Carletta. 2002. A simulation of small group discussion. *Proceedings of EDILOG 2002: Sixth Workshop on the Semantics and Pragmatics of Dialogue*, pages 117–124.

Matthias Rehm, Elisabeth Andre, and Michael Nischt. 2005. Let's come together - social navigation behaviors of virtual and real humans. In Mark Maybury et al., editor, *INTETAIN 2005*, LNAI, pages 122–131. Springer.

Albert E. Scheflen. 1975. Micro-territories in human interaction. In Adam Kendon, Richard M. Harris, and Mary Ritchie Key, editors, *World Anthropology: Organization of Behavior in Face-to-Face Interaction*, pages 159–173. Mouton, Paris.

Wei Shao and Demetri Terzopoulos. 2005. Autonomous pedestrians. In *SCA '05: Proceedings of the 2005 ACM SIGGRAPH/Eurographics symposium on Computer animation*, pages 19–28, New York, NY, USA. ACM Press.

G. Keith Still. 2000. *Crowd Dynamics*. Ph.D. thesis, Warwick University.

Mankyu Sung, Michael Gleicher, and Stephen Chenney. 2004. Scalable behaviors for crowd simulation. *Computer Graphics Forum*, 23(3):519–528.

Hannes Hogni Vilhjalmsson and Justine Cassell. 1998. Bodychat: autonomous communicative behaviors in avatars. In *AGENTS '98: Proceedings of the second international conference on Autonomous agents*, pages 269–276, New York, NY, USA. ACM Press.