Animating Idle Gaze in Public Places

Angelo Cafaro[1], Raffaele Gaito[1] and Hannes Högni Vilhjálmsson[2],

[1] Università degli Studi di Salerno, Italy
angelcaf@inwind.it, raffaele@cosmopoli.it
[2] Center for Analisys and Design of Intelligent Agents, School of Computer Science, Reykjavik University, Iceland
hannes@ru.is

Abstract. In realistic looking game environments it is important that virtual characters behave naturally. Our goal is to produce naturally looking gaze behavior for animated agents and avatars that are simply idling. We studied people standing and waiting, as well as people walking down a shopping street. From our observations we built a demo with the CADIA Populus multi-agent social simulation platform.

Key words: Social animation, gaze, idling, avatars, game environments

1 Introduction

It is well established that gaze is an important aspect of human behavior that exhibits regular patterns during social activity [1–3]. Different social situations call for different gaze patterns, and within the same situation, different personal, inter relational and environmental factors also play a role [4]. A pervasive personal factor is cognitive activity [1], which at least will always produce a base-line gaze behavior in the background. In this paper we explore and simulate this base-line behavior in two different social situations: (1) Waiting for a Bus; and (2) Walking down a shopping street. In both situations our subjects are "idling alone" to minimize external factors, but they are still part of the dynamic social environment of a public place.

2 Related Work

Animated characters in games should foster a sense of co-presence: the sense of actually being with another person rather than simply being a graphical object on a computer screen [5]. Static screen shots may look convincing, but during actual game play the behavior has to keep up with the visual realism. The complete lack of gaze behavior in online avatars was addressed by [6] in BodyChat by fully automating it based on existing theory like [2] and [3]. BodyChat focused on social interaction, but did not model idling behavior. We build on the same theory, but add our own empirical results to fill in the idling gaps. Similar to BodyChat, [5] highlighted the importance of secondary behavior, generated autonomously for an avatar. They show a technique to add automated gaze to user-defined task execution, but base the generation on informal observations, while we rely on detailed video analysis. [7] focuses on generating gaze during the execution of certain tasks. While our work relates to their "Spontaneous Looking" situation, we extend it further with new well defined social situations and new observations. [8] presented a statistical eye movement model, which is based on both empirical studies of saccades and acquired eye movement data. Unlike our work, they analyzed people having face-to-face conversations. In recent work [1] remarked that gaze plays a large number of cognitive, communicative and affective roles in face-to-face human interaction, which is led us to think more about the cognitive base-line before engaging in interaction.

3 Video Studies

We performed two studies that we will refer to as study 1 and study 2. In the following descriptions we will refer to the study subjects as Ss.

Study 1 We picked Hlemmur, the main bus terminal in Reykjavik, as our setting for waiting behavior. Ss were filmed for a duration between 1 and 2 minutes with the camera placed 30 m away from the bus station as depicted in figure 1 (left).

Study 2 We picked Laugavegur, the main shopping street downtown Reykjavik, as our setting for walking behavior. Ss were filmed along a 30 m walking-path with the camera placed 30 m away from the end of the walking-path as depicted in figure 1 (right).

In both studies we first chose a S and, then, we recorded him/her for the whole study duration with two cameras, one focused on the S's eyes and another focused on the S's surroundings using a larger field of view. We analyzed all collected

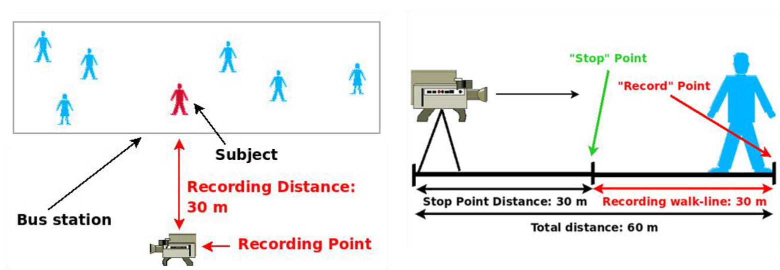

Fig. 1. The setup of the two studies. In study 1 (left), the camera recorded subjects outside a bus terminal on the other side of the street. In study 2 (right), the camera recorded a subject walking towards the camera on the sidewalk.

video frame-by-frame, annotating the following for each subject: Eye direction (combinations of up/center/down and left/center/right relative to head), head orientation (combinations of up/center/down and left/center/right relative to torso), torso facing (left, forward or right relative to line to camera), eyelids state (open or closed)[3] and target of attention. For these we also annotated duration. Moreover, for the eyes, head and torso we annotated the time interval in which movement happens and the speed of that movement (slow, neutral or fast). For study 1 we also annotated the following: Potential gaze targets for the Ss (categorized as objects or persons), proxemics area in witch the potential target appears [9] and instances when a potential target in a given area produced actual gaze movement. In study 1 we annotated 9 subjects for a total duration of 14'30", during which 142 gaze shifts were observed. In study 2 we annotated 16 subjects for a total duration of 5'40", during which 216 gaze shifts were observed (excluding a few at the camera).

The results include both general observations and empirical data extracted from the annotations. The observations for Study 1, include:

1. Ss that produce short glances were observed to keep their glances short throughout the study;
2. Conversely, Ss producing longer glances or gazes were observed to keep them long throughout the study;
3. The shorter glance Ss were observed to pick many different targets around them;
4. The longer glance Ss were picking targets from a narrower set;

The empirical data describes: The gaze attraction of potential targets that were either objects or persons, in relation to the proximity of the target. This is based on how much of the total time an object or a person spent at a given proximity was spent being looked at by the subject. In addition the duration of each gaze was recorded. This is shown in Table 1.

For Study 2, the main discovered patterns are:

1. Ss frequently look to the ground while walking down the street;
2. Related to that, Ss first choice is to look down to the ground when exercising gaze-aversion;
3. Ss usually close their eyelids just before moving their head (or changing gaze direction);
4. Ss almost never look up, up-left or up-right.

The empirical data describes: How often a target of a given type gets looked at compared to other target types. Information about timing including the average, minimum and maximum gaze length. This is shown in Table 2.

We are at an early stage of a more in-depth statistical analysis of the massive amount of data we have gathered, but the patterns we have already discovered along with the first set of numerical data has been the basis for new autonomous idle gaze behavior generation algorithms for the two social situations we studied.

[3] We annotated this only for study 2 due to the high frequency of gaze movement.

	Objects			Persons		
Proxemics Area	Time on Target	%	Avg. Dur.	Time on Target	%	Avg. Dur.
Intimate	84.11 out of 84.11	100%	6	0 out of 6.43	0%	-
Personal	10.26 out of 10.26	100%	5.13	7.3 out of 61.16	12%	2.43
Social	14.36 out of 22.56	64%	4.80	45.17 out of 96.68	47%	22.60
Public	1.33 out of 1.33	100%	1.33	6.90 out of 29.46	23%	2.30
Extra	5.90 out of 8.66	68%	5.90	9.00 out of 35.56	25%	3.00

Table 1. Observations about the relationship between gaze targets and proxemics from Study 1 where the subject is waiting alone for a bus. It is noticeable that objects at a close range receive complete attention while persons at a similarly close range are avoided. All durations are in seconds.

Targets			Durations		
Target Type	%	Time on Target	Avg.	Min.	Max.
Same Gender	3%	7.75	0.64	0.24	1.24
Opposite Gender	5%	12.92	0.72	0.29	1.15
Shops	13%	33.58	0.74	0.14	1.34
Cars	9%	23.25	0.93	0.1	1.8
Ground	25%	64.58	1.5	0.5	2.5
Camera	7%	18.08	-	-	-
Other Side	16%	41.33	-	-	-
Unknown	22%	56.83	-	-	-
Total	100%	258.32			

Table 2. Observations about where people look when walking down a shopping street from Study 2. The first two target types refer to other people on the sidewalk, of which there were equal number of males and females. Cars refer to those coming up the street and passing the subjects. All durations are in seconds.

4 Autonomous Generation of Idle Gaze Behavior

We implemented the idle gaze generation with a new tool for constructing social behaviors for avatars and agents in game environments: CADIA Populus [10]. The process was simply a matter of plugging in new steering behaviors along with conditions to activate them during the particular situations that we were modeling. The new behaviors fit nicely into the steering behavior framework of CADIA Populus, adding a new set of motivations for turning the head and eyes when appropriate.

Both of our steering behaviors follow the same working principle, they differ only in their update frequency[4]. The *idle gaze waiting* required a lower update frequency than the *idle gaze walking*. This makes intuitive sense. The goal of the entire autonomous gaze generation process is to continuously obtain a target and a gaze duration for that target. The target can be an entity of the environment (other avatars or objects), a point in space or a direction relative to the avatar itself. In order to achieve this goal the process starts analyzing the entities around the avatar, using the perception system provided by CADIA Populus, and creates a list of potential targets for the decision step. At this stage several factors are combined to obtain the final target and its duration: First the social situation dependent steering behavior, that incorporates the observation data, is

[4] CADIA Populus updates each steering behavior with a given frequency.

applied and then additional features from the avatar's profile and preferences are incorporated. We limited the profile to a simple personality variable (extrovert or introvert) and gender, while a preferences held a possible bias towards any kind of entity. The underlying perception system already provided in CADIA Populus allows access to two kinds of information: the social high-level environment and the low-level avatar visual perception. Information related to the social environment is retrieved through a social perception interface; examples of social perception include how many individuals are within a certain proximity range (using the proxemics notation intimate, personal, social and public zones). Notice that the perception of social and public space has a blind cone behind the avatar, of respectively 90 and 150 degrees [10]. The avatar visual perception is divided into central view and peripheral view. Our basic attention model is real-

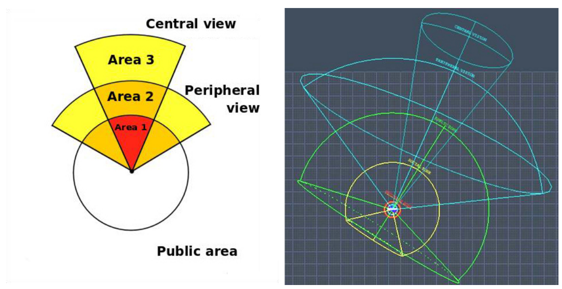

Fig. 2. The area subdivision in our basic attention model for the potential target selection (left). The social and visual perception systems in CADIA Populus (right).

ized using the intersection between the social and visual perception information, obtaining the three areas as we can see in Fig. 2. In order to produce a list of potential targets the system collects them from these areas by priority, starting from Area 1 and ending when the mentioned targets are found. Finally, only entities in remaining areas that are moving toward the avatar are added, as they represent high potential for upcoming interaction. A detailed attention model implementation is beyond the purpose of this paper, in fact ours is primarily a way to obtain a reasonable list of targets to work with.

The decision phase is a core process and is based on four values from our data: *Choice Probability*, *Look Probability*, *Minimum* and *Maximum Duration*. To use these differently for each situation. For study 1, in each proxemics area we identify two general target types: Objects or Persons. Then, fixing a target type in one specific area we assigned the four above-mentioned values. For study 2, since the setting was the same for all recordings, all the potential targets were classified into five main categories. Three of them including targets able to move (Cars, Same Gender People or Opposite Gender People), the remaining two included fixed targets (Shops and Other[5]). Finally, for each target category we

[5] We grouped in this category all the other objects present on the scene, i.e. pickets, lamps, etc. . .

assigned the four above-mentioned values. Next we pick *a single* target of possible interest by using the *Choice Probability* to select among potential targets. An example of this probability is given in Table 2. Then, for this single target, we may or may not decide to actually look at it. This decision is left to what we call a *Look Probability* for a particular type of target. An example of a probability of this sort is given in Table 1. Finally, if a gaze is produced, the duration of the gaze for that target is a random value between *Minimum Duration* and *Maximum Duration* as seen in the data.

In some cases the decision process could not select a target for two possible reasons: there aren't potential targets surrounding the avatar or the decision process chose not to produce gaze towards the most likely target (due to *Look Probability*). In these cases we generated a gaze behavior with a relative direction based on the discovered common patterns mentioned earlier. To make the gaze behavior even more dynamic, the avatars are always attentive to changes in surrounding potential targets. This means that even during a gaze shift to a chosen target. These changes are being monitored in the background. Another important detail is that repeated targets are avoided using a simple memory structure.

5 Conclusions and Future Work

This paper has introduced one approach to design and animate avatar gaze behavior for game environments based both on existing literature and on data gathered through carefully planned observation. As we can see in Fig. 3, we have constructed a virtual model of the places where we gathered our data in order to test our steering behaviors. We have not performed user studies yet to

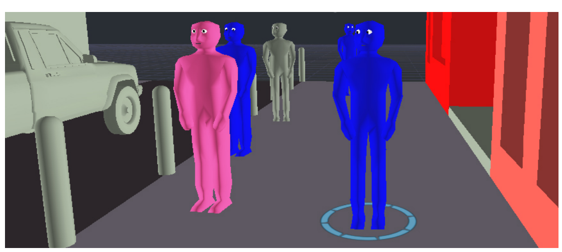

Fig. 3. Screenshot of our gaze behaviour inside CADIA Populus.

independently evaluate the believability of the resulting gaze behavior, but it is clear from the accompanying video[6] that compared to no gaze or purely random gaze, our results look promising. Some limitations should be considered. First of all, the data was gathered in a single location, which may or may not generalize to other locations or cultures. Secondly, the videos were gathered in a natural

[6] Available at http://cadia.ru.is/projects/cadiapopulus/cafaro2009.avi

setting so it was impossible to obtain detailed data on many of the personal state factors that we know can influence gaze, for example how the subjects were feeling. Thirdly, only certain types of potential targets were present so even if we tried to keep target types general some new scenes may require more data. Finally, CADIA Populus currently lacks vertical head movement, which constrains looking down to eye movement only.

This is work in progress, so future work is extensive. For example we would like to incorporate the speed of head and torso movement. Including the closing of the eyes could improve believability. We have already analyzed the data for that and just have to implement it. Regarding the video studies, the next steps include a more thorough statistical data analysis and hypothesis testing for contributing to the body of theoretical work on gaze behavior. Finally, we would like to construct new studies with different configurations to cover more factors and social situations.

6 Acknowledgements

Big thanks to the CADIA team and Claudio Pedica in particular. This work was made possible with the support of the Humanoid Agents in Social Game Environments Grant of Excellence from The Icelandic Research Fund and the ERASMUS European student exchange programme.

References

1. Lee, J., Marsella, S., Traum, D., Gratch, J., Lance, B.: The rickel gaze model: A window on the mind of a virtual human. In: Proceedings of the 7th international conference on Intelligent Virtual Agents (IVA '07), Berlin, Heidelberg, Springer-Verlag (2007) 296–303
2. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters (Studies in Interactional Sociolinguistics). Cambridge University Press (November 1990)
3. Argyle, M., Cook, M.: 5. In: Gaze & Mutual Gaze. Cambridge University Press (January 1976) 98–124
4. Argyle, M., Furnham, A., Graham, J.A.: Social Situations. Cambridge UniversityPress (1981)
5. Gillies, M.F.P., Dodgson, N.A.: Behaviourally rich actions for user controlled characters. Journal of Computers and Graphics **28(6)** (2004) 945–954
6. Vilhjlmsson, H.H., Cassell, J.: Bodychat: Autonomous communicative behaviors in avatars. In: Proceedings of the 2nd International Conference on Autonomous Agents (Agents '98), New York, ACM Press (1998) 269–276
7. Chopra-Khullar, S., Badler, N.I.: Where to look? automating attending behaviors of virtual human characters. In: Proceedings of the third annual conference on Autonomous Agents (Agents '99), New York, NY, USA, ACM (1999) 16–23
8. Lee, S.P., Badler, J.B., Badler, N.I.: Eyes alive. In: Proceedings of of ACM SIGGRAPH 2002. Volume 21., New York, NY, USA, ACM (2002) 637–644
9. Hall, E.T.: The Hidden Dimension. 1st edn. Doubleday, Garden City (1966)
10. Pedica, C., Vilhjlmsson, H.H.: Social perception and steering for online avatars. In: Proceedings of Intelligent Virtual Agents (IVA '08). (2008) 104–116