Non-verbal invitation by an embodied agent

P.M. de Koning
p.m.dekoning@student.utwente.nl

ABSTRACT
The ability of an embodied conversational agent to start the initial contact with a human with non-verbal behavior is one step closer to a more natural human agent conversation. Four different agent implementations are presented with the goal to get the attention and to avoid the attention of people by using the gaze and the smile. The associated results that are found while these agents were placed in an environment are shown. One of the agents was capable of keeping people interested longer. Another result was the surprising amount of provoking and aggressive behavior people showed towards the agents. More research is needed however to copy a believable human non-verbal invitation.

Keywords
Embodied Agent, Gaze, Close Salutation, Invitation, Eye and Head Movements, Movement Tracking, Non-verbal Behavior.

1. INTRODUCTION
People are used to communicate and interact with other people. This is why embodied agents provide a natural and easy way to let humans interact with machines. For this they have to follow the human to human conversation characteristics as close as possible [CASS00]. Embodied agents are intelligent agents that interact with the environment through a body within that environment. The embodied agent used in this paper for example consists of a camera to see and a human head on a computer screen to give feedback. The human to human conversation is a complex series of actions from both sides. And a lot of research still needs to be done to create an agent which is capable of this kind of interaction with a human. Human communication can be divided into two parts. There are verbal ways to reach another person, like choice of words and intonation. But there are also non-verbal channels for example: hand gestures, gaze, facial expressions, body posture and body orientation. In both areas a lot of research has been done already.

The invitation to start a conversation or interaction is one part of communication which has seen little research. In my research group a Virtual Dancer is being created, this agent needs a way to invite people to join the dance [REID06]. But this invitation might also be useful for advertisements (e.g. to attract people and look at a product without being annoying) or even computer games (e.g. to encourage a player to talk to an agent in a natural way) [PET6B]. The easiest way to do this is by using verbal ways. A shouting agent is bound to attract attention from people around it. Another less intrusive and more subtle way would be to use non-verbal signals. It will be interesting to see if an embodied agent is capable of inviting humans to do anything at all using non-verbal signals. According to [PETE05] a number of things can be read from the non-verbal channels in conversations: attention, perception, comprehension, internal reactions, believability, interest and agreement. For this research the attention and interest are of importance. The level of attention can be observed by looking at the gaze, the gestures and the locomotion.

Gaze or ‘direction of attention’ is a way to shift attention by reorienting the eyes, the head and the body [PETE05]. Is it possible to use gaze as a non-verbal signal to invite people to do something? According to [KEND67] gaze serves four functions in human to human conversations: it gives visual feedback, it regulates the flow of the conversation, it communicates emotions and relationships and it improves concentration by restricting the visual input. The eyes in particular are capable of giving signals and receiving signals during conversations [LEE02]: they regulate part of the conversation, they signal the search for feedback, they look for information, they show emotions and they influence the behavior of others. There is the answer: gaze and especially the eyes can influence the behavior of others and at the same time it also gives feedback about the attention of others. How this works exactly is still unknown.

Before starting an interaction the gaze on both sides makes it possible to see when another person is interested. This check is fast and subtle and reduces the social risk of engaging in a conversation with an unwilling participant [PETE05]. Without this eye-contact people need different channels to see if the other is interested, for example in [MUKA05] people are waving their hands or using words to ‘set up’ the start of a conversation. An agent using human-like gaze might be able to get the attention of others non-verbally. And this would be the first step to invite someone to do something. Peters in [PET6A] showed that embodied agents are capable of giving humans the awareness of being watched. This has been done by changing the gaze, body orientation, locomotion and gestures of the agent. Again the eyes seem to be very important, because foremost the eyes orientation and after that the head orientation and thereafter the body orientation have the biggest influence on this human awareness of being watched.

An embodied agent could, like humans, also take part in subtle process to see the other’s willingness to interact. This short negotiation is mostly a subconscious process in humans. To let an agent participate in this the ingredients of the negotiation need to be known. [KEND90] provides some information of a similar process, by describing the greeting. The greeting is also an interaction that starts in a subtle way. Every type of greeting consists of three parts. First the ‘distance salutation’ followed by the ‘approach’ and ended by the ‘close salutation’. During the ‘distance salutation’ people recognize each other explicitly and by doing this they tell each other they are ready for the greeting relation. In the ‘approach’ phase the number of withdrawals (for example looking away for a second) increases while the distance gets shorter. This might have to do with civil inattention discussed in [CARY78]. At the end of the greeting ceremony during the ‘close salutation’ they again tell each other they are ready for the greeting relation, but also the type of relation is established/confirmed.

In [KUNO04] the researchers let their agent change his facial expression to indicate he knows somebody is watching him. They made him smile, according to [MART82] this is a fitting behavior: the research subjects reported more eye gaze when there was a smile involved. One of the problems Kuno and his team encountered though was the Mona Lisa Effect: people tend to believe a two dimensional face is looking at them while

this is not the case. The solution they use for this problem is to let the head move in random directions until it is watching somebody. In the follow-up research [KUNO05] their agent tries to attract the attention of a person. Now not only the head and eye orientations are changing but also the body orientation. Test persons positioned around the agent were asked if they felt the agent was watching them on a scale from one to five. The problem they encountered this time was Mental Inertia: the neighbor of the selected person also thought the agent was watching him, because the body of the agent also oriented towards his direction. They were not able to find enough evidence that one of their solutions worked.

2. NON-VERBAL INVITATION

As indicated in the last chapter the first step to invite people is to get their attention. The main research question is thus:

- *How can an embodied agent attract attention by using non-verbal signals?*

This has been done before with non-verbal signals in an artificial environment [KUNO05]. According to the literature discussed above the smile and the gaze are fitting non-verbal signals. The researchers in [KUNO05] also used these signals and asked a group of test persons to fill out a questionnaire. Questionnaires are however not always reliable. It would be interesting to place a comparable agent in a non-laboratory setting. And to observe how many people choose to react and to observe what kind reactions show up. To measure the interest people show in the agent the number of gazes towards the agent, the gestures and the time of a person in front of the camera can be examined. When an agent has been created with these non-verbal signals, the following questions need to be answered:

- *How many responses does an embodied agent which can attract attention by using these non-verbal signals get?*
- *What kind of reactions does an embodied agent which can attract attention by using these non-verbal signals get?*
- *How long will people be interested by this agent before walking away?*

A lot of the passing people will probably glance at the agent because it's something new and has not been seen in that environment. That's why at least two types of non-verbal behavior needs to be tested to see what the difference in responses is. One engaging behavior is created to get people's attention. And another one, called the non-engaging version, is missing a certain part of that behavior. As is observed in [KEND90] between complete strangers, after mutual gaze people will most likely approach the agent. Again as stated in the last chapter gaze might be the key to getting attention. Just as in [KUNO05] one specific person at a time needs to be signaled. This way the agent is be able to pick one specific person out of a group to start an interaction with. So to be able to answer the above research questions the following sub question needs to be answered first.

- *How can Mental Inertia be avoided in a dynamic environment?*

Without doubt the agent needs to let his non-verbal signals react to the behavior of this specific person. This must be done obvious enough to let the other people know the agent is not responding to them and to reduce the Mona Lisa Effect. In the following chapter the solution to this problem and the used methods are discussed.

3. METHOD

In this chapter the different aspects of the setup are discussed. First the way the observation is done, then an introduction to the embodied agent used, followed by explanation of how the agent gets signals from the environment and how he sends signals back to this environment.

3.1 Observation

The agent is placed on one side of a corridor. This way a passerby has the choice to look or to stand still at the location of the agent or to move on. To observe the behavior of the people passing by, a camera is also placed at the agent's position. Because it is unknown which behavior to expect and it is out of the scope of this research to recognize it electronically, the captured data is being annotated and analyzed afterwards. The camera only needs to capture this footage when people are in front of the agent. So a trigger has been implemented to start and stop recording when people are in front of the agent. The time of people entering and leaving the field of view is been logged by the system and this is used while annotating the camera data. No difference is made between people moving from left to right or moving from right to left.

3.2 Embodied Agent

The embodied agent needs to have a gaze similar to humans in this context. The Virtual Dancer [REID06] is pretty complex so the embodied agent Obie [DUYB04] also internally created in the HMI Research Group (by The Duy Bui and Ahn Duc) is chosen to do the job. He is able to show and change emotions, eye and head orientations in real-time. In figure 1 the original Obie can be seen. The fact that he is a male and the other characteristics of his appearance could very well influence the behavior of passing people. But this is not part of the scope of this research. An upper body is missing, so this was created including the ability to control this body independently of the head. The animations he is capable of are: blinking with his eyes and showing small head and eye movements to simulate a real 'breathing' person. Animations for looking around, looking away and looking up when somebody comes into view had to be created.

Figure 1 - Obie showing a smile

3.3 System Input

The agent needs to react to the environment and especially to the presence and behavior of people. The camera, which is used to observe people, can also be used to track people in the field of view of the agent. This required a special wide angle camera lens, especially in corridors the agent needs to see persons coming from the 'angle of his eyes'. And the most realistic agent has a vision with the same angle as the human vision. People probably don't expect a human looking agent to respond only when they are right in front of his nose.

3.4 Mental Inertia

One problem discussed in the last chapter that needs to be solved before the experiment can be conducted is Mental Inertia [KUNO05]. Mental Inertia arises when multiple people are in front of the agent. The agent turns his gaze towards one of them and the people in the same direction think he is also looking towards them. The Mona Lisa Effect also plays a part in this, because a two dimensional face that is looking straight ahead seems to be looking at anyone who stands in front of it. The solution used in this experiment was not to get rid of the initial Mental Inertia, but to make the agent respond to the behavior of the selected person. The agent orients his upper body and to a lesser degree his head and to an even lesser degree his eyes towards the position the selected persons stands in relation to the position straight for the agent. Figure 2 depicts the two extremes, the left picture for people in the right part of the agent's view as seen from the agent's perspective and the right picture for people in the left part of the view. Notice the difference in degrees for the different parts, when the eyes are oriented further from the 'straight forward look' the agent seems to be looking at a position next to you.

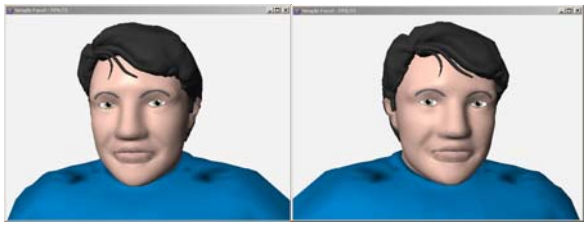


Figure 1 - The mental inertia solution

This way also the attention problem [PET6A] encountered is overcome. Peters discovered that an agent looking straight ahead with his body, head and eyes gets a surprisingly low attention score. Probably because this gaze was seen as stare behavior. A small angle between the head and eyes heightened the attention score, probably because of the bigger contrast between the body parts. Because almost nobody in this experiment stands exactly in the middle of the agent's viewport there was most of the time contrast between the eyes, the head and the upper body.

3.5 Agent Behavior

A lot of factors can influence the experiment, like the appearance of the agent, the time of the day, the location. The time of the day determines for example how many people will walk by. That's why the agent of this research switches from behavior one to behavior two after a certain time and then to behavior one again, etc. This way the differences in responses between the two behaviors can be studied while reducing different circumstances.

Because this research is about the ability of an embodied agent to get attention, behavior number one is trying to get attention (from now on called the engaging behavior) and the second behavior is the opposite (from now on the non-engaging behavior).

To overcome the Mona Lisa Effect, the agent shows idle behavior when nobody is around. The idle behavior consists of looking around in random directions with a very small chance to look straight out of the screen.

Two experiments A and B with different behaviors have been conducted. Both experiments use the gaze and the smile as nonverbal behaviors, because related research, see chapter one, indicated these are fitting behaviors for initiating interaction. A second experiment was conducted with slightly different behaviors to get different results.

The behaviors of Experiment A are switching every fifteen minutes and only when nobody is in view. The engaging version consists of idle behavior until a person enters his view. Then he looks up and follows the first person to enter his view. Also the agent shows the happiness emotion when somebody is in view. When the person is on one of the sides the happiness is 0%. When he gets closer to the position right in front of the agent the happiness emotion goes up to 35% when he stands right in front of the setup. With a happiness emotion below 36% the smile is a really subtle change in the facial expression (see figure 3) and might have a subconscious influence on the behavior of people passing by. After someone leaves the screen the agents stares for 3 seconds in the last direction and then starts again with the idle behavior. The non-engaging version consists of the same behavior except after looking for three seconds at a person the agent looks away to the other direction followed by idle behavior until the person leaves the screen. The smile behavior of the agent stayed the same, related to the position of the person in front of the agent.

Figure 2 - The subtle smile (seen right) in experiment A

In the second experiment, Experiment B, the behaviors are switching every thirty minutes when nobody is in view. The versions again start with idle behavior and a neutral expression on the face. The engaging version once more tracks the first person who enters the view. But the non-engaging version keeps looking in random directions. Both versions however start smiling when the first person who enters stands still for one second in front of the agent. After this one second the happiness emotion level goes up fast from 0% to 100% and the person is treated with a big smile, when the person starts moving however the happiness emotion level goes down to 0% again.

In both experiments the looking up gaze functions as a distant salutation. The agent tells the persons in view he is prepared to interact with them by looking at them. During the approach of attracted people the agent tracks them with his gaze. Due to time constraints no withdrawals were implemented. And the smile serves as a close salutation to establish a friendly relation.

3.6 Experiment Setup

Figure 4 shows a simplified picture of the experiment. One Pentium 4 personal computer, a firewire camera with wide angle lens and a 19" TFT screen with a large viewing zone were used. The monitor was placed between hip and head height of an average person. No audio systems are placed because this research is focused on the non-verbal behaviors.

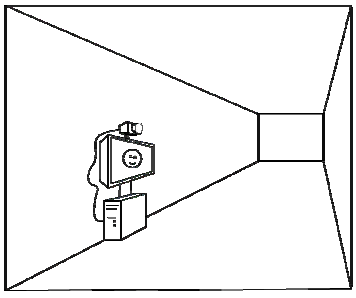


Figure 3 - The experiment setup

3.7 Annotation

The video data collected by the camera is annotated by hand using Media Player Classic and Microsoft's Excel. Media Player Classic lets users change the speed of the videos and step through the frames, which is very handy. Excel makes it easy to use formulas to calculate statistics and diagrams of data. My own actions in front of the camera, mostly for testing purposes, are not annotated. Different aspects of the observation are annotated: start time, end time, type of reaction, already seen, gender, multiple persons and non-verbal behaviors. When somebody steps out of the view and shortly after steps back into view this is annotated as a single observation, because the person was probably still standing in front of the system but just outside the camera view.

Firstly the start and end time is taken from the logs the software created. When multiple persons are in view or the background subtraction showed glitches the start and end times were being added manually. Afterwards the start and end time are used to calculate the duration of the observation which helps to answer the research question.

Secondly the type of reaction is being annotated. The type of reaction tells us something about the level of engagement. This can be "not seeing" to indicate the person has not given any proof on video he/she has seen the system. A person holding a jacket for his/her faces is not part of the "not seeing" category. This person would fit in the next category called "ignoring". Also persons giving a short glance at the agent followed by a constant look the other way are placed in this category. And persons from which is known they already passed and looked at the system and are walking by for the second, third or more time without looking. The following category is "looking at". These persons do look at the agent when passing by, but they don't take time to come closer or respond to it in another way. Persons who do actively (more then looking at) respond to the presence of the agent are placed in the "testing" category. This categorization gives an initial look at the behaviors and the attention levels of the passing persons.

For every person it is notated if the annotator has already seen this person walking by the agent and still remembers this. Some people for example employees working in the surroundings of the experiment are passing multiple times and are easy to recognize multiple times. But because a lot of people have walked in front of the agent not every single one of them has been recognized as somebody who was in front of the agent before. The fact that someone has already seen or tested the agent has influence on the behavior the second time.

The gender type is pretty straight forward. M stands for male and V (from the Dutch word for female) for female. Different responses might be discovered for the different genders.

By using the logs of the system the state of the agent for every observation has been filled in. The two possible states are engaging and non-engaging. And these states hopefully trigger different behaviors.

Also it is annotated if multiple people were in view during this observation. Groups sometimes let people behave differently.

The non-verbal behavior towards the agent/system is notated. The choice has been made not to use the Facial Action Coding System from Ekman [EKM78], because this is far to detailed for this experiment and the available time and because of the limited camera resolution not every facial expression could be seen in the required detail. However examples of behavior that is notated are a certain facial expression like a smile, shaking the head or stepping back and forth. Talking or smiling to other people is not registered. Most of the time the non-verbal behaviors are only applicable for people with the reaction type "testing".

4. IMPLEMENTATION

The implementation consists of two parts. On one side the agent Obie which is created in Java with use of JOGL to call OpenGL instructions. On the other side the camera/vision software in C++ that serves as the vision of the agent. Both are connected to each other using a client-server socket connection.

4.1 Used Development Tools

In Eclipse, a software workspace for Java developers, the agent Obie is extended with an easy to use model-view-controller pattern. The model part creates the visual embodied agent and the needed animations. The controller part handles events and lets both GUI and camera control the model. And the view at last is a simple GUI for testing and socket connection purposes.

To receive and analyze camera input the ParleVision [PAR06] framework is used. This is a framework, also internally developed in the HMI Research Group for vision software. The new processors needed for the ParleVision pipeline for this research were created using Microsoft Visual C++ 2003. Further on they are explained.

4.2 Agent body

The upper body was created by importing a body in the VRML format. Obie contained functionality to convert the VRML format to its internally used format. Afterwards the body was added to the model and placed at the right position. And simple mechanics to control the orientation of the body were added to the java code.

4.3 The Java Package

All the Java classes created for this research form a new package and this forms an extension for the Obie package. This way the existing Obie classes are kept in their original state.

The package has the model, view, controller classes discussed above. Also event and eventlistener classes are present. And a client class for the socket connection.

4.4 The ParleVision Pipeline

Below in figure 5 the pipeline that is created in ParleVision can be viewed. And in figure 6 some example frames are shown.

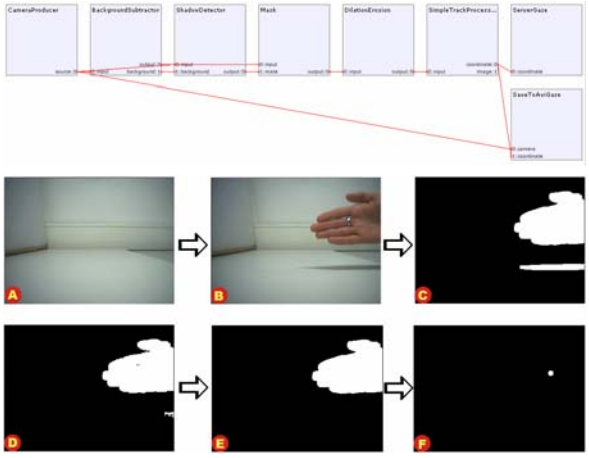


Figure 5 & 6 - The Parlevision pipeline

The pipeline starts at the left with the CameraProducer which delivers the image from the camera (B). The next processor makes it possible to subtract a captured frame (A) of the background from the current camera frames, so only the foreground image is passed on to the next processor in black and white (C). The ShadowDetector detects shadows and with the aid of an inverted mask removes them from the frames (D). This is followed by the DilationErosion processor which removes noise from the image (E). The specially for this research created SimpleTrack processor takes the white blob which was on screen first and calculates it's center of mass (F). The also newly created ServerGaze processor sends the coordinate via a socket connection to the agent. And the last processor, SaveToAviGaze, gets input from both the CameraProducer and the SimpleTrack processor and creates a video from the frames when somebody is in view.

The SimpleTrack processor is just as its name indicates simple. The creation of a tracker was not the main goal of this research and there was limited time available to create one. What it does is it remembers the last blob and compares its location to the new blobs. The closest new blob is used to calculate the center of mass, when nobody is around the coordinate (-50,-50) is send as output. When multiple people move in front of each other along the camera a problem arises. Their blobs make contact from the camera viewpoint and the processor will likely track the wrong blob. First a camera position on the ceiling was considered, this way people's blobs almost never touch each other. Unfortunately the ceilings of the corridors used in this research were not high enough to get a good overview.

The SaveToAviGaze processor looks at the coordinate the SimpleTrack processor returns. When this coordinate is not (-50,-50) somebody is in view and the processor captures the frames from the CameraProducer into a DivX encoded video file. Every hour a new video file is created, because it is known big captured video files become unreadable and shorter videos make the annotating process easier. Also every captured frame has the current time marked in the upper left corner for annotating purposes.

4.5 Standstill timer

For experiment B a timer was needed to give a signal when somebody was standing still in view for a couple of seconds. Therefore a simple timer in Java has been implemented and compares the incoming Center of Mass coordinates. When the changes in the coordinates are below a certain threshold for one second the smile animation is triggered. The first used threshold had to be heightened because the vision software caused fluctuations in the coordinates even when somebody was standing still.

5. ENVIRONMENT VARIABLES

The experiments are taken in the wild and contain a number of variables we do not have under control which could influence the behavior of the people walking by and thus the results. First these differences are discussed before presenting the results.

5.1 Location

The system for the first part of experiment A was placed on the second corridor of the EWI faculty building, called Zilverling, between the Human Media Interaction Master department and the Laboratory/Help-Desk part. In figure 7 a snapshot from the agent perspective can be viewed. The system was placed in a room behind glass looking out to the corridor. The system was also implemented in this room and multiple students were working in this room. Thus these other students had already seen some parts of the system working. The location behind the window gave an advantage because this way the camera, also placed behind the window, captured not only a part of the corridor but also the reflection of the monitor including the agent. This way the interaction between the agent's behavior and the human's behavior can directly be seen on video.

Figure 7 – Location 1 in the HMI corridor

For the second part of Experiment A the system was placed a little further down the corridor to a more visited part of the corridor (see figure 8). It was again placed in a room behind glass. Because it was placed in the same corridor a lot of people who had seen the first part of the experiment also came by the second part.

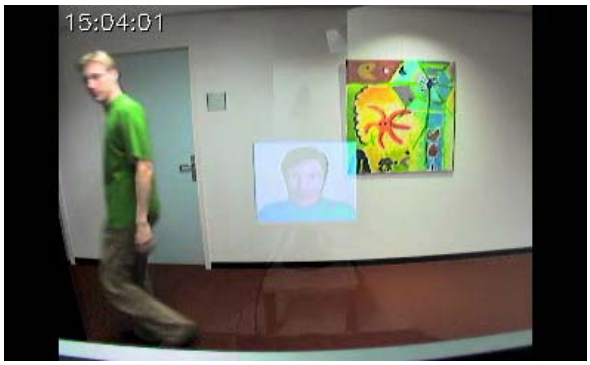


Figure 8 - Location 2 in the HMI corridor

For experiment B a more crowded place with new test persons was needed. The ideal location should not have any groups passing by, but this was hard to find. The chosen location was next to the counter of the janitor near the entrance of the cantina and many college rooms in the Waaier building. A snapshot can be viewed in figure 9 below.

Figure 9 - Location near the Waaier cantina

As the number of observations show (experiment A 220 observations in over 8 hours and experiment B 1399 observations in less then 5 hours) it was indeed much more crowded with people. A lot of groups were going to or coming from lunch or class. The system was placed on a small cart (normally used for televisions), this was a couple of centimeters higher then experiment A. Also the corridor was more then twice as wide and high as the corridor used in experiment A. Because the system was placed on a cart and not behind a window it was more prominent and could be seen from a distance. Sadly without the window the reflection of the agent was not captured on video. For safety reasons a person was required nearby to keep watch of the system. I was sitting about five meters away at a cantina table reading, eating and talking to people and as far as I observed nobody knew or indicated they knew I was watching the system.

5.2 Time of the day

Experiment A was held on a Friday and Monday afternoon. Friday afternoon the location was very quiet. This was excellent to set the system up for the first time, but it did not get many observations to analyze. Monday afternoon on the other hand got a lot more people on camera and thus observations.

Experiment B took place during the morning and afternoon. And the camera captured a lot of groups going to and coming from the cantina lunch during lunch time. Also during college switching times a lot of groups could be seen coming and going.

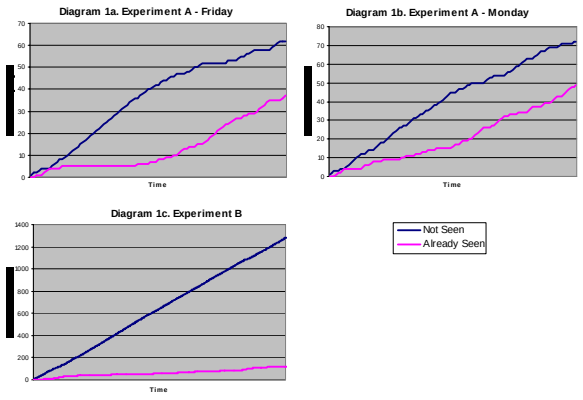
The agent does not behave correctly when groups are in view that's why this difference is indicated. During experiment A 19% of the observations and during experiment B 77% of the observations was from people moving or standing in groups in front of the system.

5.3 Public

Because of the location most of the people walking past the system of experiment A were Human Media Interaction students, professors and employees of both Human Media Interaction and the laboratory.

The location of experiment B however attracted professors, employees and students of various studies including a lot of international students. Also janitors and employees of the cantina were seen often.

Diagram 1 - Number of people that have (not) seen the agent



For both experiments the number of people who have "already seen" the system increases by time and the people who have "not seen" the system decreases. This can be seen in the following charts. Notice the higher number of people in experiment B, which might be because the observer did not to remember all the people who already had seen the experiment and the relatively lower number of people returning to the cantina or the college rooms.

The gender of the observed people was also annotated and gender differences might also have an impact on the results. However the percentage of females remained approximately equal. In experiment A 28% and in experiment B 29% of the observations was female.

5.4 Notification of the camera

Because of privacy regulations people have been warned by a sign of the camera's presence. This also could have influenced the behavior of the people walking by. For example people who did not wanted to be recognized on camera hid their faces before walking by or increased their walking speed. This notification also could have enlarged the percentage of people in the "not seeing" category.

5.5 Background subtraction glitches

During the experiments the intensity of the light in the corridor changed a couple of times, for example because outside the windows the sun was going under and because somebody opened a door and light was coming out from the room into the corridor. This caused bad background subtraction and therefore wrong agent behavior. The data collected in experiment A during these times was not taken into account for the results mentioned below. In experiment A this was 26% (77 of 297) of all the observations. In experiment B it was unknown when and how long wrong agent behavior occurred because the behavior of the agent was not recorded on video. And thus these data is used in the results below.

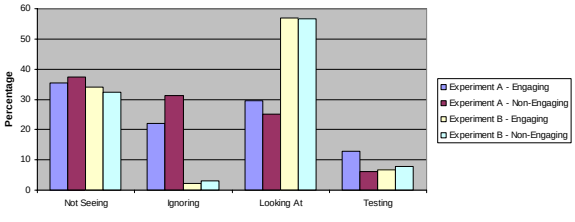
6. RESULTS

In the following diagrams the percentages represent the percentage of the total number of observations for the related experiment.

6.1 Number of Responses

In the following diagram number two the percentage of the observations for the different reaction types can be seen. As discussed above, probably because of the more prominent location and not because the agent behavior, more people have seen the agent in experiment B. The low number of "ignoring" people in experiment B forms a coherent relation with the low number of people seeing the system a second time, see the last chapter. Also the public of experiment A contained relatively more employees walking multiple times past the experiment who started ignoring the system after the first couple of times.

Diagram 2 – The reaction types



"Testing" is overall the category with the smallest percentages. It was hoped the agent's behavior would cause a revolutionary high number of people starting an active interaction but

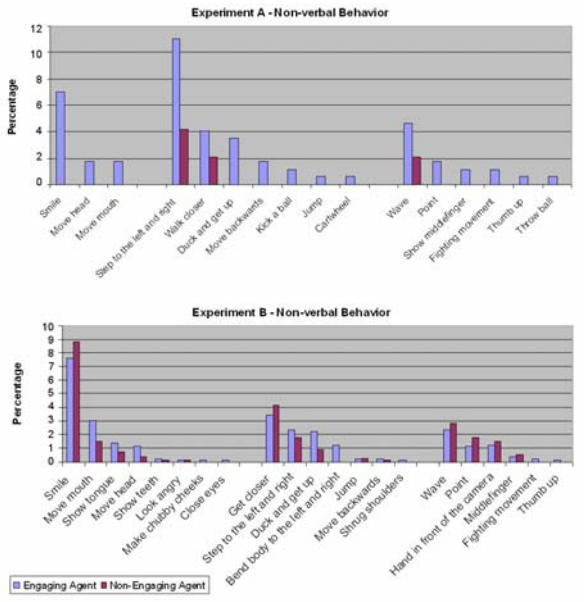
unfortunately it did not. But the low percentage might also be a good thing. [MUKA05] described people without eye-contact need different channels like waving or talking to see if the other is interested. And this might indicate most people indeed felt like they were having eye-contact with the agent.

But more important for this research are the differences between the engaging and non-engaging versions. Most of these differences are however very small and not significant. The small differences for experiment A indicate that the engaging version indeed gets a little more attention and the non-engaging version gets ignored more.

6.2 Non-verbal Behavior

Diagram 3 shows the percentage of non-verbal reactions for both experiments. The percentages are taken from the number of people who passed the engaging agent (A 172 and B 721 observations) and the non-engaging agent (A 48 and B 678 observations) respectively and not the total of people who passed the experiment. The results are presented in three groups. The first group contains non-verbal behaviors concerning the head. The second contains full body behaviors and the third has hand and arm specific movements. After the tables every non-verbal category and its results will be described.

Diagram 3 - Non-verbal Behaviors



The smile category indicates people were showing a happy face towards the agent by lifting the corners of their mouth. The agent shows in both experiments a smile. However the smile in A was hard to notice and in B was hard to miss. Remarkably during experiment A only the engaging agent received returning smiles. But the low scores of non-verbal behavior towards the non-engaging agent might have to do with the low number of people passing by the non-engaging version in experiment A. The non-engaging agent in experiment B got little more smiles than the engaging version. This might be because the non-engaging version showed no interest with his gaze but his sudden smiles and unpredictable gaze might have surprised people and made them smile. The presence of people around might also have something to do with the number of smiles. Because in experiment A 75% and in experiment B 92% of the smile observations was seen while multiple people were present in front of the system.

Move head means people were moving their whole head for example nodding, shaking or tilting the head. This was one of the behaviors people used to test if the agent was tracking them (in this case their eyes). Sadly the agent follows the center of mass of the whole body and moving the head will give little feedback of the agent tracking the person. This might explain why less people show this behavior in the non-engaging version of experiment B, because the non-engaging version of the agent clearly looks in other directions and does not require this test.

Some people opened their mouths once or multiple times while looking at the agent. Because no sound was recorded it is unknown what the reason for the opening was. Sometimes people seemed to talk, yell, yawn and once something like a belch was spotted and all these are categorized under mouth movement. In experiment B the agent visibly opens his mouth when he smiles; perhaps some people were copying this behavior.

People who stick the tongue out of their mouths while looking at the agent were categorized as show tongue. This behavior is later on discussed as a provoking behavior.

Some people opened their mouths but showed obviously their teeth and did not show signs of laughing or talking. This behavior might be comparable to aggressive dogs showing their teeth but it also could be to copy the showing of the teeth during the smile of the agent. This is called the show teeth behavior category and is also later on discussed.

The look angry category contains people who lowered their eyebrows and showed an angry look. This is also a provoking behavior later discussed.

The make chubby cheeks and close eyes (for a couple of seconds) categories speak for themselves. Both were rarely seen and not much can be said about them.

One of the often seen behaviors was the step to the left and to the right behavior. These people were moving sideward from one side to the other while looking at the screen. Why more people in experiment A then in experiment B showed this behavior is unknown. Just like moving the head this might have to do with people testing if the agent is tracking them or something else. The agent that did not track, the non-engaging version in experiment B, also shows this, it gets lower results compared to the engaging version. This also counts for the following category.

Other people did not step sideways but they bend their body to the left and right to, probably, test the system. It is unknown why nobody showed this behavior during the non-engaging version or experiment A.

Some people were walking towards the system to probably get a closer look these are part of the walk closer category. For both behavior types the agent seems to attract people to get closer, this might have to do with the system being something new in the environment. Or people might want to see a close salutation. Or people want to get out of the way of people walking by.

Also some people were moving away from the system but they stay in view, this is the move backwards category. This could be seen as a testing behavior; to see how the agent and his gaze respond to somebody standing further away. Experiment B's corridor was much wider then the corridor of A and often groups were passing behind people standing in front of the camera during experiment B that might explain why more people were observed doing this in experiment A.

People were lowering their body to probably test the tracking system but sometimes people seemed to duck to be at the agent's eye-level. Both are found in the duck and get up category. The head of the agent in experiment B was placed a little higher compared to experiment A. And this might be an explanation of the lower number of ducks in experiment B.

There were a couple people who heightened their center of mass by jumping. These are found in the jump category. The kick a ball category was spotted when two people were kicking back and forth a ball in front of the agent. And the cartwheel and the shrug shoulders categories speak for themselves. These last four had very low percentages so not much can be said about them.

Some people were waving with their hand to the agent these are found in the 'wave' category. Remarkably the engaging agent of the first experiment gets the highest percentage of waves.

A similar behavior was to put the hand in front of the camera but this was probably most of the time to block the camera's sight and to see how the agent would respond without his 'vision'. The reaction of the system while a hand covers the entire camera view is the reaction the agent would give to a center of mass in the center of the view. In experiment A this type of behavior was not observed most likely because the camera was placed at a distance behind a window.

When people were pointing to the agent with their finger this behavior was categorized as 'point'. Sometimes this behavior was done to point the presence of the agent out to a fellow person.

A couple of people showed their hand with only their middlefinger raised to the system this is the 'middlefinger' category. This is one of the provoking behaviors discussed later.

The 'fighting movement' category consists of people showing fighting stances with their arms, moving their fists or other fighting related behavior. This is also a provoking behavior discussed later.

The 'thumb up' speaks for itself. And the 'throw ball' movement was seen once when the two people were kicking a ball in front of the agent. Both have percentages too low to say something useful.

One surprising group of behaviors as seen in the video footage was the number of provoking almost aggressive behavior showing up in Experiment B. The following categories are placed in this group: show middlefinger, show tongue, fighting movement, show teeth, and the angry look. The differences of this kind of behavior between the engaging and the non-engaging versions are not that big.

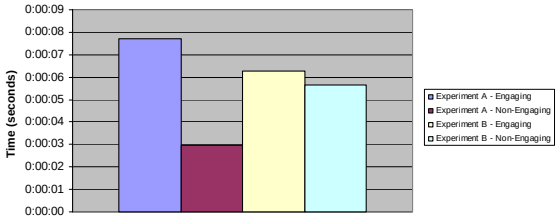
In experiment A 15% and in experiment B 26% of the testing reactions contained at least one behavior out of this group. And of this group observations showing these behaviors in experiment A 100% and in experiment B 89% was male. And of the same observations with a testing reaction and provoking behavior in experiment A 100% and in experiment B 89% of the persons was standing in a group before the camera.

It is unknown why such a relatively large group shows this provoking behavior towards this agent. And why this was mostly done by man standing in a group. It might simply be macho behavior. Or has the agent part in this because he stares at people without withdrawals? Or could it be his appearance? Or maybe people don't like being filmed by the camera? The reason why the percentage in experiment A is lower could be because the kind of public, consisting of many Human Media Interaction students and employees. They are more familiar with agents in general. Or could it be because the agent was behind glass and forms less of a direct threat?

6.3 Duration

Another research question was about how long people will stay in the agent's view. The following chart shows the average time in seconds for all the observations.

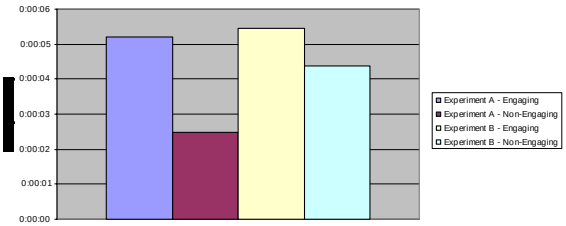
Diagram 4. Time in View and Agent Behavior



In the first experiment the engaging agent clearly lets people stand longer in view then the non-engaging agent. In the second experiment however the difference between the two agent behaviors is too small for a noteworthy result.

When only the average observation time from the "Looking At" category is observed more or less the same results can be seen. Only now the time in view of the engaging version in A is a little shorter.

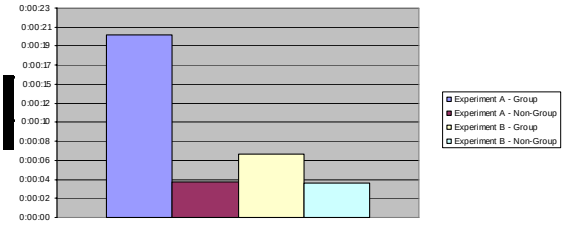
Diagram 5. Time in View and Looking At



And it's clear people who see the engaging version stick around a little longer.

Something interesting shows up when group/non-group and time in view are compared with each other. People in groups tend to stick longer in front of the agent as can be seen in the following chart.

Diagram 6. Time in View and Groups



7. CONCLUSIONS

It is unknown if the solution presented in this paper to the mental inertia, as described in the first chapter, solves the problem. A lot of people showed "testing" behaviors. For example behavior like moving or bending the body and head. This might be an indication that they were testing to see if the agent was looking at them. But this could also be to see how and if the agent responds to certain behaviors. Maybe research should not look into how to overcome this problem. But look

into how to use the mental inertia to give the appearance of personal interaction to multiple people at once.

Some other interesting results showed up during the experiments. People react to the non-verbal behavior of the agent in different ways. Sadly the number of people actively starting an interaction with the engaging agents was not as high as was hoped. The engaging agent of experiment A was however capable of keeping people longer in view in contrast to his non-engaging counter-part.

People show a lot of different non-verbal behaviors in front of the agent. Surprisingly 15% in the first experiment and 26% in the second experiment of the observations of people actively starting an interaction shows provoking or aggressive behavior towards the agent. And almost all of these behaviors were done by males standing groups.

Also the other results show that the presence of multiple people plays a big role in how people respond to the system. And this might be interesting to look into in future research. Further points of interest to look into in future research can be found in the next chapter.

8. FUTURE WORK

[PETE05] talked about the importance of mutual gaze on the attention level of both parties. In future research an agent which can detect the gaze of other people and its direction might be able to attract people more subtly, more natural and look them straight in the eye in stead of looking at their center of mass. The time of people in view during this experiment is not that long. The research done in [CARY78] points out an initial mutual gaze followed by a second mutual gaze has a positive correlation with starting and keeping a conversation. If this future agent with gaze detection also has withdrawals implemented it can create multiple mutual gazes and might increase the time people stay in view.

Also a better location would have a high ceiling which would allow a second camera to view the corridor from above. This way the distances between people and the agent, which play a role during salutations [KEND90], can be measured. And there would be less tracking problems of people moving in front of each other. And the paths of how people move through the corridor past the agent can be researched.

Another improvement would be to use a camera with a wider angle or multiple cameras. This way people standing just outside the agent's vision can still be captured on video. And the effect they have on the people standing in front of the agent can be seen.

Also the recording of the sounds and conversations people have in front of the agent would give better insight of why they sometimes open their mouth. And what their ideas are concerning the agent.

REFERENCES

[CARY78] Mark S. Cary. *The role of gaze in the initiation of conversation*. Social Psychology Vol. 41 No. 3, 269-271, September 1978.

[CASS00] J. Cassell, *More Than Just a Pretty Face: Embodied Conversational Interface Agents*. Communications of the ACM 43 (4), page 70-78, 2000.

[DUYB04] Duy Bui, *Creating Emotions and Facial Expressions for an Embodied Agents*, Master Thesis, University of Twente, 2004.

[EKM78] P. Ekman and W.V. Friesen, *Facial Action Coding System*, Consulting Psychologists Press, Palo Alto, CA, 1978.

[KEND67] Adam Kendon. *Some Functions of Gaze Direction in Social Interaction.* Acta Psychologica 32, pages 1-25, 1967.

[KEND90] Adam Kendon. *Conducting Interaction: Patterns of Behavior in Focused Encounters*. Studies in Interactional Sociolinguistics. Canbridge University Press November 1990.

[KUNO04] Yoshinori Kuno, Arihiro Sakurai, Dai Miyauchi, Akio Nakamura. *Two-way Eye Contact between Humans and Robots*. ICMI'04, State College, Pennsylvania, USA, October 2004.

[KUNO05] Yoshinori Kuno, Dai Miyauchi, Akio Nakamura. *Robotic Method of Taking the Initiative in Eye Contact.* CHI 2005, Portland, USA, 2005.

[LEE02] Sooha Park Lee, Jeremy B. Badler, Norman I. Badler. *Eyes Alive*. International Conference on Computer Graphics and Interactive Techniques, Pages: 637-644, San Antonio, Texas, 2002

[MART82] Wade Martin, Marta Rovira. *Response biases in eye-gaze perception.* The Journal of Psychology, No. 110, Pages 203-209, 1982.

[MUKA05] Naoki Mukawa, Tsugumi Oka, Kumiko Arai and Masahide Yuasa. *What is Connected by Mutual Gaze? User's Behavior in Video-mediated Communication.* CHI 2005, Portland, USA, April 2005

[PAR06] *ParleVision: Rapid Development Of Computer Vision Applications* http://hmi.ewi.utwente.nl/showcases/ParleVision, last visited on 13 December 2006.

[PET6A] Christopher Peters, *Evaluating Perception of Interaction Initiation in Virtual Environments using Humanoid Agents*, University of Paris, 2006

[PET6B] Christopher Peters, *A Perceptually-Based Theory of Mind for Agent Interaction Initiation*, International Journal of Humanoid Robotics, University of Paris, 2006.

[PETE05] Christopher Peters, C. Pelachaud, E. Bevacqua and M. Mancini. *Engagement capabilities for ECAs*. In Fourth International Joint Conference on Autonomous Agents & Multi-Agent Systems, AAMAS '05, Workshop: Creating bonds with humanoids, 2005.

[REID06] D. Reidsma, H. van Welbergen, R.W. Poppe, P. Bos, A. Nijholt, *Towards Bi-directional Dancing Interaction*, Proceedings of the 5th International Conference on Entertainment Computing, volume 4161, page 1-12, Cambridge UK, 2006.

[VILH98] Hannes Högni Vilhjálmsson, Justine Casell. *BodyChat: Autonomous communicative behaviors in avatars*. In AGENTS '98: Proceedings of the second international conference on Autonomous agents, pages 269-276, New York, USA, 1998.