**KTH Computer Science
and Communication**

Modeling of Natural Human-Robot Encounters

NIKLAS BERGSTRÖM

nbergst@kth.se

2008-04-08

Supervisor: Danica Kragic

Examiner: Danica Kragic

Abstract

For a person to feel comfortable when interacting with a robot, it is necessary for it to behave in an expected way. This should of course be the case during the actual interaction, but is in one sense even more important during the time preceding it. If someone is uncertain of what to expect from a robot when looking at it from a distance and approaching it, the robot's behavior can make the difference between that person choosing to interact or not.

People's behaviors around a robot not reacting to them were observed during a field trial. Based on those observations people were classified into four groups depending on their estimated interest in interacting with the robot. People were tracked with a laser range finder based system, and their position, direction of motion and speed were estimated. A second classification based on that information was made. The two classifications were then mapped to each other. Different actions were then created for the robot to be able to react naturally to different human behaviors.

In this thesis three different robot behaviors in a crowded environment are evaluated with respect to how natural they appear. With one behavior the robot actively tried to engage people, with one it passively indicated that people had been noticed and with the third behavior it made random gestures. During an experiment, test subjects were instructed to act according to the groups from the classification based on interest, and the robot's performance was evaluated with regard to how natural it appeared. Both first- and third-person evaluations made clear that the active and passive behavior were considered equally natural, while a robot randomly making gestures was considered much less natural.

Referat

Modellering av naturliga möten mellan människa och robot

För att en person ska känna sig trygg under en interaktion med en robot måste roboten bete sig på ett förväntat sätt. Detta gäller själva interaktionen, men även perioden innan den, vilken på ett sätt är viktigare för en persons intryck av roboten. Om man är osäker på vad man kan förvänta sig av roboten när man betraktar den på håll och närmar sig den, kan robotens beteende vara en avgörande faktor i valet om man ska interagera eller inte.

Under ett förberedande experiment observerades människors beteende runt en robot, som i detta fall var ovetande om dem. Dessa observationer användes sedan för att dela upp människor i fyra klasser beroende på hur intresserade de verkade vara av att interagera med roboten. Med hjälp av ett spårningssystem baserat på lasersensorer, uppskattades deras hastighet, position och rörelseriktning varefter de utifrån denna information delades in i nya klasser. Sedan bestämdes sambandet mellan den första indelningen och denna klassificering. Olika rörelsemönster skapades åt roboten med syfte att på ett naturligt sätt bemöta olika beteenden hos människor.

I denna rapport utvärderas tre olika robotbeteenden i en omgivning med mycket människor med avseende på naturlighet: ett beteende där roboten aktivt försöker engagera människor, ett där den passivt visar att människor har upptäckts, och ett där den slumpmässigt gör rörelser. Ett experiment genomfördes där försökspersoner instruerades att bete sig enligt den gruppering som gjordes med avseende på människors intresse, och robotens beteenden utvärderades med avseende på hur naturliga de verkade. Både första- och tredjepersonsutvärdering visade att de aktiva och passiva beteendena var likvärdiga, men en robot som rör sig slumpmässigt upplevdes mycket mindre naturlig.

Acknowledgements

This thesis was made possible by a number of people and organizations. The research was performed at Advanced Telecommunications Research Institute International (ATR) in Kyoto, Japan. I would like to express my sincerest gratitude to Prof. Henrik Christensen and Prof. Hiroshi Ishiguro, for giving me the opportunity to spend six months at ATR, and to Dr. Takayuki Kanda and Dr. Takahiro Miyashita for supervising my research. Furthermore I would like to express my sincerest gratitude to Sweden-Japan Foundation for financial support. This research was also supported by the Ministry of Internal Affairs and Communications of Japan.

Contents

1	Introduction	1
1.1	Why Robotics	1
1.2	Robots in a Human Context	2
1.3	Outline	3
2	Problem Description and Motivation	5
2.1	Background	5
2.2	Tracking Humans	6
2.2.1	Identification and Feature Extraction	6
2.2.2	Estimation of Intentions	7
2.2.3	Selection	7
2.3	Designing Behaviors	8
2.3.1	Making the Robot Act Naturally	8
3	Related Work	9
4	Field Trial	13
4.1	Observations	13
4.2	Classification	15
5	System Description	19
5.1	An Overview of the System	19
5.2	Robot Platform	20
5.2.1	Controller Mechanisms	20
5.2.2	Communication Abilities	21
5.2.3	Designing Behaviors	21
5.3	The Sensor Network	22
5.3.1	The Human Tracker	22
6	Modeling Natural Encounters	25
6.1	Classification of Humans	25
6.1.1	Classification of Speed	26
6.1.2	Classification of Direction	26
6.1.3	Classification of Position	27

6.2	Determining the Robot's Beliefs	28
6.3	Selecting Target Human and Action	30
7	Experiment	33
7.1	Setup and Implementation	33
7.1.1	Setup	33
7.1.2	Implementation	33
7.2	Evaluation	34
7.2.1	Observations	35
7.2.2	Data analysis	35
7.2.3	Discussion	36
7.2.4	Sources of Error	37
8	Conclusions and Future Work	39
8.1	Conclusions	39
8.2	Future Work	40
	Bibliography	41
A	Acronyms and Definitions	43

Chapter 1

Introduction

Autonomous robots are at an increasing rate leaving the research labs and entering people's everyday environments. Unlike industrial ones, these robots possess the ability to reason about their environment and to adapt to unexpected situations. Toy robots have been available for some years, like the robot dog [3] that can react to commands and adapt to its environment. Household robots, like autonomous vacuum cleaners [19] and lawn mowers [18], are examples where robots are expected to operate in shifting environments, necessitating abilities to make own decisions. So called humanoid robots¹ are still a rare phenomenon in daily life situations, but they can for example be seen acting as receptionists at exhibitions [20]. With these robots, other dimensions about robots' characteristics are introduced.

1.1 Why Robotics

Throughout history, man has always invented new tools for facilitating life and improving the quality of life. The research in robotics lies along this path. Industrial robots have taken over difficult and dangerous work from humans, and have rendered possible the production of goods with higher quality and at a higher pace than otherwise would have been possible. Teleoperated robots are used for tasks like disarming bombs and exploring sunken ships; situations constituting considerable danger. However, in some situations it is not possible to remotely control robots, for example ones that are sent to the planet Mars where the time for sensor data to reach Earth, and for commands to reach the robot, is too long. Required to operate autonomously, they fetch information from sensors, evaluate it and take appropriate actions in accordance with their directives. These robots are machines that operate without the need to interact with humans, so they can be constructed to best suit the situation, regardless of how they look.

However, when robots are created to operate together with humans, there are additional requirements on their behavior as well as their appearance. One example is household robots with abilities reaching beyond those of the robots just

---

¹From this point, if nothing else is said, these humanoid robots are simply referred to as robots.

performing one task, like the vacuum cleaner and lawn mower mentioned above. In Japan, where an aging population combined with a very low birthrate undermine the country's ability to take care of elder citizens, this kind of robot is seen as one solution to the problem, acting as an aid, as well as company, in people's homes.

1.2 Robots in a Human Context

Research is continuously advancing, algorithms are being renewed or refined, computers are becoming more powerful and sensors more accurate. As a consequence, robots are becoming more and more sophisticated. Their voice recognition software makes them able to make out larger portions of what is being said, their ability to navigate in their environment without running into obstacles is improving, they are becoming better at determining what they see and who they see, and not least, the two legged robots are walking better and are now even starting to run. As these capabilities get refined, the robots also gain the prerequisites for being able to behave like humans. When people look at these robots, they tend to unconsciously give the robots human characteristics. They expect the robots to perceive them, to hear them and to reply, in other words, to be aware of their presence.

The field of Human-Robot Interaction (HRI) deals with this issue, making the robots interactive and behave in a way that we would expect them to. This ranges from robots being able to move in a predictable way in a crowd, to robots being able to carry out a conversation. The type of interaction dealt with in this thesis is the latter. Most research in this field is focused on the core of the interaction, that is from the moment that the robot says something, or is being spoken to, up to the moment that either the robot or the person moves along. In contrast to other research on HRI, such as path planning where the robot must take into consideration people far away in order to pass them without running into them, this kind of research focuses on what happens when a person is standing close to the robot. However, when two people interact, the interaction might have begun before they reach the point where they are standing close to each other talking. Imagine the scenario where two people are walking towards each other, one spotting the other and trying to establish eye contact, then slowing down, then stopping when a comfortable distance for a conversation has been reached. This would of course come naturally to two humans, so for a robot it is also important to be able to detect if someone is seeking to interact with it. If unable to deal with this part of the interaction, people might be reluctant to proceed with their intentions and choose to not interact.

Most people have never seen a robot in real life, let alone interacted with one, which puts substantial requirements on the robot's behavior as they start to populate environments people visit in their daily lives, like information centers and receptions. One natural objective for the robot is to not cause damage or injury to anything or anyone. If it is acting inappropriately, like not considering people's whereabouts, thereby running into them, or making intrusive gestures with its arms,

1.3. OUTLINE

this will of course reduce people's trust in the robot. Thus, to fulfill its purpose, the robot has to act in a predictable way and to clearly signal its capabilities so that people know what they can expect from it. A vital part in order to establish a connection is that the robot shows that it is aware of people around it. As much as a behavior where the robot shows that it is aware of others will make people feel noted, elevate their confidence in the robot and make them feel more comfortable with it, not displaying this sense of awareness might have the opposite effect. People might become uncertain of how well the robot can perform, or even if its working at all. For robots having important functions, like robots giving directions at an airport, failing to encourage people to interact will render them less useful.

1.3 Outline

The thesis is structured as follows:

Chapter 2 outlines the research done for this thesis. It gives motivations for different design choices and describes the method used.

Chapter 3 gives an overview of related work and describes why the research presented in this thesis is relevant.

Chapter 4 describes the field trial where data for the experiment was gathered. A detailed description on how this data was used is given and some general observations on people's behaviors are also made.

Chapter 5 describes the framework used for the experiment. A overview of the hardware and software that was used is given.

Chapter 6 presents the controller software that was developed as well as a detailed description of the method that it is built on.

Chapter 7 treats the experiment. A description is given of how it was performed, and the results from the evaluation done by test subjects and the evaluation done by observers of the experiment are presented.

Chapter 8 draws conclusions of the results from Chapter 7 and gives suggestions to future work.

Chapter 2

Problem Description and Motivation

This chapter describes the problem studied in this thesis in detail. Design issues concerning tracking and identification of humans are dealt with, as are issues related to the creation of behaviors for the robot.

2.1 Background

The research done for this thesis focuses on identifying people's behaviors and to correlate their behaviors with their movements. With that as a starting point the task is then to make the robot behave appropriately during the first part of the interaction, up until a conversation begins. The purpose is to make the robot aware of people moving around it and to equip it with behaviors to deal with different situations that might occur. In that way people will understand that the robot is aware of them and will have a better idea what to expect from it. A method will be developed for identifying people at a distance and to map their movements to their intentions towards the robot. Observations from a field trial will serve as a base for this mapping. Three different behaviors will be designed for the robot, and will be compared against each other. To test the robot behaviors an experiment will be performed. People participating in the experiment will be asked to evaluate their experiences of how the robot was behaving, and third-person observers will be asked to give their impressions when looking at different situations from the experiment.

As mentioned in Section 1.2, an interaction does not begin when the first word is spoken. Rather it begins when a mutual understanding has been established between both parties that they intend to initiate a conversation or in some other way exchange information. The intention to interact can be two-sided, for instance if they are friends, or one-sided, for instance when asking someone for directions. When the interaction involves two humans, this process usually occurs without any problems or misunderstandings. Eye contact is often established, and there is an unspoken understanding between the two what their intentions are.

2.2 Tracking Humans

A possible application for a robot is to act as an information provider, for instance at a station or at a mall. Several studies have been done in this field, for example robots acting as exhibition guides at a museum (Shiomi et al. [13]), and robots giving directions at a train station (Hayashi et al. [6]). Neither of these studies concerns the initial part of interactions, but both present situations where inclusion of it might have proven useful. Common for these studies are that they are both set in very crowded environments. One difference is however, that in a museum, most people are there for a purpose and would be interested in the robot's services, while most people passing a train station would not. In crowded environments two people would not have any major issues initiating an interaction once they spotted each other. If however a person acts as a provider of information, several people might simultaneously be interested in asking questions. This is indicated by the people trying to seek eye contact or walking towards the person. In this case the problem to decide an appropriate behavior to make all the people feel treated well might be non trivial even for humans.¹ Which person should be prioritized? For the robot to deal with these problems, several obstacles have to be overcome.

1. The robot should identify the people around itself.
2. The robot should to assess people's intentions.
3. The robot should select one of the people it has classified as wanting to interact, and act appropriately.

2.2.1 Identification and Feature Extraction

Related to the first problem are a number of decisions that have to be made. How big is the area around the robot that should be covered? Should the robot use its own sensors to identify people, or should external sensors be used? What information about the humans should be generated? As noted by Pacchierotti et al. [11], interactions take place within distances up to 3.5 meters. The area outside that zone mainly hosts one-way communication. (Depending on the nature of the interaction, i.e. if it is between two friends or a more formal interaction, different zones are used [11][16]). Thus the robot should at least be able to identify humans within a circle with radius of 3.5 meters, but since it should be able to discover people before the core interaction begins, an even larger area should be covered. The size of the area covered, combined with the risk of occlusion suggests that the robot's own sensors might not prove to be accurate enough, and the usage of external sensors would be needed.

---

¹When approaching an information desk, it is obvious what services are provided, and so the need to establish a mutual understanding is not necessary. In regards of a robot however, it might not be obvious what its capabilities are or what services it can provide, which makes it more important to give clear signals about this.

2.2. TRACKING HUMANS

So what information should be generated? A feature that would be helpful is the rotation of the person's body, as that gives an indication of where the person might be looking. This information can be obtained with the sensors and software used for this thesis, but it turned out to perform too slowly when the number of surrounding humans increased (see Section 5.3). A required feature is of course the position of the human, and that is also the only information obtained from the sensors used in this experiment. With this solution no information on the orientation of the body is available, but when a person is moving, the direction of motion can be approximated. Since people tend to walk straight ahead, their orientation is assumed here to equal the direction of motion. Ideally the gaze should be determined, as the gaze is very important for the robot to estimate whether a person is interested in it or not [1]. This would require the data from the sensors to be fused with information from cameras, which were not used as the focus for this thesis has been on using positions of humans.

2.2.2 Estimation of Intentions

Apart from the positional information obtained from the sensors, speed and direction of motion can be estimated. The approach used in this thesis is to estimate the intentions of the surrounding humans towards the robot, i.e. the robot's beliefs about the humans, based on this set of properties. One issue is how to use the sensor values and generated values. Should they be used directly as continuous variables, or to split into discrete classes? Another issue is whether the robot's beliefs should be described as continuous variables, or as discrete ones. Furthermore is a problem that it is difficult to know the rotation of the body when someone is standing. The intentions of a person facing the robot compared to one facing another direction would be estimated rather differently.

2.2.3 Selection

Once the robot's beliefs have been established, the question remains how to choose among the ones believed to be interested, and how to act appropriately. If only one such a person exists, the problem is trivial. When having to choose between two or more humans, the choice is more difficult. One of them might seem more likely to approach, but if the robot do not get any reaction from the person it has directed its attention towards, when should it shift its attention? Furthermore, how long should it at least focus on one person before shifting, even if another person seems more interested? If that period is too short, the robot might appear confused, constantly shifting its attention to different people. Related to this problem is what types of actions that the robot should perform that are appropriate, which is dealt with below.

2.3 Designing Behaviors

The problem dealt with when the robot selects an action is twofold.

- To detect at an early stage if someone wants to interact with the robot, and to act accordingly.
- To identify if someone seems to be interested in the robot, but for some reason is reluctant to initiate an interaction, and encourage the person to approach.

To handle these two problems two types of actions are required. A welcoming one that gives an already interested person the impression that the robot has noticed her, and an encouraging one that actively seeks to make the person approach.

2.3.1 Making the Robot Act Naturally

The remaining issue is to select an appropriate action in accordance with the robot behavior when a person has been selected. For this two things are taken into consideration: how the selected person is currently moving, and how the other people around the robot, if any, are moving. The goal of the selected person-action combination is for the robot to behave as *naturally*² as possible.

Actions can be considered to have different costs attached to them. The cost attached to just turning the robot's head is low, as the time it takes to return the head to the initial position or look in another direction, is short. Thus this action can be applied without much of thought. If the target human turns out not to be interested, it will not take the robot long to switch its focus if another person seems more interested. As a contrast, to turn the entire body or to move around, a very high cost is attached. The rotation of the entire robot is slow, and the time it takes to rotate in order to face the newly selected person is much longer than to just turn its head. Therefore, for the robot to be considered behaving naturally, it must choose its actions carefully. If it were to turn the body and approach every time someone was selected it might result in the robot appearing rather confused, constantly turning and driving towards different people.

So when choosing a high cost action, the robot must be quite sure that the person is interested. It is also safer to choose a high cost action if the target human is the only one present, or the only one that seems interested.

---

²The properties of natural, as used here, are described in Section 7.2.

Chapter 3

Related Work

As humans we have the ability to read other people's signals, often unconsciously, to adapt our own behavior to others' in order to not make them feel uncomfortable. We evaluate for instance a person's posture and gaze in order to get an idea about how comfortable that person is feeling in that situation, and might adjust our own behavior to accommodate the counterpart's preferences. For a robot to be able to blend in and to avoid causing discomfort to humans, it is important for the robot as well to be able to adapt to different people and circumstances. This has been studied by Mitsunaga et al. [10], who made a robot adapt its behavior to each test subject when detecting different signals of discomfort. The purpose for this thesis is also to make people feel comfortable around the robot, but does not go so far as adjusting the robot's behavior for each person.

The ability to adjust our behavior very much depends on where in the world we have been raised. Cultural phenomena define how we interact in different social settings. When traveling to different countries it is not uncommon to end up in awkward situations, so both knowledge of the culture and an ability to adapt to new situations is essential in order to avoid them. Moreover our attitudes towards different phenomena depend on our background. People from different cultures have different attitudes towards robots. This was shown by Bartneck et al. [2] who concluded, quite contrary to common beliefs, that the Japanese presented the most negative attitude in the study.

When talking about HRI the first things that probably comes to mind is a robot talking to a person or a group of people. Much research has been and is being done concerning these rather explicit interactions, for instance [13], [16] and [17]. However succeeding with much more subtle interactions, that might happen on a person's unconscious level, is very important to get a good impression of the robot. In the case of talking to a robot, our impression is to a large extent determined by how naturally it speaks, how well it appears to understand questions, and if it seems to be talking to us and not just talking. As opposed to this, in other situations we might find the robot natural if we never notice it. Take the case where a robot is moving around among people, then we might only get an impression of the robot

if it is in our way or meeting us very closely. In this case the impression will be negative one. Hall [7] suggests that the area around a human can be divided into zones depending on the nature of the current interaction. The intimate space (less than 0.45 m) for close friends, the personal space (between 0.45 m and 1.2 m) for friends, the social space (between 1.2 m and 3.5 m) for acquaintances, and the public space (more than 3.5 m) for one way communication. For people to feel comfortable when meeting a robot in environments like a corridor, the robot should stay out of the intimate space (Pacchierotti et al. [11]). Sisbot et al. [12] takes this a bit further, also considering that the robot should be visible to the person and not suddenly appear from behind, even if staying out of the intimate space. In this thesis, the robot was positioned in a way that the people who were possible to interact with had a clear view of the robot. In continuation of the experiment presented in this thesis however, one possibility is for the robot to move around freely in its environment. In this case it is essential to act in a way that does not impose any discomfort on people.

Imagine the scenario where a robot works at for instance an airport, accompanying people from point A to point B. In this case it is of course necessary to avoid other passengers walking through the terminal at a comfortable distance, but it is equally important to make sure that the people following the robot not fall behind, to move at a pace that suits them and to keep a comfortable distance to them as well. Thus, in contrast to avoiding people, other research focuses on robots following people or people following robots. In [15] Svienstins et al. make the robot to adapt its speed depending on the preference of the person. They conclude that both the speed and the distance to the robot are highly individual characteristics, and even when looking at recordings of their experiments, noticing signs of walking with a speed differing from the preferred one is difficult to detect. Like the experiment conducted for this thesis, the only sensor used was laser range finders and the only information generated was position and speed.

When evaluating the performance of a robot in a certain setting of a human-robot interaction, it is possible either to look from the robot's perspective or the human's perspective. In the case of a robot, measurable properties like long interaction time or high interaction frequency can be measures of success, while when evaluated by people, properties like human likeness and naturalness that cannot be measured are used. The performance when measured from the robot's perspective might differ from the performance when seen from a human's perspective. This is noted in [5], where Gockley et al. compare the performance of a robot that is following a person either by going in the direction of the person or taking the same path as the person. Although the performance was equal for both methods when analyzing the time before tracking failure and followed distance, the impression from an observer's point of view was that the robot acts much more as expected and human like when going in the direction of the person. For this thesis the only things evaluated are people's impressions on how natural the robot acts, no quantifiable measurements have been used. If the experiment had been held in an environment in real life, it would have been justifiable to measure the frequency of

people approaching the robot. In a controlled experiment however, if a test subject approaches the robot, it can not be correlated to how well the robot behaved. It can also be questioned whether engaging the highest number of people possible is something to strive towards for a robot providing information at a mall.

One aspect of a human-robot interaction is whether the robot is addressing just one person or addressing a group. If belonging to a group, people might all have the same purpose for interacting with the robot, and so the robot can treat them all as participating in the same interaction. If however they were treated as individuals, all with different agendas, the robot would have to interact with them individually. Shiomi et al. [14] have developed a method to accurately determine the group state of people in order for the robot to decide what behavior to apply, if it can interact with the group as it is or if it needs to proceed in a different manner. For the experiment in this thesis people are treated like individuals and the test subjects were instructed to act as such, but if deployed in a real environment, people will arrive in groups and it will be necessary for the robot to be able to cope with both individuals and groups.

A work closely related to this thesis is [9]. Instead of having a mobile robot, Michalowski et al. used a robot placed behind a desk to act as a receptionist. Using only a monitor with pan and tilt capabilities acting as a face, and a loudspeaker, the robot's behaviors were limited to turning towards people and talking to them. The robot's task was for instance to attend to people when they were approaching, greet them in an appropriate manner and carry out an interaction. Using both a laser range sensor and a camera, they were able to determine people's positions and, if standing close to the robot, their gaze as well. Not using information on direction of motion however, they were not able to identify if people were just passing the area or if they were standing there. A majority of the people that the robot turned to were passing the robot, and often after they already had passed. This suggests that it is essential for the robot to include people's direction of motion and speed when assessing their intentions towards it, something that is considered in this thesis. Furthermore they found that over half of the greetings towards people were not followed by an interaction, and that a misdirected greeting often appeared burdening, which indicates that either a more restrictive usage of utterances or a better prediction of people's intentions should be developed.

When designing a robot like the robot receptionist, one must ask what its purpose is, and then decide what criteria to use for evaluation. Michalowski et al. have compared the robot's internal state during the experiment to a video to calculate the amount of misdirected behavior from the robot. Their goal is then to minimize those numbers and to increase the number of interactions. In the case of this thesis on the other hand, the goal is to make the robot behave as naturally as possible, so that people who want to interact feel comfortable doing so.

Chapter 4

Field Trial

In this chapter the field trial that was used as a starting point for the experiment is described. Observations on people's behaviors were mapped with their movements to be able to later create a similar environment.

The experiment for this thesis was held in a controlled environment, but due to the nature of the experiment, it would not be of any greater value if the simulated situations were not plausible. Data and observations were taken from a field trial held at a newly opened mall, called AEON, close to ATR. During six weeks, each weekday from 1 p.m. to 5 p.m., people were able to interact with the robot: ask the robot about directions to a shop or a restaurant, or play a simple game like rock paper scissors. During the first two weeks it was also possible to sign up for individual RFID-tags identifying people, to enable the robot to call people by name, and to remember if they had interacted before. The robot was placed in an area next to a pair of escalators, facing a couple of stores (Figure 4.1 and 4.2) where there was a continuous flow of people. The robot was restricted to only rotate, it was not able to move backward or forward, which meant that interested people had to approach the robot themselves in order to interact with it. To initiate an interaction, people had to step onto a grid of floor sensors, which constituted a quadratic area around the robot with two meter sides. To end an interaction, the robot said goodbye, and resumed its idling action once the person left the floor sensors. During the field trial at AEON, the robot had no information on people standing outside the floor sensors, and the effects of this could be observed.

4.1 Observations

A few observations could be made about the environment and the people moving in it. First, since the field trial was held during normal work hours, the environment quite naturally was populated with people not working. Thus a majority of the observed people were mothers with children, and elderly people. Also school children in ages from around 12 to 15 made up a large part.

Figure 4.1: Overview of the layout of a part of the second floor at AEON and a detailed figure of the setup.

People who seemed interested in the robot exhibited a large variety of behaviors. (The behaviors of people in possession of an RFID-tag were of course biased in the way that they on forehand had shown an interest in the robot and had the intention of approaching it when coming to the area.) There were both people who approached the robot without any hesitation, as well as those who walked back and forth outside the floor sensors, trying to attract the robot's attention, only to loose interest and leave. The majority of the ones who interacted with the robot were children, sometimes accompanied by their parents on the floor sensors, and sometimes on their own. The behavior among them also differed very much. Some children were very eager to interact and tried to get the robot's attention while there already was an ongoing interaction, while some hid behind their parents and required much persuasion before stepping onto the floor sensors. It should be said

Figure 4.2: The AEON environment.

4.2. CLASSIFICATION

however, that a majority of the people passing the robot did not show any interest at all and walked past it without giving it any attention.

As noticed by Hayashi et al. [6], although in that case it concerned social interactions between two robots, people are more likely to stop and watch the robot in an ongoing interaction than they are when the robot just is acting and talking on its own. This could also be noticed when observing people's behaviors while the robot was interacting. Sometimes several minutes could pass without an interaction, but once someone approached the robot and started to interact, people tended to gather around it to watch the interaction, and to interact themselves once the other interaction was completed.

The floor sensors, covered by a purple mat, seemed to be an obstacle to some people, who were reluctant to step onto them. They still tried to get the robot's attention by waving at it, and moving sideways to get it to follow them with its head. However, since people were not standing on the floor sensors, the robot was not aware of their presence and often people lost interest and moved along. This clearly indicates the need for the robot to be aware of people standing at least within its social space, and to behave in an encouraging way to engage people in an interaction.

4.2 Classification

The field trial was used for two purposes. Firstly, in order to reconstruct the AEON environment in an accurate way, but without exactly imitating different situations that occurred, a classification of people's behaviors and appeared intentions was made. These different classes of behaviors, along with their occurrences, were then used when reconstructing situations in the controlled environment. Secondly, people's movements were tracked and mapped to their behaviors. This was done in order for the robot to be able to determine people's intentions based on their movements during the experiment.

Data from the last week of the field trial was used in order to reduce the novelty effect of the robot as much as possible. However, the robot remained a big attraction throughout the entire field trial, and most of the time people were waiting in line in order to interact with the robot. The rest of the time people's behaviors were observed and classified into four groups, depending on how interested they seemed and how they were acting towards the robot. The only people considered were those actually moving in a direction towards the robot, or standing close enough, as they are the only ones that might be of interest for the robot.

As mentioned above, some people had been given RFID-tags and came to the area with the robot for a purpose. For this reason their behaviors were different from the ones coming to the area not knowing about the robot, and so they were excluded when doing the classification. Furthermore, people with no interest in the robot are of course of no interest for the robot, why people who were passing the area on the opposite side, in the robot's public space, without paying any

Figure 4.3: Trajectories of people who interacted with the robot. The black lines represent people labeled *Interested* and grey lines people labeled *Indecisive* or *Hesitating*.

attention to the robot were also excluded. These people were easily identifiable as not interested, both when looking at their behaviors and their movements, and could not be mistaken for someone who might be interested.

People were divided into four major groups according to the following:

- *Not Interested* – People coming in from the side walking towards the robot, but then diverting from the current trajectory, and passing the robot. They are not slowing down, something that could have indicated an interest in the robot.
- *Indecisive* – People walking in front of the robot, but slowing down or making a brief stop when doing so, thus indicating their interest in the robot.
- *Hesitating* – People clearly showing an interest in the robot by stopping in front of it and facing it. Someone standing for three seconds or more was classified as hesitating.
- *Interested* – People coming from any direction and directly approaching the robot to start an interaction.

The ratios of the groups above were as following:

- *Not Interested* – 53 %
- *Indecisive* – 13 %
- *Hesitating* – 27 %
- *Interested* – 7 %

Michalowski et al. [9] experienced people dividable into three groups: people with no intention of interacting, people who will interact no matter what and undecided people. There were however large variations among the group of undecided

4.2. CLASSIFICATION

Figure 4.4: People who are all labeled *Hesitating*.

people, so in order to imitate the AEON environment more accurately and to be able to better control the robot’s behavior, the undecided people have been divided into two groups, *Indecisive* and *Hesitating*, as described above. Undecided people would often shift between the two groups, but the addition of the *Hesitating* class will allow the robot to deal with standing and moving people a different way. Figure 4.3 shows three trajectories of undecided people (B1, B2 and C). C was a person walking slowly while observing the robot and decided to interact without further delay, and labeled *Indecisive* all the way. B1 and B2 were people who walked slowly but stopped for a short while (B2) and long while (B1) respectively before interacting with the robot, and thus shifted between *Indecisive* and *Hesitating*.

Figure 4.4 shows a situation where all people are considered *Hesitating*. The lady in the middle tried for over a minute to get the robot’s attention, but eventually left without interacting. If a robot with knowledge of her presence would have greeted her in some way, it might have resulted in an interaction. The people whose trajectories are drawn in Figure 4.3 interacted anyway, but for the ones stopping (B1 and B2), an interaction might have come sooner. Since person C went up to the robot without directly approaching it, a robot acting aware might not have had any impact on how soon the interaction would have started, but an indication of awareness could have improved that person’s perception of the robot.

People approaching the robot without any hesitation were labeled *Interested*. They almost exclusively approached the robot from the front, probably because the robot when idling was facing that direction. Figure 4.3 shows the paths (A1 and A2) of two people who approached the robot directly, and thus labeled *Interested*.

People labeled *Not Interested* were mostly people moving towards the robot from the side, and thus could be mistaken for approaching the robot, but then diverting from that trajectory when within about 4 meters from the robot. They then walked in a semi-circle around the robot and exited the area. Figure 4.5 shows a vector field of the direction of motion of people with this label moving from the right to

Figure 4.5: The flow of people labeled *Not Interested*.

the left. Dark lines correspond to many of people passing those coordinates, and bright lines accordingly correspond to few people.

Chapter 5

System Description

This chapter covers the tools used for the experiment. First an overview of the robot and the sensors used for tracking humans, and then a detailed description of the method on which Robovie's behavior is based.

5.1 An Overview of the System

The system used for the experiment consists of four major parts (Figure 5.1). Humans are detected by a network of laser range finders (LRF) (1), and the gathered data is processed to generate a position and an ID for each human (2). These positions are then used by a classification software to generate direction of motion and speed for each human (3a). This information is used as a basis for selecting a target human¹, and to determine the robot's behavior, which is done by a controller software (3b). Finally, when a human and an action have been selected, instructions are sent to the robot (4), which acts accordingly.

Figure 5.1: The different components used in the experiment.

---

¹“Target human” refers to the person who the robot is currently focusing its attention on.

5.2 Robot Platform

In this research the Robovie-II robot (Fig 5.2) has been used. The robot itself consists of numerous of sensors and actuators. However, since most of the sensing and processing is made outside the robot, it has been used as an actuator rather than an independent robotic entity. The only onboard sensor that was used was the gyro for determining the robot's rotation.

Robovie-II is built on a differential drive wheel base with a set of sonar sensors placed around the platform, and bumper sensors attached on the periphery of the base. Robovie's body is built on that base, a metal skeleton with arms and head attached, and the body covered with touch sensors. The inside consists of two computers, one running the robot's controller software, processing sensor data and triggering actuators, and one handling speech recognition and voice synthesis among others.

5.2.1 Controller Mechanisms

The core of Robovie's behaviors lies in a control loop as described by Kanda et al. [8]. A behavioral unit, *situated module*, describes a sequence of actions the robot can perform. Behaviors consist of sequential and iterative executions of these units. When the robot is active, it is running a loop that is continuously executing the current situated module. A new module will start executing in one of two different ways. Either the end of the current module is reached and a new one is chosen, or the another module interrupts the one currently running.

When actions and sequences of actions are designed, the situated modules, as

Figure 5.2: Overview of Robovie-II.

5.2. ROBOT PLATFORM

well as rules for transitions between them, are constructed. Situated modules consists of instructions to move the robot's actuators, for instance to make it move an arm or to turn around, and to listen for sensor information. The transition rules regulate which modules can follow which, and which one of these modules that should be selected as the next one. Furthermore they regulate if a module is interruptible and what conditions should be met for one module to interrupt another.

5.2.2 Communication Abilities

Other computers can connect to Robovie via an Ethernet interface, and send commands via its server interface. In this way it is possible to send the robot instructions or external sensor data, and to read its internal state.

5.2.3 Designing Behaviors

Two types of behaviors were created to deal with the problems described in Section 2.3. The first one, the *Active* behavior, was designed to address both the mentioned problems, and one, the *Passive* behavior, was designed to only address the first problem. The following set of actions has been created to be used in different scenarios.

- **Idle** – When no one is around the robot, or no one seems interested, the robot will just casually sway its arms and tilt its head to indicate that it is currently operating. While in the idle behavior, the robot will also reverse to its initial position and posture if not already there.
- **Look** – The robot will just look at the target human and follow him or her with its head. When looking, it is not enough to just look in the direction of that person, but also to give the impression that the robot is looking into the person's eyes. With no information available on the target human's height, the robot assumes the person to be 160 cm tall, and tilts its head accordingly.
- **Turn** – The robot is instructed to first look and then turn towards the target human. The robot will continue to turn so that it constantly faces that person if he or she moves.
- **Greet** – The robot looks at, turns towards and approaches the target human while uttering a phrase like "Let's talk!" or "Hi, I'm Robovie!". It stops before it enters the person's personal space, and if already in that space, it only turns and speaks.

While the active behavior utilizes all four of the actions above, with the passive behavior the only actions available for the robot are *Idle* and *Look*. Hence when someone seems to be hesitating it has in this case no possibilities to further encourage that person.

In addition to these two behaviors, a *Random* behavior was designed. This uses all the actions mentioned above, but the robot is fed information of a fictive target human, and so it will appear to do actions in a random manner. To generate these fictive target humans, recorded data from test runs of the experiment was used and played to the robot instead of the real data. In this way it was ensured that the ratio of the different human behavior classes, as well as the usage frequency, was equivalent to the experiments where real data was used.

The purpose of the random behavior is to evaluate whether the robot appears to act naturally just because it makes gestures, or if the naturalness comes from the robot directing its actions towards people.

5.3 The Sensor Network

For the robot to be able to detect a person early, in order to react in a suitable manner, a large area around the robot must be surveyed at all times. Thus some of the robot's onboard sensors, like eye cameras, are not suited for this task. The other sensors, the omni vision camera and the sonar sensors, do not cover the desired area with enough accuracy, and so external sensors must be used.

For this research four laser range finders, SICK LMS-200, have been used. They were each set to scan an area of 180° and a maximum distance of 8 meters, with a resolution of 0.5° . They were placed along a line with a distance of 1.6-2.2 meters apart, which enabled the network to scan an area of 8 times 21.3 meters (see Figure 5.3). However, due to the fact that the LRFs scan areas in shapes of semi circles, the entire rectangular area is not covered. The sensors were mounted 80 cm above the ground. At this level obstacles like sofas were avoided while at the same time people's arm movements could be accurately tracked. When tracking people with LRF, the most common approach is to mount the sensors just above floor level to minimize the effects of occlusion. In this case the sensors had to be mounted roughly at waist level as a consequence of the algorithm used to track humans (see section 5.3.1). This increases the risk of occlusion due to the waist, and possibly arms, covering a larger part of the scanned area. However, since four sensors were used, the risk of a person being occluded from all sensors was reduced. Something that was not taken into consideration for the field trial was carts of different sorts that were frequently used at AEON. They were tall enough for the LRFs to detect and were often identified as one or two humans by the tracking software. Since carts were not present during the experiment, and visually easily identifiable during the field trial, they did not constitute any major problem.

5.3.1 The Human Tracker

The data from the LRFs is fetched and processed by a tracking software using particle filters [4]. This software is able to follow the movements of each tracked human, and estimates the x- and y-coordinates. It is also able to estimate the rotation of the body by matching a pre-defined shape to the cross section of a

5.3. THE SENSOR NETWORK

Figure 5.3: The LRF Network. The person to the right is not seen by sensors A and B, but seen by sensors C and D, and so fully trackable.

person given by the LRF network. The pre-defined shape consists of a circle with two smaller circles on each side, representing torso and arms, which is the reason for mounting the LRFs at waist level. Designed for offline processing, the software turned out not being able to perform shape matching with a sufficient frequency, which is why that step of the algorithm had to be dropped. Consequently the information generated is the position combined with an ID for each tracked human.

Although four sensors are being used, and so the risk for occlusion is reduced, there might still be areas that are not seen by any of the sensors. However, due to the nature of the particle filter algorithm, people passing this area might still be recovered.

A side effect of using the external LRF network to track humans is that the robot gets tracked as a human as well. During the experiment this was used to update its internal state with. As a consequence problems with wheel slippage, that would cause the robot to lose track of its actual location, were eliminated.

Chapter 6

Modeling Natural Encounters

The approach used for determining the target human and what action to choose, takes two things into consideration: people's *estimated interest* in the robot, and the current *environment state*. The first property determines who to choose as target human and, together with the second property, determines which action to apply.

People are first classified by only looking at their current speed and not considering the robot's state (Section 6.1.1). They are then classified from the robot's perspective, depending on their orientation and position in relation to the robot (Section 6.1.2 and 6.1.3). These classifications constitute the model of the world that the robot uses when selecting an action (Section 6.3).

6.1 Classification of Humans

For selecting a target human, merely using the x- and y-coordinates is not enough. Therefore, between the human tracker and controller software, there is a software with the responsibility of classifying humans into classes (the module 3a in Fig. 5.1). Each time step this software gets data for all the tracked humans in form of an ID together with a position from the human tracker. Here a time step varies very much depending on the number of humans currently being tracked, since the tracking software is very processor intensive and its performance depends linearly on the number of tracked humans. With the equipment used, the time between two time steps ranged from 0.05 s to 0.2 s.

The humans are identified by the ID assigned by the human tracker, and their positions are saved over time in the classification software. These positions are used to generate three properties for each human: speed and direction of motion in the global coordinate frame, and whether the human is standing or not. The third property is necessary due to the tracking software adding a small amount of noise each time step. This results in the tracked position of people standing still randomly fluctuating around their real position. One option would have been to just threshold the speed so that motions under a certain value would have been considered as standing. But the random changes in position often result in speeds

Table 6.1: Human speed classes and their boundaries

State	Boundary (m/s)
Standing	0
Slowly walking	$> 0, \leq 0.6$
Walking	$> 0.6, \leq 2.0$
Fast walking	< 2.0

that exceed values that otherwise would not have been considered as standing. Instead the variance of the past ten positions is calculated and thresholded. If the person is standing still, the variance should be low due to the random characteristic of the coordinates, whereas if the person is moving in one direction, the variance quickly rises. The number of past positions being included, as well as the value for the threshold was manually adjusted until a good performance was reached.

6.1.1 Classification of Speed

When these three properties have been determined, a rough classification depending on the speed and standing property is made. The controller software uses this classification and not the actual speeds in its algorithm. Table 6.1 gives the classes and their boundaries.

The two values used here to distinguish *Slowly walking* from *Walking*, and *Walking* from *Fast walking*, are based on people's behaviors during the real life experiment (see Section 4). People not paying any attention to the robot and just passing the area directly from one side to the other were labeled *Walking* or *Fast walking*. However people who slowed down, and in that way showed interest in the robot, presented a large range of speeds. The condition for a person to be labeled *Fast walking* was that he or she did not slow down in front of the robot. The boundary between *Walking* and *Fast walking* was set accordingly.

Most people who seemed interested in the robot, whether finally interacting with it or not, would slow down to observe it, and were labeled *Slowly walking*. (Around 15 % would approach the robot directly without slowing down before stepping onto the floor sensors.) The 0.6 m/s boundary was a good compromise between correctly classified people, and people seemingly not interested but walking slower than 0.6 m/s.

6.1.2 Classification of Direction

The challenge is, as mentioned in Section 2.2.2, to understand a person's intentions just from the position, speed and direction of motion. The speed used to classify people in the previous section, is measured globally. When the robot looks at people, it uses this classification along with the position and direction of motion to make another classification, which constitutes the robot's belief about the environment.

6.1. CLASSIFICATION OF HUMANS

Figure 6.1: The different directions of motions specified by which zone the robot is standing in.

The directional information is used to split people into three groups, which are defined as follows (see also Figure 6.1):

- **Approaching** – The robot is standing in a 45° cone centered in the person's direction of motion.
- **Passing** – The robot is standing in a 225° cone centered in the person's direction of motion, and outside the person is not approaching.
- **Leaving** – The robot is standing outside a 225° cone centered in the person's direction of motion.

When people, no matter what their intentions are, finally moved towards the robot, they walked fairly straight towards it. However, noise in the raw data as well as noise added by the human tracker is propagated to the calculation of the direction. Based on observations at AEON, the 45° limit resulted in a good balance between the number of correctly and incorrectly classified approaches.

People passing the robot reach a point where the robot is at a 90° angle from their trajectory of motion. Assuming they look straight ahead, they will no longer see the robot and are considered as leaving the robot. Due to noise, as discussed above, the limit was increased from 180° to 225° .

6.1.3 Classification of Position

The positional information is used to calculate the distance to the robot. Here people are grouped into the ones located in the public space, *Far* and the ones located in the social space or closer, *Close*. Figure 6.2 shows the concentration of people standing and walking slowly in the scanned area, taken over one day at AEON. According to observations made at the mall, people in these two groups were most likely to interact with the robot, and as the figures show, the highest concentrations are found within the robot's social space. This suggests that it is valid to make this classification of *Far* and *Close*. The reason for the higher concentrations to the right in both figures is the sofa placed there.

Figure 6.2: Concentration over time of standing 6.2a and slowly walking 6.2b people. Darker areas mean more visits. The robot's social space is indicated with a black, 6.2a, and white, 6.2b, semi circle.

When a person is standing, it is impossible to know the orientation of the body. The person might be interested and looking at the robot, or not interested at all, looking somewhere else. This fact makes the standing class more difficult to handle. For the experiment conducted for this thesis, it is assumed that a person is facing the robot if standing in its social space. At AEON this space did not contain any natural spots for people to stop at, such as notice boards, orientation maps or shop windows, which made the robot the most natural reason for stopping in this space.

6.2 Determining the Robot's Beliefs

As discussed in Section 4.2, people largely belong to one of three groups: no intention of interacting, a firm intention of interacting, and indecisive. When looking at the behavior of people before an interaction, some people were standing for at least a few seconds before approaching, thus appearing indecisive. This originates both from people being indecisive and from the robot only being able to interact with one person at a time, which makes people who want to act no matter what, wait for their turn. After the ongoing interaction was completed, a person with firm intention would however approach the robot directly. To estimate people's interest in the robots, the four classes of beliefs have been used corresponding to the ones found in Section 4.2. The classes are accordingly:

- **Interested** – People estimated to have a firm intention of interacting.
- **Indecisive** – People estimated to not have decided whether to interact or not.
- **Hesitating** – People estimated to not have decided whether to interact or not, and has been standing still for three seconds or more.
- **Not interested** – People estimated to have a firm intention of not interacting.

In total three classifications are made (based on speed, distance and direction of motion), and with these as a starting point the robot makes up its beliefs about

6.2. DETERMINING THE ROBOT’S BELIEFS

Table 6.2: Distance to the robot $\leq 3.5\text{ m}$

	Approaching	Passing	Leaving
Slow walking	Interested	Indecisive	Not interested
Walking	Interested	Indecisive	Not interested
Fast walking	Interested	Not interested	Not interested
Standing	Indecisive (<i>If standing > 3 s, Hesitating</i>)		

Table 6.3: Distance to the robot $> 3.5\text{ m}$

	Approaching	Passing	Leaving
Slow walking	Indecisive	Not interested	Not interested
Walking	Indecisive	Not interested	Not interested
Fast walking	Indecisive	Not interested	Not interested
Standing	Not interested	Not interested	Not interested

each person. Table 6.2 and 6.3 state the relations between classifications and beliefs and are based on observations made at AEON as explained below.

From each combination of classifications a correlation has been made, that best fits the observations from AEON. People leaving are all considered not interested since they apparently are moving away from the robot. People passing or standing in the public space are also considered not interested. As mentioned in Section 4.2, at AEON people in those states were excluded when determining the ratios of the different classifications. They must however be included in model for it to be complete. Some of these people would eventually interact with the robot, but before doing so, they would have to either enter the robot’s social space or approach the robot directly, which would alter the robot’s belief about them. Observations made clear that most people in those states would not interact, and as Michalowski et al. [9] conclude, it might feel intrusive if the robot tries to get your attention when you do not want it. Therefore these people are ignored until entering a state with another belief. People in the social space approaching the robot are all estimated as interested, while people approaching while in the public space might keep approaching or deviate (for instance people coming out of the stores across from the robot). Fast walking people not approaching the robot are all estimated as being not interested. As discussed above, no information is available about the direction of people standing and they are assumed to be facing the robot while in the social space, thereby being considered as indecisive. When standing still for three seconds or more, it is assumed that people are still interested, and will be encouraged by the robot addressing them. The belief changes accordingly. Fast walking people never stopped to interact, and they are therefore believed not to be interested. People in the two other walking states presented a large variety of behaviors, from just passing the robot without giving it so much attention, to

moving back and forth outside the floor sensors for several minutes, trying to get the robot to see them. They are considered indecisive and not hesitating while they are moving since it is difficult to know their real intention. Addressing them, as done when a person is considered to be hesitating, is both costly and might feel intrusive.

6.3 Selecting Target Human and Action

The environment is classified into one of four mutually exclusive states. As opposed to the estimation of beliefs that was based on observations, this classification is made with the cost of the robot's actions in mind, as discussed in Section 2.3.1. The classification below does not take people estimated as not interested into account.

- **Single** – There is zero or one human present.
- **People Present** – Two or more people are standing or passing, no one is approaching.
- **People Approaching** – Two or more people are approaching, no one is standing or passing.
- **People Present and Approaching** – At least one person is standing or passing and at least one person is approaching.

When applying an appropriate action, the robot takes both its belief about the target human and the current state of the environment into account. The *Turn* and *Greet* actions, which both contain turning of the robot's body, will be used with more care than the *Look* action because turning of the body is far more time consuming than just turning the head. Another reason for restrictive usage of the *Turn* and *Greet* actions is that if the robot turns towards one person, another person who is equally interested might feel left out and loose interest. This fact is also considered when two people are standing. When turning while two people are standing the robot will rotate so it faces a point in between them, and then look at both of them for a while each to make them both feel noticed. The rules that are applied can be found in Table 6.4.

Table 6.4: The rules for choosing an action for the *Active* behavior. The cells labeled *Not valid* are situations that can not occur.

	Single	Waiting	Approaching	Appr. and Wait.
Interested	Turn	<i>Not valid</i>	Look	Look
Indecisive	Look	Look	Look	Look
Hesitating	Greet	Greet	<i>Not valid</i>	Look
Not Interested	<i>Not valid</i>	<i>Not valid</i>	<i>Not valid</i>	<i>Not valid</i>

6.3. SELECTING TARGET HUMAN AND ACTION

The rules above apply to the *Active* behavior. For the *Passive* behavior, the *Idle* action is kept and the rest of the actions are replaced with the *Look* action.

An outline of the algorithm used to process the classified humans and to control Robovie can be found in Table 6.5. The loops are executed as fast as possible with the bottleneck being the human tracking. When a new set of readings are available from the human tracker, these are classified and the resulting classifications are read by the controller.

When a target human has been found, the inner loop is entered. An action is selected according to the rules in Table 6.4 and a command is sent to the robot to change actions. It is then continuously fed the current location of the target human. This pattern is broken if the target human leaves, if too long time has passed, the target human has approached (which is the ultimate goal) or another person becomes more interesting.

Table 6.5: An overview of the controller algorithm

Controller Algorithm
Loop Call <i>Update</i> Select the most promising human If No target human found Then Continue Loop Call <i>Update</i> Set robot behavior If The target human leaves Then Break If Current interaction time > MAX_T Then Break If Target human has approached Then Break If Another human is more interesting Then Switch target human
Function <i>Update</i> Get new readings about the humans and update the robot's beliefs Update the state of the environment

Chapter 7

Experiment

In this chapter a description of the controlled experiment is given, and the results from first- and third-person evaluations are analyzed.

7.1 Setup and Implementation

For the experiment four test subjects, one female and three male, were recruited from ATR. They were all interns working at another lab, so they all have a background studying technical subjects and are familiar with robots, but no one had interacted with Robovie before.

7.1.1 Setup

The experiment was held in a controlled environment at ATR. The placement of the LRF network and the robot was taken from the setup at AEON, though, due to the size of the room, a smaller area was covered (Figure 7.1). Two doors were held open to let the test subjects pass back and forth in order to simulate the environment at AEON.

The test subject were instructed to behave according to one of the four groups described in Section 4.2, and were shown footage from the AEON experiment to get a better idea of how people were behaving, and what distinguished the four groups.

7.1.2 Implementation

The experiment was divided into twelve sub experiments, each ten minutes long. Before these sub experiments took place, one test run was held in order to minimize novelty effects. The three robot behaviors described in Section 5.2.3 were run four times each in a randomly selected order.

Throughout every sub experiment the test subjects entered individually through one of the doors, acted as they were instructed, and exited. This pattern, referred to here as an *episode*, was repeated until the end of the experiment. Usually each

Figure 7.1: The setup for the experiment. The jagged lines are the background as taken from the LRFs.

person performed between ten and twenty episodes during one sub experiment. The instructions were given on notes that each test subject randomly picked up before entering the room. Each note was labeled with one of the four groups described in Section 4.2 and they occurred with the ratio described in that section. The test subjects were however not informed of how the robot was set to behave, and they did not get any advance information about the existence of different robot behaviors.

7.2 Evaluation

Two evaluations of the experiment have been done. One first-person evaluation consisting of the test subjects' own impressions, and one third-person evaluation consisting of the impressions from a total of 14 people having watched a compilation of video clips from the experiment. For the video fifteen clips were selected, each clip containing one scene of 20 to 40 seconds duration. Five clips with the same robot behavior were chosen for all three of the robot's available behaviors. Each clip has two corresponding clips containing similar behavior among the test subjects, but with different robot behavior. For both third-person observers evaluating the video and the participants a seven grade scale was used, with seven corresponding to natural and one corresponding to not natural. The following criteria were given on how to grade¹:

- In general, the naturalness of the robot's behavior should be determined with the following three questions as starting points.
  1. How aware does the robot seem to be of the surrounding people?
  2. How does the robot respond to people's intentions?
  3. How does the robot react to people's behaviors?

¹The last point are directed to the persons watching the video.

7.2. EVALUATION

- What the robot does when the person has approached the robot, i.e. stands right in front of the robot, shall NOT be regarded.
- Most of the time there are several people present. You shall give your impression of how the robot behaves towards them collectively².
- The overall impression should be considered, how does the robot behave throughout the scene.

During the experiment the test subjects were instructed to evaluate each episode, and an average of 63 grades were collected from each sub experiment. For the evaluation of the video clips, the observers were asked to give one grade for each clip.

7.2.1 Observations

Although performed as a controlled experiment, with test subjects knowing they participated for a purpose, both a genuine curiosity and some skepticism among them could be seen. “I felt like hugging it”, one of the test subjects said after having been addressed, while another of them felt like retreating when approached by the robot due to a feeling of uncertainty of what the robot was going to do next.

One interesting observation, more related to human behavior, is that the test subjects when standing seemed to adjust their positions in order to stand more face to face with the robot if they felt like they were being noticed. During the sub experiments with the robot’s random behavior, this happened several times. The robot turned towards an imaginary person standing close to the real person’s position, and the test subject adjusted his or her position accordingly. When looking from the perspective of the video camera, it is difficult to see whether the robot actually is facing the person or a point close to the person, which of course also affects the third-person evaluation.

Figure 7.2 shows a situation during the experiment. The test subject to the right in 7.2a (closest in 7.2b) has been standing still and is just about to start walking. He is standing in the robot’s public space, and thus classified as *Not Interested*. The person to the left in 7.2a is leaving the area and classified as *Not Interested*. The person in the middle in 7.2a is walking slowly in the robot’s social space. She is classified as *Indecisive* and the robot is currently targeting her by looking at her. The lines from the humans indicate direction of motion, and the line from the robot indicates the direction of its head.

7.2.2 Data analysis

The data from first- and third-person evaluation have been treated separately. On both data sets a repeated measures analysis of variance (ANOVA) have been con-

---

²“Collectively” in this case refers to how the robot behaves towards them as a number of individuals, and not as a group.

Figure 7.2: Estimated positions (left) and the corresponding scene (right).

ducted and the Bonferroni method has been applied to compare the performance of the different robot behaviors.

First-person evaluation

For each sub experiment between 47 and 75 grades were collected, and for the robot's behaviors, *Active*, *Passive* and *Random*, 270, 246 and 236 grades respectively.

The ANOVA showed that there was a significant difference between the behaviors ($F(2,747) = 82.340$, $p < .001$). Multiple comparisons with the Bonferroni test revealed that the *Active* behavior was preferred to the *Random* behavior ($p < .001$) and that the *Passive* behavior was preferred to the *Random* behavior ($p .001$). The comparison of the *Active* and *Passive* behaviors showed that there was no significance difference between them ($p = 1.00$).

Third-person evaluation

The analysis of the third-person evaluation gave similar results as the first-person evaluation. The ANOVA revealed a significant difference between the behaviors in this case as well ($F(2,194) = 27.470$, $p < .001$). Multiple comparisons with the Bonferroni test revealed that the results for *Active* and *Passive* behaviors are significantly better than the *Random* behavior ($p < .001$), and that there were no significant difference between the *Active* and *Passive* behavior ($p = .737$).

The grades for both first- and third-person evaluation can be seen in Figure 7.4, while means and standard deviations are displayed in Figure 7.3.

7.2.3 Discussion

According to both first- and third-person evaluation the robot was experienced to be the most unnatural when behaving in a random manner. Although expected,

7.2. EVALUATION

Figure 7.3: Results from first- and third-person evaluation.

this clearly shows that it is not enough for the robot to just move and make gestures for it to appear natural, and thus does not add to people’s trust in the robot, or feeling of comfort around it.

The analysis in Section 7.2.2 shows that the test subjects found the active and passive behavior to be equally natural. This could indicate that in most cases, there is no need to turn the entire body for people to realize that the robot has noticed them and is aware of their intentions. There was a difference in how the test subjects commented on their experience when classified as *Indecisive* or *Hesitating*. The common opinion was that the robot sometimes appeared confused and that it was not able to catch up with the people moving around it when turning its body. At the same time, when using the *Passive* behavior there was a feeling among the test subjects that the robot was just staring at them when they were standing for a period of time. This suggests that the robot’s *Turn* action is not necessary, and just looking should be enough for most cases, but to engage a hesitating person the *Greet* action is needed.

The mean of the active and passive behaviors were about the same, but with a larger standard deviation for the active behavior. This supports the claim above that an active behavior is needed, but actions involving turning of the robot should be used more restrictively. This is also supported by Michalowski et al. [9], who experienced missed opportunities of interactions when not actively initiating contact with hesitating people, as well as people not wanting to interact feeling uncomfortable when being addressed.

7.2.4 Sources of Error

A major source of error, both during the field trial as well as the experiment, was the tracking system losing track of people. At AEON this was a minor problem, since the observations mainly were made by looking at the actual experiment, and the data generated from the tracking, such as speed and direction of motion, only were depending on the person’s state, not ID. During the experiment the test subjects

Figure 7.4: Results from first- and third-person evaluation, divided by the robot's behavior.

could be lost when for instance making sudden movements or approaching the robot very fast. This would lead to the robot changing its attention to someone else or just ignoring the lost person if the tracking software was not able to find that person again. This in turn would lead the robot to behave in a non expected way and would make the robot seem less natural.

Chapter 8

Conclusions and Future Work

8.1 Conclusions

This thesis addresses the problem of a robot of detecting people at an early stage in order to know their intentions towards the robot and to act towards them in a natural way. In a field trial, held at a shopping mall, a robot was standing in an area with a continuous flow of people with the purpose of answering people's questions about the mall or to play simple games with them. It became apparent that there is a need for the robot to not only show that it is aware of the people around it, but also to encourage people, who seem reluctant to initiate an interaction, to approach. Some people would try to get the robot's attention for several minutes before losing interest.

During the field trial people's behaviors around the robot were observed and classified depending on their interest in the robot, in order to simulate that environment in a controlled experiment. These classifications were mapped with their speed, orientation and position in relation to the robot. Different actions were also created for the robot to naturally react to the observed human behaviors.

In an experiment different robot behaviors were evaluated with respect to how natural they appeared. Three robot behaviors were developed, based on the actions above. One behavior where the robot uses its entire body and actively tries to engage people, a passive behavior where the robot only show that it has noticed people by just looking at them, and a random behavior where the robot looked in random directions and made gestures in a random fashion. The robot's different behaviors were evaluated through a series of sub experiments. Test subjects were asked to act according to the classifications made during the field trial, and to evaluate how natural the robot appeared. The entire experiment was captured on video and a series of sequences were compiled for third-person evaluation.

The results from the evaluation show that it is possible to estimate people's intentions given only their position, orientation and speed, and to successfully make the robot behave naturally towards them. According to both the first-person and the third-person evaluation the active and passive behavior were considered equally

natural, and the random behavior was considered significantly less natural than the two. From these results it is apparent that a targeted behavior feels more natural than a random one. Although the active and passive behaviors are statistically equal in terms of naturalness, it was clear from comments made by both first- and third-person evaluators that their strong and weak points were different. The active behavior could be very encouraging for a hesitating person, but made the robot sometimes not being able to follow when turning and when several people were present. This situation was handled better by the passive behavior, but it was felt not to be enough when just looking at a standing person. This suggests that a combination of the two might have been more successful, to only use the robot's head for all situations apart from where a person is hesitating.

8.2 Future Work

There are some limitations to the method used for this thesis. Apart from improvements in areas not directly connected to this thesis, for instance human tracking, there are a few areas where extensions to the method could generate improvements in the robot's behavior. As mentioned in Section 2.2.1, people's gaze tells much about their interest in the robot. This could be taken advantage of when a target human has been found: to use the robot's eye cameras to determine whether the target human is currently looking at the robot or not. Another addition that might be useful is to look at the history of a person's movements. (Currently the only history that is taken into consideration is if a person has been standing for more than three seconds.) This would show if there is any difference in people's behaviors depending on how long have been around the robot and how they have been previously classified.

Bibliography

- [1] R. Atienza and A. Zelinsky, "Active Gaze Tracking for Human-Robot Interaction", *Fourth IEEE International Conference on Multimodal Interfaces (ICMI'02)*, pp. 261, 2002
- [2] C. Bartneck, T. Nomura, T. Kanda, T. Suzuki and K. Kennsuke, "A cross-cultural study on attitudes towards robots.", *Proceedings of the HCI International, Las Vegas*, Vol. 21, No. 1-2, pp. 217-230, 2005
- [3] M. Fujita, "AIBO: Toward the era of digital creature", *Int. J. Rbot. Res.*, Vol 20, No. 10, pp. 781-794, 2001
- [4] D. F. Glas, T. Miyashita, H. Ishiguro, and N. Hagita, "Laser Tracking of Human Body Motion Using Adaptive Shape Modeling", pp. 602-608, *IROS2007*, 2007
- [5] Rachel Gockley, Jodi Forlizzi and Reid Simmons, "Natural Person-Following Behavior for Social Robots" *HRI'07, March 8-11, 2007, Arlington, Virginia, USA*, pp. 17-24, 2007
- [6] K. Hayashi, D. Sakamoto, T. Kanda, M. Shiomi, S. Koizumi, H. Ishiguro, T. Ogasawara, N. Hagita, "Humanoid robots as a passive-social medium – a field experiment at a train station" *ACM 2nd Annual Conference on Human-Robot Interaction 2007*, pp. 137-144, 2007
- [7] E. T. Hall. "The hidden dimension: man's use of space in public and private", *New York: Doubleday*, 1966
- [8] T. Kanda, H. Ishiguro, M. Imai and T. Ono, "Development and Evaluation of Interactive Humanoid Robots", *Proceedings of the IEEE, vol. 92, No. 11, November 2004*, pp. 1839-1850, 2004
- [9] M. P. Michalowski, S. Sabanovic and R. Simmons, "A Spatial Model of Engagement for a Social Robot", *AMC'06, Istanbul, Turkey*, pp. 762-767, 2006
- [10] N. Mitsunaga, C. Smith, T. Kanda, H. Ishiguro, N. Hagita, "Robot Behavior Adaptation for Human-Robot Interaction based on Policy Gradient Reinforcement Learning", *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, Vol. 24, No. 7, pp. 820-829, 2005

BIBLIOGRAPHY

- [11] E. Pacchierotti, H. I. Christensen and P. Jansfelt, "Evaluation of Passing Distance for Social Robots", *The 15th International Symposium on Robot and Human Interactive Communication (RO-MAN06)*, Hatfield, UK, September 6-8, pp. 315-320, 2006
- [12] E. A. Sisbot, R. Alami, T. Simeon, K. Dautenhahn, M. Walters, S. Woods, K. L. Koay and C. Nehaniv, "Navigation in the Presence of Humans", *Proceedings of 2005 5th IEEE-RAS International Conference on Humanoid Robots*, pp. 181-188, 2005
- [13] M. Shiomi, T. Kanda, H. Ishiguro and N. Hagita, "Interactive Humanoid Robots for a Science Museum", *EEE Intelligent Systems*, Mar/Apr, 2007, Vol. 22, No. 2, pp. 25-32, 2007
- [14] M. Shiomi, K. Nohara, T. Kanda, H. Ishiguro, and N. Hagita, "Estimating Group States for Interactive Humanoid Robots", *EEE International Conference on Humanoids*, Dec. 2007., 2007
- [15] E. Sviestins, N. Mitsunaga, T. Kanda, H. Ishiguro and N. Hagita, "Speed Adaptation for a Robot Walking with a Human", *HRI'07, March 8-11, 2007, Arlington, Virginia, USA*, pp. 349-356, 2007
- [16] T. Tasaki, S. Matsumoto, H. Ohba, M. Toda, K. Komatani, T. Okata and H. G. Okuno, "Dynamic Communication of Humanoid Robot with multiple People Based on Interaction Distance", *ROMAN 2004*, pp. 71-76, 2004
- [17] F. Yamaoka, T. Kanda, H. Ishiguro and N. Hagita, "Modeling of Positional Interaction of explainer robot", *HRI'07, March 9-1, 2007, Washington DC, USA*, 2007
- [18] "Husquarna Automower", <http://www.automower.com/>, Latest visited: March 25, 2008
- [19] "Trilobite 2.0" <http://trilobite.electrolux.com/node217.asp>, Latest visited: March 25, 2008
- [20] "Wakamaru" <http://www.mhi.co.jp/kobe/wakamaru/english/>, Latest visited: March 25, 2008

Appendix A

Acronyms and Definitions

- Action – *An individual motion carried out by the robot*
- AEON – *A Japanese mall where a field trial was held*
- ANOVA – Analysis Of Variance
- ATR – Advanced Telecommunications Research Institute International
- DOF – Degrees Of Freedom
- HRI – Human-Robot Interaction
- LRF – Laser Range Finder
- Robot Behavior – *A set of actions*