Vrije Universiteit Brussel

FACULTY OF ENGINEERING
Department of Mechanical Engineering

Development of the Huggable Social Robot Probo. On the Conceptual Design and Software Architecture.

Thesis submitted in fulfillment of the requirements for the award of the degree of Doctor in Engineering (Doctor in de Ingenieurswetenschappen) by

Jelle Saldien

October 2009

Promotors: prof. dr. ir. Dirk Lefeber
prof. dr. ir. Bram Vanderborght

Abstract

This dissertation presents the development of a huggable social robot named Probo. Probo embodies a stuffed imaginary animal, providing a soft touch and a huggable appearance. Probo's purpose is to serve as a multidisciplinary research platform for human-robot interaction focused on children. In terms of a social robot, Probo is classified as a social interface supporting non-verbal communication. Probo's social skills are thereby limited to a reactive level. To close the gap with higher levels of interaction, an innovative system for shared control with a human operator is introduced. The software architecture defines a modular structure to incorporate all systems into a single control center. This control center is accompanied with a 3D virtual model of Probo, simulating all motions of the robot and providing a visual feedback to the operator. Additionally, the model allows us to advance on user-testing and evaluation of newly designed systems. The robot reacts on basic input stimuli that it perceives during interaction. The input stimuli, that can be referred to as low-level perceptions, are derived from vision analysis, audio analysis, touch analysis and object identification. The stimuli will influence the attention and homeostatic system, used to define the robot's point of attention, current emotional state and corresponding facial expression. The recognition of these facial expressions has been evaluated in various user-studies. To evaluate the collaboration of the software components, a social interactive game for children, Probogotchi, has been developed. To facilitate interaction with children, Probo has an identity and corresponding history. Safety is ensured through Probo's soft embodiment and intrinsic safe actuation systems. To convey the illusion of life in a robotic creature, tools for the creation and management of motion sequences are put into the hands of the operator. All motions generated from operator triggered systems are combined with the motions originating from the autonomous reactive systems. The resulting motion is subsequently smoothened and transmitted to the actuation systems. With future applications to come, Probo is an ideal platform to create a friendly companion for hospitalized children.

Acknowledgments

I remember my last year of my studies in Product Development. I was looking for the perfect subject to accomplish my master thesis, when my nephew Bavo attracted my attention with wild stories about artificial muscles and walking robots. Could it be true, that some creative inventors at the VUB were actually doing research on things that I only knew from science-fiction movies? I found out soon enough when I heard they needed a Product Developer to help them with a conceptual design for a new project. It turned out that the project was even more incredible than I expected: *a robot for hospitalized children*, the idea of the youthful dreamer Ivan Hermans. Supported by a handful of VUB researchers and my promoter Dirk Lefeber, I started my new adventure in the world of robotics. Meanwhile, two roboticists named Bram Vanderborght and Björn Verrelst truly believed in this project and put their efforts in finding proper funding to realize this new concept. When they were able to convince ministers Guy Vanhengel and Benoît Cerexhe, the project was launched. And before I knew it, I was a full member of the Robotics & Multibody Mechanics Research Group (R&MM).

The first person whom I would like to gratefully acknowledge is my promoter, professor Dirk Lefeber, who gave me the opportunity to continue on my first designs and to start a PhD under his guidance. I thank him for believing in me and for all the support he gave me throughout this entire project. I would like to thank Bram and Björn for initiating this project and for introducing me into the life of a robotic scientist. I would like to thank Bram specifically for his guidance during this last months with the writing of this dissertation. Our robot Probo presented here is the result of some very good team-work. Therefore I would like to thank especially Kristof Goris, my partner in this project. His technical insights and eye for perfection were indispensable for the successful realization of our project. Starting with his master thesis, Jeroen Vandenbrande joined the team a few years later. He provided us with all the support we needed to turn new, sometimes crazy ideas into the operational systems that reside in Probo today. I would like to thank Ivan, who offered his full support and always helped where he could. I also want to thank all my colleagues of the R&MM who were always ready to help and to share their opinions, during discussions, events, conferences and the daily life at the VUB. And in particular Kristof, Rino, Pieter and Pierre for sharing their space with me in our cozy office. I would like to thank Jenny for all the arrangements and special orders, she always knew to manage with a smile. The prototype of Probo could have never been built without the support of Jean-Paul, Frederick, Stijn and Thierry. Special thanks go to all the people of the HOA16 project, especially the professors Johan Vanderfaeillie, Ann Nowé and Eric Soetens for their advice and support in all the multidisciplinary aspects during the development of Probo. And my colleagues Selma and David for their efforts in the development of improved auditory and attentional systems. My friends and family have always been a big part of my life, and provided me with the necessary motivation to go for what I believe in. Their interests and discussions about my work have learned me to share different opinions on the introduction of robots for social interaction with humans. A special thanks for Wouter, who stood by me the first year I was dropped in Brussels.

Last, but certainly not least, my most profound thanks are to my parents, my sister and brother, for the unconditional support they have given me through the best and worst of my entire education and PhD research.

Thank you all!

Jelle Saldien

Jury members

President	prof. dr. ir. Jacques Tiberghien Vrije Universiteit Brussel
Vice-president	prof. dr. ir. Rik Pintelon Vrije Universiteit Brussel
Secretary	dr. ir. Michaël Van Damme Vrije Universiteit Brussel
Internal members	prof. dr. Ann Nowé Vrije Universiteit Brussel
	prof. dr. Johan Vanderfaeillie Vrije Universiteit Brussel
External members	dr. ir. Albert Van Breemen Market Development Manager at ASML
	prof. dr. Karsten Berns Technishen Universität Kaiserslautern
Promoters	prof. dr. ir. Dirk Lefeber Vrije Universiteit Brussel
	prof. dr. ir. Bram Vanderborght Vrije Universiteit Brussel

Contents

List of Figures

List of Tables

Chapter 1

Introduction

"Imagination is more important than knowledge. For knowledge is limited to all we now know and understand, while imagination embraces the entire world, and all there ever will be to know and understand."

-Albert Einstein-

1.1 Motivation

The last decades most of the robotics research and development was done in automation and industrial robots. What seems very similar with the development of expensive mainframes in the computer business 30 years ago. Bill Gates saw this as an opportunity and introduced the concept of a personal computer (PC). Which has led to a PC revolution and has changed the world as we know it today. Recently, Bill Gates [57] predicted in his article "A Robot in Every Home"; that it is most likely that in the near future more robots will be introduced to work in our houses and offices rather than the industrial robots working at the factories. He envisions a future in which robotic devices will become a nearly ubiquitous part of our day-to-day lives. But we must realize that the working environment of an industrial robot is completely different from a robot that acts in our domestic lifestyle. A factory is an organized environment with predefined tasks, while working in close collaboration with humans requires more advanced skills. If a new generation of robots is being created to live among humans and assist us in many different aspects of our daily lives. These robots will be better accepted if they measure up to a certain standard of social interaction and human-like communication, achieving the label of "social robot". No robot today can fulfill the role of a fully social interactive robot, but new research domains such as HRI (human-robot interaction) are gaining vastly more interest. In this context there is a strong need for robotic platforms that support HRI research. One specific subclass of HRI focusses on the use of social robots for therapeutic purposes. Based on the positive effects that have been found in AAT (Animal Assisted Therapy), the first robots are being tested for similar purposes (termed RAT - Robot Assisted Therapy). These social pet type robots are being used for therapy targetting children and elderly [121, 122]. With a special focus on hospitalized children, Probo can contribute to this new research domain. Although the investment in developing such a robot might not result in a direct commercial succes, the possibilty to become a usefull companion for children in need, has been an important motivation throughout this entire project.

1.2 Outline

Our motivation is further argumented throughout the thesis; presenting the domain in which Probo resides and its position towards similar social robots. The motivations are translated into the goals and specifications for this project, that are realized in the development of the social robot Probo. A full overview of the hardware design is presented by my colleague Goris in [62]. The remainder of this thesis focusses on the presentation of the software architecture and its containing components. The outline of the next chapters is as follows:

Chapter 2 describes the position and definitions of the social robot as a new type of service robot. It presents the pioneers that started with this new research domain and a state of the art of the social robots that exist today.

Chapter 3 presents the arguments and design choices that have been taken to set the specifications for the development of the social robot Probo.

Chapter 4 introduces the platform Probo, with a focus on the software architecture and its defining structures. The control center that provides en operator with a GUI to control is presented, including the 3D virtual model for visual feedback and simulation of the robot.

Chapter 5 offers a view on Probo's perceptual system, covering the human oriented visual, tactile and auditory stimuli, and the additional object identification. A preliminary attention system is presented, using visual stimuli and active vision, to generate gaze motions.

Chapter 6 presents the emotional system, describing how Probo is able to show facial expressions and how well they are recognized. A homeostatic regulation provides an autonomous system to react on perceived actions, generating an emotional state and corresponding expressions. These systems have been implemented in the 3D game named ProboGotchi.

Chapter 7 presents the motor system that includes the animation module and handles the motion requests of all systems and sends them towards the motors.

Chapter 8 summarizes the achievements and presents the future work and road map for Probo.

Chapter 2

The Dawn of Social Robots

"The engineer is the key figure in the material progress of the world. It is his engineering that makes a reality of the potential value of science by translating scientific knowledge into tools, resources, energy and labor to bring them into the service of man ... To make contributions of this kind the engineer requires the imagination to visualize the needs of society and to appreciate what is possible as well as the technological and broad social age understanding to bring his vision to reality."

-Sir Eric Ashby-

2.1 Introduction

The word "robot" was first introduced in a czech play named RUR (Rossum's Universal Robots) in 1920, and was derived from the czech word "robota" meaning "labor". The first robots were automated machines, used to facilitate the human labor in the factories. These industrial robots increased productivity, accuracy and endurance, and decreased the costs what leaded to cheaper mass-production. Nonetheless, robots were particularly employed to do dirty, dull, dangerous or inaccessible tasks, they started to replace the jobs of some of the workman to save on wages. With companies searching for the best cost saving measures, investment in automation and robots can be a necessary move. Perhaps this has lead to the negative and somewhat threatening image western people hold against robots. This image is being exploited in many science fiction stories, predicting some kind of robotic revolution. Increasing software development and AI techniques are definitely improving the intelligence and autonomy of the robots. Robots are shifting from industrial to commercial (service) applications, changing the operational environment. A part of these service robots are meant for domestic and personal use. Introducing these personal robots into our everyday human social lives, requires these robots to adapt to human forms of communication. Robots that are adapted for these task can be tagged as a "social robot". The concept of a social robot is presented in Section 2.3. An overview of the state of the art in social robots is given in Section 2.4. But first a recent overview and future prospections on the robotics market, with the focus on service robots, are presented in the next Section 2.2.

2.2 From Industrial to Service Robots

At the end of 2007 about 1 million industrial robots (Figure 2.1) and 5.5 million service robots were worldwide operating in factories, in dangerous or tedious environment, in hospitals, in private houses, in public buildings, underwater, underground, on fields, in the air, in the space, etc. Robots are appearing everywhere! Up to the end of 2011 more than 17 million service robots and 1.2 million industrial robots will populate the world, reports the IFR (International Federation of Robotics) Statistical Department in the new study "World Robotics 2008" [39].

The IFR defines two major categories of robots: industrial robots and service robots. Industrial robots have historically represented the vast majority of robotic development, with many deployed in the automotive industry. The IFR provides the following definitions:

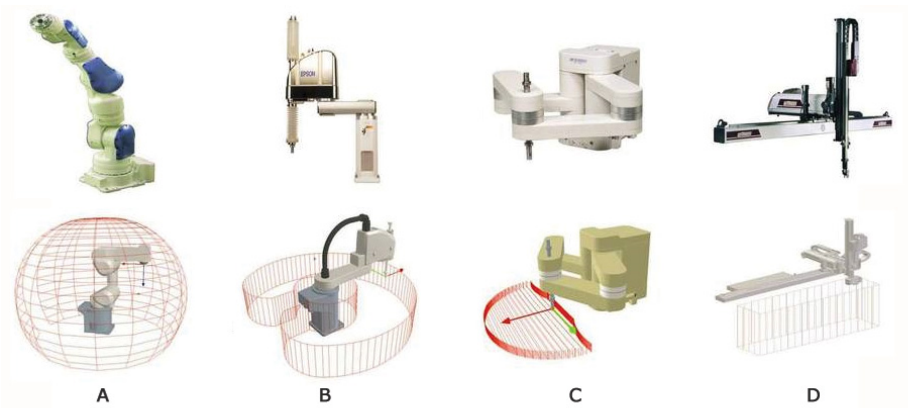

Figure 2.1: Examples of industrial robots; Articulated Robot (A), SCARA Robot (B,C) and Linear (Cartesian) Robot (D).

Figure 2.2: Examples of professional service robots for; milking, defense and exploration, underwater operations, demolition, medical procedures and industrial cleaning.

Industrial robot (as defined by ISO 8373)

An industrial robot is an automatically controlled, reprogrammable, multipurpose manipulator programmable in three or more axes, which may be either fixed in place or mobile for use in industrial automation appli-

cations.

Service robot (provisional definition)
A service robot is a robot which operates semi- or fully autonomously to perform services useful to the well-being of humans and equipment, excluding manufacturing operations.

Service robots are subdivided into professional service robots (Figure 2.2) and personal/domestic service robots (Figure 2.3). Professional service robots work in domains inaccessible to people, such as navigating abandoned mines and cleaning up nuclear waste. 49,000 units of robots for professional use were installed up to the end of 2007, as can be seen in Figure 2.4. The total value of service robots for professional use installed up to the end of 2007 was about 7.8 billion U.S. dollars. The most deployed robots are underwater systems, medical robots, milking robots and robots for defense, rescue and security applications. Turning to the projections for the period 2008-2011, the stock of new service robots for professional use is forecast to increase by some 54,000 units. An application area with strong growth might be humanoid robots mostly in the form of toys and hobby systems. However, up till now there have been no significant sales of humanoids as human companions for performing typical everyday tasks in production, office or home environments. That is probably due to the low capabilities and intelligence of these humanoids, that do not comply with most people's expectations. First shipments of these humanoid robots started in 2004, mostly in Japan or to international research laboratories and universities as research and development platforms. It is estimated that about 100 humanoid robots had been produced up to the end of 2007.

So far, service robots for personal and domestic use are mainly in the areas of domestic (household) robots, which include vacuum cleaning and lawn-mowing robots, and entertainment and leisure robots, including toy robots, hobby systems and education and training robots. About 3.4 million units for domestic use (3.3 million vacuum cleaners and more than 110.000 lawn mowers) and about 2.0 million units for entertainment and leisure sold up to end 2007, with a total value of 1.3 billion U.S. dollars. The success of these robots is due to their optimization to serve for a single task or purpose, unlike humanoids or other multi-purpose personal robots. The market for robots for handicap assistance is still small, but is expected to double in the next four years. These robots have not yet taken off to the anticipated degree, given their potential with regard to both imaginable need and the existing technological level of the equipment. A huge increase of personal robots, especially for entertainment and leisure, can be seen in Figure 2.5. It is projected that sales of all types of personal service robots will reach over 12.1 million units in the period 2008-2011, with an estimated value of 5.1 billion U.S. dollars. Many of these robots interact with people who have no special skills or training to operate a robot.

Figure 2.3: Examples of personal service robots; lawn mowing, floor cleaning, pool cleaning, companionship, hobby/education, surveillance and handicap assistance.

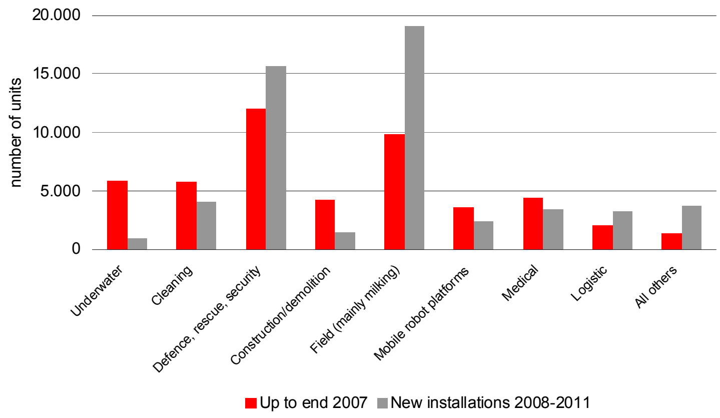

Figure 2.4: Service robots for professional use. Stock at the end of 2007 and projected installations in 2008-2011. (Source: World Robotics 2008)

The design and interaction of these robots will be critical, and raises a number of research and design challenges. A new research field that is meant to address these issues is termed "human-robot interaction".

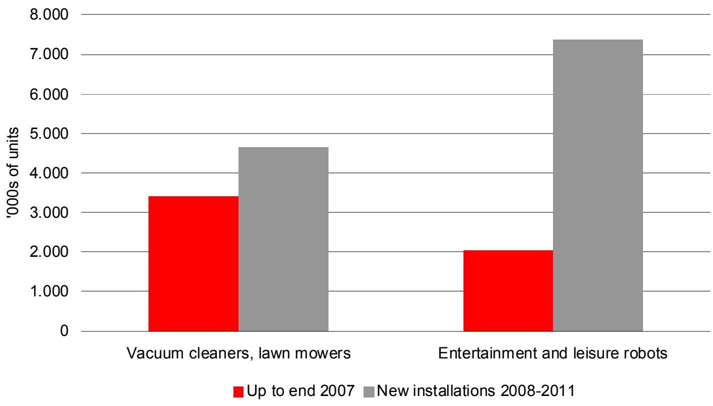

Figure 2.5: Service robots for personnel/domestic use. Stock at the end of 2007 and projected installations in 2008-2011. (Source: World Robotics 2008)

2.2.1 Human-Robot Interaction

HRI (human-robot interaction) is the study of interactions between humans and robots, it is a multidisciplinary field with contributions from human-computer interaction, artificial intelligence, robotics, natural language understanding, and social sciences. The basic goal of HRI is to define a general human model that could lead to principles and algorithms allowing more natural and effective interaction between humans and robots. Research ranges from how humans work with remote, tele-operated unmanned vehicles to peer-to-peer collaboration with anthropomorphic robots. Recent research in HCI (human-computer interaction) has highlighted the importance of robots as a new interface. Reeves and Nass [111] researched the role of computers as new interface media in the manner of TV and radio, and they proved that humans act toward computer interfaces (even a simple text-based interface) as if they were communicating with other humans. The future challenges of HRI are to comply to the human standards of communication. New multidisciplinary robot platforms, that can be used to communicate in the same way that humans do are necessary. Cassell et al. [27] showed the importance of anthropomorphic expressions, such as arms and heads on embodied agents, for effective communication with humans. This new type of robotic platforms are recently born under the name of "social robots".

2.3 Social Robots

Initially social robots were inspired by biology and based on techniques of artificial intelligence. They were used to study swarms or the behavior of insects [55]. However, in later approaches the interaction between humans and robots became more important. In contrast to most service robots, social robots are explicitly developed for the interaction of humans and robots to be like the interaction of human and human. The recent upcoming of this new type of robots lead to different definitions and classifications of social robots. The most relevant definitions and classifications that are in the line of this thesis are presented. Including a new field of study named robotherapy that focusses on the use of social robots for therapeutic purposes.

2.3.1 Definitions

In [8] Bartneck and Forlizzi propose the following definition of a social robot: *A social robot is an autonomous or semi-autonomous robot that interacts and communicates with humans by following the behavioral norms expected by the people with whom the robot is intended to interact.*
Communication and interaction with humans is a critical point in this definition. Therefore, robots that only interact and communicate with other robots would not be considered to be social robots. This definition also implies that a social robot has a physical embodiment. In recent work [63], Hegel et al. argue that the aesthetic form of a social robot communicates the social cues and signals, and that the behaviour of a social robot is mediated somehow through its physical form. But, the robot is not a social robot per se, it needs specific communicative capabilities to become a social robot. First, it implies the robot to behave (function) socially within a context and second, it implies the robot to have an appearance (form) that explicitly expresses to be social in respect to any user. From this point of view, a social robot consists of a robot and a social interface (see Figure 2.6). A social interface encloses all the designed features by which a user judges the robot as having social qualities. In [55],

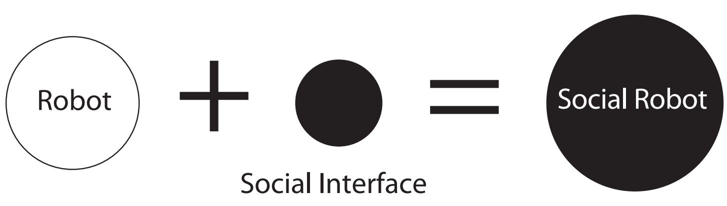

Figure 2.6: A social interface creates a social robot. (Source [63])

Fong et al. introduce the term *socially interactive robots* to describe robots for which social interaction plays a key role. Specifically considering robots that exhibit the following "human social" characteristics:

- express and/or perceive emotions
- communicate with high-level dialogue
- learn/recognize models of other agents
- establish/maintain social relationships
- use natural cues (gaze, gestures, etc.)
- exhibit distinctive personality and character
- may learn/develop social competencies

Social interactive robots can be used for a variety of purposes: as research platforms, as toys, as educational tools, or as therapeutic aids. The common, underlying assumption is that humans prefer to interact with machines in the same way that they interact with other people. Libin [82] defines the following features for an artificial creature to be an interesting and engaging communicating, gaming, educational, or even therapeutic partner for people of all ages, cultures, and life experiences.

- It imitates a real life (human- or animal-like) behavior.
- It models motor, emotional, and cognitive behaviors normally experienced by animals or humans.
- It communicates with a person on various levels: tactile-kinesthetic, sensory, emotional, cognitive, and social. These communications can be characterized using both verbal and nonverbal modes, and they can be evaluated as positive or negative.

2.3.2 Classification

The classification of social robots can be done on different levels; based on their abilities of interaction, based on their embodiment or based on their purposes. In [20], Breazeal defines four classes of social robots in terms of (1) how well the robot can support the social model that is ascribed to it and (2) the complexity of the interaction scenario that can be supported:

Socially evocative. Robots that rely on the human tendency to anthropomorphize and capitalize on feelings evoked when humans nurture, care, or involved with their "creation".

Social interface. Robots that provide a "natural" interface by employing human-like social cues and communication modalities. Social behavior is only modeled at the interface, which usually results in shallow models of social cognition.

Socially receptive. Robots that are socially passive but that can benefit from interaction (e.g. learning skills by imitation). Deeper models of human social competencies are required than with social interface robots.

Sociable. Robots that pro-actively engage with humans in order to satisfy internal social aims (drives, emotions, etc). These robots require deep models of social cognition.

Complementary to this list the following three classes have been added by Fong et al [55]:

Socially situated. Robots that are surrounded by a social environment that they perceive and react to [36]. Socially situated robots must be able to make the distinction between other social agents and various objects in the environment.

Socially embedded. Robots that are: (a) situated in a social environment and interact with other agents and humans; (b) structurally coupled with their social environment; and (c) at least partially aware of human interactional structures (e.g., turntaking) [36].

Socially intelligent. Robots that show aspects of human style social intelligence, based on deep models of human cognition and social competence [30][31].

Robots can also be classified on embodiment or the robot's morphology. In [143], Yanco and Drury present three values:

anthropomorphic Robots that have a human-like appearance.

zoomorphic Robots that have an animal-like appearance.

functional Robots that have an appearance that is neither human-like nor animal-like, but is related to the robot's function.

This division follows the morphology suggested by Fong et al [54], except that the category of "caricatured" (robots having simplified or stereo-typical representations) is collapsed into anthropomorphic or zoomorphic, depending upon whether the caricature most closely resembles a person or an animal. In [82] artificial creatures are classified from the robopsychologist's point of view. Where they are divided into two major groups: assisting robots, which are oriented

Table 2.1: Classification of interactive stimulation robots with regard to human needs and benefits (modified from [82]).

ROBOT	HUMAN	
Type	Need	Benefit
Social robots Recreational robots	To provide company To entertain	Communication Positive stimulation through entertainment
Educational robots	To stimulate engagement in educational process	Enrichment of learning skills
Rehabilitation robots	To recover from injury or to compensate for existing disability	Medical treatment and aids trough rehabilitation
Therapeutic robots	To alleviate negative mental states and psychological disfunction	Therapy of negative states and behaviors

toward industrial, military, research, medical, and service activities, and interactive stimulation robots, which are designed for social, educational, rehabilitation, therapeutic, and entertainment purposes. The classification of interactive stimulation robots with regard to human needs and benefits can be seen in Table 2.1. Robotic creatures of the interactive stimulation class have become the primary subjects of a new field of study named robotic psychology and robotherapy [83]. Robotic psychology focuses on the compatibility between humans and robots, while robotherapy concentrates on using interactive robots as therapeutic companions for people with psychological problems or limited physical, cognitive, or emotional resources.

2.3.3 Robotherapy

Social robots are slowly but steadily being investigated as remedial and therapeutic tools. A key ingredient that makes a therapeutic robot social is interactivity. Robotic wheelchairs or robots that are used for sensorimotor exercises often just look like "machines", more or less autonomous, but provide a particular functionality similar to a bicycle or a bus, in that they are clearly tools rather than interaction partners. However, in scenarios where service robots and disabled people cooperate, where the strength and weaknesses of both parties are exploited towards forming a "relationship", robots will require certain social skills [140]. Using synergetic effects emerging from robot-human cooperative activity is one advantage of using social robots. In [54], Fong et al. list the following advantages that social robots can offer for therapeutic use:

- Robots can provide a stimulating and motivating influence that make living conditions or particular treatments more pleasant and endurable, an effect that has particular potential for children or elderly people.
- By acknowledging and respecting the nature of the human patient as a social being, the social robot represents a humane technological contribution.
- In many areas of therapy, teaching social interaction skills is in itself a therapeutically central objective, an effect that is important in behavioral therapeutic programs, e.g. for autistic children, but that might potentially be used across a range of psychological, developmental or social behavioral disorders.

To date, social robots have been studied in a variety of therapeutic application domains, ranging from using robots as exercise partners [61], using robots in pediatrics [107], robots as pets for children and elderly people [121, 122, 136, 137], and robots in autism therapy [139, 35, 34]. The use of robots as pets, also opens up new opportunities in adapting therapies that are now based on interaction with animals. Animal assisted therapy (AAT) and animal assisted activities (AAA) are becoming commonly used in hospitals, especially in the United States [23]. AAT and AAA are expected to have useful psychological, physiological and social effects. Some psychological studies have already shown that animals can be used to reduce heart and respiratory rate [1], lower levels of stress [7], progress mood elevation and social facilitation. The problem with animals is that they are difficult to control, they always have a certain unpredictability, and they are carriers of disease and allergies. Therefore, the use of robots (instead of animals) has more advantages and a better chance of being accepted in hospitals. Using these social pet robots for therapeutic purposes is termed robot-assisted therapy (RAT). Autism therapy is a different, promising application domain of interactive social robots. People with autistic spectrum disorders have impairments in social interaction, communication and imagination. In the Aurora project (AUtonomous RObotic platform as a Remedial tool for children with Autism), Dautenhahn and her colleagues have been studying how autonomous interactive robots can be used by children with autism [32]. The state of the art in social robots for therapeutic use is described in section 2.4.3.4

2.4 State of the Art in Social Robots

This section presents the most relevant current and historical projects regarding the new domain of social robotics. Most of these projects were also an inspiration in various ways to help with the development of Probo. The projects are grouped in different sections. The first section covers the most successful artificial companions that have been developed in different forms, serving as a toy for children. Section 2.4.2 describes the more intelligent companions that appeared on the market embodied as robot pets. Despite the potential this was not a great success yet. Only the cheap robotic toys that are mentioned in 2.4.1 were able to have commercial success. Nonetheless are these robotic toys contributing to the future market for social robots. Results from research will be gradually implemented in these toys to enhance the interactions with the user. Most of the social robots now are used for HRI studies. The face is the most important element to express social cues. Section 2.4.3.1 describes robotic projects that use a robotic face to study HRI. Robots that also include gestures by moving the whole upper body, including arms and hands, are described in section 2.4.3.2. Section 2.4.3.3 describes, the humanoids that use their full body to interact with the humans and the environment. The final section 2.4.3.4 covers the robot projects that especially focus on RAT.

In Europe several large multi-disciplinary, and multi-partner, projects have recently been funded to further explore the potential of robotic artificial companions - most notably these include the COMPANIONS project [11] and LIREC (Living with Robots and Interactive Companions) - both of which focus on aspects of combining advanced technologies to create personal, persistent 'agents' or 'companions', which stay with the user for long periods of time, developing a relationship and getting to know its owner's preferences and wishes.

2.4.1 Artificial Toy Companions

The first steps in designing social robots were made with the idea to create artificial companions (or virtual pets), as depicted in Figure 2.7. One of the first attempts was the highly successful Tamagotchi from Bandai, which is a virtual pet that needs to be taken care for. It looks like an egg-shaped electronic device with an LCD display and buttons to enable interaction. The next big step was the development of an artificial pet, called Furby (from Tiger Electronics), without any kind of graphical interface or buttons. It interacted with the environment through sight, touch, hearing, and physical orientation. The extraordinary thing about the Tamagochi/Furby phenomenon was that people who knew better began to feel guilt about their behavior towards a small cheap and simple toy that lacks any level of intelligence. The brief history of those toys says a great deal about people and their ability to create and transfer their affections, despite their knowledge of what is really going on inside an object [141].

Recently more and more commercial virtual (screen-based) pets such as Catz, Dogz, MoPets and Nintendogs are being brought into the market. Millions of consumers worldwide have purchased these products, played with them, in-

Figure 2.7: Examples of artificial toy companions: Tamagotchi, Nintendogs, Furby, Catz, Robosapien, WowWee Alive Lion Cub and FurReal Biscuit Pup.

teracted with them, invested time in looking after them, and perhaps even become emotionally attached to them. Despite this huge financial and emotional investment by consumers, and an ongoing development and marketing investment by industry (new titles are appearing almost daily), academic interest in such products is virtually nil [80].

In 2004 WowWee released the Robosapien, a toy-like biomorphic robot, and reported that over 1.5 million units were sold in the first 5 months of sale. Now WowWee has a whole range of robotic toys including the WowWee Alive line that is competing with Hasbro's FurReal line in the growing robotic animal toy market. The WowWee Alive animals feature animated facial and vocal expressions triggered by users' touch and comes with adoption papers and a family portrait. These toy robots are a big commercial success, but are very limited and can be considered as the precursors for the more intelligent social robots.

2.4.2 Robotic Companions

Here we consider robotic companions (depicted in Figure 2.8) as more intelligent and perceptive towards the environment than the previously mentioned

pet toys. They will express a certain behavior based on biological/psychological inspired models. They perceive the environment using different kinds of sensors. What is a difference from the toy companions that will play a certain motion sequence based on the button or sensor that is activated. Despite their advanced technologies, these robotic companions are mainly sold for research purposes and are not commercially successful. Other robotic companions that are more focused on therapeutic applications are covered in Section 2.4.3.4
NeCoRo [82], manufactured by the Omron Corporation, Kyoto, Japan, was

Figure 2.8: Examples of robotic companions: Necoro, Aibo and Pleo.

designed as a first of its kind real-life-looking emotional creature. Fifteen actuators inside the robotic creature's body make its behavior believable by providing adequate responses to human voice, movements, and touch. Multiple built-in sensors together with an artificial intelligence produce NeCoRo's self-organizing behavior. The real-life-looking robotic cat creates a playful, natural communication with humans by mimicking a real cat's reactions. NeCoRo can stretch its body and paws, move its tail, open and close its eyes, and meow, hiss, or purr when it is touched or if someone speaks to it.

Sony's robotic dog AIBO [56], which in Japanese means "companion", is perhaps the most well known of all embodied virtual pets (or robotic companions). In terms of size and shape, AIBO was made to look like a puppy. It would act like a puppy by exploring its environment (somewhat slower than a real dog), wanting to play, getting tired, angry and excited, needing sustenance (a battery recharge), and could be trained by owners to do tricks and develop its own personality. Unfortunately, in 2006 Sony announced a termination of all commercial and research activity in robotics[1] leaving a gap in the market for home entertainment robots which demonstrate sophisticated social behaviors and cognitive abilities.

[1]Following the Sony Corporation FY05 3Q announcement; the production of AIBO Entertainment Robots has been discontinued as of end March 2006.

An upcoming robot toy that might replace AIBO as a companion in the near future is Ugobe's Pleo. Pleo is an animatronic dinosaur toy designed to emulate the appearance and (imagined) behavior of a week-old baby Camarasaurus. It was designed by Caleb Chung, the co-creator of the Furby. Chung selected this species of dinosaur because its body shape, stocky head, and relatively large cranium made it ideal for concealing the sensors and motors needed for lifelike animation. According to Ugobe, each Pleo will "learn" from its experiences and environment through a sophisticated artificial intelligence and develop an individual personality. Pleo is equipped with a camera-based vision system, two microwaves for binaural hearing, eight touch sensors (head, chin, shoulders, back, feet), foot switches (surface detection), infrared mouth sensor for object detection into mouth, orientation tilt sensor for body position and force-feedback sensors in the joints. Unfortunately Ugobe filed for bankruptcy in April 2009, the company probably came along at the wrong time. Why did Pleo-maker Ugobe fail? There are many reasons, but former CTO John Sosoka thinks one of them is that the one company tried to do too much, working on both hardware and software. Jetta Company Ltd. (the original manufacturer of Pleo) acquired the assets of Ugobe and has recently re-launched Pleo.

2.4.3 Social Robots for HRI Studies

2.4.3.1 Robotic Faces

A lot of social robots are designed with the purpose to study the HRI and focus on facial expressions, as depicted in Figure 2.9. Other similar projects include Doldori [81] and eddie [126].

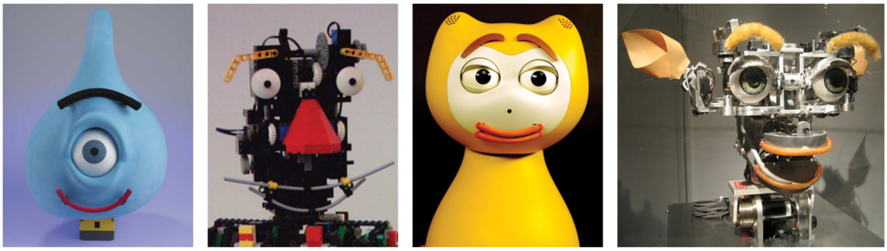

Figure 2.9: Examples of robotic faces: eMuu, Feelix, iCat and Kismet.

eMuu is a robot developed by Bartneck to function as an interface between the user and intelligent home. The user can instruct the robot to perform a number of domestic tasks [9]. eMuu is able to express three emotions; happiness, sadness and anger with one lip and one eyebrow. eMuu's lip and eyebrow are flexible and able to bend for displaying appropriate emotions. The emotion engine, which controls the emotional state and the facial expression, is based on the OCC (Ortony, Core and Collins) model [103].

Feelix, a robot built by Canamero and Fredslund [25] is constructed from LEGO Mindstorms. It is a cheap and easy method to build robotic faces. However, the robot might be perceived only as a toy due its appearance and the prior experience of playing with LEGO that many people have. Feelix has been used as a research platform for human robotic interactions study. Its face has 4-DOF (two eyebrows, two lips), designed to display six facial expressions (anger, sadness, fear, happiness, surprise, neutral) plus a number of blends. It can only sense tactile stimulation.

The iCat [132] is a research platform from Philips for studying human-robot interaction with a socially intelligent robot. The iCat has four touch sensors, two microphones, a CCD camera, and 13 servos for facial expressions (moving lips, eyebrows, eyes, eyelids and headposition). To make its movements believable the iCat makes use of the Open Platform For Personal Robots (OPPR). It is a so called "Animabotics" software platform. Animabotics is suggested by van Breemen as a new discipline in between robotics and animatronics. Animabotics should support designers in creating autonomous and intelligent robotic user interfaces, while providing tools to make these robots as enjoyable as the ones found in animatronics industry.

Kismet is a pioneer example of a social robot and is an inspiration for most of the new social robots that are being developed. Kismet was developed by Breazeal [20] as a test bed for learning social interactions involving an infant (the robot) and her caretaker (a human). This robot is a head with active stereovision and configurable expressive features (controllable eyebrows, ears, eyeballs, eyelids, a mouth with two lips, and a neck that can pan and tilt) with 18 DOF. All these features, together with an expressive vocalization system, allow Kismet to display a wide variety of emotional expressions that can be mapped onto a three-dimensional space with dimensions arousal, valence, and stance. Kismet perceives a variety of natural social cues from visual and auditory channels, and delivers social signals to people through gaze direction, facial expression, body posture, and vocalizations. The system architecture (Figure 2.10) consists of six subsystems: the low-level feature extraction system, the high-level perception system, the attention system, the motivation system, the behavior system, and the motor system. The low-level feature

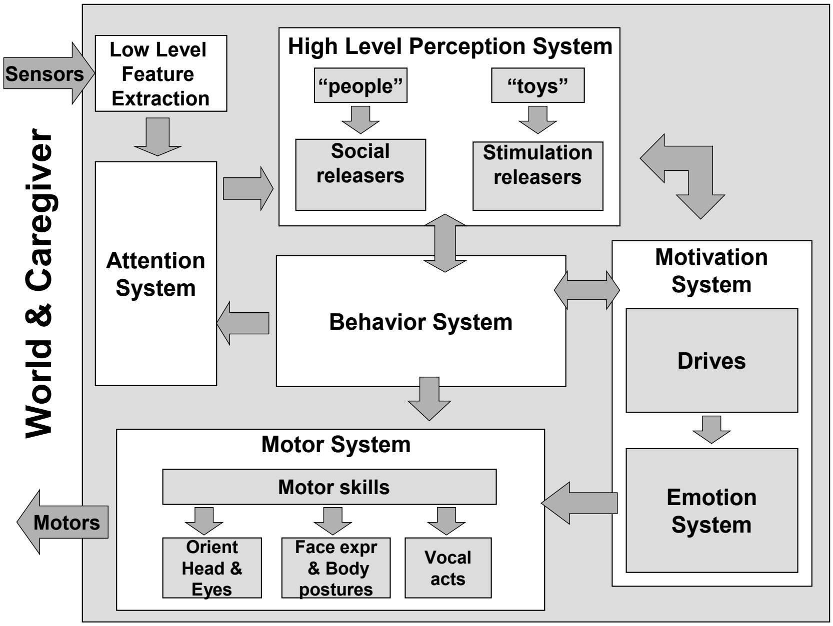

Figure 2.10: The system architecture of the robot Kismet. Six sub-systems interact to enable the robot to behave coherently and effectively.

extraction system extracts sensor-based features from the world, and the high-level perceptual system encapsulates these features into percepts that can influence behavior, motivation, and motor processes. The attention system determines what the most salient and relevant stimulus of the environment is at any time so that the robot can organize its behavior about it. The motivation system regulates and maintains the robot's state of "well being" in the form of homeostatic regulation processes and emotive responses. The behavior system implements and arbitrates between competing behaviors. The winning behavior defines the current task (i.e., the goal) of the robot. The robot has many behaviors in its repertoire, and several motivations to satiate, so its goals vary over time. The motor system carries out these goals by orchestrating the output modalities (actuator or vocal) to achieve them. For Kismet, these actions are realized as motor skills that accomplish the task physically, or expressive motor acts that accomplish the task via social signals. A caregiver exploits the learning mechanics of Kismet's motivation system, to place Kismet in either a positive affective state (positive reinforcement) when the robot does something desirable, or a negative affective state (negative reinforcement) when the robot does something undesirable. By doing so, Kismet's affective states mirror those of the caregiver [19]. These learning mechanisms bias the robot to learn and pursue behaviors that please the caregiver and to avoid those that displease her. By communicating reward and punishment information through emotive channels, the caregiver can actively help Kismet identify and pursue new behaviors as they play together, by assigning them values of goodness (i.e., pleases caregiver) and badness (i.e., displeases caregiver). As such, these mechanisms could serve as a simple form of empathy and as a starting point for teaching Kismet the value of its actions to others.

2.4.3.2 Robotic Upper Torso

Social robots can enhance their communication by the use of gestures. The robots depicted in Figure 2.11 have an actuated upper torso to transfer their expressions. Other similar projects are COG [21], wakamaru [124] and armarIII [3].

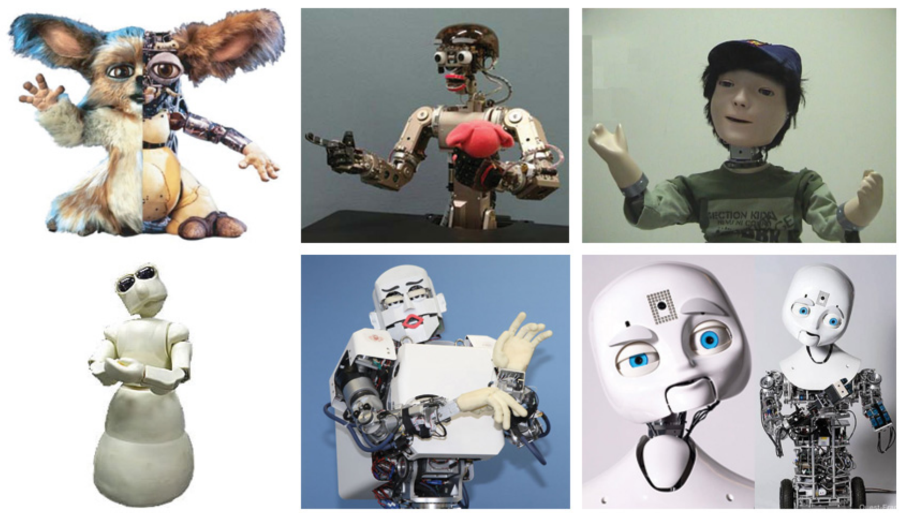

Figure 2.11: Examples of robotic with actuated upper torso: Leonardo, Infanoid, Kaspar, Robovie-IV, WE-4RII and Nexi (MDS).

Leonardo [18] is one of the most expressive robots in the world today. Leonardo has 69 degrees of freedom and 32 of those are in the face alone. As a result, Leonardo is capable of near-human facial expression (constrained by its creature-like appearance). Although highly articulated, Leonardo is not designed to walk. Instead, its degrees of freedom were selected for their expressive and communicative functions. It can gesture and is able to manipulate objects in simple ways. A real-time face recognition system for Leonardo can be trained on the fly via a simple social interaction with the robot. The interaction allows people, using a speech recognition system, to introduce themselves and others to Leonardo, who tries to memorize their faces for use in subsequent interactions. A sense of touch for Leonardo is being developed and will be useful for detecting contact with objects, sensing unexpected collisions, as well as knowing when it is touching its own body. Other important tactile attributes relate to affective content (whether it is pleasure from a hug, a ticking gesture, or pain). As a result, interacting with Leonardo is more like interacting with a creature than like interacting with a robot that is specialized for one skill. Leonardo is used to explore multiple forms of social learning and to investigate how people can teach robots. These forms of social learning include; learning by Tutelage, Guided Exploration, Spatial Scaffolding, Imitation, Perspective Taking and Social Referencing [18, 16].

Infanoid [75], which is 480 mm tall, the approximate size of a four year-old human child. With 29 actuators (mostly DC motors with digital encoders and torque-sensing devices) and a number of sensors, it can express attention (by directing its eyes, face, and hands), facial expressions (by moving its eyebrows and lips), and other hand and body gestures. Infanoid was able (1) to detect a human face and direct its gaze to the face (eye contact), and (2) to estimate the direction of attention from the face and search the attention line for any salient object (joint attention).

Kaspar [35] is a child-like humanoid robot which was designed and built by the members of the Adaptive Systems Research Group at the University of Hertfordshire to study human-robot interactions with a minimal set of expressive robot features. Kaspar has 8 degrees of freedom in the head and neck and 6 in the arms and hands. The face is a silicon-rubber mask, which is supported on an aluminum frame. It has 2 DOF eyes fitted with video cameras, and a mouth capable of opening and smiling. The robot can be used in two modes: remotely controlled as well as in autonomous operation. KASPAR is being used to study human-robot interaction as part of the European RobotCub Project and has been used in the past in the Aurora project.

Robovie [93] is a humanoid robot developed by the Intelligent Robotics and Communication Labs, ATR. They have developed communication robots, named Robovie, for the study of communication between individual humans as well as between humans and robots. The new Robovie-IV is designed as a robot with the same hight of a child. The body of the robot is covered with soft, light-colored skin for a soft look and touch. Many tactile sensors are embedded in the soft skin, to create a sense of touch. Robovie-IV can move in two modes, one is a normal wheeled robot mode with passive casters, and the other is the inverted pendulum mode. Using optical and RF (Radio Frequency) tag readers and laser range sensors the robot can identify different humans.

WE-4RII [94] (Waseda Eye No.4 Refined II) is one of the Emotion Expression Humanoid Robots that have been created since 1995 in order to develop new mechanisms and functions for a humanoid robot having the ability to communicate naturally with a human by expressing human-like emotion. WE-4RII can express its emotion using the upper-half body motion including the facial expression, arms, hands, waist and neck motion. The robot supports visual, auditory, cutaneous and olfactory sensations. WE-4RII has two color CCD cameras in its eyes. After calculating the gravity and area of the targets, WE-4RII can follow them with the eye, the neck and the waist. Regarding auditory sensor, WE-4RII has condenser microphones in each ear. It can localize the sound direction from its loudness and the phase difference in a 3D space. WE-4RII has tactile and temperature sensors to simulate a human cutaneous sensation. The robot can recognize when it is being "pushed", "stroked" or "hit", by using 2 layers of FSR (Force Sense Resistor). For WE-4RII can quickly distinguish the smells of alcohol, ammonia and cigarette smoke, using four semiconductor gas sensors. WE-4RII changes its mental state according to the external and internal stimuli, and expresses its emotion using facial expressions, facial color and body movement. The information flow into the robot is shown in Figure 2.12 [95]. There are two big flows. The one is the flow caused from the external environment, the other is the flow caused from the robot's internal state. The Robot Personality consists of the Sensing Personality and the Expression Personality. The former determines how a stimulus works the mental state, the later determines how the robot expresses its emotion. The need and the emotion are a two-layered structure. The need and emotion affect each other through the Sensing Personality.

The latest and one of the most advanced social robots from MIT is the Nexi, referred to as a Mobile/Dexterous/Social robot or "MDS". Nexi is a platform to support research and education goals in human-robot interaction, teaming, and social learning. In particular, the balancing 2 wheeled base provides a small footprint of the robot (roughly the size of a 3 year old child). This allows multiple robots to operate safely within a typical laboratory floor space. The 15 DoF face has several facial features to support a diverse range of facial expressions including gaze, eyebrows, eyelids and an articulate mandible for expressive posturing. Perceptual inputs include a color CCD camera in each eye, an indoor Active 3D IR camera in the head, four microphones to support sound localization, a wearable microphone for speech. Nexi is now commercially available as the product MDS from Xitome Design (an MIT spin-off company).

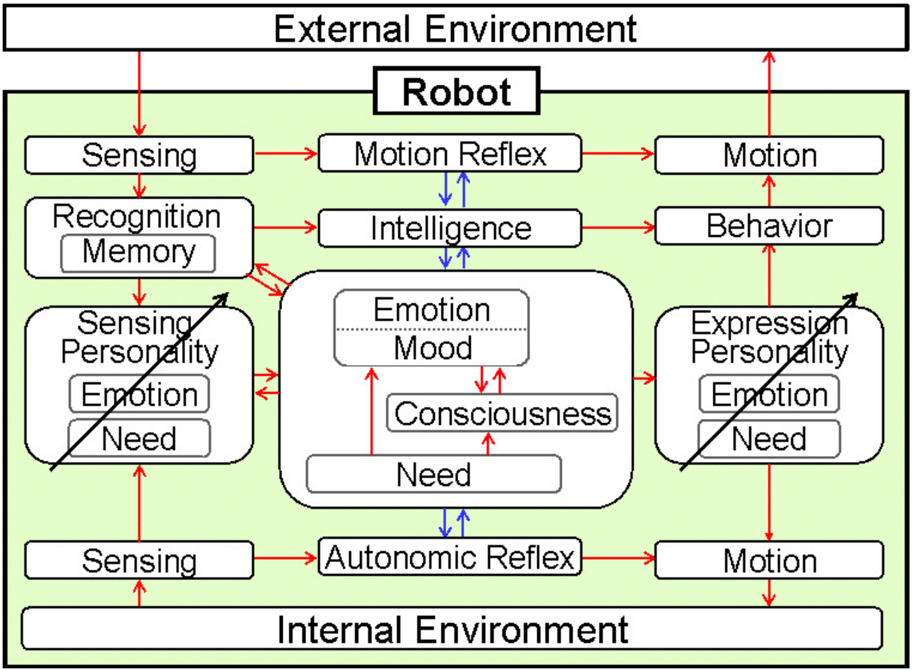

Figure 2.12: Information Flow of the Mental Modeling of WE-4RII.

2.4.3.3 Full Body Humanoids

Provided with a pair of legs, the humanoids are able to show full human-body expressions as shown in Figure 2.13. Notice that they lack the ability to show facial expressions. Other projects, such as Albert HUBO [101] and HRP [71], placed an android head on the humanoid's body. Some others have made androids to completely resemble a human such as Repliee Q2 [88].

Probably the most advanced humanoid robot in the world is ASIMO [114, 66] from Honda. ASIMO stands for Advanced Step in Innovative M0bility. Lightweight materials, like a magnesium alloy structure, combined with powerful computers and 34 servo motors throughout its body help ASIMO move smoothly with ease. Standing at 130 centimeters (4 feet 3 inches) and weighing 54 kilograms (114 pounds), the robot resembles a small astronaut wearing a backpack and can walk or run on two feet at speeds up to 6 km/h. Using the visual information captured by the camera mounted in its head, ASIMO can detect the movements of multiple objects, assessing distance and direction. ASIMO can also interpret the positioning and movement of a hand, recognizing postures and gestures. Because of this ASIMO can react to and be directed by

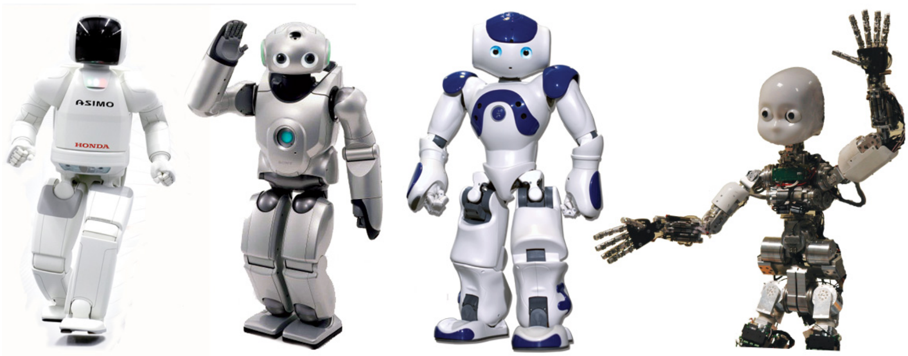

Figure 2.13: Examples of full body humanoids: ASIMO, QRIO, Nao and iCub.

not only voice commands, but also to the natural movements of human beings. This enables it to, for example, recognize when a handshake is offered or when a person waves and respond accordingly. It can also recognize movement directions such as pointing. ASIMO can recognize the objects and terrain of its environment and act in a way that is safe for both itself and nearby humans. It can respond to its name, face people when being spoken to, and recognize sudden, unusual sounds such as that of a falling object or a collision, and face in that direction. It is also able to respond to questions, either by a brief nod, a shake of the head or a verbal answer. ASIMO has the ability to recognize faces, even when ASIMO or the human being is moving.

To follow up on the success of AIBO, Sony developed and marketed (but never sold) QRIO [59], a bipedal humanoid entertainment robot. QRIO ("Quest for cuRIOsity") was developed as entertainment robot and is very well known for its impressive dancing capabilities. It has the abilities to interact with humans using speech, body posture, and gestures. For realizing the emotional influence on QRIO's behavior the EGO (Emotionally GrOunded) architecture was developed [116]. This control architecture of QRIO consists of a perception part, an emotion representation, predefined behaviors, and a memory part. In combination with voice and face recognition, QRIO is able to remember people as well as their likes and dislikes. After QRIO recieved the integrated motion control for walking, jumping and running, the world's first running humanoid robot (23cm/s) was credited in Guinness World Records in December 2003. As mentioned in section 2.4.2, Sony announced a termination of all commercial and research activity in robotics in 2006, including AIBO and QRIO.

On August 15, 2007, Nao replaced the robot dog AIBO as the standard platform for the Robocup ("Robot Soccer World Cup"), an international robotics competition. Nao is an autonomous, programmable and medium-sized humanoid robot, developed by the French company Aldebaran Robotics. Nao Academics Edition has 25 DOF since it is built with two hands with gripping abilities. All versions feature an inertial sensor and 4 ultrasound captors that provide stability and positioning within space to Nao. Nao also features a powerful multimedia system (4 microphones, 2 hi-fi speakers, 2 CMOS cameras) for text-to-speech synthesis, sound localization or facial and shape recognition amongst various other abilities. The software package includes a dedicated programming software, Aldebaran Choregraphe and Nao is compatible with Microsoft Robotics Studio, Cyberbotics Webots and Gostai Urbi Studio.

ICub [10, 92] is a full-fledged humanoid robot "child" with sophisticate motor skills and several sources of sensory information including vision, sound, touch, proprioception, and vestibular. At 104 cm tall, the iCub has the size of a three and half year old child. With a total of 53 DOF, iCub will be able to crawl on all flours and sit up to manipulate objects. Its hands have been designed to support sophisticate manipulation skills. The iCub software was developed on top of Yarp[2]. Yarp is a set of libraries that support modularity by abstracting two common difficulties in robotics: namely, modularity in algorithms and in interfacing with the hardware. Yarp helps organize communication between sensors, processors, and actuators so that loose coupling is encouraged, making gradual system evolution much easier. Yarp includes a model of communication that is transport-neutral, so that data flow is decoupled from the details of the underlying networks and protocols in use.

2.4.3.4 Social Robots for Therapy

Robots that are designed specifically for therapy are depicted in Figure 2.14.

Paro [121, 122] is a seal type mental commitment robot. It has been developed by Dr. Takanori Shibata of Advanced Industrial Science and Technology, AIST, since 1993 for those who cannot take care of real animals and those who live in places where pet-animals are forbidden. Mental Commitment Robots are designed to provide 3 types of effects: psychological, such as relaxation and motivation, physiological, such as improvement in vital signs, and social effects such as instigating communication among inpatients and caregivers. Paro is equipped with the four primary senses; sight (light sensor), audition (determination of sound source direction and speech recognition), balance and tactile sense. Its moving parts are as follows: vertical and horizontal neck movements,

[2] YARP stands for Yet Another Robot Platform.
More info: `http://eris.liralab.it/yarpdoc/index.html`

Figure 2.14: Social Robots for Therapy. Paro, Robota, Keepon and The Huggable.

front and rear paddle movements and independent movement of each eyelid, which is important for creating facial expressions. Paro weighed about 2.7 kg. Paro has a behavior generation system consisting of two hierarchical layers of processes: proactive and reactive. These two layers generate three types of behavior; (1) proactive behavior: behavior-planning layer and a behavior-generation layer, (2) reactive behavior, and (3) physiological behavior. Paro resembles a baby seal, because this animal does not bring any expectations towards the children, in contrast to a cat or dog.

The humanoid robotic doll, Robota [14], has been developed within the AURORA project. As a robotic toy it supports a rich spectrum of multi-modal interactions involving speech, music and movements. Robota is able to synthesize speech as well as processing video through a serial link connected to a PC. Each Robota robot is a mini-humanoid robot, 60cm tall, whose face is that of a commercial doll [13]. Robota has the ability to recognize human faces and to direct its gaze towards the user, the ability to understand and learn a restricted vocabulary and the ability for simple imitation of the user's motion.

Keepon [76] is a small creature-like robot designed for simple, natural, nonverbal interaction with children. The minimal design of Keepon's appearance and behavior is meant to intuitively and comfortably convey the robot's expressions of attention and emotion. Keepon has a yellow snowman-like body that is 120 mm tall. Keepon has two eyes, both of which are color CCD cameras with wide-angle lenses (120° horizontally), and a nose, which is actually a microphone. Since the body is made of silicone rubber and its interior is relatively hollow, Keepon's head and belly deform whenever it changes posture or someone touches it. These four degrees of freedom produce two qualitatively different types of actions: 1) *Attentive action*: using gaze. This action enables eye contact and joint attention. 2) *Emotive action*: rocking (left to right) and bobbing (up and down).These behaviors suggest emotions such as pleasure and excitement about the target of Keepon's attention. With these two actions, Keepon can express what it perceives and how it evaluates the target. These communicative functions of Keepon's actions can easily be understood by humans, even babies and toddlers.

The Huggable [129] is designed for use in hospitals and nursing homes with a wide range of users from the elderly to small children. As such, it is important to develop a robot that invites users of all ages to interact with it. The Huggable is based upon a fantasy animal: *the Teddy Bear*. The Huggable is a new type of robotic companion capable of active relational and affective touch-based interactions with a person. In the head of the robot is an array of microphones, two cameras in the eyes (one color and one black and white), and a speaker in the mouth. In the body, the Huggable features an inertial measurement unit (IMU), passive potentiometers in the hips and ankles for joint position sensing, and an embedded PC with wireless networking. The Huggable currently features eight degrees of freedom (DOF) - a 3-DOF neck (nod, tilt, and rotate), a 2-DOF shoulder mechanism in each arm (rotate and in/out), and 1-DOF ear mechanism. In addition, these degrees of freedom feature a quiet and back drivable transmission system. The robot features a full body, multimodal sensitive skin system [128] capable of detecting affective and social touch [127].

2.5 Summary

It is clear that social robots will have a big influence in our future society, working for and in close cooperation with humans. When designing a new social robot, decisions have to be made on embodiment, level of interaction, level of intelligence(autonomy), and off course on the purpose for the new robot. In [42] Duffy presents also an overview of the ethical problems with affective technologies regarding; privacy, emotional manipulation, replacing humans with machines, robots in our personal space and a machine's ability to tell a perfect lie. One could also pose the question; what level of responsibility has a robot designer over the acts of his own creations? All these ethical discussions are not covered in the scope of this thesis, but are therefore not less important. The need for more guidelines on ethical issues concerning social robots has been recognized by the robotics community. Noel Sharkey [120] argues that the steady increase in the use of robots in day-to-day life poses unanticipated risks and ethical problems and that the introduction of robot ethics guidelines is needed immediately. One of the aspects that is often neglected in the robots presented in this chapter is the importance of safety, which plays a key role in HRI. In the next chapter the design choices for our robot Probo are presented. Most of these choices are based on the examples, guidelines and issues, presented in this chapter.

Chapter 3

Conceptual Design of Probo

"Everything you can imagine is real."

-Pablo Picasso-

3.1 Introduction

The development of the robotic platform Probo is part of the ANTY project. The ANTY project will be used to accomplish four important goals for Probo concerning different areas of interest:

- Probo for children in the hospital.
- Probo as a multidisciplinary research platform.
- Probo as an educational stimulant for students.
- Probo as an attraction pole for the media.

At the beginning the robot was given the name: Anty. Due to design considerations discussed later in this chapter, the name was changed to Probo. A previous study on the design of a robot for hospitalized children [115], contributed to the start of the ANTY project and has been used as starting point for the design of the social robot Probo. Some of the results in this study are considered in this chapter. The overall development of Probo can be divided into three main parts:

Development of the Probo concept. This part includes the design choices that have been made for the development of the social robot Probo and is presented in this chapter.

Development of the mechatronics. This part includes the hardware platform; responsible for the mechanical design, the actuators and the electronics that provide the power and data-acquisition. Cabling, sensors, electronics and mechanical parts are fitted in a modular construction design. This part is described in the thesis "The Development of the Huggable Social Robot Probo. On the Hardware Design and Construction." by Kristof Goris.

Development of the software architecture. This part includes the different software modules going from low-level sensor analysis and motor control to attention, emotion and homeostatic systems and the high-level control center that offers a GUI for any operator to control the robot. This part is described in the following chapters of this thesis.

3.2 Design Issues

All robot systems, whether socially interactive or not, must solve a number of common design problems. These include cognition (planning, decision making), perception (environment sensing), action (mobility, manipulation), human-robot interaction (user interface, input devices, feedback display) and architecture (control, electromechanical, system) [54]. Socially interactive robots, however, must also address those issues imposed by social interaction [20, 31]:

Human-oriented perception. A socially interactive robot must proficiently perceive and interpret human activity and behavior. This includes detecting and recognizing gestures, monitoring and classifying activity, discerning intent and social cues, and measuring the human's feedback.

Natural human-robot interaction. To achieve this, the robot must manifest believable behavior: it must establish appropriate social expectations. It must regulate social interaction, and it must follow social convention and norms that are natural in human-human or human-animal interaction.

Readable social cues. A socially interactive robot must send signals to the human in order to: (1) provide feedback of its internal state; (2) allow human to interact in a facile, transparent manner. Channels for emotional expression include facial expression, body and pointer gesturing, and vocalization.

Real-time performance. Socially interactive robots must operate at human interaction rates. Thus, a robot needs to simultaneously exhibit competent behavior, convey attention and intentionality, and handle social interaction.

3.3 Positioning of Probo

The different classifications and definitions for social robots as presented in Chapter 2 will help to define the position of Probo in the field of social robotics. The position for Probo will be defined on the level of interaction, embodiment and autonomy. For the robot platform Probo two different types of users can be distinguished:

The Interacting User The child that interacts with the robot using human-like communication abilities.

The Controlling User The operators that program, control or guide the robot's interactions with respect to the child.

This complies with the goals stated in the introduction; combining the creation of a robot for hospitalized children and the creation of a multidisciplinary research platform. An interaction diagram is presented in Figure 3.1.

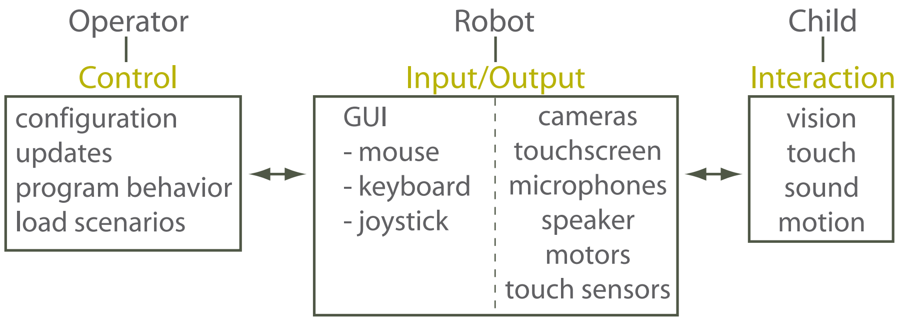

Figure 3.1: The Interaction Diagram between Operator, Robot and Child.

3.3.1 Level of Interaction

The most basic form of social interaction or human communication is nonverbal communication. Nonverbal communication is thought to have phylogenetic and ontogenetic primacy [24]. Phylogenetic primacy means that, in the evolutionary history of our species, nonverbal communication, preceded language in evolutionary time [43]. Burgoon et al. [24] argued that people tend to rely more heavily on nonverbal communication, especially in times of stress, compared with verbal communication because of our evolutionary heritage. Ontogenetic primacy means that, in the beginning of life, the importance of nonverbal communication supersedes that of verbal communication. An important part of non-verbal communication is the media that humans use to express themselves. According to Cole [28], the face plays a very important role in the expression of character, emotion and/or identity. Mehrabian [91] showed that only 7% of information is transferred by spoken language, that 38% is transferred by paralanguage and 55% of transfer is due to facial expressions. Facial expression is therefore a major modality in human face-to-face communication. In the daily life, people rely on face-to-face communication. Therefore it is most important in the development of Probo to focus on the ability to show facial expressions.

The robot Probo, will start out as a social interface as is mentioned in Section 2.3.2 and defined by Breazeal in [20], this subclass of social robots uses human-like social cues and communication modalities in order to facilitate interactions with people (i.e., to make the interactions more natural and familiar). For instance, one might adopt a performance model to communicate with others from far away using a robot avatar (giving the distal person both a physical presence and social presence to others). In this case, such a robot would need sufficient social intelligence to appropriately convey (or perform) a person's message to others, complemented with gaze, gestures, facial expression, etc. Because this class of robot tends to value social behavior only at the interface, the social model that the robot has for the person tends to be shallow (if any) and the social behavior is often pre-canned or reflexive.

3.3.2 Embodiment

The appearance of Probo is an imaginary creature combining features of ancient mammoths and anteaters. According to the classification suggested in [143] and [54], Probo's morphology can be defined as caricatured-zoomorphic. If a robot needs to portray a living creature, it is critical that an appropriate degree of familiarity be maintained. Mashiro Mori [97] contends that the progression from a non-realistic to realistic portrayal of a living thing is nonlinear. In particular, there is an "uncanny valley" (see Figure 3.2) as similarity becomes almost, but not quite perfect. At this point, the subtle imperfections of the recreation become highly disturbing, or even repulsive. Consequently, caricatured representations may be more useful, or effective, than more complex, "realistic" representations. There is no exact similarity with a well-known creature and consequently there are no specific expectations towards the behavior of this creature as would be in case of a cat or a dog. In Figure 3.2 we

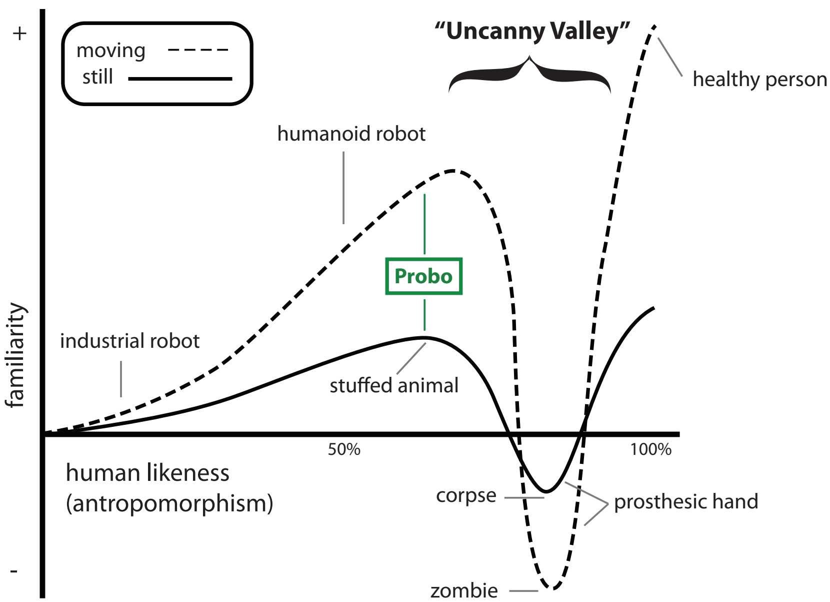

Figure 3.2: The relation between the users familiarity and the human-(or animal)-likeness of a robot with Mori's uncanny valley hypothesis (adapted from [87]).

can see that for still images stuffed animals have a very high familiarity. Most of the children have grown up in a close relationship with stuffed toys (or cuddly toys). The toys usually are a symbol of affection and are often presented as a gift to children or even adults. Until recently most of these stuffed toys were not actuated and did not react on the acts of the user. If a stuffed toy is promoted with the abilities of smooth and natural motions, familiarity will probably increase as can be seen in the design space.

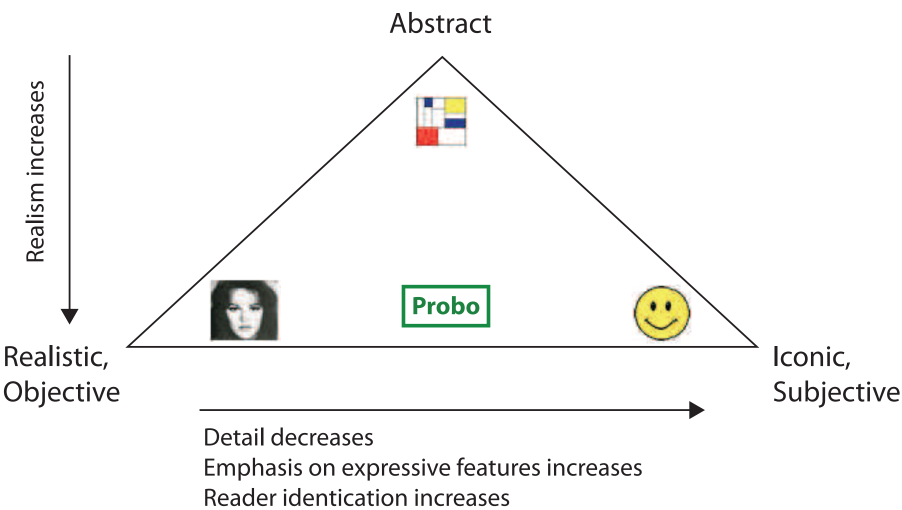

Figure 3.3: The Design space of faces in comics and narrative art (modified from [90]).

In his book "Understanding Comics" [90], Scott McCloud introduces a triangular design space for cartoon faces (Figure 3.3). The left apex is realistic, i.e. a perfect representation of reality, for example a photograph. Traveling to the right faces become more iconic, that is, the details of the face are stripped away to emphasize the expressive features; emoticons such as :-) are commonly used to express emotions in text-based communication (chat, email, sms, ...). The simplification has two effects. Firstly it allows us to amplify the meaning of the face, and to concentrate on the message rather than the medium. Secondly the more iconic a face appears the more people it can represent. Dautenhahn points out that iconography can aid the believability of a cartoon character [33]. And in [112], Reilly argued that unrealistic characters, such as in Disney animation movies, may be much more believable, because they are more abstracted and use exaggerated expressions. This explains why most of the stuffed toys are based on simplified version of real animals or on popular cartoons. Children don't wont an exact copy of a living animal, that would be almost similar to taxidermy[1] and is more suitable for museums. In [35], one of Dautenhahn's latest papers, Probo is positioned in the middle of the design space as depicted in Figure 3.3.

Both design spaces give a good evaluation for the use of stuffed toys based on the appearance of an imaginary animal. That is why the embodiment of Probo is set to have *the appearance of a stuffed imaginary animal with the ability of smooth and natural motions.*

3.3.3 Autonomy

In [8] and [143], is stated that autonomy is one of the defining factors for human-robot interaction. There is a continuum for robot control ranging from teleoperation to full autonomy; the level of human-robot interaction measured by the amount of intervention required varies along this spectrum. Constant interaction is required at the teleoperation level, where an operator is remotely controlling a robot. Less interaction from the operator is required as the robot has greater autonomy. In earlier work of Ortony, Norman and Revelle [103], the interplay of affect, motivation and cognition is considered in controlling behavior. Each is considered at three levels of information processing (Figure 3.4): *the reactive level* is primarily hard-wired and has to assure the quick responses of the robot making it look alive; *the routine level* provides unconscious, un-interpreted expectations and automotive activity; and the *reflective level* supports higher order cognitive functions, including behavioral structures and "full-fledged" emotions.

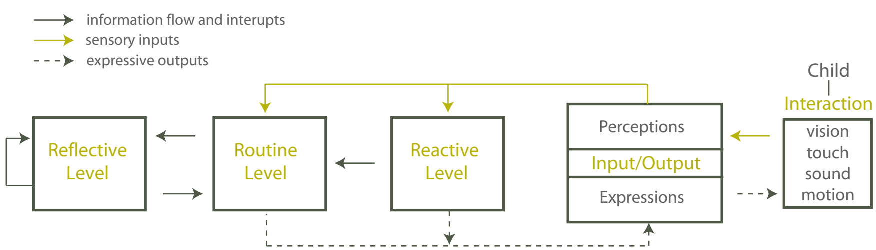

Figure 3.4: Three levels of information processing and their interconnections (adapted from [103]).

As mentioned in Section 3.3.1 the social interaction of Probo is on the reflexive or reactive level. To be operational a shared control between the operator and

[1] Taxidermy is the act of mounting or reproducing dead animals for display (e.g. as hunting trophies) or for other sources of study.

the reactive systems of the robot is necessary. The first autonomous systems should focus on primary, life-like functions such as gazing and basic emotional expressions. Gradually autonomy can increase together with the control interfaces of the operator, giving the operator higher levels of abstraction. Depending on the purpose/application, the robot must be able to perform autonomous interaction sessions, that are pre-configured and guided by an operator.

3.4 The Purpose of Probo

The social robot Probo will be used for two complementary purposes, respectively addressing both types of users.

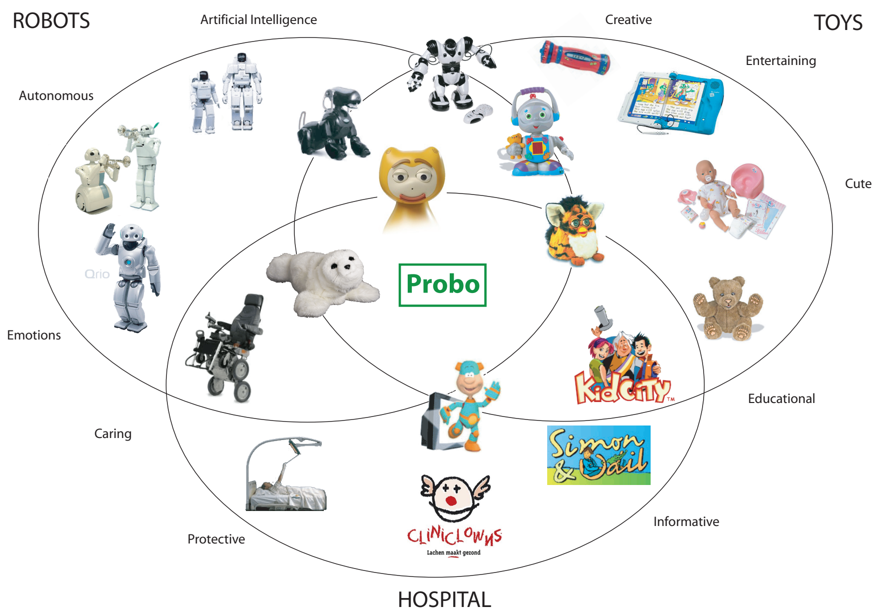

Figure 3.5: Position of Probo.

3.4.1 A Companion for Children in the Hospital

A hospitalization is a serious physical and mental occurrence. It brings children in some situations which are completely different from the experiences at home. In hospital, children's experiences are more limited and stressful due to the closed and protective environment. Hospitalized children are confronted with many difficulties and special needs [133]. These needs can be grouped in three major categories.

Distraction It is absolutely necessary for a child to be able to relax and have some amusement. Playing interactive games will distract the children so they can set their mind on something else than their illness.

Communication During the stay in a hospital a child's social life is more isolated and its normal level of communication is diminished, strongly reducing the communication with friends and family. Using technologies for distant communication, the robot can serve as an interface providing children a way to enhance their communication.

Information Children's fear is sometimes due to lack of information. The better children are prepared what they will see, hear and feel, the less fear they will have to face the surveys and examinations accompanied with their illness. New methods to present this information to children can improve their preparedness.

3.4.2 A Platform for Multidisciplinary Research

As stated before Probo is developed to serve as a multidisciplinary research platform in three major research areas:

Technological This includes mechatronics, software engineering, AI, vision, speech and audio. These research fields are extensively studied nowadays and they have a big impact on how the future robots will successfully collaborate with humans in a wide range of applications. The integration of these different components in one, smoothly and in real-time, working robot requires an adequate software architecture that is capable to distribute its processes over different processors, since they can be computational intensive tasks.

Biological/Social/Psychological This research is situated in human-robot interaction and emotional communication. Between robotics and biology often a nice symbiosis exists where models derived from humans or animals are implemented in a "bio-inspired" robot. Models that represent social interaction and human communication reach further then only the biological (or neurological) interpretations, and have to be supported from a social-psychological point of view.

Therapeutical This includes the exploration of the possibilities in Robot Assisted Therapy, such as social robots to engender mental effects, such as pleasure and relaxation to reduce stress and pain. Most of this research is discussed in Section 2.3.3.

3.5 The Development of Probo

Now that the position and purpose for Probo have been defined, there are still some issues we need to address a part from the software and mechatronic designs. That is why the *Design Drivers* are defined. These drivers are the guardians for a successful realization of the project during the development.

An Identity It is important to create a believable character that is attractive for children. The appearance should be well designed together with a connecting story board (defining name, history, family and background). This identity will help to build up a sort of relationship between a robot and a human.

Safe and Soft Interaction A robot that is in close interaction with humans and especially children should at anytime guarantee an intrinsic safe interaction. If the embodiment is based on a stuffed animal, one must feel the softness during interaction. This is what will make the robot huggable.

Modular Design Each building block or component has a different task and can be developed more or less independent from the others. This will result in a flexible architecture allowing to adapt to different users or purposes and ideal for future expansion.

User Friendly Design The operator of the robot can be anyone ranging from the medical staff in a hospital to researchers from various disciplines. Therefore a logical and easy to use GUI is needed.

Social Interaction The social capabilities for the robot should increase gradually. Following the modular design starting with a reactive level of interaction and a shared control with the operator. The focus will be on non-verbal face-to-face communication.

3.5.1 An Identity

One of the unique features of Probo, compared to other similar projects, is that this character has its own identity, which is of major importance for communication and emotional interaction with humans. Classical animators are masters at conveying intentionality through characters. In the “Illusion of Life” [130], Thomas and Johnston stress the importance of an identity for making animated characters believable. Robot identity or personality is conveyed in numerous ways. Emotions are often used to portray stereotype personalities: timid, friendly, etc [146]. There is a reason to believe that if a robot had a compelling personality, people would be more willing to interact with it and to establish a relationship with it [20][74].

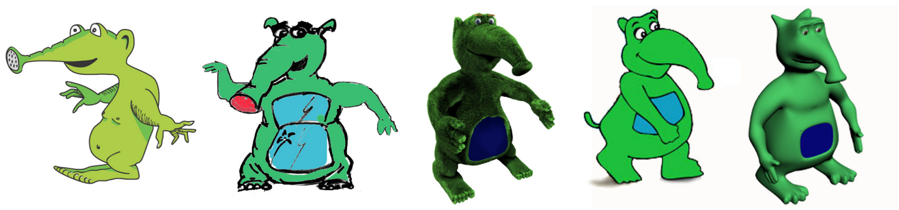

Figure 3.6: A chronological overview of the design of Probo with 2D and 3D drawings.

A number of 2D and 3D designs for Probo have been made, a chronological overview is presented in Figure 3.6. In order for Probo to become a believable character, the identity of Probo also includes a name, a family and a history. The name Probo is derived from the word *proboscidea*. Proboscidea is an order that now contains only one family of living animals, *Elephantidae* or "the elephants", with three species (African Bush Elephant, African Forest Elephant, and Asian Elephant) [142] (see Figure 3.7). During the period of the last ice age there were more, now extinct species, including a number of species of the elephant-like *mammoths* and *mastodons*. In the name Probo we can also see the word "ROBO" which emphasizes the robotic nature of Probo. Also the word "PRO" is recognized to underline the positive effects on research aspects on one side and education and welfare of children on the other side. The colour of the robot is green, because this colour evokes mainly positive emotions such as relaxation and comfort. In [72] the relationship between colour and emotion were tested, whereas the colour green attained the highest number of positive responses. The majority of emotional responses for the green colour indicated the feelings of relaxation and calmness, followed by happiness, comfort, peace, hope, and excitement. Green is associated with nature and trees, and thus created feelings of comfort and soothing emotions.

The history of Probo starts in the Ice Age where he lived among other similar species such as the mammoths and mastodons. About 12.000 years ago, warmer, wetter weather began to take hold. Rising sea levels swamped the coastal regions. Forests replaced open woodlands and grasslands across the continent. The Ice Age was ebbing. As their habitats disappeared most of the Ice Age creatures became extinct. Probo managed to migrate north and was frozen underneath the ice-cap at the North Pole. Here Probo was contained in a large ice cube, until recently the earth started to warm up again. Due to global warming the polar caps started to melt and create large floating chunks of ice drifting into open sea. Probo escaped inside such a chunk of ice and finally arrived at Mainland Europe. His quest here is to help children overcome their difficulties and diseases and to bring more joy into their lives.

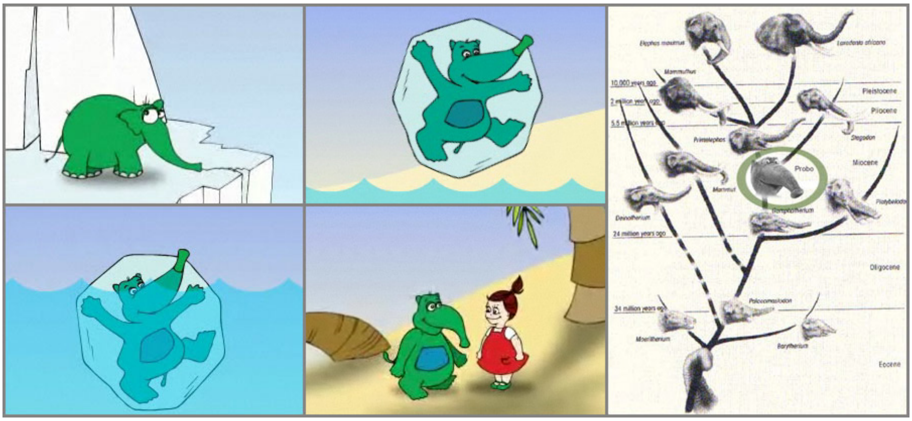

Figure 3.7: The history of Probo.

3.5.2 Safe and Soft Interaction

The need for safety in robotics was already identified by Asimov in 1942 [4]. Nowadays, for industrial robotics, standards define requirements on the robotic device and its working environment, e.g. ISO 10218-1 (2006). As domestic robotics emerges, robotic safety has become an important item on the research agenda. It is up to the robotics society to evaluate their robotic products on safety. Nevertheless, safety legislation for domestic robotic products is far behind on technological advances. As opposed to industrial applications, in domestic robotics, the design of the operational environment can not be influenced and the need for intrinsically safe robotic designs rises [138].

To build safety aspects intrinsically in the robot's hardware all the motors have a spring in series (series elastic actuation, SEA [110]) so in case of a collision the robot will be elastic and safe while providing a soft touch. A triple layered construction provides safe interactions and the soft touch for the user, see Figure 3.8. It also protects the robotics inside and gives our robot the final form. The layered construction consist of hard ABS covers mounted on the aluminium frame of the robot, shielding the internals. These covers are encapsulated in a PUR foam layer, that is covered with a removable fur-jacket. The fur-jacket can be washed and disinfected, and complies to the European toy safety standards EN71-1, EN71-2 and EN71-3. The use of the soft actuation principle together with well-thought designs concerning the robot's filling and huggable fur, are both essential to create Probo's soft touch feeling and ensure safe interaction.

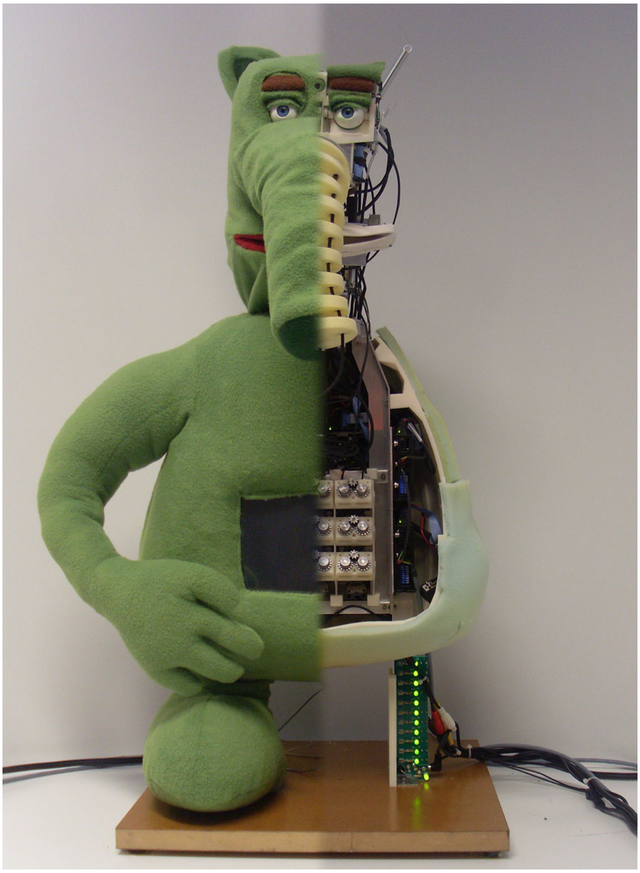

Figure 3.8: The triple layered protection covering the internal mechatronics of the prototype Probo: ABS covers, PUR foam, fur jacket.

3.5.3 Modular Design

As the comparison with the development of the personal computers has been made before. Also in the computer industry modularity in design has lead to many advantages. Besides reduction in cost (due to lesser customization, and less learning time), and flexibility in design, modularity offers other benefits such as augmentation (adding new solution by merely plugging in a new module), and exclusion [6]. Computers use modularity to overcome changing customer demands and to make the manufacturing process more adaptive to change. Modular design is an attempt to combine the advantages of standardization with those of customization. In the design of the robot's hardware the use of different building blocks or components is a must. Each of the building blocks will represent logical components such as a neck-system, the trunk-system, the eye-system, ear-system and mouth system. Each of these components is composed from smaller components (see Section 4.2).

The software controlling these block should evidently follow the same design principles. Modular programming is a software design technique that increases the extent to which software is composed from separate parts, called modules. Conceptually, modules represent a separation of concerns, and improve maintainability by enforcing logical boundaries between components. Modules are typically incorporated into the program through interfaces. The modular design in the software is described in Section 4.5.2.

3.5.4 User Friendly Design

The concept of a graphical user interface (GUI), that originated from Engelbart [47] and Alan Kay [12], changed the way that people interacted with electronic devices such as computers. A GUI offers graphical icons, and visual indicators, as opposed to text-based interfaces. A GUI uses a combination of technologies and devices to provide a platform the user can interact with, for the tasks of gathering and producing information. A series of elements conforming a visual language have evolved to represent information stored in computers. This makes it easier for people with little computer skills to work with and use computer software. The GUI developed to comply to the different users that can perform as the operator is presented in Section 4.5.3.

A robot can be regarded as a computer with a physical body that enables it to interact with the real world. Hence, if one considers the characteristics of their physical embodiment, robots can also be seen as interfaces for human beings. A robot used as an interface between the real and information worlds can be referred to as a robotic user interface (RUI) [123]. Probo can be used as a RUI to give children access to distraction, communication and information. This interface can enhance the interaction with the accessed applications and present them in a more child friendly way. The touch screen in the belly of Probo allows access to any existing computer applications. An example of a distractive game is the ProboGotchi game, described in Section 6.7.

3.5.5 Social Interaction

According to the The social interaction for Probo is focused on nonverbal communication. In order to achieve this Probo has the following abilities:

Human-oriented perception Probo's perceptional system includes the sensors needed to gather the data from the environment. This data is processed into input stimuli or low-level perceptions. This processing consists of the detection and classification performed by vision, audio and touch analysis. These low-level perceptions are the basic inputs for the attention system. The attention system together with the perceptional systems are presented in Chapter 5.

Readable social cues In order to express emotions Probo uses an emotion space to show facial expressions. A non-sense affective speech will be used to intensify the expressions of the face. The homeostatic system simulates an emotional state based on internal needs and influenced by the user's actions. The simulation and expression of the emotions are presented in Chapter 6.

A remarkable trunk To enhance the social interaction a moving trunk was introduced, giving Probo his unique identity. The trunk is made of foam and is actuated using three elastic cables and thereby provides an intrinsic safe interaction. The moving trunk is used to attract the user's attention and to provoke interaction. Additionally the trunk improves the expression of emotion in the face.

Smooth and natural motions All the interactions must be able to follow the human interaction rate, and therefore demand a high performance. The motions of the robot have to be fast and smooth. In order to achieve the motor system has to mix all the incoming motion requests in the right priority with smooth transitions. The motor system and the tools to create these motions are presented in Chapter 7.

3.6 Summary

This Chapter presented the design issues for the development of Probo (see Figure 3.9) and its position in the world of social robotics. Probo can be defined as a zoomorphic social robotic interface with a complementary purpose as both a multidisciplinary research platform for HRI, and a companion for hospitalized children. The main aspects are a huggable appearance, an attractive trunk or proboscis, animated ears, eyes, eyebrows, eyelids, mouth, neck, and finally a touch screen in its belly. The internal mechanics of the robot are covered with foam and a removable fur-jacket, in such a way that Probo looks and feels like a stuffed animal. Probo is provided with a user-friendly control center that allows an operator to share control with automated systems. Gradually more autonomy can be introduced using the modular software architecture.

Figure 3.9: The prototype of the huggable robot Probo interacting with children.

Chapter 4

The Software Platform for Probo

"The important thing is not to stop questioning. Curiosity has its own reason for existing. One cannot help but be in awe when he contemplates the mysteries of eternity, of life, of the marvelous structure of reality. It is enough if one tries merely to comprehend a little of this mystery every day. Never lose a holy curiosity."

-Albert Einstein-

4.1 Introduction

Following the design specifications described in the previous chapter, a bottom-up approach for the development of the software components was the best choice to gradually increase autonomy and provide the operator with higher levels of abstraction. In a bottom-up approach the individual base elements of the system are first specified. These elements are then linked together to form larger subsystems, which then in turn are linked, sometimes in many levels, until a complete top-level system is formed. This strategy often resembles a "seed" model, whereby the beginnings are small but eventually grow in complexity and completeness. This requires a simulation environment with a virtual model of the robot that can be connected with the first software components. In this way each of the components can be tested and evaluated without the physical robot.

This chapter briefly presents the hardware platform of Probo, followed by the virtual model of Probo and its implementation in a software component. After a view on existing robotics software platforms, the choice was made to work with the C# programming language in the .NET environment that can implement the virtual model using the XNA framework. An overview of the software concepts and the developed architecture is presented later in this chapter.

4.2 The Hardware Platform

An overview of Probo's hardware architecture is presented in Figure 4.1. The prototype of Probo has a fully actuated head with a total of 20 DOFs. The head is composed of several anthropomorphic systems that are presented as the motion and expression layer. In order to create motions, each of these systems makes use of the following layers; Actuation (the actuators), Low level drive (the EPOS motor controllers), Supply (the electronics) and finally the High level drive or the software Motor System. Two special actuation systems are introduced to comply with the hardware design specifications:

Non Back Drivable Servo (NBDS) is a custom made servo motor system that mainly consists of a DC motor, a non back drivable gear train and a control unit.

Compliant Bowden Cable Driven Actuator (CBCDA) is a custom made passive compliant servo motor system which transmits motion over a relative long distance compared to its own size, using a Bowden Cable mechanism. The use of the CBCDAs creates the opportunity to group and to isolate the different actuators and to place them anywhere in the robot.

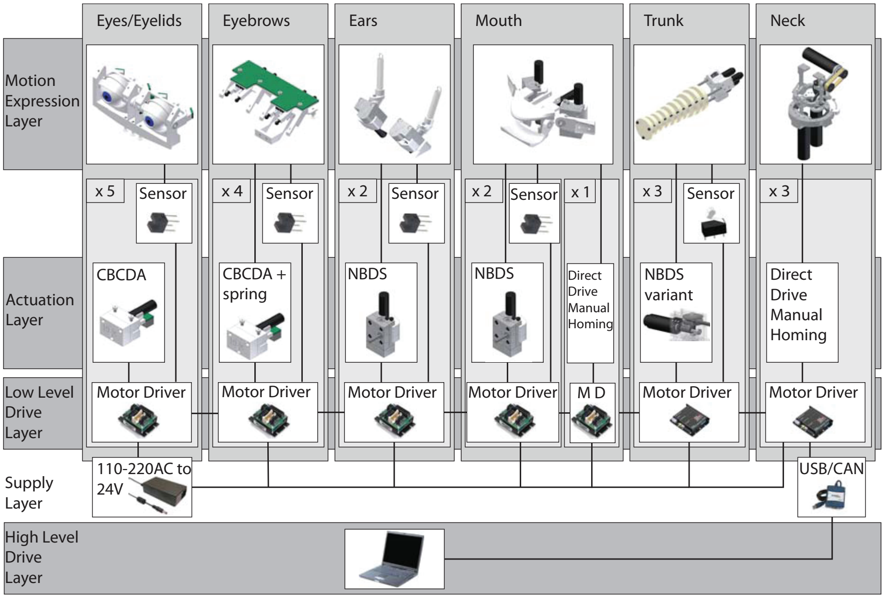

Figure 4.1: The hardware architecture of Probo.

The anthropomorphic systems used for the expressions of the robot are:

The Eyes and Eyelids [5 DOF] are actuated using the CBCDA mechanism as depicted in Figure 4.2. The eyes are able to pan separately but they tilt together conform to human-eye motions. This gives the robots its gazing ability. The upper eyelids are actuated separately, allowing the robot to wink, blink and express emotions.

The Eyebrows [4 DOF] play a very important role in facial expressions. Each eyebrow is connected with two actuated rods. By moving these rods both translation and rotation of the eyebrows can be obtained (see Figure 4.3). The rods are both actuated using the CBCDA mechanism.

The Ears [2 DOF] are actuated helical springs. Because the ears are likely to be touched during interaction, safety is very important. A flexible foam core in the shape of the ear is placed over the spring to ensure intrinsic safety. The NBDS actuator is used to point the ear's opening to the front (to look attentive) and point it to the ground (to make the ears lie flat to the back) as depicted in Figure 4.4.

The Trunk [3 DOF] consists of a foam core with segmented extension discs as depicted in Figure 4.5. Axial to the centerline, three flexible cables are

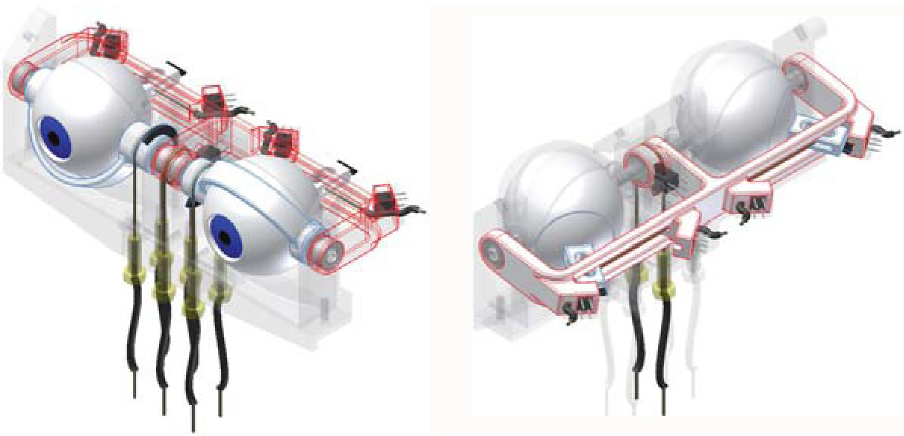

Figure 4.2: The design of eyes and eyelids.

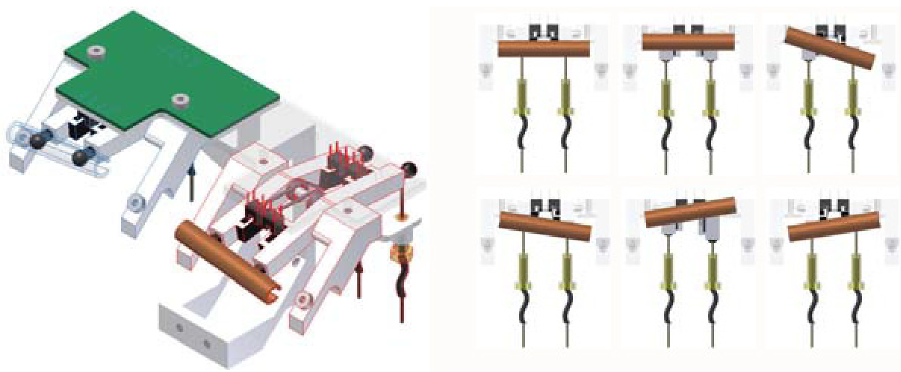

Figure 4.3: The design of the eyebrows.

guided through the discs and attached to the front disc. The end of each cable is attached to a reel. A coordinated winding and unwinding of the reels results in a curling motion of the entire trunk. A high compliance and durability of the trunk is ensured by using a foam material actuated with flexible cables. Interaction with this trunk will be intrinsic safe both for the user, that can not be hurt, and for the motors, that can not be broken. The actuators use an NBDS variant to move the reel.

The Mouth [3 DOF] contributes mainly to the robots facial expressions by moving the corners of the mouth. In addition to the expressions, the mouth also serves to enhance the affective speech by performing basic lip-sync movements. The mouth exists out of silicon that is attached to the four parts. Three of them are actuated; the lower lip (direct drive)

Figure 4.4: The design of the ears.

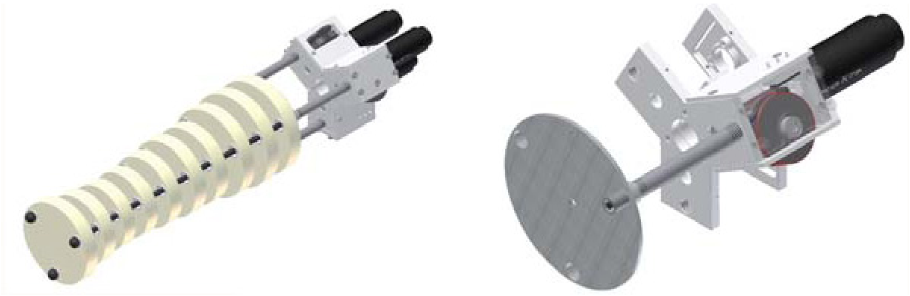

Figure 4.5: The design of the trunk.

and the two mouth corners (NBDS) as can be seen in Figure 4.6.

The Neck [3 DOF] provides the motions for the head (including the previously described systems). The neck supports human-like motions i.e. rotation, nodding and canting. The motions are direct driven and combined with balanced springs for gravity compensation. The design of the neck is presented in Figure 4.7.

All these systems are assembled together in a compact but modular head design as depicted in Figure 4.8. The moving parts are presented in a yellow color. For more information on the hardware design and construction see [62].

4.3 The Virtual Model

Based on the designs for the appearance of Probo (as seen in Figure 3.6), a 3D model has been created. The model was constructed in 3DS Max, using

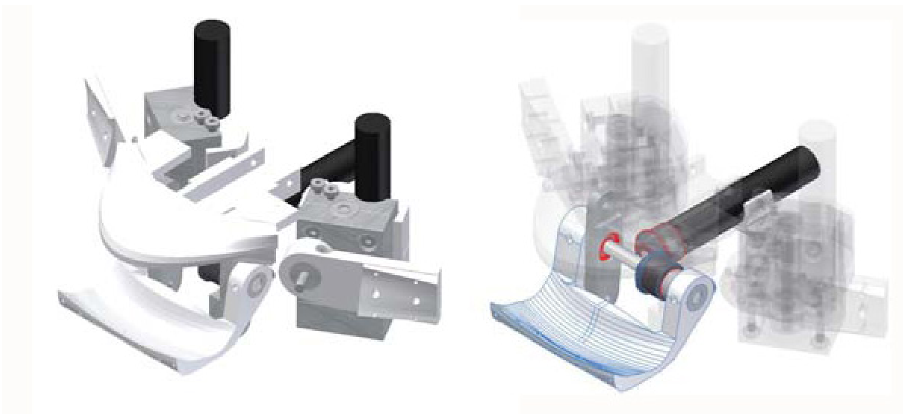

Figure 4.6: The design of the mouth.

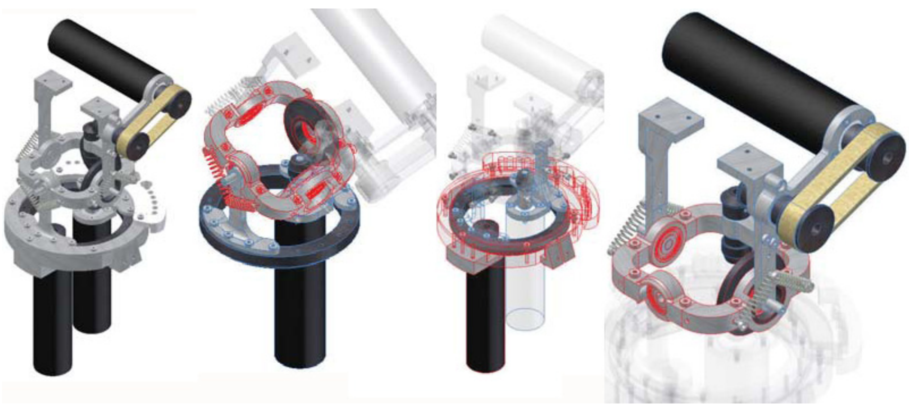

Figure 4.7: The design the neck.

techniques of professional video and game animation such as skeletal animation and skinning. Skeletal animation is a technique in computer animation in which a character is represented in two parts: a surface representation used to draw the character (called the skin) and a hierarchical set of bones used for animation only (called the skeleton). As the character is animated, the bones change their transformation over time, under the influence of some animation controller. Each bone in the skeleton is associated with some portion of the character's visual representation. Skinning is the process of creating this association. Portions of the character's skin can normally be associated with multiple bones, each one having a scaling factor called vertex weights, or blend weights. The movement of skin near the joints of two bones, can therefore be influenced by both bones. On the left in Figure 4.9, the hardware platform

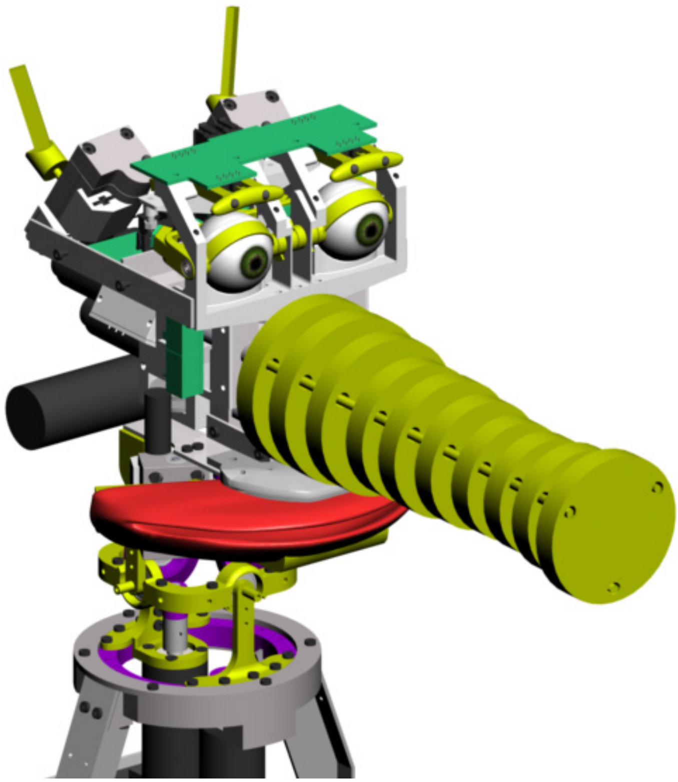

Figure 4.8: The design of the head of Probo.

with moving parts or bones (in yellow) is presented. The plastic covers, shown in the middle, encapsulate the hardware and are designed to provide the base support for the skin, responsible for the appearance of Probo. Adjustments on all parts were made to make everything fit together.

4.3.1 Simulation

To simulate realistic deformation of the model, the mechanical parts need to be associated with certain vertices of the skin. An example can be seen in Figure 4.10. On the left side the vertex weights associated with the head(bone) are shown followed by the vertex weights associated with right ear(bone). The influence (ranging from 0% - 100%) that the moving part (or bone) will have on the deformation of the skin is presented by a colour gradient. Moving the right ear will deform the skin according to the distribution of the weights of the vertices, shared between the ear and the head. A good distribution of the

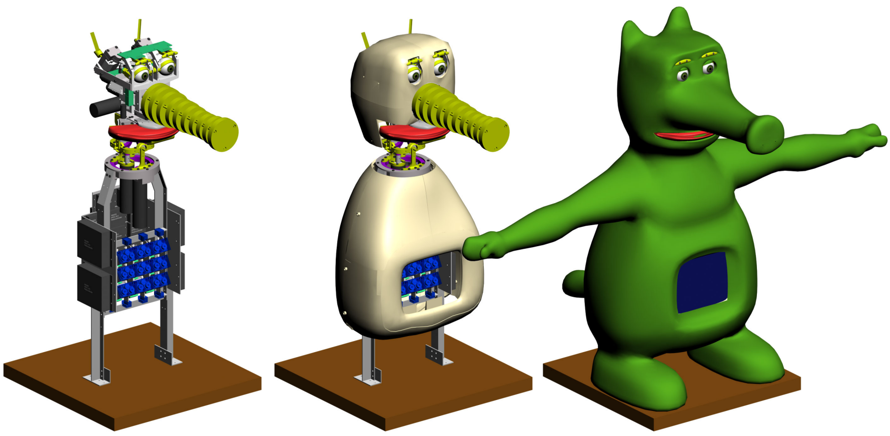

Figure 4.9: The design of the layered structure of Probo.

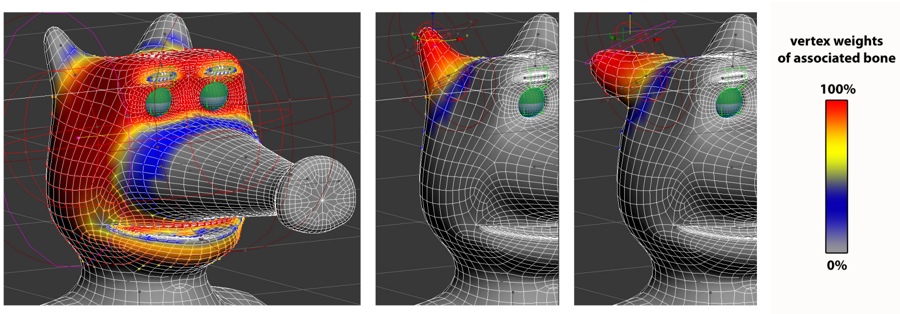

Figure 4.10: The weights of the skin's vertices associated with the head and with the right ear.

vertex weights will determine the quality of the simulation. All solid moving parts can be simulated this way. But the simulation of the motion of the trunk requires a different approach. The foam core that is actuated using three elastic cables provides curling (trunk-like) motions. To simulate these motions, the trunk is represented by 6 joints or bones as seen in Figure 4.11. Each of the joints will rotate according to the length of the elastic cables (C_i). A simple conversion for the rotation angles of each joint is given by Equations 4.1,4.2 and Figure 4.12. Changing the length of the cables will cause a rotation on each of the joints, simulating the desired curling motion. When all skinning is done every vertex of the skin will be associated with one or more bones.

$$\alpha = (C_1/2 - C_2/2) \tag{4.1}$$

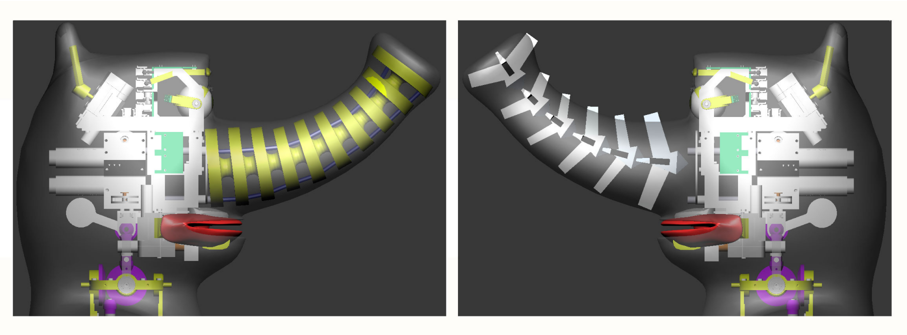

Figure 4.11: The curly motion of the trunk (left) is simulated using 6 bones (right).

$$\beta = (C_1/2 + C_2/2) - C_3 \tag{4.2}$$

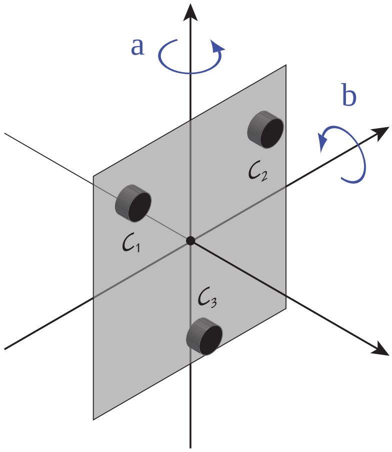

Figure 4.12: Conversion graph for the trunk motion.

The mechanical designs for the hardware structure have been imported and used as the skeletal structure of the model. The mechanical parts are converted to bones and linked together to obtain realistic kinematical movements. Now each DOF can be animated, simulating the actual motions as in the mechanical designs. The result is that we can not only simulate the mechanical motions, but also have a visual feedback on the deformation of the skin. The movements can be controlled by using sliders to set the desired angle for each DOF and simulate actuation of the parts (Figure 4.13). The resulting motions and expressions are used in an iterative design of the hardware parts. The expressions are rendered to obtain pictures that are used for user evaluations (see Section 6.4).

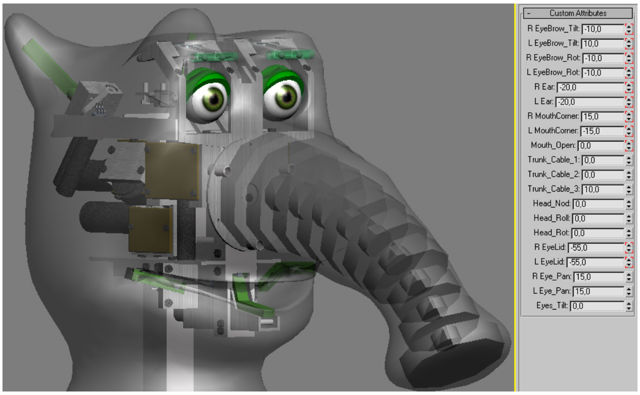

Figure 4.13: Virtual model with control slider for the DOF.

4.3.2 Implementation

The virtual model is imported into a software component using Microsoft's XNA Game Studio. XNA Game Studio is an integrated development environment (IDE) for development of games, providing a set of managed classes that are compatible with the .NET Framework. This means that the virtual model can now be controlled from our software and rendered in a real-time 3D window. The whole model is imported, including the skin mesh, the mechanical (bone) structure, the skin data (vertex weights) and textures. The inverse kinematics of the mechanical structure are also included, giving easy access to all positions of the model's body parts at any given time in the 3D space. All the motions of the robot are to be smooth and natural. So to avoid brusque movements and high inertia, the DOF are filtered before they are send to the motors (see Section 7.6). These filters are provided by the DOFProboFilter component. To simulate the similar motions in the virtual model as in the physical robot, the same (low-pass) filters from this component are applied before the virtual model receives the DOF positions. Additionally the provided TrunkInertia and TrunkGravity filter are applied to simulate both gravity and inertia that is visible for this non-rigid body part.

Based on the design of the virtual model a stuffed model (depicted in Figure 4.14) of Probo was made in two sizes. A scaled model (25cm) is used for promotional purposes, while the real sized model (75cm) is to evaluate the appearance and touch of Probo. Sensors are first introduced in the stuffed model for evaluation and later implemented on the real robot. In combination

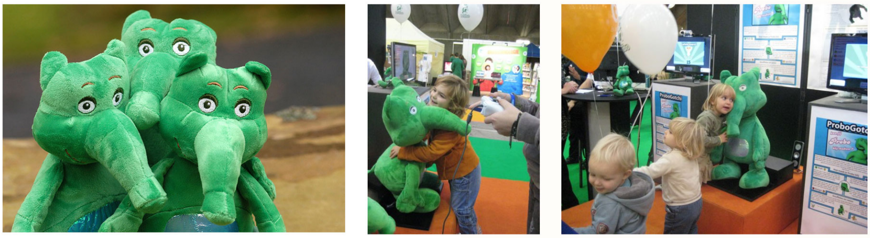

Figure 4.14: The stuffed models of Probo used in interaction test with children.

with the virtual model a new 3D game (ProboGotchi) was developed using the stuffed model as an interface. This setup provides an easy way to perform interaction test to evaluate new software units, with a new appealing gameplay for the children. The ProboGotchi game is described in more detail in Section 6.7.

4.4 The Software Platform

A good software platform provides us with a set of reliable software components that are integrated into a unified platform. The software platform for Probo had to comply with the following demands: easy to use, free of charge, support for social interaction, support for modular programming and connectivity with a simulation environment that supports the virtual model (bone animation and skinning). Most of the robotics platforms are intended for industrial robots (manipulating robot arms) or mobile robot platforms (navigation). Important platforms that can be used for service robots are:

Evolution Robotics Software Platform (ERSP) [98] is a commercial software development kit (SDK) from Evolution Robotics. It incorporates a robust visual pattern recognition system (ViRP) and a visual-based localization and mapping (vSLAM) system. The platform comes with a graphical toolkit that can be used for building programs based on a number of reusable building blocks. ERSP doesn't provide a simulation environment.

Open Robot Control Software (Orocos) [22] is an open source, general-purpose and modular framework for robot and machine control. It incorporates four C++ libraries: the Real-Time Toolkit, the Kinematics and Dynamics Library, the Bayesian Filtering Library and the Orocos Component Library. It is optimized for real-time applications and comes with a set of drivers for selected robot hardware. Orocos doesn't provide a unified graphical development or simulation environment.

Player/Stage/Gazebo [77] The Player is an open source network server for a number of robot hardware platforms, providing a simple interface to the sensors and actuators over the IP network. Stage and Gazebo are both complementary simulators for multi-robot outdoor environments, compatible with the Player server. Stage is designed to simulate a very large robot population with low fidelity, Gazebo is designed to simulated a small population with high fidelity.

Microsoft Robotic Developer Studio (RDS) [68] is a Windows-based environment for robot control and simulation. It is free of charge for academic and hobbyist developers, supporting a wide variety of robot hardware. It includes a lightweight services-oriented runtime that does most of the messaging and thread management tasks. RDS provides easy access to the sensors and actuators via a .NET based concurrent library implementation and supports a number of languages including C#. It comes with a simulation environment (based on the XNA framework) and a graphical service development toolkit.

Robotic platforms that give more support for social interaction are;

OpenHRP [70] (Open Architecture Humanoid Robotics Platform) is a software platform for humanoid robotics, and consists of a dynamics simulator, view simulator, motion controllers and motion planners of humanoid robots. OpenHRP has been developed by AIST, the University of Tokyo and MSTC. OpenHRP is integrated by CORBA, and each module, e.g. the dynamics simulator, of OpenHRP is implemented as a CORBA server. The users can develop their own software on OpenHRP as well as replace its building blocks by their own ones.

URBI [5] (Universal Real-Time Behavior Interface) is a robotics platform built by Gostai. Instead of creating a graphical service development environment, URBI introduced its own language (urbiScript) optimized for creation of parallel and distributed event-driven services. Gostai's partners provide components such as computer vision and speech recognition. The URBI Studio, a graphical development kit, is still under development. URBI doesn't have its own simulation environment, but it has been integrated with Webots, a popular commercial robotics simulation environment.

YARP [52] stands for Yet Another Robot Platform and is part of the RobotCub project. It is an open source software platform, providing a set of libraries, protocols, and tools to keep modules and devices cleanly decoupled. YARP supports building a robot control system as a collection of programs communicating in a peer-to-peer way, with a family of connection types that meet the diverse, sometimes contradictory, and always changing needs of advanced robotics. YARP does not provide a graphical development or simulation environment.

It is hard to find a robotic platform that suits our demands. Another approach is to look at a more general platform or middleware and build our own dedicated software on top of that. Middleware is computer software that connects software components or applications. It sits “in the middle” between application software working on different communication and operating systems, giving programmers the means to implement component-oriented applications. Some of the robotic platforms consider themselves as middleware for robotics (URBI, YARP). Others (Orocos, OpenHRP) are based on general-purpose middleware, for example CORBA. Another general-purpose middleware is Mircosoft’s .NET framework (Figure 4.15). The 2.0 version supports interoperation between different languages and consists of the Base Class Library (BCL), Windows Forms, ASP.NET and ADO.NET. The .NET framework includes the BCL in order to encapsulate a large number of common functions, such as user interface, data and data access, database connectivity, cryptography, web application development, numeric algorithms, and network communications, which makes the programmer’s job easier. Windows Forms provides GUI components for the .NET framework and is described in Section 4.5.3.

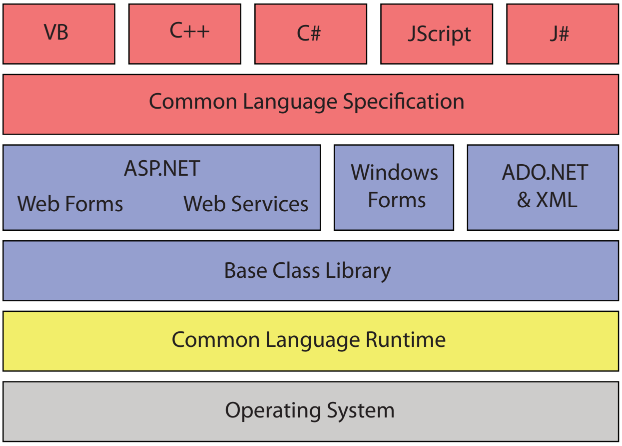

Figure 4.15: The .NET Framework 2.0 Stack.

In 2006 Microsoft released “XNA Game Studio Express”, based on the .NET Framework 2.0, as a free toolkit for 3D game developers. XNA offers the tools to build a 3D simulation environment and supports the skinning techniques needed for the virtual model of Probo. The professional IDE to develop .NET applications is Microsoft's Visual Studio. Visual Studio offers a good programming environment to satisfy our demands, with a smart interface for component-oriented development of the GUI, Virtual Model and hardware-drivers. Besides the easy use of Visual Studio, the Microsoft Developer Network (MSDN) provides information, documentation, and discussion by Microsoft itself and by the community at large. All the software can be written in managed code. Managed code will be managed at runtime, providing core services such as memory management, thread management, and remoting, while also enforcing strict type safety in the code. These services are provided by the common language runtime (CLR), also see section 4.5.2.

Working with Visual Studio and the .NET Framework does not exclude the use of robotic platforms in the future.

4.5 Software Concepts

Based on the design drivers defined in the previous chapter, the focus during the development of the software should be; a modular and user-friendly design to provide the operator an interface for social interaction. The first step for a modular design is object oriented programming (OOP), now commonly used in the mainstream software applications. One step further is the use of component-oriented programming grouping the classes and interfaces into separate components, that are again contained within assemblies. Both programming paradigms support the development of a user-friendly GUI.

4.5.1 Object Oriented Design

OOP has become the dominant programming methodology since the mid-1990s and is now supported by many programming languages, including C#. OOP focuses on the objects we want to manipulate rather than the logic required to manipulate them. The core attribute of OOP is the use of a class as a definition to create objects of that class. A class defines the abstract attributes (properties) and behaviors (methods) of the possible objects. During runtime one or more objects are created based on the class definition, named an instance of that class. This instance will have his own set of values for each of the attributes and the behaviour defined in the object's class. For example: The "Lassie" object is an instance of the Dog class, with "brown" as a value for the furColour property and a ChangeFurColour method to modify that property value. As in the case of the example, most of the time the software objects are representations of real world items. Another important aspect of OOP is inheritance. Derived classes are more specialized versions of a class, which *inherit* attributes and behaviors from their parent classes (or base classes), and can introduce their own.

4.5.2 Component Oriented Design

The software architecture for Probo is a structure of components that are dynamically linked to each other. Components can be seen seen as reusable building blocks, delivering services trough defined interfaces. Each component groups classes that belong toghether to obtain a logical building block. As recommended by Lowy in [85] the best way to manage components in .NET is the use of assemblies. An assembly is the basic packaging unit in .NET, assembling multiple physical files into a single logical unit. .NET assemblies contain code in Common Intermediate Language (CIL), which is usually generated from projects in .NET languages, and then compiled into machine language at runtime by the Common Language Runtime (CLR) just-in-time compiler (see Figure 4.16). Components can reside in either an application assembly or a library assembly. An application assembly project will be compiled to a executable file (EXE), while a library assembly project will be compiled to a dynamic-link library (DLL). Most of the components for Probo are contained in library assemblies. The linking between assemblies is done by adding references, allowing one assembly to use the components that are defined in other assemblies. This actually means that after compilation our components (contained in DLL's) can be used by any other project in a .NET programming language. In Visual Studio all the assembly projects are grouped in one "solution". In Visual Studio parlance, a solution is a set of code files and other resources that are used to build an application. Hence, a solution needs to have one startup project, an application assembly, that uses the references to other (library) assemblies to gain access to their components (see Figure 4.17).

Most of the assemblies developed for Probo are based on abstract representations of psychological, social and biological models related to social interaction. For good understanding these abstract representations will be named *systems* and their software counterparts will be named *software units*.

4.5.3 Graphical User Interface

Each software unit has a separate GUI, to control the parameters of its own system. In the .NET framework, the graphical application programming interface (API) is given the name Windows Forms. It provides access to the native Windows interface elements wrapped in managed code. The Windows Forms include common GUI components such as buttons, textfields, sliders, etc. With this approach the user (operator) is presented with a familiar interface. The difficult part of the development of the GUI is to give the user a good overview

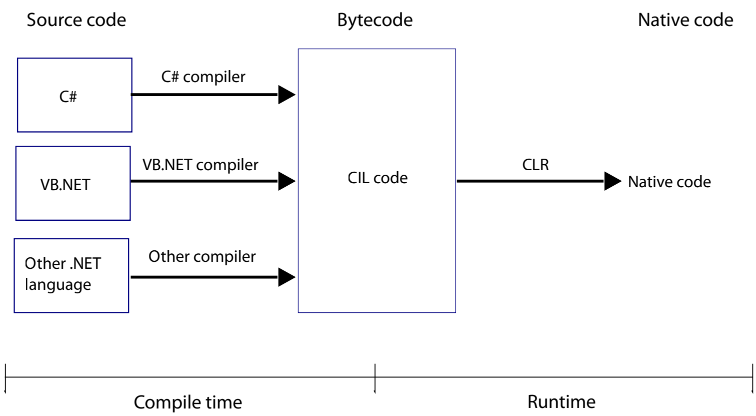

Figure 4.16: Common Language Runtime diagram.

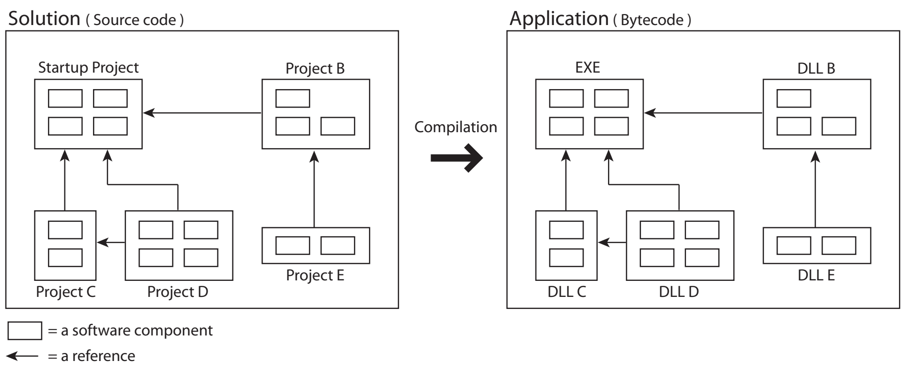

Figure 4.17: Component Oriented Design.

of all the systems that are used to control the robot. The modular design principle is applied here to offer a customizable interface suited for each type of user operating the robot. The global GUI consists of a container that is used for the layout of the different windows provided by each software unit. The windows can be arranged as desired; by dragging windows and dropping them on the panels for appropriate docking. The layout is automatically saved after closing, custom layouts can be saved and loaded at any time. An example of the two possible layouts is given in Figure 4.18.

The links between the GUI and the underlying logic are event-based. If the operator clicks on a button an event is fired to execute the corresponding

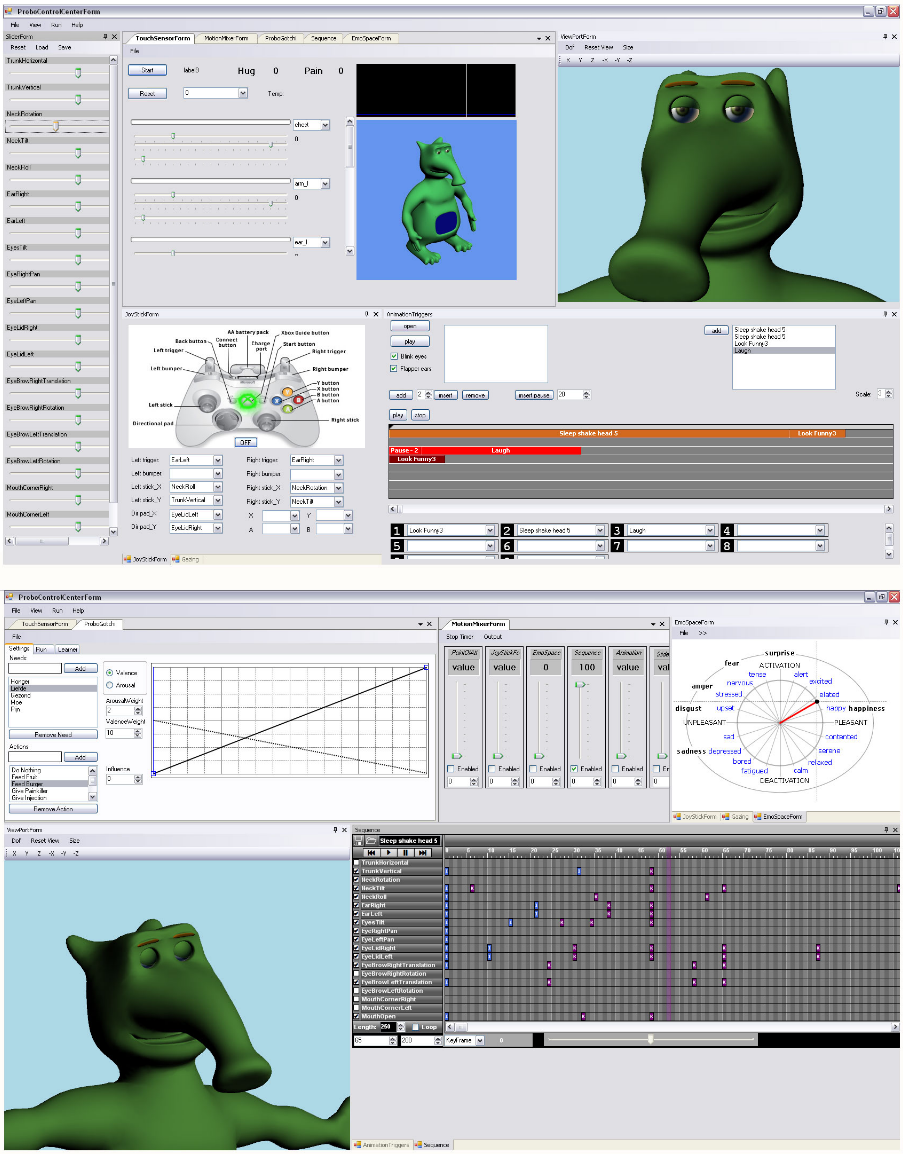

Figure 4.18: The GUI of Probo's software presented in two different layouts.

function of that button. Other systems will run timer-based or thread-based providing different levels of autonomy. These systems are influenced by the settings an operator has defined and the events that are triggered from the

GUI. In this way the GUI provides the operator with a shared control over the robot.

4.6 Software Architecture

With the component-oriented approach presented in Section 4.5.2, the following design methodology can be applied. If a new system is introduced, a system design is made and translated into software components. The components and their relationships are contained in an assembly or software unit. By referencing other developed units, their components and services can be used. If needed the assembly is first compiled as a stand-alone application for better testing and debugging. Subsequently, the assembly is added to the general solution were it can be tested in cooperation with the others.

4.6.1 The System Design

Following on the concept of social interaction defined in Section 3.5.5, an abstract system architecture is designed. The architecture and interaction with the users is presented in Figure 4.19. All features in black represent the systems that are currently implemented in the software. The Audio and Speech-system are under development by the Digital Speech and Audio Processing (DSSP) research cell of the VUB and will be implemented later. A preliminary attention system is implemented, pending to a more advanced system that is under development (See Section 5.2). Future development will focus on a behavior system, including the ability for an operator to program scenarios.

The interaction of both types of users (child and operator) is presented. Considering the interaction with a child, three main systems of the architecture can be noted; input (perceptional system), processing (control system) and output (expressional system). The perceptual system is responsible for capturing the child's actions. It consist of a group subsystems, with the human-oriented systems (visual, auditory and tactile) and the Object Identification (ID) system. These systems need to analyze the sensor data in order to detect social cues. The vision analysis includes the detection of human faces and later also objects and facial features such as facial expressions. Audio analysis will include the detection of the direction and intensity of sounds and the recognition of emotions in speech. Touch analysis gives the location and force of touch, that is classified into painful, annoying or pleasant touch. A larger classification of haptic interactions will be developed later. The Object ID system uses RFID technology to identify (symbolic) objects. These social cues or stimuli are passed on to the control systems; attention and actions. The attention system uses the stimuli to determine the robot's focus of attention. The stimuli can also

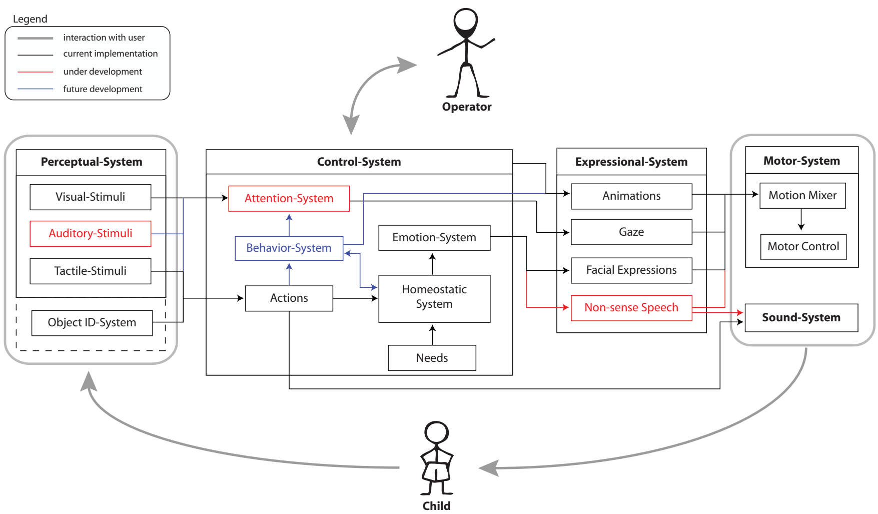

Figure 4.19: The system architecture of the robot Probo.

be recognized as certain actions, influencing the homeostatic system. Based on the defined needs and the perceived actions, the homeostatic system modifies the emotional state. Next Probo's emotional state and focus of attention are passed on to the expressional systems. The focus of attention is expressed using a gaze direction by giving motor commands for eyes and head motion. This motion is combined with the motor commands for facial expressions based on the emotional state, and the motor commands for animations triggered by the operator. Additionally the motions for lip synchronization derived from the non-sense affective speech will be included later.

4.6.2 The Software Design

The software architecture in Figure 4.20 presents the dataflow between the software units, starting and ending with the robot's hardware.

All the software units are initiated from the startup unit named the Control Center and thereby accessible from the GUI for the operator to control. All the dataflow connections represented here are linked when the Control Center is loaded. These connections are made using interfaces, public properties and events. The dataflow closely follows the system design as described in the previous section. Starting with the hardware sensors that are accessed trough the dlls provided by the manufacturers. And ending with the EPOS Motor Con-

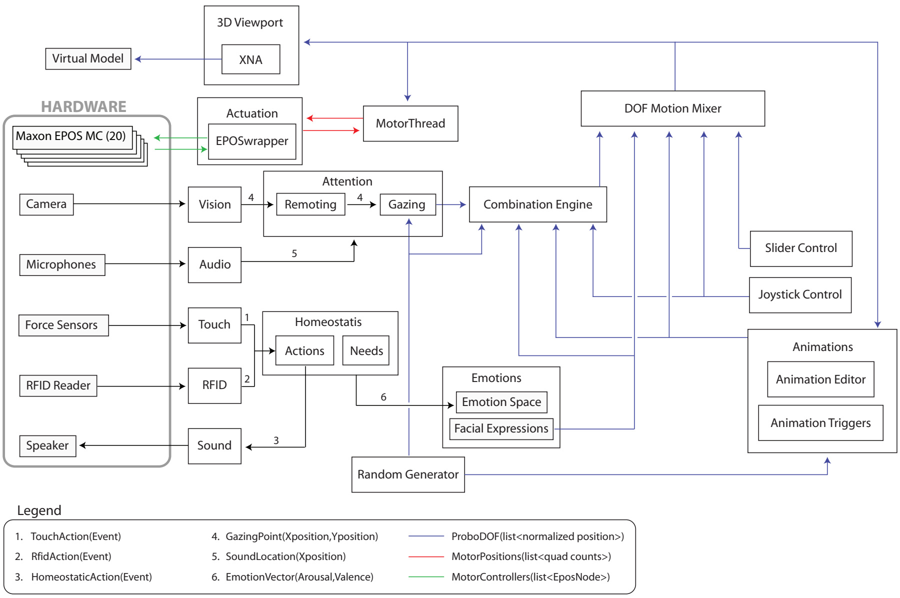

Figure 4.20: The software architecture of the robot Probo.

trollers that will drive the motors to provide motion for the hardware systems. The software units in between are organized in three groups;

The Perception and Attention System presented in Chapter 5.

The Homeostatic and Emotion System presented in Chapter 6.

The Animation and Motion System presented in Chapter 7.

The interaction between the units is obtained trough the Control Center, described in the next Section.

4.6.3 Control Center

The Control Center is the main software unit. It is an application assembly, providing the execution file to startup the platform. The unit contains only one component named “ProboControlCenterForm”, but has references to most other units. This component has the ability to use all the components inside these referenced units. The task of this component is to initialize all the components and provide them with the necessary links to each other. Most components will receive data, process it and offer the results to others. The windows form of the Control Center is set to be a multiple-document interface

(MDI) container. This means that the window can now contain the windows of other components. The implementation of the DockManager, a component from the WinFormsUI of Weifen Luo [86], provides the easy layout as described in Section 4.5.3. The different GUIs or window forms from the initialized components are now all available to the operator. Interaction with these windows is event-based; after an Event is fired (e.g. by pressing a button) the corresponding EventHandler will call the function to handle the Event. This is no problem for the GUIs because the Event and the EventHandler are located in the same component. But if for example a pain stimulus is detected by the touch component, it needs to send an event to trigger the action component of the homeostatic system. So if a component from one unit wants to send an event to another unit, the Control Center is needed. Because the Control Center holds references for both units it can handle the events coming from one component and call the handling function of another component. Another method for linking is the use of interfaces. This method is described in Section 7.5 and also here the Control Center provides the link between the units.

4.6.4 3D Viewport

The 3DViewport holds the virtual model of Probo described in Section 4.3. It is the simulation environment for the robot and provides visual feedback of the robot's motions to the operator. This unit contains three important components;

GamePart The GamePart is derived from the XNA framework Game class and has all the functions to load and render the 3D content. One of its properties is an object of the ProboModel class.

ProboModel The ProboModel is responsible for updating the bone positions resulting in the motion of the virtual model. Matrix translation and rotations are used based on the DOFProbo to create the correct motions in 3D space.

ViewPortForm The actual viewport is created using a derived windows form class, ViewPortForm, that contains the render window from the GamePart. By doing so the viewport behaves as a windows form and can be implemented in the Control Center.

An additional component of the GamePart is the Camera that is used to change the view in the 3D environment. It is a first person shooter (FPS) camera, allowing the user to walk (using keyboard) and look (using mouse) around in the 3D space. The camera provides a projection with a perspective field of view (near plane of 10 and a far plane of 10000). All objects in the viewing frustum will be rendered and are displayed in the ViewPortForm as presented in Figure 4.21.

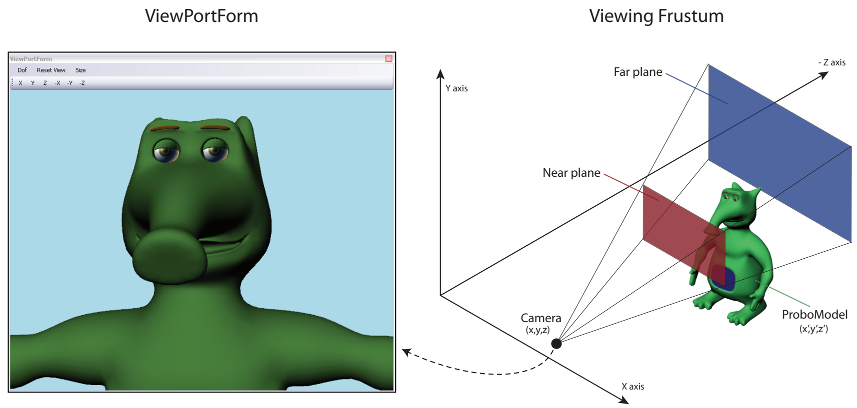

Figure 4.21: The viewing frustum of the camera and the corresponding viewport.

4.7 Summary

The software platform presented in this chapter defines the structure for the other building blocks that will be presented later. The control center has the task to manage this structure and provide an overview of the entire platform during operation. By using the 3D model of Probo, a real time feedback of the results that the platform produces is presented to the operator. Besides the operational feedback, the 3D model has proven to be a great help during the design and evaluation of both hardware and software parts. In compliance with the possibility for future expansion the architecture supports a shared control between the operator and autonomous systems. These autonomous systems are influenced by the social interactions with a user. It is the perceptual system, presented in the next chapter, that uses its sensors to capture the interactions and produce useful stimuli for the other systems.

Chapter 5

The Perceptual System

"After this, there is no turning back. You take the blue pill - the story ends, you wake up in your bed and believe whatever you want to believe. You take the red pill - you stay in Wonderland and I show you how deep the rabbit-hole goes."

-Morpheus in The Matrix-

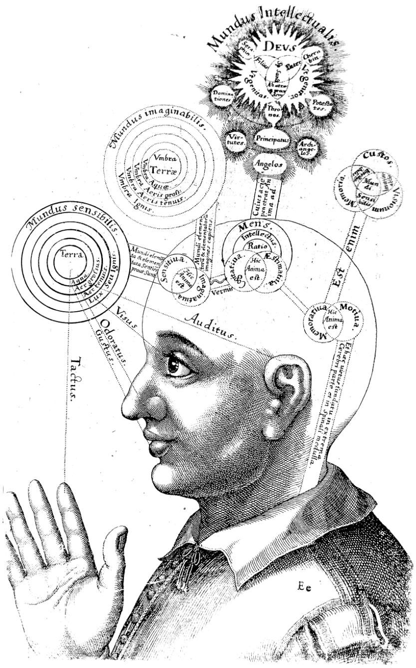

Figure 5.1: Robert Fludd's depiction of perception (1619).

5.1 Introduction

To obtain a meaningful social interaction with humans, a social robot must be able to perceive the world in the same way as humans do, i.e., sensing and interpreting the same phenomena that humans observe. This means that, in comparison to the perception required for robots with conventional functions (localization, navigation, obstacle avoidance), social robots must possess perceptual abilities similar to humans. In particular, social robots need perception that is human-oriented: optimized for interacting with humans and on a human level [60]. The perceptional system for Probo consists of a visual, tactile and auditory system to comply with human's social interaction. Additionally a non-human perceptive subsystem is introduced providing an identification of selected objects. The task of these perceptive systems is to generate stimuli that can be used in higher level systems. One of these systems closely linked with the perceptual system is the attention system, providing a point of attention for a human-like gazing system. This chapter presents the perceptual systems of Probo and the development of the attention system.

5.1.1 The Human Perceptual System

In [105], human perception is defined as the active reception and coordination of information received through our sensory systems in order to make sense of the environment and to behave effectively within it. In contrast with the direct and immediate sensations actually received and transmitted, perception is the transformation of that information into nerve cell activity that is transferred to the brain where further processing occurs. Our perceptual systems do not passively receive stimuli from the world, instead they actively select, organize, interpret, and sometimes distort sensory information. The real world then may not be the same as the one we perceive. Broadly, perception can be said to be the study of the human organism's relation to the physical world. The perceptual systems use our senses to perceive different aspects of this physical world:

1. The Kinesthetic System - position of body parts (muscles and joints)
2. The Vestibular System - balance and orientation
3. The Savory System - taste and smell
4. The Visual System - sight
5. The Auditory System - hearing
6. The Tactile System - pressure, temperature and pain

The perceptual systems are closely linked with our sensory system. The sensory system is a part of the nervous system responsible for processing sensory information. A sensory system consists of sensory receptors, neural pathways, and parts of the brain involved in sensory perception. Commonly recognized sensory systems are those for vision, hearing, somatic sensation (touch), taste and olfaction (smell). Sensory systems code for four aspects of a stimulus; type (modality), intensity, location, and duration. A stimulus modality (sensory modality) is a type of physical phenomenon that can be sensed.

5.1.2 Probo's Perceptual System

The goal for Probo is to have a successful non-verbal communication with children. The vestibular and savory system are not implemented, because they don't provide an added value to the robot or its functionalities. The kinesthetic system is represented by the simulation of the actuated body parts in the virtual model. Additionally the motor positions are monitored by the motor control system. The most important perceptual systems that are gradually implemented in the robot are the visual, auditory, tactile and (non-human) object ID system. Similar with the human perceptual system, sensors are used to sense the environment. All the necessary components to read out the sensors are contained in the DLLs (provided by the hardware manufacturers) that are referenced from the software units of the perceptional system.

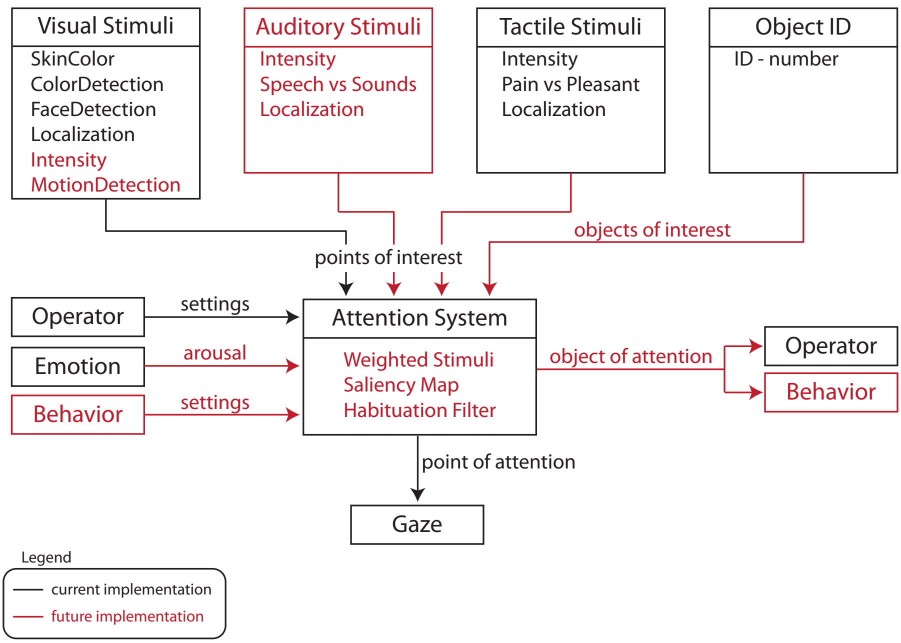

Figure 5.2: The architecture of the Perceptual system in collaboration with the Attention system.

Figure 5.2 presents the different perceptual systems and the detectable stimuli that will be used to determine a point of attention for the robot. This point of attention is the target point for the gaze system. In the current implementation the robot's gaze is influenced by the visual stimuli (see Section 5.3) and the operator's settings. The auditory stimuli are under development and are described in Section 5.4. Currently the tactile stimuli (Section 5.5) and the object ID (Section 5.6) are only serving the homeostatic system, described in the next Chapter. First the different steps to develop an advanced attention system are presented, including the visual attention system that is currently implemented.

5.2 The Attention System

A multidisciplinary team of researchers from machine vision, audio processing, cognitive/biologic psychology, orthopsychology and robotics have joined forces to study and implement a (joint) attention mechanism for Probo in order to improve natural human-robot communication. The proposed attention system is based on a common mechanism for exogenous and endogenous attention. A schematic overview of this model is depicted in Figure 5.3 and described in described in [64] and [58]. This advanced attention model combines the psychological aspects of attention with the computational modeling of visual attention in machine vision. Further integration and translation of this model is needed before the implementation in the robot Probo can be started. Following our

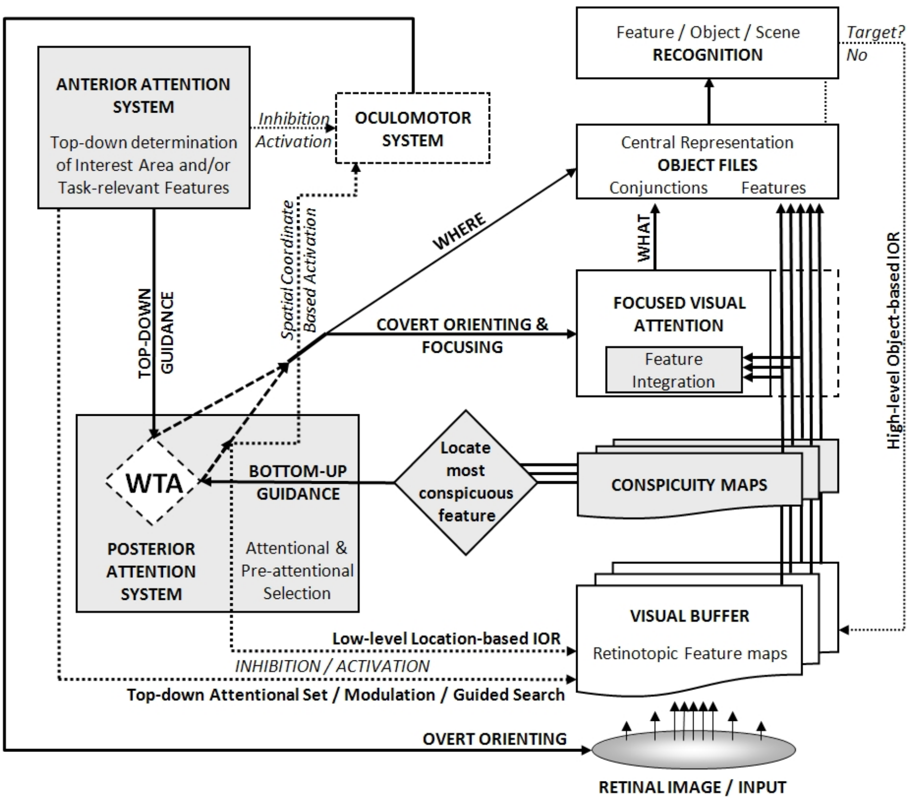

Figure 5.3: The advanced attention model of Henderickx, Maetens, Geerinck and Soetens [64].

design strategy, gradually more components are implemented building up towards a more advanced model. The current design for future implementation (see Figure 5.2) is data-driven and focusses only on the bottom-up or exogenous attention (except from any operator interventions). The purpose of this exogenous attention system is to pick out low-level perceptual stimuli that are particularly salient or relevant at that time, and to direct the robot's attention and gaze toward them. This provides the robot with a locus of attention that it can use to organize its behavior. A perceptual stimulus may be salient for several reasons. It may capture the robot's attention because of its sudden appearance, or perhaps due to its sudden change. It may stand out because of its inherent saliency, such as a load noise or a red ball may stand out from the background. Or perhaps its quality has special behavioral significance for the robot, such as being a typical indication of danger. The endogenous attention is currently carried out through the control settings from the operator. But the existence of a future behavior system can share this control by for example focussing on face detection rather than on a red ball if a scenario for communication with a human is activated.

5.2.1 Visual Attention

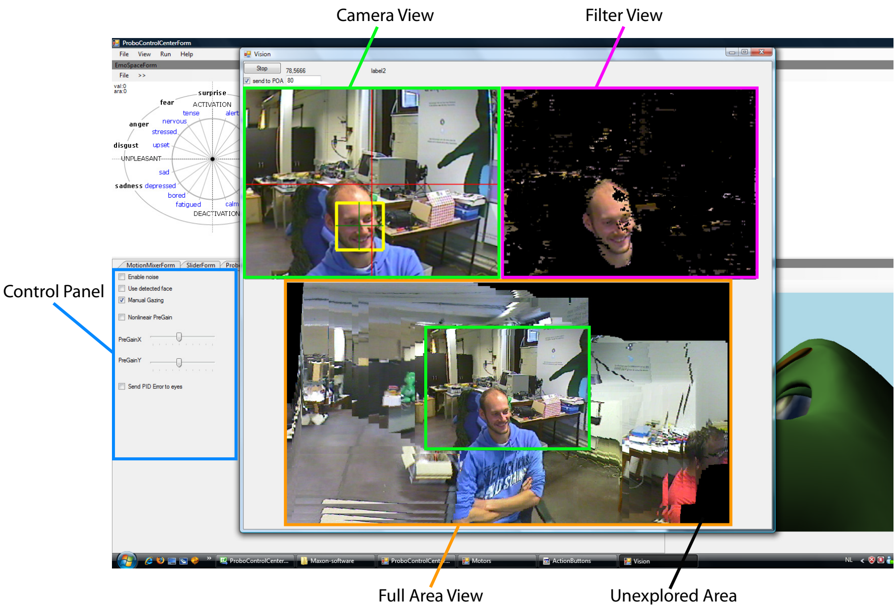

Figure 5.4: The GUI of the current attention system.

In the current implementation the attention system is limited to visual attention. The operator is presented with the GUI depicted in Figure 5.4. The GUI is divided into four parts:

Control Panel This panel holds the controls that an operator can use to alter the settings of the attention system.

Camera View This window renders the actual camera image. Detected faces or objects are displayed with a surrounding color frame.

Filter View Depending on the settings, a specific filter is applied on the original camera image. The result is rendered in the Filter View.

Full Area View When moving the camera a mosaic image is constructed to obtain a full area view, covering the full visual range of the robot.

Using the checkboxes from the Control Panel the operator can choose from five options:

Manual Gaze Using the Full Area View the operator can click with the mouse on the desired point of attention.

Random Gaze The point of attention is set after a random time interval, using a random point. These random objects are made using the Random Generator, within the limits of natural motions.

Color Detection After choosing the desired color, a color filter is applied on the camera image and is presented in the Filter View. The center of the largest color area or blob is set as the current point of attention.

Skin Color Detection A special case of color detection is the skin color. The accompanying filter will focus on the skin-tone and the center of the largest area will be set as point of attention.

Face Detection With this option all frontal faces will be detected. The center of the largest face will become the center of attention.

The corresponding point of attention is passed on to the Gazing unit. The Vision analysis that is performed to calculate the desired point of attention can be very demanding on the computational power. Especially if it is to be expanded in the future. The point of attention is therefore passed on through a Remote Object (see Section 5.3.4). This allows the vision processing to take place in a different process or even a different computer, so it will not influence the computational resources for the other systems. The Gazing unit needs to calculate the positions of the eyes, eyelids and head to obtain a simplified human-like ocular motor control.

Table 5.1: Classification of Eye Movements (adapted from [100]).

Eye Movement Type	**Function**
	Gaze Stabilization
Vestibulo-ocular	Initiated by vestibular mechanisms during brief/rapid head movement
Optokinetic (vestigial in humans)	Initiate by visual mechanisms during slow head movement
	Gaze Shifting
Vergence	Adjusts for different viewing distance
Smooth Pursuit	Follows moving visual target
Saccade	Directs eyes toward visual target

5.2.2 Gaze System

Normal visual perception requires the proper functioning of ocular motor systems that control the position and movement of the eyes to focus the image of the object-of-interest (i.e., the visual target) on corresponding areas of the retinas of the two eyes [100]. The gaze systems control the eye movement to direct the eyes towards a visual target and to follow the movements of the visual target. They coordinate the movement of the two eyes to ensure that the images on the two retinas fall on corresponding areas of the binocular field. There are two functional classes of eye movements (Table 5.1): those that stabilize the eye when the head moves or appears to move (gaze stabilization) and those that keep the image of a visual target focused on the fovea (a.k.a., foveation) when the visual target changes or moves (gaze shifting). Two gaze stabilization systems operate during head movement: the vestibulo-ocular and the optokinetic systems. Vestibulo-ocular and optokinetic movements are conjugate movements in which both eyes move in the same direction. Voluntary or guided saccades are eye movements initiated to bring an object-of-interest into view or initiated under direction (e.g., to the command "eyes left"). Saccades consist of short, rapid, jerky (ballistic) movements of predetermined trajectory that direct the eyes toward some visual target. Smooth pursuit (tracking) is an eye movement elicited by a moving visual target that the eyes follow voluntarily or under direction (e.g., the request to "watch a moving ball"). Pursuit movements are described to be voluntary, smooth, continuous, conjugate eye movements with velocity and trajectory determined by the moving visual target. By tracking the movement of the visual target, the eyes maintain a focused image of the target on the fovea. Notice that a visual stimulus (the moving visual target) is required to initiate this eye movement.

For the simplicity of the design there is only one camera for the visual system. This generates enough data to perceive the most relevant stimuli. The implementation of a stereo camera would unnecessary increase design complexity and computational power. The camera is mounted between the eyes and is therefore controlled along with the head motions. As a result the gaze of the eyes is simulated without feedback from the camera. The current implementation doesn't allow depth measurement to control the vergence of the eyes. So after a visual stimuli is introduced, first the eyes will target the point of interest, by using rapid eye-movements (saccades). The head follows using gaze stabilization to keep the eyes locked on the target. To achieve this, the gaze control for the robot supports two different approaches. The *filtered gaze control* and the *proportional-integral-derivative (PID) gaze control* depending on the control option, chosen by the operator. The filtered approach serves to comply with the manual and random gaze. In both cases a new target location (Target(x, y)) is suddenly introduced. The change from old to new target can be defined as a step function. The step-response from a low-pass filter determines the head positions and the inverse is used for the eye positions. A representation of the gaze control reacting on a target step-response is depicted in Figure 5.5. With this filtered gaze control the eyes will saccade to the newly introduced target and stay locked on the target while the head moves more slowly into position until the eyes are centered with respect to the head. Notice that the new targets can be anywhere in the full area view, even outside of the active camera view. The PID approach serves the cases of visual stimuli (color and face). Besides acquiring the target, the PID also provides some basic tracking for moving visual targets. There are two software PID components created to control both pan (motion on the x-axis) and tilt (motion on the y-axis). The PIDs are set to follow the target positions, with a continuous feedback from the camera view. Both PIDs have an error factor $e_x(t), e_y(t)$, defining the difference between the target and the current position (See Figure 5.6). The error factor is used to control the eye positions to ensure that the eyes are always locked on the target. The result is a system able to track moving stimuli; fast moving targets or changing targets will first be tracked with fast eye-movements, followed by the slower moving head.

5.3 The Visual System

If a robot is intended to interact with people, it needs an active vision system that can serve both a perceptual and communicative function [17]. An active vision system is a computer vision system where the viewpoint of the camera can be manipulated in order to investigate the environment, rather than passively observing it (passive vision). An active vision system operates on sequences of images rather than on a single frame and since it can scan over the scene, the range of the visual scene is not restricted to that of the static view. This allows the construction of a full area view. The perceptive stimuli

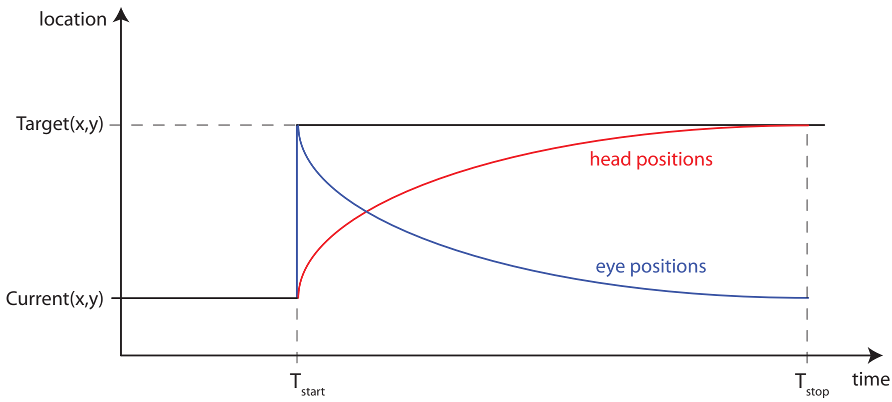

Figure 5.5: Reaction of the filtered gaze control on a new target (step-response).

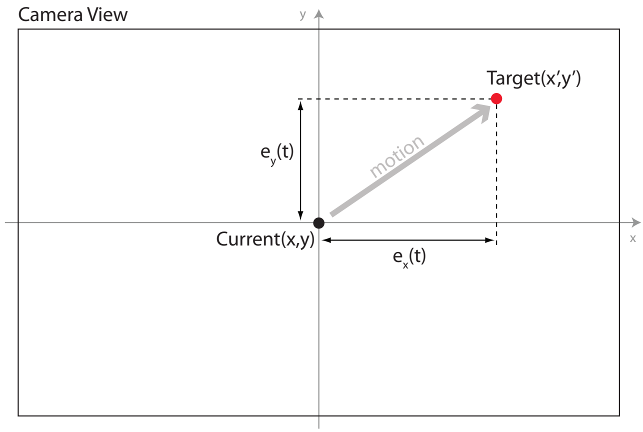

Figure 5.6: The target position for the PIDs controlling the head and the error factors controlling the eyes.

are obtained through vision analysis of the active camera view.

5.3.1 Camera

The human eye is a complex biological device. The functioning of a camera is often compared with the workings of the eye, mostly since both focus light from external objects in the visual field onto a light-sensitive medium. In the case of the camera, this medium is film or an electronic sensor; in the case of the eye, it is an array of visual receptors. With this simple geometrical similarity, based on the laws of optics, the eye functions as a transducer, as does a CCD camera. A CCD camera converts optical brightness into electrical amplitude signals using charge coupled device (CCD) image sensor. The vision sensor used in the robot Probo is a CCD snake camera (type: CM-R3010CP4, Ø16mm) mounted just above the eyes as depicted in Figure 5.7. Motions of the camera are thereby coupled with the head movements. A capture device is used to convert the analog data from the camera and send it via USB to the computer. The capture device is accessed from the vision software unit to acquire the captured images for further processing.

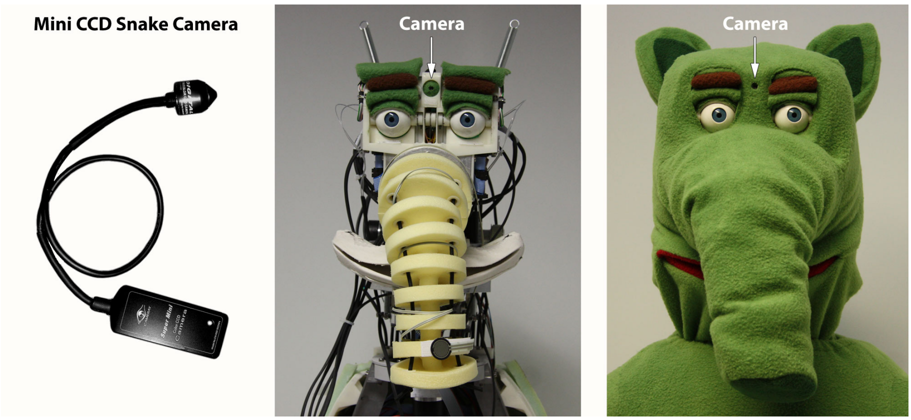

Figure 5.7: The camera and its placement just above the eyes of the robot.

The camera can be modeled using the pinhole camera model. To represent the mapping of three dimensions (an object) onto two (its image), the perspective projection (see Figure 5.8) of the pinhole camera model is used. A perspective projection is the projection of a three-dimensional object onto a two-dimensional surface by straight lines that pass through a single point [49]. Simple geometry shows that if we denote the distance of the image plane to the center of projection by f, then the image coordinates (x_i, y_i) are related to the object coordinates (x_o, y_o, z_o) by $x_i = \frac{f}{z_o} x_o$ and $y_i = \frac{f}{z_o} y_o$. The camera for Probo has a super cone pinhole lens with $f = 3.7mm$.

The pinhole lens will also distort the image, especially on the edges of the image/lens. This radial distortion can be corrected by displacement of the image's pixels location. The lens distortion can be roughly considered as a Barrel distortion, that is corrected using the inverse displacement (Pincushion distortion) as depicted in Figure 5.9. This is a rather simple approach, but it yields a acceptable result. A more advanced and better inverse distortion correction is presented in [37].

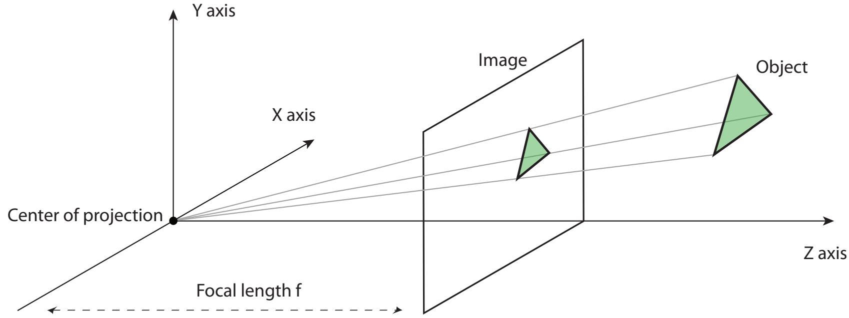

Figure 5.8: The perspective projection of the pinhole camera model.

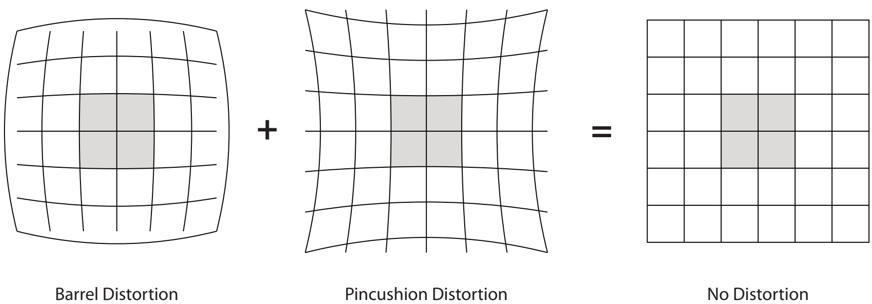

Figure 5.9: The correction of the lens distortion.

5.3.2 The Full Area View

To obtain the Full Area View, the vision system builds an overview of the full visual area using an image mosaic system. The full visual area is defined by the full range of the camera motion (see Figure 5.10). The image mosaic system links the actual camera image with its position, updating the full area view with new images when the camera is at a certain position. Figure 5.11 demonstrates the use of the mosaic system. In Frame 1 the person in red is in the active camera view and in the bottom right of the full area view for Probo. In Frame 2 Probo changes his active view by moving his head; now the person in blue is in the active camera view and in the center of the full area view. This means that the person in blue is in front of Probo, the last image of the active camera view when Probo was looking down on his right side is also visible in the full area view. In Frame 3, the person in red moves to the front of Probo and enters the active camera view. This person is also still present in the bottom right of the full area view, because this is the image taken from the last time Probo was looking at that area. This mosaic system provides an updating visual memory map of the environment. That can be used by the operator to direct the robot's gaze to objects that are located outside of Probo's view.

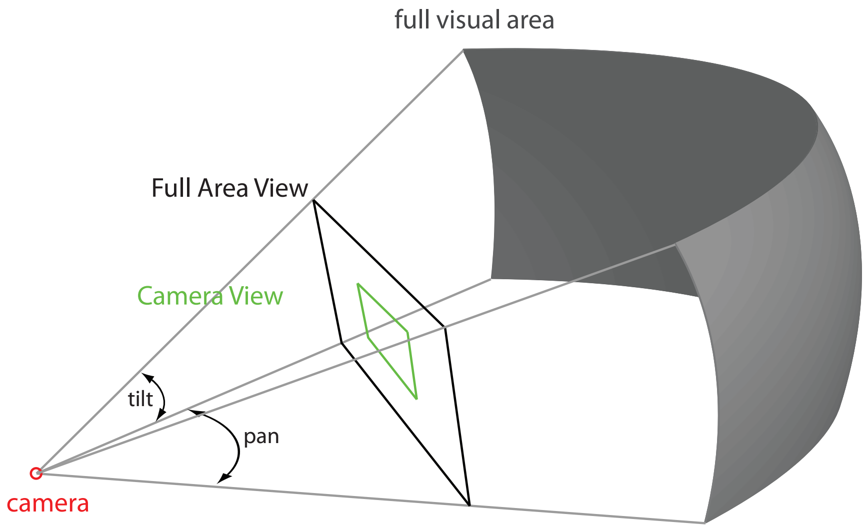

Figure 5.10: The construction of the Full Area View.

5.3.3 Visual Analysis

For the visual analysis of the captured images, techniques from "computer vision" are applied. Computer vision is a scientific discipline aimed at the development of artificial systems that can obtain relevant information from images. It is used for all kinds of applications, including robot vision. A lot of software tools have been developed providing useful functions for computer vision. Some of these tools have been evaluated for Probo:

RoboRealm [113] is a robotic vision software application for use in computer vision, image processing, and robot vision tasks. It has a very easy to use GUI providing the tools to build up your own complex image analysis using pipelines. The creators of RoboRealm aim at basic robotic vision issues such as navigation, rather than social interaction. Face detection is thus not included in the software yet.

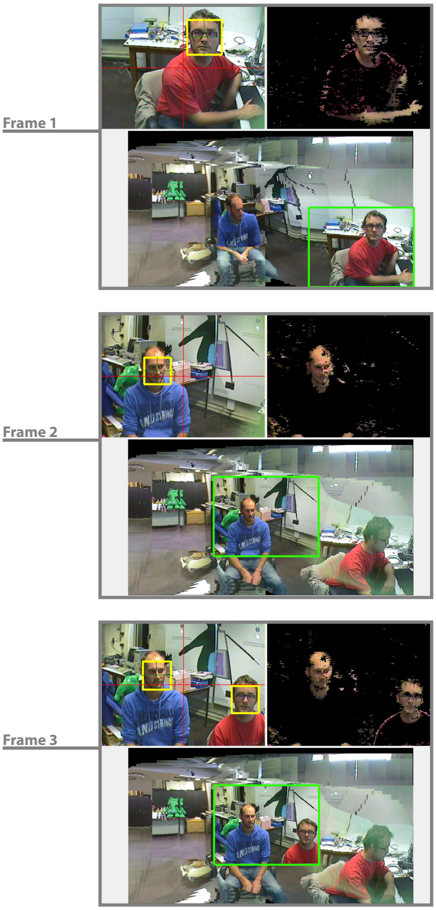

Figure 5.11: The GUI of the vision system at three consecutive moments in time.

Skilligent Robot Vision [125] is a software component which implements powerful object recognition and object tracking algorithms. The system is specifically designed for robotics applications including visual object recognition and tracking, image stabilization, visual-based servoing, human-to-machine interaction and visual localization. The vision software comes with an image database system, a searchable storage of visual information about objects, optimized for object identification and content-based image retrieval applications. The database stores visual information about physical rigid objects.

OpenCV [15] (Open Source Computer Vision) is a library of programming functions for computer vision, originally developed by Intel. It is free for commercial and research use under the open source BSD license. It is designed for computational efficiency and with a strong focus on realtime applications. Example applications of the OpenCV library are Human-Computer Interaction (HCI); Object Identification, Segmentation and Recognition; Face Recognition; Gesture Recognition; Motion Tracking, Ego Motion, Motion Understanding; Structure From Motion (SFM); Stereo and Multi-Camera Calibration and Depth Computation; Mobile Robotics.

AForge.NET [78] is a C# framework designed for developers and researchers in the fields of computer vision and AI - image processing, neural networks, genetic algorithms, machine learning, robotics, etc. The work on the framework's improvement is in constants progress, what means that new feature and namespaces are added regularly.

RoboRealm has been used for testing different filters in the beginning of the project, but it is not open source or freeware and does not provide software components. Skilligent looks very promising, but is also not open source or freeware. So for vision analysis in Probo we use components from AForge.NET and OpenCVDotNet (an OpenCV wrapper for the .NET framework). This allows us to use computer vision functionalities that we need, from within our managed software environment. The task of the visual analysis is to provide visual stimuli that can be used to control the gazing for Probo.

5.3.3.1 Color and SkinColor Detection

For the color and skin color detection a bitmap filter is used. This filter takes the bitmap or pixmap image, that can be seen as a spatially mapped array of pixels. The pixels are than converted to 24 bits per pixel (8 bits for each channel; Red, Green and Blue). After a specific color range is filtered, the remaining adjacent pixels are grouped into pixel blobs. The center of the largest blob will be selected as the point of attention. This is achieved with the help of the ExtractBiggestBlob filter provided within the AForge.NET framework. To detect the skin color over a wide range of lighting conditions, the possible RGB values for pixel are not skin color if they comply with Equation 5.1 and 5.2 (adapted from Kismet [20]). The result of the SkinColor filter is visible in the Filter View in Figures 5.4 and 5.10.

$$(R < (1.1 * G))||(R < (0.9 * B))||((R > (2.0) * G) \tag{5.1}$$

$$(R > (2.0) * B))||(R < 20)||(R > 250) \tag{5.2}$$

5.3.3.2 Face Detection

Face to face communication plays an important role in human to human communication. Thus it is natural to assume that an important quantity of non-verbal information can be obtained for human-robot interaction by watching faces [29]. A recognition process can be much more efficient if it is based on the detection of features that encode some information about the class to be detected. This is the case of Haar-like features that encode the existence of oriented contrasts between regions in the image. A set of these features can be used to encode the contrasts exhibited by a human face and their spacial relationships.

The object detector of OpenCV has been initially proposed by Viola [134] and improved by Lienhart [84]. A Haar classifier (namely a cascade of classifiers working with Haar-like features) is trained with a few hundreds of sample views of a particular object (i.e., a human face), called positive examples, and random images, called negative examples. After a classifier is trained, it can be applied to a region of interest in an input image.

OpenCVDotNet provides this Haar classifier as a software component named CVHaar. The classifier is loaded with trained settings for frontal face detection. The CVHaar component is used to obtain a list of rectangles containing the detected faces in an image. If more than one face is detected the face with the largest rectangle is selected. Then the center location of the rectangle is passed on to the Gazing unit using a Remote Object.

5.3.4 Remote Object

Remoting enables software components to interact across application domains. The components interacting with each other can be in different processes and systems using different transportation protocols, serialization formats, object lifetime schemes. Using object references to communicate between server objects and clients is the core of remoting. In the previous chapter event-based communication between the different components in the same application domain was described. Remoting allows us to use similar techniques by referencing to objects that are not residing in the same application domain (process) or even on the same computer. Figure 5.12 demonstrates how a remote object is used to pass the GazingPoint, originating from the visual attention system towards the Gazing component that will calculate the corresponding head and eye movements. In this way the computational load of the vision analysis can be excluded from the rest of the system. In the current implementation the vision analysis resides in a separate process using an Inter Process Communication (IPC) channel to communicate with the Control Center. If the necessary computational power for the vision would increase, the process can be run on a separate computer using a Transmission Control Protocol (TCP) channel to communicate over an ethernet network.

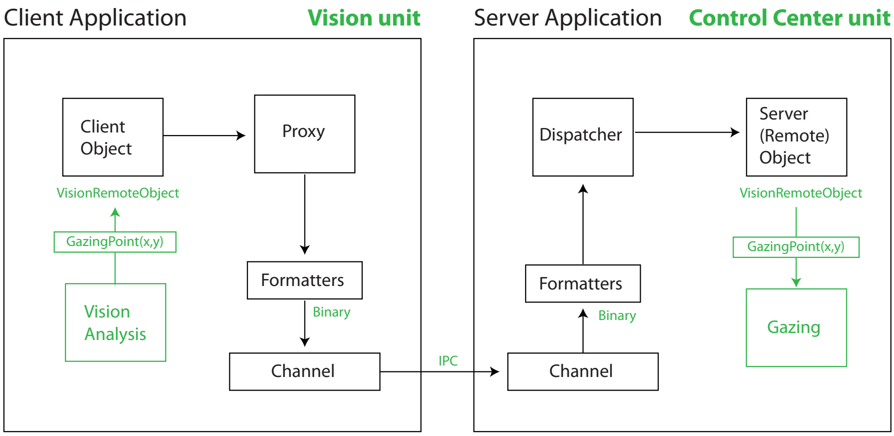

Figure 5.12: Remoting in .NET.

5.4 The Auditory System

Behaviors for social interaction may include to pay attention to a new sound source, to move toward it, or to keep face to face with a moving speaker [102]. Some sound-centered behaviors may be difficult to attain, because the mixture of sounds is not well treated. As part of our multidisciplinary team the DSSP research group is responsible for the auditory attention. Preliminary tests with commercial low-cost microphones showed that a professional microphone array is necessary to obtain descent sound localization. Current development of the auditory system focusses on:

Signal Activity Detection A signal activity detector detects when a signal is above background noise and thus determines which parts of the signal can be classified as background noise and which parts contain an active signal. In order to distinguish the active and non-active parts, the detector makes use of an estimation of certain characteristics of the ambient noise signal [38].

Audio Classification A silence/speech/music/other classifier is developed, where the first stage is a threshold based RMS silence detector. At the next stage the signal is passing through a speech/music/garbage classifier. Performance of this classifier was evaluated and described in [104]. Currently the classifier contains four categories which can be extended and adapted for different applications in the existence of enough data for training the classifier.

Sound Localization Sound localization and sound quality enhancement is obtained through the use of a microphone array. A microphone array is any number of microphones spatially distributed and operating in tandem. Two general techniques for sound direction estimation are taken into consideration [40]; Steered Response Power (SRP) or beam forming versus Time-Delay On Arrival (TDOA). Implementation of one of these systems requires further research.

Auditory Saliency Detection A bottom-up or saliency driven attention allows the brain to detect nonspecific conspicuous targets in cluttered scenes before fully processing and recognizing the targets [69]. In [119] it is suggested that there exists a unified framework for both visual and auditory sensory processing. Analogous to visual saliency maps, a saliency map for audition was proposed in [73]. This saliency map is currently being evaluated for future implementation.

These system can be used to assist the visual attention in determining the most interesting point or object that deserve Probo's focus of attention. For example if Probo is watching a conversation between two people, the attention will follow the turn taking between speaker and listener. Assuming that there are no other stimuli gaining interest over the conversation.

5.5 The Tactile System

Touch is the most developed sensory modality when we are born and it continues to play a fundamental role in communication throughout the first year of our life [50]. The study of the communicative functions of touch has received increasing attention over the past decades. But in contrast with the studies on vision and audio, the study of tactile communication has always been neglected [65]. Touch interaction is also slowly making its introduction into the robotic world. Most notable are the social robots Paro [136], Leonardo [18], the Huggable [127], Robovie-iv [93] and The Haptic Creature [145]. The sense of touch is rather unique: the skin is the largest organ in the human body. The human sensory system for vision and audio can obviously be imitated using cameras and microphones. But a straightforward solution that can imitate the human sense of touch is yet to be discovered. In our opinion, the most important attribute of human touch for social interaction is the measure of a force applied to a specific region. A simple technology that allows us to measure force is the Force Sense Resistor.

5.5.1 Force Sense Resistor

As their name implies, force sensing resistors (FSRs) use the electrical property of resistance to measure the force (or pressure) applied to a sensor. A force sensing resistor is made up of three parts: (Figure 5.13 shows this configuration)

An active area consisting of a pattern of conductors, which is connected to the leads on the tail to be charged with an electrical voltage.

A plastic spacer which includes an opening aligned with the active area, as well as an air vent through the tail.

A flexible substrate coated with a thick polymer conductive film, aligned with the active area.

When external force is applied to the sensor, the resistive element is deformed against the substrate. Air from the spacer opening is pushed through the air vent in the tail, and the conductive material on the substrate comes into contact with parts of the active area. The more of the active area that touches the conductive element, the lower the resistance. Over a wide range of forces, it turns out that the conductivity is approximately a linear function of force ($F \alpha C$,$F \alpha \frac{1}{R}$) as depicted in Figure 5.14.

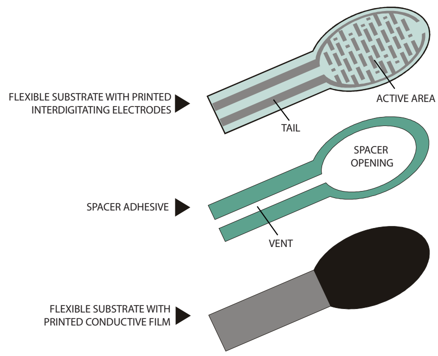

Figure 5.13: The FSR construction.

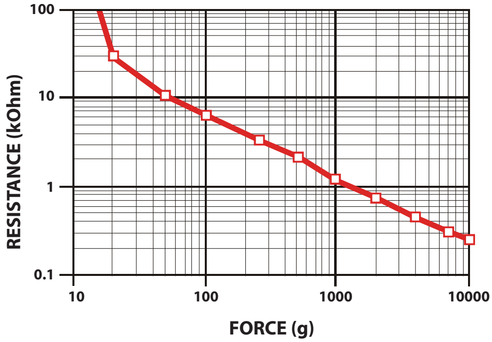

Figure 5.14: The FSR force-resistance characteristics.

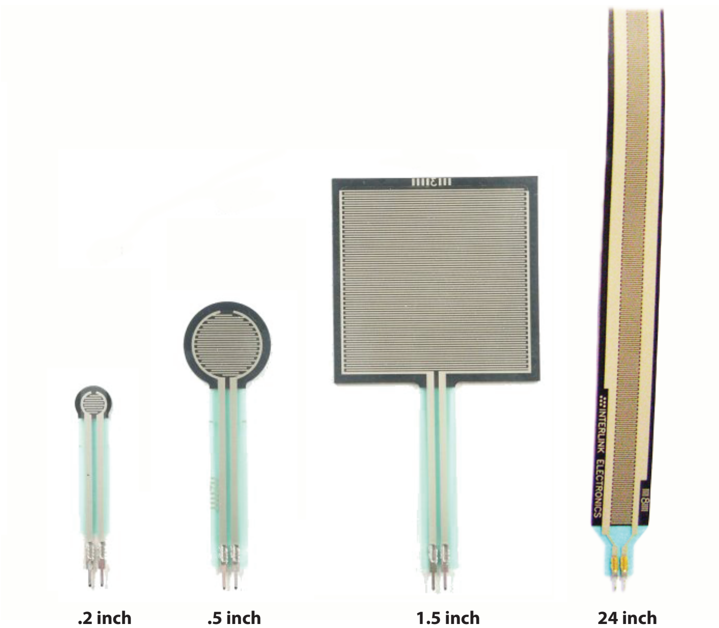

Figure 5.15: Different forms of FSRs.

For the implementation in Probo different forms of FSRs are used as depicted in Figure 5.15. Depending on the area of Probo's skin that needs to be sensitive, a different form of FSR is chosen. Each FSR is connected with a voltage divider to provide a signal (0-5V) for the Phidget Interface Kit (see Figure 5.16). This kit is connected with the computer via USB and comes with a DLL for .NET to read out the sensors. The Touch software unit uses eventHandlers to connect

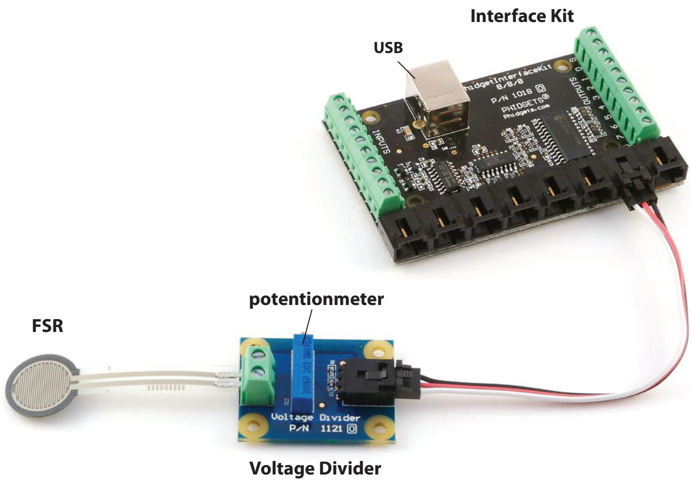

Figure 5.16: The use of a voltage divider to connect the FSR with the Phidget Interface Kit.

with the Phidget events. In case a Phidget's SensorChange event is fired, the eventHandler updates the force values that are used for the tactile analysis.

5.5.2 Tactile Analysis

In the current implementation the tactile analysis can only distinguish between normal, unpleasant and pleasant touch. The touch is considered pleasant when it is a gentle pulsating motion such as; rubbing, petting, scratching or massaging. If during the interaction the force gets too high (even if it is for a very short moment), the touch is considered as unpleasant such as; squeezing, pinching, hitting or poking. To analyze this affective touch, the detected force is considered as a function of time ($F(t)$). The classification is in different stages (See Figure 5.17), starting with the "pain threshold". If this threshold is exceeded, the touch interaction will be classified as unpleasant. An additional envelope detector (Envelope1), with a high raise but a low fall time, is used to cover the signal measured straight after a pain stimulus, so these forces can't be classified as pleasant touch. The higher the signal, the longer it takes the envelope to go under the pain threshold. If the force signal remains below the pain threshold the signal is further analyzed; first the signal is put through a high-pass filter to concentrate on the signal changes during the interaction, next an envelope detector (Envelope2) is used to obtain the intensity of the alternating force. If this result exceeds the "hug threshold", the touch interaction is classified as pleasant. If there is no unpleasant or no pleasant touch detected the touch interaction is classified as normal touch.

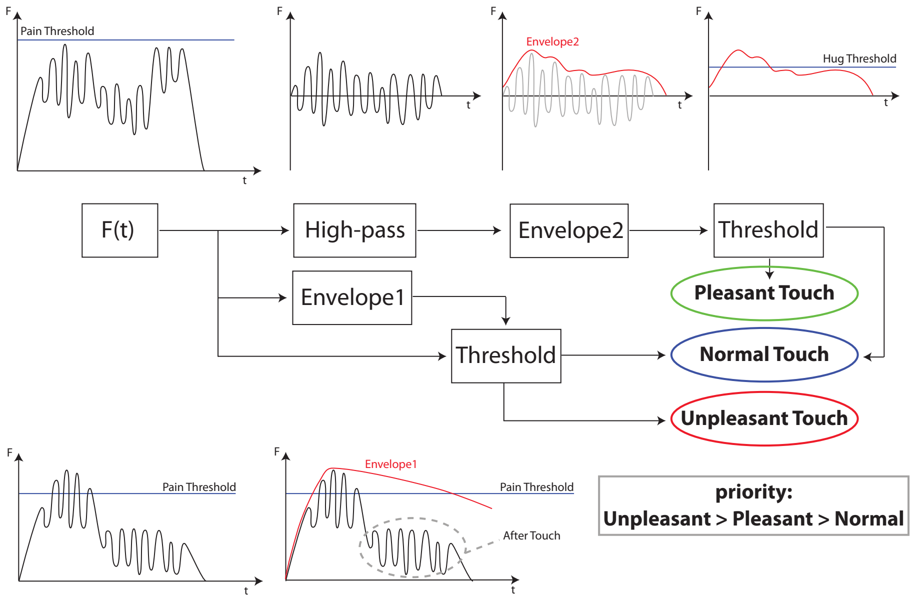

Figure 5.17: The stages for classification of affective touch.

The GUI of the Touch software unit presents the operator with a visual feedback of the touch interactions. Figure 5.18 shows the GUI during touch interaction; a pleasant touch (or hug) is detected and Probo's left arm and right ear are being touched. When a body part is touched the corresponding area on the 3D picture will light up. The intensity of the light represents the amount of force measured. The different body parts of Probo that can sense a touch interaction include: the left ear, the right ear, the top of the head, the trunk, the chest, the left arm and the right arm.

5.6 The Object Identification System

Object identification is most useful for applications such as asset tracking (e.g. libraries, animals), automated inventory and stock-keeping, toll collecting, and similar tasks where physical objects are involved and the gap between the physical and the "virtual" world must be bridged. In a world of ubiquitous computing, unobtrusive object identification enables the seamless connection between real-world artifacts and their virtual representations [135]. Identification of objects can facilitate social interaction, especially in the case of therapy,

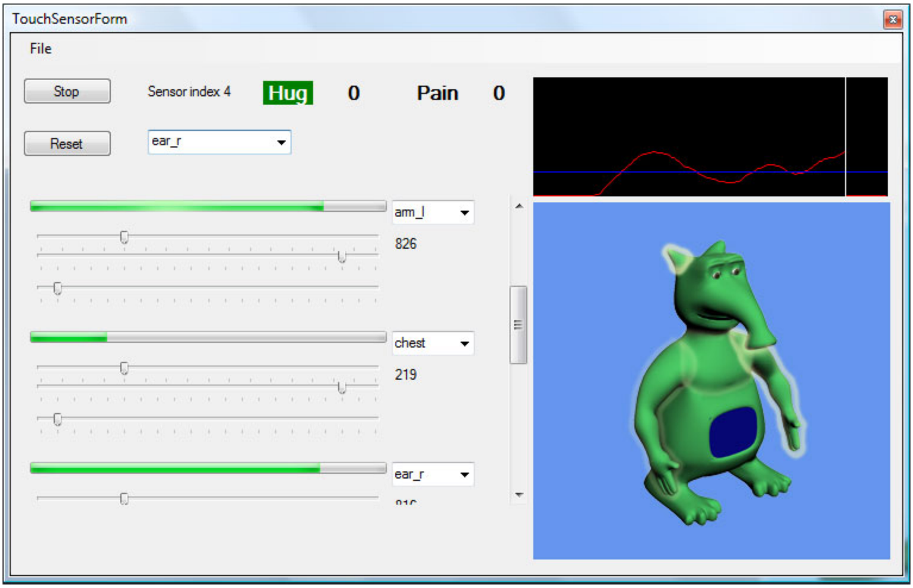

Figure 5.18: The GUI of the Touch software unit.

where interactive scenarios are applied.

5.6.1 Radio Frequency Identification

Radio-frequency identification (RFID) is always made up of two components. The *transponder* is located on the object to be identified. The detector or *reader* can be a read or write/read device, depending upon design and the technology used [51]. The transponder (typically referred to as an RFID tag) can be applied to or incorporated into a product, animal, or person for the purpose of identification and tracking using radio waves. Some tags can be read from several meters away and beyond the line of sight of the reader. There are generally three types of RFID tags:(1) active RFID tags, which contain a battery and can transmit signals autonomously,(2) passive RFID tags, which have no battery and require an external source to provoke signal transmission and,(3) battery assisted passive (BAP) RFID tags, which require an external source to wake up but have significant higher forward link capability providing great read range. The passive RFID tags are the most commonly used and are recommended in [135] for powerful object identification. The transponder (passive tag) is only activated when it is within the response range of a reader. The power required to activate the transponder is supplied to the transponder through the coupling unit (contactless) as is the timing pulse and data. The transponder and reader that are used for Probo are depicted in Figure 5.19.

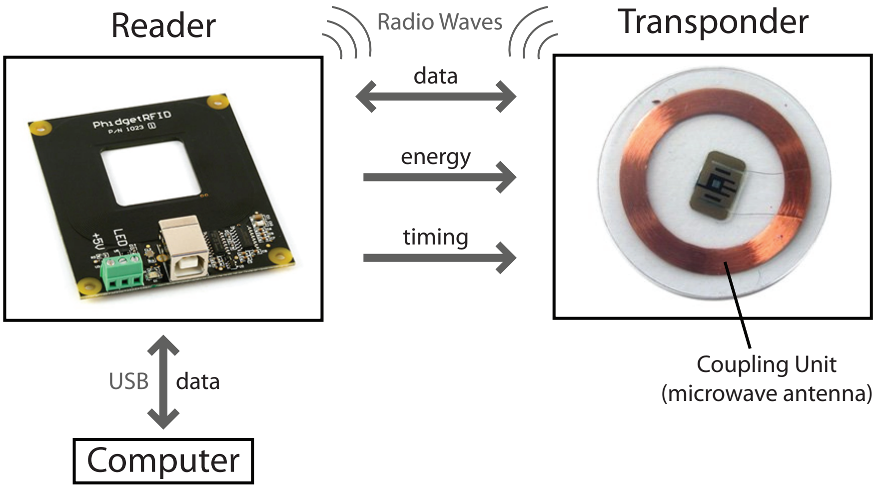

Figure 5.19: The RFID communication between computer, reader and transponder.

5.6.2 Implementation

Currently the RFID system is being evaluated in a stuffed model of Probo as an input device for the homeostatic system for the ProboGotchi game (see Section 6.7). The transponders are embedded in iconic cards (see Figure 5.20), representing four possible actions that can be performed on Probo. The RFID reader is placed inside the stuffed model at the height of Probo's mouth. The user is than explained to feed the iconic cards to Probo by bringing them in close proximity of its mouth. The Object ID software unit has the simple task to obtain the ID-number from the reader when a tag is detected and trigger the corresponding action from the homeostatic system. The Phidgets RFID reader can be accessed in the same way as the Interface Kit used for the touch analysis. When a tag is detected by the reader an event is fired. The corresponding EventHandler defined in the software unit will check the ID-number to set the action, and fire a new event in its turn. This event is further processed by the homeostatic system. RFID is becoming increasingly prevalent as the price of the technology decreases. The future of RFID has enormous potentials in identification of all objects and persons. Identification with RFID is very fast and asks barely any computational power. While vision is useful for the detection of objects (for example a human face, or a red ball), the identification

Figure 5.20: The iconic cards, with embedded RFID transponder, representing special actions.

of unique objects is more difficult and asks much more computational power. A combination of both technologies will be profitable in anyway.

5.7 Summary

This chapter presented the perceptual systems that are inspired on the biological human systems. A visual attention system is implemented and will generate a point of attention based on the operator's choice. This point will be used to control the gaze of our active vision system. While the eyes and head are directed to this point of interest, the active vision system builds an image of the full visual area. The modular design of the components allows future incorporation of auditory and tactile stimuli into a more advanced attention system. The tactile analysis uses FSRs to classify touch interactions into pleasant or unpleasant touch. In combination with the actions that are identified using embedded RFID tags, the tactile stimuli activate the homeostatic system. The next chapter describes how the homeostatic system uses these actions to generate an emotional state for Probo, and how this state is expressed towards the outside world.

Chapter 6

The Emotional and Homeostatic Systems

"The emotions aren't always immediately subject to reason, but they are always immediately subject to action"

-William James-

6.1 Introduction

In recent years, emotion has increasingly been used in interface and robot design primarily for the reason that humans tend to treat computers in the same way that they treat other humans. Nass et al. [99] conclude that individuals's interactions with computers are inherently natural and social. Since emotional communication is natural between people, we should interact more naturally with computers (or robots) that recognize and express affect [106]. Everyone expresses his emotions in its own way depending on his personality, age, culture, etc. Nonetheless there are some social clues that can be used to recognize the emotions. These social clues consist of perceptible indicators such as facial expression, voice intonation, gestures and movement, posture and pupil dilation. Based on the design drivers for social interaction (see Section 3.5.5), readable social cues for Probo are developed. In this chapter we present facial expressions (Section 6.2), a major modality to convey social cues. The emotion space provides a way to link the emotional state of the robot with a corresponding facial expression (Section 6.3). User-tests have been performed to define the recognition rate of some basic emotions in comparison with other similar social robots (Section 6.4). Next the homeostatic system (Section 6.6) is presented, an autonomous system that generates an emotional state based on internal needs and perceived actions (triggered from the perceptual system). An overview of all systems is presented in Figure 6.1. A fully autonomous 3D game, named ProboGotchi (Section 6.7), is developed to test these systems during social interaction with children.

6.2 Facial Expressions

In the daily life, people rely on face-to-face communication. Following on the arguments presented in Section 3.3.1; facial expression is the most important modality in human face-to-face communication. To support facial expressions with a robot, a system for coordinating the motions of each DOF needs to be developed. Most of the motions in a human face are based on the Action Units (AU) defined by the Facial Action Coding System (FACS) developed by Ekman and Friesen [46]. AU express a motion of mimic muscles as 44 kinds of basic operation, with 14 AU to express the emotions of anger, disgust, fear, joy, sadness, and surprise. Which are often supported as being the 6 basic emotions from evolutionary, developmental, and cross-cultural studies [44]. The AU are used to define the motions for Probo's DOF. Because Probo does not have a human face and for simplifying the design, some of the AU are missing, others are replaced and some are added. Table 6.1 shows the AU in Probo that are used for facial expressions. The lack of the lower eyelid and a fixed upper lip lead to missing AU, the AU regarding the nose movements (AU 9 and 11) will

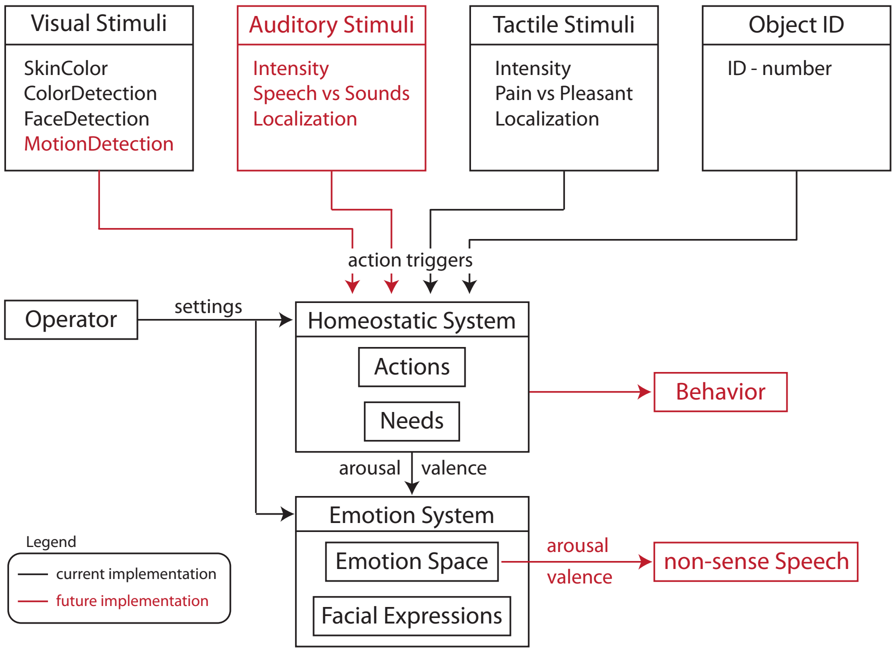

Figure 6.1: The architecture of the Perceptual system in collaboration with the Homeostatic and Emotion system.

be replaced by the movement of the 3 DOF trunk. Table 6.2 shows the AU that are triggered for each of the basic emotions. The movement of the ears and the greater visual influence of the trunk will add extra gestures to express the emotions. In Table 6.3, the DOFs in Probo's head are compared with other prominent robot heads like: Kismet [20], iCat [132] and Eddie [126]. These robotic heads make use of eyes, eyelids, eyebrows and a mouth to conform with the AU.

6.3 Emotion Space

To realize a translation from emotions into facial expressions, emotions need to be parameterized. Fong et al. [54] present three primary theories to describe emotions. The first approach describes emotions in terms of discrete categories or as basic emotions [44]. The second approach characterizes emotions using continuous scales or basis dimensions, such as arousal and valence [118], [109]. The third approach, componential theory, acknowledges the importance of both categories and dimensions [108]. In the robot Kismet [20], facial expressions

Table 6.1: A description of the Action Units used for facial expressions with Probo.

AU	**Definition**	**Involved muscles**
AU 1	Inner Brow Raiser	Frontalis (pars medialis)
AU 2	Outer Brow Raiser	Frontalis (pars medialis)
AU 4	Brow Lowerer	Corrugator supercilii, Depressor supercilii
AU 5	Upper Lid Raiser	Levator palpebrae superioris
AU 9	Nose Wrinkler	Levator labii superioris alaeque nasi
AU 11	Nasolabial Deepener	Zygomaticus minor
AU 12	Lip Corner Pull	Zygomaticus major
AU 15	Lip Corner Depressor	Depressor anguli oris (also known as Triangularis)
AU 17	Chin Raise	Mentalis
AU 26	Jaw Drop	Masseter, relaxed Temporalis and internal pterygoid
AU 43	Eyes Closed	Relaxation of Levator palpebrae superioris;
		Orbicularis oculi (pars palpebralis)
AU 45	Blink	Relaxation of Levator palpebrae superioris;
		Orbicularis oculi (pars palpebralis)
AU 46	Wink	Relaxation of Levator palpebrae superioris;
		Orbicularis oculi (pars palpebralis)

Table 6.2: The relationship between emotions and the AUs.

Emotion	Ekman and Friesen (2002)	Probo
anger	4, 4+7+17+23	4
disgust	9, 4+6+9+10+17+22	9, 4+9+17
fear	1+5+25/26	1+5+26
joy	12/13, 6+11+12/13	12, 11+12
sadness	1+4, 1+4+15/17	1+4, 1+4+15/17
surprise	1+2, 1+2+5	1+2, 1+2+5

Table 6.3: DOF, ranges and AU of the actuated joints of Probo's head in comparison with other prominent non-humanoid robot heads.

Kismet	Eddie	iCat	Probo			
(DOF)				Range [°]		AU
Eyes (3)	Eyes (3)	Eyes (3)	Eyes (3)	Pan Tilt	100 80	
Eyelids (2)	Eyelids (4)	Eyelids (2)	Eyelids (2)		150	5,43,45,46
Brows (4)	Brows (4)	Brows (2)	Brows (4)		45	1,2,4
Ears (4)	Ears (4)		Ears (2)		90	
Yaw (1)	Yaw (1)		Mouth (3)	Yaw	45	17,26
Lips (4)	Lips (4)	Lips (4)		Lipcorners	60	12,15
	Crown (1)		Trunk (3)		360	9,11

are generated using an interpolation-based technique over a three-dimensional, componential *affect space* (arousal, valence, and stance).

In our model two dimensions; valence and arousal are used to construct an emotion space, based on the circumplex model of affect defined by Russell [109], which has as well been implemented in the robot Eddie [126]. This emotion space will provide a continuous range of emotions, including the basic emotions defined by Ekman [44]. In the emotion space a Cartesian coordinate system is used, where the x-coordinate represents the valence and the y-coordinate the arousal, consequently each emotion $e(v, a)$ corresponds to a point in the valence-arousal plane (Figure 6.2). In this way the basic emotions can be specified on a unit circle, placing the neutral emotion $e(0, 0)$ in the origin of the coordinate system. Now each emotion can also be represented as a vector with the origin of the coordinate system as initial point and the corresponding valence-arousal values as the terminal point. The direction α of each vector defines the specific emotion whereas the magnitude defines the intensity of the emotion. The intensity i can vary from 0 to 1, interpolating the existing emotion $i = 1$ with the neutral emotion $i = 0$. Each DOF that influences the facial expression is related to the current angle α of the emotion vector. An adjustable interface is developed to define the specific value for each angle $(0° - 360°)$ of each DOF. When selecting one DOF, a value for each basic emotion is set on the unit circle. To attain a contiguous relation, a linear interpolation between the configuration points is applied.

By adding more (optional) points or values the curve can be tuned to achieve smooth, natural transitions between the different emotions. An example is shown (Figure 6.3) for the DOF controlling the eyelid, extra points were added in the first half of the emotion space to achieve smoother transitions. The graph is respectively starting and ending with $\alpha = 0° = 360°$.

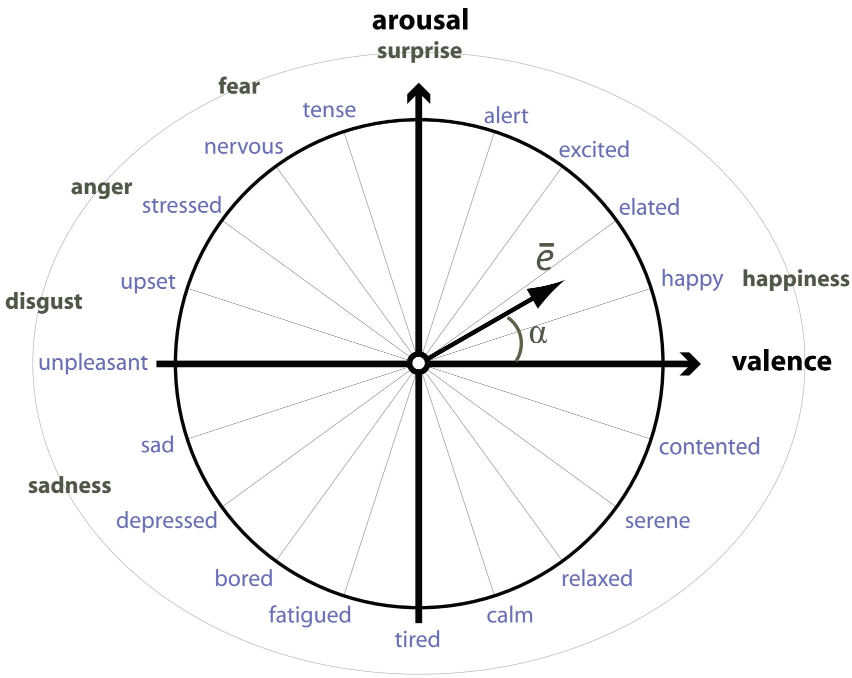

Figure 6.2: Emotion space based on the circomplex model of affect defined by Russell [109].

6.3.1 The Emotion Space GUI

The GUI of the EmotionSpace software unit (Figure 6.4) offers the operator an emotional interface. The left side displays the emotion vector in the emotion space. The emotion vector can be moved in real-time by dragging it over the emotion space. In the middle a list of all the DOF is presented, each in a different color. After selecting a DOF, its corresponding line will become solid in the graph on the right. This active line can now be adjusted by altering the points depicted as little rectangles. After adjustments have been made the emotion mapping for all the DOFs can be saved in a file. In this way different emotional mappings can be created and loaded when needed.

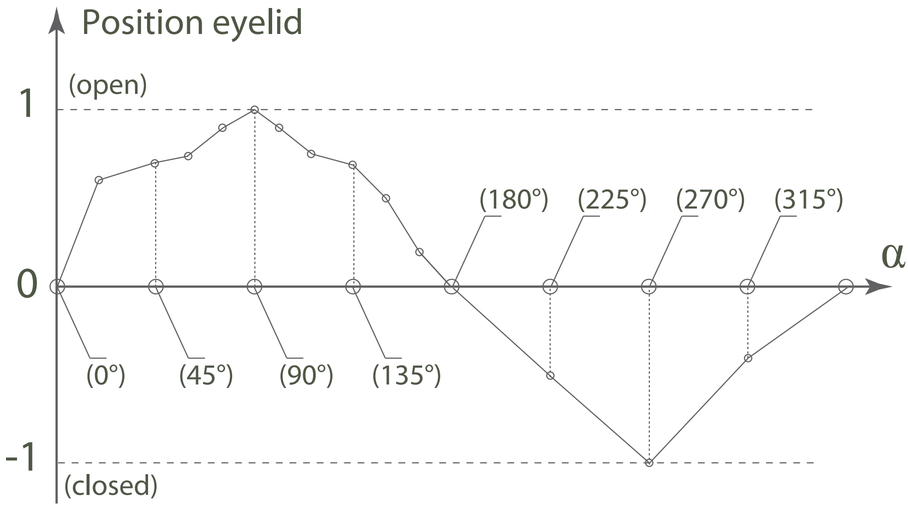

Figure 6.3: Adjustable interface for defining the value off the DOF (controlling the position of the eyelid) for each emotion (angle α).

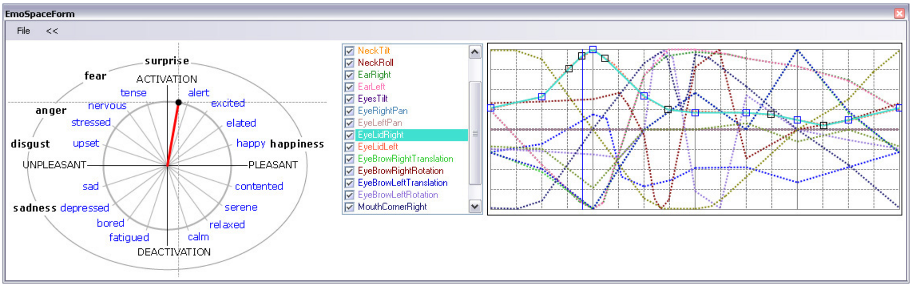

Figure 6.4: The GUI of the EmotionSpace software unit.

6.4 Tests and Evaluation

6.4.1 Recognition tests

To test the recognition of facial expression the virtual model was used in a first user-study. The study was based on a survey performed by Breazeal evaluating the expressive behavior of Kismet [20]. The subjects were asked to perform a comparison task where they compared 8 color images of the virtual model (Figure 6.5) with a series of line drawings of human expressions (Figure 6.6). A second study consists of a multiple-choice-test, in which people were asked to match the 8 color images with 8 given answers (words), identifying each emotion. A paper version of the first test has been filled out by children and adults. For the second study a paper version was used to test the children and an electronic version was used to test the adults. The electronic version was an online survey giving the possibility to test more people.

To measure a certain level of agreement between the users the Fleiss' Kappa [53] value was calculated. This value is a statistical measure of inter-rater reliability that can be calculated on and is given by Equation 6.1 with $\bar{P}$ (Equation 6.2) and $\bar{P}_e$ (Equation 6.3). The factor $1 - \bar{P}_e$ gives the degree of agreement that is attainable above chance, and, $\bar{P} - \bar{P}_e$ gives the degree of agreement actually achieved above chance. The scoring range is between 0 (complete random) and 1 (complete agreement). In our studies: $i = 1, ...N$ represents the participants, n is the number of pictures of Probo (with n_{ij} the number of ratings per picture) and $j = 1, ..k$ represents the possible answers (words or drawings). An interpretation of the κ values has been suggested by Landis and Koch [79], and is presented in Table 6.4. This table is however not universally accepted, and can only be used as an indication.

$$\kappa = \frac{\bar{P} - \bar{P}_e}{1 - \bar{P}_e} \tag{6.1}$$

$$\bar{P} = \frac{1}{Nn(n-1)}(\sum_{i=1}^{N}\sum_{j=1}^{k} n_{ij}^2 - Nn) \tag{6.2}$$

$$\bar{P}_e = \sum_{j=1}^{k} p_j^2 \tag{6.3}$$

Twenty-five subjects (6 - 8 years of age) filled out the questionnaire for the first test. The children were presented an image of our virtual model representing one of the 8 emotions. For each of those images they had to choose the best matching sketch representing human emotions. The results are shown in Table 6.5. During the test, the observation was made that the children were really seeking for a visual resemblance without recognizing the underlying emotions. The same test is performed on sixteen adult people (20 - 35 years of age), the results are shown in Table 6.6. The correct answers were similar with the children's test, with the exception of *surprise*, giving an overall match of 67% for the adults to 60% for the children. Where the children had difficulties identifying the emotion of *surprise* most of the adults (81%) had a positive match.

Table 6.4: Guidelines for strength of agreement indicated with κ values (values from Landis and Koch [79]).

κ	**Interpretation**
< 0	Poor agreement
0.0 - 0.20	Slight agreement
0.21 - 0.40	Fair agreement
0.41 - 0.60	Moderate agreement
0.61 - 0.80	Substantial agreement
0.81 - 1.00	Almost perfect agreement

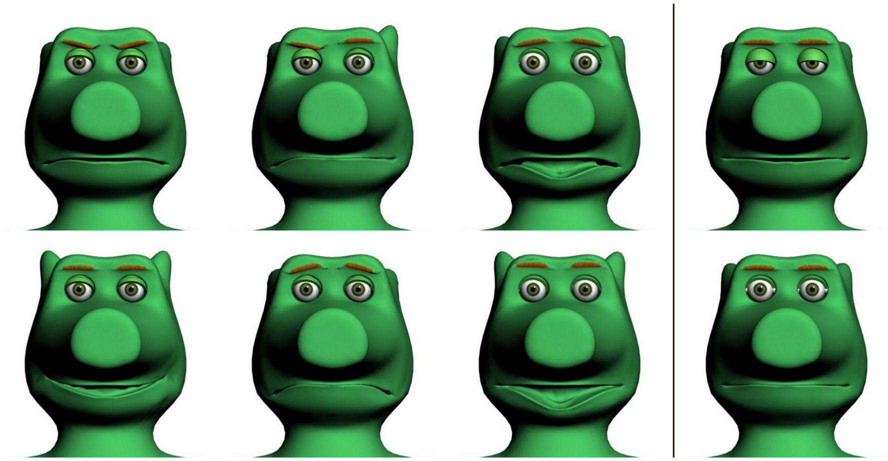

Figure 6.5: Facial expressions of the virtual model used in a user-study. The 6 basic emotions (anger, disgust, fear, happy, sad and surprise) on the left and the emotions tired and neutral on the right.

The hypothesis was made that some of the adults, first try to recognize the underlying emotions rather than just look for a graphical similarity, resulting in better matches. In the next tests the sketches are excluded and replaced by words to force the subjects to define the underlying emotions.

In the second user-study 20 children (age 8-12) were asked to identify the emotion expressed by the virtual model using a multiple-choice questionnaire. In this way the children are obliged to look for the underlying emotion represented by the facial expression. The results of this study are shown in Table 6.7. These result are significant better (overall 78% against 60%) than with the first test. In this user-study $\kappa = 0,56$ which means there is a moderate agreement between the answers that the children gave. The multiple-choice questionnaire is repeated in an online survey, where 143 adults (age 18-60) completed the

Table 6.5: The result of the comparison test with children shown in percentage match.

% match	happy	sad	disgust	anger	fear	tired	surprise	neutral
happy	**54**	0	7	0	0	0	18	0
sad	0	**74**	9	7	15	2	0	0
disgust	0	4	**62**	4	3	0	0	4
mad	1	2	2	**66**	3	9	0	16
fear	0	0	0	0	**48**	0	29	0
tired	0	4	5	2	0	**87**	3	4
surprise	0	0	0	0	9	0	**28**	0
sly grin	5	0	2	11	5	0	0	0
stern	0	12	9	0	2	0	0	40
anger	2	0	0	3	0	0	7	4
repulsion	2	4	0	7	3	0	0	0
pleased	36	0	4	0	12	2	15	32

overall %	60	Fleiss' Kappa	0.46

Table 6.6: The result of the comparison test with adults shown in percentage match.

% match	happy	sad	disgust	anger	fear	tired	surprise	neutral
happy	**56**	0	0	0	6	0	13	0
sad	0	**88**	0	0	44	13	0	6
disgust	0	6	**63**	0	0	0	0	0
mad	0	0	6	**69**	0	0	0	6
fear	0	0	0	0	**44**	0	0	6
tired	0	0	6	6	0	**81**	0	44
surprise	0	0	0	0	0	0	**81**	6
sly grin	19	0	6	0	0	0	0	0
stern	0	6	19	19	6	0	0	19
anger	0	0	0	6	0	0	0	0
repulsion	0	0	0	0	0	0	0	0
pleased	25	0	0	0	0	6	6	13

overall %	67	Fleiss' Kappa	0.46

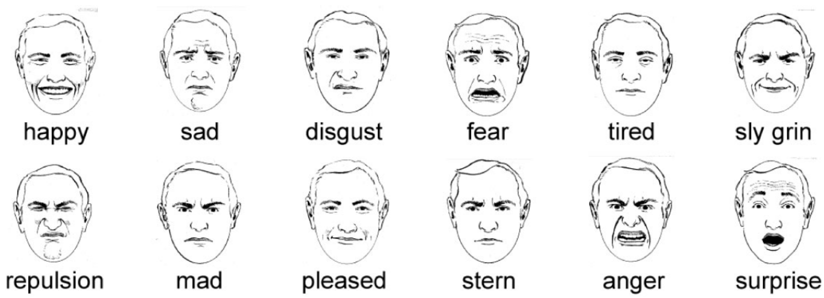

Figure 6.6: The sketches used in the evaluation, copied from Kismet's survey, adapted from (Faigin 1990) [20].

Table 6.7: The result of the multiple-choice questionnaire with children shown in percentage match.

% match	happy	sad	disgust	anger	fear	tired	surprise	neutral
happy	**95**	0	0	0	0	5	10	0
sad	0	**85**	5	0	20	0	0	0
disgust	0	0	**50**	5	15	0	0	10
anger	0	0	15	**90**	0	0	0	0
fear	0	5	0	5	**65**	0	0	10
tired	0	0	0	0	0	**70**	0	5
surprise	5	10	10	0	0	0	**85**	5
neutral	0	0	20	0	0	25	5	**70**

overall %	78	Fleiss' Kappa	0.56

questionnaire. The results are shown in Table 6.8 and are slightly less (5%) then the test with children. In both tests the recognition of the emotional states *disgust* and *fear* is the lowest.

6.4.2 Comparison Recognition tests

To compare with other similar robot projects, only 6 basic facial expression were included in the last 3 tests. The trunk was used to emphasize the expressions, based on the AU of the human nostrils. In the first test, pictures of the virtual model (Figure 6.7) were used in a user study with 143 children. This resulted in a overall identification rate of 88% with a substantial agreement ($\kappa = 0.75$) between the children. With the same settings pictures of the prototype of the robot were taken and tested. The results of the first test with the uncovered robot (Figure 6.8, are presented in Table 6.10). Next the results

Table 6.8: The result of the online multiple-choice questionnaire with adults shown in percentage match.

% match	happy	sad	disgust	anger	fear	tired	surprise	neutral
happy	**89**	1	1	0	0	0	9	0
sad	0	**81**	1	2	19	1	0	2
disgust	1	4	**43**	7	17	0	2	3
anger	0	0	34	**88**	0	1	0	0
fear	0	12	3	1	**51**	0	0	10
tired	1	1	2	1	1	**80**	0	1
surprise	1	0	11	0	11	0	**86**	14
neutral	8	1	4	1	1	18	3	**70**

overall %	73	Fleiss' Kappa	0.57

Table 6.9: The result of recognition test with the pictures of the virtual model (Figure 6.7).

%match	happy	sad	disgust	anger	surprise	fear
happy	**92**	0	0	0	7	0
sad	0	**97**	2	0	0.5	4
disgust	1	2	**81**	2	5	8
anger	0	0	1	**97**	0.5	1
surprise	7	0	12	0	**82**	5
fear	0	1	4	1	5	**82**

overall %	88	Fleiss' Kappa	0.75

with the fur covered robot (Figure 6.9 are presented Table 6.11). For each of the tests 23 different children were participating. Both tests follow the positive outcome of the test with the virtual model, giving an identification rate of respectively 83% and 84%.

Evaluations for the projects EDDIE [126], Kismet [20], Aryan [96] and Feelix [25] are compared in Table 6.12. These results show that the recognition of the emotional states of Probo are significantly higher than the others. In all the projects the recognition of fear has the lowest score.

A recent study [48] on the deficits in facial expression recognition, includes re-

Table 6.10: The result of recognition test with the pictures of the uncovered robot Probo (Figure 6.8).

%match	happy	sad	disgust	anger	surprise	fear
happy	**79**	0	0	0	21	0
sad	0	**96**	9	0	0	0
disgust	0	4	**88**	0	0	4
anger	0	0	0	**100**	0	0
surprise	18	0	0	0	**69**	26
fear	4	0	4	0	9	**70**

overall %	83	Fleiss' Kappa	0.67

Table 6.11: The result of recognition test with the pictures of the fur covered robot Probo (Figure 6.9).

%match	happy	sad	disgust	anger	surprise	fear
happy	**100**	0	0	0	0	0
sad	0	**87**	0	0	0	9
disgust	0	0	**87**	4	4	4
anger	0	9	4	**96**	0	0
surprise	0	0	9	0	**70**	22
fear	0	4	0	0	26	**65**

overall %	84	Fleiss' Kappa	0.68

Table 6.12: Emotion recognition rate of different robot faces.

	Probo	Kismet	Eddie	Aryan	Feelix
happy	100	82	58	-	60
sad	87	82	58	-	70
disgust	87	71	58	-	-
anger	96	76	54	94	40
surprise	70	82	75	71	37
fear	65	47	42	41	16
overall %	84	73	57	69	45

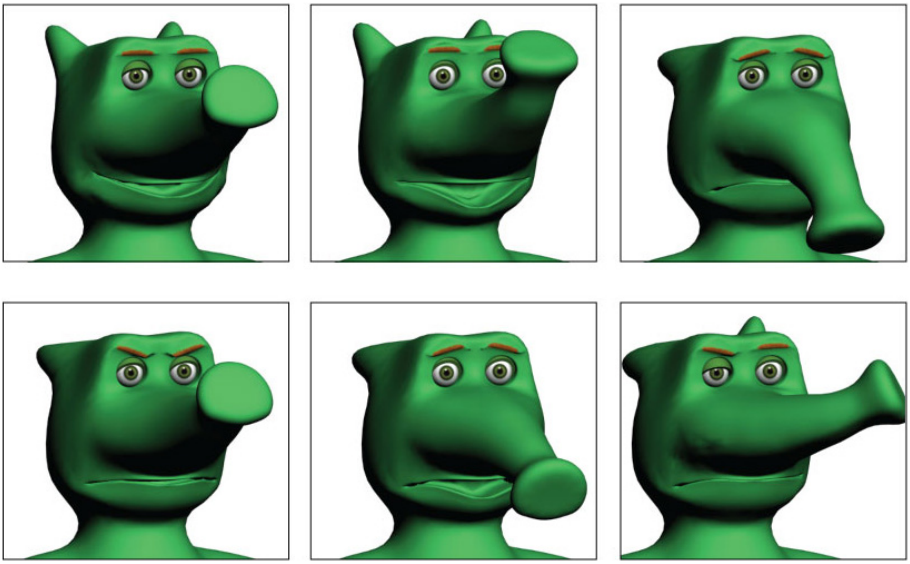

Figure 6.7: The 6 basic facial expressions of the virtual model used in big user-study (happy, surprise, sad, anger, fear and disgust).

Figure 6.8: The 6 basic facial expressions of the uncovered prototype used in a user-study (happy, surprise, sad, anger, fear and disgust).

Figure 6.9: The 6 basic facial expressions of the covered prototype used in a user-study (happy, surprise, sad, anger, fear and disgust).

sults for recognition of human facial expressions posed by model JJ from Ekman and Friesen [45]. These facial expressions are morphed (blended) across continua that span the following six expression pairs: happiness-surprise, surprise-fear, fear-sadness, sadness-disgust, disgust-anger, and anger-happiness (Figure 6.10). The results for healthy participants are presented in Figure 6.11. An estimation of the overall recognition rate is 84%.

6.4.3 Conclusions

From the first tests, we can conclude that it is better to use a list of words rather than a list of drawings to measure the recognition of emotions in facial expressions. Using words forces the user to look for the underlying emotion from the displayed facial expression. For the next tests we concluded that after incorporating the trunk movement in the facial expressions an overall recognition rate of 88% is achieved for the virtual model. Connecting the prototype resulted in a recognition rate of 84% for the physical robot. In comparison with other similar projects the recognition rate for Probo is higher. The recognition of emotions in facial expressions is based on social cues. Therefore it is important to emphasize the facial features that play an important role regarding the expression of emotions, such as the eyebrows, eyelids and mouth. Additional features such as ears and trunk increase the recognition rate. The

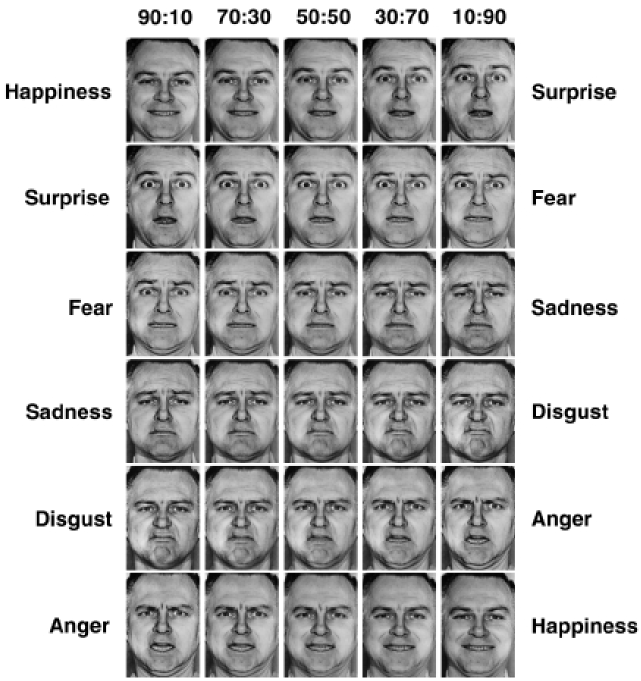

Figure 6.10: Facial expression continua posed by human model JJ (adapted from [48]).

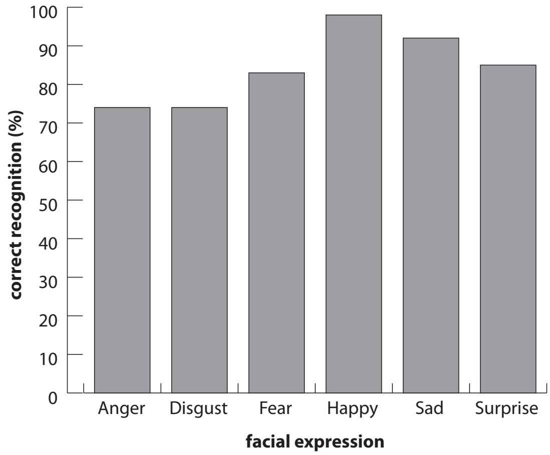

Figure 6.11: Accuracy of facial expression recognition posed by human model JJ (modified from [48]).

other projects mentioned in Table 6.12 did not cover their hardware, resulting in a mechanical look and the absence of a skin. This makes it more difficult to extract the facial features and explains the lower scores. In all the robotic projects the recognition of fear in facial expressions tends to be the most difficult. In comparison with recognition of human facial expressions (based on the results presented in [48]), the overall recognition rate is equal.

6.5 Nonsense affective Speech

A nonsense affective speech is currently under development by the DSSP research group of the VUB. This nonsense affective speech, which will be a cross-cultural and understandable language for most people regardless of their own native language. In one of the current approaches [144], this speech is produced by using a database with natural expressive speech samples and a database with neutrally spoken speech samples, both recorded with a professional speaker. From the neutral speech examples, carrier sentences of the non-existing language for Probo will be produced by firstly segmenting the recorded utterances into several nonsense syllables and then concatenating them in the same syllabic structure as the desired emotional prosodic template, which is selected from the expressive database. To produce emotional speech for that non-existing language, the same pitch and timing structure as found in the prosodic template are copied on the nonsense carrier phrase, a process that is known as prosodic transplantation and that effectively provides the synthetic output with a same intonation pattern as the natural expressive example. Later this affective speech can be linked with the emotional state of Probo, using the arousal and valence dimensions to control the speech. In this way the nonsense affective speech will contribute to the expression of emotions.

6.6 Homeostatic System

In psychology, much research is conducted on human needs and various theories are proposed such as the “Maslow’s hierarchy of needs” [67] and the classification of fundamental human needs by Manfred Max-Neef [89]. Some of these theories are used as an inspiration to define needs for robots (for example AIBO and Kismet) in order to have an artificial homeostasis regulation [2]. To provide a configurable system that makes Probo react emotionally, a model based on actions and internal needs is developed. The idea is that a user can perform different actions on Probo and that Probo will react accordingly by changing his facial expression. This model gives an operator the ability to configure the robot’s homeostasis by creating the desired actions and internal needs, determining the emotional state of the robot in real-time.

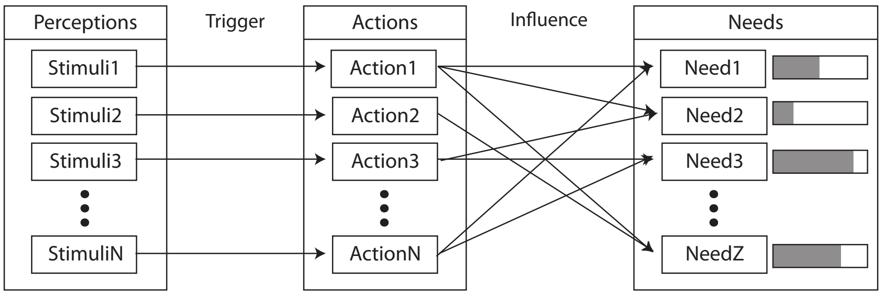

Figure 6.12: The actions and needs of the homeostatic system.

6.6.1 Actions

The software Actions can be considered as certain stimuli that are detected by the perceptual system. These stimuli represent actions towards the robot coming from users during social interaction. The operator can define the actions and how they are triggered. In the current implementation a set of actions has been created to comply with a children's game. Each action has a name and a set of influence factors. The influence factor will change the value of the corresponding need each time the action is triggered. An example of this configuration is depicted in Figure 6.12. Each action is linked with a sound effect, to provide auditory feedback as a reaction to a triggered action. A "hurt probo" action is for example triggered by a pain stimulus originating from the touch analysis, providing an auditory "Aaauuw!!" as feedback to the user.

6.6.2 Needs

Every internal need is visually represented as a progress bar holding the value of the need at a certain moment in time. The needs can be seen as the internal regulators of Probo's emotional state. They are continuously changing based and the "internal clock" and the influence from the triggered actions. This internal clock is a timer that triggers a "do nothing action" influencing the needs accordingly. For example; if nothing happens the need for affection grows slowly with each clock cycle. The continuously changing value of each need is mapped and multiplied with a weight factor (N_weight) to contribute to the total arousal and valence, as depicted in Figure 6.13. By adding all the valence and arousal contributions of all the needs, the total valence and arousal value is obtained (given by equation 6.4 and 6.5), defining the emotion vector in the emotion space.

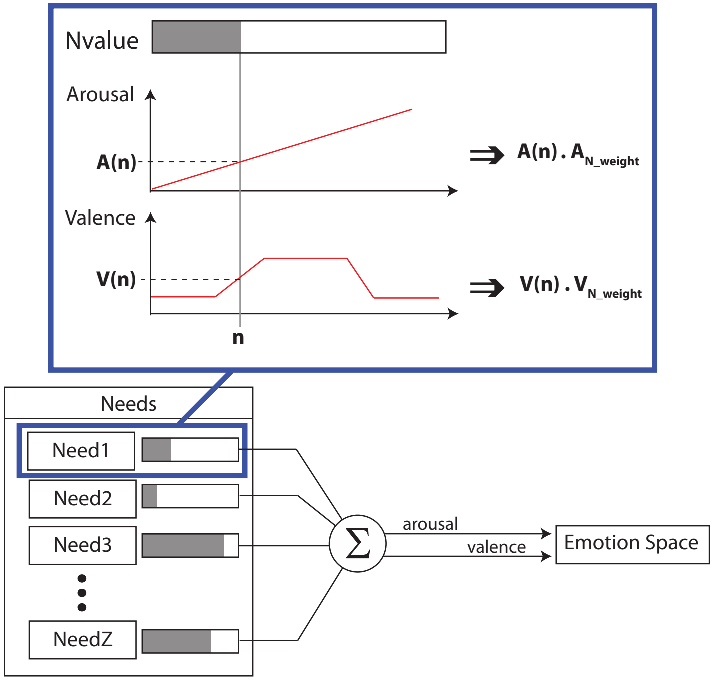

Figure 6.13: The calculation of the valence and arousal based on the different needs.

$$Arousal_{total} = \frac{\sum(A(n).A_{Nweight})}{\sum A_{Nweight}} \tag{6.4}$$

$$Valence_{total} = \frac{\sum(V(n).V_{Nweight})}{\sum V_{Nweight}} \tag{6.5}$$

6.6.3 The Homeostatic GUI

The Homeostatic software unit contains all the actions, the needs and their relationships. This unit provides access to these components through its GUI. The operator is presented with two modes:

Configuration gives the operator the possibility to configure the homeostatic system using the *Settings* panel (depicted in Figure 6.14). These settings include; adding actions and their influence on the needs, adding needs

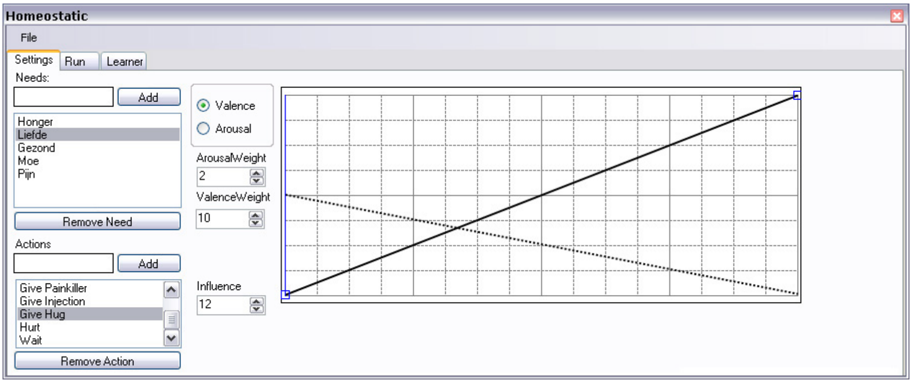

Figure 6.14: The Settings panel of the homeostatic GUI.

with their weights and function maps for valence and arousal. The function maps can be altered in the same way as with the facial expressions (see Section 6.3.1). All settings (list of actions, needs and maps) can be saved in a file, to be loaded when needed.

Operation gives the operator the possibility to monitor and interact with the homeostatic system using the *Run* panel (depicted in Figure 6.15). The Run panel is generated dynamically based on the defined actions and needs. Progress bars on the left show the status of the different needs. Buttons in the middle can be used to manually trigger the actions. The progress bars on the right present the resulting arousal and valence that can be send to the emotion space. The internal clock's interval (or "do nothing action") can be set in milliseconds.

The tab for a third panel named *Learner* can be noticed in both figures. This panel is used to control a Reinforcement Learner (RL), which is currently used for testing purposes. The RL is able to learn an optimal strategy for taking actions, by experiencing them without the knowledge of their semantics. The reinforcement comes from a reward function that gives a higher reward as the emotional state reaches complete happiness. This RL has been implemented to compete with children in the ProboGotchi game (see Section 6.7).

An overview of the total data flow is depicted in Figure 6.16. The data flow starts with the stimuli that are originating from the perceptual system after analysis of the sensors. When these stimuli are detected they will trigger the corresponding action. This action will subsequently influence the connected needs. Next the value of the needs is mapped to arousal and valence dimensions. The arousal and valence define the emotion vector, used to represent an emotion in the emotion space. This emotional vector is then mapped to the corresponding facial expression. This facial expression consists of a list of

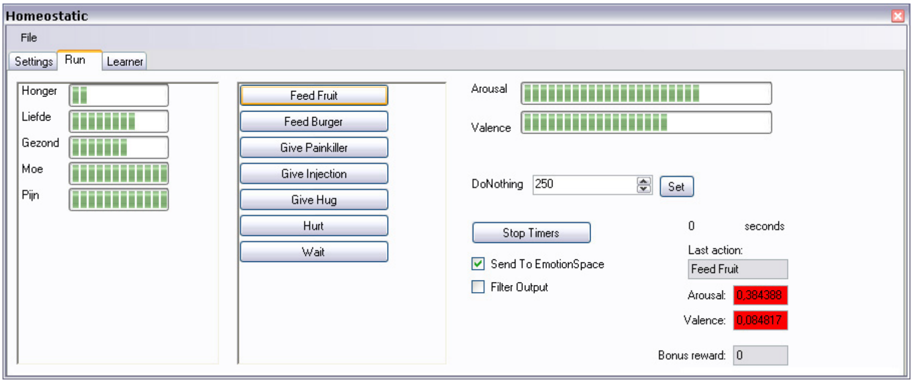

Figure 6.15: The Run panel of the homeostatic GUI.

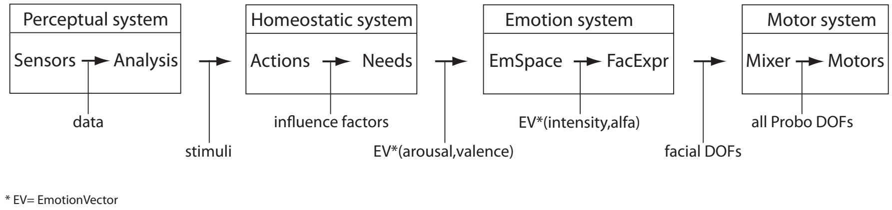

Figure 6.16: The dataflow starting from sensory input data to facial expressions as motor output.

positions for the facial DOFs. These DOFs are then finally transmitted to the motor system.

In comparison with the systems used for AIBO and Kismet, this system is not predefined to simulate a certain animal or infant-like behavior. This system provides flexibility and can be easily configured and connected to obtain a homeostasis system that is suitable for a specific application. A good example is the Probogotchi game, described in the next section.

6.7 Probogotchi

The inevitability of technological progress reflects our dependency on it. There are two aspects to this: our dependency on machines and the feeling we have when they are dependent on us. The success of the Tamagotchi [26] (Bandai, 2000) and its many derivatives is based on our associations with dependency (the needs pets may have that we would fulfil) and the attraction of perceived unpredictability (not knowing when the pet will be hungry, sleepy, or in a

Figure 6.17: A screen shot of the Probogotchi game.

playful mood) yet reliability (if we played with it, it will be happy). An artificial system like the Tamagotchi is designed to constantly ask for our attention through some "feeding" or "loving" mechanisms. Our interactions with the device develop through this artificial notion of dependency and social action and reaction. In the case of the Tamagotchi, users are not "tricked" into thinking that the device is a real pet with real needs. Its success is in its ability to seduce users into a captivating scenario. The scenario is captivating precisely because it plays on the user's sense of satisfaction of being depended on. The fear of emotional manipulation is reduced here when the artificiality is obvious and the user knowingly, voluntarily, and even seeks to participate in these artificial scenarios [42]. The Probogotchi game is a 3D game where children have to interact with Probo in order to make him happy. Probogotchi is based on the idea to care for a virtual pet and make it dependent of the users' actions. The Probogotchi game uses the described homeostatic system to simulate his emotional state, displayed by the facial expressions. The internal needs are presented to the users in the form of health bars shown in Figure 6.17. For this game the homeostasis system has been configured with 6 different needs (hunger, affection, health, energy, pain and hygiene) that are influenced by six different interactions. Four are triggered using the RFID iconic cards (feed fruit, feed burger, give a pill, wash with soap). Two other interactions are triggered from the touch analysis that can detect pleasant touch (hugging or tickling) and unpleasant touch (poking or hurting). To set the timing of the homeostasis system the "do nothing" action is triggered by a timer after a certain interval. By doing the right actions at the right time the emotional state of Probo will change towards a happy state (valence and arousal > 0.5). The game has different levels of rising difficulty obtained by decreasing the time interval

Figure 6.18: Children playing the Probogotchi game.

for the “do nothing” action. The game has been tested with many children on different expositions that can be seen in Figure 6.18. Most of the children were very fond of the new way they had to interact with a game. They had real empathy with the feelings Probo showed, what motivated them to take good care for him. Probogotchi is an innovative combination of different modules that have been developed within this project. With Probogotchi a new type of game controller is introduced, where a user interacts with a virtual character through a stuffed representation of that animal. It is a perfect example of what is possible with the modular design principles used during development.

6.8 Summary

The success of the non-verbal face-to-face communication for Probo depends on its facial expressions. The emotion space provides an easy way to set an emotional state that automatically translates into a corresponding facial expression. The mapping between the emotional vector and the corresponding positions for each DOF can be easily configured by any operator. A universal configuration to express the basic emotions based on the FACS has been tested in different user-studies. At first the virtual model was used, later expressions with the physical robot were evaluated. The results showed that the addition of the trunk had a positive effect on the recognition rate. The recognition rate is slightly higher when using pictures of the virtual model (88%) rather than pictures of the physical robot (84%). This means that improving the fur-covering, to look more like the virtual model, could lead to better results. After comparison with other similar robots, we conclude that Probo's facial expressions are well recognized. To link the emotional state of the robot with the social interactions, a homeostatic system is developed. This systems uses perceived actions to influence internal needs for the robot. The arousal and valence to determine the emotional state, are calculated based on the values of the different needs. The homeostatic system can be fully configured and monitored by the operator. In the Probogotchi setup, the homeostatic system is configured to act as a pet that needs to be taken care of. This setup was successfully tested and demonstrated at various science fairs. The next chapter will present the animation and motor systems that are responsible for creating Probo's life-like motions.

Chapter 7

The Motor System

"Limitations live only in our minds. But if we use our imaginations, our possibilities become limitless."

-Jamie Paolinetti-

7.1 Introduction

Creating a believable imaginary creature with human-like characteristics and smooth, natural motion has been successfully realized for many years. Starting with the first hand-drawn 2D animation films in the late 1930s and continuing with the more recent successes of 3D animation. In [117], Scheeff et al. discus how techniques from traditional animation can be used in social robot design. The principles employed in realizing a successful "illusion of life" in Walt Disney's famous cartoon characters are very well described in [130]. While inspiration can be constructively drawn on how to apply similar strategies to designing social robots and create the illusion of life and intelligence, the problem for the functional design of the social robot is much more complex, of course, than cartoon characters as behind each character is a puppet master [41]. That is why in our implementation we try to share the control of the robot with an operator, and provide him with the tools to act as a kind of puppet master for the robot. Robotic movements are considered to be very abrupt, in contrast to the smooth, elegant motion displayed by humans and animals. Another technique to create this illusion of "life" is to implement some form of unpredictability in the motion and behavior of the robot to make it appear more "natural". In [131] Van Breemen proposed to apply principles of traditional animation to make the robot's behavior better understandable. Stating that a preprogrammed set of carefully designed animation motions requires a change in the software architecture of the robot. And a general architecture that merges preprogrammed motions and feedback control loops is needed.

Based on the design drivers for social interaction (see Section 3.5.5), the software tools for creating smooth and natural motions, compliant with the human interaction rate, are developed. This chapter presents the necessary tools to create motion sequences or animations that make the robot look more "alive". The Sequence Editor (Section 7.2.1) is used to create motion sequences that can be used by the Animation Triggers (Section 7.2.2), including the Animation Player and Animation Keys. The previous chapters described how the perceptual stimuli are processed to provide attentional gaze and facial expressions. The different outputs from these systems are combined with the outputs from the additional Direct Motion Control tools (Section 7.4), using the Combination Engine (Section 7.3). Resulting in the shared control between the operator and the autonomous systems reacting on the input stimuli. The Motion Mixer (Section 7.5) provides the operator with an additional control over the different motion outputs of all systems. The output of the Motion Mixer holds the normalized positions of the DOFs that are send to the motors. The Motor Thread (Section 7.6 will first filter these DOF positions and transform them into motor positions. Next it will use the tools, provided by the Actuation unit (Section 7.7, to send the motor commands over the CAN bus to the EPOS motor controllers. An overview of the Motor Systems is presented in Figure 7.1.

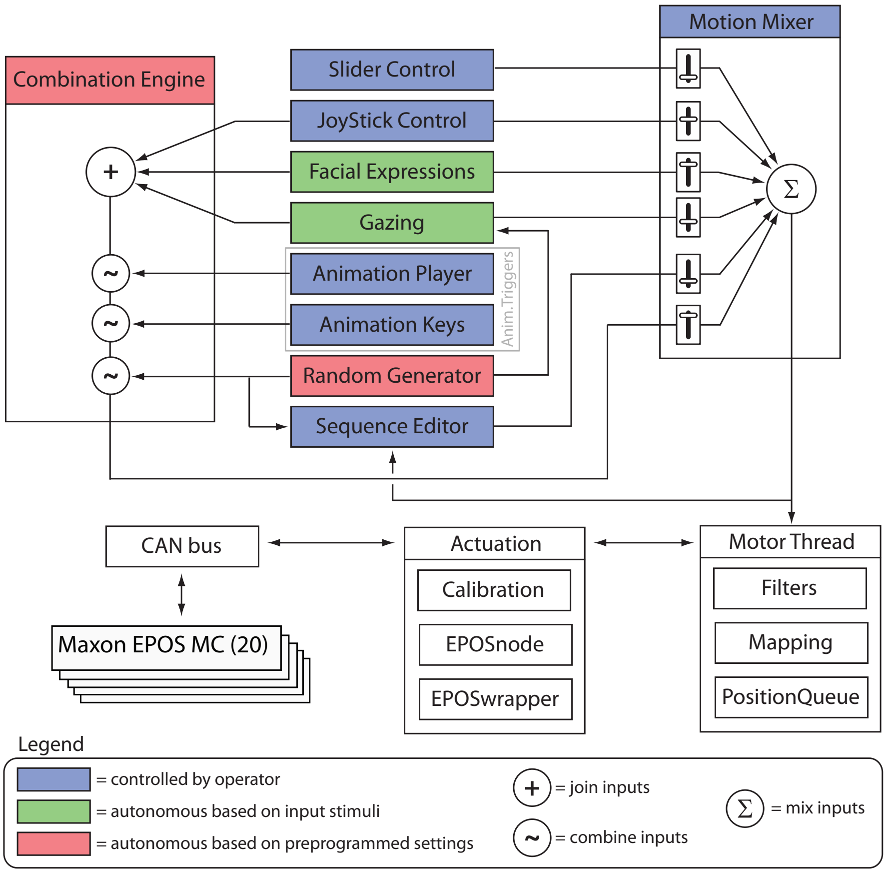

Figure 7.1: The architecture of the Motor system.

7.2 Animations

Traditionally, the control of robotic body parts is carried out by feedback control loops. This results, however, in rather machine-like behavior that does not reveal much about what the robot is doing or thinking [131]. Van Breemen argues that in order to bring robots to life - such that they show behavior that can be naturally understood and anticipated - principles known from the field of character animation should be applied. We share this point of view and developed a Sequence Editor, inspired from the principles of character animation, to create motion sequences. These sequences can subsequently be used by other software tools, such as the Animation Triggers and the Combination Engine, to trigger them and combine them into believable animations.

7.2.1 Sequence Editor

In computer animation, most of the animators manipulate the animation variables (avars) that control the position of a part of an animated object (e.g. a character). Rather than set avars for every frame, they usually set avars at strategic points (frames) in time and let the computer interpolate or "tween" between them, a process called keyframing. Keyframing puts control in the hands of the animator, and has roots in hand-drawn traditional animation. A key frame in animation and filmmaking is a drawing that defines the starting and ending points of any smooth transition. They are called "frames" because their position in time is measured in frames on a strip of film. A sequence of keyframes defines which movement the spectator will see, whereas the position of the keyframes on the film, video or animation defines the timing of the movement. Because only two or three keyframes over the span of a second does not create the illusion of movement, the remaining frames are filled with inbetweens. The Sequence Editor (see Figure 7.2) is developed using the same

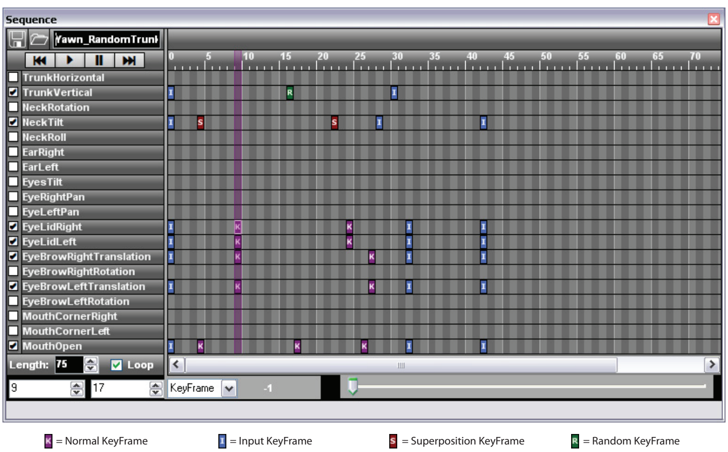

Figure 7.2: The GUI of the Sequence Editor.

principles that are being applied in computer animation. The Sequence Editor has a time line that is composed out of a sequence of frames. Identical as with a video film, each frame can be seen as a still picture and after pressing play the frames are shown at a certain frame rate. A linear interpolation is used to fill in all the frames between two keyframes, providing a smooth transition. The editor has a separate time line (or track) for each of the DOFs. If certain DOFs are not used in the sequence they can be set to non-active, so they will not be taken into account when the sequence is later combined with other motions. A loop can be defined in every sequence by setting a start and end frame. When this loop is active, only the frames between the start and end frame will be played. A keyframe can be created by clicking on a certain frame in a DOF track. By default a Normal Keyframe will be created. The editor supports four types of keyframes:

The Normal Keyframe has a fixed value that is automatically set to match the output value of the Motion Mixer at the time the Keyframe is created. Optionally this value can be changed using the slider at the bottom of the GUI.

The Input Keyframe copies the value of the other motions that are running at the time the sequence is being played.

The Superposition Keyframe adds its fixed value with the value of the other motions that are running at the time the sequence is being played. Its fixed value can be set in the same way as with a Normal Keyframe.

The Random Keyframe takes a random value each time the sequence is being played.

Each Keyframe can be dragged and dropped or copy and pasted. To create a smooth motion sequence loop the following steps need to be considered. First, Input Keyframes are used before the start frame of a loop, to guide any previous motion smoothly to the Normal Keyframes (start frames of the loop). Using Normal Keyframes inside the loop will override any underlying motion at the same DOF track. To finish, it is best to use Input Keyframes again at the end of the sequence (outside the loop) to return smoothly to any underlying motion that is generated by other sequences or systems. This is depicted for with an example of a "Yawn" sequence in Figure 7.2. The Superposition Keyframe is useful if you want to add a certain motion sequence on top of another motion. For example if the gaze is directed towards a face, a sequence for nodding 'yes' or 'no' has to be added to the underlying head motion to maintain the gaze direction. The linear interpolation to calculate the position at a certain time ($P(t)$) between two keyframes ($p_1(t_1)$ and $p_2(t_2)$) on a single DOF track is given by Equation 7.1 and depicted in Figure 7.3.

$$P(t) = p_1 + \frac{(p_2 - p_1).(t - t_1)}{t_2 - t_1} \tag{7.1}$$

If other types of keyframes are used than the Normal Keyframe, p_1 and/or p_2

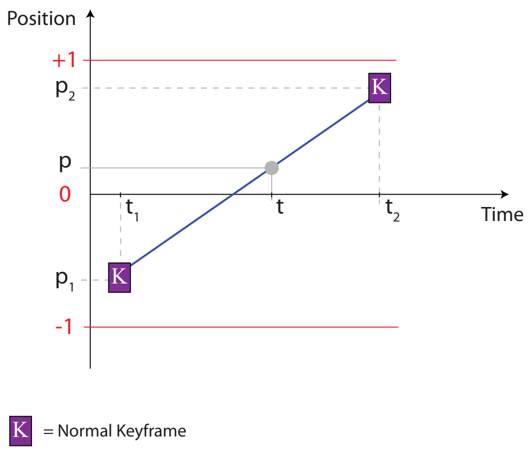

Figure 7.3: The interpolation that determines a frame's position at a certain time between two keyframes.

are replaced (according to the type) with: the current motion value (Input), the keyframe value added with the current motion value (Superposition), a random value (Random). An example of the result of Input Keyframes and Superposition Keyframes on an underlying motion (lower priority) is depicted in Figure 7.4. After using these keyframes to create smooth motion sequences, each sequence needs to be saved for use in the Animation Triggers unit.

7.2.2 Animation Triggers

The Animation Triggers unit is used to control the sequences that have been created using the Sequence Editor. The operator is presented with the GUI of the Animation Triggers unit, depicted in Figure 7.5. This GUI provides the operator with the ability to control the different components contained in this unit, as there are:

The Random Generator controls , which are simple checkboxes to toggle the automatic blinking of the eyes and flappering of the ears on or off. Both sequences are triggered with changing random time intervals.

The List with Loaded Sequences contains the sequence files that are loaded for use with the Animation Player or Animation Keys. The sequence files are created with the Sequence Editor.

The Animation Player allows the operator to combine different sequences into longer animations. The sequences that are loaded in the list, can

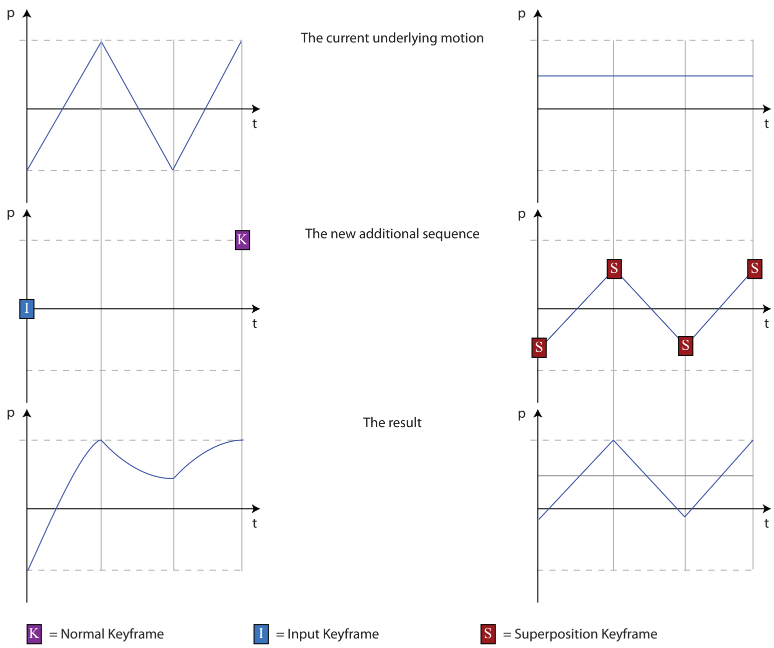

Figure 7.4: The influence of the type of keyframe on underlying motions.

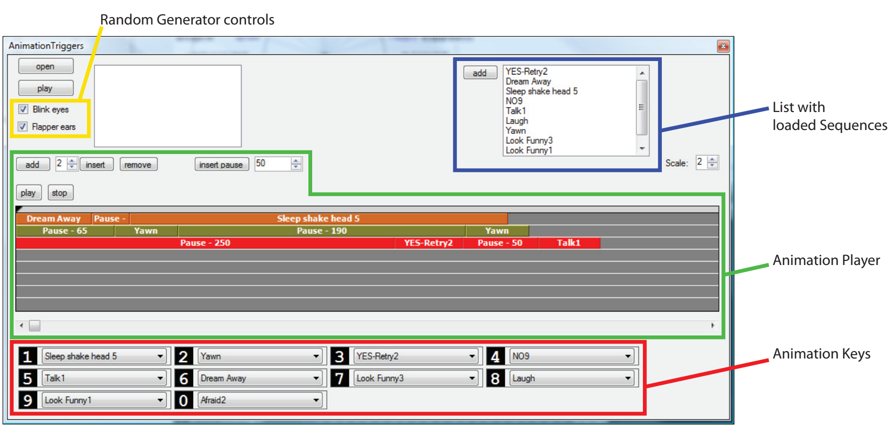

Figure 7.5: The GUI of the Animation Triggers unit.

be inserted at a certain level (or priority) on the time line. When two sequences overlap, the sequence with the highest level will gain priority during the time they overlap. The sequence with the highest level can override the motion data of all the common active DOFs, following the override rules defined by its keyframes' types.

The Animation Keys give the operator the ability to use the keyboard for triggering different sequences. In the current implementation each number of the keyboard can be linked with a sequence from the loaded list. If a key is pressed the corresponding sequence will start to play. If a loop has been defined for this sequence it will play in that loop as long as the key is pressed. After releasing the key, the sequence will continue to play till the last frame (ignoring the loop's end frame).

Given the Sequence Editor, to create a database of motion sequences, the Animation Triggers can be used by an operator to act as a "puppet master". To combine these animations with the motions originating from the attentional and emotional systems, a Combination Engine is introduced.

7.3 Combination Engine

In [131], Van Breemen states that a mixture between pre-programmed motions and feedback loops is required. The pre-programmed motions are designed to make the robot better understandable, whereas feedback loops let the robot react to stimuli from the environment. To comply with the design specifications on autonomy (described in Section 3.3.3), we developed the Combination Engine to allow a shared control between the operator and the reactive systems of the robot. The dataflow of the Combination Engine is depicted in Figure 7.6.

Different DOF tracks that are generated from the facial expressions (EyeBrows, EyeLids, Ears and Mouth), the gaze (Eyes and Neck) and the joystick control (Trunk) are joined into one list of temporary DOFs. In the next stage the DOFs of the eyelids are corrected to keep them relative with respect to the position of the eyes. The sequences that are being played by the Animation Player are then combined according to their priority level and their keyframe types. The resulting motion is subsequently combined with the temporary DOF list. If one key is triggered on the keyboard, the corresponding sequence will be combined with the temporary DOF list. If more keys are triggered, the sequences will first be combined with each other, following a priority from left to right, before they are combined with the temporary DOF list. Finally, at random time intervals, a sequence to blink the eyes and flap the ears is combined with the DOF list. This DOF list is then the output of the Combination Engine.

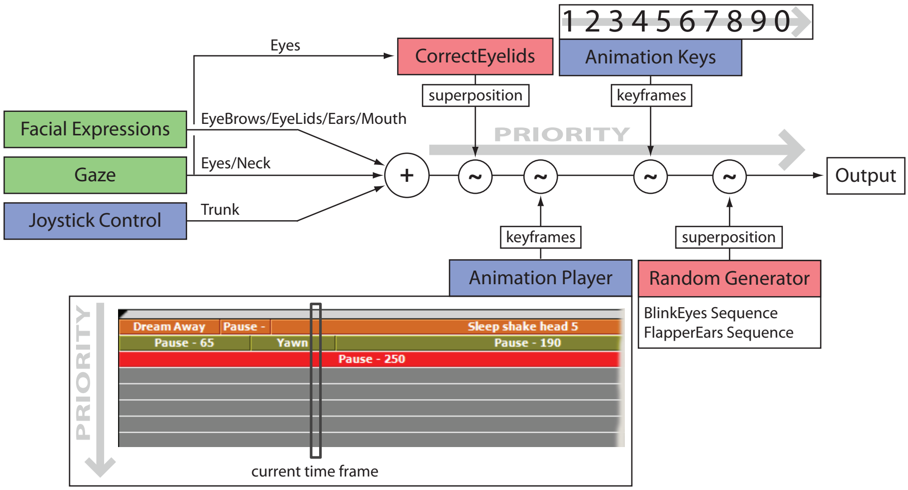

Figure 7.6: The dataflow of the Combination Engine.

7.4 Direct Motion Control

The operator is provided with two tools for direct control of the DOFs:

The Slider Control can be used to change the value of each DOF separately and is mainly used for testing. The GUI is depicted on the left side in Figure 7.7. The sliders have a resolution between -1000 and 1000, that is transformed to the normalized DOF (between -1 and 1)

The Joystick Control is to be used with a hardware Xbox 360 controller (both wired and wireless are supported). The Xbox 360 controller is in fact a gamepad (and not really a joystick). The gamepad has a set of buttons, two triggers, two thumb sticks and a directional pad. The GUI allows the operator to link each of the DOFs with the gamepad's controls as depicted on the right side in Figure 7.7. The gamepad is an efficient tool for an operator to gain direct control over some the robot's DOFs. The wireless controller gives the ability to tele-operate some features of the robot, without being noticed. This can be very useful during Wizard of Oz testing.

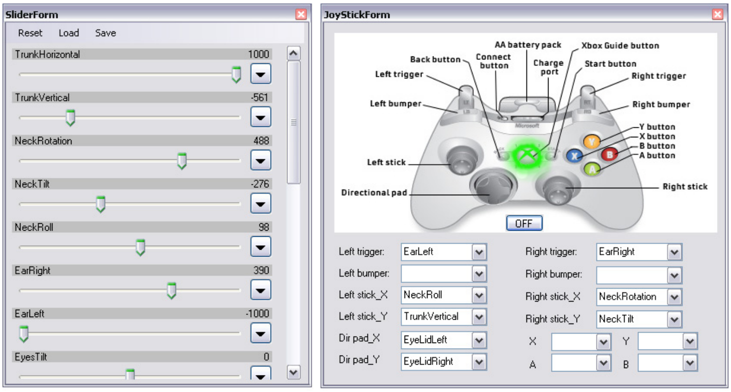

Figure 7.7: The GUIs for the Slider and the Joystick control.

7.5 Motion Mixer

The Motion Mixer is developed based on the principles of an audio mixer. The Motion Mixer will dynamically create motion channels for each of its inputs. All the inputs are provided with a slider to control their gain. Each of the active DOF tracks of all inputs are added together according to their gain. The mean value is subsequently set as the output of the mixer (see Equation 7.2). The GUI for the Motion Mixer is depicted in Figure 7.8.

$$Output = \sum_{i=0}^{n} \frac{Gain_i.Input_i}{n} \tag{7.2}$$

Every software unit (client) that wants to provide an input for the mixer can implement IMixerInput. IMixerInput is a software interface from the Motion Mixer unit demanding the client to provide the necessary (public) methods, so the mixer can process their input. One of these methods is a public getter for an object of the DOFProbo class. The DOFProbo class contains a list of all the DOF (objects) that can be controlled for the robot Probo. Each DOF object has the following properties; *name* (string), *position* (double) and *active* (bool). The position of a DOF is normalized; -1=minimum, 0=neutral, 1=maximum. Notice that only a DOF object with his active property set to “true” will be added to the mixing. The Motion Mixer produces an output DOF (also a DOFProbo object) that is update continuously. An IMixerOutput interface is provided for other software units that want to use the output of the mixer.

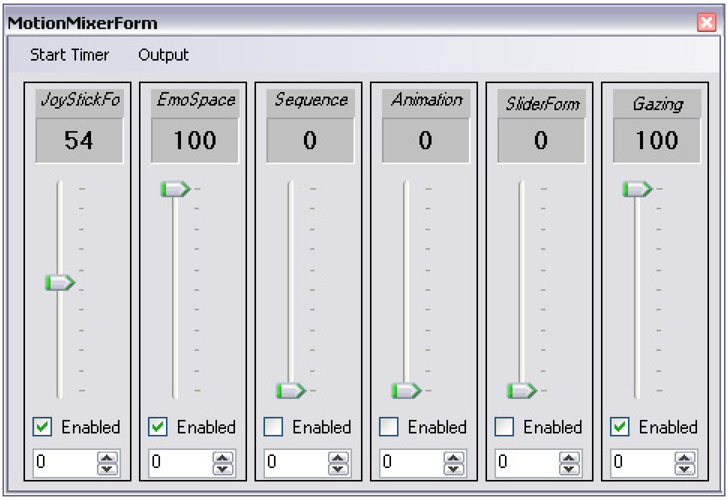

Figure 7.8: The GUI for the Motion Mixer.

7.6 Motor Thread

The Motor Thread unit is responsible for transformation of the normalized DOF values into motor positions. First the DOFProboFilter is applied to ensure smooth motion. Next a Round Robin system is used to update the DOFs, which are divided into four groups. All the (normalized) DOFs are subsequently converted into motor positions (Mapping) and are put on the PositionQueue. The PositionQueue is the buffer serving the MotorWorkerThread. The requested motor positions are compared with the actual motor positions. If there is a difference, the velocity is calculated to reach the requested position. Finally, the new target motor positions are send to the EPOS motor controllers. This is depicted in Figure 7.9.

The DOFProboFilter contains all the filters for each of the DOFs. All filters are low-pass filters with adapted to provide smooth, natural motion. Each filter consists of a cascade of first order software low-pass filters as depicted in Figure 7.10. Different α values are used taken into account the motor transmission and the body part that needs to be actuated.

The Round Robin was introduced after some timing issues occurred. Instead of updating the full DOF list, only one of the four groups is updated at each timer tick. Each group combines the DOFs that need to be updated at the same time. Because only the updated positions will be processed, the MotorWorkerThread can now speed up communication with the EPOS motor controllers.

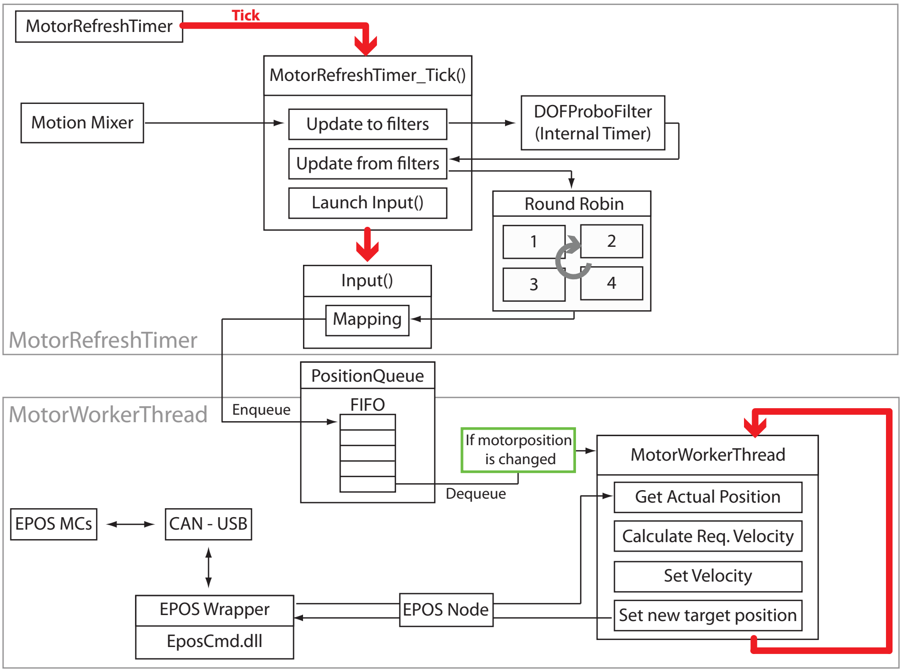

Figure 7.9: The working of the Motor Thread unit.

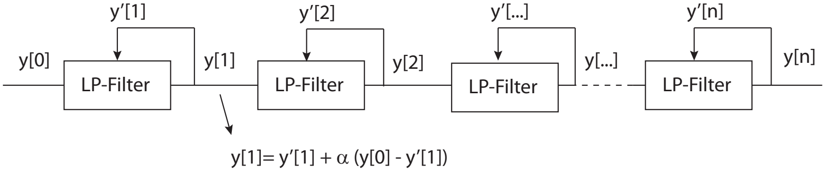

Figure 7.10: A cascade of low-pass filters to provide smooth, natural motions.

The Mapping converts the filtered DOF values into quad counts. All Maxon positioning systems evaluate the rising and falling signal edges. With regard to encoder pulse count, this results in a four times higher positioning precision. This is what is referred to as quad counts. Converting the normalized DOF to quad counts is a simple multiplication for most of the DOF with exception of the eyebrows and the trunk. The software representation of the DOFs of the eyebrows consist of a translation (up/down - T) and a rotation (clock/counterclock-wise - α). Their mechanical implementation consists of a combination of a translation on the left(ML) and on the right(MR) side of each eyebrow. The conversion is given by Equation 7.4.

$$ML = T - \tan(\alpha.\frac{\Pi}{6}) \quad (7.3)$$

$$MR = T + \tan(\alpha.\frac{\Pi}{6}) \quad (7.4)$$

The transformation, from 2 DOF to 3 motor positions, to actuate the trunk is more complex; going from horizontal and vertical curling motion, towards the rotation of the three motors that wind and unwind the elastic cables (C_1, C_2, C_3). First the horizontal and vertical DOF are used to define a trunk vector: $V(r, \alpha)$. This vector is used to calculate the cable length (CL) that needs to be (un)winded (see Figure 7.11). Finally the rotation (θ) of the motor is calculated based on the radius (R) of the reel that is used to wind up the cables.

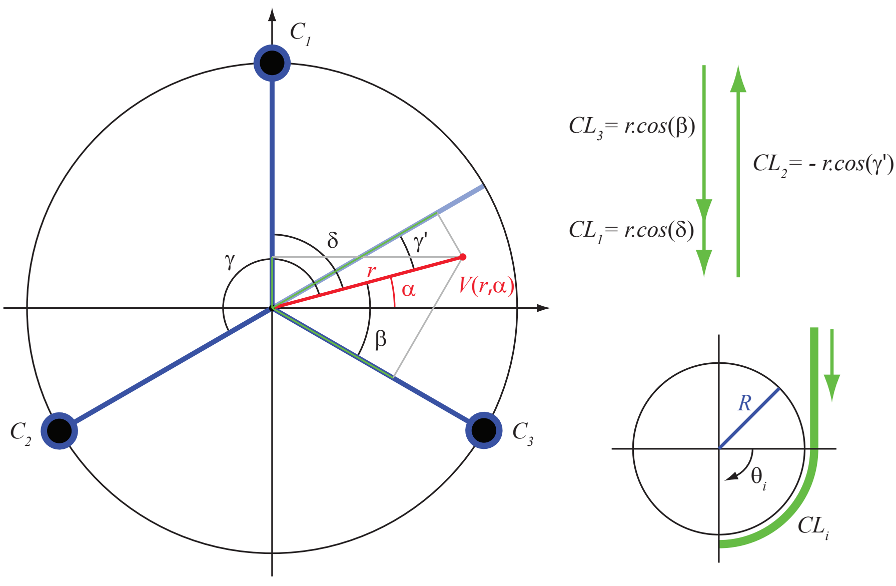

Figure 7.11: The conversion from the trunk vector $V(r, \alpha)$ into the motor rotation angle θ_i.

The Position Queue serves as a buffer between a timer-based system that produces the motor positions and a thread-based system that wraps these positions into readable packets and sends them to the motor controllers. The Position Queue is a First In First Out (FIFO) buffer with no more than five slots available. If the five slots are occupied when a new set of positions is ready to be enqueued, the oldest set will be discarded to make room for the new positions. This is to prevent latency from building up and to create a soft real-time system. The MotorWorkerThread, responsible for dequeuing the positions, will use the components of the Actuation unit to communicate with the EPOS motor controllers.

7.7 Actuation

The Actuation unit is a set of components, used by the Motor Thread unit, to communicate with the Maxon motors. Communication with Maxon motors is established via the EPOS motion controllers. EPOS is a modular-designed digital positioning system suitable for DC and EC Maxon motors with incremental encoder. The built-in CANopen interface allows networking to multiple axis drives and online commanding by a CAN bus master unit. In our implementation the CAN bus master unit is provided by a CAN-USB interface from National Instruments. The Actuation unit has three major components:

The EPOS Node is the software representation of a motor and corresponding EPOS; including its name, address, calibration settings and all of the motion parameters (Mode, Profile, MaxVelocity, MaxAcceleration, PID,...) that are needed to fully control the motor's output.

The EPOS Wrapper is a wrapper for the EPOS communication in managed .NET code. It uses Maxon's EposCmd.dll to provide access to the EPOS control functions. These functions are used to send commands and receive data from the EPOS motor controllers over the CAN bus.

The Calibration provides the tools for calibrating the motor settings after the robot is powered up. The encoders used to control the motors' positions are not absolute and therefore need to calibrated each time the power is turned off. The calibration or homing of a motor is done to define the minimum and maximum position. Optical and mechanical switches are incorporated into the mechanical design of the robot, to indicate the minimum and maximum position of each DOF.

7.8 Summary

Inspired by the creators of computer animation, the tools to create life-like animations for Probo are developed. The operator can use this toolkit to create keyframed motion sequences and use them to build larger animations. A GUI is provided to manage these motions that are combined by the Combination

Engine with the motions originating from the autonomous systems presented in the previous chapters. All motions for Probo are represented using the DOFProbo object, containing a list with all (19) DOFs. After the Motion Mixer has mixed all incoming DOFs, the MotorThread will process the DOFProbo to obtain smooth motions in the form of new motor positions. After calibration all (20) motors are ready to receive their new positions. The new velocities for the motors are calculated based on the actual positions. The delay that is experienced after the request for the actual positions, leads to high latencies. To solve this problem a round robin system is introduced that uses a FIFO buffer to produce updated motor positions. Currently a new communication protocol is being tested that would significantly reduce the mentioned delay and thereby cancel the problem (and the corresponding round robin system). This and other future work to improve Probo is further described in the next chapter.

Chapter 8

Conclusions and Future Work

"What we think, or what we know, or what we believe is, in the end, of little consequence. The only consequence is what we do."

-John Ruskin-

8.1 Conclusions

All concepts and systems that are presented in the previous chapters are successfully implemented into the robots hardware. That resulted in a working prototype of the social robot Probo. This prototype has been presented to the press on the 21th of April 2009 as depicted in Figure 8.1.

Figure 8.1: Probo at the press conference on the 21th of April 2009.

International news reports, interviews and other media coverage can be found on our website: http://probo.vub.ac.be/press/. The press coverage has contributed to spread the word about this new robotic creature. By this we positioned Probo among the state of the art in social robots as was aimed in the beginning of this project. We also believe that Probo holds a unique position serving a social interaction with children and a research platform for HRI. In our opinion most of the goals (presented in Chapter 3) addressing two types of users; interactive (child) and controlling (operator), have been achieved.

8.1.1 Probo for Children

To address the children, Probo has been given a “fairytale”-like identity and history that matches his embodiment and remarkable trunk. We were able to achieve the predefined appearance of a stuffed imaginary animal with the ability of smooth and natural motions (see Section 3.3.2). Aspects of safety and soft touch or “huggability” are successfully implemented. A triple layered modular structure (depicted in Figure 8.2) ensures this softness and protects

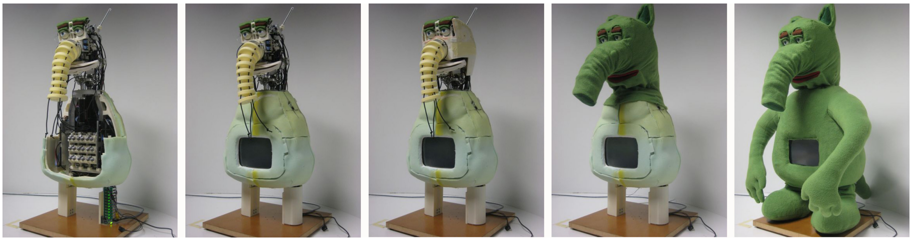

Figure 8.2: The modular layered structure of the robot Probo.

inside mechatronics from the outside world and vice versa. To comply with human standards of hygiene, the outside layer is a kind of jacket that can be removed and cleaned. Compliant actuation and the use of flexible materials contribute to the aspect of safe and soft interaction. The 3D virtual model of Probo can be used to depict Probo in different poses in pictures or animated movies. Pictures of the facial expressions, simulated with the virtual model, were used in user-tests to see how the recognition can be improved. The latest test, incorporating the trunk, resulted in an overall recognition rate of 88% for the virtual model and 84% for the physical robot. In comparison with other similar projects the recognition rate for Probo is very good and equals the recognition rate of human facial expressions. Based on the virtual model, a stuffed version of Probo is created for public relations and evaluations with the children. Although the likeability of both the virtual and the stuffed model are presumed to be good, specific likability tests have not been taken yet, but most children reacted very positive after interaction with Probogotchi (as can be seen at http://probo.vub.ac.be/press/television/). Probogotchi combines both models into an interactive game for the children. The game also incorporates the homeostatic and emotional system that are used to define an emotional state and corresponding facial expressions based on the perceived actions that are triggered by the children. The homeostatic system is configured to simulate pet-like needs, motivating the children to take care for an artificial pet.

8.1.2 Probo for Operators

In its current implementation, Probo's social skills are limited to a reactive level with two major autonomous systems: the attention system and the homeostatic system. These systems share the control over the robot with an operator. The operator is accommodated with a control center to take care of the higher-level decisions, like in a Wizard of Oz setup. The control center manages all software components and composes a flexible GUI to control all individual systems. The 3D virtual model is added as a simulator of the robot's motions and a viewport for the operator. The modular structure of the software architecture allows an easy introduction of new components. By gradually increasing autonomy, the operator receives an ever higher level of abstraction. In Section 3.3.3, we stated that the robot must be able to perform autonomous interaction sessions, that are pre-configured and guided by an operator. Although the operator is given a toolkit to create and trigger smooth motion sequences, a higher-level of scenario/behavior system is needed.

8.1.3 Social Interaction

The social interaction for Probo was predefined to focus on non-verbal communication (see Section 3.5.5) using the following abilities:

Human-oriented perception has been achieved by the following systems:

- Vision analysis includes the detection of color blobs, skin-color regions and human faces.
- Audio analysis will include the detection of the direction and intensity of sounds. (under development by DSSP - Digital Speech and Signal Processing - VUB in HOA16: see Appendix B)
- Touch analysis gives the location and force of touch, that is classified into unpleasant, pleasant or normal touch.
- Object identification is used to identify embedded RFID tags in symbolic cards that represent actions towards the robot.

Readable social cues are expressed by the following systems:

- Gaze direction is conveyed using the point of attention provided by the attention system.
- Facial expressions are displayed using a graphical mapping of the emotional vector in the emotion space, into a position for each of the facial DOFs.
- Sound effects are produced for auditory feedback when an action from the homeostatic system is triggered.
- Non-sense affective speech will be produced using the emotional vector as input source. (under development by DSSP in HOA16)

Smooth and natural motions are realized using the following systems:

- The animation system provides a toolkit to create motion sequences and combine them into larger animations.
- The combination engine takes into account the motions originating from the autonomous systems and combines them with the motions originating from the operators commands.

- The motor system filters the motions to provide life-like movements before they are transformed into motor positions and passed on to the motor controllers.

These abilities are depending on the following underlying autonomous systems:

The attention system is preliminary, and in its current implementation limited to visual stimuli. By use of an active vision system, the operator is presented with an image of the full view area and the configuration of the attention system. A more advanced attention system is under development by the multidisciplinary team of the HOA16 project.

The homeostatic-emotional system is used to react on perceived actions towards the robot. Once an action is triggered it will influence the needs of the robot that are used to determine an emotional state by calculating the position of the emotion vector in the emotion space. The emotion space uses the dimension of arousal and valance to define an emotional state.

Demonstrations of these abilities can be seen at http://probo.vub.ac.be/Pics_Video/ and http://probo.vub.ac.be/press/television/.

8.1.4 Probo for research

Probo is developed to serve as a multidisciplinary research platform in three major research areas:

Technological This includes mechatronics, software engineering, AI, vision, speech and audio.

Biological/Social/Psychological This research is situated in human-robot interaction, including the models and theories to simulate social interaction inspired by humans and animals.

Therapeutical This includes the exploration of the possibilities in robot assisted therapy.

The research activities with VUB departments in these areas are grouped in the HOA16 project. A complete list of the collaborations and project proposals is given in Appendix B.

8.2 Future Work

8.2.1 Updates and Problem Solving

The current implementation of Probo is still in a prototype phase, and can not be regarded as an end-user product. But to ensure a good operational state the following updates should be noticed:

The software architecture makes use of the .NET 2.0 framework, which is released in 2005. An update to .NET 3.5 will provide more options, especially for the improvement of the GUI. During this update all components should be reviewed to comply with the rules for full component oriented design. This mainly concerns the improvement on the interfacing. Also a connection with Microsoft's Robotic Studio (or other popular robotic softwares) should be taken into consideration.

The appearance of Probo that is realized by its outside layer, its "jacket", has to be improved. This must be performed by specialists in the creation of puppets and animal-like animatronics.

The MotorThread needs to be updated with a new communication protocol to resolve a latency issue that is experienced in the current implementation. This is now being tested and will make the current Round Robin system superfluous in the future.

8.2.2 New Developments

To ensure the future for Probo as a multidisciplinary research platform, a new proposal for "the consolidation, innovation and validation of the robotic research platform" has been submitted for further development or integration of the following systems:

The attention system needs to be gradually improved by incorporation of additional perceptual stimuli, saliency maps and habituation effects. This is currently under development by the HOA16 project.

A behavior system will improve the global social interaction level. It is the next step to provide more autonomy to the robot and a higher-level of abstraction to the operator. A master thesis has started for the design of a first implementation.

A gesture system will contribute to the expression of readable social cues. In combination with the hardware development for arms and hands, a gesture system has to be developed and incorporated into the existing architecture. This development is part of the submitted project proposal.

8.2.3 Validations and Applications

To support the future deployment of Probo for children in the hospitals, additional validation test have to be performed and useful applications have to be implemented. The implementation for cognitive behavioral therapy is currently under development in collaboration with ORPS (Orthopsychology - VUB) in the HOA16 project. An implementation for post-procedural pain measurement with children is under development in collaboration with ORPS and KHK (Katholieke Hogeschool Kempen). Besides applications with therapeutical use, the development for applications that focus on the distraction, communication and information (see Section 3.4.1) should be promoted. The deployment of Probo in the hospitals will be further supported by our co-initiator; the Anty Foundation. Being a non-profit organization, the Anty Foundation has the purpose to improve the life circumstances of vulnerable children by placing technology at the service of children. A goal shared by all members of this project.

Appendix A

List of Publications

Journal Articles

- J. Saldien, K. Goris, S. Yilmazyildiz, W. Verhelst, and D. Lefeber. On the Design of the Huggable Robot Probo. *Journal of Physical Agents, Special Issue on Human Interaction with Domestic Robots*, June, 2008. Red de Agentes Fisicos, 2(2):3-12, ISSN 1888-0258.

- J. Saldien, K. Goris, B. Vanderborght, J. Vanderfaeillie, and D. Lefeber. Expressing Emotions with the Social Robot Probo. *International Journal of Social Robotics, Special Issue on Social Acceptance in HRI*, (submitted August 2009).

Reviewed Conference Proceedings

- K. Goris, J. Saldien, and D. Lefeber. Probo, a Testbed for Human Robot Interaction. In *Proceedings of 4th ACM/IEEE Internal Conference on Human Robot Interaction* San Diego, California, March 11-13, 2009. ACM, pp. 253-254 ISBN 979-1-60558-404-1.

- K. Goris, J. Saldien, I. Vanderniepen, and D. Lefeber. The Huggable Robot Probo, a Multi-disciplinary Research Platform. In *Eurobot 2008 Conference*, Heidelberg, Germany, May 22-24, 2008. MatfyzPress, pp. 63-68, ISBN 978-80-7378-042-5.

- K. Goris, J. Saldien, B. Vanderborght, and D. Lefeber. The Huggable Robot Probo: Design of a Robotic Head. In *The 2nd AISB Symposium on the Role of Virtual Creatures in a Computerised Society: The Reign of Catz and Dogz*, Aberdeen, Scotland, April 1-4, 2008. The Society for the

Study of Artificial Intelligence and Simulation of Behavior, 1(1):23-29, ISBN 1 902956 60 5.

- J. Saldien, K. Goris, B. Vanderborght, and D. Lefeber. On the Design of an Emotional Interface for the Huggable Robot Probo. In *The 2nd AISB Symposium on the Role of Virtual Creatures in a Computerised Society: The Reign of Catz and Dogz*, Aberdeen, Scotland, April 1-4, 2008. The Society for the Study of Artificial Intelligence and Simulation of Behavior, 1(1):1-6, ISBN 1 902956 60 5.

- K. Goris, J. Saldien, B. Vanderborght, B. Verrelst, R. Van Ham, and D. Lefeber. The Development of the Eye-System for the Intelligent Huggable Robot ANTY. In *9th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, Brussels, Belgium, September 12 - 14, 2006. pp. 168-171.

- J. Saldien, K. Goris, B. Vanderborght, B. Verrelst, R. Van Ham, and D. Lefeber. ANTY: The Development of an Intelligent Huggable Robot for Hospitalized Children. In *9th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, Brussels, Belgium, September 12 - 14, 2006. pp. 123-128.

- D. Lefeber, B. Verhelst, R. Van Ham, P. Beyl, B. Vandenborght, M. Van Damme, J. Naudet, R. Versluys, J. Saldien, K. Goris. Compliant actuators for robots in direct contact with humans. In *9th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, Brussels, Belgium, September 12 - 14, 2006. pp. 771-777.

- B. Vandenborght, B. Verhelst, R. Van Ham, M. Van Damme, J. Saldien, K. Goris, D. Lefeber. Trajectory Generator for the Pneumatic Biped Lucy based on Inverted Pendulum Model. In *9th International Conference on Climbing and Walking Robots and the Support Technologies for Mobile Machines*, Brussels, Belgium, September 12 - 14, 2006. pp. 53-58

Reviewed Book Chapters

- K. Goris, J. Saldien, and D. Lefeber. Probo, an Intelligent Huggable Robot for HRI Studies with Children. In *Advances in Human-Robot Interaction*, ISBN 978-953-7619-X-X.

- K. Goris, J. Saldien, I. Vanderniepen, and D. Lefeber. The Huggable Robot Probo, a Multi-disciplinary Research Platform. In *Research and Education in Robotics - EUROBOT 2008*, Germany, September, 2009. Springer, 33:29-41, ISBN 978-3-642-03557-9.

- K. Goris, S. Yilmazyildiz, J. Saldien, B. Verrelst, W. Verhelst, and Dirk Lefeber. Probo, a Friend for Life? In *Brave New Interfaces: Individual, Social and Economic Impact of the Next Generation Interfaces* Brussels, Belgium, 2007. VUB Brussels University Press, pp.253-274, ISBN 978-90-5487-416-4.

Magazines

- J. Saldien, and K. Goris. Probo, een knuffelrobot voor gehospitaliseerde kinderen In *De Connectie*, Amstelveen, Nederland, November 2008. KvK Utrecht, 3(4):23-25, ISSN 1871-3807

Appendix B

List of Probo Related Research

B.1 Project Proposals

B.1.1 Accepted

- ANTY project - Brussels Capital Region

  Name: Ontwikkeling van het ANTY concept: technologische innovatie met een sociaal karakter. Design en bouw van de kindvriendelijke robot ANTY . Development of the ANTY concept: technological innovation with a social aspect. Design and construction of the huggable robot ANTY

  This project consists of two sub-projects:

  - CIBG (working)
    - Period: October 2005 - October 2008 (extended to December 2009)
    - Budget 350.000
  - IWOIB (personnel)
    - Period: January 2007 - December 2009
    - Budget 243.100

  Origin: Brussels Capital Region

  The main goal was to develop a first prototype of the robot. Focus was to show emotions by facial expressions of the head and leave the body unactuated. In the belly a touch screen is placed to provide extra visual information. The robot is in real time controlled by an operator. Focus of the project is the future use of Probo with hospitalized children. All the objectives of the project are attained within the four years of the project with the addition that a virtual model is provided as well.

- HOA16

  Name: Ontwikkeling van natuurlijke mens/robot communicatie. Studie en implementatie van (joint) attention mechanismen. Development of a natural human-robot interaction. Study and implementation of joint attention mechanisms.

  Period: 2+2 years (2008-2011)

  Budget: 371.388 euro: 311.388 euro personnel and 60.000 euro working

  - 2007-2009: 160.000 euro received
  - 2009-2011: 180.000 euro asked

  Origin: HOA16-VUB

  Partners:

  - Robotica & Multibody Mechanica (R&MM), vakgroep Toegepaste Mechanica (MECH), Faculteit Ingenieurswetenschappen (IR) . Prof. Dirk Lefeber - Dr. Bram Vanderborght
  - Digitale Speech and Signal Processing onderzoeksgroep (DSSP), vakgroep electronica (ETRO), Faculteit Ingenieurswetenschappen (IR). Prof. Werner Verhelst
  - Machine Vision onderzoeksgroep (IRIS), vakgroep electronica (ETRO), Faculteit Ingenieurswetenschappen (IR). Prof. Hichem Sahli
  - Departement Cognitieve en Biologische Psychologie (COBI), faculteit Psychologie en Educatie (PE). Prof. Eric Soetens, Prof. N. Deroost
  - Departement Orthopsychologie (ORPS), faculteit Psychologie en Educatie (PE). Prof. Johan Vanderfaeillie
  - Computational Modeling lab (COMO), Departement Informatica (DINF), Faculteit Wetenschappen (WE). Prof. Ann Nowé

  One of the essential research topics regarding “Cognitive Interactive Robotics” is the realization of natural communication possibilities between the robot and the user (e.g. children). The three main objectives of this project are to develop an attention system in function of a natural communication between robot and a human (child) and evaluate these mechanisms using the robot Probo. Also an automatic treatment of chronic pain will be developed. This project develops the cognitive abilities of the robot to increase its autonomy and come to useful applications of the robot. A strong multidisciplinary team is established to achieve these goals. Within this project no funding is foreseen for the further development of the robot itself.

  Workpackages

- Development of an attention system in function of a natural communication between robot and a human (child). (COBI & ETRO)
  - * Development of a working attention model based on (Henderickx, Maetens, Soeten)-model (COBI)
  - * Incorporation of visual attention model (exogenous) (ETRO-IRIS)
  - * Incorporation of auditory attention (exogenous) (ETRO-DSSP)
  - * Incorporation of emotional state (endogenous) (COBI & COMO)
  - * Implementation of Joint Attention in the developed model (COBI)
- Validation scenario to test models. (User-centered design) (ORPS & MECH)
  - * Scenario(s) and user group (of hospitalized children AZK Jette) must be defined (ORPS)
  - * Implementation of models on Probo platform (MECH)
  - * Tests and evaluation with user group (ORPS & MECH)
- Development of automatic treatment of chronic pain. (Cognitive behavior Therapy) (ORPS & MECH)
  - * Development of software interface for this therapy (ORPS & MECH)
  - * Implementation of specific tools that link the virtual model of Probo with the interface (ORPS & MECH)

B.1.2 Submitted

- Concerted Research Action 2010-2015

  Name: Consolidation, innovation and validation of the robotic research platform Probo for cognitive human-robot interaction.

  Period: 5 years (2010-2015)

  Budget: 1.154.333 euro: 838.333 euro personnel and 316.000 euro working

B.1.3 Rejected

- FP7: ARTIPETS - Europe FP7-ICT-2007-3.

  The ARTIPETS project focuses on the development of artificial pets capable of maintaining believable multi-modal, affective interaction with a child over a long period of time (1 week to several months) in therapeutic settings.

  Total budget:

1,3 million - 4 years - 48 MM for VUB

Partners:

- HSR (Prof. Alberto Sana), TXT (ICT-bedrijf), - IT, Milaan
- MIT (Prof. Cynthia Breazeal) - USA, Massatschusets
- Univ. Plymouth (Tony Belpaeme), Distance Lab - UK
- Univ. KaisersLautern(Prof. Karsten Berns), Univ. Siegen - DE
- Univ. of Budapest - HONG
- Univ. Of Eindhoven (Prof. Cees Midden) - NED
- Philips CL iLab (Kees Tuinenbreijer) - NED

- BOAP: Interactive virtual 3D interface for realistic animated simulation of mechanical steered models. - Universitaire Associatie Brussel

  Total budget:

  203.155 euro - 4 years - 12 MM for Ehb Partners:

  Partners:

  - Robotica & Multibody Mechanica (IR) - VUB
  - Animatie lm, Audiovisuele Kunsten - EHB - Rits
  - GroepT (external expert)

- Microsoft Research - Robots Among us: Research and development of a mobile generic, graphical user interface for Probo - PHL (Provinciale Hogeschool Limburg)

  Total budget: $ 70.000 - 12 MM for PHL

  Partners:

  - Robotica & Multibody Mechanica (IR) - VUB
  - Research Institute TINFO - PHL

B.2 Educational/Collaborational Research

B.2.1 Small Projects

2005-2006

- Two International Engineer Students (Roemenia), promotor Dirk Lefeber, Development of trunk mechanism for Probo

2006-2007

- Informatics (KTA Dendermonde): will develop ICT programs for the robot Probo, contact Tom Rydant
- The Hogeschool Gent, Fashion Department, 2nd bachelor fashion technology, made different designs for the external look of the robot Probo, contact Els Janssens
- The Hogeschool Gent, Department Mechanics, 2nd bachelor electro-mechanics, study to develop a robot arm for Probo, contact Antoine De Henau
- Informatics (KTA Dendermonde): Development of interactive story about Probo, contact: Didier Van Maalsake
- Pedagogy (KTA Dendermonde): Study over the experience of a child in a hospital and the link with the robot Probo, contact: Elise De Boeck

2007-2008

- Informatics (KTA Dendermonde): will develop ICT programs for the robot Probo, contact Tom Rydant
- Informatics (KTA Dendermonde): Development of interactive story about Probo, contact: Didier Van Maalsake
- Fotonics (VUB): Implementation of sensors in fibre optics which wil be used to detect touch, contact Thomas Geernaert.
- Computer science (VUB-WISE): Implementation of a virtual model and control in high leve software for a user-friendly user interface, contact Bram Pellens.
- KHK (Katholieke Hogeschool Kempen): The use of the robot Probo as pain measurement for children, contact Prof Louis Peeraer.

2008-2009

- KHK (Katholieke Hogeschool Kempen): The use of the robot Probo as pain measurement for children, contact Prof Louis Peeraer. In collaboration with ORPS (Orthopsychology) and the Katholieke Hogeschool van de Kempen (KHK) an electronic version to measure post-procedural pain with children is under development. The current version is a web survey made in Flash and is coupled with a database to store all the data. After testing with these procedures, the electronic surveys can be redesigned as scenarios on the robot platform and implemented on the scenario module of the robot's control center.

B.2.2 Master Theses

2004-2005

- Jelle Saldien, Hogeschool Antwerpen, Productontwikkeling, Voorgaande studie van de robot Probo en zijn toepassingsgebieden.

2005-2006

- Selma Yilmazyildiz, DSSP, Digital sound and signal processing, VUB, promotor Werner Verhelst, Voorgaande studie van de spraak voor Probo, Communication of Emotions for E-creatures.
- Caroline Davidovic, Haute Ecole Paul-Henri Spaak, Programmatie voor de bewegingen van het hoofd, Commende de la Tête du robot social Probo.

2006-2007

- Arul Mary Sudha Thaines Marian, DINF, Artificial Intelligence, VUB, promotor Ann Nowé, Implementatie van een emotioneel model dmv AI technieken.
- Jeroen Vandenbrande, EHB, Erasmus hoge school Brussel, promotor Kris Steenhout. Programma voor de omzetting van emotie naar gelaatsuitdrukkingen bij een robot.
- Sofie Du Four, ORPS, Orthopsychologie, VUB, promotor Johan Vanderfaeillie, Globale studie over het gebruik van IT en een robot zoals Probo in het hospotaal.
- Tak Long Leung, Haute Ecole Roi Baudouin, Mons, promotor Björn Verrelst, Implementatie van een I2C bussysteem, Le development d'un bus de communication destine à commander le robot Probo.

2008-2009

- Aurelie De Coster, Vrije Universiteit Brussel, Werktuigkunde, Voorgaande studie ontwikkeling arm-grijpmechanisme knuffelrobot Probo.

2009-2010

- Sybren Van Der Straeten, Vrije Universiteit Brussel, Werktuigkunde, Design of a Behavior System for the Social Robot Probo.

Bibliography

1. KM Allen, J. Blascovich, J. Tomaka, and RM Kelsey. Presence of human friends and pet dogs as moderators of autonomic responses to stress in women. *Journal of personality and social psychology*, 61(4):582–589, 1991.

2. R.C. Arkin. Homeostatic control for a mobile robot: Dynamic replanning in hazardous environments. *Journal of Robotic Systems*, 9(2), 1992.

3. T. Asfour, K. Regenstein, P. Azad, J. Schr
"oder, A. Bierbaum, N. Vahrenkamp, and R. Dillmann. ARMAR-III: An integrated humanoid platform for sensory-motor control. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids 2006), Genoa, Italy*, 2006.

4. I. Asimov. Runaround. *Astounding Science Fiction*, 29:94–103, 1942.

5. J. Baillie. Urbi: A universal language for robotic control. *International Journal of Humanoid Robotics*, pages 7–29, 2004.

6. C.Y. Baldwin and K.B. Clark. *Design rules. Vol. 1: the power of modularity*. MIT Press, 2000.

7. G. Ballarini. Pet therapy. Animals in human therapy. *Acta Biomed Ateneo Parmense*, 74(2):97–100, 2003.

8. C. Bartneck and J. Forlizzi. A design-centred framework for social human-robot interaction. In *13th IEEE International Workshop on Robot and Human Interactive Communication, 2004. ROMAN 2004*, pages 591–594, 2004.

9. C. Bartneck and O. Michio. eMuu-An Emotional Robot. *Demonstration at Robo Festa.*

10. R. Beira, M. Lopes, M. Praça, J. Santos-Victor, A. Bernardino, G. Metta, F. Becchi, and R. Saltarén. Design of the robot-cub (icub) head. In *Proc. IEEE International Conference on Robotics and Automation, ICRA, Orlando.* Citeseer, 2006.

11. D. Benyon and O. Mival. Introducing the companions project: intelligent, persistent, personalised interfaces to the internet. In *Proceedings of the 21st British HCI Group Annual Conference on HCI 2008: People and Computers XXI: HCI... but not as we know it-Volume 2*, pages 193–194. British Computer Society Swinton, UK, UK, 2007.

12. T.J. Bergin Jr and R.G. Gibson Jr. *History of programming languages—II.* ACM New York, NY, USA, 1996.

13. A. Billard. Robota: Clever toy and educational tool. *Robotics and Autonomous Systems*, 42(3-4):259–269, 2003.

14. A. Billard, B. Robins, J. Nadel, and K. Dautenhahn. Building Robota, a Mini-Humanoid Robot for the Rehabilitation of Children with Autism. *The RESNA Assistive Technology Journal*, 2006.

15. G. Bradski and A. Kaehler. *Learning OpenCV: Computer vision with the OpenCV library.* O'Reilly Media, Inc., 2008.

16. C. Breazeal, A. Brooks, J. Gray, G. Hoffman, C. Kidd, H. Lee, J. Lieberman, A. Lockerd, and D. Chilongo. Tutelage and collaboration for humanoid robots. *International Journal of Humanoid Robotics*, 1(2):315–348, 2004.

17. C. Breazeal, A. Edsinger, P. Fitzpatrick, and B. Scassellati. Active vision for sociable robots. *IEEE Transactions on Systems, Man, and Cybernetics, Part A*, 31(5):443–453, 2001.

18. C. Breazeal, G. Hoffman, and A. Lockerd. Teaching and working with robots as a collaboration. In *Proceedings of the Third International Joint Conference on Autonomous Agents and Multiagent Systems-Volume 3*, pages 1030–1037. IEEE Computer Society Washington, DC, USA, 2004.

19. C. Breazeal and J. Velásquez. Toward teaching a robot 'infant' using emotive communication acts. In *Proceedings of the 1998 Simulated Adaptive Behavior Workshop on Socially Situated Intelligence*, pages 25–40, 1998.

20. C.L. Breazeal. *Designing Sociable Robots.* Mit Pr, 2002.

21. R.A. Brooks, C. Breazeal, M. Marjanovic, B. Scassellati, and M.M. Williamson. The Cog project: Building a humanoid robot. *Lecture Notes in Computer Science*, pages 52–87, 1999.

22. H. Bruyninckx. Open robot control software: the OROCOS project. *Language*, 551:14.

23. M. Burch. Animal-assisted therapy and crack babies: A new frontier. *Pet Partners Program: A Delta Society Newsletter*, 1991.

24. J.K. Burgoon, D.B. Buller, and W.G. Woodall. *Nonverbal communication: The unspoken dialogue.* McGraw-Hill New York, 1996.

25. L.D. Canamero and J. Fredslund. How does it feel? emotional interaction with a humanoid lego robot. *Socially Intelligent Agents: The Human in the Loop. Papers from the AAAI 2000 Fall Symposium*, pages 23–28, 2000.

26. L. Cardelli. Can a Systems Biologist Fix a Tamagotchi?

27. J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson, and H. Yan. Embodiment in conversational interfaces: Rea. In *Proceedings of the SIGCHI conference on Human factors in computing systems: the CHI is the limit*, pages 520–527. ACM New York, NY, USA, 1999.

28. J. Cole. *About Face.* Mit Pr, 1998.

29. J.L. Crowley and J. Coutaz. Vision for man-machine interaction. *Robotics and Autonomous Systems*, 19(3):347–358, 1997.

30. K. Dautenhahn. Getting to know each other–artificial social intelligence for autonomous robots. *Robotics and autonomous systems*, 16(2):333–356, 1995.

31. K. Dautenhahn. The art of designing socially intelligent agents: Science, fiction, and the human in the loop. *Applied artificial intelligence*, 12(7-8):573–617, 1998.

32. K. Dautenhahn. Robots as social actors: Aurora and the case of autism. In *Proc. CT99, The Third International Cognitive Technology Conference, August, San Francisco*, pages 359–374, 1999.

33. K. Dautenhahn. Design spaces and niche spaces of believable social robots. In *Proceedings of the International Workshop on Robots and Human Interactive Communication.* Citeseer, 2002.

34. K. Dautenhahn and A. Billard. Games children with autism can play with robota, a humanoid robotic doll. *Universal Access and Assistive Technology*, pages 179–190, 2002.

35. K. Dautenhahn, C.L. Nehaniv, M.L. Walters, B. Robins, H. Kose-Bagci, N.A. Mirza, and M. Blow. KASPAR–A Minimally Expressive Humanoid Robot for Human-Robot Interaction Research.

36. K. Dautenhahn, B. Ogden, and T. Quick. From embodied to socially embedded agents–implications for interaction-aware robots. *Cognitive Systems Research*, 3(3):397–428, 2002.

37. J.P. de Villiers, F.W. Leuschner, and R. Geldenhuys. Centi-pixel accurate real-time inverse distortion correction. In *Proceedings of SPIE*, volume 7266, page 726611, 2008.

38. T. Dekens, M. Demol, W. Verhelst, and F. Beaugendre. Voice Activity Detection based on Inverse Normalized Noise Likelihood Estimation. *CIE 2007*, pages 18–22.

39. IFR Statistical Department. *World Robotics 2008.* 2008.

40. J.H. DiBiase, H.F. Silverman, and M.S. Brandstein. 8 Robust Localization in Reverberant Rooms. *Microphone arrays: signal processing techniques and applications*, page 157, 2001.

41. B.R. Duffy. Anthropomorphism and the social robot. *Robotics and Autonomous Systems*, 42(3-4):177–190, 2003.

42. B.R. Duffy. Fundamental Issues in Affective Intelligent Social Machines. *The Open Artificial Intelligence Journal*, 2(1):21–34, 2008.

43. R. Dunbar. *Grooming, gossip, and the evolution of language.* Harvard Univ Pr, 1998.

44. P. Ekman. Are there basic emotions? *Psychological review*, 99(3):550–553, 1992.

45. P. Ekman and W.V. Friesen. Pictures of facial affect, 1976.

46. P. Ekman and W.V. Friesen. *Facial Action Coding System.* Consulting Psychologists Press, 1978.

47. D.C. Engelbart and W.K. English. A research center for augmenting human intellect. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 395–410. ACM New York, NY, USA, 1968.

48. G. Fairchild, S.H.M. Van Goozen, A.J. Calder, S.J. Stollery, and I.M. Goodyer. Deficits in facial expression recognition in male adolescents with early-onset or adolescence-onset conduct disorder. 2009.

49. O. Faugeras. *Three-Dimensional Computer Vision A Geometric Viewpoint.* MIT press, 2006.

50. T. Field. *Touch.* The MIT Press, 2003.

51. K. Finkenzeller. *RFID handbook: radio-frequency identification fundamentals and applications.* Wiley, 1999.

52. P. Fitzpatrick, G. Metta, and L. Natale. Towards long-lived robot genes. *Robotics and Autonomous Systems*, 56(1):29–45, 2008.

53. J.L. Fleiss et al. Measuring nominal scale agreement among many raters. *Psychological Bulletin*, 76(5):378–382, 1971.

54. T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots: Concepts, design and applications: Technical Report CMU-RI-TR-02-29, 2002.

55. T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3-4):143–166, 2003.

56. M. Fujita. AIBO: Toward the era of digital creatures. *The International Journal of Robotics Research*, 20(10):781, 2001.

57. B. Gates. A robot in every home. *Scientific American Magazine*, 296(1):58–65, 2007.

58. T. Geerinck, H. Sahli, D. Henderickx, I. Vanhamel, and V. Enescu. Modeling Attention and Perceptual Grouping to Salient Objects. In *Attention in Cognitive Systems: International Workshop on Attention in Cognitive Systems, WAPCV 2008 Fira, Santorini, Greece, May 12, 2008 Revised Selected Papers*, page 166. Springer London, Limited, 2009.

59. L. Geppert. Qrio, the robot that could. *IEEE Spectrum*, 41(5):34–37, 2004.

60. M. Ghaffari and L. Ernest. Hall. Robotics and nature, from primitive creatures to human intelligence. *IEEE Intelligent Systems*, 20(11):169–176, 2004.

61. J. Goetz and S. Kiesler. Cooperation with a robotic assistant. In *Conference on Human Factors in Computing Systems*, pages 578–579. ACM New York, NY, USA, 2002.

62. K. Goris. *The Development of the Huggable Robot Probo. On the Hardware Design and Construction.* Phd Thesis, 2009.

63. F. Hegel, C. Muhl, B. Wrede, M. Hielscher-Fastabend, and G. Sagerer. Understanding Social Robots. *Proceedings of the 2009 Second International Conferences on Advances in Computer-Human Interactions*, pages 169–174, 2009.

64. D. Henderickx, K. Maetens, T. Geerinck, and E. Soetens. Modeling the Interactions of Bottom-Up and Top-Down Guidance in Visual Attention. *Lecture Notes In Artificial Intelligence*, pages 197–211, 2009.

65. M.J. Hertenstein, J.M. Verkamp, A.M. Kerestes, and R.M. Holmes. The communicative functions of touch in humans, nonhuman primates, and rats: A review and synthesis of the empirical research. *Genetic, Social, and General Psychology Monographs*, 132(1):5–94, 2006.

66. R. Hirose and T. Takenaka. Development of the humanoid robot ASIMO. *Honda R&D Technical Review*, 13(1), 2001.

67. W. Huitt. Maslow's hierarchy of needs. *Educational Psychology Interactive*, 2004.

68. J. Jackson. Microsoft robotics studio: A technical introduction. *IEEE Robotics & Automation Magazine*, 14(4):82–87, 2007.

69. O. Kalinli and S. Narayanan. A saliency-based auditory attention model with applications to unsupervised prominent syllable detection in speech. *Proc. Interspeech, Antwerp, Belgium*, pages 1941–44, 2007.

70. F. Kanehiro, H. Hirukawa, and S. Kajita. Openhrp: Open architecture humanoid robotics platform. *The International Journal of Robotics Research*, 23(2):155, 2004.

71. K. Kaneko, F. Kanehiro, S. Kajita, H. Hirukawa, T. Kawasaki, M. Hirata, K. Akachi, and T. Isozumi. Humanoid robot HRP-2. In *IEEE International Conference on Robotics and Automation*, volume 2, pages 1083–1090. Citeseer, 2004.

72. N. Kaya and H.H. Epps. Relationship between Color and Emotion: A Study of College Students. *College Student Journal*, 38(3):396–406, 2004.

73. C. Kayser, C.I. Petkov, M. Lippert, and N.K. Logothetis. Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology*, 15(21):1943–1947, 2005.

74. S. Kiesler and J. Goetz. Mental models of robotic assistants. In *Conference on Human Factors in Computing Systems*, pages 576–577. ACM New York, NY, USA, 2002.

75. H. Kozima. Infanoid: An experimental tool for developmental psycho-robotics. In *International Workshop on Developmental Study*. Citeseer, 2000.

76. H. Kozima, M.P. Michalowski, and C. Nakagawa. A Playful Robot for Research, Therapy, and Entertainment. *Int J Soc Robot*, 1:3–18, 2009.

77. M. Kranz, R.B. Rusu, A. Maldonado, M. Beetz, and A. Schmidt. A player/stage system for context-aware intelligent environments. *Proceedings of UbiSys*, 6:17–21.

78. A. Krillov. AForge .NET framework: http://www.aforgenet.com/, 2009.

79. JR Landis and GG Koch. The measurement of observer agreement for categorical data. *Biometrics*, 33(1):159, 1977.

80. S. Lawson and T. Chesney. The reign of catz & dogz at CHI 2009. In *Proceedings of the 27th international conference extended abstracts on Human factors in computing systems*, pages 4787–4790. ACM New York, NY, USA, 2009.

81. H.S. Lee, J.W. Park, and M.J. Chung. An Affect-Expression Space Model of the Face in a Mascot-Type Robot. In *Proc. of IEEE/RAS Int. Conf. on Humanoid Robots, to be published*, 2006.

82. A.V. Libin and E.V. Libin. Person–Robot Interactions From the Robopsychologists' Point of View: The Robotic Psychology and Robotherapy Approach. *Proceedings of the IEEE*, 92(11):1789–1803, 2004.

83. E. Libin and A. Libin. New diagnostic tool for robotic psychology and robotherapy studies. *CyberPsychology & Behavior*, 6(4):369–374, 2003.

84. R. Lienhart and J. Maydt. An extended set of haar-like features for rapid object detection. In *IEEE ICIP*, volume 1, pages 900–903, 2002.

85. J. Lowy. *Programming .NET components.* O'Reilly Media, Inc., 2005.

86. W. Luo. DockPanel Suite: http://sourceforge.net/projects/dockpanelsuite/, 2006.

87. K.F. MacDorman. Subjective ratings of robot video clips for human likeness, familiarity, and eeriness: An exploration of the uncanny valley. In *ICCS/CogSci-2006 Long Symposium: Toward Social Mechanisms of Android Science*, 2005.

88. D. Matsui, T. Minato, K.F. MacDorman, and H. Ishiguro. Generating natural motion in an android by mapping human motion. In *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2005.

89. M. Max-Neef. Development and human needs. *Real-Life Economics: understanding wealth creation, edited by: Ekins, P. and Max-Neef, M., Routledge, London*, pages 197–213, 1992.

90. S. McCloud. Understanding comics: The invisible art. *New York: HarperPerennial*, 1994.

91. A. Mehrabian. Communication without words. *Psychology Today*, 2(4):53–56, 1968.

92. G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The iCub humanoid robot: an open platform for research in embodied cognition.

93. N. Mitsunaga, T. Miyashita, H. Ishiguro, K. Kogure, and N. Hagita. Robovie-iv: A communication robot interacting with people daily in an office. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 5066–5072, 2006.

94. H. Miwa, K. Itoh, M. Matsumoto, M. Zecca, H. Takanobu, S. Roccella, M.C. Carrozza, P. Dario, and A. Takanishi. Effective emotional expressions with emotion expression humanoid robot WE-4RII. In *IROS, 2004*.

95. H. Miwa, T. Okuchi, K. Itoh, H. Takanobu, and A. Takanishi. A new mental model for humanoid robots for human friendly communication introduction of learning system, mood vector and second order equations of emotion. In *IEEE International Conference on Robotics and Automation, 2003. Proceedings. ICRA'03*, volume 3, 2003.

96. H. Mobahi. Building an Interactive Robot Face from Scratch. *Bachelor of Engineering Final Project Report, Azad University, Tehran-South Campus, Tehran, Iran, Maggio*, 2003.

97. M. Mori. *The Buddha in the Robot*. Tuttle Publishing, 1981.

98. M.E. Munich, J. Ostrowski, and P. Pirjanian. ERSP: A software platform and architecture for the service robotics industry. In *Proc. IEEE International Conference on Intelligent Robots and Systems*, 2005.

99. C. Nass, J. Steuer, and E.R. Tauber. Computers are social actors. In *Proceedings of the SIGCHI conference on Human factors in computing systems: celebrating interdependence*, pages 72–78. ACM New York, NY, USA, 1994.

100. J. Nolte and J. Sundsten. *The human brain: an introduction to its functional anatomy*. Mosby St. Louis, 2002.

101. J.H. Oh, D. Hanson, W.S. Kim, I.Y. Han, J.Y. Kim, and I.W. Park. Design of android type humanoid robot albert HUBO. In *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1428–1433, 2006.

102. H.G. Okuno, K. Nakadai, and H. Kitano. Social interaction of humanoid robot based on audio-visual tracking. *Lecture notes in computer science*, pages 725–735, 2002.

103. A. Ortony, D.A. Norman, and W. Revelle. 7 Affect and Proto-Affect in Effective Functioning. *Who Needs Emotions?: The Brain Meets the Robot*, 2005.

104. Y. Patsis and W. Verhelst. A Speech/Music/Silence/Garbage/Classifier for Searching and Indexing Broadcast News Material. In *Database and Expert Systems Application, 2008. DEXA'08. 19th International Conference on*, pages 585–589, 2008.

105. Perception. Science Encyclopedia: http://science.jrank.org/pages/5094/Perception.h 2009.

106. R.W. Picard. *Affective computing*. The MIT Press, 2000.

107. C. Plaisant, A. Druin, C. Lathan, K. Dakhane, K. Edwards, J.M. Vice, and J. Montemayor. A storytelling robot for pediatric rehabilitation. In *Proceedings of the fourth international ACM conference on Assistive technologies*, pages 50–55. Citeseer, 2000.

108. R. Plutchik. A general psychoevolutionary theory of emotion. *Theories of emotion*, 1, 1980.

109. J. Posner, J.A. Russell, and B.S. Peterson. The circumplex model of affect: An integrative approach to affective neuroscience, cognitive development, and psychopathology. *Development and Psychopathology*, 17(03):715–734, 2005.

110. G.A. Pratt and M.M. Williamson. Series elastic actuators. In *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS-95)*, volume 1, pages 399–406, 1995.

111. B. Reeves and C. Nass. *The media equation: How people treat computers, television, and new media like real people and places.* Cambridge University Press New York, NY, USA, 1996.

112. W.S.N. Reilly. *Believable social and emotional agents.* PhD thesis, University of Birmingham, 1996.

113. RoboRealm. Robotic Machine Vision Software: http://www.roborealm.com/, 2009.

114. Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki, K. Fujimura, H.R.D.C. Ltd, and J. Saitama. The intelligent ASIMO: System overview and integration. In *IEEE/RSJ International Conference on Intelligent Robots and System, 2002*, volume 3, 2002.

115. J. Saldien. Probo: a friend for life. Master's thesis, Artesis Hogeschool Antwerpen, 2005.

116. T. Sawada, T. Takagi, and M. Fujita. Behavior selection and motion modulation in emotionally grounded architecture for QRIO SDR-4XII. In *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004.(IROS 2004). Proceedings*, volume 3.

117. M. Scheeff, J. Pinto, K. Rahardja, S. Snibbe, and R. Tow. Experiences with Sparky, a social robot. In *Proceedings of the Workshop on Interactive Robot Entertainment*. Springer, 2000.

118. H. Schlosberg. Three dimensions of emotion. *Psychological review*, 61(2):81–88, 1954.

119. S. Shamma. On the role of space and time in auditory processing. *Trends in Cognitive Sciences*, 5(8):340–348, 2001.

120. N. Sharkey. COMPUTER SCIENCE: The Ethical Frontiers of Robotics. *Science*, 322(5909):1800, 2008.

121. T. Shibata, T. Mitsui, K. Wada, A. Touda, T. Kumasaka, K. Tagami, and K. Tanie. Mental commit robot and its application to therapy of children. *Advanced Intelligent Mechatronics, 2001. Proceedings. 2001 IEEE/ASME International Conference on*, 2, 2001.

122. T. Shibata, K. Wada, T. Saito, and K. Tanie. Robot Assisted Activity for Senior People at Day Service Center. *Proceedings of International Conference on Information Technology in Mechatronics*, pages 71–76, 2001.

123. N. Shimizu, N. Koizumi, M. Sugimoto, H. Nii, D. Sekiguchi, and M. Inami. A teddy-bear-based robotic user interface. *Computers in Entertainment (CIE)*, 4(3), 2006.

124. S. Shiotani, T. Tomonaka, K. Kemmotsu, S. Asano, K. Oonishi, and R. Hiura. World's First Full-fledged Communication Robot" wakamaru" Capable of Living with Family and Supporting Persons. *Mitsubishi Juko Giho*, 43(1):44–45, 2006.

125. Skilligent. Skilligent Robot Vision System: http://www.skilligent.com/products/computer-vision.shtml, 2009.

126. S. Sosnowski, A. Bittermann, K. Kuhnlenz, and M. Buss. Design and Evaluation of Emotion-Display EDDIE. *Intelligent Robots and Systems, 2006 IEEE/RSJ International Conference on*, pages 3113–3118, 2006.

127. W.D. Stiehl and C. Breazeal. Affective Touch for Robotic Companions. In *First International Conference on Affective Computing and Intelligent Interaction (Currently Under Review)*. Springer, 2005.

128. W.D. Stiehl and C. Breazeal. A Sensitive Skin for Robotic Companions Featuring Temperature, Force, and Electric Field Sensors. In *Proceedings of the 2006 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2006)*, 2006.

129. WD Stiehl, J. Lieberman, C. Breazeal, L. Basel, L. Lalla, and M. Wolf. Design of a therapeutic robotic companion for relational, affective touch. *Robot and Human Interactive Communication, 2005. ROMAN 2005. IEEE International Workshop on*, pages 408–415, 2005.

130. F. Thomas and O. Johnston. *Disney animation: The illusion of life.* Abbeville Press New York, 1981.

131. AJN Van Breemen. Bringing robots to life: Applying principles of animation to robots. In *Proceedings of Shapping Human-Robot Interaction workshop held at CHI.* Citeseer, 2004.

132. AJN van Breemen. iCat: Experimenting with Animabotics. *Proceedings, AISB 2005 Creative Robotics Symposium*, 2005.

133. J. Vanderfaeillie and Vandenplas Y. *Handboek orthopedagogische hulpverlening. Deel 1. Een orthopedagogisch perspectief op kinderen en jongeren met problemen*, chapter Zieke kinderen en jongeren, pages 199–237. Leuven/Voorburg: Acco, 2005.

134. P. Viola and M. Jones. Rapid Object Detection using a Boosted Cascade of Simple. In *Proceedings of CVPR2001*, volume 1, 2001.

135. H. Vogt. Efficient object identification with passive RFID tags. *Lecture Notes in Computer Science*, pages 98–113, 2002.

136. K. Wada, T. Shibata, T. Saito, and K. Tanie. Analysis of factors that bring mental effects to elderly people in robot assisted activity. In *IEEE/RSJ International Conference on Intelligent Robots and System, 2002*, volume 2, 2002.

137. K. Wada, T. Shibata, T. Saito, and K. Tanie. Effects of robot-assisted activity for elderly people and nurses at a day service center. *Proceedings of the IEEE*, 92(11):1780–1788, 2004.

138. M. Wassink and S. Stramigioli. Towards a novel safety norm for domestic robotics. 2007.

139. I. Werry, K. Dautenhahn, B. Ogden, and W. Harwin. Can social interaction skills be taught by a social agent? The role of a robotic mediator in autism therapy. *Lecture Notes in Computer Science*, pages 57–74, 2001.

140. DM Wilkes, A. Alford, RT Pack, T. Rogers, RA Peters, and K. Kawamura. Toward socially intelligent service robots. *Applied Artificial Intelligence*, 12(7):729–766, 1998.

141. Y. Wilks. Artificial companions. In *InSTIL/ICALL Symposium 2004*. Springer, 2004.

142. D.E. Wilson and D.A.M. Reeder. *Mammal Species of the World: A Taxonomic and Geographic Reference*. Johns Hopkins University Press, 2005.

143. HA Yanco and J. Drury. Classifying human-robot interaction: an updated taxonomy. In *2004 IEEE International Conference on Systems, Man and Cybernetics*, volume 3, 2004.

144. S. Yilmazyildiz, W. Mattheyses, Y. Patsis, and W. Verhelst. Expressive Speech Recognition and Synthesis as Enabling Technologies for Affective Robot-Child Communication. *Lecture Notes in Computer Science*, 4261:1, 2006.

145. S. Yohanan and K.E. MacLean. The Haptic Creature Project: Social Human-Robot Interaction through Affective Touch. In *Proceedings of the AISB 2008 Symposium on the Reign of Catz & Dogs: The Second AISB Symposium on the Role of Virtual Creatures in a Computerised Society*, volume 1, pages 7–11, 2008.

146. S.Y. Yoon, B.M. Blumberg, and G.E. Schneider. Motivation driven learning for interactive synthetic characters. In *Proceedings of the fourth international conference on Autonomous agents*, pages 365–372. ACM New York, NY, USA, 2000.

A new generation of robots is being created to live among humans and become a nearly ubiquitous part of our day-to-day lives. These robots will be better accepted if they measure up to a certain standard of social interaction and human-like communication, achieving the label of "social robot". No robot today can fulfill the role of a fully social interactive robot, but new research domains such as HRI (human-robot interaction) are gaining vastly more interest. In this context there is a strong need for robotic platforms that support this HRI research.

This dissertation presents the design and software development of a huggable social robot named Probo. Probo embodies a stuffed imaginary animal, providing a soft touch and a huggable appearance. Probo's purpose is to serve as a multidisciplinary research platform for HRI focused on children. In terms of a social robot, Probo is classified as a social interface supporting non-verbal communication. Probo's social skills are thereby limited to a reactive level. To close the gap with higher levels of interaction, an innovative system for shared control with a human operator is introduced. The software architecture defines a modular structure to incorporate all systems into a single control center. The robot reacts on basic input stimuli that it perceives during interaction with children. These stimuli will influence the robot's attention and emotional state, which are communicated by its gaze and facial expressions. To facilitate interaction with children, Probo has an identity and corresponding history. Safety is ensured through Probo's soft embodiment and intrinsic safe actuation systems. Smooth life-like motions contribute to the huggable robotic companion that Probo has become.

Jelle Saldien