Interacting With a Ball-Playing Entertainment Robot

Thesis

Submitted for the degree of
Bachelor of Science (B.Sc.)

at the

University Bremen
Digital Media

1. Examiner:	Udo Frese
2. Examiner:	Robert Porzel
Submitted by:	Laura Martha Spillner
Matriculation number:	4018535
Date of submission:	01.10.2018

Abstract

The robot Doggy is able to play a simple ball game with a human. In the scope of this thesis, the robot's interactive behaviour was expanded to perform emotional gestures (both during and outside of the ball game) based on its internal emotional state. This behaviour design was implemented both as a simulation software and on the robot itself. It was tested on the robot and evaluated in terms of how it impacted the test participants' interaction with and perception of the robot: Emotional expression was experienced as a positive addition to the interaction; but while the participants were generally able to easily understand the robot's gestures, the number of distinct expressions that could be distinguished was limited.

Contents

Chapter 1

Introduction

1.1 Motivation

When thinking of ball-playing entertainment robots, one of the first things that might come to mind is the RoboCup, the global robot football tournament. Its goal is to promote research in AI and robotics through its publicity (RoboCup Federation Inc, 2016), and it has indeed become an entertaining public event not only for researchers, but also for hobbyists and others who were not previously familiar with robotic sports. However, as Laue et al. (2014) discuss in “An Entertainment Robot for Playing Interactive Ball Games”, RoboCup is still subject to certain limitations, because it continues to be mainly a research-driven endeavour. While it is able to stimulate more research in the area of sport and entertainment robotics, it does not at this point produce directly commercially viable, flexible solutions which can be employed for example as entertainment devices (that is, maybe apart from Sony’s Aibo, the soccer playing robot dog which enjoyed some popularity as a kind of household entertainment item) (Laue et al., 2014, p. 171). Mainly hindering in this is that the rather complex mechanisms used in these robots depend on a rigidly defined environment, and are not flexible enough to adapt to changing surroundings, with varying colours, lighting, and people interacting with them in an unpredictable manner.

This was the motivation for the Multi-Sensor Interactive Systems Group of the University Bremen in developing the robot Piggy (Laue et al., 2014), which would serve as an entertainment robot able to play interactive ball games with people. The goal of this project was to build a "direct, commercially realistic application of sport robotics technology in the entertainment industry", a minimalist system which would be able to "[serve] at events, such as office parties, fairs or open house presentations" (Laue et al., 2014, p. 171). Piggy, the robot that was developed, served with great success as an exhibition at the CeBIT computer fair in Hanover in 2012, as well as at the Open Campus event at the University Bremen.

Piggy's main functionality was this: A person throws a ball, the robot detects the flight path of that ball, and moves a kind of bat (which comprises the main upper body of the robot) in order to intercept the ball, hitting it back to the player. Since this was supposed to be a commercially viable as well as flexible system, able to be used in any environment, it was subject to a number of further requirements: The robot needed to be safe for lots of people to interact with it (especially seeing how there would be a large number of people interacting with it outside of a carefully controlled manner of use), be flexible in term of changes in the environment, light and background (which would interfere with the recognition of the ball), have reasonably low costs, and be fast and reactive enough to be able to work with a large number of ever-changing players without the game becoming monotonous.

The robot that was developed consisted of a body (which housed its actuators: controls and motors; as well as sensory equipment such as cameras) and, connected to this, a bat - a rod with a Styrofoam sphere at its end. This sphere is manoeuvred so that it meets the ball thrown by the player. In its first iteration, this system was "dressed" as a highly stylized blue pig (maybe for the reason that the game idea drew some inspiration from the children's game "Schweinchen in der Mitte"): a large round head serving as the body of the robot (with eyes where the cameras were situated), and a white hat housing the bat. The second iteration of the robot's outer design, which was developed in *New Dog, New Tricks - Human-Robot Interaction and Appearance Design of a Ball Playing Robot* by Tzeng (2013), was changed to the body of a stylized plush dog, and henceforth called Doggy. Here, the upper body of the dog is the bat. Its head is the sphere at the top of the bat, so that Doggy hits balls with its head. Added to this was a tail, which can be moved independently of the body. Pictures of both can be seen in Chapter 2.1.

Piggy, in its first iteration, was already able to express a number of simple gestures, dependent on a few states. Its reiteration in Doggy expanded on that, adding a wider range of possible gestures, and making use of a higher number of degrees of freedom and the addition of the movable tail. Furthermore, Tzeng (2013) developed a software that would allow the design and visualization of new gestures - however, this software is unfortunately no longer usable at this time. In addition to this, Bartsch (2015) designed a system for Doggy to hear and react to sounds played in its vicinity, acting depending on where a sound comes from, which opened up new possibilities to integrate this functionality in interactive behaviour and maybe more complicated games.

1.2 Aim of This Thesis

The goal of this thesis is to build on the existing system and its previous extensions, in order to expand the interactive behaviours Doggy exhibits.

- The first step is the design of Doggy's interactive behaviour. Doggy's actions are based on a behaviour loop: a finite state machine controls the robot's behaviour, looping through two base states as long as the robot is not interacted with, reacting after a ball is played, and returning to its idle state afterwards (Tzeng, 2013). Doggy's new interactive design will include the expansion of the existing state machine, as well as a set of gestures that Doggy can play both when idle, and when interacting with a human player (based both on existing behaviours from the work of Laue et al. (2014), as well as the gestures that were planned by Tzeng (2013) when the robot was redesigned as Doggy). One particular aspect will be the reaction to sound, the recognition of which was implemented by Bartsch (2015). Ideal would be the design of other games in addition to the existing ball game, or at least other means of interacting with the robot apart from throwing a ball.
- The second goal is to develop software that will allow the visualization of this extended behaviour and the new gestures. This software can then be used to evaluate the interactive behaviour on how natural the gestures look without having to test directly on the robot. In this way, small changes can be quickly tested and previewed, in order to optimize the gestures before implementing them in the robot's software.
- The third step is the integration of the developed behaviour into the existing system on the robot, in order to test it with real human players. The current working system should be able to work with the new gestures, as well as with the modules for sound recognition.
- If that is achieved, the behaviour will be evaluated in a user test on the actual robot, by designing a test scenario with which to assess the impact of the interactive behaviour on users' enjoyment of game play and interaction with the robot.

Chapter 2

Previous and Related Work

2.1 Piggy and Doggy

2.1.1 The First Generation: Piggy

Piggy was the first iteration of the robot, its conception and implementation are detailed in "An Entertainment Robot for Playing Interactive Ball Games" (Laue et al., 2014). As was described in Chapter 1.2, the character Piggy was the highly stylized head of a blue pig, which used a white, pointed hat as the bat.

Figure 2.1: The first design of the robot's exterior, as shown in Tzeng (2013).

The blue head of the pig houses the computer and power supply. At the junction between this and the bat (the pig's white hat), two servo motors control the bat's movement. Special consideration in designing the robot was taken to lower both the overall cost, as well as the weight of the finalized construction:

> "The key idea in our robot design is to use as few degrees of freedom (DOF) as possible. The design of

> conventional 6 DOF robots is dominated by the fact that lower motors have to support and accelerate the weight of higher motors. Avoiding this set-up dramatically reduces weight and costs and increases safety and reactivity." (Laue et al., 2014, p.173)

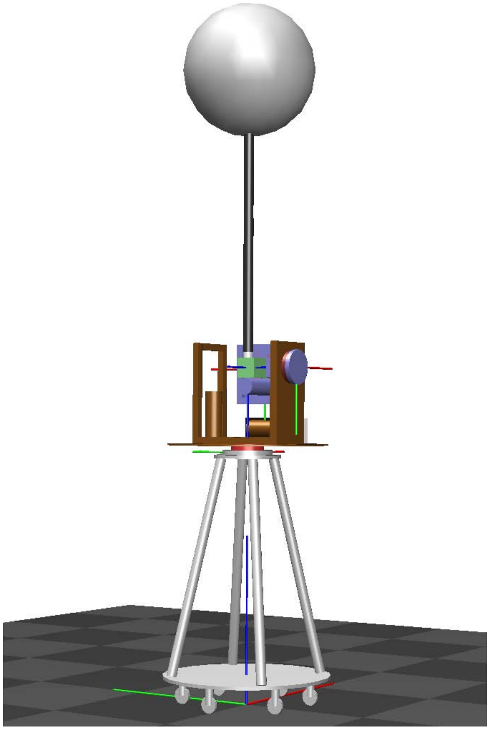


Figure 2.2: A model of the robot's interior, as shown in Schuethe and Frese (2014).

Because of this, the robot was designed with only two degrees DOF, which is enough to play a basic ball game. Piggy could move its bat forwards and backwards, and turn it from side to side. This allowed it to reach any point on a sphere (or, in practice, the front half of that sphere). Since at the end of the bat there is a ball, it is not necessary to control the orientation of the bat (Laue et al., 2014, p.5). Therefore it could always meet the ball as long as it passed within its reach.

Even in this first iteration, Piggy offered a considerable amount of interactive behaviour. It was able to play several gestures, which were controlled in a simple finite state machine: The robot would start out sleeping, and could be awoken by a person manually lifting the bat, which made it transition into a waiting state. It would play a cheering gesture, in order to motivate people to interact with it, if no ball had been thrown in a certain amount of time. When a ball had been thrown badly, so that it passed outside the robot's reach, Piggy would play a complaining gesture. Of course, any gesture would be interrupted in order to hit a ball when one was thrown.

2.1.2 Improving Piggy: The Creation of Doggy

In Tzeng (2013), the physical appearance of the robot was changed to the one it is now, Doggy, and the robot gained a few additional features.

For the design of this version, the body of the stylized blue pig was replaced by that of a simplified plush dog. Its lower body, which housed the internals of the robot and up to this point used to be Piggy's head, is now just a simple cylinder, while the upper body encases the bat. It ends in the sphere that is used to intercept balls, which is designed to be Doggy's head: This results in a ball game in which the robot dog plays the ball back to the player with its head.

Doggy has three DOF where Piggy only had two: It can still turn from side to side, it can bend down forwards or backwards, and can also bend down to either side. This redundancy allows for two things: Firstly, it is possible to reach a required position in more than one way, and therefore potentially faster than before. And secondly, it allows for more distinct gestures (since it is now Doggy's body that moves and it can be more subtly controlled), which achieves more varied and natural movement and as such more possibilities for interaction. Furthermore, Doggy was given a tail, which can be moved independently from the body. This has two DOF as well, it can be moved up or down, as well as to the left or right. It is fixed to the body in such a way that it turns with the rest of the body.

Figure 2.3: Front view of Doggy, as shown in Tzeng (2013).

As a whole, Doggy stands at approx. 2 m tall, and bends approx. 1 m from the ground. The head has a radius of 20 cm. Since Doggy is taller than most humans, it is important to design the gestures in such a way that they are not too rapid or 'jerky', because the robot can seem quite imposing, especially when it moves quickly.

Similar to Piggy before it, Doggy was, as is detailed in Tzeng (2013), already able to play a number of gestures, switching between some general states: It played a standby gesture when there was no interaction, a happy gesture after successfully hitting a ball, or a sad one after failing to do so. These gestures were designed to make use of the added possibilities Doggy offers. As such, there were a number of dancing gestures, rotating the body or moving it in a zigzag motion. Additionally, Tzeng (2013) provided software that could be used to design new gestures in a 2D-animated editor. Unfortunately, this software does not work any more at this point, making it necessary to either rework it or start anew. Although the behaviour based on the gestures developed by Tzeng (2013) is not integrated any more in Doggy's current state, the files that describe these motions can still be used as a starting point for the current project.

2.2 Related Work

First of all, it is important to examine to what extend gestures and emotional or social expression are important in human-computer-interaction (HCI) or, in this case, human-robot-interaction (HRI), and how gestures change the way in which humans interact with robots.

Figure 2.4: Side view of Doggy, showing its tail, as shown in Tzeng (2013).

There have been projects similar to what is hoped to achieve with Doggy: In "Playing catch with robots: Incorporating social gestures into physical interactions", Carter et al. (2014) investigate the effects of social gestures in a somewhat similar kind of ball game between a human player and a robot. Kerepesi et al. (2006) explore interaction between humans and a dog-like robot in game play in their study "Behavioral comparison of human-animal (dog) and human-robot (AIBO) interactions". Carter et al. (2014) show that "social gesturing of a robot enhances physical interactions between humans and robots" (Carter et al., 2014, p. 231), as has been suggested by previous studies that they reference (Carter et al., 2014, p. 232). They reason that this is due to the fact that communication in humans (and, in extension, between robots and humans) happens not only through speech but also through bodily gestures (Carter et al., 2014, p. 231), and that the robot's gestures positively affect the way it and the game are perceived by the study participants. Kerepesi et al. (2006) had their study participants interact with the (rather famous, not only because it is also used in robot football) entertainment robot

AIBO, which is like Doggy modelled after a dog, but much closer to an actual dog both in appearance and movement. Their research shows that humans' interaction with a dog-like robot is fundamentally very similar to that between a human and a puppy (Kerepesi et al., 2006, p. 97), which again lends credence to the idea that emotional expression and communication through non-verbal gestures should be an important part of HRI.

At this point, it is expedient to take a step back and look into why humans tend to react positively and naturally to emotional and gesture expression by a robot; and survey the theoretical background of what constitutes a *social robot*, existing frameworks for how its expression should be realized, and how it can be used to complement existing HRI (like Doggy's ball game):

As Fong et al. (2003) state in "A Survey of Socially Interactive Robots", "[t]he common, underlying assumption is that humans prefer to interact with machines in the same way that they interact with other people" (Fong et al., 2003, pp. 145-146). Breazeal (2003) laid important groundwork in studying this phenomenon in "Toward sociable robots". In exploring this topic, she distinguishes four classes of what she defines as social robots and explains how those differ from non-social ones. A case study was done using the robot Kismet, which interacts with humans through "vocal turn-taking behaviour" (Breazeal, 2003, p. 167). Breazeal stresses the importance of understanding humans' characteristically being a "profoundly social species", because, as she explains: "Our social-emotional intelligence is a useful and powerful means for understanding the behaviour of, and for interacting with, some of the most complex entities in our world - people and other living creatures" (Dennett, 1989 as cited in Breazeal, 2003, pp. 167-168). Human means of social and emotional communication have to be taken into account when designing robots, because humans tend to apply their social models not only when communicating with other humans, but also when faced with a robot or otherwise sufficiently complex entity, in order to "explain, understand, and predict their behaviour as well" (Reeves and Nass, 1996 as cited in Breazeal, 2003, p. 168). Similarly, some socially interactive robots also model their human partner's emotional state and are actively trying for social interaction; others, however, can be perceived as being socially interactive solely through humans attributing to them mental states and emotions that are not actually implemented (Fong et al., 2003, pp. 145-147). Fong et al. (2003) also provide in their paper a comprehensive overview concerning existing definitions of social robots, starting with those of Breazeal as well as of Dautenhahn and Billard, who define them as "embodied agents that are part of a heterogeneous group: a society of robots or humans" (Dautenhahn and Billard, 1999 as cited in Fong et al., 2003, p. 144). Based on that, they develop their own definition of the term socially interactive robots, more generally as “robots for which social interaction plays a key role” (Fong et al., 2003, p. 145) (in contrast to robots that interact with humans in the more general realm of HRI, but do not rely on any kind of social part in our understanding of that interaction). A closely related term is used by Shibata (2004): Human interactive robots, which they specify as being designed and used for psychological enrichment, are “a type of service robot that provide a service by interacting with humans while stimulating their minds and we, therefore, tend to assign high subjective values to them” (Shibata, 2004, p. 1750).

Other works concentrate on specific aspects of how emotions can be formalized and simplified for robots’ interaction and internal representation. Dryer (1998) proposes a two-factor model for emotion in HCI of Dominance and Valence, which is based on psychological research. He tested his model in a study which specifically examined human participants’ emotions when interacting with a computer game. Their results further support the underlying assumption that “human-computer interaction may be fundamentally similar to interpersonal interaction” (Dryer, 1998, p. 79), which in this case specifically means that they were able to model it in the dimensions of dominance and valence. Another approach is that of Moshkina and Arkin (2005), who developed a framework they call TAME (Traits, Attitudes, Moods and Emotions). They again underline how humans will apply their social models to non-human entities (including both animals and objects like robots or even inanimate things). They base their ideas also on the work of Nass et al. (1997) (as cited in Moshkina and Arkin, 2005), whose research established humans’ propensity to treat computers, robots and other objects as “social actors” - “even [with] minimal cues [evoking] social responses” (Moshkina and Arkin, 2005, p. 1444).

In addition to these, other papers that were consulted are directly concerned with game play between humans and robots, in a variety of different contexts. While this work on Doggy is aiming to design gestures for robots that can be understood by human players, Brooks et al. (2004) worked on teaching robots to recognize gestures and emotions for communication and understanding new games. They mention why game playing in particular can be useful in computer research and development: “Games have always been primary motivators of computer-generated entertainment. Indeed, game development and computer architecture exhibit something akin to a bidirectional symbiosis” (Brooks et al., 2004, p. 1). Similar work was done by Liu et al. (2006), who also focus on robots understanding human emotions and affective cues. They stress how “[a]s robots and people begin to co-exist and cooperatively share a variety of tasks, ‘natural’ human-robot interaction with an implicit communication channel and a degree of emotional intelligence is becoming increasingly important" (Liu et al., 2006, p. 285). An altogether different approach is taken by Robins et al. (2010). Their work concerns robot-assisted play for children with special needs, but this paper is especially interesting for its detailed description of the design process (Robins et al., 2010, pp. 876-878, 882), as well as its literature review on robot systems and robot play. Furthermore, this work also showcases another potential use for socially emotive robots and robots as a kind of social teaching toy. Of tangential interest were other papers that were consulted for their work on catching or ball games with robots, similar to what is possible with Doggy. Not emotional gestures in addition to game play, but specifically natural movement while playing games, is, for example, the focus of Riley and Atkeson (2002).

Chapter 3

Interaction Design

3.1 Previous Interactive Animations

Doggy's predecessor Piggy was already equipped with a number of interactive behaviours, as is described in Laue et al. (2014), in order to "give users some kind of appealing feedback as well as the impression of an intelligent robot" (Laue et al., 2014, p. 179). These behaviours were formalized in a finite state machine, as can be seen in Figure 3.1. The robot's default state is a *sleeping* state, from which it will play a ball if one is thrown, *cheer* if nothing happens for a set amount of time, or *complain* if a ball was thrown but did not reach it's workspace so that it could not be played back.

Both the *cheering* and the *complaining* state involved the robot moving its bat and playing an accompanying sound.

Later, two additional behaviours were added: a reaction to when the robot is hit or pushed by a person (consisting of an 'Ouch!' sound), as well as the ability to follow the player with its bat while waiting, in order to convey attention, by using a Microsoft Kinect.

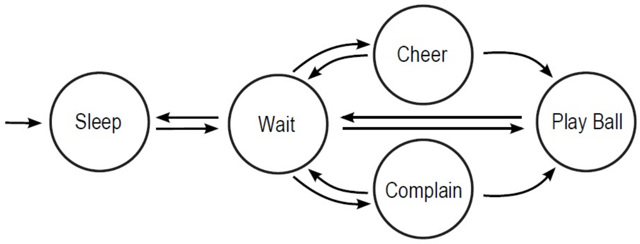


Figure 3.1: The state design of the original robot Piggy, as shown in Laue et al. (2014).

When Piggy was redesigned as Doggy in Tzeng (2013), this state machine was further expanded upon, which can be seen in Figure 3.2.

This iteration distinguished between three types of gestures: *standby* (which was played while no balls were thrown), *happy* (which was played after a ball was thrown and hit) and *sad* (which was played after a ball was thrown but could not be hit). These gestures consisted of a combination of torso movement and tail movement (Tzeng, 2013, p. 30), of both of which there were several different variants. Torso movement consisted of moving the torso back and forth at different speeds, depending on whether the gesture was supposed to be happy or sad; while tail movements emulated the way a dog might wag its tail.

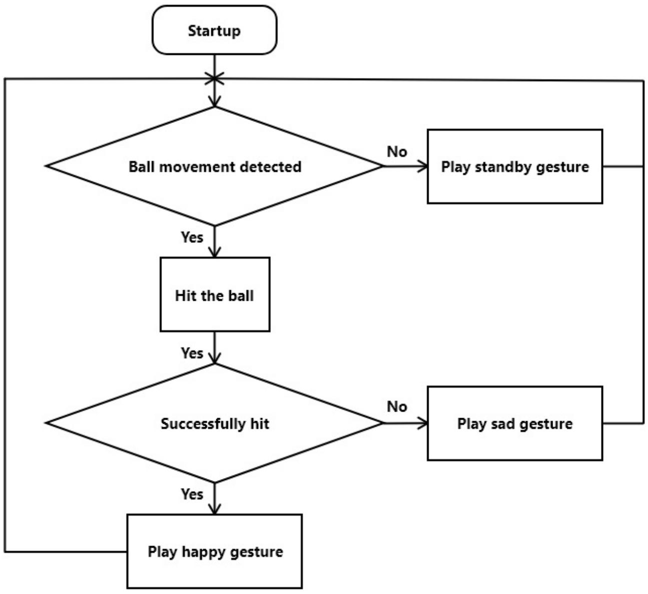


Figure 3.2: The previous state design of the robot Doggy, as explained in Tzeng (2013).

3.2 Theoretical Background

As was discussed in Chapter 2, there is already a lot of research that supports how important gestures and emotional communication are for HRI. As Carter et al. (2014) state: “Emblematic gestures convey information without the need for concurrent speaking; cultures have agreed-upon meanings for these specific motions. Common emblems include head nods, hand waves, and shoulder shrugs” (Kendon, 1992 as cited in Carter et al., 2014, p. 231). There are many gestures that have certain commonly understood meanings; however, it is often difficult to exactly define the specific meaning of one movement, even though we can easily understand the intended meaning intuitively. One goal of this work is to design a number of new gestures for Doggy, that will both be useful for communication in a new interactive game, and can be easily understood despite Doggy’s limited range of motion. It is therefore important to more closely examine how specific robotic gestures can be understood by humans, what concepts already exist for the design of sociable/interactive/gesturing robots, and how these can be implemented on Doggy specifically.

In “Playing catch with robots: Incorporating social gestures into physical interactions”, Carter et al. (2014) explain:

> “Previous research has investigated how physical movements, including gestures, change viewer perception of synthetic agents. Humans tend to respond socially to non-human objects like computers, and this can be generalized to robots. Moreover, the level of social response can be manipulated by agent characteristics. One goal of human-robot and human-computer interaction is to determine how to design robots and agents that respond to humans in a social way, assuming it is appropriate given the context.” (Reeves and Nass, 1996; Dautenhahn, 2007 and Goetz et al., 2003 as cited in Carter et al., 2014, pp. 231 - 232)

In their research, Carter et al. (2014) used physical gestures to denote the robot’s disappointment when it had not been able to catch a ball thrown to it. Their robot’s design allowed for several humanoid gestures, including shrugging its shoulders, shaking its head, and throwing up its hands in apparent frustration. They found that “the incorporation of these gestures made the throwing and catching interaction more positive for the human participants. The survey results indicated that gestures made the robot seem more human-like, engaging and responsive, presumably because it demonstrated awareness of failure. The gestures also elicited increased smile scores from the participants, implying that the gestures successfully conveyed an appropriate and somewhat humorous acknowledgement of failure by the robot” (Carter et al., 2014, p. 235).

There are different ways to approach the design of a robot that is to “respond to humans in a social way” (Carter et al., 2014, p. 232), particularly when it comes to how the recognition of the human partner’s social cues and the choice of answer for the robot will be represented internally. Breazeal (2003) coins the term *social robot*, which she defines as such:

> “We argue that people will generally apply a social model when observing and interacting with autonomous robots. Autonomous robots perceive their world, make decisions on their own, and perform coordinated actions to carry out their tasks. As with living things, their behaviour is a product of its internal state as well as physical laws. Augmenting such self-directed, creature-like behaviour with the ability to communicate with, cooperate with, and learn from people makes it almost impossible for one to not anthropomorphise them (i.e. attribute

> human or animal-like qualities). We refer to this class of autonomous robots as social robots, i.e., those that people apply a social model to in order to interact with and to understand." (Breazeal, 2003, p. 168)

In this class of social robots, there are still several distinctions to be made - for example, to what extent the robot's internal design supports the social model that a human will apply to it and how it is mapped internally, as well as to what extent the robot itself understands the humans reaction and applies a social model to that as well. Breazeal defines four subclasses of social robots, which differ in the sophistication of their "interaction scenario" (Breazeal, 2003, p. 169):

- **Socially evocative:** The robot "is designed to encourage people to anthropomorphise the technology in order to interact with it, but goes no further", "the human attributes social responsiveness to the robot, but the robot's behaviour does not actually reciprocate".
- **Social interface:** It "uses human-like social cues and communication modalities in order to facilitate interactions with people" (which can be e.g. a robot avatar, an interface model which conveys information like museum tour guide); the robot "[performs] a person's message to others"; it uses "pre-canned social behaviour".
- **Socially receptive:** The robot also "benefit[s] from interactions with people" (i.e., learning); but is "socially passive" (there is no "pro-active engaging").
- **Sociable:** The robots are "socially participative 'creatures' with their own internal goals and motivations"; they "pro-actively engage people in a social manner not only to benefit the person (...) but also to benefit itself (...)".

(Breazeal, 2003, p. 169)

These distinctions show that even without an internal model to interpret human behaviour and react accordingly, a robot can still act in a social manner simply by the social characteristics that a human will imbue it with; either only through it's anthropomorphic design, or through "pre-canned" social behaviour - e.g., pre-defined gestures that are used in certain situations, without the robot actually learning or having a changing internal model of its goals, its environment, or its human partner. Doggy's design (although not purely anthropomorphic, but also modelled after an animal in a more cartoonish way) does very much encourage someone to read its behaviour as emotional and as social cues. Doggy's gestures should therefore also be designed in such a way as to be easily readable, particularly because Doggy has only five DOF (which is including the wagging of the tail).

It can of course not enact facial expressions or movements of the limbs like Breazeal's robot Kismet was able to do. In choosing which gestures to express, Doggy will react to its environment and the state of the game it is playing, but not to social cues from human partners. Therefore, while Doggy does offer some social behaviour, in terms of this definition it is only a social interface and does not learn from human reactions, imitate human behaviour, or model such social cues.

Fong et al. (2003) refer to Breazeal's 2003 definition of social robots in their review of the field in "A Survey of Socially Interactive Robots". Their notion of *socially interactive robots* has been adapted both from Breazeal's definition and distinct classes, as well as from several other models, notably Dautenhahn and Billard's (1999) definition of social robots. They recount this definition of social robots as "embodied agents that are part of a heterogeneous group: a society of robots or humans", which are able to "recognize each other and engage in social interactions, they possess histories (perceive and interpret the world in terms of their own experience), and they explicitly communicate with and learn from each other" (Fong et al., 2003, p. 144). As such, it becomes apparent that while Doggy is to be endowed with some socially interactive qualities, it is not a social robot as per this definition, missing the sophistication of the features that social robots are to exhibit according to this (particularly regarding their histories, explicit communication and learning abilities). Fong et al. (2003) also expand on the definition by Breazeal (2003) by adding three more distinct classes to these: Firstly, they define robots as *socially situated* if they "are surrounded by a social environment that they perceive and react to", which includes that they have to be able to recognize other social robots or different social agents, as opposed to non-social objects in their environment. Secondly, they describe the class of socially embedded robots, which includes robots that fulfil the following requirements: They are "situated in a social environment" similarly to socially situated ones, in which they "interact with other agents and humans"; additionally they have to be "structurally coupled with their [...] environment; and lastly have to have an internalized 'understanding' of the social norms of human interaction. All of these combine to form an entity that is not only situated in a social environment, but has the ability to actively take part in the same and is therefore integrated in it." Lastly, they add the class of socially intelligent robots, which require even more precise internal models of social interaction based on human interaction, and can therefore be classified as artificial social intelligence (Fong et al., 2003, p. 145). While Doggy is indeed situated in a social environment, and does perceive and react to parts of it (both the ball in play, as well as some actions of the human player, should be things Doggy reacts to), it does not in fact recognize other social robots, differentiate between other social agents and non-social objects, and has no further understanding of its environment or its interactions or interactive partner.

Fong et al. (2003) go on to describe the design considerations that have to be taken into account when creating a socially interactive robot. For this thesis, it is important to explore how the restrictions that are given by the existing design of Doggy will influence possible interaction and emotional gestures. Doggy is a robot that Fong et al. (2003) would describe as being functionally designed, which means that its social intelligence is defined by how its gestures and interaction are perceived by the human player, while the internal design does not simulate those emotional states found in biology. As they explain, “this approach assumes that if we want to create the impression of an artificial social agent driven by beliefs and desires, we do not necessarily need to understand how the mind really works. Instead, it is sufficient to describe the mechanisms (sensations, traits, folk-psychology, etc.) by which people in everyday life understand socially intelligent creatures” (Persson et al., 2001 as cited by Fong et al., 2003, p. 148). Moreover, the reasons for this design align with those considerations Fong et al. (2003) propose - in particular, Doggy is constrained by its environment and the need for a functional physical design (for example, having movable facial features would prohibit the use of the head as the end of the bat). This is also in line with the context of the interaction, which will usually be of a shorter time span and limited number of repeats, which they note rather only requires superficial social competence (Fong et al., 2003, p. 148).

Following this, Fong et al. (2003) discuss several categories of design constraints that should be taken into account: Firstly, they argue that a socially interactive robot “must proficiently perceive and interpret human activity and behaviour”, meaning it has to be able to not only act on its own, but also in some way react to the actions of its human partner - by recognizing the social cues connoted in their behaviour. Secondly, the interaction would have to be natural, meaning that the interaction between human and robot needs to follow “social conventions and norms”. And thirdly, the robot must exhibit what Fong et al. (2003) call readable social cues, which are gestures acted out by the robot that a human interactive partner can read as representing particular internal states (Fong et al., 2003, p. 148). Analysing the points made by Fong et al. (2003) in terms of the robot Doggy, it becomes apparent that, while natural, readable gestures can be implemented and can help in forming a natural way of interaction between human and robot, it will not be possible to extend the variety of social cues the robot is able to understand beyond the scope of what is possible at the moment. The current state only allows the robot to react either to physical re-positioning of the bat by a human, to a ball being thrown by a player, or possibly to audible commands. Implementing the recognition of gestures or other expressions by the human player (for example in using a Kinect, which has been used with Piggy/Doggy in this project previously) will not be achievable in the scope of this thesis. Furthermore, the main focus lies on providing readable cues in form of gestures enacted by Doggy to the human.

Concerning the physical design (*embodiment* and *morphology*) of the robot, the following should be noted: Concerning embodiment Fong et al. (2003) quote the definition “that which establishes a basis for structural coupling by creating the potential for mutual perturbation between system and environment” (Dautenhahn, Ogden, et al., 2002 as cited by Fong et al., 2003), which means, as they explain, that “the more a robot can perturb an environment, and be perturbed by it, the more it is embodied” (Fong et al., 2003, p. 149). Doggy’s embodiment is one of the largest constraints on the level of defined interaction that is possible. It offers two main ways of influencing it’s environment somewhat indirectly: its movement (through which it can hit a ball or play gestures) and sound; but it is not able to physically perturb the environment in ways other than the hitting back of a thrown ball. Moreover, looking at Doggy’s morphology, we can see how this influences in what way social interactions with the robot will be perceived by a human. Fong et al. (2003) note that the morphology of a robot “helps establish social expectations” and therefore “biases interaction”, in particular: “A robot that resembles a dog will be treated differently (at least initially) than one which is anthropomorphic” (Fong et al., 2003, p. 149). Looking at Doggy, this robot combines aspects of several possible types of morphology into one. It is at the same time zoomorphic in the design of its appearance, anthropomorphic in that it is ‘standing upright’ and mimics the design of a clearly anthropomorphised dog, as well as caricatured in that it is a highly stylized version of a dog. Mainly, it’s morphology was designed to be functional and is constrained by the goal that it shall accomplish. Furthermore, the movements that are accomplishable by the robot do not in particular emulate that of a dog (neither a human, for that matter). One main particularly zoomorphic and dog-like feature that can be made use of is the dog’s ability to wag its tail. Other than that, the emotion of gestures will have to be conveyed through the way in which the upper body moves from side to side, and through the velocity at which it does so.

As is explained in Fong et al. (2003):

> “Artificial emotions are used in social robots for several reasons. The primary purpose, of course, is that emotion helps facilitate believable HRI. Artificial emotion can also provide feedback to the user, such as indicating the robot’s

> internal state, goals and (to an extent) intentions. Lastly, artificial emotions can act as a control mechanism, driving behaviour and reflecting how the robot is affected by, and adapts to, different factors over time." (Fong et al., 2003, p. 151)

Fong et al. (2003) discuss several theories of conveying emotion in robotics, and which design considerations should be taken into account. They explain:

> "Three primary theories are used to describe emotions. The first approach describes emotions in terms of discrete categories (e.g., happiness) [...]. The second approach characterizes emotions using continuous scales or basis dimensions, such as arousal and valence. The third approach, componential theory, acknowledges the importance of both categories and dimensions." (Fong et al., 2003, p. 151)

Emotion can be conveyed through speech, facial expression, and body language (Fong et al., 2003, pp. 152-153) - this thesis concentrates, of course, on the latter. Facial expression is not possible in the current iteration of Doggy's design (although it is interesting to note that some unintentional movements due to Doggy's design, mostly the flapping of the ears, are easy to interpret for a human as something of a 'facial' expression of a dog). Sound could be incorporated though, however, this will not be the goal of this work. Different again from emotion is *personality*. This attribute of a robot is something that can be conveyed through the emotions it expresses, as well as through its embodiment, motion, and communication manner (Fong et al., 2003, p. 154). A human interacting with a robot could easily read it as having a personality - whether it was intended to or not. Therefore it is expedient to consider the different personality types that Fong et al. (2003) distinguish between: *Tool-like* (robots that operate as 'smart appliances' and serve specific tasks on command, this correlates with traits like dependability, reliability, etc.); *pet or creature*; *cartoon* (caricatured personalities, such as seen in animation - exaggerated traits); *artificial being* (artificial, e.g. mechanistic, characteristics as might be inspired by literature and film); *human-like* (Fong et al., 2003, p. 155). Doggy of course is fundamentally pet- or creature-like, since it being a dog is the first association a human will have. However, in addition to that its personality can also be seen as a cartoon - because of its simplified and stylized appearance and motion, exaggerated movements will be needed to make gestures understandable, and its design in and of itself is very much inspired by animation (Tzeng, 2013, pp. 7-8). Fong et al. (2003) discuss a number of other design considerations, which are however less relevant to the design of gestures for Doggy. These include: human-oriented perception (which is not the case here - while Doggy can perceive sound, it does not parse speech), user modelling as well as learning and imitation (which are similarly not the case here), intentionality, which is attention and expression of goal-oriented behaviour (Doggy does have simple goals, even though they might not be explicitly implemented as such, like getting a human to play with it and hitting a ball it is thrown - this is only very loosely relevant though, since it will not be able to express the intentionality of its actions through further gestures).

A different concept of robotic emotions is the previously mentioned one of Dominance and Valence (see Chapter 2.2), which is proposed by Dryer (1998). As he explains, “central to the design of some future computer systems are the questions of how emotions can be specified, expressed, and accounted for by machines” (Dryer, 1998, p. 76). He proposes the two-factor-model of *dominance and valence*, in which he suggests that facial features can be located on the same interpersonal dimensions of dominance and affiliation (even basic expressions e.g. of emoticons) (Dryer, 1998, p. 78). Doggy is of course not equipped to display different facial expressions, but Dryer (1998) goes on to explain:

> “Although a face is used for the current demonstration, an advantage of this system is that emotions need not be represented by a human-like face. Dominance and affiliation might be represented in various ways for different kinds of representations. A chat avatar in the form of a dog, for example, might use a wagging tail to communicate high affiliation, a tail between its legs to communicate low dominance, and so on. Because the system uses the underlying dimensions of emotion, it is not tied to any particular kind of expression. The expression could be people, dogs, or abstractions, and animated characters, still images, or icons.” (Dryer, 1998, p. 78)

Another model for robotic emotion and personality is the TAME framework (as mentioned in 2.2), which was introduced by Moshkina and Arkin (2005). This framework is also in part based on Breazeal’s work with the robot Kismet: “Kismet also has a motivational system composed of drives and emotions, in which the affect space is defined along three dimensions: arousal, valence and stance” (Moshkina and Arkin, 2005, p. 1444). It incorporates a number of theories of personality, mood, emotion and attitudes, but the authors clarify that “it is not intended to be a cognitive model of affect and personality, but rather serves as a framework for modelling personality and affect in behaviour-based autonomous robotic systems” (Moshkina and Arkin, 2005, p. 1445). They define four interrelated components for the TAME model: “Personality Traits, Affect-Based Attitudes, Moods, and Emotions” (Moshkina and Arkin, 2005, p. 1445):

- **Trait:** based on the Five-Factor-Model of personality (McCrae et al, 1996)

- **Emotion:** “Emotional phenomena are divided into a set of basic emotions, such as joy, interest, surprise, fear, anger, sadness and disgust. These are recognized by a number of theorists, such as Ekman, Friesen, Izard, Plutchik and Tomkins” (Watson et. al, 2000 as cited by Moshkina and Arkin, 2005)
- **Mood:** “Thayer views moods as a background feeling that persists over time, that emphasizes meaning and enhances or reduces pleasure in our lives”; [...] “the two categories of moods included in the Mood Component are positive affect and negative affect, which are fairly independent of each other” (Thayer et. al 1996, as cited by Moshkina and Arkin, 2005)
- **Attitudes:** “a general and enduring positive or negative feeling about some person, object or issue” (Breckler et. al, 1989 as cited by Moshkina and Arkin, 2005)

Moshkina and Arkin (2005) further describe their design in detail on the basis of their partial implementation of the *Personality* and *Emotion* modules, which was implemented on the physical platform of Sony’s AIBO ERS-210A (Moshkina and Arkin, 2005, p. 1446). They conducted a study in which they compared participants’ interactions with AIBO both in the emotional and non-emotional state, and conclude that those who believed that the robot displayed emotions and/or personality also believed that these features made their interaction more pleasant, the negative mood was reduced in the emotional condition, and women were more susceptible to emotional cues and more willing to attribute emotions to the robot (Moshkina and Arkin, 2005, p. 1450).

In “An overview of human interactive robots for psychological enrichment”, Shibata (2004) also talks about different iterations of social robots. Specifically, he uses the term *human interactive robot*, and defines the following categories: “There are four categories of human interactive robots for psychological enrichment in terms of their relationship with humans: 1) performance robots; 2) teleoperated performance robots; 3) operation, building, programming, and control robots; and 4) interactive autonomous robots” (Shibata, 2004, p. 1750). While at a first glance we might want to argue that Doggy is an interactive robot and should therefore belong in the fourth category, Shibata (2004) does explain that “an interactive autonomous robot behaves autonomously using various kinds of sensors and actuators and can react to stimulation by its environment, including interacting with a human. The human perceives the behaviour of the interactive autonomous robot by using his senses. Interaction with the robot produces mental stimulation in the human. He then interprets the meaning of the robot’s behaviour and evaluates the robot with reference to his own knowledge and experiences” (Shibata, 2004, p. 1750). This is only true for Doggy in the most simplistic terms, in that its sensors (both sound and movement) are looking for very specific input and will therefore only produce a reaction in a few specific cases, outside of which it does not interact. How it reacts is also strictly pre-defined, and will not be changed through learning with its experiences. In addition to the aforementioned categories, Shibata (2004) differentiates between short-term and long-term interaction (Shibata, 2004, p. 1750)). Doggy is of course intended firstly for short-term use, which is apparent in that it does not need a learning function to prevent its behaviour from becoming too repetitive. Instead, its appearance and first impression are more important.

3.3 Proposed Interaction Design

The first step in furthering Doggy's socially interactive behaviour is to expand upon the existing gesture animation, as well as the state machine that describes Doggy's internal emotional state (see Chapter 3.1). The goal is to have Doggy interact with potential players both during and outside of game play, specifically in a way that will entice people to interact with it and play with it. In particular, this means that sound will be recognized and will also lead to engagement from Doggy's side, for which the work in *Sound of Interest. Ein Ballspielroboter hoert stereo* (Bartsch, 2015) will be used as the basis. However, it will not be possible within the scope of this work to try to have Doggy recognize its human partner's gestures or analyse their emotional state, or to have Doggy also keep an internal model of its partner's emotions. This might be one of multiple ways to keep working on Doggy in the future: more sophisticated sound recognition and reaction to it, internal representation of not only Doggy but also the people it interacts with, or other ways of reading its partner such as reading human gestures/facial expressions/etc. In addition to extending its internal state machine, more distinct emotional states will also be developed - at the moment, there are two types of gestures, which represent two distinct states: *happy* and *sad*.

Doggy is what could, going by the definitions of Breazeal (2003), be called a *social interface*: A person interacting with the robot will ascribe to it social attributes (even if only because of it's external design), and Doggy itself will provide social cues (gestures) in order to facilitate interaction - these gestures are, as Breazeal says, 'pre-canned behaviour'. It is necessary for any gestures that we design for Doggy to be both feasible for Doggy to perform within its existing three/five DOF, as well as for them to be easily understandable still, especially because Doggy will not be able to understand the human reaction.

In designing new gestures and states for Doggy's emotional expression, two main ideas will be considered: firstly, its overall *personality* (Fong et al., 2003, p. 154); and secondly, its momentary *emotions* (Fong et al., 2003, p. 151).

As discussed in Chapter 3.2, Doggy is both a *pet/creature* and a *cartoon character* in Fong et al.'s 2003 understanding of personality. This is mainly defined by its embodiment (the dog design was, as Tzeng (2013) explains in *New Dog, New Tricks - Human-Robot Interaction and Appearance Design of a Ball Playing Robot*, very much inspired by animated characters), as well as its possible motion (since this will also have to be highly stylized in order to work with Doggy's embodiment and DOF). Because of this, the goal is to design gestures for Doggy that convey a personality that mimics that of a cute cartoon dog. In terms of the "Big Five" personality traits that Fong et al. (2003) reference, that could be realized as: sociable and outgoing (on the scale of *extraversion*), definitely *agreeable* (instead of unpleasant or unfriendly), emotionally maybe more stable than nervous (in terms of *neuroticism*), and more curious instead of cautious (in terms of *openness*) - *conscientiousness* probably being the factor to be the least defined by the emotional range that is possible for Doggy. In practical terms, Doggy would appear generally happy and easily excited, especially happy to be interacted with. This personality will persist through its varying emotional states, and will inform how these different emotional gestures are designed. In addition, Doggy's motion should never seem for example threatening, too slow or lethargic, angry, or similarly unwelcoming. 'Negative' emotional states (disappointment at not succeeding at the current game, or sadness at not being interacted with at all) should hence still invoke the image of an adorable cartoon character or pet that someone will then want to cheer up and interact with, instead of seeming cold or hostile.

Doggy's *emotions* on the other hand are not consistent like its underlying personality, but change according to the situation. That means that while the personality is at the basis of *how* the different gestures are designed and which emotional states there are, the emotions themselves define *which* gesture is chosen in which moment.

It seems feasible for this project to only distinguish between these two concepts, instead of for example using the *TAME* model (Moshkina and Arkin, 2005) which defines not only emotion opposite trait (which corresponds to what here has been referred to as personality), but also attitude and mood. Unlike in the TAME model, for Doggy, trait, attitude and mood do not change. Trait/personality is therefore still useful to have as a baseline for its different emotional states and expressions. The general mood however will not change (considerably): For example, it is not necessary or useful for Doggy to be more excitable in the morning and more tired towards the end of the day, or less easily trusting after several disappointing encounters - since every game is relatively short-term and no single person will get to notice the difference, as well as due to the wish to keep the experience of playing with Doggy relatively consistent from one player to the next. Similarly, Doggy does not have to have markedly different attitudes towards different things, since Doggy will not be confronted with (or rather, be able to respond to) a vast variety of outside stimuli, but will rather always be in what amounts to basically the same situation (waiting to play a game or playing a game). Its attitude (that it wants to play and seeks human interaction) is therefore relatively consistent and can be used as part of its personality.

In order to define which emotional states will be useful for Doggy's interaction, one can study which external or internal stimuli Doggy will have to react to and change its internal state for. There are three main external cues that Doggy will be able to identify:

- The main one cue is that a ball has been thrown towards it, which is still the basis of Doggy's game play. When this happens, Doggy attempts to hit the ball back to the player. The most important thing here is that this overrides any other emotion or interaction, so that Doggy can be fast enough to actually hit the ball; therefore it stops any current gesture in favour of moving towards the ball. After the ball has either been hit successfully or not, Doggy will respond appropriately (hitting the ball makes it happy, not doing so disappointed or sad).

- The second external cue is sound; specifically, people should to be able to get Doggy's attention by clapping and Doggy should be able to react according to the direction from which the sound comes. Imitating a real dog, Doggy will get excited when someone indicates that they want to interact with it; Doggy's excited reaction is in hopes of facilitating further interaction. If this does not however follow, Doggy will again be sad about not getting to play, which leads to the third external cue:

- A specific period of time in which no interaction happens. This is not exactly an outside event like the other two options, but can still be used as an attribute of its environment that causes a certain reaction. Specifically, that means that Doggy when not being interacted or played with waits for any kind of cue or interaction - during this default state, it makes sense for it to perform an encouraging, positive gesture without prompting, in order to invite interaction (as was used in Tzeng, 2013). If, however, after a certain amount of time no reaction occurs, Doggy does not keep making default/happy gestures, but changes into a sad or discouraged state for a while or until a different outside cue occurs. The same thing happens if Doggy had been excited by a sound cue (which does tell it that there is someone there that might play with it), but no other interaction followed even after trying to encourage such.

Figure 3.3 illustrates these internal emotional states in a state machine.

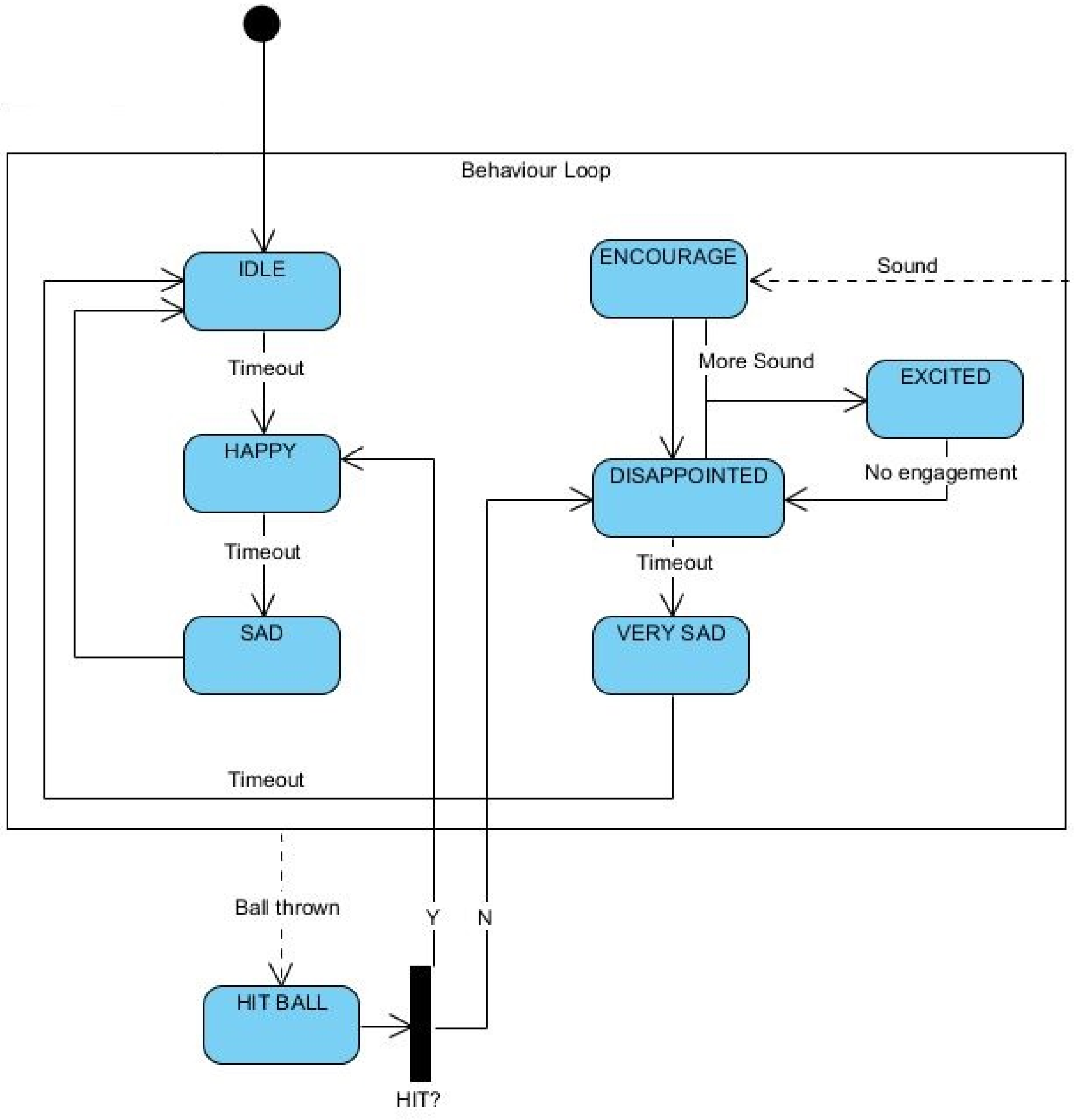


Figure 3.3: Proposed state machine for Doggy's behaviour.

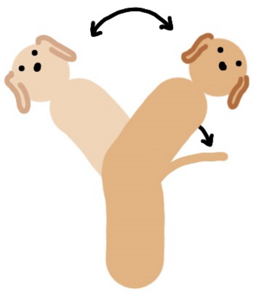
Out of the states shown in Figure 3.3, the *idle* state as well as the one when hitting a ball are the only states in which no gestures are performed. The other states all have at least one gesture associated with it, but could potentially choose from a variety of more or less similar gestures, that convey the same emotion but allow for less predictable expression.

A starting point for these gestures is provided by the work on this project by Tzeng (2013). While these gestures do not at the moment run on the robot, the files that were provided with that thesis do still provide a (formalized) description of these gestures. That includes both emotional gestures (‘excited’, ‘fearful’, ‘sad’), as well as a number of movements (‘dancing’, ‘dizzy’, ‘spiral’, ‘zig-zag’).

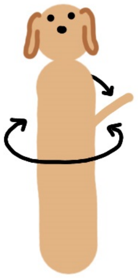
The gesture files describe the movements as an array of positions over time, a position consisting of the angles of the motors that move Doggy’s body and tail. After implementing the program that would simulate Doggy’s movements as well as possible gestures (see Chapter 4), it was possible to read these files in order to actually see the movements they defined. Before that, however, a number of relatively simple gestures for the different states was sketched out first, which would allow the preview of the different states without having to parse the existing gesture files. These first gestures were each described by singular movements or simple back and forth movement (see Figures 3.4a - 3.4d):

- **Happy:** moving left to right & wagging tail.
- **Excited:** rotating around itself from left to right & wagging tail quickly.
- **Encourage:** nodding towards the player & wagging tail.
- **Sad:** letting its head/upper body & tail hang down.

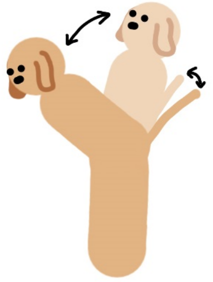
After implementing the test program and simulating the gestures (see Chapter 4), these were adapted based on the rendering. The four initial gestures were all kept and be used for the evaluation, they were however tweaked in terms of exactly how fast and how far in either direction the gestures were performed (see Chapter 4.3).



(a) Happy



(b) Excited



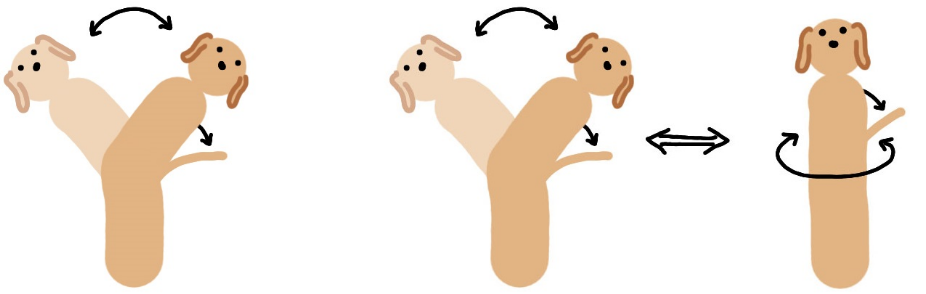
(c) Encourage

(d) Sad

Figure 3.4: First iteration of gestures.

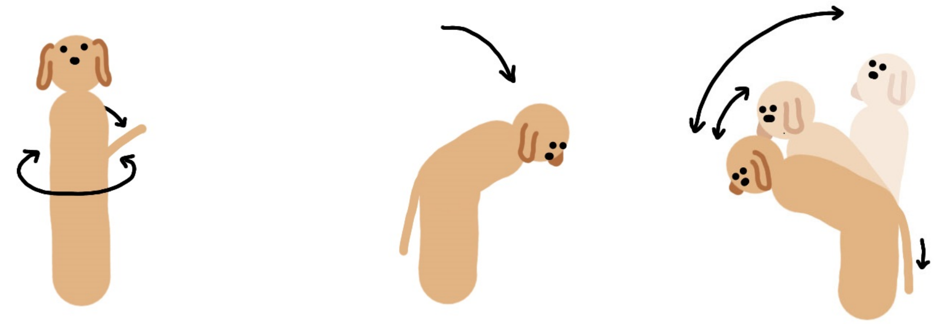
In addition to the initial simple gestures, which all involve movement around only one axis, new gestures were added that are comprised of different types of movement (for example rotation and sideways movement in *happy (2)* and *excited (2)* (see list below)). Gestures involving more complicated movements and more variation during one gesture were considered, but were abandoned for not being easily readable even in the simulated test. Additionally, there are variations of some gestures that only differ in their speed - the same movement, for example simple sideways back and forth movement, could be perceived as happy, but calm, up to quite excited and agitated, depending on the speed with which the movement was performed. Therefore, *happy (2)* and *excited (2)* only differ in speed and scale of their movement. Figures 3.5a through 3.5i illustrate the different gestures that are listed below. Attached with the digital copy of this work are additional animated renditions.

- **Happy (1):** moving left to right & wagging tail.
- **Happy (2):** moving left to right + rotating around itself & wagging tail.
- **Happy (3):** moving left to right (fast) & wagging tail.
- **Excited (1):** rotating around itself from left to right & wagging tail quickly.
- **Excited (2):** moving left to right + rotating around itself (fast) & wagging tail quickly.
- **Sad (1):** letting its head/upper body & tail hang down (slowly).
- **Sad (2):** lowering its head/upper body → up a bit → back down (slowly), and letting its tail hang down.
- **Disappointed (1):** moving upper body & tail down (faster).
- **Disappointed (2):** moving upper body & tail down → shaking head.
- **Encourage (1):** nodding up and down towards the player & wagging tail.
- **Encourage (2):** nodding towards the player + moving left to right (fast)& wagging tail.
- **Encourage all:** nodding in different directions & wagging tail.



(a) Happy (1) & (3)

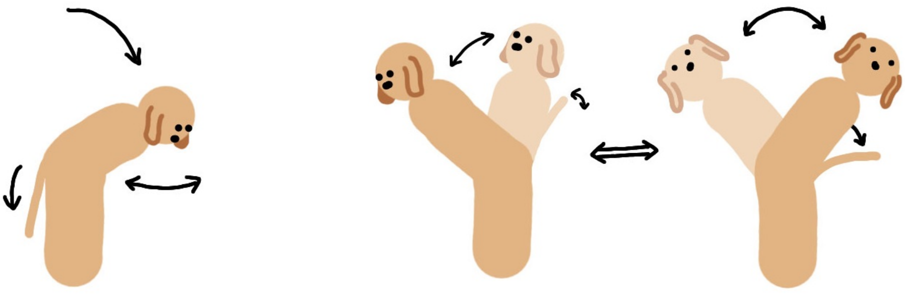
(b) Happy (2)/excited (2)



(c) Excited

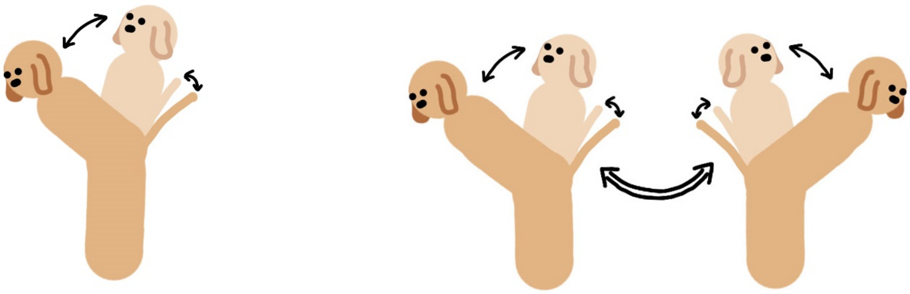
(d) Sad (1)/disappointed (1)

(e) Sad (2)



(f) Disappointed (2)

(g) Encourage (2)



(h) Encourage

(i) Encourage all

Figure 3.5: Final gesture designs, that were to be tested on Doggy itself.

Chapter 4

Implementation of the Test Program

4.1 Internals of the Doggy Robot

In order to simulate possible gestures for Doggy to perform, the first step is to look at the internal make-up in order to see which movements Doggy is actually capable of, and how these might be represented schematically when implementing the animations. Figure 4.1 illustrates the internals of the robot, Doggy's 'skeleton' so to say.

As was discussed in Chapter 2.1, Doggy (specifically, its body; the moving bat that will hit the ball), has three DOF, which are achieved by three motors - the joints around which the robot moves its upper body. These are situated approximately in the middle of the body, the lower body being static. The lower body is 77.89 cm tall, at which point there is the first motor (ACS1). This allows it to rotate around its vertical axis (the z-axis), thereby rotating its upper body from side to side and changing the direction in which Doggy faces. 23.34 cm above this, at a height of 101.23 cm, are the next two motors. The second one (ACS2) is 17.15 cm left off centre (i.e. along the negative y-axis in terms of the world coordinate system), and allows it to bend over forwards or backwards (rotating around the y-axis). The third one (ACS3) is 20.55 cm off centre in front (i.e. along the x-axis in the world-coordinate system) and lets it bend down to either side (rotating around the x-axis). In Figure 4.1, the x-axis is marked red, the y-axis is green, and the z-axis is blue. The world coordinate system is situated at the very bottom. For each of the motors, its own coordinate system is illustrated (they rotate around their own z-axis). Since all of Doggy's rotational axes intersect at one point (point C), there is a redundancy in DOFs. Doggy's head (the Styrofoam ball at the top of the bat) therefore also moves on a sphere around Doggy's centre.

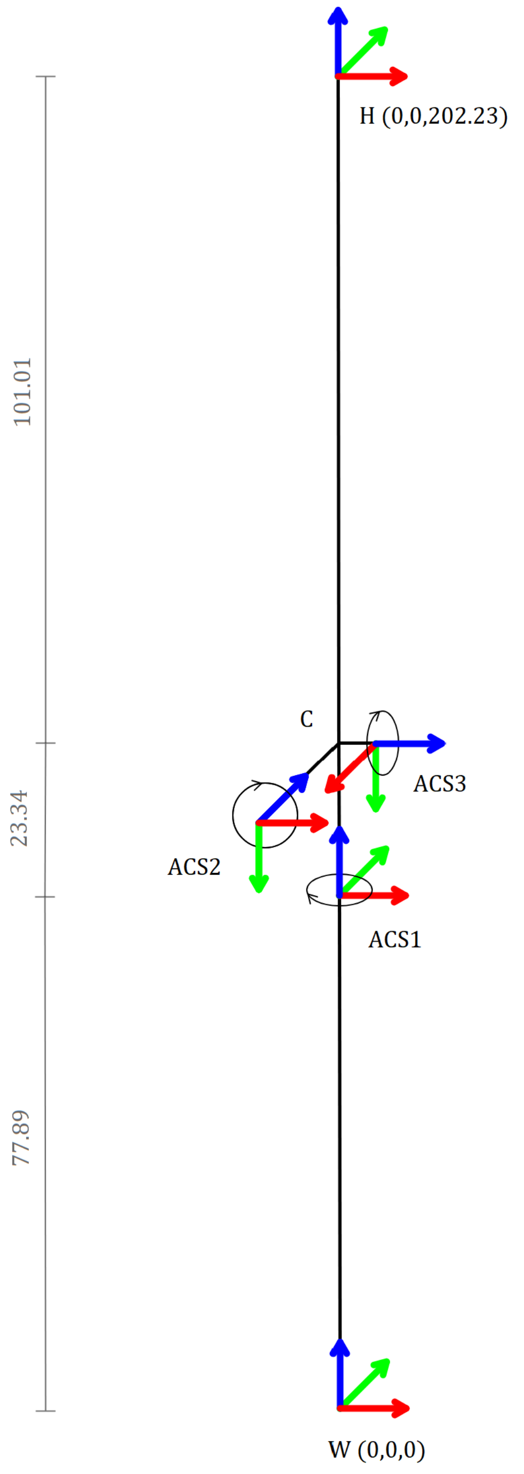


Figure 4.1: Illustration of Doggy's internal make-up.

Technically, Doggy could reach all positions on that sphere with only two DOF (since a position on a globe can be sufficiently described with two variables - consider latitude and longitude), as was done with Piggy originally (Laue et al., 2014, p. 4). However, as was mentioned in 2.1.2, the additional third 'joint' allows for more distinct gestures for Doggy to perform, as well as overall more natural motion - since Doggy has a clear 'front' (marked by the face - whereas Piggy's hat was the same from every side), the position of its head is not only defined by the position on its sphere of possible motion, but also by the orientation of the face. This can change the expression of emotion considerably, and can be exactly controlled only with three DOF (with two DOF, Doggy is simply unable to bend down to the side).

Furthermore, there is the tail, which was added with the robot's 'costume change' by Tzeng (2013). This has two additional DOF: it can be moved up or down (around an axis parallel to Doggy's y-axis), or to either side (around the x-axis). This allows Doggy an even larger range of emotional expression - the position and movement of a dog's tail (even a partly humanoid and partly cartoon-like version as Doggy is) is one of the things a human will immediately use to gauge its emotional state (for example if it is wagging its tail, holding it up, or letting it hang down).

In order to translate the interaction and emotional expression concepts that were discussed in Chapter 3 onto the actual robot, the first goal is to build a virtual model with which to test what different movements might look like. Going through many iterations of possible gestures on the actual robot would not only consume a lot more time, but could even be potentially dangerous if the velocity and abruptness of certain movements were judged incorrectly, or if a movement should not be possible on the robot in the first place because its body cannot be moved in that way (rotating completely around itself, for example). In implementing a program to test these possible movements, Doggy's internal structure as seen in Figure 4.1, particularly the position of the different motors, will be represented exactly. While the same range of motion could be achieved by having all movements be a rotation around one centre point, using the exact position makes it possible to visualize the layout of the motors, and makes it easier to adapt the program to possible future changes.

4.2 The Test Program

The simulation program was implemented in C++ (the entire program code is included on the attached CD). The code is based on other projects, including the code programmed in the course 'Computergrafik' at the University Bremen under Prof. Gabriel Zachmann (Zachmann, 2015), as well as two open source projects (Beucler, 2018 and User MaxH, 2010), and uses the open source vector math library (Sony Computer Entertainment Inc., 2007).

There are two main parts to this implementation: Firstly (and most importantly for the transfer of this onto the actual robot) the control of Doggy's movements, and secondly the drawing of the 3D model, which was imperative for the first evaluation of possible gesture designs and the creation of the extended set of gestures. Additionally, it is possible to visualize the technical movements 'skeleton' similar to Figure 4.1.

4.2.1 Movement Control

Note: In the program code, in order to make the implementation less abstract and therefore somewhat easier on the programmer, Doggy's movements are designated as roll, pitch, and yaw (as is known from the rotational movement of an aircraft around its principal axis). *Yaw* signifies the rotation around Doggy's z-axis, in Doggy's case, the rotation of its upper body ‘around itself’; *pitch* is the rotation around the y-axis, which here means Doggy bending its body down to the front or to the back; and *roll* refers to the rotation around the x-axis, which is Doggy tilting its body down either to its left or right side. The order of these operations is important: pitch and roll are applied to the coordinate system after it has been rotated by the yaw angle, roll is performed after the pitch angle has been applied.

It is possible to either control Doggy’s movements directly, or to let it play out its animations throughout its internal states as it would when active, but not receiving any interaction. Direct control is achieved using keyboard commands: the *a* and *d* keys control the roll angle (sideways rotation), the *w* and *s* keys control the pitch angle (forward/backward rotation), and the *q* and *e* keys control the yaw angle (rotation around itself). In addition to the main body movement, the tail can be moved individually with the *j* and *l* keys (for sideways movement) or the *i* and *k* keys (for up/down movement). The arrow keys are used to rotate the entire world in order to shift the view (left and right control the view angle to either side, up and down control the tip angle, tipping the screen up or down).

Additionally, instead of ‘steering’ Doggy, a specific position (in terms of the three angles) can be input, and Doggy will move directly into that position. This is achieved in the *moveToDestination* function, which is also used in Doggy’s own movements: It receives a destination position as a vector, and a movement speed; and unless the current angle is already (approximately) the same as the destination angle, each angle is then increased or decreased (depending on the necessary direction) by the speed value. If all angles are already the same as the destination, the function returns *true*; otherwise *false* - while it returns *false*, this function will be executed with every redraw, which will stop once the desired position is achieved. Using this function, the movement between two positions is roughly interpolated (it is not necessarily exactly proportional due to small differences in how much time elapses between redraws, but in practice this proved to be not noticeable), and not, as in the previous project, defined step by step. The same is done in the *moveTailToDestination* function, which gets passed a two-dimensional vector with the two tail angles, as well as the speed.

If Doggy’s movement is not controlled directly, it moves through a number of states and performs specific animations based on those. The states and animations are based on what was designed in Chapter 3.3, however, throwing the ball or having Doggy hit it was not incorporated at this point. The happy and disappointed states were still implemented, since these can be called manually in order to simulate the accompanying movement.

- **Idle:** Doggy moves back to the centre position. When more than 10s have passed since the last animation was completed, the state changes to *happy*.
- **Happy:** A happy gesture is performed, and after 25s of further inactivity, Doggy changes into the *sad* state.
- **Excited:** An excited animation would be performed after Doggy successfully hit a ball. If no further interaction follows for 25s, Doggy changes into the *sad* state.
- **Sad:** The sad animation is performed. Then, if Doggy is not already in centre position, it moves to that; and changes back into the *idle* state.
- **Disappointed:** If Doggy was not successful in hitting a ball, it performs at this point a disappointed gesture. After that, it moves back to its default position and *idle* state
- **Encourage:** Doggy performs one of the encouraging animations; but after 15s changes into the state *disappointed*.
- **Encourage all:** Doggy performs the 'encourage all' animation (which, other than encourage, is played in alternating directions towards several potential players), and after 15s also switches to *disappointed*.

There are two further modes of interaction of Doggy's simulation, which are realized with two more states. Firstly, it is possible to have Doggy move towards a 'sound' cue - for this simulation, this is imitated using keyboard commands, by letting the user choose a direction of origin for the cue by pressing a key on the bottom row of letter keys (y through m, which can be imagined as being situated in a half circle around Doggy's front). These put Doggy into the state:

- **Turn to:** Doggy moves towards where the simulated sound cue is coming from. If then for 25s no interaction follows, it changes into the *encourage* state.

Secondly, it is possible to choose directly one of the states and have Doggy perform its gesture. These can be chosen with the number keys, and initiate the state:

- **Animation:** The requested states' gesture is played. Doggy stays in this state until the 0 key is pressed, which puts it back into the *idle* state.

In the first iteration of the program, this same state was used to manually choose one of the animation files from Tzeng (2013) in order to view these. It was changed to switch between the new, proposed gesture animations while designing and implementing these. A few of the states have alternative gestures (for example happy (1), (2) and (3)), which can be exchanged manually.

4.2.2 Drawing the 3D Model

To render the 3D model for this simulation, OpenGL and GLUT (the OpenGL Utility Toolkit) are used. Normally, the z-axis would be the one facing towards the user, it being the depth axis; the x-axis points to the right and the y-axis to the top. Instead of translating Doggy's internal make-up into this coordinate system, the first step rotates the OpenGL matrix so that x points forward (Doggy is facing the user) and z upward. In the second step, the view angle and tip angle are applied (which is the user changing the orientation of the entire view) - the scene is rotated around the y-axis by the tip angle, and around the z-axis by the view angle. Then, in this now rotated matrix, Doggy's actual body is drawn. The 3D model for the body was created in Blender, and is loaded as a new mesh. The lower and upper body are drawn individually, as only the upper body moves. The lower body is drawn at this point.

There is one other possible step before the upper body is called, which calls the function *Draw_Clapping* - this was used to visualize the direction from which a sound cue was simulated, to be able to verify that Doggy turns in the correct direction. Other necessary steps happening even before this (initializing OpenGL, setting material, lighting and shading parameters and such) shall not be explained in further detail here, but can be referenced in the included source code.

The main work in drawing the 3D model (specifically, its moving parts) is done in the *Draw_Dog* function, which is called at this point. This moves and draws the upper body and tail in the following steps:

The coordinate system is translated upwards along the z-axis to where the upper body begins. There, it is rotated around the z-axis (around 'itself') by the yaw angle. Then, the tail is moved and drawn: The coordinate system is translated to a point at the edge of Doggy's body on the x-axis, and then rotated by the sideways rotation angle around the z-axis, and by the up/downwards rotation angle around the y-axis; after which the tail is drawn and the matrix returned to its former orientation. It is then rotated into the relative coordinate system of ACS2, where the pitch angle is applied (which bends the upper body down forwards or backwards); and is after that rotated into ACS3's coordinate system, where the roll angle is applied (which bends the upper body down to either side). After that, the coordinate system is returned from the relative motor coordinate system to that of Doggy, keeping of course the applied movement angles, and here the mesh of Doggy's upper body is loaded, material parameters are set and the body is drawn.

Alternatively, it is possible to draw only the 'skeleton' of the robot, which illustrates the body itself only by lines, but adds coordinate systems and the locations of the different motors, like they are visualized in Figure 4.1.

4.3 Changes Based on Simulated Tests

The DoggyTest program allows for a first, rough evaluation of the initial concepts of Doggy's states, and was also used when designing the proposed gestures. After these have been transferred onto the physical robot, it will be possible to conduct a more expansive evaluation with users, to have them gauge the impression in terms of which 'emotion' is being performed, and draw conclusions on how these gestures influence the interaction with the Doggy robot. The DoggyTest program, on the other hand, will not be used as a user test, but instead allows the programmer to conduct a sort of simplified 'expert' prototype test with which to explore two main questions: firstly, and most importantly, to ensure the designed animations are correctly formalized and use a range of motion and velocity that can safely be performed on the robot; and secondly, to get a first impression of what possible animations will look like and change the initial designs that seem to be obviously unsuitable.

There were two different types of animations that were to be simulated in the test program. Firstly there were the animations that were first designed in the scope of the Master thesis by Tzeng (2013) - these were provided as comma separated value documents, in which each line provides a time, and the exact angles of each joint. This format allows for animations which are variable in speed, and makes it unnecessary to calculate sub-steps between positions. However, this also makes it very difficult to design and formalize new movements without using the design program that was implemented by Tzeng (2013). Therefore, new animations were simplified to only have to provide the outer points of each movement, and no smaller steps on the way between two positions. The interim positions are calculated using the *moveTo* function as explained in 4.2.1, and only the overall speed can be set, no acceleration. The angles for the gestures are simply provided in degrees.

Based on the simulation prototype, a number of changes were made to the first draft animations, and general movement ranges were defined:

- Speeds were set to
    - **Slow:** 2° per second
    - **Normal:** 5° per second
    - **Fast:** 10° per second
    - **Very Fast:** 20° per second
- Quick forward/backward movement confined to ±20°; forward movement, for example when sad was reduced to 60°
- Range of motion in rotating the entire body was confined to a total angle of 120°
- The range of motion for the tail in height is 20°(down) to 100°(up)

Additionally, the action *encourage all* was added at this point, for cases in which several sound cues were received in quick succession (otherwise, Doggy would only give quick jerks or twitches, or remain motionless - because every new sound cue interrupts the animation that is in progress). Finally, while the original animation taken from Tzeng (2013) combined tail and body movement (the pre-set animations for either could be freely combined, but it had to be one of the specified and saved combinations that were to be played), this was changed to separating them completely, so that every state calls two animations.

Chapter 5

Evaluation of the Design with the Robot Doggy

5.1 Integrating the Implementation with ROS

Doggy uses *ROS*, the open-source *Robot Operating System*, which provides operating system services for robotics "including hardware abstraction, low-level device control, implementation of commonly-used functionality, message-passing between processes, and package management" (Open Source Robotics Foundation, 2018). As is explained in the *ROS Documentation*, ROS software is typically organized in packages, which then include for example libraries, data sets or other files, as well as nodes, which constitute the individual processes. Structurally, there is a ROS master, which allows the nodes to communicate via messages, and organizes services that can be published or subscribed to.

Doggy's interaction system and emotional states were transferred to this model for the robot to use. This was realized by implementing a new package for Doggy's animation, which was the easiest way for it to be integrated with the existing packages that control Doggy's main movement and the ball game. In *Sound of Interest. Ein Ballspielroboter hoert stereo* (Bartsch, 2015), a sound recognition system was implemented in a similar way. This was connected to Doggy's main software as well as to the new interaction animations, which would allow the integration of Doggy's reaction to sound cues. The basis for the implementation was the work provided by Dennis Schüthe (Schuethe and Frese, 2014). In particular, this were the nodes that allowed the communication between computer and robot through Ethernet or microcontroller.

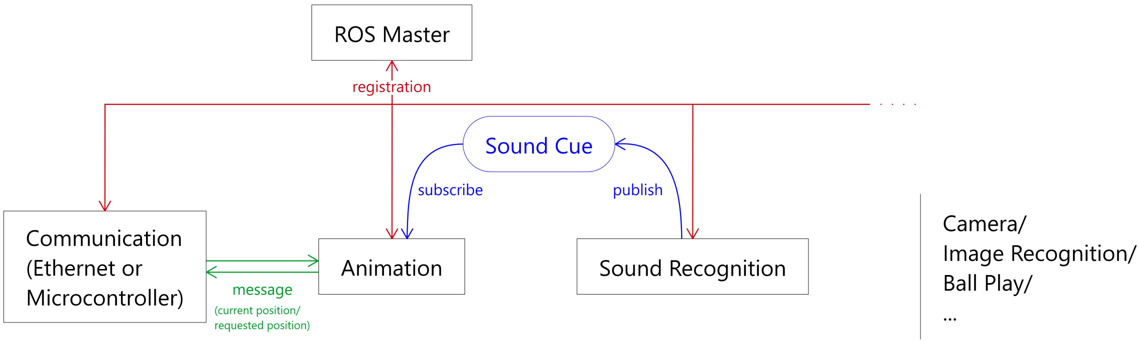


Figure 5.1: Structure of Doggy's ROS network

Figure 5.1 shows the structural network of the implementation, the computation graph, focusing on those nodes that are relevant for gesture animation and interaction. Additional nodes are those that control the cameras, process the images, recognize a ball that was thrown, calculate the necessary movement to hit the ball back and control that part of the game play. The *master* organizes the registration of all the nodes, without it, nodes would not be able to use services or send messages. The *communication* node (it is possible to use either Ethernet or microcontroller, which are two distinct nodes but for clarity are here summarized as communication) organizes the communication directly with Doggy: it initiates and controls the device connection, and processes user input (via keyboard) with which the user can choose to receive, log or send raw data, and get the current angles or set new angles. With this, the *animation node* exchanges messages when wanting to read the current position or send the required angles. The *sound recognition* node publishes when a sound cue was received; the animation node subscribes to that topic and thus learns when there is a new cue. Unfortunately, having this function reliably was not achieved while working on this project (especially since the goal was specifically to try to recognize where a sound cue comes from). When performing the user tests, this had to be taken into account, but the problem could be circumvented in order to still be able to test the interaction (see Chapter 5.2). The animation node itself is what was implemented based on the DoggyTest program. It controls Doggy's internal emotional state, and which animated gestures are performed. It also reads keyboard input. Like in the simulation program, Doggy can be steered using keyboard controls, and like in the simulation, this can be used to provide a pseudo-sound cue from a specified direction. In addition to this, an option was added to use a gaming controller (a typical X-Box controller or one with equivalent button layout) to control Doggy: The joysticks allowed for a much more natural way of 'steering' the robot's body. While this will not usually be utilized by people interacting and playing with Doggy in a casual manner, it proved very useful for testing the range of motion, ‘pre-viewing’ ideas for possibles movements, as well as in the later user tests.

5.2 Test Design

In evaluating Doggy’s interaction design, the goal was to obtain a qualitative assessment of the motions and gestures that Doggy performs, and of how its emotional states are understood. Main questions in this were:

- How intuitive is the interaction with Doggy? How much instruction is needed for people to interact with Doggy, and what kind of interaction does it afford?
- How are the emotional gestures/states understood? Firstly, *do* users understand the (after all quite simplistic) movements as expressions of emotion; and secondly, how understandable are the gestures in terms of which emotional state is being conveyed? What kind of reaction do these emotional performances facilitate, or do they lead to confusion?
- How much fun is it for someone to interact or play with Doggy? Does the expression of different emotions improve or diminish the experience, or does it maybe not impact it at all?

For this evaluation, different kinds of tests were considered. While it would have been interesting to conduct a somewhat larger scale user test including for example a survey of how they perceived the interaction, this ended up not being feasible due to time constraints. Instead, it was decided to conduct a *discount usability test*. This is a concept that was introduced by Nielsen (1989) in “Usability Engineering at a Discount”: The main idea is to conduct a test with only a small number of participants in a simplified user test, utilizing the method of thinking-aloud, and focusing on the qualitative evaluation over the gathering of quantitative data. Nielsen (1989) also advocated the use of narrowed-down prototypes, for example paper prototypes that could be easily constructed at different stages of the implementation process; as well as a heuristic expert evaluation. In this case, however, a simpler prototype was not necessary as this project builds on the Doggy project which already provided the working robot, and one of the main reasons for this test is to evaluate the gestures not only on a simulation, but in how they are perceived when performed on the actual robot.

Therefore, this evaluation focuses specifically on the tests with a small number of users. During the test, users were asked to 'think aloud' and voice their thoughts, questions, and opinions during their interaction. These were protocolled along with the observation of the participant's interaction. Afterwards, the participants were interviewed on how they perceived the interaction.

The emotional states were in working order when the tests were conducted, however, there were some technical problems that impacted the conception as well as the execution of the tests. As was explained in Chapter 5.1, the sound recognition was not usable as would have been preferred. In order to still include it in the interaction, this was realized with something of a *'wizard of oz'* solution: The program already provided the functionality of giving a pseudo-sound cue by using either the keyboard or the controller to choose a direction from where the cue was supposed to be detected. During the test, the participant was kept under the impression that Doggy would be able to perceive loud sounds like claps, while in actuality the direction was given by the administrator that was standing on Doggy's side of the room. This same solution was used to show the participants different variations of gestures - it was only possible to implement a simplified version of the interactive design in ROS, so to compare how the participants would interpret the different iterations that were designed for most states, this was animated manually instead. Unfortunately, while the ball game on its own functioned perfectly, the current implementation of the communication nodes did not allow both the ball game and the interactive behaviour or sound recognition to be executed at the same time, and the adaptation of it to make this possible was not achieved in this project. Because of this, both the game and the interactive behaviour was included in the interaction tests, but were tested separately; there robot was manually switched between the ball game mode and the interaction mode.

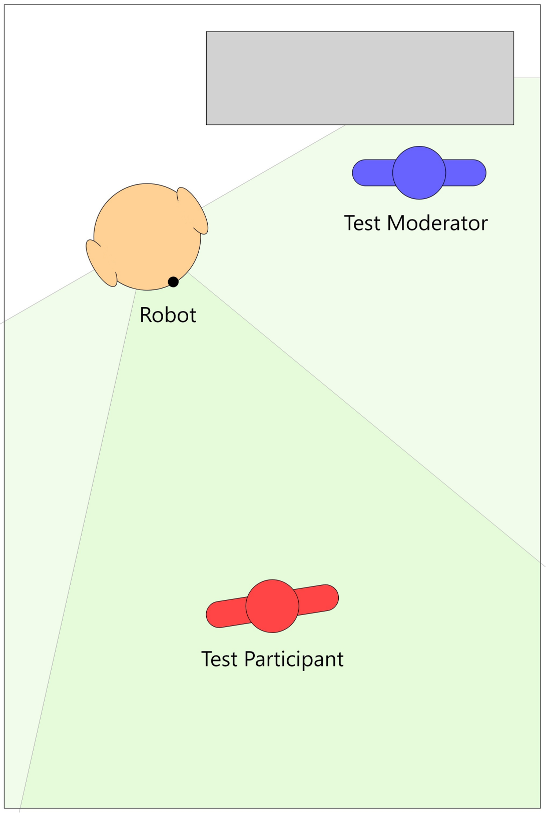


Figure 5.2: Layout of the room in which the test was conducted

Therefore, a significant part of the behaviour design that was dependent on the game itself could not be tested as such, but only simulated.

In September and October of 2016, this test was performed with three participants (individually), all of which were current or former university students between the ages of (at the time) 25 and 29. They were studying or had studied subjects ranging from Biology to German studies and Art for Education - none of them had any experience or preconceived notions of interacting with robots (two of the participants however did own or had owned a dog). The tests were conducted at the University Bremen, in the office in which Doggy was stationed. The room, while not particularly spacious, was large enough to comfortably throw a ball to the robot and have it hit it back, although only in an angle of about 60°around Doggy - the layout of the room can be seen in 5.2, the area from which a ball can be thrown is marked in light green; the larger, lighter area being Doggy's main range of motion. The participants knew that the project involved interacting with a 'Ballspielroboter' (a ball-playing robot) which they knew was in the form of a dog, but only one participant knew before the test what exactly that entailed - that they could throw a ball towards the robot and it would hit it back with its head. None of the participants were given any indication of how exactly they were supposed to interact with the robot or what Doggy would react to, unless they specifically asked for direction.

The test was structured in the following way:

1. **Introduction:** The participant was welcomed to the test and given a general explanation of the project itself and of the test that was to be performed. While it was not explained how exactly the robot worked before the test, it was explained that the focus of the project was interaction between the robot and a human, including using the provided ball to interact with it. The participant was instructed to think aloud while testing the robot.

2. **Familiarisation:** The participant was allowed to interact with the robot in whatever way they wished or could think of in order to familiarise themselves with what it would be able to do. When they initiated the game play, the robot was switched to that mode.

3. **Game Play & Interaction:** The participant engaged with the robot in the typical ball game, and afterwards through other interactions. They were generally allowed completely free action, it was however verified whether or not the robot had arrived at each possible state at least once.

4. **Judging different gestures:** After the main game play, the participant was also shown different iterations of the gesture animations, and asked to describe how they perceived the gesture (in particular, without the context of which situation in the game a gesture had occurred at).
5. **Conversation:** Lastly, the participant was interviewed on how they perceived the robot itself and the interaction, and whether they believed that the robot's design and motions influenced how they interacted with it.

5.3 Test Results

Following is a summary of the results obtained during the usability evaluations. The tests were all conducted in German, so the participants' descriptions of how they perceived the interaction with Doggy was recorded in German and translated into English for this analysis of the results.

Affordance - what kind of interaction does Doggy invite? How intuitive is the interaction?

The participants were provided with a ball, but were not told exactly what kind of movement the robot was capable of or in what way it might react to the ball. One participant concluded that the robot could 'see' the visually easily identifiable ball and tested whether the robot would rotate to follow the ball (which it did not do). This participant additionally tried to touch or lightly move Doggy, and while it is possible to move Doggy, this did not lead to any particular reaction. They did not dare to move it with more force (which would have also been possible and would not have lead to any damage, although it wouldn't have led to a different kind of interaction either). The participants quickly asked if they were supposed to simply throw the ball in the general direction of the robot - considering Doggy's external design, it was clear to either of them that the robot would not be able to move about the room, and that it lacked any limbs to interact with the ball. Notably, all of the participants were very careful with the robot, fearing to potentially break something or interact with it in the 'wrong' way. They asked for confirmation before throwing the ball for the first time (even the participant that was already familiar with the theoretical concept of how the game would work and what the robot could do), and both explained that they would throw it carefully and relatively low in order to see how the robot would react.

None of the participants immediately tried to interact with the robot through sound - neither by making any kind of loud noise, nor by trying to talk to it. All of them imagined that it might ‘see’ them and react to their movement, but neither supposed that it might be voice controlled or activated. That is even though voice commands are one of the main means of interaction that Doggy lacks in comparison to an actual dog. One participant wondered whether the robot would see them and rotate to follow them about the room. At that point, they were prompted by asking how they might otherwise get a dogs attention: their first idea was to whistle sharply, which was considered as enough of a sound cue to have Doggy react and turn towards them. The other two participants only used the sound cue when they learned that they could clap or produce another loud sound to get a reaction, which they were told after asking for further instruction. None of them however doubted that Doggy was indeed turning according to the sounds - even though the wizard of oz method employed for this meant that the direction was not entirely accurate at times, they easily accepted this as the robot approximating their position based the sound alone.

Emotional gesture expression

Out of the available gestures, some were very easily understood, while others were more confusing. Table 5.1 lists how the participants described the different emotional gestures in their own words (translated into English from German).

Very clear was in particular the sad expression (in which Doggy lowers both its upper body and lets its tail hang down), which all three participants understood instantly and which lead all of them to express their sympathy or pity for Doggy. There were different gestures for sadness and disappointment (distinct in how quickly, how far, and for how long Doggy leant down). The reactions to these where not notably different - the disappointment gesture was also understood as a negative or sad emotion immediately, but was generally categorized as being the same as or similar to the sadness it had shown in the beginning when there had been no interaction. One gesture in which the robot ‘shakes its head’ (moves from side to side while in the downwards position) was more readily understood as disappointment. These gestures similarly lead to the participants being sympathetic.

Similarly, obviously happy expressions were quite well understood. In these, Doggy moves its body from side to side and wags its tail. Unfortunately, while all the participants easily recognized the tail wagging motion, this gesture is not as clearly implemented as might be preferred - because of the physicality of the tail, it was not possible for the robot to hold it in as high a position as was planned, and it easily sagged downwards - one of the participants commented on that specifically.

The third type of gesture is that intended to encourage interaction. One version is for a single sound cue and leads Doggy to turn to that particular direction, the other one is for when there are too many opposing cues and it is not clear where to turn to. In these, Doggy typically moves its head forwards and backwards and rotates to different directions when there are several sound cues. While the encouragement during the interaction phase did lead to the participants producing more sounds, they generally found Doggy moving its head towards them more confusing than other gestures. They did verbalize that they understood it to want them to interact with it (or throw the ball towards it), but one participant in particular was unsure of whether they were supposed to do something specific.

When comparing different iterations of the gestures, it became clear that the participants did not immediately distinguish between gestures that were very not obviously different. Instead, they tended to categorize all expressions as either happy or sad - while they did use different descriptors during the interaction (for example ‘disappointed’ was clearer then), the difference or the exact meaning was not clear outside of that context. They did try to find different terms for different gestures, seemingly understanding that the different animations were most likely intended to convey slightly different emotional expressions - however, the comparison between their descriptions shows that they were not generally in agreement for example which gesture was rather ‘happy’ and which one was more ‘excited’ (they used the term ‘aufgeregt’ for either gesture). Additionally, all participants very much preferred exaggerated gestures over more subtle ones, and often preferred simplified ones (clear movement in only one or two directions) over more complicated ones (different combinations of all possible movements).

Interaction experience

All three participants noted that they found both the costume and the movements endearing. Generally, both positive and negative emotions were received with some amusement (e.g. when Doggy was sad in the beginning of the interaction when nothing had happened for a while, this was received with sympathy for the robot but also some mirth at the exaggeratedly miserable motion). However, none of the participants ever tried to provoke

Gesture		**Interpretation by**		
Gesture	**Description**	**Participant 1**	**Participant 2**	**Participant 3**
Happy (1)	sideways	happy		
Happy (2)	sideways + rotation	cheering	happy cheerful	agitated/excited
Happy (3)	sideways (fast)	happy very happy?	agitated	happy
Excited (1)	rotation	cheerful excited	happy agitated/excited	excited?
Excited (2)	sideways + rotation (fast)	cheering	happy excited	agitated excited
Sad (1)	down (slowly)	sad	unhappy	
Sad (2)	down (slowly), up a bit, down	very sad	disappointed	unhappy sad
Disappointed (1)	down (faster)	sad scared?	still unhappy? angry?	frustrated
Disappointed (1)	down, shake head	sad	sad, disappointed	
Encourage (1)	nodding towards player	‘wants me to interact with it’		
Encourage (2)	nodding + sideways (fast)	excited, ‘wants me to interact with it or to excite me’		"wants to interact"
Encourage all	nodding in different directions	very excited, ‘wants us to interact’		

Table 5.1: Participants’ interpretation of different gesture iterations.

a negative reaction. Notably, one participant, after having concluded the test and the interview (they asked for further information on the project, so the robot was returned to its interactive behaviour while the participant was in conversation with me), kept clapping every now and then - in order to keep the robot engaged and from switching into a negative emotion over and over.

They also agreed that this was very much a short-term game, and that the gestures would work well in order to get potential players to interact with the robot in the first place, but not offer considerable incentive to keep someone engaged in comparison with only the ball game. One of the participants noted that Doggy was ‘on a loop’. Asked to elaborate, they explained that they understood that Doggy’s actions depended on their own actions and on time passed between them, but that they had ‘figured out’ how the robot would react in which situation.

Chapter 6

Discussion

In Chapter 2.2 and Chapter 3.2, it was discussed how several other works in this or related areas found the addition of gestures conveying emotion would improve the interaction experience for participants; most notably Carter et al. (2014), who imbued their robot with humanoid gestures which they evaluated both with surveys and by measuring smile scores during the interaction. While in the scope of this work it was not possible to conduct a test large enough to scientifically compare interaction with and without the addition of gestures, the testimonies from the test participants support that the addition of gestures was part of what made their experience enjoyable. Even though the gestures are rather simplistic in their design due to the constraints of Doggy's embodiment, the participants were not only able to accurately judge the expressions most of the time, but reacted overwhelmingly positive to the gesture animations. Furthermore, their reactions seemed to be very much influenced by Doggy's morphology and physical design not only as a dog, but specifically as a cartoonish, animated character: They were amused by and adored both positive and negative reactions, very similar to how one might enjoy the behaviour of a somewhat frustrated pet (or maybe even a toddler), or a similarly 'cute' and decidedly non-threatening cartoon-character. They expressed that they found it endearing, sometimes particularly so when Doggy appeared sad. Just as Fong et al. (2003) stated in "A Survey of Socially Interactive Robots", the participants' expectations and therefore their interaction was very much shaped by the robot's morphology.

The emotional gestures were designed in order to work easily with the available range of motion and DOF, and where therefore stylized and often over-exaggerated, in keeping with the motions that might be used in animated characters on film. The participants generally found it easy to understand the particular motions as the emotion they were intended to convey. When showing different iterations of the gestures, they all preferred the more exaggerated versions, and expressed their belief that these were more easily understood and fit better with the character - which supports the notion that both the physical design created by Tzeng (2013) and the corresponding animated, cartoonish gestures are a good way of dealing with a limited variety of possible motions. Of course, the more exaggerated a gesture has to be to be understood, the fewer distinct gestures can be implemented. In this test, the participants already mainly distinguished between positive ('it's happy') and negative ('it's disappointed/sad/frustrated') emotions, and only tried to find finer differences when asked about that specifically. Originally, I had been hoping to be able to create a broader variety of gestures after a first round of testing on the robot (which ended up not being possible in that time-frame), but considering these findings, I do not think that it is feasible to try to expand on that even further without losing clarity and ease of understanding. Interestingly, in spite of its design, another emotion that can easily be conveyed is actually aggression: when first transferring the code to the robot, I found that it was able of surprisingly fast and abrupt motions, which had a quite startling effect when it moved rapidly forwards towards someone. Of course, this is not something that the robot is intended to project, which is why I did not include such gestures in the eventual user test, and adapted the gestures in such a way as to prevent this occurrence. Overall, it should be positively noted not only that the participants reported that they enjoyed the animations, but particularly that they found them to be encouraging in their interaction with the robot. After initially being careful, all three participants found Doggy's first gestures, as well as the ones after a pause of some time, to be an appeal to interact with it.

In their initial approach however, all three participants were at first curious but cautious. While a ball easily affords to be thrown, a robot - regardless of its appearance - does at least in this case seem to have people wary of potentially damaging something expensive, or behaving 'incorrectly'. They did however try to interact with it by touch, which is invited by the plush costume of the dog. The original robot was indeed able to react to someone moving its bat by hand - I was not able to keep this feature when implementing this project, however, this might be a promising avenue for further work.

Problems in Design and Implementation

What was produced in the scope of this work is, on the one hand, the new interactive behaviour through Doggy's states and emotional gestures and, on the other hand, the implementation of a simulation program for the robot Doggy, and the addition to the ROS project that manages the internal emotional states and can be used to play different animations in accordance with those states. Initially, I had been hoping to be able to expand interaction with Doggy more significantly - either by using movements and gestures as the basis for other games, or by increasing the variety of emotional expressions to make a more distinct and sophisticated internal emotional life of the robot possible. But as the results discussed above indicate, it does seem that the best way of having Doggy perform emotions is to stick with a smaller number of very distinct states and gestures. Therefore, it seems that it might have been more viable to work in the opposite direction of the one I decided on: instead or expanding Doggy's expression and actions, to improve its recognition of outside cues and reactions to that. The work on the sound recognition which was implemented by Bartsch (2015) might have been a good starting point for this. It is unfortunate that I was not able to achieve the integration of that project. This (and other modes of communication/recognition) could be a starting point for future projects on this robot.

When implementing the first test program, I approached this mainly from the perspective of how to control the state machine, and how to render and animate the motions. While I included the manual steering of the robot from the beginning (mostly as a way to test the animation of the motions in the correct directions), I did not originally consider the inclusion of different ways in which a (simulated) person could interact with Doggy; and only added the sound cue since that was one existing project that I specifically wanted to integrate. In retrospect, I suspect that due to this conception of the first program, I constrained myself to concentrating on the robot's movements and animations, instead of considering other possible means of interaction with which a person might trigger Doggy's reactions. In transferring the design to the actual robot, I very much stuck to translating the DoggyTest program to the new framework. At this point, it might have been possible to look into which other possible means of affecting the robot there are from the outside (touch, sight, etc.). I considered this only after having conducted the first user tests. Unfortunately, I had quite underestimated the time I would need to implement the animations on the ROS system in the first place (I was not familiar with ROS beforehand, and did not allow for enough time to familiarize myself with this or the programming environment). Because of this, I did not make any changes or produce any further work on the system after these tests.

The tests themselves provided very helpful insight not only in how someone perceives the emotional expressions of the robot, but particularly how a person approaches interaction with the robot in the first place. The participants' different ideas for interacting with it were what made me consider how it might be valuable to expand especially the means in which the robot can be externally affected. Apart from that, looking back on the test design, I do think that there are interesting possibilities for future tests left open: Firstly, while the think aloud protocol was very useful for recording the participants' thoughts and assessments of the system during the interaction, I would like to have taken a more structured approach to the interview afterwards, designing a survey and maybe including the evaluation on Likert-type scales. Notably, a test like this might make it possible to more accurately compare results, if there was a test with enough participants to distinguish between two groups (one interacting with the robot displaying the gestures, the other one either not seeing any such gestures or using the previous interaction model). While the participants' responses are in line with the idea that their experience in interacting with the robot is improved through the inclusion of emotional gesture expression, another test might produce data that clearly supports this.

Future Work

I believe the most promising course of action to further expand upon and improve the interactive experience with the Doggy robot is to increase the variety of sensors, and build Doggy's possible reactions on those. There are several avenues that could be explored:

The basis for sound recognition already exists. Assuming the technical problems in the sound recognition itself are solved, a first integration with the gestures is already conceived. Of course, sound can be utilized in a much broader scope than 'only' for directional recognition. With voice control becoming more and more common in our daily lives as natural language recognition is more and more integrated into all manner of 'smart devices', it stands to reason that it will become increasingly natural for users to want to talk to computers in general and robots in particular. In Doggy's case, I would argue that it very much makes sense to try to teach it a number of voice commands - inspired, of course, by how a dog might be taught a number of spoken commands. This manner of interaction I believe people might quickly be able to apply to Doggy.

Another manner of interaction that afforded itself to the participants in my test was to touch Doggy. Of course, in terms of Doggy's senses, this rather corresponds to its kinaesthetic sense - it wouldn't be feasible to imbue the robot with a large number of touch sensors; it is however a lot easier for it to sense how its body is being moved and what position it is in. One version of this was already used in the original Piggy project, where it was possible to 'wake up' the robot by moving its bat. This opens up the opportunity for a number of other interactive behaviours as well: Going again by how one might interact with an actual dog; a first idea could be for Doggy to react by bowing its head down towards someone that is touching it in order to be ‘petted’, or movements/‘tricks’ could be taught to Doggy by first showing it manually how to move in the required manner.

Furthermore, sight is another sense that the robot already uses in some capacity, when recognizing a ball and moving towards it. If it is possible to ensure that a ball can be tracked reasonably well why being held, then based on this it could be used for Doggy to track an active player and move accordingly (inspired by how a dog might recognize a ball or similar toy). Additionally, this might be used as a possible way to convey commandos for other tricks - a dog can be taught to understand and react to hand gestures just as to audible commands, and while it would be a lot more difficult to implement a sufficiently sophisticated solution that recognizes a player’s hand position, something like the ball as a tracking point might enable the recognition of bigger gestures. Alternatively, a solution like the Microsoft Kinect could be used in a similar manner if integrated into the robot.

Finally, when discussing the theoretical background in Chapter 2.2, I established that I would model Doggy’s internal emotional state, but could not attempt to also have it recognize a human’s emotional expressions and model its partner’s probable internal state. Continuing the project in this direction would not only expand the means of interaction with Doggy, but also let Doggy become a ‘more social’ robot in the theoretical sense. Right now it expresses itself, but does not try to understand the person interacting with it - further work on the robot’s different senses could make it possible to also pursue this direction in the future.

List of Figures

Bibliography

References

Bartsch, Michel (2015). *Sound of Interest. Ein Ballspielroboter hoert stereo.*

Breazeal, Cynthia (2003). “Toward sociable robots”. In: *Robotics and Autonomous Systems* 42.3. Socially Interactive Robots, pp. 167–175.

Brooks, Andrew G. et al. (2004). “Robot’s Play: Interactive Games with Sociable Machines”. In: *Comput. Entertain.* 2.3, pp. 10–10.

Carter, E. J. et al. (2014). “Playing catch with robots: Incorporating social gestures into physical interactions”. *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pp. 231–236.

Dautenhahn, Kerstin (2007). “Socially intelligent robots: Dimensions of human-robot interaction”. In: *Philosophical transactions of the Royal Society of London. Series B, Biological sciences.* Vol. 362, pp. 679–704.

Dautenhahn, Kerstin and Aude Billard (1999). “Bringing Up Robots or&Mdash;the Psychology of Socially Intelligent Robots: From Theory to Implementation”. *Proceedings of the Third Annual Conference on Autonomous Agents.* AGENTS ’99. Seattle, Washington, USA: ACM, pp. 366–367.

Dautenhahn, Kerstin, Bernard Ogden, and Tom Quick (2002). “From Embodied to Socially Embedded Agents - Implications for Interaction-aware Robots”. In: *Cogn. Syst. Res.* 3.3, pp. 397–428.

Dennett, D.C. (1989). *The Intentional Stance.* A Bradford Book.

Dryer, Christopher (1998). “Dominance and valence: A two-factor model for emotion in HCI”. In: *Emotional and Intelligent: The Tangled Knot of Cognition. Papers from the 1998 AAAI Fall Symposium*, pp. 76–81.

Fong, Terrence, Illah Nourbakhsh, and Kerstin Dautenhahn (2003). “A Survey of Socially Interactive Robots”. In: *Robotics and Autonomous Systems.* Vol. 42, pp. 143–166.

Goetz, J., S. Kiesler, and A. Powers (2003). “Matching robot appearance and behavior to tasks to improve human-robot cooperation”. *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.* Pp. 55–60.

Kendon, Adam (1992). “Some Recent Work from Italy on "Quotable Gestures (Emblems)"”. In: *Journal of Linguistic Anthropology* 2.1, pp. 92–108.

Kerepesi, A et al. (2006). “Behavioral comparison of human-animal (dog) and human-robot (AIBO) interactions”. In: *Behavioural processes.* Vol. 73, pp. 92–9.

Laue, Tim et al. (2014). “An Entertainment Robot for Playing Interactive Ball Games”. *RoboCup 2013: Robot World Cup XVII.* Ed. by Sven Behnke et al. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 171–182.

Liu, C., P. Rani, and N. Sarkar (2006). “Human-Robot Interaction Using Affective Cues”. *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp. 285–290.

Moshkina, L. and R. C. Arkin (2005). “Human perspective on affective robotic behavior: a longitudinal study”. *2005 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1444–1451.

Nass, Clifford I. et al. (1997). “Human Values and the Design of Computer Technology”. In: ed. by Batya Friedman. Stanford, CA, USA: Center for the Study of Language and Information. Chap. Computers Are Social Actors: A Review of Current Research, pp. 137–162.

Nielsen, Jakob (1989). “Usability Engineering at a Discount”. *Proceedings of the Third International Conference on Human-computer Interaction on Designing and Using Human-computer Interfaces and Knowledge Based Systems (2Nd Ed.)* Boston, Massachusetts, USA: Elsevier Science Inc., pp. 394–401.

Persson, Per, Jarmo Laaksolahti, and P Lonnqvist (2001). “Understanding socially intelligent agents-a multilayered phenomenon”. In: *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans* 31.5, pp. 349–360.

Reeves, Byron and Clifford I. Nass (1996). *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places.* New York, NY, USA: Cambridge University Press.

Riley, Marcia and Christopher G. Atkeson (2002). “Robot Catching: Towards Engaging Human-Humanoid Interaction”. In: *Autonomous Robots* 12.1, pp. 119–128.

Robins, Ben et al. (2010). “Human-centred design methods: Developing scenarios for robot assisted play informed by user panels and field trials”. In: *International Journal of Human-Computer Studies* 68.12, pp. 873–898.

Schuethe, Dennis and Udo Frese (2014). *Task Level Optimal Control of a Simulated Ball Batting Robot.*

Shibata, T. (2004). “An overview of human interactive robots for psychological enrichment”. In: *Proceedings of the IEEE* 92.11, pp. 1749–1758.

Tzeng, Yi-reh (2013). *New Dog, New Tricks - Human-Robot Interaction and Appearance Design of a Ball Playing Robot.*

Online References

Beucler, Sylvain (2018). *OpenGL Programming/Post-Processing - Shaders.* URL: `https://en.wikibooks.org/wiki/OpenGL_Programming/Post-Processing#Shaders` (visited on 09/24/2018).

Open Source Robotics Foundation (2018). *ROS Documentation.* URL: `http://wiki.ros.org/ROS/` (visited on 09/24/2018).

RoboCup Federation Inc (2016). *About the RoboCup - Objective.* URL: `https://www.robocup.org/objective` (visited on 09/29/2018).

Sony Computer Entertainment Inc. (2007). *Vectormath_aos.h.* URL: `http://hackage.haskell.org/package/bullet-0.2.2/src/bullet/vectormath/sse/vectormath_aos.h` (visited on 09/28/2018).

User MaxH (2010). *Rotate Box With Arrow Keys.* URL: `https://www.opengl.org/discussion_boards/showthread.php/169966-How-to-control-cube-rotation-using-Keyboard-keys` (visited on 09/29/2018).

Zachmann, Gabriel (2015). *Computergraphik - WS 14/15.* URL: `http://cgvr.cs.uni-bremen.de/teaching/cg1_1415/index.shtml` (visited on 09/29/2018).

Eigenständigkeitserklärung

Hiermit versichere ich, dass ich die vorliegende Bachelorarbeit selbstständig und nur unter Verwendung der angegebenen Quellen und Hilfsmittel verfasst habe. Die Arbeit wurde bisher in gleicher oder ähnlicher Form keiner anderen Prüfungsbehörde vorgelegt.

Bremen, den 01.10.2018

Laura Spillner

Appendix A

A.1 Animated Renditions of Gestures

These (simplified) animations illustrate the possible gestures explained in Chapter 3.3. Note: in order to view the animated graphics, it is necessary to use a PDF reader capable of displaying those.

Figure A.1: Happy (1)

Figure A.2: Excited (1)

Figure A.3: Sad (1)

Figure A.4: Sad (2)

Figure A.5: Disappointed (1)

Figure A.6: Encourage (1)

Figure A.7: Encourage (2)