Universität Bremen

Integration and Evaluation of an Interactive Ball-Playing Robot

Bachelor Thesis

Submitted for the Attainment of the Degree

Bachelor of Science (B.Sc.)

at the

University of Bremen

1. Examiner:	Udo Frese
2. Examiner:	Robert Porzel
Submitted by:	Pascal Weidenbach
Matriculation number:	4300303
Course of Study	Digital Media
Target Date:	23.09.2019
Date of submission:	20.09.2019

Abstract

While most robots use sounds and animations, the effect of the integration of these stimuli into the system are often times not analyzed. In this thesis, the gameplay routine of the ball-playing robot Doggy is extended to integrate emotional expressions with sounds and animations. A study was executed to compare the base version to the version with integrated sounds and animations. Qualitative observations reveal that mentioned stimuli are able to evoke emotional responses from users matching the emotional expressions of the robot. Sounds and animations alone are able to entertain users during downtime of gameplay. The integration of sounds and animations was evaluated to be a beneficial addition for the robot. However, the effect of the integration on the system could not be determined as planned, due to hard- and software issues during the study.

Table of Contents

List of Figures

1. Introduction

1.1 An Introduction to the Robot Doggy

The robot used for this thesis is Doggy, which was created by the workgroup "Multisensorische Interaktive Systeme" at the university of Bremen (Bartsch, 2015). Doggy is an interactive ball-playing robot used for entertainment at bigger events such as exhibitions or firm events. The stationary robot is able to react to balls thrown to him by intercepting them. For this, the robot moves its upper body towards the predicted interception location. The balls are tracked based on stereo images from two cameras.

Its predecessor is Piggy, which is also able to interact with balls (Laue, Birbach, Hammer, & Frese, 2014). Piggy has a state machine with five finite states. This allows the robot to switch between expressing joy in the cheering state and sadness in the complain state. Doggy is based on Piggy's design philosophy of safety, flexibility and easy appliance, reasonably low costs, reactivity and throughput. The design of Doggy is of an anthropomorphized cartoonish dog (Tzeng, 2013). The robot was designed with a new state machine that is able to express sadness and happiness with body animations. The attached, controllable tail is also used to express these emotions. Later, two microphones were added to the robot, which allow the robot to localize sound sources spatially (Bartsch, 2015). The design of animations was expanded on and an updated state machine was created (Spillner, 2018). Eight emotional states with three expressiveness levels were added to the newly designed state machine. The animations for the states were sketched out and implemented onto Doggy. However, the animations were not compatible with the gameplay. Furthermore, the implementation of the animations was not available for this thesis.

1.2 Aims and Objectives

Several scientific works have been drafted to improve Doggy in various aspects. This thesis will build upon the current state of Doggy by adding new behavior layers to the gameplay.

While most robots use sounds and animations, the effect of the integration of these stimuli into the system are often times not analyzed. This thesis aims to integrate sounds and animations into the gameplay routine of the robot, to improve the appeal and likeability and to raise the interest of users during interaction. It also aims to create a comparable result of a robot with the integration of mentioned stimuli and without.

Initially, the background of Doggy will be examined, to gain insights into its previous design. Particularly relevant is the previous design philosophy regarding animations and state machines.

Secondly, related work will be analyzed to specify the current state of art regarding robots using sounds and animations. Related work regarding the design of sounds and animations with relevancy to robots will be evaluated.

Afterwards, with the background and research in mind, new sounds and animations will be proposed for the robot. Furthermore, the state machine will be adapted to fit to the analyzed issues.

Then, the proposed design will be implemented and documented in detail, in order to simplify future work on Doggy and to showcase the implementation process. A focus will be placed on merging the created sounds and animations to the gameplay module of Doggy.

Finally, a study will be conducted to evaluate how the integration of sounds and animations into the system affects various aspects regarding the perception of users.

2. Background of the Ball-Playing Robot Doggy

2.1 Doggy's predecessor Piggy

Piggy (cf. Figure 2-1) is Doggy's predecessor and is explained in detail in Laue et al. (2014). The title "An Entertainment Robot for Playing Interactive Ball Games" shows the general design idea. Piggy was intended as a short and exciting experience for events like firm parties, exhibitions, fairs etc. to attract people by being fun, interactive, social, clever and a hands-on robot experience.

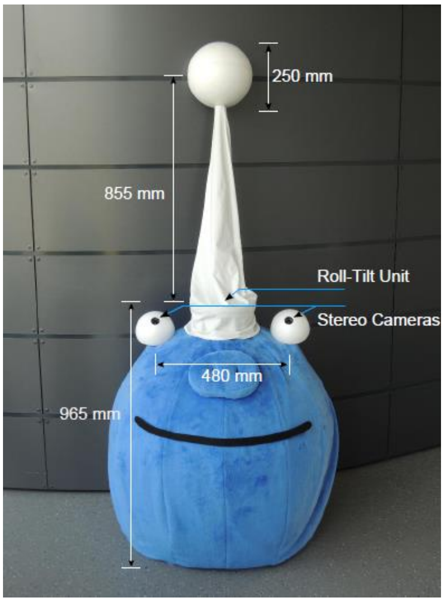

Figure 2-1: The measurements of Piggy.

The robot depicts the head of a blue pig with a white bat as its main interaction tool for the ball playing game. Piggy's roll-tilt unit has two degrees of freedom (DOF), an unconventional amount that was chosen because it "dramatically reduces weight and costs and increases safety and reactivity" (Laue et al., 2014, p. 4), yet is enough to reach balls for the game. The head contains a computer, motors and a power supply. The balls are tracked by Piggy's eyes, which contain two cameras, allowing him to process images spatially. Apart from the ball playing game, other interactions are possible such as waking Piggy up from the sleeping state by lifting up his bat or by hitting him, which causes a loud "ouch" reaction.

The main concepts for the design process were safety, flexibility and easy appliance, reasonably low costs, reactivity and throughput. Laue et al. (2014) explain them as following:

> "1. Safety. For interacting with humans, the system has of course to be safe. It must be guaranteed that no person gets injured by the robot.
>
> 2. Flexibility and easy appliance. Similar to existing entertainment devices, the hardware and software must operate under various conditions. Amongst others, this concerns elements in the perceivable environment, lighting conditions, or the behavior of humans participating in the game. The system must be transportable and a non-expert must be able to set it up.
>
> 3. Reasonably low costs. The costs for the construction and maintenance of many current robot systems are not economical for any commercial activities. A successful entertainment robot should have a prize comparable to current non-robot entertainment devices.
>
> 4. Reactivity. A robot that interacts with people in a game (e.g. some kind of ball game) must have a level of reactivity comparable to that of humans. A significantly lower performance would result in a boring game that does not challenge the humans.
>
> 5. Throughput. Often events have many visitors and a large number of people should interact quickly with the system." (Laue et al., 2014, p. 2)

Piggy uses a state machine to express emotional states (cf. Figure 2-2). The behavior of Piggy is designed to prioritize the ball playing state, so the states cheer and complain can be interrupted when a ball with a reachable trajectory is detected. Two animations exist for the states cheer and complain. In the waiting state, the bat is set to follow the player to highlight Piggy's attention. This person tracker was created with a *Microsoft Kinect* sensor.

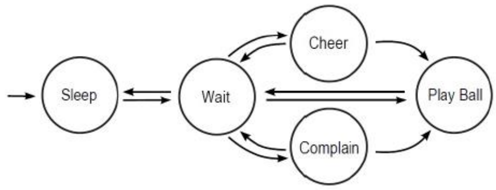

Figure 2-2: The five possible states of Piggy

2.2 Appearance and Animation Design by Tzeng

In "Human-Robot Interaction and Appearance Design of a Ball Playing Robot" Tzeng (2013) designed a new shape, costume and animations for the descendent named Doggy (cf. Figure 2-3). Like Piggy, Doggy is also able to interact with balls. The same appearance design strategy is used, since it is following the idea of a stuffed animal. This has advantages regarding the familiarity according to the uncanny valley. The improved version is able to move its torso and is equipped with a steerable tail, fitting to the appearance of a dog. The bat, which is now the robot's head, is better incorporated into the design, as the robot is an anthropomorphized dog. However, since the eyes of the robot are now in the moving head of the robot, the location of the cameras had to be changed. They are now in Doggy's waist and are hidden in a belt.


Figure 2-3: The robot's new design.

Tzeng (2013) added a "Classroom" for Doggy, with which Doggy can be "taught" to move. Animations can be recorded by using a Wii Classic Controller. Two joysticks are used to control four dimensions of Doggy. Torso and tail animations can be recorded, which form a "comby gesture". The created animations can be played in the "Classroom" tool. The design idea for the creation of animations is to create multiple takes of an animation. Then the most congenial attempt can be chosen. Tzeng (2013) evaluated this approach, as opposed to one with keyframing on a timeline, to be more intuitive and efficient.

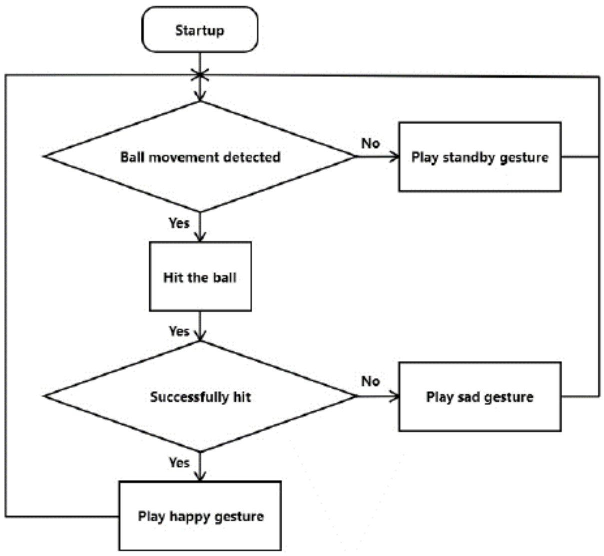

Figure 2-4: The state machine created by Tzeng (2013).

The state design for Doggy includes the three types happy, sad and standby, in which the prerecorded animations are played (cf. Figure 2-4).

2.3 Emotional Expression Design by Spillner

In "Interacting With a Ball-Playing Entertainment Robot" Spillner (2018) expands upon the gesture design of Doggy by designing a personality matching for Doggy from which a new internal state design derived. This design consists of eight states to express emotions. These have up to three expressiveness levels. The three levels of happiness are described as:

> "• 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." (Spillner, 2018, p.26)

Spillner (2018) proposed several new animations for Doggy to display emotions in varying intensity levels. These were sketched out as visible in Figure 2-5. A software was developed to quickly visualize new animations, similar to the design tool created by Tzeng (2013). Based on C++, the tool renders a 3D model of Doggy by using OpenGL and GLUT. This model can be controlled with a keyboard to allow an interactive design of animations. Then the designed animations were to be transferred onto the real robot by adding an animation package to the ROS network. Unfortunately, this package was not available for the majority of this thesis and only found later in an unfinished state in a backup folder.

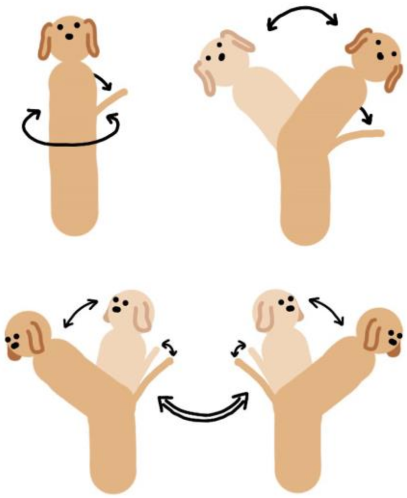

Figure 2-5: The animations Excited, Happy(1)/(3) and Encourage All.

Spillner (2018) attempted to implement the sound cues of Bartsch (2015) into the logic of the animation package, so that Doggy would turn to a user when a sound such as a clap or a loud whistle was detected. However, due to implementation problems this was only added into the evaluation as a Wizard of Oz type of interaction. The evaluation of the system was separated into two parts due to problems with the implementation: *a gameplay part* and a *sound and gesture part*. Generally, the perception of the gestures was correct. The participants often times chose fitting or the exact emotion that was expressed by Doggy. "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" (Spillner, 2018, p. 46). However, since the gestures were only evaluated outside of the actual gameplay, further research needs to be conducted on how the gestures affect the entertainment factor of the gameplay.

2.4 Stereo Sound Detection by Bartsch

In his master thesis "Sound of Interest - Ein Ballspielroboter hört stereo" Bartsch (2015) focusses on allowing Doggy to hear in stereo, to detect sounds. With the help of a phase only matched filter and a frequency-learning filter as well as common algorithms, sounds that surpass the general detected noise level will be categorized as interesting. The filters used by Bartsch (2015) even work in noisy environments, matching to Doggy's intended use cases such as exhibitions. When an interesting sound is found, its position is calculated with the help of the stereo speakers. With the found position, Doggy is able to react to the sound by turning to the position of origin with a deviation of 2 degrees.

As previously mentioned, Spillner (2018) tried to add the sound cues of Bartsch (2015) to the gameplay of Doggy, but due to time issues this was not achieved. This work will focus on adding gestures and sound to the gameplay and due to time limitations adding the work of Bartsch (2015) will no be pursuit.

3. Related Work

Adding sounds and animations to Doggy is expanding upon the interaction between the human and the robot, thus falls in Human Robot Interaction (HRI). Dautenhahn (2007) coins HRI as a "highly interdisciplinary area, at the intersection of robotics, engineering, computer science, psychology, linguistics, ethology and other disciplines, investigating social behavior, communication and intelligence in natural and artificial systems". Thus, it is a highly complex area that can be analyzed in a myriad of ways. This thesis, however, only focusses on HRI topics relevant to the addition of sounds and animations onto Doggy.

Initially, various robots will be analyzed that are relevant regarding the design of sounds and animations. In addition, it serves to give an overview of the current state of the art. This chapter also addresses possible design approaches for socially designed robots. The following two related work sections approach the topics sounds and animations for robots in detail.

3.1 Related Robots

Sony's AIBO is a pet dog robot which was designed with the goal to "maximize the lifelike appearance" (Fujita, 2001, p. 1). To achieve this, a highly sophisticated architecture was built. It includes a behavior model, randomness, instincts/emotions, learning ability, long term adaptations and various motions. AIBO is able to interact with people by processing sound and images or by using its touch sensors. A hit on the head is recognizable and causes AIBO to emit a surprised sound. The general design idea is that the more detailed the robot is, the more it will feel like an actual living creature. The over 130.000 purchase requests from all over the world substantiate the interest AIBO sparks in people (Fujita, 2001). For Doggy, the design philosophy of creating an unpredictable, detailed robot seems to be relevant for the design of sound and gestures. Manifold expressions mixed with randomness should therefore appeal more to visitors.

Paro is a robot embodied by a baby seal covered with white fur. Ubiquitous surface tactile sensors allow the recognition of human contact through the fur. Light sensors constitute a visual interpretation. Speech recognition and the detection of the direction of sound is possible through a microphone. Additionally, Paro is equipped with balance sensors. The robot is a tool for robot assisted therapy (RAT). Paro is used in care houses for elderly people. Consequently, natural social interaction is of importance. Paro is able to express behavior through facial expressions and emotional sound cues. Behind the behavioral states is a sophisticated system with proactive, reactive and physiological behavior. For example, it consists of a long-term memory, which uses reinforcement learning, to change its behavior to receive more stimulation, such as stroking. Reactive behavior, for example, will occur when a sudden, loud sound was detected. Paro then looks towards the detected direction of the sound (Wada & Shibata, 2007). Paro's intended interaction period exceeds Doggy's by far; consequently, its hardware and software are more sophisticated, allowing for extended socially acceptable interaction.

The social robot Probo (Saldien, Vanderborght, Goris, Van Damme, & Lefeber, 2014) also focusses on RAT. In therapy users have to interpret the states of the robot with great ease, thus the authenticity of emotional expression is of great importance. To create natural motions, keyframes are used to interpolate animations to create an "illusion of life" (Saldien et al., 2014, p. 1). Additionally, a complex emotional system decides how Probo responds to interactions such as petting. For Doggy, the focus on natural motions seems particularly relevant, as detailed, varied animations are more appealing and more entertaining.

The robot Daryl (Embgen, Luber, Marco, Vanessa, & Kai, 2012) makes use of robot specific modalities "such as certain sounds, light, colors, shape, or robot-specific body parts." (Embgen et al., 2012, p. 1). They argue that robots should interact with humans in a robot-specific way, just like animals interact in their own specific way with tail movements, sounds etc. Hence, Daryl uses colored light to indicate emotional states. However, since Doggy is more of an anthropomorphized, cartoonish dog, its interaction modalities should relate to those of a dog.

With the robot RAMCIP Antona et al. (2019) focus on emotional state interpretation based on facial and sound related expressions. A screen allows RAMCIP to visualize a human-like face and speakers play RAMCIP's synthesized voice of a human. Especially interesting for Doggy is the creation of RAMCIP's voice, as adaptations to the vocal and verbal part of sound allow for different state interpretations. The study also raises questions regarding cross-modal influences between sound and visual stimuli. For Doggy, these influences have to be analyzed regarding sound and gestures.

3.2 Social Robot Design

In "A Survey of Socially Interactive Robots: Concepts, Design, and Applications" Fong, Nourbakhsh, & Dautenhahn (2002) offer a classification to design robots and to create system parts for socially interactive robots. They state that for socially interactive robots, regardless of the use case, humans will interact with them, as they would do with other humans. This is due to the natural desire of humans to interact socially. The human social interaction has certain expectations that, when met, result in a satisfying interaction, which also will allow the human to feel capable and confident.

3.2.1 Design Approaches

Fong et al. (2002) mention one robot area that is desirable to have social interaction, based on the research of Fogg (1999), who identifies three types of computer persuasion: the computer as a *tool*, *medium* or *social actor*. A digital pet is at the peak of the *social actor* side of the triad, thus should create relationships with users and persuade by "changing behavior, feelings or attitudes of humans" (Fong et al., 2002, p. 4). Thus, it is the robot designer's task to allow the human to feel this way. A robot has to be designed to be capable of handling interactions in a social manner and fitting to the human's expectation, so that a social bond between the two entities can be built. In the following, two design approaches will be analyzed, which are helpful to create social robots.

3.2.1.1 Biological Design

One design approach is the biologically inspired one. The robot is designed to recreate the characteristics of a biological creature to imitate the social behavior of an animal. Humans have repeated contact with animals in their life; nature itself is the standard that the expectations of humans regarding robots are measured against. It is assumed that in order for a robot to feel natural "it must have a naturalistic embodiment, it must interact with its environment in the same way living creatures do, and it must perceive the same things that humans find to be salient and relevant" (Fong et al., 2002, p. 5).

As HRI is an extraordinarily interdisciplinary area, even ethology, the scientific study of animal behavior in their natural environment, can be considered to improve the interaction between robots and humans. In their journal article "An Ethological and Emotional Basis for Human-Robot Interaction" Arkin, Fujita, Takagi, & Hasegawa (2003) focus on ethology "to provide the appearance of life-like activity" for sophisticated robots, like Sony's AIBO, that are to be entertaining for longer periods.

In addition, another relevant biologically inspired topic is the theory of mind. Humans build mental images of their own and of other people's "beliefs, goals, perceptions, feelings, and desires" (Fong et al. 2002, p.6). According to Ziv & Frye (2003), beliefs, desires and actions stand at the core of the people's daily theory of mind. Their "analysis implies that people typically engage in *actions* because they *believe* those actions will satisfy their *desires* (Wellman, 1991)" (Ziv & Frye, 2003, p.1). Humans also project their theory of mind onto other biological creatures and onto social robots. Therefore, it is important for robots to indicate interest in order for humans to deduce robot desires. Fong et al. (2002) describe mutual points of interest as excellent initiators to build a theory of mind. Finally, after an action was performed, it is crucial to react with appropriate feedback on whether the robot's desire was fulfilled, to allow humans to reinforce on their theory of mind.

A research paper from Kerepesi, Kubinyi, Jonsson, Magnusson, & Miklósi (2006) compares behavior between human-animal and human-robot interactions and shows the success of biologically inspired designs. "Previous questionnaire studies on human-robot interaction showed that people describe their relationship with AIBO similar to a relationship with a dog puppy (Kahn et al., 2003), attribute animal characteristics to the robot and view it as a family member (Beck et al., 2004)" (Kerepesi et al., 2006, p. 96). Additionally, their research indicates that AIBO was as entertaining and arousing as a real dog puppy for their subjects.

3.2.1.2 Functional Design

Apart from the biologically inspired design, there is the functional design. This design is less sophisticated, as it does not base on scientific work or biology. Fong et al. (2002) explain that it is not mandatory to grasp how the mind functions to create a robot that appears to be socially intelligent. By adding traits to the robot, which humans tend to identify in social intelligent beings, the impression of an independent social robot can be built. However, because this approach is less complex than the biologically inspired one, robots designed with functional design tend to specialize in a smaller range of tasks. Therefore, the robot also focusses on providing a limited, but well-defined experience, as opposed to the biological design. A counterexample is the previously mentioned, biologically inspired AIBO, since it offers richer and deeper interaction possibilities.

The advantages of functional design lie in its simplicity. For example, short interaction intervals with non-repeating users allow a rather unsophisticated social intelligent agent to still appear as socially intelligent to its users. Then even simple prerecorded sounds of dialogue or animals may suffice as social interaction (Fong et al., 2002).

3.2.2 Robot State Machine Design

In order to express emotions robots often use state machines. All of the previously mentioned robots use some kind of state machine to behave more socially. It is the robot's internal representation of emotion or behavior and allows external or internal events to change the behavior of the robot. Bartneck (2002) created a five step model for emotion processing, which is helpful for designing state machines for social robots.

> "Classification**:** What do I feel about what just happened?
> Quantification**:** How much do I feel about it?
> Interaction**:** How does this affect what I was already feeling?
> Mapping**:** What should I do to express this feeling?
> Expression**:** How should I do that?" (Ribeiro & Paiva, 2012, p. 4)

3.3 Related Work Regarding Sounds

The interaction with an entertainment robot such as Doggy should evoke an emotional response, such as happiness or surprise. Sound, as another stimulus, has the possibility to enhance the experience Doggy provides. Therefore, research is analyzed to see what robots use sound, how sounds are created and which impact sounds have in aiding people to determine a robot's expressed emotion.

3.3.1 Usage of Sounds by Robots

The correct localization and interpretation of sound has gotten a lot of attention by scientific authors, one of which being the previous work on Doggy by Bartsch (2015). Okuno, Nakadai, Hidai, Mizoguchi, & Kitano (2001) focus on tracking multiple speakers in a noisy environment, while suppressing the robots own voice. Pineau, Montemerlo, Pollack, Roy, & Thrun (2003) present their nursing robot Pearl, which uses speech recognition to interact with elderly people. Paro, the seal robot detects the direction of sound and uses speech recognition (Wada & Shibata, 2007). By using eight microphones Valin, Michaud, Rouat, & Letourneau (2004) detect sound omnidirectionally. There are also several robots that output sound, such as AIBO, which expresses joy by playing a sound of laughter (Fujita, 2001); Piggy (Laue et al., 2014); Pearl (Pineau et al., 2003); Paro, the seal robot, which plays sounds of a seal (Sharkey & Wood, 2014); My Real Baby, which plays infant sounds (Fong et al., 2002) or RAMCIP, which uses a synthesized voice (Antona et al., 2019). As shown, robot sound recognition, localization, synthetization and playback is used in various amounts of robots and are therefore state of the art.

3.3.2 Expressing Emotions by Emitting Sounds

In “My robot is happy today: how older people with mild cognitive impairments understand assistive robots’ affective output” Antona et al. (2019) target the expression of emotions with the assistive robotic platform RAMCIP with the goal to create a helpful, empathic robot for elderly people with mild cognitive impairments or an Alzheimer disease. As the robot’s goal is to support people in everyday life, social interactions are essential for the ease of use. Thus, emotional responses are key for a quick interpretation of the robot’s state, facilitating the ease of use.

The anthropomorphic robot RAMCIP has a screen at head level, on which eyes, eyebrows and a mouth are animated. The face combined with a synthesized voice allows RAMCIP to express his emotions. The facial expressions were chosen based on an experiment in which users chose the most appropriate images from a large set to convey the desired emotional states. To express emotions by speech, Antona et al. (2019) focused on changing its vocal part, as merely changing the (semantical) verbal part leads to an unnatural “robotic” voice. Prosodic cues are used to add emotion to the vocal part of speech. By adjusting the volume, rate and pitch of the voice, the intonation of the voice is adapted to fit to the desired emotional state. The choice of the cue parameters for the various expressions is not explained, however the values are reasoned with as for example: “the robot speaks with [a] slightly higher speaking rate and [a] significantly higher pitch to expresses[sic] excitement” (Antona et al., 2019, p. 421).

An experiment was done to test whether the intended spoken emotions could be matched to the intended facial expressions. The verbal part always matched emotionally to the vocal part of the speech. The participants listened to a sentence and had to choose the most fitting facial expression from a set of eight images with eight different emotional states. The recognition of emotional valence turned out to “yield good results”(Antona et al., 2019, p. 423), however the average recognition of the exact emotion was only ~17.8%, that is ~37% lower than the average recognition rate for the facial expressions. Some emotions such as happiness and anger were more easily identified correctly as for example disappointment or anxiety (Antona et al., 2019).

This study shows that it can be difficult for participants to deduce exact emotions based on speech, even when both vocal and verbal parts were designed to represent the desired emotion. However, since the participants had to choose from the facial images of RAMCIP instead of directly choosing emotions by words, this task was more difficult. The same argument can be made for the emotional valence. In addition, the setup was not optimized as the “configurations of faces and prosodic cues […] need further fine tuning and testing.” Nevertheless, this study demonstrates that audio output alone has the potential for users to identify emotions, although it is unclear which role the verbal or vocal part had in the participant's choice of emotion.

Fong et al. (2002) specify speech as a "highly effective method for communicating emotion" (Fong et al., 2002, p. 12) and list parameters to change the emotional factor of speech: "loudness, pitch (level, variation, range), and prosody", which Antona et al. (2019) also adjusted for their study. Research reveals that the vocal adjustments regarding emotions are similar amongst speakers (Fong et al., 2002), and can therefore be mimicked by robots to become more social.

3.3.3 Effects of Audiovisual Stimuli

Humans are trained to interpret emotions with all available senses. Thus, sound should not be analyzed in isolation only. The following part focuses on audiovisual stimuli and how they mutually influence each other. This research is especially relevant, since the addition of sounds to Doggy's gameplay will result in audiovisual stimuli. Unfortunately, the research regarding the effectiveness of sound in HRI is rather thin. Not many scientific works focus on the impact sounds of robots have on the user, especially for pet robots. However, science that does not specifically focus on robots can be evaluated to measure the effects of sounds on humans, since it can still be relevant for HRI.

Stock, Grèzes, & Gelder (2008) state that "there is now considerable evidence that multisensory stimuli presented in spatial or temporal proximity are bound by the brain into a unique perceptual gestalt" (Stock et al., 2008, p. 185). Okuno et al. (2001) identify sound as key to improve visual enjoyment and HCI (human computer interaction). In their study, Gelder and Vroomen (2000) analyzed the effect tone of voice has on facial emotional expression. Individuals were presented with a sad or a happy facial expression, accompanied by a sentence that had low emotionally influencing semantics. The sentence itself was spoken in either a happy or a sad tone of voice. Participants were told to ignore the voice and to categorize the facial expression into either sad or happy. "The results indicated a clear crossmodal bias, e.g. a sad facial expression paired with a happy voice was recognized more as happy, compared to when the same facial expression was paired with a sad voice" (Stock et al., 2008, p. 186). It can be concluded that a human's behavioral recognition is influenced by the tone of voice, and even does so when humans are told to ignore it.

In their study "Body Expressions Influence Recognition of Emotions in the Face and Voice" Stock, Righart & Gelder (2007) conclude that "when observers make judgments about the emotion conveyed in a voice, recognition is biased toward the simultaneously perceived whole-body expression." In a similar study, Stock et al. (2008, p. 186) "investigate the influence of human and environmental emotional auditory information on the recognition of emotional body expression." In this study, they present body expressed emotions with either human or animal auditory emotions. Both perception channels induced either happy or fearful emotions. By mixing them, all possible constellations were tested. Their results reveal, that "recognition of body expressions is influenced by non-verbal vocal expressions" (Stock et al., 2008, p. 187). When a happy body language was accompanied with happily singing birds, participants were more likely to identify the body language as happy, as compared to the body language on its own. In contrast, the recognition was worse when sounds of aggressively barking dogs were played. Additionally, this effect was stronger for emotions conveyed by a human voice.

Stock et al. (2008) hypothesize, that crossmodal influences are stronger for matching audiovisual stimuli, such as a matching human body and voice. Chen & Spence (2017) show, that the trigger that causes this effect was observed and discussed by many authors before and was coined as the unity assumption. "According to [this], an observer assumes that two different sensory signals refer to the same underlying multisensory event" (Vatakis & Spence, 2007, p. 1).

Vatakis & Spence (2007) conducted a study and found empirical support for the unity assumption. Participants perceived a human voice and a human mouth. Both voice and mouth were matching in terms of what and how it was said, but the gender of the pairs was not always matching. The participants had to determine which stimulus was perceived first, however, the stimuli were desynchronized by up to ±300ms. Participants performed worse when the gender of the audiovisual stimuli matched, which the unity assumption supports. The matching audiovisual stimuli more convincingly caused the participants to believe that both stimuli originated from one unique event (unity assumption), therefore causing trouble in the ability to differentiate the two.

Taking advantage of the unity assumption in HRI is desirable to allow users to interpret social robot states more congruently and to add to the authenticity of the robot.

3.4 Related Work Regarding Animations

"Robots appearance has made great progress over the years, but movements are often an afterthought in the design process" (Balit et al., 2016, p. 1). This often leads to social robots that move mechanically and fail to achieve what is called the illusion of life. In several pieces of current literature it is mentioned that a key factor for a social robot is creating the illusion of life (Saldien, 2014) (Balit et al., 2016). The illusion of life refers to the viewer perceiving the animated character as something that appears to be alive as opposed to something that is controlled and moves more like a work-oriented robot (Saldien, 2014). In a framework regarding animacy by Piaget the "importance of movement and intentional behaviour" are highlighted to create the illusion of life (Bartneck, Kulić, Croft, & Zoghbi, 2009, p. 74).

Fong et al. (2002) reveal that humans in general are influenced emotionally by body movement, such as dance. "Computer games, such as The Sims, Creatures, or Nintendo Dogs show that lifelike creatures can deeply involve users emotionally. This involvement can then be used to influence users" (Bartneck et al., 2009, p. 74). The same applies to robots. Existing design guidelines for animating virtual characters can be used for robot animation design. "In order to bring robots to life - such that they show behavior that can be naturally understood and anticipated - principles known from the field of character animation should be applied"(Breemen, 2004). Ribeiro & Paiva (2012) apply the twelve animation principles of Disney to robots. In the following, the most relevant parts of the twelve principles by Ribeiro & Paiva (2012) will be explained.

3.4.1 Twelve Animation principles by Ribeiro & Paiva

Squash and Stretch

Non rigid bodies squash and stretch but keep their volume when in motion, however this rule doesn't apply to most robots as they usually use rigid components.

Anticipation

The anticipation of movements and actions by the viewer is desirable. The animated actor will appear more natural if signs, such as eye movements, foreshadow their animations because it allows the viewer to predict their behavior.

Staging

Staging guides the focus of the viewer onto the most relevant parts in the scenery. The use of lights, camera angles, sounds or additional objects can accomplish this. For robots, a multi-modal expression is possibly more adequate.

Straight Ahead and Pose-to-Pose Straight ahead action means the animator starts out at a pose and designs the frames from that onwards, so the design process is very open without a clear end position in mind. Pose-to-Pose means animating towards a planned sequence of poses. First, key positions are determined which are then connected by breakdown poses that create a realistic transition between the key poses.

Follow-Through and Overlapping Action
After the anticipative movement and the action, a final motion is added which acts as a counterpart to the anticipative movement. The follow-through movement is caused by the action itself. If an actor were to jump he would first attempt to gain momentum by leaning backwards with body and arms (anticipative movement), then perform the jump (action) and after the impact try to stabilize himself (follow-through). Animating in this way creates the impression that the actor follows the laws of physics thus makes him appeal as if he is part of our world.

Slow In and Out
Most movements in the real world start and end slower than the motion in between due to acceleration and deceleration. It appears natural and feels right for the viewers, leaving a more realistic impression. This is especially relevant for animating robots because they usually move at a constant speed causing the typical robotic vibe.

Arcs
Natural movements follow arcs, not straight lines. When a ball bounces over a table the position by time will form a graph that is shaped like an arc. The acceleration, however, will look like a zigzag graph. The animator’s task is not to model the acceleration graph but the positional graph. Arcs are the result.

Secondary Action
The word animation derives from the Latin word animo meaning "give life to". Living beings do small motions, which are seemingly unimportant for the animation process, such as scratching or breathing, but create an illusion of depth. The believability of an actor is increased by secondary actions.

Timing
Timing is key in animating because it has to match the motions humans are accustomed to if the animation should appear natural. However, there is a certain range of tolerance in timing that portrays the same motion in a different way. For example, the speed at which a hand is waved can express the sadness of the actor when saying goodbye to someone but it could also express happiness. This allows the reusage of the same motion by applying different time factors to create motions that cause different feelings in the viewer.

Exaggeration
While not as easily applied to the real world, exaggerating animations causes the actor to feel livelier, engaged and interesting. Extending the motion range of an actor makes him more noticeable, thus gets more attention.

Solid Drawing
While symmetry is a sign of beauty, asymmetry has its place in design too since it causes a more realistic animation. Humans usually don't stand stiff with both legs perfectly symmetrical. The weight distribution deviates, which causes movements towards one side. This may lead to one leg being further extended as another. In general, slight asymmetries can appear beautiful.

Appeal
The way an actor is animated changes how he is perceived. If an actor should be likeable, then the motions should be fluent and beautiful. A stiff animated actor with awkward movements will more likely be perceived as negative and boring. To keep the appeal of the viewer, the animations should be understandable so that the viewer is able to build a connection with the actor.

While most of these guidelines are certainly helpful in the design process of robot animation, certain limitations apply to robot animation in comparison to character or cartoon animation. Since robots are objects existing in the real world, physical laws apply to them. Consequently, movements might be limited in speed or precision. Animations have to be designed with caution, to not damage the robot. Movements may cause unwanted noise from motors, gears etc. "All these constraints specific to robotics can modify the expressiveness of a movement if they are not accounted for (Balit et al., 2016, p. 1). The design process of robot animation therefore comes with more challenges than regular animation design (e.g. design for game or cartoon characters).

3.4.2 Expressiveness of Emotions by Individual Body Parts

Research reveals that certain parts of our body are better fit to express certain emotions. In a study by App, McIntosh, Reed, & Hertenstein (2011, p. 605) "participants favored the body for embarrassment, guilt, pride, and shame; the face for anger, disgust, fear, happiness, and sadness; and touch for love and sympathy". The two emotions used in the behavioral model of Spillner (2018) happiness and sadness, therefore are best expressed by the face. Excitement and disappointment are very close to happiness and sadness and likely best to be expressed by the face as well. For the robot Daryl, Embgen et al. (2012) mention that the movements for sadness and disappointment are similar and therefore easy to misinterpret. Thus, it appears that the desired emotional expression for a robot like Doggy might be limited due to the lack of facial expression. However, Spillner (2018, p. 46) mentions that "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." The expression of emotion seems to be possible, yet not optimal with the limitations regarding Doggy's facial expression.

4. Proposed Design

The following part consists of the proposed design for Doggy. The reviewed scientific work and the background of Doggy will be considered for the design of sounds, animations and a state machine.

4.1 Proposed Design of Sounds

The research presented indicates a positive impact regarding the perceived social ability of sound emitting robots. Sound can add to the emotional expressiveness of a robot, which increases the accurateness of state interpretation. Therefore, users are able to interact more naturally with the robot, enhancing the ease of use and enjoyment. Sound also has the potential to build upon the theory of mind of people interacting with the robot. Doggy is designed to be a ball-playing robot and therefore people might have the theory of mind that this robot, just like other dogs, wants to catch balls. Research has shown that people are satisfied when reality turns out to match their theory of mind. Hence, when Doggy catches the ball, a sound of excitement can allow people to feel this satisfaction, increasing the enjoyment experienced in the interaction. Similarly, an expression of sadness, when no interaction is taking place, can lead to people building a theory of mind. They might start to think the robot dog is sad, because no interaction is taking place and, as humans are empathic creatures, will likely try to cheer Doggy up by playing the ball game. The creation and fulfillment of the theory of mind is a central element in designing sound (and also gestures) for Doggy.

Another aspect that comes with the addition of sound is that Doggy, as a robot designed for large exhibitions, will gain more attention and stand out more as an attraction. A silent robot is more easily missed than an attention grabbing one. Therefore, more people will interact with the robot. Attracting new people to the robot is important to keep the attraction going. The gameplay is Doggy's main aspect and without people to interact with it, there can be extended times of no interaction. The downtime of gameplay can therefore extend itself. Sound can help to reduce the downtime and therefore prevent the spiral of gameplay downtime, to keep Doggy interesting as an exhibition robot.

The aforementioned aspects related to sound lead to the conclusion that the addition of sound is desirable for Doggy.

As seen in the related work, robots emit all kinds of sounds. Thus, it is necessary to decide which type of sound to use for Doggy. As Doggy is an anthropomorphized, cartoony dog, a natural source of sound should be used. Robotic sounds would therefore misfit the appearance of Doggy. In several cartoon movies, such as *the Lady and Tramp* or *the 101 Dalmatians*, dogs use voices of human actors. In *Mickey Mouse*, the anthropomorphic dog Goofy is able to speak, while Mickey Mouse's dog Pluto sounds like a real dog. The film industry demonstrated that both human voices and dog sounds are feasible for cartoon dogs. Doggy is a robot dog and not a cartoon character, but certainly, both options are possible for this being, too.

A human voice could allow the creation of a character with more depth. It also allows the expression of emotions in a verbal way. Additionally, voice could be used to target visitors more directly. For example, Doggy's state "encourage" could be a lot clearer with a vocal call to action. Doggy could also be more entertaining, by adding funny or witty utterances such as jokes. In general, the addition of a voice opens up plentiful design possibilities.

Using dog sounds could appear more natural and since Doggy is not only a cartoonish dog, but also a robot, the addition of a human part could be confusing and dubious. Furthermore, dog sounds are more in line with the unity assumption. Even though Doggy is an anthropomorphized dog, its appearance is more of a dog than that of a human. According to the unity assumption, crossmodal influences are more impactful with matching audiovisual stimuli, such as Doggy's appearance combined with dog sounds. Therefore, for Doggy, a gesture that on its own would be difficult to interpret by users could be understood more intuitively with dog sounds and vice versa. This is not just because both stimuli express the same emotion simultaneously, but also because of crossmodal influences, which aid the expressiveness of the conjunction of stimuli.

It appears that both choices are possible additions for Doggy. For now, both types will be considered in order to find the best solution for Doggy.

The previously mentioned robot RAMCIP uses a synthesized voice to utter, while My Real Baby uses recorded bits of a real infant making sounds. For Doggy, both options have to be considered, for both the human voice and the dog sounds.

A synthetic voice (synthetic voice will henceforth also include synthetic dog sounds for simplicity) is unlimited in possible utterances. For every possible emotional state, a matching sound can be synthesized by adjusting the volume, rate and pitch, as well as the prosody for spoken language. This allows the robot to be less repetitive and possibly to increase the expressiveness and natural feel. By creating a synthetic voice, the design would match the biological design more than the functional design. This is because it allows the robot to interact with its environment in possibly infinite ways through sound, just like a real dog. A sophisticated synthetic voice could fully adapt to any emotional state, however the interaction period of Doggy is only designed to be short-term. A synthetic voice might be too complex for the limited time individual users will interact with the robot. Therefore, the functional design seems to be the right approach for Doggy. Fong et al. (2002) state, that for short interaction periods prerecorded sounds might be sufficient for social interactions.

To achieve a conjunction of sound and gestures, the sounds have to be added into the state machine. Then sounds can be played at the same time as gestures start. Starting both stimuli at the exact time is desirable, as the study of Vatakis & Spence (2007) showed that a mismatch of around ±200ms already causes more than 80% of participants to correctly identify which sound was played first. This could likely decrease the benefits of crossmodal influences that base on the unity assumption.

A robot that is repetitive can quickly appear dull to its viewers, therefore plenty of sounds are required. Furthermore, the addition of sounds attempts to increase the sociability of Doggy. However, if the emitted sounds seem repetitive, the viewer could get the impression of an unrealistic and unbelievable robot instead. The goal should be to create an amount of sounds that at the minimum do not get repetitive in the length of an average user experience. Additionally, during the creation of sound, an emphasis has to be made on states that are more likely to occur than others.

While a completely synthesized voice would be too sophisticated for the use time of doggy, adapting the recorded audio files could prove to be useful in order to add more expressiveness to

Emotion	Speech intonation parameters
NEUTRAL	Volume: 8/10, Rate: +1, Pitch: +0
EXCITED	Volume: 10/10, Rate: +1, Pitch: +4
SAD	Volume: 8/10, Rate: -2, Pitch: -3
SLEEPING	N/A
TIRED / CONFUSED	same as neutral
FOCUSED	Volume: 8/10, Rate: +0, Pitch: +0

Figure 4-1: Speech intonation parameter by Antona et al. (2019).

the utterances. However, this depends on each sound and might not be needed at all. As seen in the research conducted by Antona et al. (2019) regarding RAMCIP, emotions can be targeted specifically, by adjusting the volume, rate and pitch, as seen in Figure 4-1.

The thematization of volume raises another important issue. The typical use scenario of Doggy is an exhibition or a firm party, where a loud environment is likely. To keep the sound of Doggy relevant, while not being disturbing, a fitting volume has to be used.

4.2 Proposed Design of Animations

For her thesis “Interacting With a Ball-Playing Entertainment Robot”, Spillner (2018) laid the foundation for the gestures of Doggy. Twelve full body animations were designed for the five states happy, excited, sad, disappointed and encourage. Since these animations are well designed, it appears to be correct to transfer them onto Doggy. However, as research shows, certain things need to be considered when transferring animations onto a robot. Furthermore, research has shown that vast amounts of animations reduce repetition. Additional animations increase depth and believability. Hence, the existing twelve gestures designed by Spillner (2018) will be transferred onto Doggy but also extended by a rich array of new gestures.

Just as sounds attract people and keep their attention, so do gestures. Therefore, it is important to keep Doggy moving. As of now, Doggy only moves to express emotions and to play the ball. However, a social being does not change nor does it express emotions all the time. A state in which Doggy does not express emotions, but keeps on moving, needs to be added as a transition for emotional states. The movements should only illustrate that Doggy moves, to create an illusion of life, since these movements can be classified *as secondary action*. This adds depth and believability to Doggy. It also demonstrates that Doggy is a robot that is able to move. Otherwise, viewers might believe that Doggy is just an oversized doll of a dog.

The transfer of gestures onto Doggy and the creation of new animations have to be executed with great care in order to create the illusion of life. This is sought after, since it is the foundation of a natural and socially acting robot. In the following parts, the remaining eleven animation laws will be analyzed regarding their relevance for Doggy’s animations, to work towards the illusion of life.

The main difference between designing animations for an existing robot and a virtual character is that physics apply for the robot. Thus, some design laws are granted by default, as humans are accustomed to movements that abide physics. Affected by this are the three animation laws *squash and stretch*, *follow-through and overlapping action* and *slow in and out.* These physical laws are given; however, the design of animations can be purposefully crafted to complement them. For Doggy, this is especially relevant for the laws *squash and stretch* and *follow-through and overlapping action* as non-rigid parts are attached. The laws affect doggy’s ears, tongue and tail. Therefore, these parts are perfectly animated by nature itself. This effect can be abused by adding animations that quickly ac- and decelerate Doggy, such as an animation that rapidly changes movement directions.

Robots are often associated with stiff, unnatural and heavily optimized movements. However, the exact opposite is desired for Doggy. Hence, design should prevent the association with a factory robot. To achieve this, the movements of Doggy have to be in line with the law *Slow in and out*. In the implementation, a focus has to be placed on adhering to this law.

In contrast, the fulfilment of the animation law *exaggeration* is limited due to the existence of physics. This is problematic as Doggy, with its cartoony design, could be expected to behave similarly to actual cartoon characters. However, the significant differences in appearance presumably reduce expectations viewers may have. Regarding Doggy, Spillner (2018, p. 47) evaluated that exaggerated gestures are “more easily understood and fit better with the character”. Exaggerations create a vivid, involved and more exciting robot. Exaggerated movements are therefore of importance. The exaggerated movements should still allow Doggy to appear approachable, not terrifying.

The anticipation of movement by viewers can help with the creation of a theory of mind. Furthermore, Doggy appears more natural, when behavior can be predicted based on anticipatory movements. For example, slowing down Doggy's movements, as he begins to see a ball in a person's hand, could demonstrate Doggy's interest for a ball in a believable way.

Staging is already given through the addition of sound into Doggy's robots. However, the gestures, too, serve as staging for Doggy, as they accentuate state changes and the corresponding emotions. Additionally, emotional expressions underscore the ball game after a ball was played.

The adaptation of time (*timing)* allows for a varying emotional expression for an animation, therefore resulting in a quicker implementation process.

The animation technique *pose-to-pose* is chosen over *straight ahead*, due to simplicity. By using the pose-to-pose technique, animations can be designed to start and end at the null position of Doggy. This allows for natural and fluent transitions between animations, since all animations start and end at the same position. Therefore, an interpolation between gestures is not needed.

Regarding *solid drawing* for robots, Ribeiro & Paiva (2012, p. 3) state: "the main concept to get from this principle is asymmetry." While symmetry is usually desirable, perfect symmetry is rare in nature. Hence, a social robot should allow asymmetry in its motions, to appear more natural. Therefore, the design technique should include some movement deviations in otherwise symmetrical movements. Another deviation, from an otherwise perfectly linear but unnatural movement, is the *arc.* Slight *arcs* should be incorporated into animations.

Lastly, the animation law *appeal* is about adding beauty to the character so that viewers are interested. The design should add charisma and fit the character. This law, however, is arguably the most subjective, as the aspects mentioned might appeal to one person while another dislikes it. For Doggy, it is reasonable to build upon the cartoony design of the costume, as explained, by using exaggerations in gestures. Generally, fluent and energetic movements should be preferred in order to give the impression of an active and playful robot that is willing to interact with viewers.

4.3 Proposed Design for the State Machine

The proposed design of animations is based on Spillner (2018). This suggests to base the proposed design of the state machine as well on the design of Spillner (2018). It can be seen in Figure 4-2. The state machine was designed not to interfere with the gameplay; therefore, any expression is skipped once a ball is detected. It is also based on recognizing sounds. However, since the recognition of sounds is not within the scope of this thesis, the states encourage and excited would not be reachable with the current design. Additionally, the proposed idle state changes need to be incorporated into the state machine.

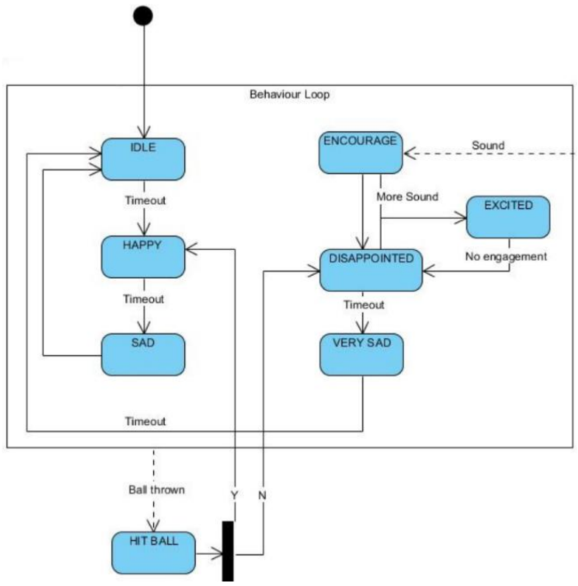

Figure 4-2: The state machine created by Spillner (2018).

One big concern with the state machine is that Doggy enters the states disappointed and very sad too quickly. Firstly, when a sound was heard but no interaction followed. Secondly, when the ball was not hit. In addition, the state change from happy to sad feels too abrupt. There needs to be a better transition between the two states. With the general state design, Doggy might appear overly melancholic for an entertainment robot that is supposed to be cheerful and happy about any interaction. With these issues in mind, a new state machine derived, that allows Doggy to appear happier and incorporates transitions between states that feel more natural (cf. Figure 4-3).

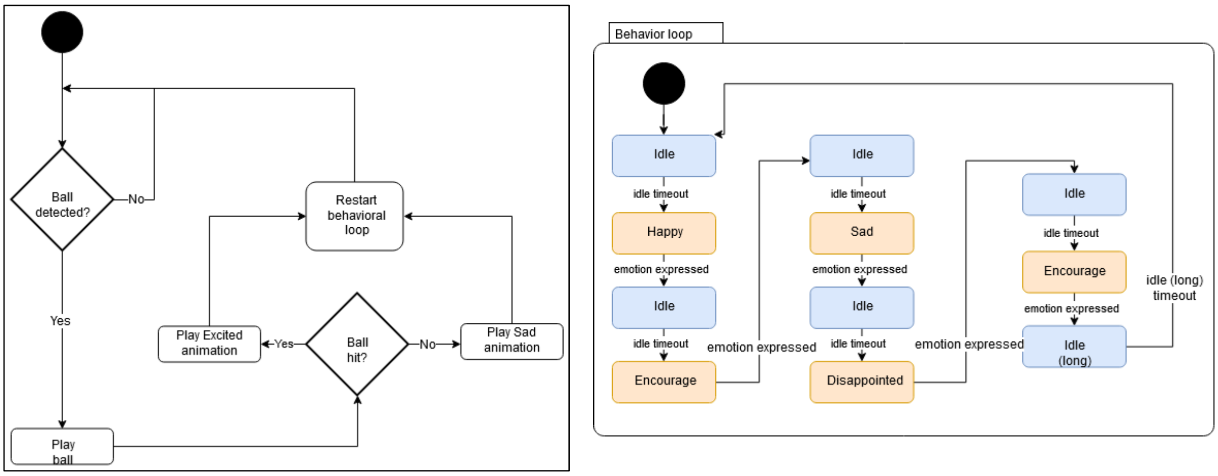

Figure 4-3: New state machine with gameplay routine and behavior loop

One thing to point out is the change of emotional expression after a ball was detected, but not hit. The sad state has been chosen over the disappointed state. This was done because expressing disappointment could leave the impression that Doggy complains about how the ball was thrown. It is better to express that the robot sees the failure in itself, not to blame anyone. Additionally, the interaction caused by a played ball results in the behavioral loop to restart. The behavioral loop is structured to express positive emotions in the beginning. Consequently, the ball playing game causes Doggy to be happy in either outcome. This behavior is more fitting to the behavior of a dog. Viewers that played catch with dogs before, might apply their theory of mind onto Doggy. Then, this state machine better reinforces their theory of mind. Dogs like to interact with balls; Doggy is designed to do the same. As it is Doggy's core concept, the new state machine also purposefully prioritizes the ball game over any expression.

The state change from happy to encourage to sad feels more authentic than a direct transition from happy to sad, as it was the case in the state machine before. Now the robot expresses happiness, then encourages viewers for more interaction and then is sad after no interaction followed. This state design is more in line with the five step model of emotional processing of Ribeiro & Paiva (2012), as it is a more authentic behavioral order in regards to the question "how does this affect what I was already feeling?" (Ribeiro & Paiva, 2012, p. 4).

5. Completion of the System

5.1 Current State of the System

The 2.1 m tall robot is powered by a Linux machine, which is located in the 77.89 cm tall lower part of the construction (cf. Figure 5-1). The machine uses the operating system Ubuntu in version 16.04. The Intel CORE I7-2600k CPU allows for sufficient calculation power. At the top of the lower part are the ball-tracking cameras "Guppy" by AVT. Doggy has three motors located above the lower part. These motors allow for three DOF, however with redundancy. ACS1 is responsible for z-axis movement, ACS2 for y-axis movement and ACS3 for x-axis movement. The end-effector is a styrofoam ball located at the very top of the robot and Doggy's head (H). At the core of Doggy is a microcontroller (µC), designed by Schuethe & Frese (2014). Its task is to control the motors. For the ball playing game, it can calculate a trajectory for the end effector with a smart usage of the DC motors' torques and joint angles in order to reach a given position in an acceptable amount of time. The microcontroller sends data via USB and Ethernet, and is controllable via the open source *robot operating system* (ROS). Currently, the setup uses version "Lunar Loggerhead" (published May 23, 2017) of ROS. ROS is used for a myriad of robots all over the world to allow quick and easy communication between modules of a robot. Often times data from various robot parts are needed to execute a task, such as the joint angles for kinematics or the camera data for the detection of a ball for a ball playing game. By using a publish and subscribe model, nodes in ROS can stay independent. Information is sent without the need to know which node might receive it. The subscription to a node is similarly modular. With the roscore a master node is created that is managing the registrations from publishing nodes and the subscriptions from listening nodes. Reusability of nodes is granted, since the amount of subscriptions is not limited, which is one of ROS' biggest advantages. While the node communication is certainly one of the key features of ROS, a huge amount of tools is also available. As an example, data can be recorded and plotted, GUIs can be developed with ease and many handy libraries are in place (Open Source Robotics Foundation, 2019).

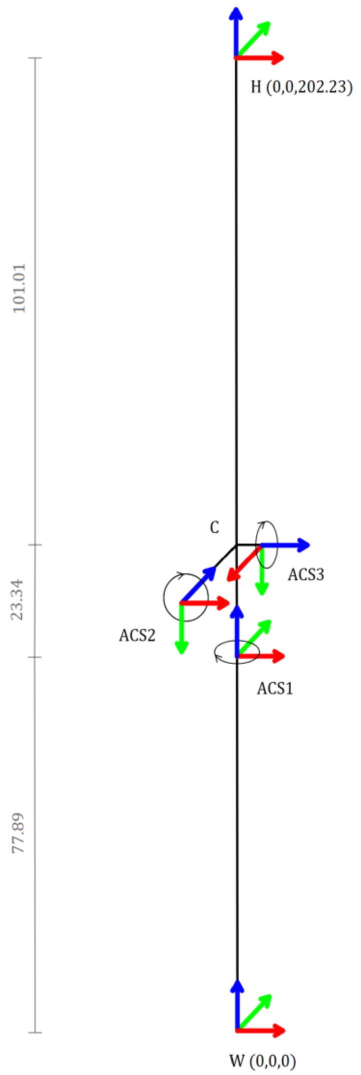

Figure 5-1: Illustration of Doggy's internal make-up by Spillner (2018).

The current setup in ROS is programmed in C++, although Python and Lisp are also available for developing with ROS. Ubuntu is ROS' designated OS as other systems, like macOS or Windows, are only partially supported.

The main part of the gameplay routine is located in the ROS package BatControlPiggy1. For the gameplay, this node is subscribed to "*AutoBallTracks*", which sends information about detected balls in the camera view. Once a ball track is considered to be a moving ball, possible bat trajectories are calculated to find an intersection between the two entities. Then only a possible and probable intersection trajectory is carried out. However, for safety reasons, the calculations are limited by joint boundaries. The three axes are limited by the two variables `jointLimitsLow` and `jointLimitsHigh`. The implementation includes a behavioral model with three states: `INIT`, `READY` and `PLAYING`.

5.2 Implementation

5.2.1 Implementing Sounds

5.2.1.1 Expanding the System to Emit Sound

The addition of sounds presupposes the integration of a speaker system onto Doggy. Therefore, capable components need to be found and added in order to deliver a sufficiently high volume. Since Doggy's PC is not able to produce a high enough volume, a solution that does not use the mainboard's line out audio port is needed. An externally powered speaker, such as an active speaker, was considered; however, since the energy supplying cables were wrapped shortly before, this option was not optimal. Finally, the thought of using a 12V final stage, supplied by the power output of the PC, arose. The final stage Basetech AP 2100 is able to provide enough power for two 50W speakers with an impedance of 4Ω. Although possible, one speaker was chosen over two, as a potential stereo effect would be minimal, since the speakers cannot be attached with a sufficient distance between each other. One Caliber Audio Technology CSB3B speaker was attached centrally between the two cameras (cf. Figure 5-2). Before mounting, the setup was tested and a strong hissing noise was emitted by the speaker. After some tests, a solution for this problem was found by using a ground-loop-isolator, two capacitors and a coil (cf. Figure 5-3).

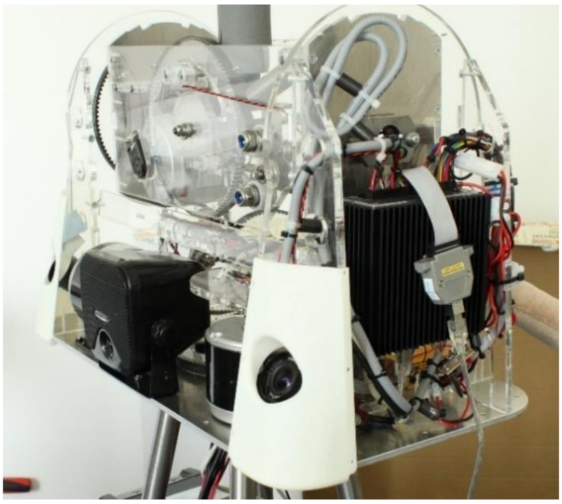

Figure 5-2: The internals of Doggy with the added Caliber Audio Technology CSB3B speaker.

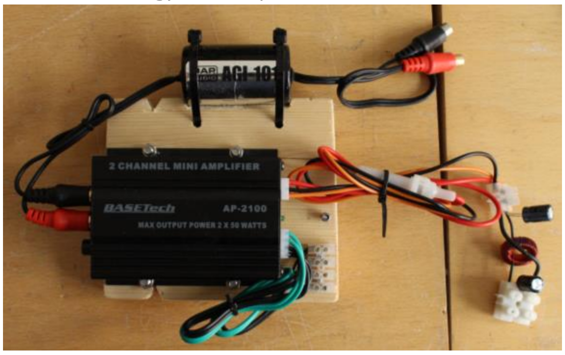

Figure 5-3: The assembly of the Basetech AP2100 with the ground-loop-isolator, two capacitors and coil.

The most fitting place for the final stage was considered to be inside of the PC's case. A medium was needed to fixate it inside. The four screw holes of the end stage were used to fixate it onto a piece of wood. Additionally, the ground-loop-isolator was attached at the top. Finally, the wooden piece was attached to the PC's case with cable binders and screws (cf. Figure 5-4).

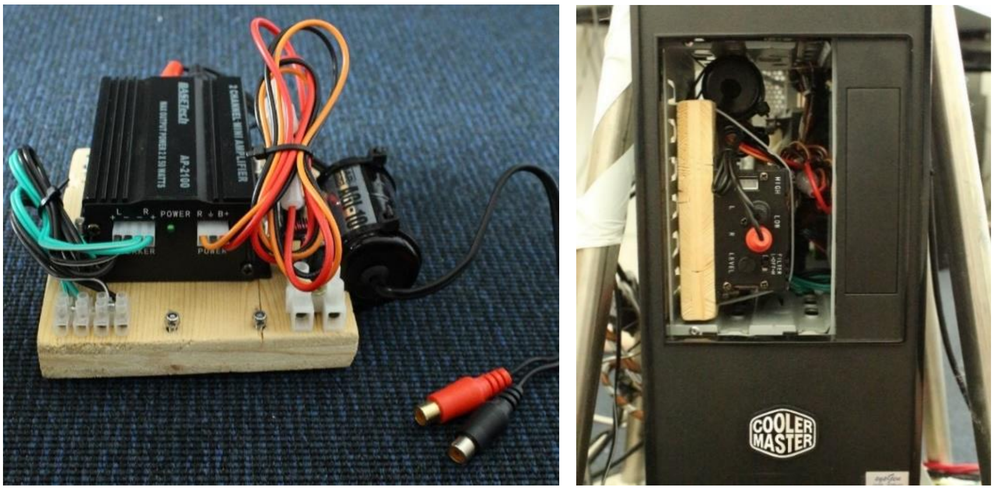

Figure 5-4: The final assembly of the construction.

5.2.1.2 Creating Sounds

As discussed in the proposed design of sound, both human voice and dog sounds were considered for Doggy to see which fit better. It was decided to start with the design of human sounds. A Zoom H5 was used to record voice in a sound absorbing environment. A young, tall, male speaker with a deep voice accepted to lend Doggy his voice. The voice actor was requested to speak with the cartoony character of Doggy in mind. The speaker was told to utter with a prosody that reflects activeness, excitement and enjoyment for the states happy, excited and encourage. The opposite was the case for the states sad and disappointed. The idle state merely contained panting without any particular emphasis regarding prosody. Additionally, the sentences used were designed to be funny and sometimes, in case of the sad expressions, even pitiful. The sounds were cut, so that the relevant part starts immediately (cf. Figure 5-5). In addition, the audio was normalized to create equal sound levels. The adaptation of prosody in post-production was attempted. However, as the speaker already included his personal prosody regarding the desired emotional effects, this was not needed and sounded unnatural.

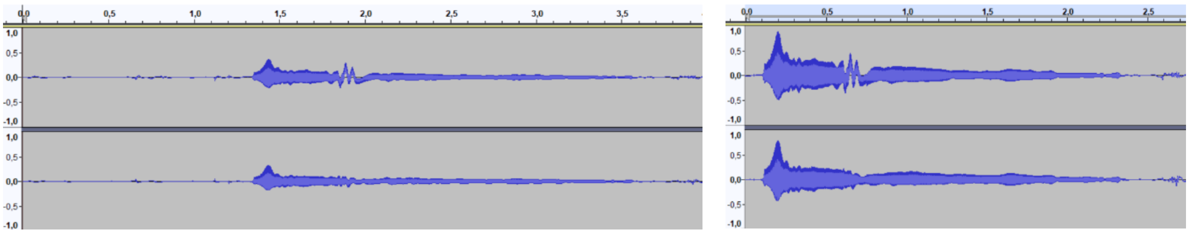

Figure 5-5: An example of sound editing. The original is on the left and the edited version on the right.

In total 55 sounds were created: 6 idle, 8 happy, 17 excited, 9 sad, 6 disappointed and 9 encourage sounds. Contrary to the proposed priority for idle sounds, only 6 were created as the diversity of panting sounds was difficult to achieve, therefore a high amount of idle sounds was not needed. For the excited state, however, by far the most sounds were created. This is because this state is entered when a ball was successfully hit; it is an event to emphasize on. Having a great diversity for this state reduces repetitiveness for people that play the ball game for a larger amount of time.

After finishing the creation of the human sounds, it was time to analyze whether to accept them for Doggy or to design new sounds based on a dog. One big concern was the clarity of the utterances. Firstly, because Doggy is intended for an environment with heavily fluctuating sound levels. Secondly, because the recorded sounds were partly spoken with a mumbled voice. Another concern was the adequateness of the speaker’s voice for the appearance of Doggy. The voice did not match Doggy perfectly. A deeper voice was needed. For the reasons given, it was decided to use sounds of a dog instead.

The dog sounds were created by using open source sounds from the internet. However, the limited amount of sounds resulted in an issue. To create a sufficient amount of sounds, various dogs had to be considered. Nevertheless, since the sounds are played spread out in time and due to the interaction time of participants being relatively low, the usage of multiple dogs as a sound source was accepted. However, during the search for sounds, now a focus had to be placed onto finding similar sounding dogs. This further limited the search for sounds. In the end, 24 sounds were created. As explained in the proposed design of sound, the focus was placed on creating sounds for the states idle, sad and excited. In total 11 idle, 7 excited, 6 sad and 2 encourage sounds were created. Due to the difficulty in finding appropriate sounds, 2 excited sounds and 2 sad sounds were reused for the states happy and disappointed.

Similarly to the human sounds, a panting noise was considered to be fitting for the idle state. This sound is very typical for dogs, hence adds to the authenticity of Doggy. Additionally, barking sounds were added for this state. To increase the amount of variation, new idle sounds were created with

Audacity. This was done by moving and copying parts of an original sound file, to create a new sequence of dog barks or panting sounds. When creating the sounds, pauses were added to add tranquility to the idle state, which is supposed to be restrained in movements as well.

For the sad state, authentic and fitting sounds could be found. Therefore, it was not needed to adapt the intonation. However, there was a lack of sounds for the excited state. As it was desired to have an above average amount of sounds for this state, more sounds had to be created. This was done by using barking and panting sounds from the idle state. The intonation was adapted by raising the volume, speed and pitch of certain idle sounds, as proposed for the design of sounds. An example for this can be found in the audio_edit_example folder on the CD.

5.2.2 Implementing Animations

5.2.2.1 Adding Doggy's Tail

Doggy has three DOF. To add more diversity to the robot's animation and to add a more natural appearance, the damaged tail of Doggy was repaired and reattached (cf. Figure 5-6, Figure 5-7). The tail is a non-rigid construct made out of polyethylene, similar to a swim noodle. It has a fabric coating matching the rest of the costume. The tail is controlled by two servomotors that allow for x-axis and y-axis movements. These motors can be controlled with the microcontroller of Doggy.

After initial tests, it was apparent that the tail was hanging too loose, which limited design possibilities. It also gave a false impression of sadness, since a hanging tail could be interpreted in that way. To maximize design possibilities, a slight upward adjustment had to be made to the tail's zero position. To achieve this, a glass fiber pole was inserted into the fabric coating. This added more rigidness, but still allows the tail to abide to the laws *squash and stretch* and *follow-through and overlapping action.*

As it turned out, using the servomotors of the tail causes a high-pitched sound with every movement. This had to be taken into consideration for the design of the animations, as it could negatively influence the experience of viewers. The tail was designed to accentuate animations. However, adding tail movements with loud, high-pitched sounds to the calm idle state caused an expressive mismatch. The hectic sounds of the tail would attract too much unwanted attention; therefore, it was decided not to use tail animations for this state. This is unfortunate, since small tail animations could accentuate the liveliness of Doggy.

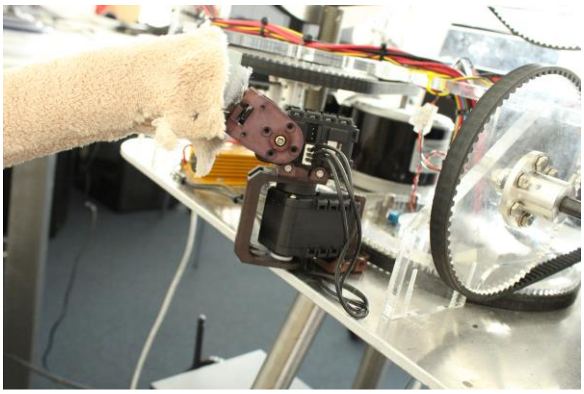

Figure 5-6: The two servomotors and the tail attached to the back of Doggy's internals.

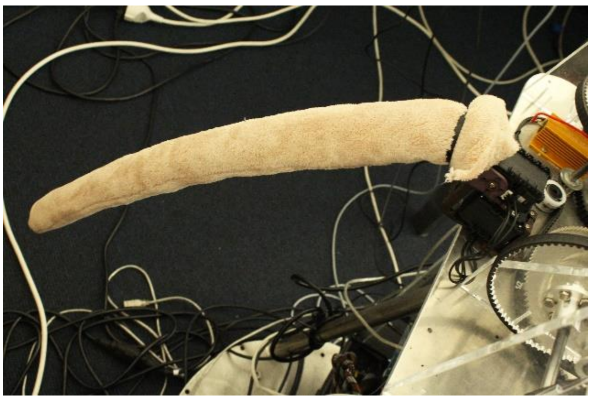

Figure 5-7: The attached tail. A cable binder was added after the coating fabric started to move outwards because of the motions.

5.2.2.2 Design Tool: ROS Package "gestures"

For the creation of animations, a tool was necessary that is capable of creating natural appealing movements. The two options considered were keyframing and to design by steering Doggy with a gamepad controller.

Keyframing is typical for animating. The process involves the creation of *keys* for certain *frames* on a timeline. The keys are spaced out on the timeline and can represent data such as position, angles, velocity etc. Then the keys can be interpolated with methods such as a linear, Bézier or a cubic interpolation. Keyframing is highly adaptive and can be used to create realistic and natural looking animations.

While usually used for games to steer fictional characters, a gamepad controller could do the same with Doggy. By mapping the stick axes to the joint angles, the two gamepad sticks can theoretically steer up to four DOF. A controller was previously used for Doggy, as Spillner (2018, pp. 37-38) states, that a controller "allowed for a much more natural way of 'steering' the robot's body. […] it proved very useful for testing the range of motion, 'pre-viewing' ideas for possibles [*sic*] movements, as well as in the later user tests". The usage of a controller would also be more in line with the design philosophy of Tzeng (2013), as his idea was to create a recording studio in which the best of various takes would be chosen. Tzeng (2013) preferred this type of animating over a keyframe based method, since it was considered to be more intuitive and efficient. Indeed, the addition of a controller to Doggy appears to be a much faster implementation than the creation of a keyframe based system and it allows for quicker design iterations. Furthermore, implementing an approach that uses keyframing might be overly sophisticated for the very limited interaction time of participants with the system. The fact that Doggy only has five DOF also limits expressiveness. This also warrants the use of a less sophisticated tool to model its animations.

Hence, I agree with Tzeng (2013) and Spillner (2018) that a recording studio approach is more intuitive than a keyframing approach and that a controller is a helpful tool to design natural looking animations. In addition, while a keyframing approach might be able to result in more refined animations, this level of perfection might not be required for a robot intended for short-term entertainment.

A ROS package with the name "gestures" was added to allow the creation of animations for Doggy. Hereafter, the usage of the package and the creation process of animations for Doggy are explained.

5.2.2.2.1 Mapping

In order to quickly preview ideas, reiterate gestures and to record and replay them, a free mode, in which Doggy can be steered with a controller, needed to be implemented. An Xbox 360 controller was chosen for steering Doggy. Theoretically, up to four DOF can be mapped to the sticks of the controller. However, with the addition of a tail, Doggy has five DOF. Therefore, a split was made and two steering modes, one for the body and one for the tail, were added (cf. Figure 5-8). This idea is following the approach of Tzeng (2013), who created a "Classroom" for Doggy, in which body and tail animations could be created separately and added to a combined animation.

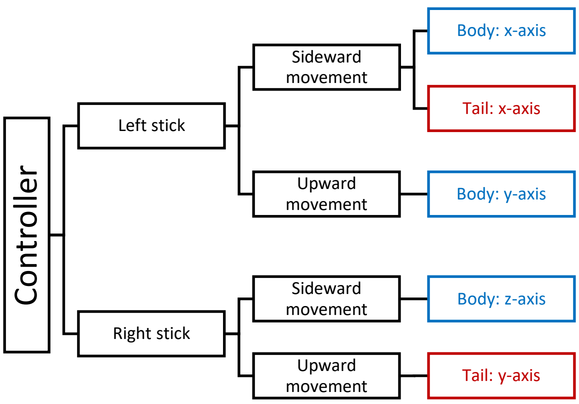

Figure 5-8: Stick mapping.

The stick ranges were mapped to the axes' ranges of Doggy. Moving a stick to its limit also causes Doggy to move to its limit in that axis and in the same direction. The maximum possible movement ranges of Doggy were taken from the ROS package BatControlPiggy1, since they were already defined there. To create animations, an option to record the movements in free mode was added. Additionally, a playback menu was created. It is used to cycle through all created animations and to replay any previously created animation.

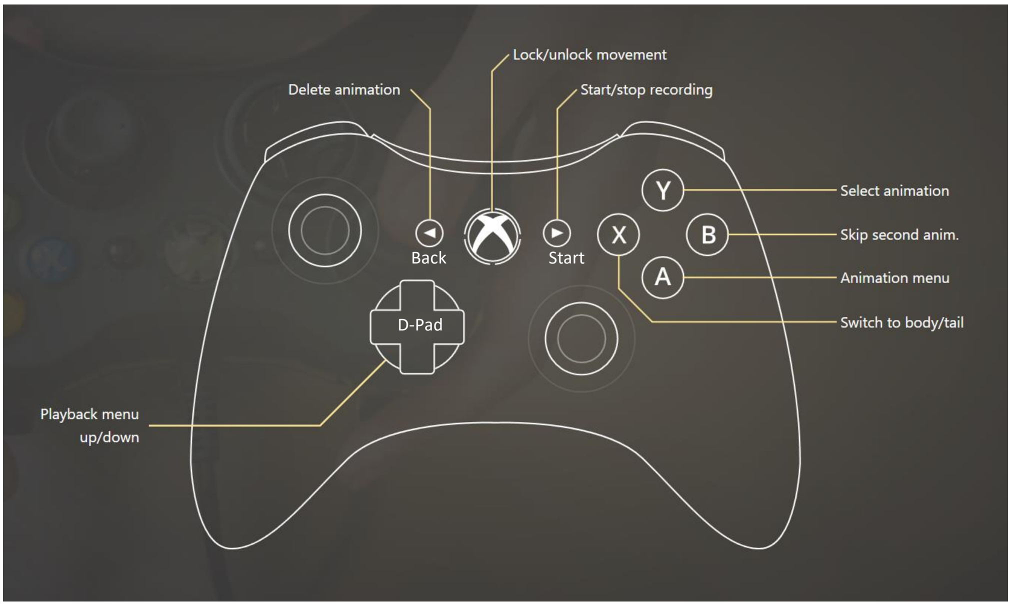

Figure 5-9: The final mapping of the controller (excluding the mapping of the sticks).

In free mode, the Xbox button locks/unlocks all movements from the sticks. The Start button starts and stops the recording of an animation. The X button is used to switch the controls of the sticks to either body or tail movement.

With the A button, the user enters and leaves the animation menu. The menu is used to play animations (cf. Figure 5-10). The D-Pad can be pressed up and down to navigate through the animation menu. Then a press on Y will select the current animation. Now another animation of the other type can be selected, so that both a body and tail animation are chosen. Alternatively, the B button can be pressed to skip the selection of a second animation. Then only the first selected animation will be played. Otherwise, both animations will be played. After navigating to an animation with the D-Pad, the Back button can be used to delete animations from the playback menu. Note that the animation will be deleted immediately without a yes or no prompt.

The final mapping can be seen in Figure 5-9.

```
            [ INFO] [1568033943.40]: Loading Animations
            [ INFO] [1568033943.54]: Finished loading Animations
            [ INFO] [1568033943.54]: To start free movement press the central XBOX X button.
A           [ INFO] [1568033944.00]: Entering animation playback menu
            [ INFO] [1568033944.00]: Animations available: 42
            [ INFO] [1568033944.00]: Select animation to playback with cross key.
            [ INFO] [1568033944.00]: Press Y to start and B to delete the animation.
D-Pad down  [ INFO] [1568033944.00]: Animation: Tail | ID: 1 | Name: crazytailwiggle
D-Pad down  [ INFO] [1568033944.70]: Animation: Tail | ID: 2 | Name: shortwiggle
D-Pad down  [ INFO] [1568033944.89]: Animation: Tail | ID: 3 | Name: tailwiggle
D-Pad down  [ INFO] [1568033945.06]: Animation: Tail | ID: 4 | Name: wagging
D-Pad down  [ INFO] [1568033945.23]: Animation: Tail | ID: 5 | Name: waggingExcited
Y           [ INFO] [1568033945.39]: Select second animation
            [ INFO] [1568033945.39]: Cancel selection of second animation with B, select with Y
D-Pad down  [ INFO] [1568033946.16]: Animation: Tail | ID: 6 | Name: tailDown
D-Pad down  [ INFO] [1568033946.37]: Animation: Body | ID: 7 | Name: happy2
D-Pad down  [ INFO] [1568033946.54]: Animation: Body | ID: 8 | Name: happy3
D-Pad down  [ INFO] [1568033946.72]: Animation: Body | ID: 9 | Name: excited2
Y           [ INFO] [1568033947.25]: Initiating animation.
            [ INFO] [1568033952.32]: Finished playing back animation.
```

Figure 5-10: Example console log for playing an animation. On the left are the buttons pressed, these are not included in the original log.

5.2.2.2.2 Recording and Playback

Once the Start button is pressed in free mode, a recording is started (cf. Figure 5-11). With the second press on the button, the recording is stopped. Then the program prompts for a unique name and a type for the animation. Afterwards, the animation will be attached to the end of the “animations.txt” file. Each animation contains a header with “Animation”, a name, a type and an emotion (cf. Table 5-1). The body of an animation contains timestamps with the positions of the axes (cf. Table 5-2). The stored values are line break separated. For the playback of animations, the body animation always takes priority. As long as the body animation is not finished, the tail animation will loop.

Description	Value
Header Indicator	Animation
Name	disappointed1
Body:0 / Tail:1	0
Behavior type	DISAPPOINTED

Table 5-1: Header of an animation.

Description	Value
Timestamp	0.000000
x-axis	0.07135
y-axis	0.01242
z-axis	0.00581
Timestamp	0.044003
x-axis	0.07132
y-axis	0.01220
z-axis	0.01468

Table 5-2: Body of an animation.

```
[ INFO] [1568035012.38]: Recording started.
[ INFO] [1568035015.08]: Enter name of animation:
circling
[ INFO] [1568035025.38]: Please select a behavior type for this
animation from the following list:
[ INFO] [1568035025.38]: 0 | IDLE.
[ INFO] [1568035025.38]: 1 | HAPPY.
[ INFO] [1568035025.38]: 2 | SAD.
[ INFO] [1568035025.38]: 3 | EXCITED.
[ INFO] [1568035025.38]: 4 | DISAPPOINTED.
[ INFO] [1568035025.38]: 5 | ENCOURAGE.
[ INFO] [1568035025.38]: 6 | UNDEFINED.
0
[ INFO] [1568035036.32]: Recording ended.
```

Figure 5-11: Example console log for recording an animation.

5.2.2.2.3 Creation of Animations

Initial movement tests made clear, that the movement with the sticks resulted in hectic, drastic and robotic movements, which is the opposite of what was required by design. The problem is that Doggy’s movement was designed to minimize the time the bat needs to reach a potential intersection with a ball. Hence, the robot accelerates a lot and overshoots targeted positions. While this is fine for the ball-playing game, the creation of natural, non-robotic animations is made more difficult. To solve this issue, an interpolation was created that is more in line with the animation law *slow in and out*.

Instead of sending the position pointed at by the stick of the controller, the robot needs to be told a position somewhere in-between the desired position and its current position. A function was created, that initially weights the current position more than the goal position. However, as the sticks’ positions continue to point at the same goal position, the weighting of the goal position starts to increase. This allowed the creation of more natural movements. Additionally, the transition speed could be changed to vary the general speed of Doggy’s motions. This was particularly useful for creating animations for specific emotional states. By adapting the transition value, for example, calm animations could be created for the idle state and hectic animations for the excited state. Furthermore, by changing the maximum movement ranges of axes, gestures could be purposefully limited, as visible in Figure 5-16.

The creation of animations for the idle states had to be limited to moving only the x-axis and y-axis (cf. Figure 5-16). This was done, because z-axis movements cause issues with the ball-tracking unit of the robot. Additionally, the movements had to be limited in velocity, since shaking also causes the ball-tracking unit to lose accuracy. These restraints, however, were not an issue, as the design of idle animations was supposed to be calm anyway. More axis movement limitations were added, based on the variables `jointLimitsLow` and `jointLimitsHigh`in BatControlPiggy1 (cf. 5 Completion of the System) to prevent damaging the robot.

The animations designed by Spillner (2018) were recorded, partly with slight adjustments (cf. Figure 5-13, Figure 5-12, Figure 5-14). During the recording of the animations a focus was placed to make use of the laws discussed in section 3.4.1 (Twelve Animation principles by Ribeiro & Paiva). For example, Figure 5-15 shows the tail animation waggingExcited. This animation is *exaggerated*, due to the high frequency and amplitude of the graphs. Since the graph is not perfectly symmetrical, this animation abides by the *solid drawing* law. The length and rhythm of the idleLong animation seen in Figure 5-16 abides by the law *secondary action* as it contains slow movements that only slightly move Doggy (note that this animation is around 63 seconds long). Animations from Figure 5-13 to Figure 5-15 have a timestamp frequency of ~2.27/s, while the animation in Figure 5-16 uses a frequency of ~248.82/s. Videos of the animations encourageAll and sad1 were added to the folder video_clips on the CD, among others.

Figure 5-13: Animation encourageAll: Body animation graph, encourageAll sketch by Spillner (2018), tail animation graph.

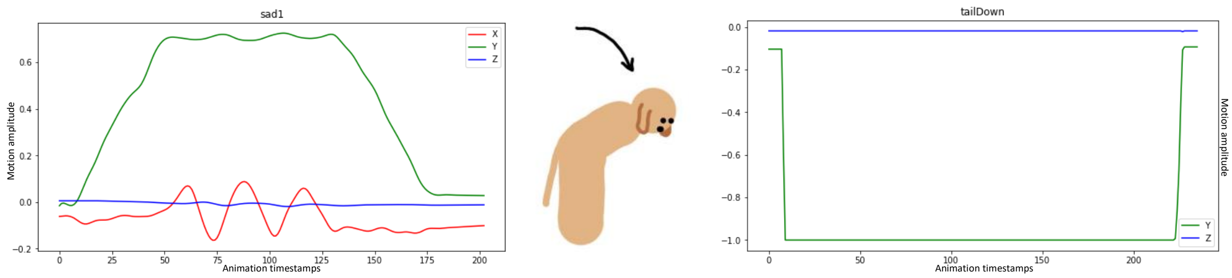

Figure 5-12: Animation sad1: Body animation graph, sad1 sketch by Spillner (2018), tail animation graph. This animation was adapted to include a small headshake. This is caused by the x-axis movements in the center.

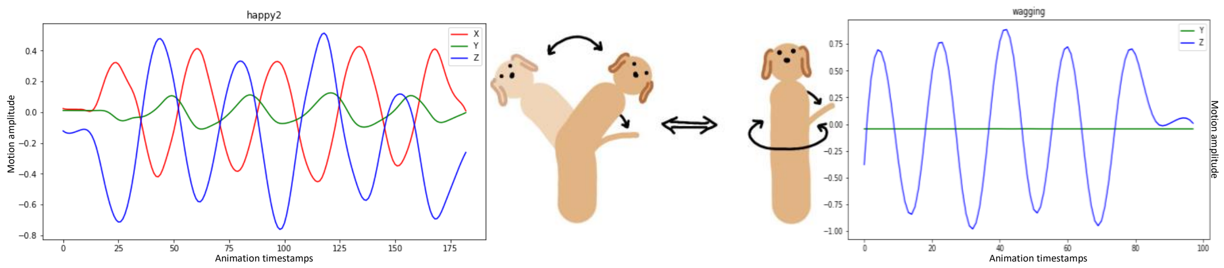

Figure 5-14: Animation happy2: Body animation graph, happy2 sketch by Spillner (2018), tail animation graph.

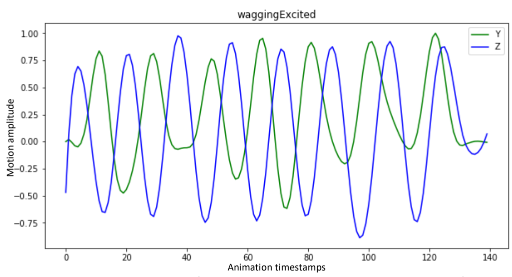

Figure 5-15: Tail animation waggingExcited.

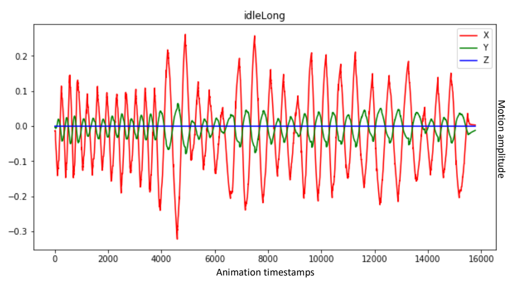

Figure 5-16: Body animation idleLong.

5.2.2.3 Post Processing of Animations

Certain animations were unsmooth around the edges of peaks, which caused faltering movements. Post processing was considered to smoothen out these graphs. A python script was created to apply a Savitzky–Golay filter to the animation graphs in the "animations.txt" file. For each point in the graph, the filter sets a window around the point and fits all points inside this window to a polynomial. Then the x-value of the point is inserted to the polynomial to obtain a new y-value for the point (Schafer, 2011). The filter is able to smoothen edges quite effectively (cf. Figure 5-17). However, for most of the animations this was not needed, as peaks were already smooth. The filter can even cause issues with certain animations. The original tailDown graph is visible in Figure 5-12, the edited in Figure 5-18. Additionally, for best results the filter's polynomial and window parameter need to be tuned for every animation. Furthermore, the filter reduces the amplitude of a graph, therefore reduces the movement range of an animation. This negatively affects the design, as an animation is purposefully crafted. In the end, it was decided not to use the script, since possible advantages are outweighed by possible disadvantages.

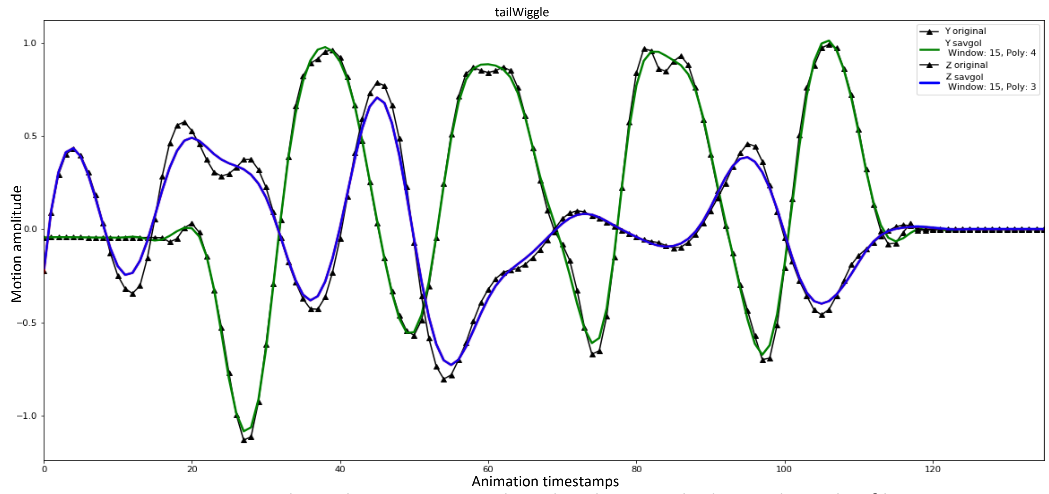

Figure 5-17: tailWiggle animation with and without applied Savitzky-Golay filter.

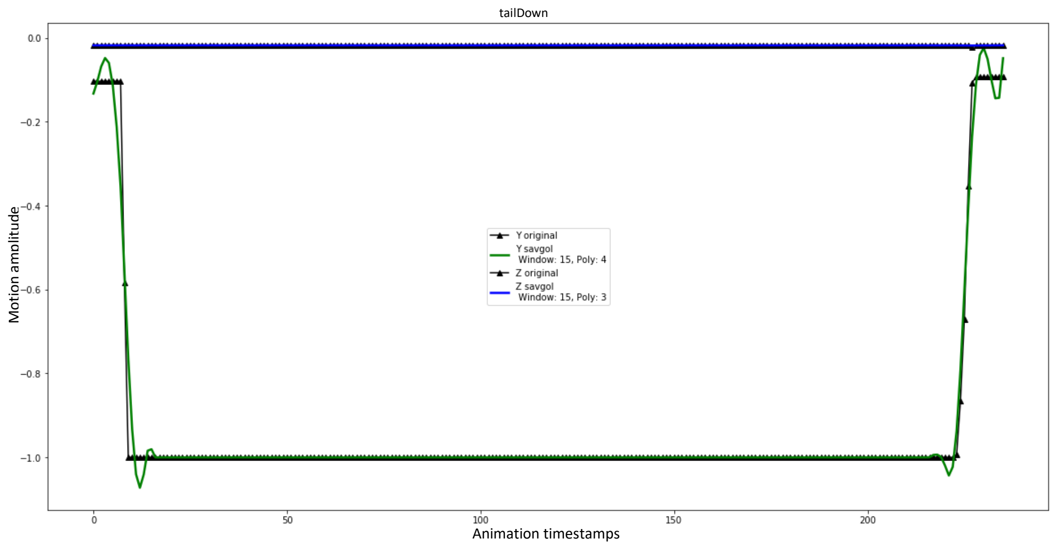

Figure 5-18: Disadvantageous adaptations caused by the Savitzky-Golay filter for the tailDown animation.

5.2.3 Integration of Modules

This part is about the integration of sounds and animations into the gameplay routine of Doggy. The gameplay logic is defined in the package BatControlPiggy1, as well as the state machine. The state machine was updated to match the proposed design. To incorporate animations into the gameplay, an interface needed to be created for the gestures package, which allows external packages to start and stop animations. A ROS subscriber listens to animation requests on the ROS topic “Animations”. Additionally, the available animations are made public for the ROS parameter server. A ROS server was created to stop animation playbacks. It listens to the ROS topic “StopAnimating”. The server sends a response once the request to stop any ongoing animation is completed. This is important for the gameplay, as it needs to be ensured that no ongoing animation will overwrite the position set by the gameplay logic. Before setting the position for the gameplay, the response of the server needs to be received. This causes a slight delay for the gameplay; however, this delay is extremely small and does not have an impact on the accuracy of ball hits.

As explained earlier, the z-axis movements are not accounted for by the ball-tracking unit of Doggy. However, plenty of emotional expressions use the z-axis. Therefore, a compromise was made. It was decided only to track balls during the calm idle state, in which accurate ball tracking is possible. It was decided to use idle state timeouts of 25 seconds (45 seconds for the state idle long) to allow for enough interaction.

Once a new behavior state is entered, a random sound and animation will be picked from the pool and played at the same time, to make use of the unity assumption. For the idle state, this is looped until the timeout is reached, or a thrown ball was detected. An example of the robot reacting to ball hits can be seen in the video ball_hits_clip to the folder video_clips on the CD. Furthermore, the unedited videos throwing_balls and gesture_cycle were added to the folder video_raw on the CD. Each of the unedited videos document ~10 minutes of Doggy’s behavior.

5.2.4 Adding a Hit Tracker

Having Doggy react to the outcome of the ball-playing game is desirable. Hence, the implementation of a hit tracker that accurately recognizes the impact of a ball onto Doggy’s bat was needed. This tracker is only needed while Doggy is trying to play a ball, thus the ball impact has to be detected while Doggy’s bat is moving. The microcontroller publishes state data in ROS. Ideally, the IMU values of the bat could be used for this; however, the microcontroller does not publish this data. As time was limited, an alternative had to be found.

An experiment was carried out to find useful data for the hit tracker. The four animations idleLong, idleSleep, idleCircle and idleCircleOut were played in this order with 10-second breaks in between.

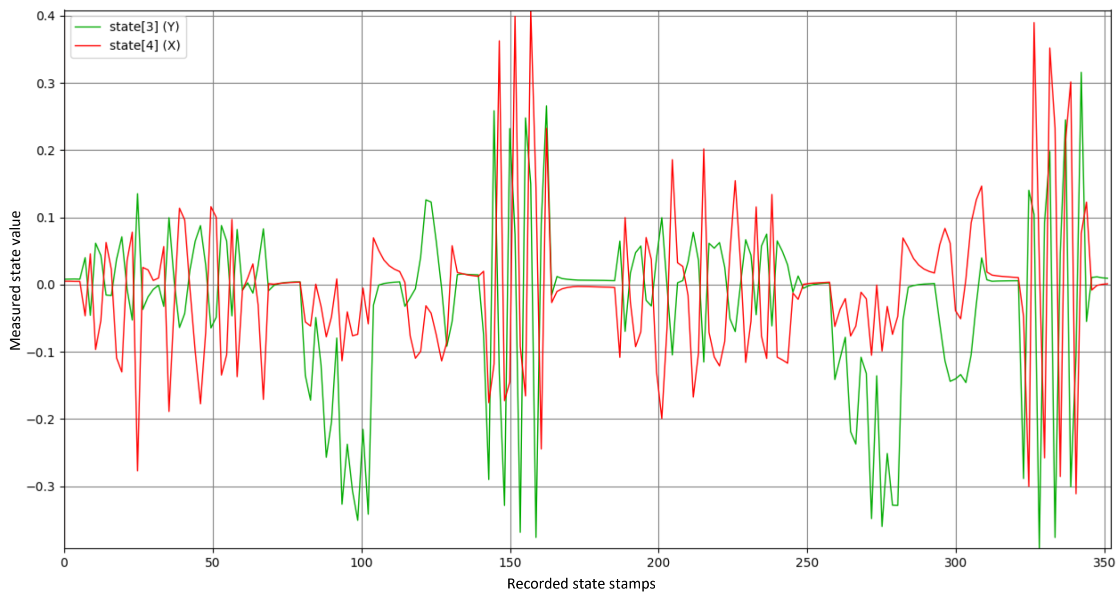

Figure 5-19: The state data of the axes X and Y (the Z-axis is not used for idle animations) for the experiment.

Then, after a 20 second break, the animations were played in the same way again, but this time every 5 seconds a ball was thrown onto the robot (once a ball was missed by Doggy). Figure 5-19 shows the state data of the axes Y and X during the experiment.

The gravity data (cf. Figure 5-20) of Doggy appeared to be the most promising in detecting anomalies regarding ball impacts. The hit tracker compares previous gravity data and considers differences above a threshold as ball hits. This approach is far from perfect and does not work in many cases. However, due to limited time, a more sophisticated approach could not be pursued.

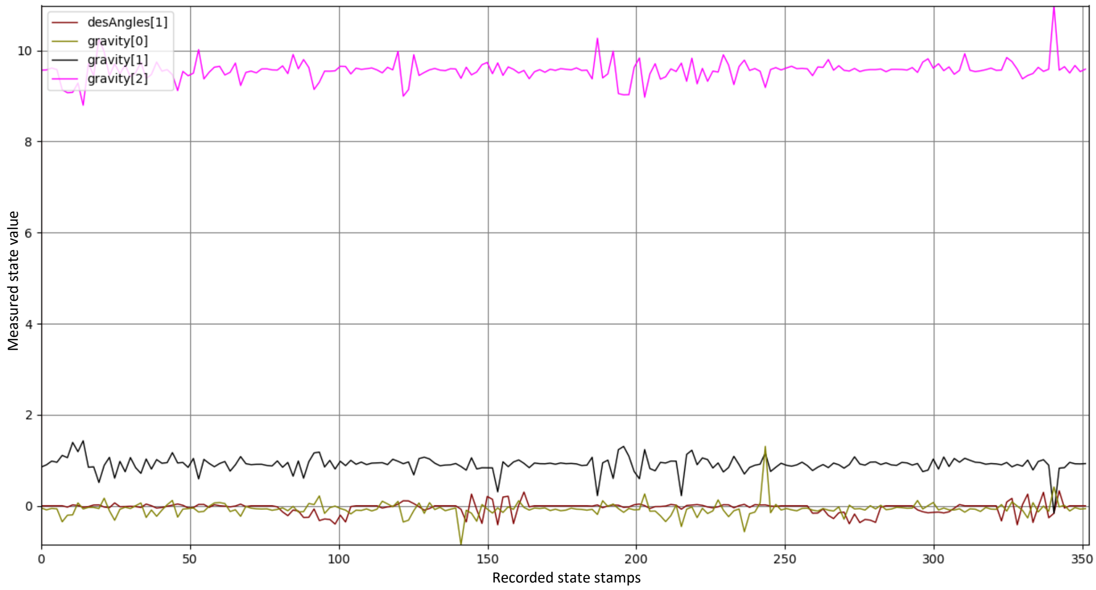

Figure 5-20: The gravity data for the experiment.

Since the results of the hit tracker were often false positives and false negatives, there had to be consequences regarding the design of Doggy's behavior. Having Doggy react with sadness about a played ball might be extremely confusing for viewers. They could, for example, believe that the robot does not want to play anymore, that the ball was thrown with too much force and hurt the dog or that the robot is broken. However, the consequences are not as bad when Doggy reacts with excitement to a missed ball. They might even be positive. People could interpret the excitement for happiness about the fact that the dog gets attention and interaction. Alternatively, they might think the robot wants to signal that it was close to hitting the ball, which already gets him excited. Additionally, people could think that Doggy wants to play more. It appears that a false negative has a big negative impact on how Doggy would be perceived. Therefore, it seems logical to remove the response sad from the reactions for the ball-playing game.

5.2.5 Changes to the State Machine

An updated version of the state machine was created (cf. Figure 5-21). It includes the condition for Doggy to be in the idle state in order to play a ball, since the ball tracking might be heavily manipulated by ongoing animations. If in idle state, a request to the “StopAnimation” topic is sent. Once the animation is stopped, the ball can be played. The playback of the sad animation was removed due to inaccuracies regarding the hit tracker.

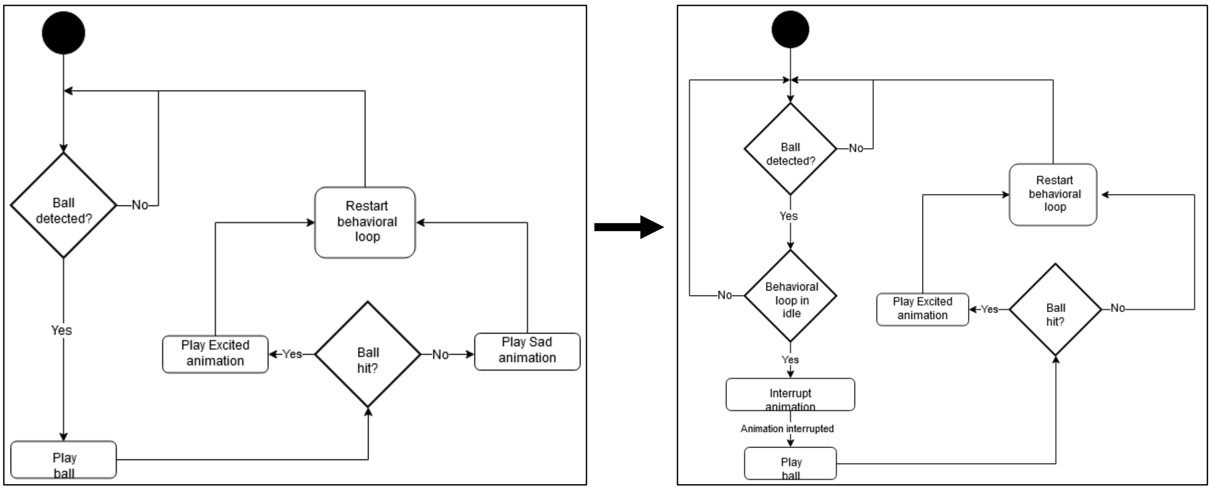

Figure 5-21: The proposed design (left) and the updated version (right) of the state machine.

5.2.6 Changes to Doggy’s Costume

While working with Doggy, some changes were made to the costume to adapt it to new requirements that arose from the usage of animations.

Firstly, the costume has cut outs for the two ball-tracking cameras. However, with the addition of animations, sometimes the costume started to cover the cameras. Then the ball-playing game was impossible, as no tracking data could be used to find balls. To solve this issue, Velcro was added to the costume, surrounding the cut out for the cameras. Then fleece stickers were attached to the surroundings of the camera (cf. Figure 5-22). Now the costume of Doggy can be fixed around the cameras in order to prevent the costume from blocking the camera’s view.

During the various repair tasks, the costume had to be taken off several times. This was a cumbersome process, as the costume is too tight for the current internals of Doggy. Since Doggy’s costume has an internal part, that is tighter than the rest, it became apparent to cut along the inner costume to loosen up the costume. Now the costume can be put on much quicker.

Figure 5-22: Both cameras are surrounded by fleece stickers.

6. Evaluation of the System

A study was conducted to evaluate the effects of the addition of sounds and animations on Doggy's gameplay and how it changes the way in which Doggy is perceived by users.

6.1 Study Planning

To evaluate the impact the addition of sounds and animations had on Doggy, a study was planned and executed on the 15th of July 2019 during the *Open Campus* of the University of Bremen (cf. Figure 6-1). The plan was to divide a field study into two parts: one with the integration of sounds and animations into Doggy's gameplay and one with only the gameplay. Then the effects of the gameplay additions could be analyzed.

Figure 6-1: Doggy on a previous Open Campus. Picture by Schüthe (2016).

For the field study, it was important to create a short evaluation process, as the open campus is an exhibition where people do not remain in the same place for a long time. A long evaluation could severely reduce the willingness of people to participate. Therefore, a short questionnaire with only seven questions was created.

The goal of this questionnaire for the evaluation is to find out, how the addition of sounds and animations affects the experience of Doggy's users. The integrated parts were designed to create a more social robot that would appear more intelligent and natural to its viewers due to emotional states with social cues. Spillner (2018) evaluated the animations she created for Doggy and found that the expressed emotions have a high recognition rate. With the addition of sound and the focus on creating natural animations, it is yet again interesting to evaluate the expressiveness of emotional states. However, Spillner (2018) evaluated with a case study, therefore the comparison with a field study likely has a low meaningfulness, as different conditions apply to the two research methods.

Doggy, as an entertainment robot, was created to allow users to enjoy themselves. Hence, the experience the users have with the system is key. With the addition of sounds and animations, the experience should therefore be enhanced. However, sounds and animations also might deter users from using the robot, as a 2.1 m high robot with hectic movements might be fear inducing. The fear perceived might be correlated to which extent Doggy appears to be robotic. Nevertheless, the fluid, natural movements and socially designed interaction should reduce the deterring factor of Doggy.

For the creation of the questionnaire, several scientific works were analyzed concerning how robots were evaluated before. For their evaluation of the robot Probo Saldien et al. (2014) used the Godspeed questionnaire. The questionnaire was designed by Bartneck et al. (2009). It was created to unify the evaluation process of HRI questionnaires. In order to achieve comparable results between studies, a standardized measurement tool is needed. Bartneck et al. (2009) compared the questionnaires of various studies and evaluated their meaningfulness based on Cronbach's Alpha while also splitting the questionnaires thematically into five categories: Anthromorphism, Animacy, Likeability, Perceived Intelligence and Perceived Safety. Saldien et al. (2014, p. 3) agree with Bartneck et al. (2009) that "a standardized measurement tool for HRI studies is necessary to make progress in this field and to be able to compare the results from different studies." To make use of the existing data, to add to it and because the categories are highly relevant for Doggy, the Godspeed questionnaire was chosen for the field study. However, as discussed before, the length of a questionnaire has to be reduced to a minimum. Therefore, only a small selection of the seven seemingly most insightful attribute pairs were chosen (cf. Table 6-1). Table 6-2 shows from which Godspeed category which pair originates. The questionnaire can be seen in Figure 6-2.

Attribute A	***Attribute B***	***Usage***
Mechanical	*Organic*	*Illusion of life / Fluidity of animations*
Unintelligent	*Intelligent*	*Impact of the integration on perceived intelligence*
Moving rigidly	*Moving elegantly*	*Fluidity of animations*
Artificial	*Lifelike*	*Illusion of life / Fluidity of animations*
Dead	*Alive*	*Illusion of life / perceived vividness / exaggeration of animations*
Awful	*Nice*	*Impact of the integration on general likeability / Detect how terrifying the robot is*
Dislike	*Like*	*Impact of the integration on general likeability*

Table 6-1: The usage of the attribute pairs for the evaluation.

Anthromorphism	Moving rigidly/Moving elegantly	-	-
Animacy	Mechanic/Organic	Dead/Alive	Artificial/Lifelike
Likeability	Dislike/Like	Awful/Nice	-
Perceived Intelligence	Unintelligent/Intelligent	-	-
Perceived Safety	-	-	-

Table 6-2: The seven attribute pairs used for the five categories.

Bitte beurteilen Sie Ihren Eindruck des Roboters auf diesen Skalen:

Mechanisch	1	2	3	4	5	Organisch
Unintelligent	1	2	3	4	5	Intelligent
Bewegt sich steif	1	2	3	4	5	Bewegt sich flüssig
Künstlich	1	2	3	4	5	Realistisch
Tot	1	2	3	4	5	Lebendig
Furchtbar	1	2	3	4	5	Nett
Gefällt mir nicht	1	2	3	4	5	Gefällt mir

Figure 6-2: The questionnaire used for the evaluation. The translation was taken from the website of Bartneck (2008, March 11).

The evaluation of the field study should not be limited to the questionnaire. Bartneck et al. (2009, p. 72) state that "a much more reliable and possibly objective method for measuring the users' perception and cognition is to observe their behavior [6]. If, for example, the intention of a certain robot is to play a game with the user, then the fun experienced can be deduced from the time the user spends playing it". Hence, the observation of participants can grant additional insights. Certain aspects, such as the participants' fun or emotional state interpretation, could be analyzed by an outstanding, focused observer. As exhibition visitors often talk freely about what they see and experience to the people accompanying them, their expressed thoughts can be used for a qualitative evaluation. These expressions are unforced and natural, therefore more authentic than any answers given for a scientific survey. Hence, qualitative observations will be made during the execution of the study.

6.2 Study Execution

During the open campus exhibition, Doggy was presented to an audience of all ages. Visitors could interact with Doggy by using the provided balls. Occasionally, visitors that interacted with Doggy were asked to participate in the study. The questionnaire was filled in for the participants during a short interview.

Initially, Doggy was presented with the integration of sounds and animations. The goal was to reach around 30 participants for each part of the study. However, during the execution of the study, Doggy's hardware stopped working properly and rain caused downtime. In the end, Doggy was functional for only around 90 minutes. Five hours were planned. This caused severe problems for the evaluation. Firstly, because of a lack of time to interview and observe participants, the amount of evaluation data was limited. Secondly, since the robot needed to be fixed during the exhibition, participants might have gotten a negative impression of the robot before interacting with it. A defect robot may be perceived as less intelligent as a working robot. Furthermore, the fact that Doggy is a robot was highlighted, since the costume was taken off and its internals were revealed. Viewers might therefore be less likely to perceive Doggy as organic and natural. In total, it is very likely that the hardware problems influenced the evaluation data negatively. Since the open campus is an exhibition, it was likely that participants saw Doggy in the defect state. In addition, the data is exclusively relevant for the first part of the study, as there was no time to present Doggy without the addition of sounds and animations. In the end, only 13 participants could be interviewed.

Another problem was caused by the tail, which broke off during the start of the exhibition. The attempt to fix the tail with cable binders during the exhibition failed. The tail was removed. Therefore, the enhancement of emotional expressions by the tail was missing. This could have negatively affected the organic appeal and the naturalness of Doggy, since a dog without a moving tail is not as authentic.

With the exception of the tail, Doggy's hardware was working for around 90 minutes. The software, however, also caused issues. During the day, the weather was fluctuating severely. The mix of clouds and sunshine caused the settings of the cameras to be poorly adapted. This had an impact on the accuracy of the ball-tracking unit. Oftentimes balls were not detected. Hence, the gameplay was not working properly. It also caused Doggy to proceed in its emotional state hierarchy. This created situations, where users were throwing balls at Doggy and since no interaction was detected, Doggy would change into the states sad, disappointed and encourage. This created a dissonance in perception. Users might have been confused about the negative and prompting responses of the robot, even though they were interacting with it. The robot, however, did not detect any interaction and was trying to induce it. This dissonance likely negatively affects the sociability of Doggy, causing it to appear less organic, intelligent and natural. In general, it is probable that the likeability was negatively affected by every issue mentioned.

6.3 Study Results

The results of the questionnaire can be seen in Table 6-3. Since only the first part of the study could be conducted, there is no table for the second part of the questionnaire. Therefore, no comparison can be made to see the influences of the integrated system onto Doggy. Now only the system as a whole can be evaluated. Especially noteworthy are the high values for the attributes *nice* and *like*, while the value for the attribute *organic* is strikingly low.

Proband	Organic	Intelligent	Moves fluently	Lifelike	Alive	Nice	Like
1	2	4	2	4	4	5	4
2	1	3	5	3	3	5	5
3	1	2	5	3	3	5	5
4	1	3	3	3	2	4	5
5	2	4	2	4	5	5	5
6	2	4	3	3	5	5	5
7	3	4	3	4	2	5	5
8	3	3	4	2	4	4	4
9	2	3	3	1	2	5	5
10	1	5	5	1	1	5	5
11	1	5	5	5	5	5	5
12	1	3	3	2	5	5	5
13	1	3	3	2	5	5	5
Average	1.6	3.5	3.5	2.8	3.5	4.8	4.8

Table 6-3: Results of the questionnaire used for the first part of the study. For brevity, column descriptions only contain the right part of an evaluation pair. Therefore, a 5 completely agrees with the attribute, while a 1 completely disagrees.

	Organic	**Alive**	**Lifelike**	**Nice**	**Like**
Doggy	1.6	3.5	2.8	4.8	4.8
Probo	3.0	3.8	3.1	4.25	4.4

Table 6-4: Comparison between Doggy and Probo

Since the Godspeed questionnaire is standardized, research can be considered to find comparable values from other robots. However, it is difficult to find usable data, as most studies simply provide averages for the five categories of the Godspeed questionnaire (for example (Syrdal, Dautenhahn, Koay, Walters, & Ho, 2013), (Salem, Lakatos, Amirabdollahian, & Dautenhahn, 2015), (Haring, Matsumoto, & Watanabe, 2013)), use a variation of it (Hashimoto, Kobayashi, Polishuk, & Verner, 2013) or are too different in appearance and interactions (Ho & MacDorman, 2010). The robot Probo is not an entertainment robot, yet it was designed with the philosophy of creating an illusion of life. Probo also expresses emotions with its body and face and by using "non-sense speech" (Saldien et al., 2014, p. 2). Furthermore, Probo is "inspired from the principles of character animation" (Saldien et al., 2014, p. 4). The robot was also evaluated with the Godspeed questionnaire. The results of the categories animacy and likeability were presented in detail. Therefore, these specific attributes can be compared (cf. Table 6-4). Doggy performs worse in the first three attributes, especially regarding the attribute *organic*. Part of this could be because Probo expresses emotions also with facial expressions. Doggy performs better in the last two attributes *nice* and *like*, which could be explained since it is an entertainment robot. Nevertheless, the two robots' performances are relatively similar regarding all attributes except for the attribute *organic*. In the end, the results of the comparison might be limited due to different study modalities and also due to a lack of study participants for Doggy's questionnaire.

Qualitative observations were made during the execution of the study. In general, Doggy was well perceived. This was most noticeable during the interview for the questionnaire. When participants were asked to rate the attribute pair *dislike/like,* they heavily emphasized on how much they like the robot. The animations and sounds raised the interest of people as they walked by. Frequently, people stopped to observe Doggy. For most of the time the robot was working, a crowd of people was observing Doggy. There were times when the size of the crowd caused a blockage of the exhibition's path. Mostly families stopped, as children wanted to interact with the robot. Even though Doggy is a tall robot with a moving bat, children were not afraid of it. Since children often times did not want to stop interacting with Doggy, they were urged by their parents to move on. The general interaction time was around 10 minutes, while some viewers even stayed for over 20 minutes. After the robot played a ball and reacted with excitement, there were comments like "oh, he is happy" or "look how excited he is now". Even when there were problems regarding the ball tracking, the robot kept the viewers interested, as sounds and animations were played nevertheless. When Doggy advanced to the sad state and expressed sadness there were people commenting that they believe the dog is sad and expressed their empathy. Once a person walked by with his dog and as Doggy played sounds, the dog stopped, turned to Doggy and started to bark at it. The crowd reacted with amusement. The idle sounds were also noted by the viewers, since they commented on the panting and barking of Doggy.

6.4 Conclusion

A direct comparison between the two study parts was not possible. Therefore, insights regarding the effect of the addition of sounds and animations onto the system cannot be gained by comparing the study parts. The study results are also influenced by several issues, which likely affect the meaningfulness. Nevertheless, the study showed that the current state of Doggy is well perceived. Doggy is able to entertain a crowd and does so even when the ball-playing game does not work, as sounds and animations alone kept viewers interested. The participants often commented on Doggy's emotional expressions and were able to interpret the states correctly. The comparison with the robot Probo revealed that for the five comparable attributes both robots have rather similar results apart from the attribute *organic*. Doggy appears more *mechanic* than *organic*, even with the integration of sounds and animations. Doggy's evaluated results for the attributes *nice* and *like* are very satisfactory and are even higher than the results for Probo.

The results of the questionnaire and the qualitative observation revealed that Doggy is likeable and keeps viewers interested longer than expected. In the end, the integration of sounds and animations into the robot appears to be a beneficial addition for Doggy. However, exact statements regarding the influence of the integration cannot be made without the results of the second part of the study.

7. Discussion

As the study was troublesome due to several hard- and software issues, the results of it are obfuscated and limited. Consequently, the evaluation is not as extensive as desired. Since the results of the second part of the quantitative study is missing, no comparison can be made between the two study parts. Therefore, it is unclear to which extend the integration of sounds and animations affected the questionnaire. The value for the attribute pair *mechanic/organic* heavily favors the attribute *mechanic*. Now it is uncertain what caused the value to be this low and which impact the animations had. Contrary to the attribute *mechanic*, the animations were designed to appear natural and vivid based on the twelve animation laws and the illusion of life (cf. 3.4 Related Work Regarding Animations, 3.4.1 Twelve Animation principles by Ribeiro & Paiva). The evaluated value for the attribute *organic* is higher for Probo than for Doggy. This could be because Probo's animations were implemented more smoothly. It could also be, because the robot is able to express emotions with its face, because of its appearance or due to other reasons. For Doggy it will remain unclear how the attribute *organic* was influenced by the integration of sounds and animations. The same is the case for other attribute pairs. This is unfortunate, as the quantitative study was mainly designed to identify how the expressions affect the system. Further research is needed.

The qualitative observation of users, however, can be used to evaluate specifically regarding the integration of sounds and animations. It demonstrates that users were emotionally influenced by the expressions of Doggy. As expected, the expressions seemed to raise the interest of users regarding the robot. In section 3.2.1 (Design Approaches), two design approaches were analyzed. For this work, most of the time the functional design (cf. 3.2.1.2 Functional Design) approach was used. The usage of pre-canned sounds and animations turned out to result in emotional expressions that were sufficient for the interaction period of users, which on average was even longer than expected. The interaction of children with the robot was surprising, as they often stepped into the area of movement of Doggy. The robot did not cause any fear in children, even with hectic and abrupt movements. This has a negative side effect. Since children moved so close to the robot, the ball tracking was not able to recognize the balls thrown onto the robot. This added onto the already problematic ball tracking due to the changing weather conditions. As stated in section 5.2.3 (Integration of Modules), a compromise was made only to play balls in the idle state, as opposed to every state, due to issues with the compatibility between z-axis movements and the ball tracker. After some interaction, users started to notice this behavior and henceforth waited for Doggy to finish its expressions before throwing a ball. However, the state transition is rather smooth, so there were cases when users waited for Doggy to finish its expression yet threw the ball to early anyway, as Doggy still was not in the idle state. This is problematic, as it adds to the confusion created by the issues of the ball tracker.

The integration includes the addition of both sounds and animations, resulting in audiovisual stimuli. Crossmodal influences have been analyzed in section 3.3.3 (Effects of Audiovisual Stimuli). Research shows, that congruent audiovisual emotional expressions are more easily understood by humans and can create a more authentic socially acting robot. The integration of sounds and animations was evaluated to be a beneficial addition. Part of this could be because of the crossmodal influences. More research can be conducted to show the individual impact of either stimulus. Then the individual influences can be compared to the influence of the conjunction of stimuli to identify the impact of crossmodal influences.

Even with various problems during the execution of the study, Doggy was able to entertain many exhibition visitors. While the gameplay of Doggy likely will always be its central aspect, the integration of sounds and animations added new nuances to the behavior of it. Doggy appears a lot more interesting because emotions are expressed in a social way. The interest of viewers speaks for the visual appeal of the animations and the design of sounds. In general, the very positive reaction of viewers raises the importance of the integration of sounds and animations for entertainment robots.

7.1 Issues and Lessons Learned…

As not everything went according to plan, this section seeks to reveal what went wrong during the timeframe of the thesis and what can be learned from these issues. Particularly relevant is the section 7.1.2 (… Regarding the Study), since it addresses the problems that caused the study to yield insufficient data for the evaluation.

7.1.1 … Regarding the Implementation

The implementation process started out slowly, as an unknown operating system was the requirement for the development process, since ROS is only optimized for Linux. Furthermore, the usage of ROS had to be understood, which also took some time. On top of that, the unfamiliarity regarding the programming language C++ caused additional delay. Before the implementation process could be started, existing issues regarding the CMake build system had to be fixed. During the whole implementation process, there were issues with the PC of the robot, as a lack of hard drive space and RAM caused slow response times. An upgrade to the setup is recommended, since it only has 4GB RAM and ~50GB HDD space.

During the final stages of the implementation before the study, the robot's costume was put on again. Afterwards, the robot did not start, as initialization checks caused errors (ZERO_SWITCH_WRONG, CALIBRATION_INDEX_SWITCH_DISTANCE). This removed the possibility of improving certain animations before the study. During the design of the animations, a boundary limit was set for the axes (cf. 5.2.2.2.3 Creation of Animations). This is needed to prevent damage to the robot. However, if the robot is set to move to one side of an axis, while being on the opposite side, the velocity can become so high, that Doggy is moved outside of the boundary zone. Then the microcontroller automatically blocks all attempts to move the robot for a couple of seconds and slowly returns Doggy to the zero position. This also happens during gameplay of the robot. Sounds are still played normally, as the sound playback is unrelated to the microcontroller. An example of this is in the video microcontrolle_reset in the folder microcontroller_reset on the CD.

In section 5.2.2.2 (Design Tool: ROS Package "gestures") a controller was chosen as a design tool. A big part of the implantation process was spend on creating natural and fluid animations to allow Doggy to appear organic and lifelike for the illusion of life. However, the time spend to map the controller, implement the recording and playback of animations and especially the process of recording decent animations was way higher than initially expected (cf. 5.2.2.2.1 Mapping, 5.2.2.2.2 Recording and Playback, 5.2.2.2.3 Creation of Animations). Looking at the graphs in section 5.2.2.2.3 (Creation of Animations) it seems plausible that these animations could be created with a keyframing tool instead. In hindsight it might have been better to use a keyframing approach instead. However, the implementation of such a tool could also take a long time, but the process of creating animations will likely be quicker.

7.1.2 … Regarding the Study

In the morning of the day of the Open Campus, on which Doggy was to be presented and evaluated, the robot still was in a defect state. The initialization errors were temporarily fixed by readjusting the elastic band around the cogs of the internals. However, while repairing the robot, the tail broke off, because the robot initialization caused the tail to move while the tail was stuck in the costume. It was attempted to fix the robot before presenting Doggy by using a new fixation metal plate and cable binders. The quick fix did not last long, as initial gestures during the study caused the tail to fall off again.

The transport of the robot also caused issues. It was tedious to move Doggy, since the stand only has four very small wheels, which frequently get stuck when moved over uneven surfaces. Additionally, it is very easy for Doggy to fall over when hitting an edge. Therefore, a long, even way was preferred over a short, bumpy one. Before transporting Doggy, the cables have to be fixated properly. Otherwise, the transport will be more cumbersome. During one of the transports, a cable fell out and was damaged by the wheels. A plastic foil is used to cover Doggy in case of rain. This foil is a bit too long, so it can get stuck under the wheels. For the future it is recommended to change the form of transportation. For example, a cart could provide a quicker and safer transport if Doggy is fixated properly.

The weather caused various changes in lighting during the study. The camera settings had to be adjusted according to the current illumination multiple times. When clouds appear, objects might be too dark if the camera was adjusted for sunlight and vice versa. It would be optimal to place Doggy in a location that avoids direct sunlight. Furthermore, there is potential to change the ball tracker to dynamically adapt the camera settings depending on lighting conditions.

During the study people frequently asked for the purpose of the robot is and where to throw the ball at. The setup revealed a monitor with the tracking of the balls, the cameras of the robot and the balls, which were used for the gameplay. Some people were able to deduce the purpose of the robot from this, however this was not clear to all viewers. The setup could be adapted to allow for an easier interpretation of the goal and purpose of the robot. For example, a net behind the robot could reveal that it is expected that balls are thrown in the robot’s direction. There were also issues with the distance from which balls were thrown. The optimal distance is about 4 meters from the robot. A marker on the ground could indicate the position from which the ball is best thrown onto the robot.

While Doggy was presented, the initialization issues returned. The robot could be fixed temporarily. However, after some time the camera stopped working or the issued movement commands were not executed anymore. Therefore, the robot had to be restarted. Then the initialization issues reoccurred. This happened multiple times during the day and it limited the time that could be used to explain the robot to viewers or perform the evaluation. At least the viewers partly reacted with understanding, as they stated that Doggy is just too hungry or not in the mood.

After the study, it was attempted to locate the cause of the errors. The error ZERO_SWITCH_WRONG could be removed by restarting the robot from different angle positions. The problem of the CALIBRATION_INDEX_SWITCH_DISTANCE had to be solved differently. Due to the elastic hysteresis of the rubber bands used for the cogs, a threshold was defined in the microcontroller for acceptable switch distance calibration. The assumption for the error was that there is fatigue in the material of the rubber bands caused by the heavy usage for the animation design process. Furthermore, it was assumed that the added costume caused additional friction during the calibration. In order to pass the initialization check, the internal threshold of the acceptable switch distance was increased and flashed onto the microcontroller. Now the microcontroller is able to pass this initialization check and can start normally. However, material fatigue could cause issues in the future again. After the problem was fixed, the tail of Doggy was repaired and now is also functional. The repair of the robot revealed that the accessibility of internal parts is cumbersome due to the non-modular design of the robot. Additionally, the costume should be changed so that it can be put on and off easily. A zipper throughout the whole costume is desirable.

7.2 Future Work

Several improvements can be made for the system. The ball tracker can be changed to dynamically adapt lighting settings fitting to the current situation. This allows Doggy to be more robust regarding fluctuating weather conditions. Consequently, the robot will be able to play more balls during an exhibition. Also, the ball tracker needs to be adapted to allow for z-axis movements during the ball tracking. This creates more freedom regarding the design of animations.

Doggy was designed to react to the outcome of a played ball. However, since the hit tracker does not work properly, there were often times false positives and negatives. Due to this, the reactions had to be limited and expressed excitement was partly misplaced. The improvement of the hit tracker can fix these issues. The hit tracker could make use of the microphone data, which was not available for this thesis, as the microphones used by Bartsch (2015) were no longer attached to the robot.

A design tool was implemented for the creation of animations for Doggy by using a controller. There is potential to improve the process of animation creation and the quality of animations by creating a keyframe based design tool. Then the created animations could be compared and evaluated. Similarly, the creation of sounds could be adapted. As various robots use synthesized voices, the addition of a synthesized dog voice might result in a wider spectrum of sounds, which allows Doggy to express emotions in a better way. Related robots use facial expressions to communicate emotions. Research revealed that certain emotions are better portrayed by facial expressions than body expressions. The addition of facial expressions could build upon the expressiveness of Doggy, since more visual stimuli are added.

More research has to be conducted regarding various aspects. As the animations of Spillner (2018) were added and extended with more animations and sounds, it is again necessary to evaluate the expressiveness of each emotion expressed by Doggy. Since the second part of the planned study was not executed, more research needs to be conducted about the impact of the integration of sounds and animations into the ball playing part of the system. Furthermore, to see which individual impact sounds or animations have onto the system, research needs to be conducted that compares the integration of sounds to the integration of animations. This not only demonstrates the effectivity of either stimulus, but also helps to see which impact the crossmodal influences have on the robot.

Bibliography

Antona, M., Ioannidi, D., Foukarakis, M., Gerłowska, J., Rejdak, K., Abdelnour, C., … Roberto, N. (2019). My robot is happy today : how older people with mild cognitive impairments understand assistive robots ’ affective output. *Proceedings of the 12th ACM International Conference on Pervasive Technologies Related to Assistive Environments PETRA ’19*, 416–424.

App, B., McIntosh, D. N., Reed, C. L., & Hertenstein, M. J. (2011). Nonverbal Channel Use in Communication of Emotion: How May Depend on Why. *Emotion*, *11*(3), 603–617. https://doi.org/10.1037/a0023164

Arkin, R. C., Fujita, M., Takagi, T., & Hasegawa, R. (2003). An Ethological and Emotional Basis for Human-Robot Interaction.

Balit, E., Vaufreydaz, D., Reignier, P., Balit, E., Vaufreydaz, D., Reignier, P., … Reignier, P. (2016). Integrating Animation Artists into the Animation Design of Social Robots Integrating Animation Artists into the Animation Design of Social Robots Author version. *Human-Robot Interaction*, 417–418.

Bartneck, C. (2002). Integrating the OCC model of emotions in embodied characters. *Proceedings of the Workshop on Virtual Conversational Characters*, 39–48.

Bartneck, C., Kulić, D., Croft, E., & Zoghbi, S. (2009). Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International Journal of Social Robotics*, *1*(1), 71–81. https://doi.org/10.1007/s12369-008-0001-3

Bartsch, M. (2015). Sound of Interest Ein Ballspielroboter hört stereo.

Breemen, A. Van. (2004). Bringing robots to life: Applying principles of animation to robots. *… of Shapping Human-Robot Interaction Workshop Held …*, *4*, 1–5.

Chen, Y. C., & Spence, C. (2017). Assessing the role of the “unity assumption” on multisensory integration: A review. *Frontiers in Psychology*, *8*(MAR), 1–22. https://doi.org/10.3389/fpsyg.2017.00445

Dautenhahn, K. (2007). Socially intelligent robots : dimensions of human – robot interaction, (February), 679–704. https://doi.org/10.1098/rstb.2006.2004

Embgen, S., Luber, M., Marco, C. B., Vanessa, R., & Kai, E. (2012). Robot-Specific Social Cues in Emotional Body Language, 1019–1025.

Fogg, B. . (1999). Persuasive Technologies Persuasive Technologies. *Communications of the ACM*, *42*(5), 26–29.

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2002). A Survey of Socially Interactive Robots : Concepts , Design , and Applications, (November).

Fujita, M. (2001). AIBO: Toward the era of digital creatures. *International Journal of Robotics Research*, *20*(10), 781–794. https://doi.org/10.1177/02783640122068092

Gelder, B. De, & Vroomen, J. (2000). The perception of emotions by ear and by eye, *14*(2000), 289–311. https://doi.org/10.1080/026999300378824

Haring, K. S., Matsumoto, Y., & Watanabe, K. (2013). How do people perceive and trust a lifelike robot. *Lecture Notes in Engineering and Computer Science*, *1*, 425–430.

Hashimoto, T., Kobayashi, H., Polishuk, A., & Verner, I. (2013). Elementary science lesson delivered by robot. *ACM/IEEE International Conference on Human-Robot Interaction*, 133–134.

https://doi.org/10.1109/HRI.2013.6483537

Ho, C. C., & MacDorman, K. F. (2010). Revisiting the uncanny valley theory: Developing and validating an alternative to the Godspeed indices. *Computers in Human Behavior*, *26*(6), 1508–1518. https://doi.org/10.1016/j.chb.2010.05.015

Kerepesi, A., Kubinyi, E., Jonsson, G. K., Magnusson, M. S., & Miklósi, Á. (2006). Behavioural comparison of human-animal (dog) and human-robot (AIBO) interactions. *Behavioural Processes*, *73*(1), 92–99. https://doi.org/10.1016/j.beproc.2006.04.001

Laue, T., Birbach, O., Hammer, T., & Frese, U. (2014). An entertainment robot for playing interactive ball games. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8371 LNAI*, 171–182. https://doi.org/10.1007/978-3-662-44468-9_16

Okuno, H.G., Nakadai, K., Hidai, K.-I., Mizoguchi, H., & Kitano, H. (2001). Human-robot interaction through real-time auditory and visual multiple-talker tracking. *IEEE International Conference on Intelligent Robots and Systems*, *3*, 1402–1409.

Okuno, Hiroshi G, Nakadai, K., & Hidai, K. (2001). Human-Robot Interaction through Real-Time Auditory and Visual Multiple-Talker Tracking, 1402–1409.

Pineau, J., Montemerlo, M., Pollack, M., Roy, N., & Thrun, S. (2003). Towards robotic assistants in nursing homes: Challenges and results. *Robotics and Autonomous Systems*, *42*(3–4), 271–281. https://doi.org/10.1016/S0921-8890(02)00381-0

Ribeiro, T., & Paiva, A. (2012). The illusion of robotic life, (1937), 383. https://doi.org/10.1145/2157689.2157814

Saldien, J., Vanderborght, B., Goris, K., Van Damme, M., & Lefeber, D. (2014). A motion system for social and animated robots. *International Journal of Advanced Robotic Systems*, *11*(1), 1–13. https://doi.org/10.5772/58402

Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Would You Trust a (Faulty) Robot?, 141–148. https://doi.org/10.1145/2696454.2696497

Schafer, R. W. (2011). What is a savitzky-golay filter? *IEEE Signal Processing Magazine*, *28*(4), 111–117. https://doi.org/10.1109/MSP.2011.941097

Schüthe, D., & Frese, U. (2014). Task level optimal control of a simulated ball batting robot. *ICINCO 2014 - Proceedings of the 11th International Conference on Informatics in Control, Automation and Robotics*, *2*, 45–56.

Sharkey, A., & Wood, N. (2014). The Paro seal robot: demeaning or enabling? *Proceedings of AISB 2014*.

Spillner, L. (2018). Interacting With a Ball-Playing Entertainment Robot.

Stock, J. Van Den, Grèzes, J., & Gelder, B. De. (2008). Human and animal sounds influence recognition of body language. https://doi.org/10.1016/j.brainres.2008.05.040

Syrdal, D. S., Dautenhahn, K., Koay, K. L., Walters, M. L., & Ho, W. C. (2013). Sharing spaces, sharing lives - The impact of robot mobility on user perception of a home companion robot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8239 LNAI*, 321–330. https://doi.org/10.1007/978-3-319-02675-6_32

Tzeng, Y. (2013). New Dog New Tricks.

Valin, J.-M., Michaud, F., Rouat, J., & Letourneau, D. (2004). Robust sound source localization using a microphone array on a mobile robot, (1), 1228–1233. https://doi.org/10.1109/iros.2003.1248813

Van den Stock, J., Righart, R., & de Gelder, B. (2007). Body Expressions Influence Recognition of Emotions in the Face and Voice. *Emotion*, *7*(3), 487–494. https://doi.org/10.1037/1528-3542.7.3.487

Vatakis, A., & Spence, C. (2007). Crossmodal binding: Evaluating the “unity assumption” using audiovisual speech stimuli. *Perception and Psychophysics*, *69*(5), 744–756. https://doi.org/10.3758/BF03193776

Wada, K., & Shibata, T. (2007). Living With Seal Robots — Its Sociopsychological and Physiological Influences on the Elderly at a Care House, *23*(5), 972–980.

Ziv, M., & Frye, D. (2003). The Relation Between Desire and False Belief in Children’s Theory of Mind: No Satisfaction? *Developmental Psychology*, *39*(5), 859–876. https://doi.org/10.1037/0012-1649.39.5.859

Open Source Robotics Foundation (2019). ROS official website. Retrieved from https://www.ros.org/

Bartneck, C. (2008, March 11). The Godspeed Questionnaire Series. Retrieved from https://www.bartneck.de/2008/03/11/the-godspeed-questionnaire-series/

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

Pascal Weidenbach