Karlsruhe Institute of Technology

A Social Robot's Emotion-Adaptive Feedback

Master's Thesis
of

Bettina Weller

KIT Department of Informatics
Institute for Anthropomatics and Robotics (IAR)
Humanoids and Intelligence Systems Lab

Referees:	Prof. Dr.-Ing Rüdiger Dillmann Prof. Dr.-Ing Rainer Stiefelhagen
Advisors:	Prof. Dr.-Ing Rüdiger Dillmann Dr. Sarah Cosentino

Duration: November 15th, 2017 – May 14th, 2018

Statement of Authorship

I hereby declare that I have developed and written the enclosed thesis completely by myself and have not used sources or means without declaration in the text. I have followed the respectively valid KIT statutes for safeguarding good scientific practice.

Karlsruhe, May 14th, 2018

Bettina Weller

Acknowledgement
First of all, I want to thank CLICS, Professor Dillmann and Professor Takanishi for their support and giving me this opportunity to do research at Waseda and by the way fulfilling one of my dreams (visiting Japan and getting to know more about the culture). Next, I want to thank the people at Takanishi lab for their guidance and lots of funny moments. I hope you get to have some free time! Of course I also thank my neighbours for going around the different parts of Tokyo end enjoying food with me. I have really enjoyed the time with you all. Besides I want to thank my family and friends for looking after me and for exchanging about what is happing. Last, I am extremely thankful towards all participants of my online survey and the people helping me to find more! All in all, I got to know that the so many social and friendly people are inhibiting our world! Thank you!

Abstract:

To achieve good human-computer interaction it is essential for both sides to understand the intentions of the other side. Alongside the spoken dialogue emotional information is hidden. Sometimes this information is needed to understand spoken and unspoken intentions. Humans learn to perceive and process this information further during growing up as it is a crucial part of communication. Therefore, for a good human-computer interaction this skill is also desired to make a dialogue seem more natural and flexible to the current situation.

Some approaches tried to use body gestures, others tried to extract emotions from audio data and others from video data of one's face. We want to focus on emotion recognition by facial expressions and use the gained information to enhance communication between human and robot. For our study, we are evaluating the emotion recognition of the Affectiva SDK (3.4.1) regarding face detection and emotion recognition in general. Further, we compare the performance for Caucasians and for the predominant ethnicity in Japan.

Depending on the current mood of the user a robot's feedback will be different dependent on what the user currently needs. As target user we have chosen elderly in a care home and hence focus on human-robot interaction in their daily life providing some basic use cases. Further, we have designed a survey to gather information regarding the acceptance of an empathetic robot.

Kurzzusammenfassung:

Eine Grundvoraussetzung für gelingende Mensch-Maschine-Interaktion ist gegenseitiges Verständnis. Neben ausgesprochenen Informationen liegen einem Dialog weitere Informationen bezüglich Emotionen zugrunde. Diese werden für ein volles Verständnis während der Interaktion benötigt. Menschen lernen diese Informationen wahrzunehmen und zu verarbeiten, ansonsten sähe Interaktion zwischen Menschen anders aus als wir sie heute kennen. Daher ist diese Fähigkeit auch für die Interaktion zwischen Mensch und Maschine gwünscht, um jene natürlicher an die momentane Situation anpassen zu können.

Einige Ansätze nutzen Gestik, andere Audiodaten oder Gesichter, um Emotionen aus diesen zu extrahieren. Wir fokussieren uns auf Emotionserkennung anhand von Mimik und nutzen diese Informationen für bessere Kommunikation zwischen Mensch und Roboter. Hierfür evaluieren wir die Emotionserkennung von Affectiva SDK (3.4.1), allgemein für Gesichtsdetektion und Emotionserkennung, sowie im Vergleich zwischen kaukasischen und in Japan vorherrschenden Gesichtszügen.

Abhängig vom Gemütszustand des Nutzers soll ein Roboter sein Verhalten anpassen, um diesem zu entsprechen. Als Zielperson haben wir Senioren in Pflegeheimen gewählt und konzentrieren uns daher auf Mensch-Maschine-Interaktion in ihrem Alltag. Hierzu zeigen wir Anwendungsfälle und nutzen eine Umfrage, um die generelle Akzeptanz eines mitfühlenden Roboters zu ergründen.

Contents

1. Introduction **1**
1.1. Motivation . . . 1
1.2. Objectives . . . 3
1.3. Field of application . . . 4
1.4. Structure of this work . . . 5

2. Background **7**
2.1. Robotics . . . 7
2.2. Emotions . . . 8
2.3. Robotics and emotion . . . 11
2.4. Related work . . . 12

3. Emotional feedback **15**
3.1. Emotion extraction . . . 15
3.2. Behaviour . . . 15
3.2.1. Humanoid . . . 16
3.2.1.1. Gesture . . . 16
3.2.1.2. Face . . . 17
3.2.1.3. Voice . . . 17
3.2.1.4. Clothes . . . 18
3.2.2. Other . . . 18
3.2.2.1. Colour . . . 18
3.2.2.2. Screen . . . 19
3.2.2.3. Distance . . . 19
3.2.2.4. Music and other sounds . . . 19
3.2.3. Context . . . 20
3.2.3.1. Time . . . 20
3.2.3.2. Emotion change and stability over time . . . 20
3.2.3.3. Target user . . . 21
3.2.3.4. User profile . . . 21
3.2.4. Robot's emotions . . . 21
3.2.5. Use cases . . . 22
3.2.5.1. Use case 1: Greeting a happy person . . . 22
3.2.5.2. Use case 2: Chased off . . . 22
3.2.5.3. Use case 3: Birthday . . . 23
3.2.5.4. Use case 4: Night patrol . . . 23
3.2.5.5. Use case 5: Sad user . . . 23
3.2.5.6. Use case 6: Good night . . . 23
3.2.5.7. Use case 7: Confused person during interaction . . . 23
3.2.5.8. Use case 8: Pepper being bored . . . 23

4. Pepper - A social robot **25**
4.1. Overview . . . 25
4.2. Pepper's feedback possibilities . . . 26

5. Tools and implementation **29**
5.1. Tools . . . 29
5.1.1. Boost 1.65.1 . . . 29
5.1.1.1. Building and linking . . . 29
5.1.2. Opencv 3.3.0 . . . 29
5.1.2.1. Linking . . . 29
5.1.3. AffdexSDK 3.4.1 . . . 29
5.1.3.1. Requirements and dependencies . . . 30
5.1.3.2. Embedding in Visual Studio 2017 . . . 30
5.1.4. Rapidjson . . . 30
5.1.4.1. Linking . . . 30
5.1.5. Choregraphe 2.5.5 . . . 30
5.1.5.1. Troubleshooting . . . 32
5.2. System for emotion recognition . . . 32
5.2.1. Requirements . . . 32
5.2.2. Integrating Affectiva . . . 33
5.2.3. Main emotion . . . 34
5.2.4. SVM for neutral . . . 35
5.2.4.1. Training . . . 35
5.2.4.2. Integration . . . 36
5.2.4.3. Evaluation . . . 36
5.2.5. Threshold . . . 36
5.2.6. Enums . . . 37
5.3. Programming Pepper . . . 37
5.3.1. Birthday . . . 37
5.3.2. User starts interaction . . . 38
5.4. Connecting . . . 39

6. Experiments and studies **41**
6.1. Face detection and emotion recognition . . . 41
6.1.1. Face detection . . . 41
6.1.2. Emotion recognition . . . 41
6.2. Interaction . . . 42
6.2.1. Structure of the survey . . . 42
6.2.2. Scenarios . . . 46
6.2.3. Conduction . . . 47

7. Evaluation **49**
7.1. Databases . . . 49
7.1.1. Head Pose Image Database . . . 49
7.1.2. KDEF database . . . 49
7.1.3. Extended Cohn-Kanade dataset . . . 50
7.1.4. JAFFE database . . . 50
7.1.5. Mixed subset . . . 50
7.2. Affectiva SDK . . . 50
7.3. Survey . . . 56
7.3.1. Demographics . . . 57
7.3.2. Qualitative . . . 57
7.3.2.1. Goal of the interaction . . . 57
7.3.2.2. Feedback about the behaviour . . . 59
7.3.2.3. General feedback . . . 60
7.3.2.4. Most liked . . . 60
7.3.2.5. Least liked . . . 60

7.3.2.6. Wishes 60
7.3.3. Quantitative 61
7.3.3.1. General 61
7.3.3.2. All scenarios 61
7.3.3.3. Only scenario 1 65
7.3.3.4. Only scenario 2 65
7.3.3.5. Only scenario 3 66
7.3.3.6. Only scenario 4 66
7.3.3.7. Overall 66
7.3.3.8. Limitations 68

8. Conclusion and future work **69**

A. Appendix **77**

1. Introduction

Thinking about human-computer interaction there are several ways of interaction possible. It may be a rigid system where the user can choose from options and is directed and controlled by the system at the same time. On the other hand, the user may have more freedom at the cost of fail safety on the system side, which can cause inadequate interactions or even failures. Depending on the intended environment both options have their advantages and disadvantages. A given choice already provides some basic information about how to interact and how to use the system. In addition, many humans act towards machines similarly as they would towards humans ([71]). Therefore, it is fail safer but may need some initial training for use whereas the other way of system design is more flexible and expects less initial training from the user. Furthermore, the system may seem much more human-like and thus less invasive just by the way it reacts.

On one hand, using the system is more natural and intuitive through being more human-like, also for people who are usually not using such services. On the other hand, it causes the system to be more liked and less being seen as something invasive ([65]). Focusing on daily use acceptance is a very important point for making it possible to integrate computers and robots in particular as well as guaranteeing a positive feeling of utilisation. Therefore it should be focused on suitable feedback systems, meaning suitable for the environment and the user. In some cases this contains emotional sensitivity, e.g. to lessen a user's frustration and make the interaction more enjoyable ([69]).

1.1. Motivation

During rational human-human interaction the aim is gaining information ([81]). It is characterised by logical thinking and extracting possible information that way. Here, the logical thinking is shaped by one's beliefs and knowledge. For example, when you hear the sentence "my neighbour is visiting his sister" you can conclude that the neighbour is male and has a sister. In case you know the neighbour and further you know that he does not have a sister you can conclude that the statement is wrong. The information can be extracted directly out of what is said and what you know but it does not give hints about how to react towards the said. However, emotional communication is different. Hereby the communication is not only on basis of what is said but also on other channels, verbal and non-verbal. This communication gives hints about how to react ([83]), e.g. if someone simply wants you to listen or expects a certain kind of reaction. In figure 1.1 person A is telling person B that he wants something. The

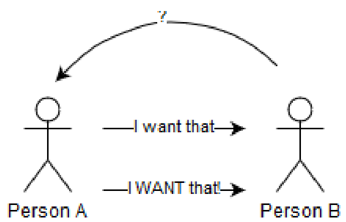


Figure 1.1.: An example for rational and emotional communication. "I want that" contains a rational statement while "I WANT that!" contains additional emotional information by accentuating *want*. Various emotional information is possible and this is just representative.

first case "I want that" is a statement, while the second case "I WANT that!" contains further information. It could be that Person A wants the thing now or thinks of it as necessary. In case A is a child and B is its parent then B could decide to buy the wanted thing for the child because it seems to be that important. He can also decide not to buy it but the child may start to cry and B has to be prepared for that. The possible reactions of A can be guessed by B but may be wrong. Such situations occur in daily life.

Thinking about daily life everyone goes through situations where he does not know how to react to another person during interaction. For example if a person is telling you about something really bad happening to him, what would be the best way to respond to him? Hugging and consoling or ignoring or trying to change the topic and to increase his mood, to mention some possible reactions. Hereby it makes a huge difference whether you know the other person or not and if you like him. Those aspects are illustrative for the different factors which influence your actions. As an appropriate reaction is already strenuous to decide for humans it is even harder for machines. First it needs to comprehend the person's emotions or other context. Next, it needs to choose an appropriate way of handling the grasped situation.

There are multiple methods for machines to guess a human's current emotions. For example direct active input by the user or someone else ([72]). Alternatively indirectly, e.g. through analysis of face, analysis of voice or language, sensors measuring physiological data or even randomly guessing to name some ([56], [70], [78]). Most humans can perceive other people's signals and interpret them ([54]). Some of those interpretations vary according to culture and therefore machines also should differentiate. The same applies for the acceptance of the reaction.

In each culture or probably even for each single person the acceptance for a specific behaviour varies. For example, the distance between two conversation partners. In some countries the normal distance is much closer or further apart than in other countries so you would feel anxious if you think someone is too close while the other person thinks the distance is normal. Another example are gestures. Some gestures may be friendly in one country while being offensive or not even known in others ([12], [35], [49]). Therefore it is important to evaluate feedback strategies through a cultural lens to make it possible to integrate them in various places over the world.

Sometimes the signals humans send were actively decided by them to communicate certain messages to their conversation partner. For example an angry look by your mother when you have done something bad. On the other hand some of those signals are done unconsciously ([54]). For instance a short shriek when being caught off guard by someone startling you. Such signals convey emotional information, in our example the surprise of the startled person. Further this information is based on the person's inner state as well as the external stimuli by being startled. Analysing such signals can help while choosing appropriate reactions as well as evaluating whether an action has been appropriate or not. In our example the next emotion of the startled person may be anger and we can deduct it may not have been the best idea to disturb him. On the other hand he can smile being happy to see you. Next, your reaction depends on this emotional reaction.

A good approach for emotion recognition is analysing one's face as we tend to show emotions clearly on our faces ([38]). In addition, some facial expressions seem to be recognised cross-culturally ([36]). Therefore using facial expressions to recognise emotions would simplify testing general and more specific feedback approaches as well as enable transferring the results.

If a robot successfully recognises a human's emotion and chooses a suitable feedback strategy it can be used in several ways. As a basic conversation partner or companion who socializes with a human as well as a service robot or for health surveillance. In various environments it has to react according to a human's emotion. For example in the role of a social robot he could show sadness if he is told something sad while in the area of health surveillance he should call for medical care if a nearby human has sudden pain. In case of health surveillance getting physiological data of the humans would be useful to help the robot decide whether an emergency occurs or not. Some of this data can also be used for emotion recognition, for example your pulse, which can change depending on your psychological state.

We are using facial expressions to detect a human's emotions. In case other data is used as well the same feedback strategies can still be applied by using the same interface or adapting it. Therefore the decision process has to provide a fixed interface about what kind of factors are involved in the process but those factors can be delivered from various sources. This makes the stimuli sources exchangeable. For instance in dark surroundings it may be harder to extract emotions according to facial expressions alone and the voice could be used as an additional feature. The result would still be the current emotions which are expected as input for the reaction finding.

The shown reaction of a human is usually not only communicated in one way but may use several features like posture, voice and eyes. Mehrabian said that besides the spoken information the underlying communication is important as well especially for the way humans interpret the spoken information which makes it necessary to communicate on multiple channels. For instance if a person says "you are hurting me" it makes a huge difference how this is said. If the speaker looks angry he probably really is hurting. On the other hand, if he smiles it could be irony as a reaction towards someone else. For a correct analysis context is needed. Another reason for using multiple channels is communication with impaired.

By providing the information on multiple channels we can address more people. If a posture which communicates anger is shown to a visually impaired person it is possible that the posture is not recognised at all. If at the same time the speaker's voice becomes louder and faster this communicates signs for anger and will likely be better recognised. Whatever can be used to express emotions can also be used for feedback. Hence it makes sense not to use only one way of feedback during communication but to use multiple ways.

In total, appropriate feedback depends on the goal of the interaction as well as the other party. However, restricting target user and environment makes it possible to develop strategies which are suitable for them. By choosing facial expressions to conclude the user's current emotions and suggesting strategies about how to handle situations we provide a basic approach to interaction between a robot and elderly people. Further, a well designed interface allows using other features than facial expressions or even other systems to provide information about a person's emotions. With this information the same process for feedback decision can be used.

1.2. Objectives

During social interaction it is important to detect the other party's emotions to understand the interaction as well as for making the other party feel understood ([83]). Often those emotions can be concluded from different channels during the interaction for example shape of voice, facial expressions or gestures. For instance, a confused face will request you to explain again or in detail. The fast interpretation of such signals increases information exchange. Furthermore it decreases miscommunication though it may not be fail-safe and multiple interpretations are possible. This example further shows that sent signals can expect certain reactions. In this case it is an explanation. In other cases it may be laughing or asking questions or even leaving. Depending on how the situation should end the choice of reaction differs. Humans have learned possible reactions but still have to decide for each interaction how to react. If we want to integrate robots in daily life they have to learn some strategies, too. We have solved this problem in a basic way which can guide future work in this area.

In general, it needs to be decided on what kind of stimuli one should react and what other features will be used to decide on the kind of reaction. Further, the whole interaction is shaped by an ultimate goal from making another person happy or motivate him to do something up to doing nothing. However, this goal mainly depends on the actual field of application and is restricted by its domain.

In this work we want to show the performance of some emotion recognition tools which are based on facial expressions as well as present a small framework for using one of those systems. Further we want to present an appliance which uses emotions as input and decides on a reaction towards these emotions. The decision process is among others influenced by the recognised emotions. The performance for emotion recognition will be compared between the faces of Caucasians and the predominant ethnicity in Japan. In the following we will refer to them as *western* and *japanese* faces. Besides the emotions a goal for the outcome of the interaction and some other components influence the choice of reaction. We will present some basic use cases and design a survey to gather opinions about possible interactions. For those interactions we are using a robot, namely a Pepper. However, our work can be transferred to other platforms. The systems for the emotion recognition and possible reactions are loosely coupled and individually exchangeable as well as integrable on various platforms. Figure 1.2 shows the basic process during interaction, which leads us. Recognizing sent emotions and using those emotions to decide on feedback. As target user we have decided on elderly. Nonetheless, our results are still applicable for

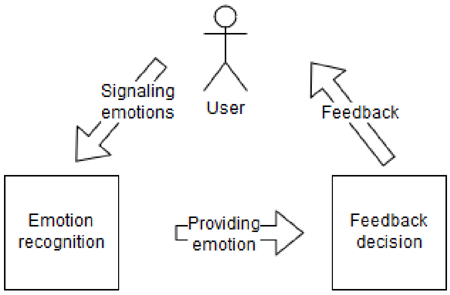


Figure 1.2.: Overview of our system

other users.

According to OECD the elderly population has been growing over the last years. In Japan it has been about 25% in 2013 and in Germany about 21% in 2014 and is still growing, which results in a rising need of caregivers ([6]). To keep up we see the need to further support care taking of elderly as well as providing entertainment and social interaction in daily life. Hence, we have chosen elderly people in a care home as our target of interaction. This mainly effects our use cases and the choice of provided strategies as the target user influences the goal of the interaction and the feedback. However, we want to note that elderly are not the only ones who could use a robot assistant, for example people in need of care after an accident at work. Further, the acceptance from family and friends is also important as they might feel uncomfortable about leaving someone with a robot.

1.3. Field of application

Since our work is partitioned into two parts which can individually be exchanged the field of application is quite wide. The whole application can mainly be used where visual input of faces is provided. For instance during interaction between any kind of robot which uses a camera and a human party. A special case would be health care where people can be consoled and calmed by a robot and the robot can estimate his success based on facial expressions as well as choose future options to further succeed. In case it does not succeed or is overburdened as well as in emergency cases the robot can alarm personnel and get additional aid by humans. Hence he supports medical staff and eases their work. A similar scenario would be a robot acting as a receptionist and doing simple booking tasks as well as answering questions of its customers. In special cases it can get support from human staff. During such interactions, it would make sense not only to use visual input and extract facial expressions but other types of input as well.

Instead of only visual input, other input, as well as combinations of multiple sources could be used. An example would be phoning a hotline where a system is taking care of you. However, in case your voice sounds angry instead of the system a real person might come to take care of you. Another possibility may be a change of the system's strategy while handling you. If other features like heart rate would be used in addition it would be easier to determine your actual emotional state as one feature is not enough to be 100% sure.

Instead of using the provided information to generate reactions it can also simply be forwarded. For example to help visually impaired to grasp the situation[1] or to analyse videos and produce data for other kinds of training. For instance, a film could be analysed and the spoken text could be emotionally tagged. This data can be used for chatbots. On the other hand, it can simply be provided in subtitles for deaf people instead of a human tagging them manually. Furthermore, it can be used during studies regarding emotion expression or for analyzing interviews.

In total human-robot interaction based on emotions has a huge application possibility for supportive actions especially in the health sector but also in other sectors. Momentarily only basic actions are

[1] Although they usually are trained in using other stimuli.

plausible. However, after more research in this area robots could become further supportive and get closer to an actual personal robot.

1.4. Structure of this work

This work is structured as follows: After an introduction to our work, some information about the background is presented in chapter 2. This includes information on robotics, emotional theories and the recognition of emotions based on facial expressions, followed by information about emotions and robotics combined. We conclude with an overview of related work regarding emotion recognition and human-robot interaction.

In chapter 3 we describe how we extract emotions and how we handle those. Hereby we are using Affectiva to get emotional information of facial images. As a next step, we show what kind of feedback can be used to show emotions and describe the feedback we suggest to use. Further, we present some use cases. In chapter 4 we present some details about the robot, Pepper, and what it is able to do.

Our implementation is shown in chapter 5 and starts with the tools we have used, followed by the integration and usage of the Affectiva SDK. Afterwards, a description of strategies to extract neutral from the emotional information follows, concluded by an overview of our programme.

For evaluation purposes, we have designed some experiments which are described in chapter 6. In a first phase occupying with face and emotion recognition, in a second phase about feedback strategies. The recognition is tested with labelled datasets while some feedback strategies were filmed and afterwards evaluated through a survey.

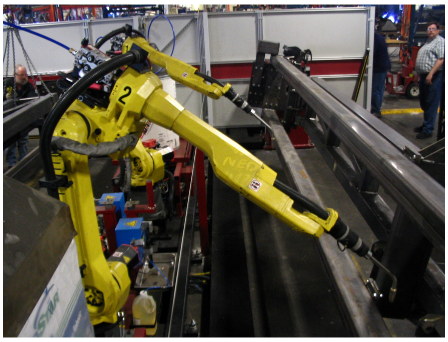
To show the performance of our programme we evaluate it and show the results in chapter 7. Following is the evaluation of our feedback strategies as described in chapter 6. Our work is summarized in chapter 8.

2. Background

In the following, we want to give some background information needed for our work. This includes information about robotics and emotional theory especially about their connection and human-robot interaction. A well designed human-robot interaction is important for integrating robots in daily life but still dependent on the circumstances the robot will be used in because they determine required ways of reacting. Up to now many projects for human-computer interaction have already been conducted therefore we refer to some of them.

2.1. Robotics

When thinking of robotics one of the first things coming to mind might be machines used in factories for production like shown in image 2.1(a). Another possible option would be the robot R2D2 from the movie Star Wars. At first glance both might be totally different as one is always repeatedly doing exactly the movements it was programmed to do while the other is also taking steps on its own. But what they have in common is that they have been designed for a specific purpose[1], the work they are doing. In general this means supporting humans during specific tasks. In the industrial robot's case it is assembling. In R2D2's case it is following the tasks its owner gives it and further support him. However, both are using



(a) An industrial robot for welding ([3])

(b) The robot R2D2 from the Star Wars movies

Figure 2.1.: Two robots - an industrial robot and R2D2

sensors to sense their surroundings and to act according to this context and what they were designed to do.

In the history of robotics multiple robots have been developed and can be used in different areas. Some are simply for amusement, some for working and supporting during multiple tasks. Examples for amusement would be Jacques de Vaucanson's mechanical flute player ([33]) or Waseda University's flutist robot WF-4RIV ([77]), shown in figure 2.2.

Both are playing a flute while on one hand the first is only mechanical and on the other hand the second one is also digitally programmable. Hereby, it is noticeable that controlling robots changed over time. In addition, it is recognisable that more autonomous robots were developed as the WF-4RIV can play together with a human musician and adapt its tempo.

[1]We know that R2D2's purpose may be more related to film industry and audience but we relate to his use in the storyline.

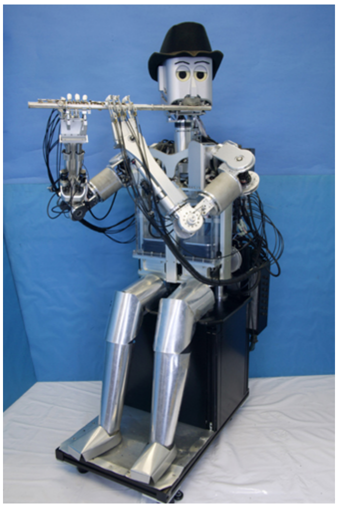


Figure 2.2.: The flutist robot WF-4RIV from TakanishiLaboratory.

Interactive robots can be distinguished through their ability to interact with their surroundings. This can include humans as well as other robots. For instance *Honda's ASIMO* can face people and recognise them as well as answer questions amongst other things ([2]). Here, no concrete surrounding is predefined and hypothetically the interaction could take place anywhere. All in all the development of robots changed from robots for only very specific tasks to more general tasks which are more dependent on context.

2.2. Emotions

Emotions play a big role in everyday life. The Oxford dictionary defines emotion as following: "A strong feeling deriving from one's circumstances, mood, or relationships with others" ([16]). So, emotions are reactions towards intrinsic and extrinsic stimuli. Reactions towards such stimuli are used to communicate our thoughts and feelings towards others. In particular reactions towards extrinsic stimuli are important during communication as they provide feedback to the conversation partner about one's feelings and thoughts during the conversation. So, hopefully the conversation partner understands those signals. Mehrabian's *7-38-55 rule* relates to this ([62]). How people understand spoken words is not only depending on their content but also the way they are communicated. An example would be distinguishing between being ironic and being serious. With only the content it is hard to decide. However, the tone of the voice or the speaker's gestures may express this more clearly. The way a person is expressing himself is also greatly determined by his emotions.

Multiple categorisations for emotions exist. One example is *Plutchik's emotion wheel* shown in figure 2.3. Plutchik uses 8 bipolar emotions, namely joy-sorrow, anger-fear, acceptance-disgust and surprise-expectancy. Similar emotions are close while opposite emotions are placed on opposing sides of the emotion wheel. Those are just the primary emotions while other emotions are a mix made up of the primary emotions. Hereby all primary emotions can be mixed not only neighbouring ones. For further distinguishing Plutchik uses a 3rd dimension to describe the intensity of emotions. The intensity of the emotions reaches its peak at the centre and gradually decreases in alignment with the distance to the centre. Russell uses a similar approach. Russels's circumplex model uses the two dimensions arousal, to determine activation or deactivation, and valence, to determine pleasure. Here we can recognise that some emotions are quite passive while others make us more active.

Emotions are an inner state but they manifest in external signs like shaping your voice or in facial expressions. Ekman grounds basic emotions amongst other things on facial expressions for being fast

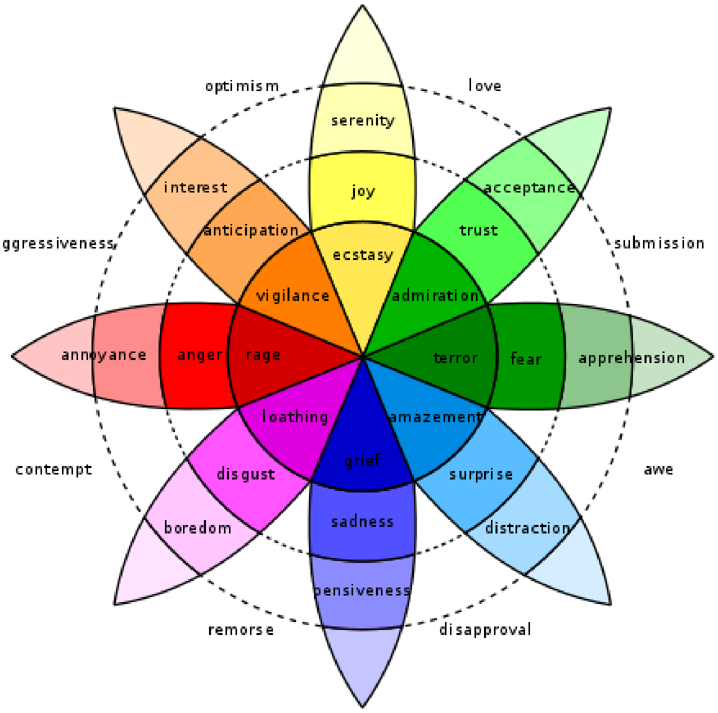
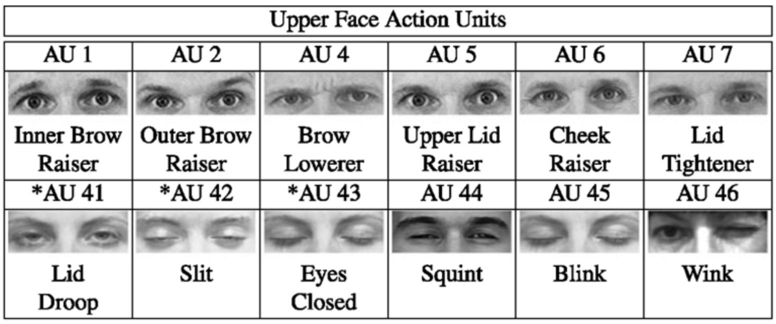


Figure 2.3.: Plutchik's wheel of emotions ([13])

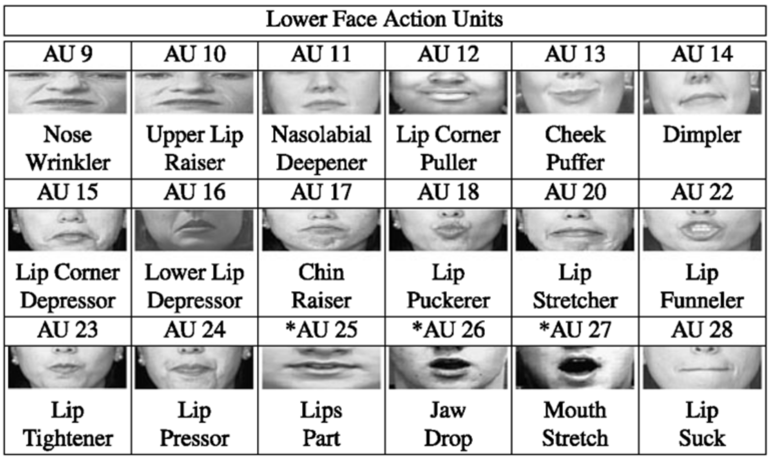
automatic reactions based on emotions as well as being universal. Ekman uses the 6 basic emotions anger, fear, disgust, happiness, sadness and surprise which are assumed to be innate. They can be classified and detected through the usage of the *Facial Action Coding System*, FACS ([37]). The movements and the corresponding intensity of facial muscles is categorised in *Action Units*, AUs. Simultaneously used AUs make up certain expressions which correspond to Ekman's basic emotions. Hereby AUs are mainly partitioned in *Upper Face Action Units* which are shown in image 2.4 and *Lower Face Action Units* which are shown in image 2.5. Upper Face AUs are focused on the eye area while Lower Face AUs focus on the mouth area. When searching for AUs it should be considered that the basic features of



Upper Face Action Units					
AU 1	AU 2	AU 4	AU 5	AU 6	AU 7
Inner Brow Raiser	Outer Brow Raiser	Brow Lowerer	Upper Lid Raiser	Cheek Raiser	Lid Tightener
*AU 41	*AU 42	*AU 43	AU 44	AU 45	AU 46
Lid Droop	Slit	Eyes Closed	Squint	Blink	Wink

Figure 2.4.: Upper Face AUs ([82])

human faces differ depending on ethnicity, sex, age and others. Such features also have influence on face detection. For example, when searching for skin colour it should be considered that skin colours differ



Lower Face Action Units					
AU 9	AU 10	AU 11	AU 12	AU 13	AU 14
Nose Wrinkler	Upper Lip Raiser	Nasolabial Deepener	Lip Corner Puller	Cheek Puffer	Dimpler
AU 15	AU 16	AU 17	AU 18	AU 20	AU 22
Lip Corner Depressor	Lower Lip Depressor	Chin Raiser	Lip Puckerer	Lip Stretcher	Lip Funneler
AU 23	AU 24	*AU 25	*AU 26	*AU 27	AU 28
Lip Tightener	Lip Pressor	Lips Part	Jaw Drop	Mouth Stretch	Lip Suck

Figure 2.5.: Upper Face AUs ([82])

and the algorithm should be robust regarding this. Another challenge is occlusion of one's face. This may be because of glasses or a beard as well as head rotation. In all cases some part of the face is covered which can restrict the results of face detection and especially of emotion recognition. For instance if you are wearing glasses it would be hard to determine whether you have closed eyes or not depending on the light's reflection. Another possible case would be when your head is rotated it may be harder to determine the momentary AUs of your face and thus your emotions. However, the face is not the only way of expressing emotions. They can also be expressed in gesture and posture amongst others. In figure

Figure 2.6.: Example of body language ([1])

2.6 some examples for body language through posture are depicted. Your posture can reveal that you are nervous or confident, that you may be hiding something or be genuine. It influences how other people see you and how they interpret your words and actions. Thus, it also influences their reactions and their reactions again influence yours.

All in all emotion recognition is important in our daily life and we are constantly using it. It is still difficult even for humans to always recognise them correctly. However, even in case of correctly recognised emotions they are not rigid but changing over time as they are based on a dynamical system reacting on feedback.

2.3. Robotics and emotion

A long time ago the philosopher Seneca already believed that control over one's emotions is the key in life ([4]). His message is to live according to cardinal virtues and withstand affects which might endanger such a kind of life. As a good life not only affects you but also the state and thus all people belonging to the state. The same goes for bad behaviour.

To decide whether behaviour is good or bad you first need some guidelines to define what is good and what is bad and in what kind of setting. This is something humans learn as they grow up taught by family, friends and society. Comparable robots get their behaviour from humans but in a different way. Either they can be hard programmed or they can be given the possibility to learn by themselves as well as mixes from those two variants. Their purpose has to be decided before other design processes can take place. In case of interaction a robot somehow has to interact with the other party.

Regarding the type of interaction multiple categories are possible. For instance textual, spoken, tactile or visual. Which ones should be chosen and how they should be used depends again on the goal and target user of the interaction. For example, when interacting with mainly unknown people we will have a greater distance towards them or use politer words. In case of known people we come closer and use more casual words. Even small interactions like greetings differ according to the situation and can be more or less formal. Then not only knowledge about the other person matters but also whether you are in a business meeting or taking a walk in a park. Robots do not know such specific details but this kind of information can be provided to them or can already be taken into account while designing the robot's behaviour.

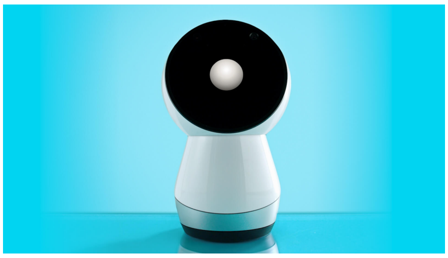


Figure 2.7.: Jibo - a personal robot for your home ([7])

The robot Jibo (figure 2.7) is a robot for your home ([8]). So typically its geographical location does not change much. It is learning how you look and who your friends are as well as gathering information on your preferences. It uses this information for interaction with you, for example when talking with you it knows who you are. Through interaction it is trying to establish a relationship to you and thereby shows emotions through posture as well as the shape of its eye or its voice. For us this is a very natural way hence it is easy to interpret.

It appears for social robots the establishment of a relationship is useful for further interaction and integration in daily life. Therefore it needs to know how to react and based on what it should react. As for humans it is easier to relate to behaviour similar to ours because we are used to it. Therefore, a robot's reactions could be inspired by human interaction und thus make the interaction and interpretation easier. However, the target user and environment should be considered during design phase.

2.4. Related work

Our work is focusing on emotion recognition based on facial expressions and on the feedback towards the recognised emotion in particular. This feedback should enable a system to react according to given emotional information about a user or conversation partner. Therefore, after extracting the emotional information contextual information has to be grasped to decide upon suitable feedback. The project *Romeo2 project*, described in [68], uses a similar approach. Upon collection of data from multiple sensors they extract information and use this for higher level tasks following the goal of providing an everyday aid in form of a robot. Unlike us they do not focus on the user only but on the whole environment.

Focusing on a human way of emotional feedback Cohen et al. have evaluated the performance of facial expressions as well as postures of a robot. They have developed facial expressions for anger, fear, happiness, sadness and surprise. Furthermore, they have done the same for postures. During experiments they have evaluated the effects of those facial expressions and postures. All of them were recognised quite well. In addition, they have concluded that the emotion sadness is better expressed through facial expressions than postures.

Kirby et al. focus on the interaction based on emotional feedback through facial expressions and the tilt of the head. Hereby, they model the robot's emotions, mood and attitude. The emotions and the attitude are results of dialogues with humans, while the mood results of the emotions experienced over the day. For evaluation they have implemented their model as a receptionist robot. Based on their evaluation they concluded that humans are able to detect even slight changes in the robot's facial expressions and to recognise its emotions. Besides, a human's interaction changed depending on the current shown mood of the robot. Thus, they conclude that their human-robot interaction is similar to human-human interaction.

Arkin et al. determine two needed aspects for long-term human-robot interaction. One is an ethnological part, which should make a robot's behaviour more predictable for humans. In case of an animal robot this means that in some way the robot should react as we would expect it from the animal it depicts. The other aspect is a motivational behaviour, which is comprehensible for humans. Both aspects serve establishing an emotional relationship between human and robot. Our work differs from this work as we do not want to perfectly mimic human behaviour but focus on the interaction with a certain goal, for example motivating the user. Further, Arkin et al. applied their assumption on a dog robot, AIBO. As motivational behaviour they chose emotions. In contrast to our approach instead of using the 6 basic emotions they chose to use a model with the three dimensions pleasant, arousal and confidence. Besides implementing part of the basic behaviour of a dog they have also added the possibility to learn and recognise new objects as well as the influence of the objects on the robot as the internal states can be changed by objects, for example hunger can be changed to feeling full. In a next step, they have transferred their idea to a humanoid robot, which can interact with humans in simple dialogues and remember faces as well as experienced emotions with specific people, which influenced the interaction. Other approaches have used the ethnological component, too.

Kidd et al. use a seal robot, Paro, for interaction with elderly people in care homes. Paro already offers some seal-like behaviour therefore the study focuses on the interaction part. Paro not only makes the people interact with it but furthermore with each other as they try to understand what it does and gives them a novel topic of interest, which seems worth to spend effort on. Still, human mentoring is needed to organise the subject's interaction time. Additionally, Wada et al. concluded that interaction with Paro helps elderly to become more energetic.

Fong et al. are providing an overview about socially interactive robots nowadays and how they are perceived by humans. They point out that personality and social interaction make a robot more attractive for interaction while it is more important that the robot provides what is expected from it. Humans are biased on their expectations based on previous experiences, Science Fiction or even the outer design of the robot. It makes them expect how to interact with robots and what they can expect regarding the robot's functions and reactions. Hereby feedback via facial expressions or movements, which show attention, seems to be well liked.

de Graaf and Allouch have analysed what kind of factors are important to make robots be liked. They have identified enjoyment as main motivation for using the robot, while the user's attitude towards the robot is influenced through usefulness amongst others. Hereby it is important to note that a social robot may not have a special goal but the engagement is its goal. In this case users focus more on other aspects like its adaptability to their needs. It was concluded that motivation for further interaction rises when humans start to bond with the robot. However, more research in this area is needed as relationships form over time and short-term studies are affected by the novelty of the used robots.

3. Emotional feedback

The basic idea is to first extract a human's emotions from an image or a video stream and afterwards to generate a social reaction of a robot. Hereby, the first step is extracting those emotions which can be done by self-implemented algorithms as well as by existing software. The next step is deciding on a suitable feedback reaction. However, suitability is context dependent and subjectively perceived. Therefore, we transfer our theories on a robot, Pepper in this case. More details about Pepper can be found in chapter 4.

3.1. Emotion extraction

For the emotion recognition we are using the Affectiva SDK (chapter 5.1). Besides the Affectiva SDK other possible services exist and also self-implementation is possible. We have decided to use this SDK for several reasons. First, it is free to use for research. Second, it is available for multiple operating systems and makes transferring our system less complicated. Further, no internet connection is needed for the usage of the SDK as it is for example when using Microsoft Azure. The SDK can be used either on the target platform or on another platform, which communicates with the target platform. Hereby the developer can decide on the way of communication and is not restricted but has to implement it himself or use another service. Besides the SDK some basic examples of how to use it are provided which makes integration easier. Also, its performance is convincing as Affectiva is faster than VGG'16 and has only about 1% less accuracy ([18]). Especially calculation time is important because we are planning on using our application in real time interaction.

A big disadvantage of the Affectiva SDK is that a classification for the absence of emotions, meaning *neutral*, is not yet provided in the developer version for Windows. However they seem to provide it for some versions so it may be provided for Windows in the future, too ([19]). To get the emotion neutral we are trying the approaches svm and thresholding. Our problem is defined as having 6 emotion scores and deciding whether one of those emotions or neutral is the main emotion. In general, we assume that the emotion having the highest confidence score is the current main emotion.

Besides the Affectiva SDK we have tried to train classifiers with IBM Watson[1] which seemed to be unsuccessful and was therefore stopped. We have used a free trial account and were therefore restricted in its usage (just 2 classifiers could each be trained once; number of samples was restricted). Thus, we cannot guarantee that our estimation of Watson not being suitable for emotion recognition is correct without such restrictions, too. Table 3.1 shows some possible services for emotion recognition based on facial images.

3.2. Behaviour

As described in chapter 1 the way one reacts towards expressed emotions of others is quite important and influences further interaction. Besides, perceived emotions and other features influence the reaction. For instance, if we know a person we will react differently than in the case of not knowing him. Therefore,

[1] https://www.ibm.com/watson/
[2] https://cloud.google.com/vision/?hl=en
[3] https://azure.microsoft.com/en-us/services/cognitive-services/emotion/
[4] https://www.ibm.com/watson/
[5] https://www.affectiva.com/
[6] https://www.kairos.com/
[7] https://aws.amazon.com/de/rekognition/
[8] http://doc.aldebaran.com/2-4/index_dev_guide.html

name	emotions	comments
Google Vision API[2]	Joy, sorrow, anger, surprise	REST-API; not only face detection; 1000 units per month for free
Microsoft Azure[3]	Anger, contempt, disgust, fear, happiness, neutral, sadness, surprise	API; 30000 transactions per month free, up to 20 per minute; face recognition possible; provides other APIs
Watson[4]	none but classifiers can be trained	API; limited test version; provides other APIs
Affectiva[5]	Anger, contempt, disgust, fear, happiness, sadness, surprise	SDK or API
Kairos[6]	Joy, ssurprise, sadness, anger, disgust, fear	API or SDK; limited free use of API (upt to1500 transactions per day and 25 per minute)
Amazon Rekognition[7]	Happiness, sadness, anger, confusion, disgust, surprise, calmness, unknown	API or SDK; provides further image analysis
Pepper (NAOqi)[8]	Neutral, happiness, anger, sadness	integrated; hard to test
Implement	various	needs time and data

Table 3.1.: Alternatives for emotion recognition based on facial expressions

we have identified some features, which seem to be useful for the decision process of how to react. Further, we have different ways of communication. Because those depend on the used platform we have partitioned them in humanoid and other. An important aspect is that the concurrently communicated verbal and non-verbal signals should not be contradicting but work together ([62], [85]). Additionally protection for human and robot like proposed in Asimov's laws should be ensured.

3.2.1. Humanoid

We will describe some possible features a humanoid robot could use to communicate its emotional feedback to a user. Not every robot will support these features. In this case an additional filtering will be needed. Still, we provide a basic guideline for possible interaction.

3.2.1.1. Gesture

As *gesture* is disclosing much about a human's current emotions ([45]) it makes sense to transfer this to a humanoid robot. Another advantage is that probably most humanoid robots will be able to use gestures as they are having hands or arms. Full body postures are another possibility. We will not cover this but suggest using [29] or Laban movement analysis ([28]) In addition, some gestures are culture-specific. Therefore, other gestures than suggested by us may be more suitable depending on the used environment.

Firstly, pointing gestures can be integrated in the dialogue to support basic dialogue (showing directions or pointing at things) as it seems to be naturally understandable and less culture-specific ([49]). Big movements might surprise the other person or accidentally hurt someone. Therefore, we suggest not using them unless environment and users are suitable. Static postures of anger, happiness and sadness are recognised similarly well as static facial expressions for those emotions ([30]). However, disgust is not very well recognised. In table 3.2 we suggest static postures for the upper part of the body to express the six basic emotions ([30], [80]). This includes the posture of the upper body, head, chest and elbows.

emotion	upper body	head	chest	elbow
anger	erect	backwards	forwards	backwards
disgust	collapsed	backwards	straight	straigth
fear	collapsed	backwards	forwards	backwards
happiness	erect	backwards	backwards	straight
sadness	collapsed	forward	forward	backwards
surprise	-	forward	straight	straight

Table 3.2.: Static postures based on[30] and [80]

3.2.1.2. Face

Similar to gestures the face is something naturally used by humans to convey emotional information. It may be even more natural than gestures and therefore easy to understand. However, not every humanoid robot may support facial feedback despite of having a face. In such cases, the expressions may also be transferred for example to an emoji on a screen.

In case the face of a robot is able to use expressions according to FACS we suggest using them for emotional encoding ([36], [37], [38]). Alternatively we suggest to orientate oneself by current research like [29]. These possibilities provide a way of emotional feedback humans are used to interpret and therefore need no further teaching. If the face is not expressive and a screen is available we suggest using emojis as they pick up typical features of emotional expressions. Thanks to the enormous amount of available emojis a huge variety of emotions and connotations can be expressed and understood like a human's expression ([47]). To decide what kind of emojis should be used it may be helpful to analyse used messenger programmes or to conduct studies. However, elderly seem to use less emojis or are less used to them than younger people ([23], [64]). Thus, emojis may currently not be suitable for elderly as the familiarness is not guaranteed. In this case, modelling a face or not using a face at all seems more suitable.

As we are using a Pepper, which is not facially expressive, we cannot use FACS. Furthermore, we target elderly and thus see emojis as not suitable because they may not be used to them. Another reason is the low height of the tablet. Pepper is a small robot and its tablet is attached to its chest. Looking there may be uncomfortable for taller people. Next, humans are used to look at the face so during a dialogue they will more likely look at Pepper's face than at its tablet.

3.2.1.3. Voice

emotion	pitch	tempo
anger	high, decreasing	depends on arousal
disgust	low	depends on arousal
fear	high	depends on arousal
happiness	high, increasing	depends on arousal
sadness	low	depends on arousal
surprise	high	depends on arousal

Table 3.3.: Suggestions for voice changes based on ([63] and [25])

The voice is something many living beings are using to transfer information. Voice itself is already using multiple channels. The most obvious one is the directly spoken message. For instance in case a close person is telling you about successfully finishing a project you may tell him that this makes you happy. However, depending on the other channels your voice is using he may or may not believe you. The shape of the voice, meaning pitch, intonation and speed, changes according to our body's current condition which is also influenced by our emotions ([63]). Therefore, in case of a very monotone voice the other person may not believe you. Of course, this does not only depend on your voice but other features, too.

Pitch is one of the main features of the voice for conveying emotions, e.g. the mean pitch or its change over time express emotions ([25], [63]). It does not only depend on the emotion but also on the intensity of the emotion. While anger decreases, pitch over time joy increases it over time. However, both as well as fear result in a higher pitch than usual. On the other hand, sadness results in a lower pitch. Further, high physical arousal results in a higher pitch while lower physical arousal results in a constant lower pitch. Thus, we suggest using a higher pitch for anger, fear, happiness and surprise and using a low pitch for disgust and sadness. Regarding arousal, we suggest increasing and decreasing the tempo of the voice. In our case, we are using a slower base tempo to make the robot easier to understand for elderly. Similarly, we suggest using higher volume if the user is hearing-impaired. Table 3.3 summarises our suggestions regarding voice adjustment. However, the pitch of Pepper's voice cannot be adjusted like this.

3.2.1.4. Clothes

Clothes, make up and accessories amongst other things determine how we perceive others ([54]). We have certain expectations how someone has to look, for example a businessman has to wear a suit, a doctor has to wear a white coat. On the other hand, we are using those expectations to guess about another person's personality. For instance if someone is wearing high quality clothes he may have to be successful and hard working. However if the other party is still quite young we would expect his parents to be the successful ones. All in all the visual appearance influences our perception of others. In addition, we are interacting differently according to our perception. In case of a doctor we would be more polite and formal but in case of a student more informal when interacting. Thus clothes (or in case of a non-humanoid robot its shape) are an important feature of the target platform for interaction which depends on the target usage scenarios and target users.

Goetz et al. conclude that the appearance of a robot influences people's perception of it. Further, it influences their expectations of it as well as the expectations of its role and its suitability. Therefore, a hospital's robot should wear a hospital's uniform to seem more trustworthy while it might not seem as suitable when wearing a hoodie for example. In our case of communicating with elderly at a care home a nurse's uniform might be suitable if he were to conduct some of the nurses tasks. However, the robot is not able to carry out complex medical tasks and is planned to be used for social interaction. Therefore, clothes that are more casual seem more suitable to show that the robot is not part of the staff. Especially in case of Pepper we think it is better to dress him as his usual white colour is very neutral and similar to medical staffs' clothes.

3.2.2. Other

We want to present some features, which are not necessarily only for a humanoid robot. Although some humanoid features may also be used for non-humanoid robots additional features are possible. Those features are intended to aid during the design process of possible interactions.

3.2.2.1. Colour

Similar to spoken language colour is a kind of language on its own. We associate colours with certain meanings though those associations can depend on culture and context. Colours can also be used to express emotions or simply to increase appeal and result in higher trust ([17], [32], [51]). Hereby it is important to note that colours do not simply have one meaning or are exclusively positive/negative. However, the colour yellow seems to be disliked although it can also express cheerfulness. In table 3.4 you can see some colours and their relating emotions. In our scenarios we have used the following colours: Yellow for joy, blue for sadness, white to calm down, green as peaceful and serious. Although yellow is not liked we decided to use it for joy because it stands out from the other colours and can symbolise cheerfulness. Further, Pepper was already programmed to use colours to show its attention.

colour	emotion
green	vivid, relaxing, boring, exciting, peaceful, sad, purity, serious, auspicious, inauspicious, depressive, cheerful
white	good, weak
red	strong, active, striking
black	strong, passive
grey	passive, weak
blue	cold, peaceful, dull
yellow	plain, sad, classical, dynamic, cheerful, weak
pink	warm, striking, romantic, enjoying, cheerful

Table 3.4.: Colour coding based on [61], [51] and [17]

3.2.2.2. Screen

As previously mentioned a screen can be used to display emojis and thus transfer emotions. It can also show written language and imitate the shape of a voice through textual shape. Hereby colours can be used to express certain emotions as well as size of the text or even text style. Another possible usage would be showing pictures which induce a certain kind of feeling. For instance, when showing idyllic places people might calm down.

How to use the screen seems to depend on the used platform, e.g. it does not make sense to show emojis if the robot can already show facial expressions on the face. Besides, it depends on the target user. For example, a screen may be more flexible about its position and can therefore be used to show things to smaller children or taller people. We suggest using a screen to give written feedback to further support the interaction. Hereby colour coding of emotions (3.2.2.1) as well as font and size can be used to transfer emotional information and make reading more comfortable ([26], [55], [74]). In regards to the font size the to be expected distance between robot and user should be considered.

In case of Pepper the screen is a tablet. This way it can not only be used to support the speech of the robot but also to enable the user to give input via the tablet. Especially in case the user is unsure about how to use Pepper and what to say this can be useful. In addition, the tablet can be used for games and thus engage people.

3.2.2.3. Distance

Depending on who we speak to we are standing closer or further apart ([48]). In general, you can say the closer we stand the closer we are emotionally. The same applies for the opposite. However, the exact distances depend on your culture and your upbringing as some cultures have much closer distances than others.

From distance you can extract how close a relationship is but in some cases you automatically come closer. For example, when telling a secret or other confidential things. Another case are actions which result in touching the other party. The used space is called *intimate distance* ([44]). *Personal distance* is not as close but still gives you the possibility to reach the other person. Next, *social distance* is more impersonal and provides the possibility to go off or even ignore the other person and further away is the *public distance*.

In the beginning we suggest for the robot to stay in the social or in the outer personal distance except people come closer on their own. Over time this distance can be adapted based on previous interactions showing how the different users feel comfortable. However, the robot should not back off if someone is coming closer as the other party knows what distance is comfortable for him.

3.2.2.4. Music and other sounds

So far we have covered the usage of the voice in general (3.2.1.3). As some non-speech vocalisations can transfer information, like screaming may show fear or happiness, the same can be done with other sounds. For instance those sounds can be music or sounds we are associating with certain things, e.g. the sound of the sea to calm one down and thereby influence one's emotions ([14], [50]). It can also directly mimic the human voice by adapting pitch and tempo. This way the typical changes of a voice possibly caused by emotions can be imitated.

Dependent on whether the robot is giving other audible feedback we propose different approaches. In case the robot is using other audible feedback we suggest using only short pieces of music. Thus, the music and other audible feedback do not interfere too much and both are still perceivable. In addition, the time needed for both is not unnecessarily long and thus avoiding long waits or robot and user speaking simultaneously. The later one could result in problems during speech recognition processes though it is possible to increase the accuracy of the recognition ([39], [60]). On the other hand if the robot is not using other audible feedback except for one it can use sounds over a longer period of time. What kind of specific audible feedback to use now depends on the interaction itself and what kind of robot is used. A musical robot will use music while other robots will use speech, music or other sounds. Even so, the volume of the robot should not trouble other people. We are using Pepper to speak with a single person. Hereby, we are not using speech and other sounds at the same time. Besides our implementation Pepper already gives some audio feedback. It is sometimes making short sounds, for example to show it has heard your voice.

3.2.3. Context

So far, we have only considered how to convey emotions through possible feedback ways. However, what kind of emotions should be conveyed has to be decided before using those. Therefore, the perceived emotion of the other party matters as well as other context of the situation. We have identified some interesting context.

3.2.3.1. Time

The current time is not a necessarily needed context. Nonetheless, it can have an impact on the interaction. For example, you would not be very happy to be disturbed during typical sleeping hours and react accordingly unhappy when being disturbed by a robot during those hours. On the other hand, if people disturb you during such hours it will probably be something important and you would react differently. In this case, time gives you information about importance because of breaking the daily rhythm.

We think the feature time is important to recognise abnormal states especially in the case of surveillance and thus alarming human personnel. In addition, some people have a bad mood in the morning or the afternoon. Hence, it is reasonable to track time and compare knowledge about the other person's behaviour and the current time. This way it is possible to detect abnormal occurrences in daily life. For example, some people are nearly always grumpy during early morning hours. In this case, we see grumpiness as the normal mood in the morning while it would be strange for another person which is usually not grumpy in the morning to be grumpy. If the behaviour differs from usual it makes sense to ask about the reason. If not it may be seen as usual. Next, certain times can be marked as times during which one does not want to be disturbed. For example, during sleeping hours or during lunch. Besides, the opposite can be done and certain times can be marked as times suitable for interactions.

3.2.3.2. Emotion change and stability over time

On one hand the emotion change of the interaction partner on the other hand the change of one's own emotions over the time are important. In general you should react to your partner's changes and show your attention which can be done by showing emotions amongst other reactions. Next, the stability of the emotion is important, meaning how long the same emotion is expressed as well as which emotions are shown in succession and over what time.

Based on your interaction partner's emotion change over time you can guess how your reactions are liked. For instance when liking your reaction the other party may smile or when being confused he may look surprised. Thus, emotions should be tracked over time and be compared. In addition, the emotion change during the interaction time can give information about the tendency the other party's mood is developing to. This could help for evaluating reactions and interaction strategies. In case of health surveillance the emotional stability or instability can be an indicator about your health, for example being bipolar or depressive ([31]).

3.2.3.3. Target user

Previously we have mentioned multiple times that the reactions of the robot should depend on the target user. This helps restricting the field of possible feedback as well as the specific design of continuous feedback. The reason for this is that every feedback is used to fulfill a certain goal which is dependent on the target user and possible use cases.

Wrede et al. studied the interaction with impaired people and noticed that they try to conceal their impairment. Further, they observed that the nursing staff is trying to detect such cases and provide clarification if needed. However, it was also noticed that clarification is preferred in cases of a task oriented dialogue while concealment is preferred in cases of a social interaction.

As target user we have chosen elderly in a care home. Thus, our use cases are inspired by scenarios possible in a care home. During the design we have thought of elderly as interaction partners. This also inspires possible features for a user profile.

3.2.3.4. User profile

As humans differentiate between unknown and known people and even in the degree of knowing interaction also changes. If we know a person is generally angry or easy to irritate maybe we will not always ask for the reason but simply accept his personality. Especially when one's personality conflicts with the aim of the interaction it is important to keep track of that and adapt feedback strategies. In case of the angry person it may be plausible to see him as calm despite still looking angry after some time and not endlessly trying to calm him down. The same goes for constant misclassification of the user's emotions. As a result the system becomes more robust. Besides, knowing someone can supply topics for conversations and possible use cases.

When asking a care taking staff of elderly in an interview we got the feedback that elderly like it when someone remembers their birthday and congratulates them. Therefore we think knowing the birthday of interaction partners is important and the robot should congratulate them. Additionally we were informed that elderly like to be reminded of their past, e.g. through playing some music which was popular during their youth. Their birthday helps concluding from which time the music should be chosen. When elderly are reminded of their past they may start talking about their memories. In this case it is important to listen and sometimes communicate your attention. Further you can learn about topics a person likes or troubles him and adapt future interactions.

Tracking peculiarities could further support interactions. For example typical morning mood and times one does not want to be disturbed or not liking physical closeness. Those can be integrated into the interactions and make each interaction more personal. The same goes for illnesses which need extra care so the robot can adapt its interaction and ease the medical staffs workload by surveillance.

3.2.4. Robot's emotions

As we have identified possible ways of feedback a robot can use as well as things that should be considered we now want to take the robot's emotions into account. This includes deciding which emotion a robot should display as well as how this emotion should change over time. The first step is deciding on the goal of interaction. In our case it is to make the interaction enjoyable and motivate the user to engage more often. Thus, the interaction needs to be interesting and variable. In addition, we want the user to become happy during interaction. This gives the direction for emotional changes of the robot. Some basic ideas were integrated into use cases.

3.2.5. Use cases

Our basic scenario is a Pepper robot residing in a care home together with elderly people. The interaction between Pepper and the elderly should be as natural and as social as possible and therefore entertain during everyday life. Some basic interactions would be greeting and saying good bye. Though, those can already vary depending on the user's current mood. For instance you can just greet a grumpy person and then back off or further engage. In both cases the other person can react in various ways. In the following we are depicting some basic use cases which could occur during interactions. A rough overview is given in diagram 3.1.

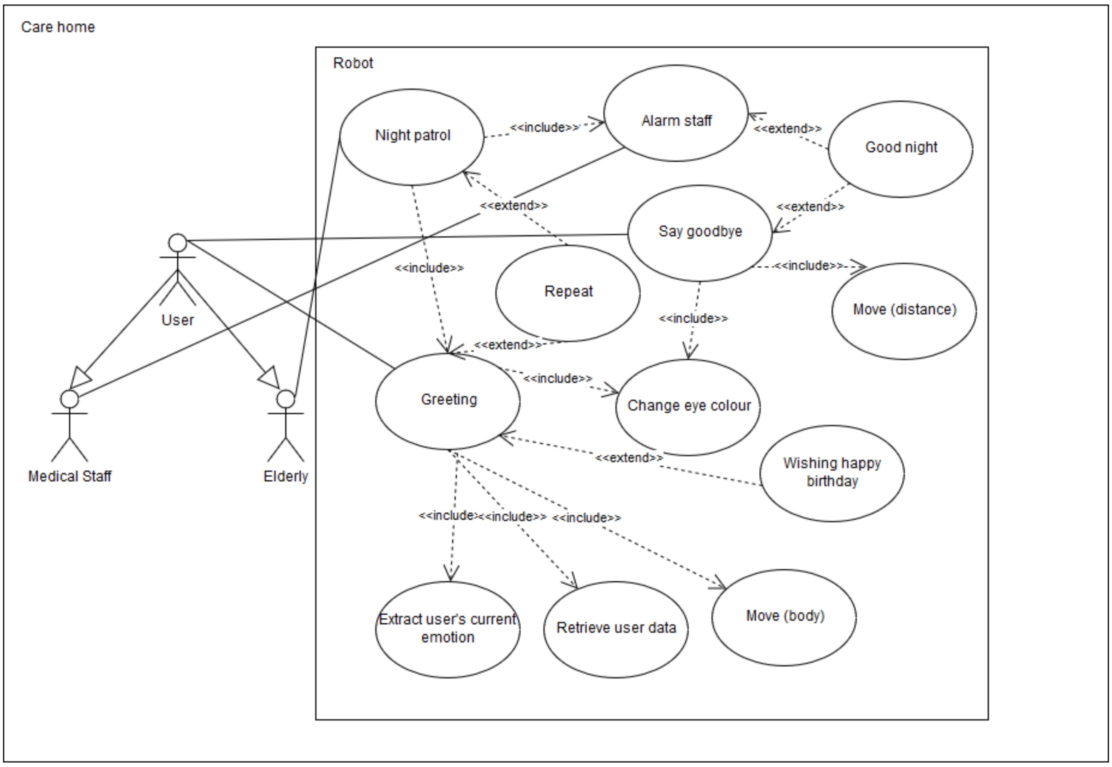


Figure 3.1.: Use case diagram for a robot interacting at a care home. It contains some basic use cases but not all.

3.2.5.1. Use case 1: Greeting a happy person

Pepper is not engaged in an interaction. A person comes closer to it and it starts looking at the person. When being close to Pepper its eyes blink once and it greets the person using the person's name. Pepper is waiting while the person is smiling. After 5 seconds it says "You are looking very happy. Has something good happened today?". The person keeps smiling and answers Pepper's question. Pepper exclaims joyfully and raises its arms "How nice!". Its eyes turn yellow.

3.2.5.2. Use case 2: Chased off

A user is looking angrily at Pepper and shouts "Go away" Don't look at me!". Pepper's eyes turn blue and says "That's not nice but ok". He turns around and moves a little bit.

3.2.5.3. Use case 3: Birthday

Pepper is not engaged in an interaction. A person is passing. Pepper detects the person and says "Happy birthday <person's name>!" while moving its arms. It starts singing and shows balloons on its tablet. The person thanks Pepper and Pepper asks what the person is going to do today. The next time Pepper sees the same person on this day it is not congratulating again.

3.2.5.4. Use case 4: Night patrol

Pepper is standing alone in a dark hall. It is 2 a.m. in the night. A person is approaching and Pepper looks at him. Its eyes turn white. The person looks confused. Pepper is greeting the other person and asks whether it can help or not. After some minutes medical staff is coming and gets the confused person.

3.2.5.5. Use case 5: Sad user

Pepper is engaged with a person. It asks how the person feels and gets the answer sad. Pepper reacts by turning its eyes blue and its body ,slumping down a little bit. It asks "Can I help you somehow?" and changes the eye colour to green. After confirming it offers to listen to the person. Afterwards it offers a hug.

3.2.5.6. Use case 6: Good night

Pepper is visiting a person's room in the evening. The person wants to sleep now. With green eyes Pepper is saying "Ok. then I'll be leaving. Sleep well and see you tomorrow!". Pepper leaves the room. It is visiting some other rooms wishing the people living there a nice rest.

3.2.5.7. Use case 7: Confused person during interaction

Pepper is engaged with another person. It is explaining the rules of a game. The other person suddenly looks confused. Pepper is asking whether it should explain the previous step again. The person agrees and Pepper explains the last step again.

3.2.5.8. Use case 8: Pepper being bored

Pepper is not engaged in an interaction. He starts playing some soft music and videos. After some time it stops and starts driving in circles for some time.

In case of Pepper it has already been programmed to randomly perform some actions when being bored. For this the autonomous mode has to be activated.

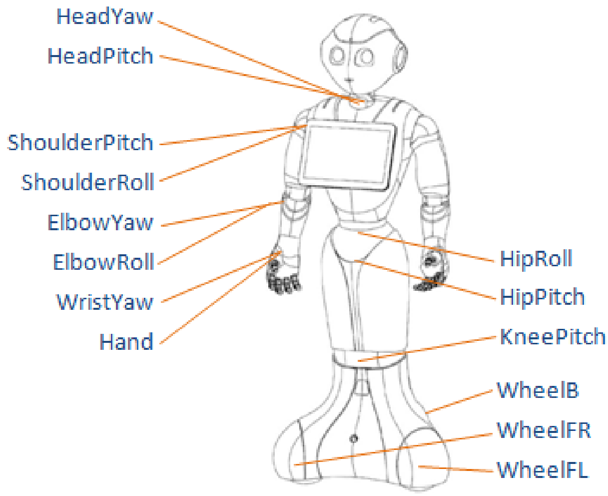
4. Pepper - A social robot

Especially in present-day Japan you can see Pepper frequently. A small childlike robot with big eyes welcoming you at a bank or even going to school ([43]). In other countries it is used as well like in Belgium to greet patients or bring them to specific departments in hospitals ([67]).

4.1. Overview

Pepper was developed by Aldebaran and SoftBank. One of their development goals is natural interaction with humans ([76]), which may be one of the reasons for its design and multiple sensorial equipment. Figure 4.1(a) shows Pepper. Table 4.1 presents a rough overview of it and its equipment. On its head it

(a) An image of Pepper ([75])



(b) A depiction of the motors used to move Pepper ([9])

Figure 4.1.: The robot Pepper

has 4 directional microphones which enable it to locate your position by your voice. Further, the tone of your voice can be analysed to evaluate your emotions. Besides, it is using cameras (3D camera, 2 HD cameras) to identify your emotions by image, meaning using your facial expressions as well as your head orientation. The robot is able to detect joy, sadness, anger and surprise. In addition, it is able to get to know you and change its behaviour accordingly. Another important feature for interaction is the tablet which is positioned on Pepper's chest. It can be used to show the spoken text or provide games among other things. For tactile interaction the robot has touch sensitive areas on its head and on the back of its hands ([46]). Further, Pepper is equipped with sonars and lasers to detect obstacles around it. The status LEDs on Pepper's shoulders show its current status (nothing, information, warning, error

Height	1,21 m
Status LED	2 on shoulders; RGB; nothing, information, warning, error unusable
Eyes LED	2 on the head; RGB
Ears LED	2 on the head; 16 blue levels
hline 2D cameras	2 in the forehead
3D sensor	1 in the forehead
Lasers	3 near legs
Infra-Red	2 near legs
Sonars	2 near legs
Microphones	4 on the head
Loudspeakers	2 on the head
Tactile sensors	1 on the head, 1 on each hand

Table 4.1.: Overview of Pepper based on [46] and [9]

unusable) followed by a voice message. Aside from Pepper's provided autonomous life additional apps can be installed and its behaviour can be programmed.

For programming Pepper, an SDK and tutorials[1][2] are provided. This SDK already provides an API with useful basic methods and a GUI which enables people with less experience of programming to still programme Pepper. Further methods can be written as well as imported. In addition, applications for Pepper can be downloaded from as well as uploaded and offered to the community. To download free apps from the store you have to be a subscriber of the Pepper Basic Plan ([46]).

To start Pepper you have to press the button behind its tablet once ([46]). It is ready for use after saying "ognak gnuk". For shutting down press the button 4 seconds and Pepper will say "gnuk gnuk" after finishing. To upload and execute programmed behaviour you have to establish a connection to Pepper. Hereby, Ethernet and Wi-Fi are usable.

After shortly describing Pepper's basic features we want to present our reasoning for using Pepper in social interaction. According to Goetz et al. humans prefer more human-like robots for human jobs. Usually social interaction is a process between humans. Therefore, we would classify social interaction as a human job and suggest the usage of a humanoid robot. Another reason for using Pepper is its childlike appearance, which makes it seem harmless and thus easier acceptable in daily life. Although Pepper itself is already often seen in Japan in particular and hence more accepted this may not be true for other countries ([43]). However, it shows the possibility for further acceptance as well as the possibility to use Pepper for projects and studies taking place in Japan. Next, Pepper is equipped with multiple sensors and actuators which enable collecting contextual information as well as actions done by Pepper. Combined with the provided functions by the SDK Pepper is easily usable for interactions.

4.2. Pepper's feedback possibilities

Previously we have depicted some feedback possibilities as well as an overview about Pepper. Now, we want to describe some possible feedback Pepper can do.

- Eyes: predefined (pink (detects you); spinning blue (listening to you), green (understood you)); further usage (colour coding of emotions during interaction)
- Voice: Speed, voice shaping, to be spoken text, language
- Music: Beforehand defined music can be played

[1] https://www.stsa.net.au/single-post/2017/04/19/Pepper-Tutorial-1-SDK-installation-Application-developmentexecution
[2] file:///C:/Program%20Files%20(x86)/Aldebaran/Choregraphe%20Suite%202.5/share/doc/index_dev_guide.html

- Tablet: The tablet can show videos or images and can be used to show output as well as for asking for input or playing games
- Posture and movements: Upper body (neck, shoulders, elbows, arms, hands) and lower body (waist, knees) can be used to depict provided postures as well as to generate new ones
- Distance: Pepper can move by itself as specified (distance per direction); engagement zone definition is provided
- Face learning: Pepper can learn faces and recognise them. Can be used for user specific interaction. However, it cannot learn names but needs manual input
- Body: The body is white but can be dressed with a jacket

Some of our suggestions are static, e.g. clothes, others are variable, e.g. colour coding. We would suggest first deciding on which static features to use and how and afterwards on variable features. The variable features can be basic ones at first and later on they can be upgraded or changed.

5. Tools and implementation

After describing our theories we now focus on the realisation. This includes tools and dependencies as well as our own implementation. First, we will describe the used tools. Afterwards follow details about the implementation using the Affectiva SDK and Choregraphe.

5.1. Tools

In this section we describe the used tools, their dependencies and their integration for *Windows 10*.

5.1.1. Boost 1.65.1

Boost provides free libraries for C++ ([5]). Some of those are already included in the C++17 Standard. We are using *boost 1.65.1* to manage directories. Especially during the evaluation of the Affectiva SDK we are using boost to coordinate the different image databases.

5.1.1.1. Building and linking

We have downloaded an executable for Windows which installs boost 1.65.1. Hereby some libraries can be used afterwards while some need to be compiled. For this purpose you need to run *bootstrap.bat*. Afterwards we need to link Boost to Visual Studio. This includes adding the path of the boost root directory to project settings in Visual Studio under *C++/General* as well as adding the path to *lib64-msvc-14.1* under *Linker/General*.

5.1.2. Opencv 3.3.0

Opencv is a free available library for computer vision and machine learning, hence it provides various image manipulations ([10]). It is provided for multiple programming languages, C, C++, Java, Python and MATLAB. In addition, it supports the operating systems Windows, Linux, Android and Mac OS.

The Affectiva SDK is using opencv_ffmpeg.dll which is provided after installation. Further, we are using Opencv 3.3.0 to load images during evaluation as well as for training and using a linear svm.

5.1.2.1. Linking

First, we downloaded an executable which was used for the installation of OpenCV on Windows. Next, we added the path of the root directory of OpenCV to the system variables, which makes the linking in Visual Studio more flexible regarding changes of the root directory. For the linking we added *$(OPENCV_DIR)/include* under project settings *C++/General* and *$(OPENCV_DIR)/x64/vc14/lib* in *Linker/General*. Further, we added the dynamic libraries *opencv_world330.lib* and *opencv_world330d.lib* in *Linker/Input*.

5.1.3. AffdexSDK 3.4.1

The Affectiva SDK is an SDK for emotion recognition based on facial expressions. We are using the current 64 Bit version of the Affectiva SDK for developers under Windows which is version *3.4.1*. An online guide for first steps is provided ([21]). Besides Windows other Affectiva SDKs exist for Android, iOS/macOS, Linux, Unity and JavaScript. A detailed description about provided classes and functions can be found here [15]. As far as we have experienced multiple image formats are supported, jpg, tif and png.

The Affectiva SDK provides classes to analyse spontaneous facial expressions in images, videos or streams. During the process algorithms search for landmarks, for example the corners of the eyebrows ([20]). Next, classifiers are used on the areas surrounding those landmarks to determine facial expressions and AUs. Based on FACS those are mapped to emotions.

Besides the Affectiva SDK other services provide emotion recognition on images (azure, kairos, Google landmark extraction, amazon, Pepper, self-made). One major benefit of Affectiva is not needing a connection to a cloud but providing the possibility to integrate the processing into the used device. So you do not have to care about internet connection and can integrate the processing conveniently into your programme. In addition it is independent from the to be used device. Another advantage is the performance which is still being tried to be enhanced as well as the possibility of real time usage ([18]).

5.1.3.1. Requirements and dependencies

Affectiva recommends the usage of a processor with 2 GHz, 1GB RAM and minimal disk space 950MB. Runtime requirements are Visual C++ Redistributable runtime for VS 2013 and Microsoft .NET framework v 4.0 which are installed automatically during installation of the Affectiva SDK. The needed operation system is Windows 7 or higher. For tracking multiple faces they recommend using one thread per to be tracked face as else real-time usage may not be possible.

5.1.3.2. Embedding in Visual Studio 2017

For development we are using *Visual Studio 2017*. After installation of the Affectiva SDK we have to configure Visual Studio for using the SDK. According to [21] the project configuration has to be set to *release* and *x64*. Further we have to add the libraries and includes. In *C/C++/General* we have to add the include directory of the Affectiva SDK. In *Linker/General* we have to add the lib directory (choose release) as well as in *Linker/Input* we have to add the lib and copy the dll (native) into our build directory.

5.1.4. Rapidjson

RapidJSON is a open source JSON parser and generator for C++ ([11]), which does not use any external dependencies. In addition, it works on multiple platforms (Windows, Linux, Android, Mac OS) and is conform to RFC7159/ECMA-404. We are using RapidJSON to format and to save the results of the image analysis with the Affectiva SDK.

5.1.4.1. Linking

After extracting the downloaded zip[1] we have to link RapidJSON to our Visual Studio project. Hereby we add the path *<RapidJSON root>/include/rapidjson* to the project settings under *C++/General*.

5.1.5. Choregraphe 2.5.5

Choregraphe is an SDK for programming the robots Pepper, Nao and Romeo. We are using version 2.5.5 to programme a Pepper. It is provided by softbankrobotics and can be downloaded after creating an account. Further, pynaoqi-python2.7-2.5.5.5 has to be installed.

[1] https://github.com/Tencent/rapidjson/blob/master/license.txt

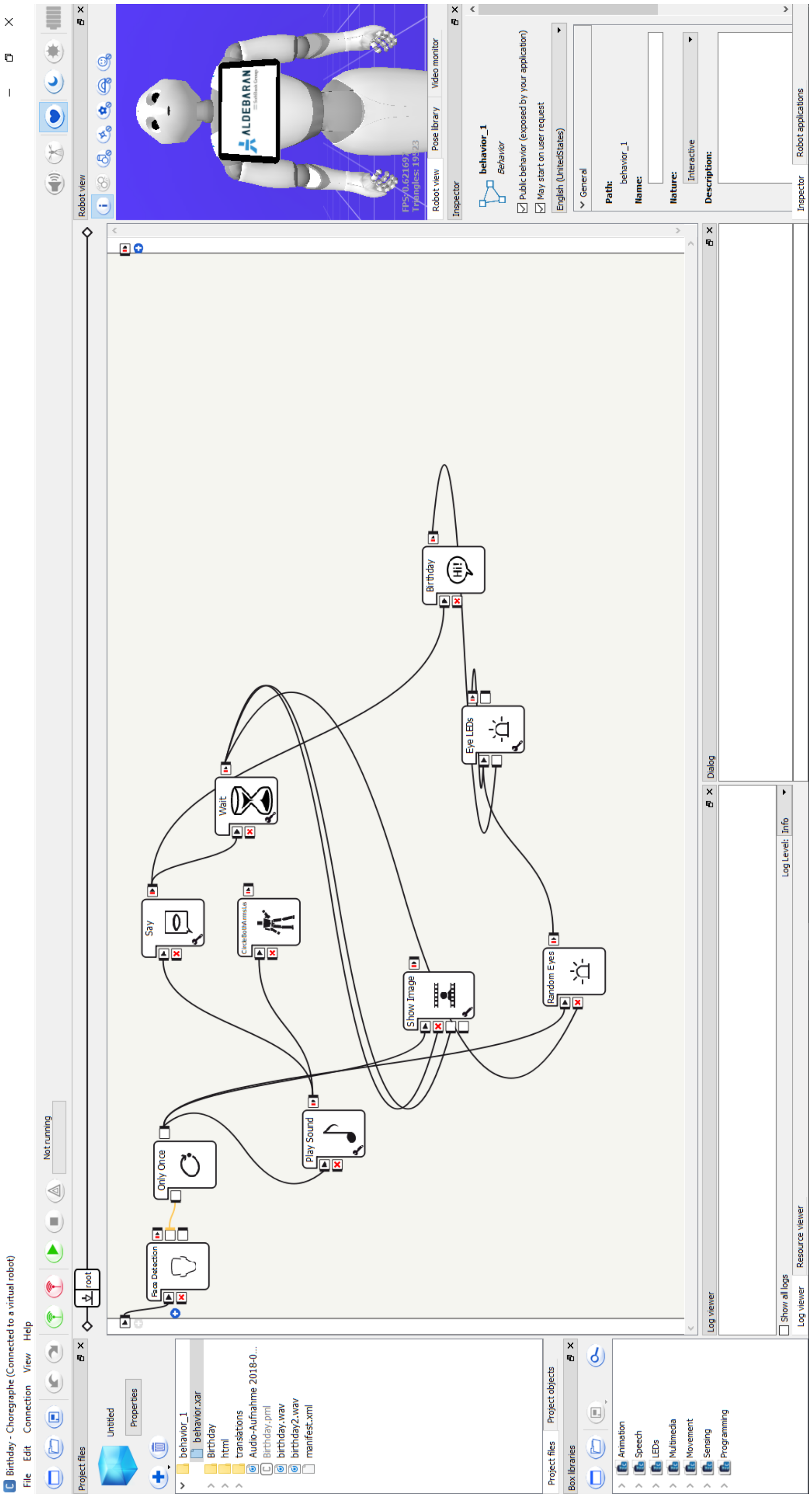


Figure 5.1.: The GUI of Choregraphe showing a simple programme.

Figure 5.1 depicts Choregraphe and an application programmed with it. For testing your code and some functions you can connect to a virtual robot. In the window on the right you can see this virtual robot and its movements if specified. The big window in the middle shows the programmed behaviour built up of boxes and their connections. On the left you can manage the files of the current project as well as choose provided boxes for it. Further management of your project is possible through the options in the top left corner. There you also find the connectivity options.

5.1.5.1. Troubleshooting

Choregraphe is sometimes not able to connect to Pepper. After restarting Pepper and the computer it should work again.

Sometimes Choregraphe is crashing. Restarting solves this problem.

Choregraphe starts processes for each uploading but does not stop them, which unnecessarily increases the workload of your CPU. They have to be stopped manually.

Sometimes the connection between computer and Pepper gets lost. Depending on the situation different solutions exist. First, try reconnecting. If this does not work close Choregraphe and restart it. Try connecting again. In case this is still not working restart Pepper and try again.

5.2. System for emotion recognition

After describing the used tools we now want to give an overview of our system for emotion recognition and how the tools were embedded. First, we give a general overview. Thereafter, we go into detail. Next, we shortly describe programming Pepper.

5.2.1. Requirements

Regarding requirements for the emotion extraction the following were given beforehand:

1. Accept images or videos as input
2. Extract information about a person's emotions based on his facial expressions
3. Return information about a person's emotions in a suitable way
4. Decide on a main emotion
5. Real time usage
6. Provide a service to a server

The requirements 1, 2 and 5 are satisfied through the usage of Affectiva.

For requirement 3 we have chosen to return the extracted information in JSON format one time in a terminal and one time written in a file. Our reasoning is that both humans and computers can understand formatted JSON data and therefore process further. In addition, the library we are using is supporting multiple operating systems and thus unlocking the potential of using it in different environments.

Details about requirement 4 are shown in 5.2.3.

Currently, our system is not providing a service towards a server and thus not fulfilling requirement 6. It is extracting data out of images put inside its image directories and returning the data JSON formatted in a terminal as well as written in a file.

5.2.2. Integrating Affectiva

We are using the Affectiva SDK 3.4.1 for emotion recognition from faces. Hereby we are concentrating on images though videos or streams are supported by Affectiva as well. For each input type another class is provided. *PhotoDetector.h* is used for photos/images, *VideoDetector.h* for videos, *CameraDetector.h* for camera streams and *FrameDetector.h* for frame streams. Currently we are using *PhotoDetector.h*. However, the methods and usage of those classes vary a bit resulting in the need of an adapter to change between the different input possibilities. Our adapter is the virtual class *Processor* and its subclasses shown in 5.2. Additional subclasses need to be added if other input methods are preferred. Also, *CameraProcessor* is currently only a mockup for the future.

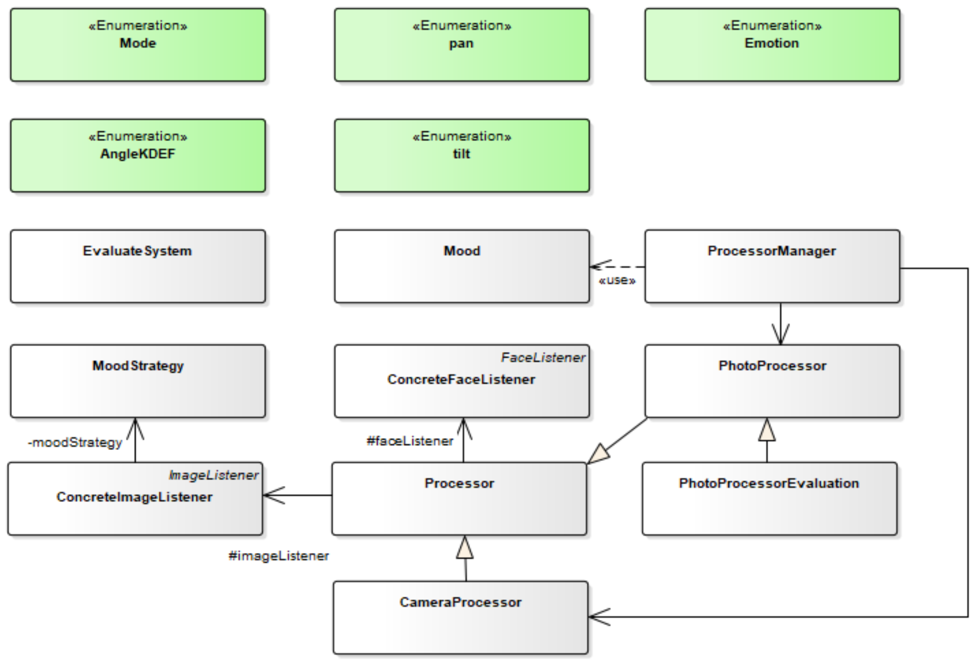


Figure 5.2.: An overview of our implementation.

On one hand, we are evaluating Affectiva's performance meaning comparing results and labels. On the other hand, we want to integrate it simply for emotion recognition meaning only extracting the results. Therefore, we are using inheritance to separate the functionality for testing in *PhotoProcessorEvaluation* from the general emotion extraction in class *PhotoProcessor*.

Regarding the delivery of analytical results the *Detector classes* of Affectiva use callback functions from the interfaces *FaceListener.h* (when a face is found or lost) and *ImageListener* (when an image is analysed). We have implemented them as *ConcreteFaceListener* and *ConcreteImageListener*.

Our goal is to get emotional information per image. Hereby, we are using the callback function *ConcreteImageListener::onImageResults(std::map< FaceId, Face > faces, Frame image)* and generate a JSON object out of the information. Which information you want Affectiva to extract can be adjusted in an instance of one of the *Detector classes*. You can also disable sending analytics to Affectiva.

As a controller to manage the different types of detectors as well as a facade to provide emotion detection for different inputs we are using the class *ProcessorManager*. In addition, it contains convenience methods for evaluating the performance on labelled images. If further segregation is preferred it ist possible to declare another subclass containing needed methods for evaluation. How to use the extracted information of Affectiva is defined in *MoodStrategy* which encapsulates needed functions like the design pattern visitor.

5.2.3. Main emotion

After the emotion analysis the emotion which should influence the behaviour of the robot has to be decided. Further, this emotion can be used for the evaluation of the performance of the Affectiva SDK in case of given labelled input. In general you would say that an emotion with a higher confidence is much more pronounced than one with a lower score and thus the dominant emotion. Therefore our first approach is using the emotion which possesses the highest confidence score.

```
1 Emotion useMax(float anger, float disgust, float fear, float happy,
      float neutral, float sad, float surprised)
2 {
3 Emotion curEmotion = UNKNOWN;
4 if (anger >= max(surprised, neutral, sad, fear, happy, disgust))
5 {
6 curEmotion = ANGER;
7 }
8 if (disgust >= max(surprised, anger, fear, happy, neutral, sad))
9 {
10 curEmotion = DISGUST;
11 }
12 if (fear >= max(surprised, disgust, anger, happy, neutral, sad))
13 {
14 curEmotion = FEAR;
15 }
16 if (neutral >= max(surprised, disgust, fear, happy, anger, sad))
17 {
18 curEmotion = NEUTRAL;
19 }
20 if (sad >= max(surprised, disgust, fear, happy, neutral, anger))
21 {
22 curEmotion = SAD;
23 }
24 if (surprised >= max(anger, disgust, fear, happy, neutral, sad))
25 {
26 curEmotion = SURPRISED;
27 }
28 //happy has best performance -> use it if it has the highest score
29 if (happy >= max(surprised, disgust, fear, anger, neutral, sad))
30 {
31 curEmotion = HAPPY;
32 }
33 return curEmotion;
34 }
```

Listing 5.1: Getting the main emotion based on its score; a higher score equals higher incidence

Listing 5.1 describes the basic algorithm for getting the emotion with the highest score. It accepts 7 parameters of the type float one for each of the 6 basic emotions and neutral. In case no score is given for one of them 0.0 can be passed as parameter value. Basically, we are checking which value is the highest and compare one emotion score with all other scores. Hereby, we are prioritising via the order of checking. If some emotions have the same score the one which is checked later is preferred. We have ordered the checks according to which are easier categorised correctly. The emotion which is seen as the main emotion after all checks is returned. Especially happiness has a high chance of being correct if possessing the highest score and thus is checked the latest (lines 29-32).

The algorithms to get the main emotion are located in *MoodStrategy* which is used by *ConcreteImage-*

Listener. Our reasoning is the callback method in *ConcreteImageListener* returning information about processed images. When returning this information it can be processed further if needed before writing it in a JSON object. So, the further processing needs to be in between getting the first results and passing those results.

Besides using the highest score to determine the main emotion other methods might be suitable as well. Those can be implemented and changed in *MoodStrategy*. So far we have implemented one other method using the emotion with the lowest score. We base this on cases where one wants to conceal his emotions. The performance in such a case still has to be tested. Another reason is that it could be interesting which emotion is the least probable. For example, when adding further thresholding to decide the main emotion.

5.2.4. SVM for neutral

5.2.4.1. Training

As previously described Affectiva does not provide a score for being neutral. You can say neutral equates the absence of any other emotions and hence a decision based on the scores of all emotions. Thus, we decided to train a linear svm for the decision of a face being neutral or not.

For training an svm for the decision of being neutral we have generated a database, which consists of the images in the mixed dataset (described in 7.1.5), meaning *all images* from the *CK+* dataset and *all frontal images* of the *KDEF* dataset. One part of this dataset is used for training and one for testing the svm so both parts are disjunct. Those parts are randomly chosen. We train multiple svms with a different amount of data ranging from 90% to 40%. In total we train 6 svms (using 90, 80, 70, 60, 50 or 40%). The reason for the minimal percentage of used data is simply that otherwise the minimal amount of required data for training is not achieved. Furthermore only images where a face is found by Affectiva are used for training or evaluation. This way we can ensure having the correct needed emotion scores for training and not falsifying the evaluation. The svm itself only decides whether a face is neutral or not.

We are using a linear svm with the two classes *neutral* or *not neutral*. Our reasoning behind this is as following. Since Affectiva already provides the 6 basic emotions we are interested in there is no further reason to explicitly train the recognition of those. Hence, we are using Affectiva to generate the scores for the 6 basic emotions and afterwards decide by using our svm if the main emotion is neutral or not. If not one of the 6 basic emotions is the main emotion. The result after usage of the svm is either 1 or -1 for neutral or not neutral.

The training of the svm takes place in class *Mood*. Listing 5.2 depicts the method used for the training.

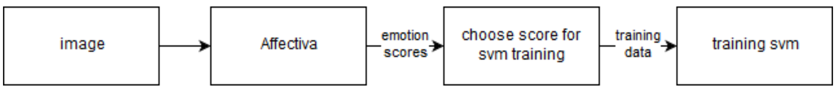


Figure 5.3.: The process for training a linear svm.

The accepted parameters contain the emotion scores, their corresponding labels and the name of the to be trained svm. Scores and labels are formatted for further use (lines 4-6). Regarding the training we are using the auto train method (line 10).

As input for the svm we are using the scores of all 6 basic emotions. We decided to use all scores because we want to detect the absence of emotions. This case seems more likely if all emotions have low scores and not just one emotion. Therefore, we either have to look at all scores or at the maximum of all scores which can be implying low scores for the other emotions. We chose all scores as there may also be a connection to the distribution of the scores speaking if all scores are very similar or if the variation is bigger. If all scores are below a certain threshold this could be seen as the non-existence of all emotions. However, at the same time all scores should be quite similar low as the existence or absence should be equally probable. For comparison we have trained svms one time with all 6 scores as input and one time with only the highest score as input. The results can be seen in section 7.2.

```cpp
void trainThresholdNeutralAll(std::vector<std::vector<float>> trainingData, std::vector<int> trainingLabel, std::string name)
{
//format data for training
cv::Mat label(trainingLabel.size(), 1, CV_32S);
makeLabel(trainingLabel, &label);
cv::Mat data = makeData(trainingData);

cv::Ptr<cv::ml::TrainData> td = cv::ml::TrainData::create(data, cv::ml::ROW_SAMPLE, label);
td->shuffleTrainTest();
svm->trainAuto(td);
}
```

Listing 5.2: Shows training data of a svm in class *Mood*

For the implementation we are using the svm class provided by Opencv3.3.0 and programme with Visual Studio 2017 in C++.

To separate the training of the svm from the rest the methods for it are put in the class *Mood*. As the training is dependent on input data *ProcessorManager* acts as a mediator and transfers the data.

5.2.4.2. Integration

Except for the need of input data the training of the svms is independent of the Affectiva SDK. Therefore, we have put it in a single class, *Mood*, which does not directly interact with the other classes.

After Affectiva extracted the emotion scores we use *PhotoProcessorEvaluation* to pass them via *ProcessorManager* to *Mood*. Since we are training multiple svms with different portions of the data but still need the results of all data we are processing all data at the beginning and keep them during the training instead of processing only the currently needed data. By using those scores and the image labels we generate the training data for the svm. Hereby first all images are processed and *ProcessorManager* gets results and labels. Next, n different random numbers are generated to decide which data is going to be used for training. The number n depends on the amount of data to be used from the total number of data.

5.2.4.3. Evaluation

Before training a svm the data is divided into training and testing data. Further only the training data is given to *Mood*. After finishing the training and saving a svm its performance is evaluated. Therefore, the testing data is used.

To get the path of an image we iterate over the directories containing the images until the index is the searched one. This image is evaluated by using the svm and the maximum score as main emotion. However, the result cannot only be neutral and not neutral but neutral or one of the 6 basic emotions. Results can be seen in section 7.2.

5.2.5. Threshold

Learning about absence or presence of emotions implies thresholding. On one hand if all emotions are absent all of them should have low scores. On the other hand if certain emotions have a low score it could already mean neutral. An underlying reason could be that some emotions can be detected with higher accuracy than others. So, you can already say if the score is not high, this emotion is absent while other emotions are possible but could also be absent. In this case we need to identify towards which emotions neutral often defaults and find thresholds for those emotions. In an optimal case the accuracy of the thresholded emotion will not suffer greatly as the emotion needs to have a high score to be identified as main emotion. This would object a low threshold.

In section 7.2 we could identify emotions which are taken as main emotion instead of being neutral. We have tried multiple values for thresholding. Integrating a threshold can be done as shown in listing 5.3.

```
1 if (evaluateNeutral)
2 {
3 if (emotionScore < threshhold) //the threshold for neutral images
    beeing classified as certain emotion
4 {
5 curEmotion = NEUTRAL;
6 }
7 }
```

Listing 5.3: Integrating threshholding

First, a boolean is checked to make sure whether the threshold should be checked or ignored (line 1). Next, the score of an emotion is compared with the found threshold and the current main emotion is adapted (lines 3-6). It is important to notice that this value can still be changed as checks for other emotions can take place afterwards. As coming below the threshold for one emotion is not guaranteeing that other emotions are absent this gives the opportunity to check.

5.2.6. Enums

Especially for the evaluation but also for training we have created some convenient enums. *Emotion* is used to state the main emotion. It contains the 6 basic emotions, neutral and unknown but can be further adapted. The other enums are only used for evaluation. *AngleKDEF* is mapping the different rotations of heads in the KDEF database to numbers. The enums *tilt* and *pan* do something similar. They provide the possible rotations, used in the head pose image database, separated as tilt and pan. Especially *EvaluateSystem* is using those enums. It is convenience class for the evaluation.

5.3. Programming Pepper

So far, we have described the process for the emotion recognition roughly. The next step would be a description of the interactions between humans and robot. As robot we have chosen Pepper.

Pepper itself already provides a basic possibility of emotion recognition. However, we are not using it and only focus on possible interactions which are based on the scenarios from subsection 3.2.5. We have implemented 3 cases. First, wishing a happy birthday. Next a sad and a happy user. Hereby, we have used Choregraphe 2.5.5 (5.1.5) which can be seen in figure 5.1.

Our implementation was done categorised by behaviour. Hence, they are not combined but two individual programmes one for the birthday and one for the happy and sad user.

5.3.1. Birthday

The birthday case is initiated by Pepper when it detects a face. The recognition still has to be integrated but is provided. However, first Pepper has to learn faces and data like birthdates which have to be checked when detecting a face. We have not done this because it is not important for showing the basic functionalities of the scenario. After congratulating you have to make sure Pepper is not doing the same all day long and save this as done. We have avoided this by a "do only once" condition which does not work in actual deployment.

Currently during congratulating we are using the tablet to show an image, the speakers to play a song, speech to congratulate, the eyes with changing colours and arm movement. However, other actions are possible as well and choosing from multiple actions would provide greater variety. After Pepper has finished it is starting a dialogue. The whole interaction is roughly depicted in figure 5.4. In this case

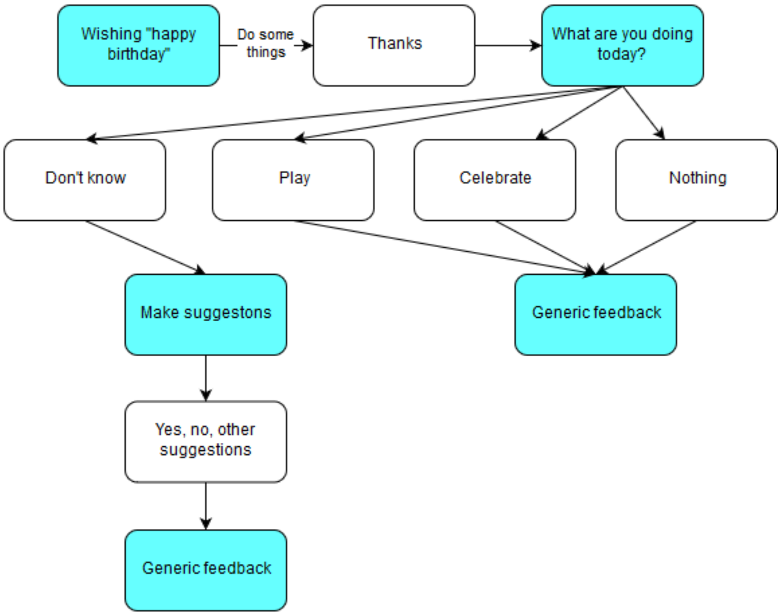


Figure 5.4.: Depicts the dialogue of Pepper wishing a happy birthday. Some parts have been simplified. The white boxes represent the user, the turquoise boxes represent Pepper.

Pepper and the user take turns in speaking when pepper inquires about the user's plan for the day. Using the right key words the dialogue can be stopped or restarted anytime.

For the dialogue we use the diologue box of Choregraphe. The box creates a *<name>_enu.top* file (for english dialogues) where you can define rules and grammars. For instance you can define dialogues with multiple layers like asking what to do in figure 5.4. To simplify and enrich them you can use delimiters[2], for example to map multiple words to one word which is used in the rules or to make words optional.

5.3.2. User starts interaction

Unlike the previous situation (birthday) this scenario is initiated by the user greeting Pepper and is followed by Pepper greeting, too. Hereby we are using a predefined way of waving Pepper's hand but it is also possible to define your own gestures and movements. In the test programme we are using the provided facial emotion recognition to simulate the integration of emotion recognition. Thus, either facial emotion recognition or the answered emotion decide the path of the dialogue. However, during the simulation using a virtual robot it is not possible to use or test the facial emotion recognition of Pepper. At the moment we have implemented possible reactions for a sad or a happy user as well as the possibility for Pepper to be chased off. Based on the emotion recognition or the spoken input the next steps are decided as shown in figure 5.5, which is slightly simplified. For the cases sad and happy we have put the behaviour inside boxes to encapsulate them from the rest of the behaviour.

In case of a sad user we make Pepper express sadness as well through its body and use another dialogue just for this situation (*SadUser_enu.top*). It is offering help in the form of listening or hugging. For a happy user we have a separate box and dialogue as well (*HappyUser_enu.top*). Hereby, Pepper is mainly

[2] file:///C:/Program%20Files%20(x86)/Aldebaran/Choregraphe%20Suite%202.5/share/doc/naoqi/interaction/dialog/dialog-syntax_full.html#delimiter-rule

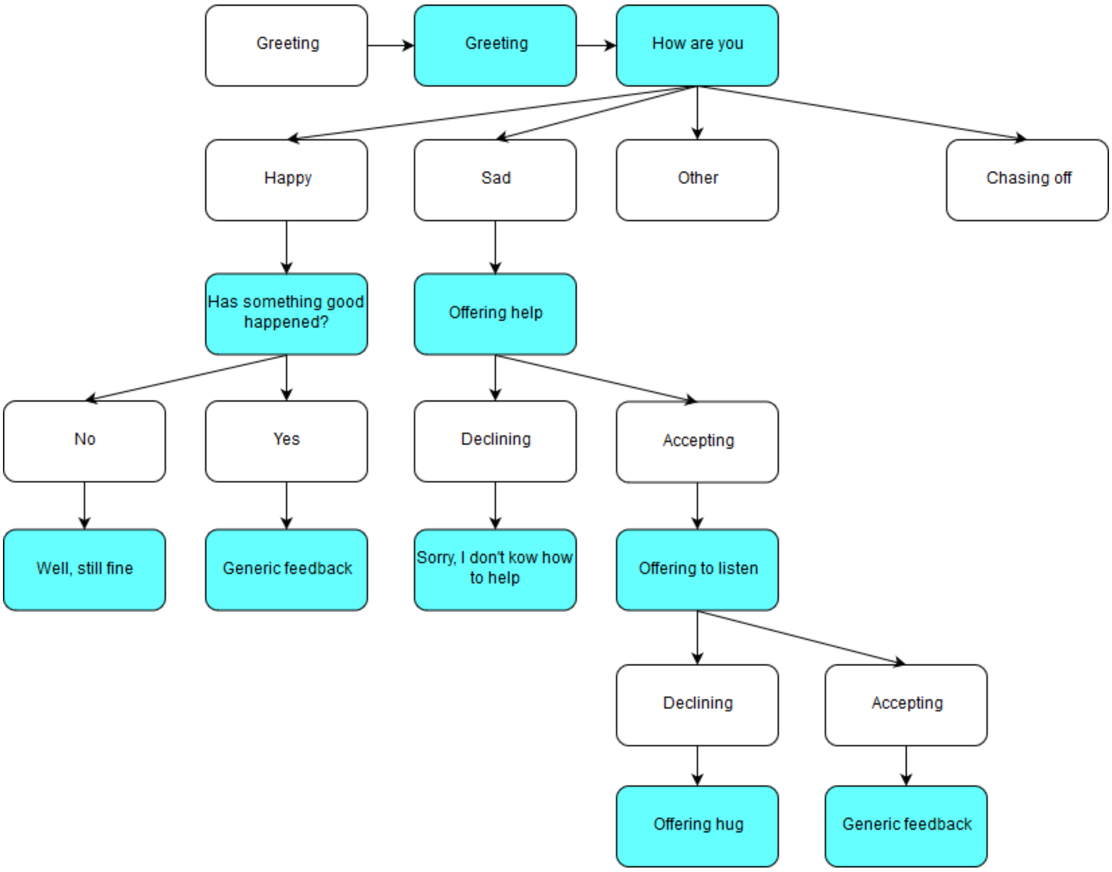


Figure 5.5.: Depicts the basic interaction of a user with Pepper. Some parts have been simplified. The white boxes represent the user, the turquoise boxes represent Pepper.

listening and using generic feedback plus sometimes expressing excitement through its movements. The current dialogues can be further enhanced and others can be added as well as more possible reactions.

5.4. Connecting

This far we have gone into detail about emotion recognition and programming Pepper. However, both still have to be connected to exchange images and the results of the emotion recognition. We have not implemented this part even so we want to present an option.

Our system runs independently of Pepper on a different computer and provides the service to analyse images. The behaviour of Pepper on the other hand runs on Pepper and may sometimes need another service to analyse videos. We suggest using ROS for them to communicate with each other as this possibility is officially supported. Further, some tutorials exist[3][4][5].

[3]http://doc.aldebaran.com/2-5/dev/ros/index_ros.html
[4]http://wiki.ros.org/pepper/Tutorials
[5]http://wiki.ros.org/pepper

6. Experiments and studies

So far, we have described and implemented some theories on emotional feedback systems but we do not know how realistic and therefore useful they are. In a normal situation, a robot should react similarly to a human conversation partner or at least he should seem to be fairly natural. Thus, one important step is to detect a person and to correctly categorise his emotions. For example, if a person is standing near a robot the robot should react with a greeting. Furthermore, if a person is crying the robot should not wrongly detect happiness else the interaction could be quite uncomfortable.

One problem with face detection and emotion recognition is the rotation a face has towards the camera. It is harder to search for specific features, like a movement of the eye corners, if it is already hard to identify them or their neutral position. Therefore, this is an important feature needed for emotion recognition.

After correctly recognising an emotion the robot should also choose a suitable reaction. For example, if someone is telling it he is feeling sad the robot should not ignore him and start talking about the weather. You may not want such a conversation partner.

In this chapter, we are explaining the experiments we have conducted for an evaluation of emotion recognition and some basic interaction scenarios.

6.1. Face detection and emotion recognition

The emotion recognition is partitioned into two main parts. Fist, detecting the wanted features and parts needed for the emotion recognition, which is the face in our case. Second, extracting the emotions. For both parts, we have conducted some basic experiments.

6.1.1. Face detection

Regarding face recognition, we have evaluated the recognition of rotated faces. Hereby, we have used the *Head Pose Image Database*, described in 7.1.1. It contains faces with different tilt and pan combinations.

Our reasoning for this experiment is as follows. Without being able to detect a face facial emotion recognition is not possible. However, the cameras, providing images, can be static as well as moveable but still restricted in its movements. For such cases it is interesting to know how a programme is performing and thus deciding whether it makes sense to use it or not.

6.1.2. Emotion recognition

Despite detecting a face this does not guarantee successful emotion recognition. A reason may be the expression being only faint or occlusions caused by a rotated head. Thus, the performance of a programme for detecting emotions in a rotated face is interesting, too. To gain some insight into a programme's performance we are using the *Karolinska Directed Emotional Faces* database, described in 7.1.2. However, we recommend repeating this experiment with different databases again because this database only contains faces with different pans but the same tilt. We have used this database since it was freely available. Further, we have analysed the performance for frontal faces more thoroughly.

By using the *extended Cohn-Kanade dataset*, described in 7.1.3, we want to make it possible to compare the results of multiple programmes with each other, using the results as a baseline. We are using 310 emotion labelled images and 327 images of neutral faces. Further, we want to compare the results for evaluating western faces and Japanese faces using the *Japanese Female Face Expression* database, described in 7.1.4, and a mixed subset of the *extended Cohn-Kanade* and the frontal images from the

Karolinska Directed Emotional Faces dataset. The *Japanese Female Face Expression* database consists of 213 images. A reason for this comparison is the possible difference in expressing oneself depending on your culture and growing up. This also refers to the intensity of expressing. Besides showing an emotion it is possible to not show any.

Not every programme supports returning *no emotion* after analysing for emotions. The same goes for the Affectiva SDK we are using. Hence, we have tried to find a workaround. One approach is training an svm and another one thresholding as described in 5.2.4 and 5.2.5. The results are described in chapter 7. Both approaches are dependent on getting confidence scores for emotions in our case for the six basic emotions. It should be noticed that actual usage can differ as the quality of the images varies and some are black and white images while others are coloured. Hereby, this evaluation is to find eventual potential.

6.2. Interaction

We have conducted an online survey to evaluate our basic interaction scenarios as well as for gathering opinions about alternative behaviour of a robot. This survey was conducted by the usage of *Google Forms*[1]. It was intended to be done once per participant and all at once. In the following, we will describe the structure of the survey and go into detail as well.

6.2.1. Structure of the survey

The basic structure of the survey is as follows. First, the participants are made aware of some basic information about the study, containing enabling the sound of their devices, the whole study being in English, finishing it in one go and the data handling. The next four sections show videos of scenarios and ask questions per scenario. Afterwards, two sections ask questions about all scenarios together. We conclude with one section asking information about the participant.

Now, we describe the four sections containing the scenarios in detail. At the beginning, a video of one scenario is shown. Next, the inquiry about the video follows. The basic questions for each video are the same.

1. Do you think the robot acts staged or natural?
   Answer: very staged - staged - a little staged - neither staged nor natural - a little natural - natural - very natural

2. Please read the following statements and rate them.
   a) The robot is creepy.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree
   b) The robot would make a good companion.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree
   c) The voice of the robot is suitable.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree
   d) I would feel like the robot controls the situation.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree
   e) The answers of the robot make sense.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

[1] https://docs.google.com/forms/u/0/

f) The robot shows emotions.
Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

g) The eye colour of the robot changes.
Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

h) The robot reacts according to the other person's emotions.
Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

i) The eye colours of the robot show emotions.
Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

j) The posture and the gesture of the robot show emotions.
Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

3. What do you think is the goal of the robot during this interaction?
   Answer: text field

4. Do you think the robot accomplishes its goal?
   Anwer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

5. Do you think people will like/dislike such a robot? Please rate the following

   a) You
   Answer: strongly dislike - dislike - slightly dislike - neither like nor dislike - slightly like - like - strongly like

   b) Elderly
   Answer: strongly dislike - dislike - slightly dislike - neither like nor dislike - slightly like - like - strongly like

   c) Children
   Answer: strongly dislike - dislike - slightly dislike - neither like nor dislike - slightly like - like - strongly like

   d) Tech-savvy
   Answer: strongly dislike - dislike - slightly dislike - neither like nor dislike - slightly like - like - strongly like

   e) Tech-averse
   Answer: strongly dislike - dislike - slightly dislike - neither like nor dislike - slightly like - like - strongly like

   f) Others apart from you
   Answer: strongly dislike - dislike - slightly dislike - neither like nor dislike - slightly like - like - strongly like

6. Do you have any suggestions for the robot's behaviour in this kind of situation?
   Answer: text field

7. Any feedback to this video? :)
   Answer: text field

Question 1 and question 2 help to rate the interaction shown in the video and ask about the participant's opinion. We mainly concentrate on the whole interaction and the acceptance of Pepper and its behaviour. Further, we ask about some details like perception of the changing eye colours. Question 3 should give us information about the interpretation of the interaction while question 4 asks how well this interpretation is fulfilled. Each interaction was video taped as a specific scenario which we did not explicitly name in the survey so we can gain insight into other people's views without biasing them too much. Next, question 5 is asking about the acceptance of a Pepper as shown in the video, based on different user groups (participant, elderly, children, tech-savvy, tech-averse and others apart from the participant). Hereby, we want to evaluate whether the scenario is usable for those groups. Question 6 and 7 are feedback questions about the scenario, one for general feedback, the other specific for the behaviour of the robot. Both are voluntary while all other questions are mandatory. Further, each scenario contains a few additional questions about the shown interaction.

Scenario 1 contains the following questions as part of question 2.

1. I could hear a song.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

2. I could see balloons on the tablet.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

3. The robot is wishing a happy birthday.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

4. The robot seems happy.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

5. Playing a song is a good idea.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

As the video was not taken professionally some things may be hard to hear or see in the video. Therefore, we want to make sure by asking those questions. Further, we ask if our intention to show a happy robot is conveyed and ask about the participant's opinion of the robot playing a song.

The section of scenario 2 contains a control question as additional question.

1. Which colour is closest to the woman's shirt?
   Answer: red - blue - green - yellow - I don't know - other

The actual answer would be something like berry red. However, this colour was not listed as a choice with the intention to make the participant think about it. Besides, it checks whether the participant concentrated on the video or not and how perceiving he was. Therefore, we accept all answers but expect mainly blue and red.

Regarding scenario 3 we added two questions to question 2.

1. The robot seems sad after hearing that the other person is happy.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

2. The robot seems happy after hearing that the other person is happy.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

Hereby, we want to test whether our intended emotional impression of the robot was conveyed. We ask about our intended emotional expression and its opposite.

For scenario 4 we add to question 2.

1. I would hug the robot in this situation..
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree
2. The robot seems sad after hearing that the other person is sad.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree
3. The robot seems happy after hearing that the other person is sad.
   Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - agree - strongly agree

Those questions are similar to the added questions of scenario 3 to test the conveyed emotional expression of the robot. However, we also ask whether the participant would hug the robot in the shown situation. This question provides an insight into the participant's mind about his acceptance of the offered hug and the robot. The next two sections contain questions about all interactions.

For this questions we have asked the participants to think about all interactions altogether as well as to imagine being the user shown in the videos.

1. Do you think the robot acts staged or neutral?
   Answer: very staged - staged - a little staged- neither staged nor natural - a little natural - natural - very natural
2. Would you feel comfortable using the robot?
   Answer: very uncomfortable - uncomfortable - a little uncomfortable - neither comfortable nor uncomfortable - a little comfortable - comfortable - very comfortable
3. Please rate the following statements and rate them.
   a) I would feel watched by the robot.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - strongly agree
   b) I would feel like having control over the robot.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - strongly agree
   c) The interaction works fine for short people.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - strongly agree
   d) I would feel comfortable in the robot's company.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - strongly agree
   e) I would feel like the robot controls the situation.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - strongly agree
   f) The interaction works fine for tall people.
      Answer: strongly disagree - disagree - somewhat disagree - neither agree nor disagree - somewhat agree - strongly agree
4. Do you think a robot should react towards your current emotion/mood?
   Answer: yes - no
5. What did you like the most about the robot's interaction?
   Answer: text field
6. What did you like the least about the robot's interaction?
   Answer: text field

7. What kind of reactions would you like from the robot?
   Answer: text field

8. Any kind of feedback about the interactions :)
   Answer: text field

Question 1, 2 and 3 are about the naturalness and acceptance of the robot, considering all scenarios. Additionally, question 3c and 3f ask how well the interaction works out for tall and short people, which will be helpful for possible environments of interactions. For this, the videos should have given an impression as one of the actors was tall while the other one was short.

Question 4 is the general question whether a robot should react towards emotions or not. Besides giving a reason for this research it gives some indication about the acceptance of emotional intelligent machines.

Question 5, 6 and 7 help to gather information about what people like the most or the least and therefore what can be improved and what can be used more often.

Question 8 gathers any kind of feedback referring to the interactions. This question was obligatory while the others were mandatory.

The second of the section referring to all shown interactions consists of the first 9 SUS questions ([27]) inquiring about the usability. We omitted the last question *I needed to learn a lot of things before I could get going with this system*. Our reasoning is that none of the participants actually used the system and therefore would have to guess, which might lead to unreliable results.

The last section of our survey contains demographic questions about the participants. We have refrained from asking for nationalities as we did not expect a great variance of nationalities which could make it possible to guess the identities of some participants.

6.2.2. Scenarios

The first four sections of our survey contain four videos of interactions between a human and a robot, Pepper. Those interactions are based on the implementation with Choregraphe described in 5.3. Now, we want to describe those interactions in detail. However, during all interactions, the autonomous behaviour of Pepper was activated and you have to mind the preprogrammed colour changes of Pepper. Further, camera view and audio may be improvable as video taping was not professional. The same goes for the actors. In addition, it needs to be noted that the scenarios are just basic interactions and not very complex.

The first interaction (scenario 1) is based on the use case *birthday*. A woman stands before Pepper. It starts showing balloons on its tablet and playing the song *Long may you live*, which is a translation of the German birthday song *Hoch sollst du leben*. While doing so its eyes slowly blink in random colours and turn yellow for 5 seconds when they finish. When the song ends Pepper congratulates the woman using her name and widening its arms. She thanks it and Pepper asks about her plans for the day. She responds with plans about a party and Pepper wishes her a nice party. Next, the woman says goodbye to Pepper.

Scenario 2 shows a short greeting. A woman greets Pepper and it responds with a greeting and hand waving.

In scenario 3 a man interacts with Pepper. Pepper asks how he feels and gets the answer "great". While moving a little back (space and posture) Pepper turns its eyes yellow for 2,5 seconds. It asks whether something special has happened and gets a positive reply. Afterwards, Pepper requests more details and thanks after getting an answer. This interaction is based on the use case *Greeting a happy person*.

Scenario 4 is based on the use case *sad user*. Pepper asks a man how he feels today. After getting the answer sad it responds with being sorry for him and its body crumbles a bit while having blue eyes for 1,5 seconds. Pepper offers to help which is accepted. Meanwhile, its eyes shortly turn green. Next, it offers to listen, which is declined. Instead, Pepper offers a hug, which is accepted. It raises its body and opens its arms offering the hug. The man bows down and hugs Pepper. Afterwards, Pepper offers another one, which is declined, and asks how the man feels now. He answers to feel better.

6.2.3. Conduction

The survey was built and conducted with Google Forms[2]. All of it is in English, the videos, too. It was active over the period of three weeks and we attained 72 participants. In search of participants, we have reached out via social media, which results mainly in younger participants.

The whole survey takes about 20-30 minutes and participants were asked to fill it in at one go. For participating no previous knowledge was needed.

[2]https://docs.google.com/forms/u/0/

7. Evaluation

In this chapter we will first discuss the performance of *Affectiva SDK*. For the evaluation of emotion recognition we have used several databases, which are described in 7.1. While using some of them as provided we also built some sets suitable for direct comparison. The main points of this evaluation are:

- Detection of rotated faces (pan and tilt combinations)
- Emotion recognition for rotated faces
- Emotion recognition in general
- Western vs. Japanese faces

The next part of our evaluation is about the usage of some basic feedback strategies. Therefore we have created some scenarios and built an online survey containing videos of these. In this chapter we elaborate the results of our online survey described in section 6.2. We have partitioned the analysis in two parts, qualitative and quantitative analysis.

7.1. Databases

In this section we describe the databases used for evaluation.

7.1.1. Head Pose Image Database

The *Head Pose Image Database* ([42]) contains 2790 colour images of people from various ethnicities. These images show 15 different faces of multiple combinations of tilt and pan. Possible configurations for *tilt* are *90, 60, 30, 15* (positive and negative) and *0* degrees, while the configurations for *pan* are *90, 75, 60, 45, 30, 15* (positive and negative) and *0* degrees. However, not every possible configuration was used while taking pictures. Therefore, for a tilt of plus minus 90 degrees only images with a pan of 0 degrees exist. For each used configuration 30 images were made in total.

To compensate for any kind of measuring inaccuracy between positive and negative rotation we have mirrored the images using BIMP[1], which is a plugin for batch operations with GIMP[2]. We have combined the original and the mirrored images into one database. Therefore, we have 60 images per used configuration and a total of 5580 images. All images are of the type JPEG.

The names of the original images contain their rotation label at the end. When we refer to the rotation of a face we will use camera view as a base. For example, if we say the face is rotated to the left than a spectator in front of the face would say so though the person, who was photographed, would say he moved to the right.

7.1.2. KDEF database

The *Karolinska Directed Emotional Faces* database, *KDEF*, ([58]) consists of 4900 images, which are emotion labelled. 70 amateur actors, 35 females and 35 males, show 7 different emotions, which are *anger*, *disgust*, *fear*, *happiness*, *neutral*, *sadness* and *surprise*. In total we have 700 images per emotion. All images are coloured.

[1] http://registry.gimp.org/node/26259
[2] https://www.gimp.org/

In addition to emotion labels, the images also have labels for pan. Each person was photographed for a *pan* of *+90, +45, 0, - 45* and *-90* degrees. This results in 140 images per pan and emotion.

For each person, 2 photo sessions were made. Therefore, we have 70 images of different people per pan and emotion. We are using only the images from session A, meaning 70 images per person, pan and emotion and 2450 images in total. Some images were defective or could not be recognised by our system. Those we have exchanged for their correlating images of session B.

7.1.3. Extended Cohn-Kanade dataset

The *extended Cohn-Kanade database*, *CK+*, contains 327 sequences of images ([57]). All of them start with a neutral facial expression and end with an expression which can be categorised as one of the 7 basic emotions, namely anger, contempt, disgust, fear, happiness, sadness and surprise. Some of the images are coloured, some are black and white.

We have extracted the first image of every sequence as a neutral image. Furthermore, we have extracted the last image of every sequence and categorised them accordingly to [57]. During sorting of the images, we noticed that not every name is unique. Thus, we have added a number at the end of each name to ensure that each name is unique per emotion label. Overall, we have 654 images consisting of *45 angry* faces, *18 contemptuous* faces, *59 disgusted* faces, *25 frightened* faces, *69 happy* faces, *327 neutral* faces, *28 sad* faces and *83 surprised* faces. As other datasets, we are using, are labelled for only 6 basic emotions we are not using the images labelled with the emotion *contempt*.

7.1.4. JAFFE database

The *Japanese Female Face Expression* database ([59]), *JAFFE*, consists of 213 images of 10 Japanese female models. All images are black and white. This database is labelled for the 6 basic emotions and *neutral*, meaning for *anger*, *disgust*, *fear*, *happiness*, *neutral*, *sadness* and *surprise*. Each image was classified according to its main emotion.

For each of the labelled emotions, the number of images varies because not every face has the same number of images per emotion. We are having *30* images for *anger*, *29* images for *disgust*, *32* images for *fear*, *31* images for *happiness*, *30* images for *neutral*, *31* images for *sadness* and *30* images for *surprise*.

The original data type is tiff but we have converted all images to jpg. During all experiments the converted versions have been used.

7.1.5. Mixed subset

In 7.1 we describe several databases we are using for evaluation. As they have different numbers of images per emotion a direct comparison of the results would be falsified. Thus we have mixed a dataset like shown in figure 7.1.

To make our own database we have combined the CK+ database and the frontal images of the KDEF database. From those images, we have randomly selected a number of images per emotion. The numbers were decided by the JAFFE database because we want to compare the results of our self-built database and the results of the JAFFE database. Therefore, we have selected *30* images for *anger*, *29* images for *disgust*, *32* images for *fear*, *31* images for *happiness*, *30* images for *neutral*, *31* images for *sadness* and *30* images for *surprise*. The remaining images were not used for evaluation purposes.

7.2. Affectiva SDK

In the following we will show our evaluation of the *Affectiva SDK*. As a first step we have tested what kind of rotated faces are still recognised. Therefore, we have used our combined database made from the *Head Pose Image Database* and its *mirrored images*. Hereby, we use the Affectiva setting *LARGE_FACES* as the used images contain single large faces. In table 7.1 you can see our results. One of the first things, which surprises us, is that the resulted matrix is not symmetrical even though we have mirrored all images

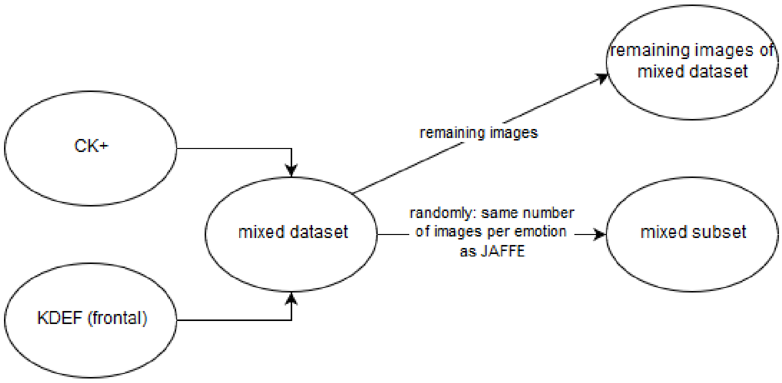


Figure 7.1.: The mixed subset is randomly chosen from CK+ and KDEF (only frontal faces). The result contains the same number of images per emotion as JAFFE.

tilt \ pan	-90	-75	-60	-45	-30	-15	0	+15	+30	+45	+60	+75	+90
-90	x	x	x	x	x	x	13.3	x	x	x	x	x	x
-60	1.7	1.7	0	3.3	23.3	38.3	46.7	43.3	28.3	10.0	1.7	0	1.7
-30	0	0	1.7	18.3	45	71.7	95	86.7	58.3	25.0	10.0	1.7	0
-15	1.7	0	10.0	38.3	75	91.7	100	93.3	83.3	56.7	18.3	1.7	0
0	0	3.3	8.3	46.7	95	98.3	100	98.3	95	80	41.7	5.0	0
+15	0	0	8.3	33.3	91.7	96.7	96.7	96.7	93.3	66.7	21.7	1.7	0
+30	0	0	3.3	25	83.3	91.7	91.7	91.7	78.3	38.3	6.7	3.3	1.7
+60	0	0	0	1.7	10	40	43.3	38.3	13.3	0	1.7	0	0
+90	x	x	x	x	x	x	1.7	x	x	x	x	x	x

Table 7.1.: The results for testing the mirrored and the original images of the Head Position Image Database regarding face detection by the Affectiva SDK (in percent). Dark grey means 90% or higher, middle grey 80% or higher and light grey 70% or higher.

to compensate for possible measurement inaccuracies. One possible reason could be that not every part of the images is checked for a face but that some parts are checked according to the used algorithm. The face detection matrix for only the original images (see table 7.2) is similar. It is even less symmetrical and shows better as well as worse performance for some angles.

As you can see, the matrix for all images is fairly symmetric for less rotation, meaning up to 30 degrees of tilt and pan. In case of pan the Affectiva SDK performs better, when the face is rotated to the right. One reason for this performance may be that higher rotation has a high influence on the recognition rate making it unpredictable. Another possibility could be the training data which was used to train the Affectiva SDK. We don't have details about those but as Affectiva mentions that they are confident for good results up to 25 degrees we can guess that the SDK was not trained for extreme rotations. Thus, the recognition rate decreases and varies for higher rotation.

The recognition rate for frontal images is quite good but rapidly decreases with higher rotation. Hence, during usage people should confront the camera as much as possible and should be instructed to do so before. Further, for the camera position the target user's height should be considered so pan and tilt can be reduced. Another possibility would be to have a moveable camera which can adjust height and rotate itself in order to get frontal images.

In all images only single and big faces where shown. Thus we chose the setting *LARGE_FACES* provided by Affectiva. If we change the modus to *SMALL_FACES* the results are similar but detection is lower and the matrix is still not symmetrical. We conclude that Affectiva uses different algorithms to find small and large faces. Further, we conclude that for both cases not every part of the image is checked

tilt \ pan	-90	-75	-60	-45	-30	-15	0	+15	+30	+45	+60	+75	+90
-90	x	x	x	x	x	x	13.3	x	x	x	x	x	x
-60	0	0	0	6.7	30	40	53.3	43.3	20	6.7	0	0	3.3
-30	0	0	3.3	20	50	73.3	96.7	83.3	56.7	23.3	3.3	0	0
-15	3.3	0	13.3	43.3	86.7	90	100	96.7	80	46.7	13.3	0	0
0	0	6.7	16.7	50	96.7	100	100	96.7	96.7	76.7	36.7	0	0
+15	0	0	16.7	40	90	96.7	96.7	96.7	93.3	70	13.3	0	0
+30	0	0	6.7	26.7	73.3	90	93.3	93.3	90	46.7	6.7	6.7	3.3
+60	0	0	0	3.3	6.7	33.3	33.3	36.7	10	0	3.3	0	0
+90	x	x	x	x	x	x	3.3	x	x	x	x	x	x

Table 7.2.: The results for testing only the original images of the Head Position Image Database regarding face detection by Affectiva SDK (in percent). Dark grey means 90% or higher, middle grey 80% or higher and light grey 70% or higher.

but only some dependent on the details of the use algorithms.

After having analysed the ability of Affectiva SDK to detect rotated faces we are interested in its ability to recognise emotions for rotated faces. Hereby, we are using a part of the KDEF database (7.1.2), meaning only the rotated faces and not the frontal faces. As an evaluation for frontal images will follow later on we are not using frontal faces but just 45° and 90° rotated faces. Nearly all images with 90° rotated faces were not recognised which was expected in regards to our evaluation of face detection for rotated faces. Therefore, we are only using the results of 45° rotated faces. They are shown in table 7.3.

emotion \ images	Half left TPR	Half left FPR	Half right TPR	half right FPR
anger	0	0	0	0
disgust	0	0	0	0.8
fear	0	0	0	0
happiness	100	99.6	98.0	99.6
neutral	0	0	0	0
sadness	0	0	0	0
surprise	6.3	0	0	0

Table 7.3.: Evaluation of the emotion recognition in rotated faces by Affectiva. The 45° rotated faces of the KDEF database have been used. For measurement we use true positive rate (TPR) and false positive rate (FPR).

Nearly all emotions were wrongly recognised but at the same time they were mostly recognised as the same wrong emotion, namely happiness. Therefore, you could think happiness is recognised quite well, however it seems more like guessing as happiness is found nearly all the time. We guess a reason could be that the Affectiva SDK is searching for certain expressions, which correspond to emotions, and those are hard to find in a rotated face. For example, if someone has raised the inner eyebrow it is possible that you are not able to see it depending on your angle towards him. The same goes for an algorithm.

To show the general performance ot the Affectiva SDK we use the CK+ database. In table 7.4 you can see the results, when searching for the main emotion. *Happiness* and *surprise* are quite well recognised (over 90%). Hardly any other emotion is mistaken as *happiness* while fear is often mistaken as surprise (40%). Further, *disgust* and *sadness* are well recognised (87,5% and 70,4%) other emotions are mistaken for them as well. Only 39,5 % of *anger* is recognised and it is often misclassified as *disgust* and *sadness*. *Neutral* is always misclassified because Affectiva does not support it. It mainly defaults to *disgust*, which is useful to know for thresholding.

In contrast to the main emotion table 7.5 depicts the confusion matrix according to the least probable

expected \ found	anger	disgust	fear	happiness	neutral	sadness	surprise
anger	39.5	34.9	2.3	0	0	20.9	2.3
disgust	6.25	87.5	0	4.2	0	0	2.1
fear	8	12	20	4	0	12	44
happiness	0	0	0	98.6	0	0	1.4
neutral	0	75.5	16.8	0.6	0	0.9	6.1
sadness	3.7	3.7	7.4	0	0	70.4	14.8
surprise	0	4.4	1.5	0	0	0	94.1

Table 7.4.: The confusion matrix for emotion recognition for the CK+ database. As main emotion the emotion with the highest score has been used.

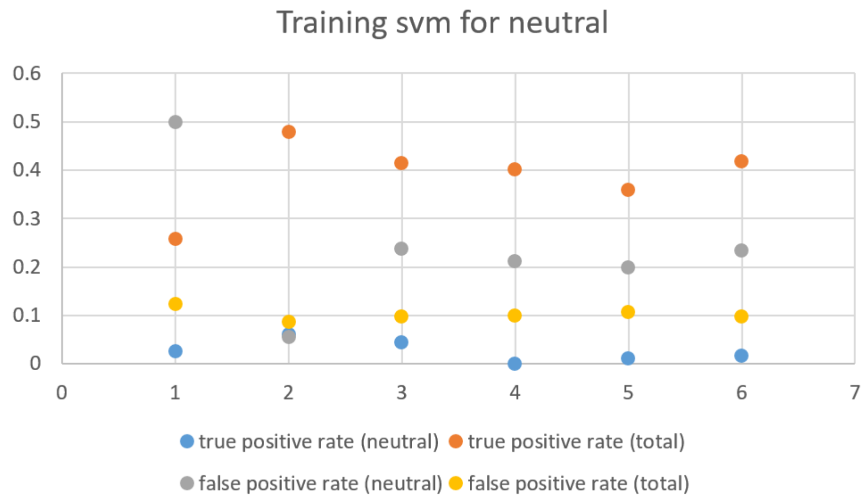
emotion, meaning the one with the lowest score. You can see to which emotion each emotion and neutral default to the *least often*, *anger* to *happiness*, *disgust* to *happiness*, *fear* to *happiness* and *anger*, *happiness* to *sadness*, *neutral* to *happiness*, *sadness* to *happiness* and *surprise* to *sadness*. Overall, none of the emotions has a high score for themselves, which is good because they should not be the least probable result but the most probable one.

Most emotions show big scores for *happiness*, which means that it is well definable. Combined with the low score of *happiness* for itself we can conclude a high probability for *happiness* if it is chosen as main emotion. Further, *neutral* seldom defaults to *happiness* (70,9%), which supports our conclusion. On the other hand, *neutral* shows a score of 0% for *surprise* and *disgust*. Thus, we guess in case of deciding for one of them as the main emotion it makes sense to check for *neutral* as well, for example, by thresholding. In case of *disgust* the thresholding could be clearer because less other emotions are misclassified as *disgust*.

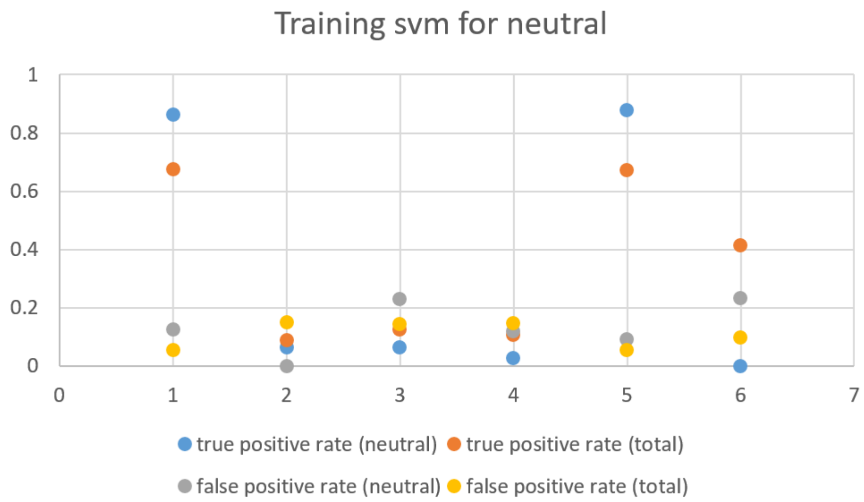
expected \ found	anger	disgust	fear	happiness	neutral	sadness	surprise
anger	2.3	0	0	97.7	0	0	0
disgust	0	0	22.9	52.7	0	25	0
fear	32	0	8	36	0	24	0
happiness	7.2	1.4	13.0	0	0	78.3	0
neutral	9.2	0	2.4	70.9	0	17.4	0
sadness	29.6	0	3.7	66.7	0	0	0
surprise	4.4	0	4.4	38.1	0	52.9	0

Table 7.5.: The confusion matrix for emotion recognition for the CK+ database. The least probable emotion has been used as main emotion.

As previously mentioned the Affectiva SDK does not provide a score for *neutral*. Because of this, we have trained a svm to decide whether the current main emotion is *neutral* or one of the *6 basic emotions*. In subsection 5.2.4 we have further described our method of training. We trained the svm several times with a different amount of data, *90%, 80%, 70%, 60%, 50%* and *40%*. The remaining data was used to evaluate the trained svm. Hereby, the data which was used for training was randomly chosen from the total training data. In 7.2(a) the corresponding scores for *false positive* and *true positive rates* for the class *neutral* as well as for *all classes* are shown. We do not see significant differences. The *true positive rate* for *neutral* is always below 10% and one time the *false positive rate* is even about 50%, which is about the same level as guessing. The *false positive rate* for all classes is always below 15% while the corresponding *true positive rate* is at least over 25%. Hence, we get a first impression about the svm. It is probably not changing a lot of previously decided labels to neutral or else the scores would differ. Further, the *accuracy* for *neutral* is quite bad as the *false positive rate* is nearly always above the corresponding *true positive rate* (the svm trained with 80% of the data is the only exception).



(a) Scores of linear svm trained with all 6 scores as input (one per emotion). 1 represents the first svm (90% training data) and 6 the last trained svm (40% of training data).



(b) Scores for a linear svm trained with only the maximal score as input.

Figure 7.2.: Scores of linear svms trained to classify neutral. 1 represents the first svm (90% training data) and 6 the last trained svm (40% of training data).

As mentioned in 5 it is possible that a connection between the absence of emotions and the variation in emotion scores exists. Therefore, we have trained a svm with only the maximal score as input. The results can be seen in 7.2(b). Compared to the svm trained with all scores the *true positive rates* look much more unstable. Sometimes they are above 70%, sometimes below 20%. The *false positive rates* seem more stable (from about 0% to 25%). Overall, the svm does not seem dependable and especially less dependable than the one trained with all scores as input.

For actual usage we prefer the svm trained with all emotion scores as input. We chose the svm which was trained with *80%* of the training data as it has the *highest true positive rate* and the *lowest false positive rate* in regards to *all emotions* as well as towards only *neutral*. We have also used this svm on the rotated faces from the KDEF database, but no significant change was noticeable.

expected \ found	anger	disgust	fear	happiness	neutral	sadness	surprise
anger	48.3	17.2	0	3.4	0	31.0	0
disgust	3.7	81.5	0	14.8	0	0	0
fear	9.4	15.6	21.9	3.1	0	15.6	34.4
happiness	0	3.2	0	93.5	0	0	3.2
neutral	0	63.3	20	0	0	0	16.7
sadness	10	13.3	6.7	0	0	60	10
surprise	0	7.1	0	0	0	0	92.9

Table 7.6.: The confusion matrix for the combined database of Western faces (mixed subset) (in percent). The svm for neutral was not used.

expected \ found	anger	disgust	fear	happiness	neutral	sadness	surprise
anger	48.3	17.2	0	3.4	0	31.0	0
disgust	3.7	77.8	0	14.8	3.7	0	0
fear	9.4	15.6	15.6	3.1	9.4	15.6	31.3
happiness	0	3.2	0	90.3	3.2	0	3.2
neutral	0	63.3	13.3	0	6.7	0	16.7
sadness	10	13.3	6.7	0	0	60	10
surprise	0	7.1	0	0	0	0	92.9

Table 7.7.: The confusion matrix for the combined database of Western faces (mixed subset) (in percent). The svm for neutral was used.

Table 7.7 depicts the results for the frontal Western faces when using the svm to classify *neutral*. Only 6,7% of the *neutral* images are classified correctly while about 63,3% are classified as *disgust*. Further, parts from the class *disgust*, *fear* and *happiness* are misclassified as *neutral*. Overall, the performance of the svm is not convincing.

In subsection 5.2.5 we have described *tresholding* as an alternative to classify *neutral*. Motivated by our guess that *disgust* might be well definable from *neutral* by thresholding we have tried to find a threshold. Indeed, the threshold is quite clear. For our data the result is *neutral* when it has been classified as *disgust* with a score 0,46. This proves the possibility of successful thresholding. However, for actual application it should be defined on realistic data as the quality of some of our images may not be adequate.

For an evaluation on Japanese faces and expressions we have used the database JAFFE (described in subsection 7.1.4). During processing of the images *20* of them were *not recognised* by Affectiva as containing a face, while only *193* faces were *recognised*. *Not recognised* were *2* images for *disgust*, *3* images for *fear*, *2* images for *happiness*, *1* image for *sadness* and *12* images for *surprise*. Therefore, our evaluation refers to *30* images for *anger*, *27* images for *disgust*, *29* images for *fear*, *29* images for *happiness*, *30* images for *sadness* and *18* images for *surprise*. Table 7.8 shows the result of the

expected \ found	anger	disgust	fear	happiness	neutral	sadness	surprise
anger	0	70	10	0	0	0	20
disgust	21.4	50	0	0	0	17.9	10.7
fear	0	13.8	24.1	0	0	0	62.1
happiness	0	3.4	20.7	72.4	0	0	3.4
neutral	0	23.3	50	0	0	0	26.7
sadness	0	36.7	6.7	0	0	43.3	13.3
surprise	0	5.6	0	5.6	0	0	88.9

Table 7.8.: The confusion matrix for the JAFFE database (in percent). The svm for neutral was not used.

processing. Surprisingly *anger* was always misclassified, mainly as *disgust*. *Happiness* is quite well recognised (72,4%) but *surprise* even better (88,9%). The *other* emotions have scores between 24% and 50%. *Neutral* is never classified as *anger, happiness* or *sadness* but often as *disgust* (63,3%), which indicates possible thresholding. Overall, some emotions are well recognised but not all. Further, each emotion has some false positives.

expected \ found	anger	disgust	fear	happiness	neutral	sadness	surprise
anger	0	70	10	0	0	0	20
disgust	21.4	50	0	0	0	17.9	10.7
fear	0	13.8	6.9	0	17.2	0	62.1
happiness	0	3.4	13.8	72.4	6.9	0	3.4
neutral	0	23.3	33.3	0	16.7	0	26.7
sadness	0	36.7	3.3	0	3.3	43.3	13.3
surprise	0	5.6	0	5.6	0	0	88.9

Table 7.9.: The confusion matrix for the JAFFE database (in percent). The svm for neutral not used.

Though, the *svm* was trained by using Western faces we are testing in on Japanese faces, too. The results can be seen in table 7.9. In contrast to the Western faces a slowly higher percentage of the *neutral* images are classified correctly but still not even half of them. In both cases it is instead classified as *disgust, fear* or *surprise* but in different proportions. Regarding *false positives* the classes vary, too. For Western faces this refers to *disgust, fear* and *happiness* while it is *fear, happiness* and *sadness* for the Japanese faces. Further, in both cases the performance of the svm is not convincing.

To show the differences and similarities in the recognition rate for Western and Japanese faces we take a look at the tables 7.6 and 7.8. In general, the recognition scores for Western faces are higher. Further, in both cases *happiness* is recognised the best. Also, *neutral* mainly defaults do *disgust* indicating the possibility to use thresholding. While *anger* is never recognised for Japanese faces it is still partly recognised for Western faces. On the other hand, *fear* is slightly better recognised but overall the performance for Western faces is more successful. We cannot be sure that this is the general case because it is possibly caused by data scarcity of images with Japanese faces.

7.3. Survey

To gather opinions about our basic interaction scenarios we have composed an online survey (described in section 6.2), which contains qualitative and quantitative questions. First, we will concentrate on the qualitative results and afterwards analyse the quantitative results.

7.3.1. Demographics

Our survey had *72* participants, *45 male*, *26 female* and *1 other*. They are part of the age ranges *18-24*, *25-34*, *45-54* and *55-64*. Nearly all participants are between *18* and *34*. *20* participants have graduated from *high school*, *39* from *Bachelor's studies*, *9* from *Master's studies*, *2* hold a *PhD* and *2* did not specify their highest level of education. The fields for these degrees are *computing (35)*, *engineering (10)*, *healthcare (2)*, *mathematics (3)*, *medicine (1)*, *science (4)*, *technology (6)* and *other (11)*. The participants' current deployment consists of the categories *student (59)*, *employed for wages (11)*, *self-employed (1)* and *unable to work (1)*. They either live *alone* or with up to *29* people. On average, they live with *3* other people. Further they perceive themselves as *tech-savvy* and like to use *new technical developments*.

7.3.2. Qualitative

Qualitative results are in the form of textual answers and therefore not restricted. We use them to gather opinions about our basic scenarios as well as suggestions for adjustments. Further, we expect general opinions as we did not specify the intended environment beforehand. In the following, we will cite some answers anonymously not changing form or content.

7.3.2.1. Goal of the interaction

When filming the interactions, we had specific scenarios in mind. However, we did not mention those in the survey to prevent biasing the participants. Another reason is that we want to verify if the intention of the interaction is comprehensible.

Scenario 1 Scenario 1 involves a robot congratulating a woman on her birthday. The speculations of the interaction goal range from quite direct ones (*"wishing a happy birthday"*) to abstract ones (*"interaction with a human"*) to more specific but still abstract ones (*"Goal of this interaction could be to make the person happier for this day"*). About 36 % of the participants guess the interaction goal is congratulating, which matches our scenario. Other 51 % think of some way of interaction. This should prove the perceptibility of the basic principle that the robot wants to interact and is not trying to prevent this or get rid of the human interaction partner. However, one participant comments *"making small talk to appear the minimum level of polite while also keeping the talking as short as possible so it can go home and watch netflix"*, which shows that the whole conversation can seem rather superficial. In the context of this work we have expected this kind of feedback as the interactions equal just basic interactions. For future designing it is an important aspect which collides with not having enough contextual knowledge and keeping the requirement of contextual knowledge low.

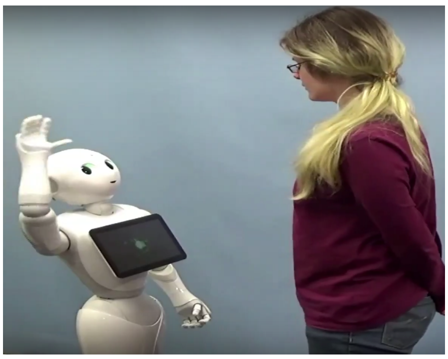


Figure 7.3.: Pepper greeting a human while waving its hand.

Scenario 2 Scenario 2 shows a robot greeting a woman while waving its hand (figure 7.3). Most participants (over 50%) guess greeting as the intended goal, which is correct. Some add features like trying to be human-like, natural or responding accurately in general. In addition, some people think about our intended use of combining speech and gesture (hand waving). However, the gesture is not welcomed by all. Comments like *"annoying people directly in front of it with waving"* and *"greeting someone? It just seems like he try to imitate human gesture tu say "hi!". But badly. I mean, he does that well, but it's not natural at all. "Good morning" was enough for that goal."* voice those concerns.

One comment, *"Cheating a little"*, we find hard to comprehend. It could mean the robot is trying to act like a human but is not really human and therefore it is cheating or faking. Another one could be that the participant thinks the robot has no clue about the situation and the dialogue but is simply mimicking the human and thus cheating. However, the wording sounds negative and we can conclude the shown interaction appears to be negative in some way.

Overall, we see the need to further enhance gesture during interactions between robots and humans.

Scenario 3 In scenario 3 a robot is interacting with a happy man asking about what has happened. Hereby our goal was to make the robot seem empathetic and interested. As this scenario is a bit more complex than scenario 1 and 2 we expect a bigger range of different answers. Still many of them remain reserved by describing the interaction. Some participants added more details like *"To show an appropriate and friendly reaction to the conversation partner"* or *"Raise positive emotions, doing smalltalk (how the day was)"*. In general, people were able to comprehend that the robot tried to act according to its conversation partner. Another participant states *"Creeping information's form the you're to help him psychological"*, which could mean that the robot tried to obtain information from its conversation partner so it can psychologically support him.

We had intended for the man to seem happy but got the feedback *"find out why the human seems depressed"*, which we trace back to the man not showing much emotion and leading to this mix-up. Here, we have to note that the man had no acting training before.

Scenario 4 Scenario 4 shows a robot interacting with a sad man, trying to encourage him (talking, hugging). Figure 7.4 shows an excerpt from the video while hugging. We expect the answers to be

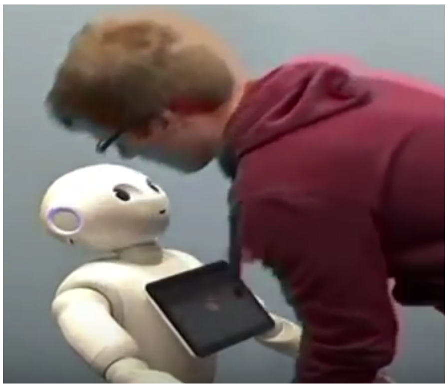


Figure 7.4.: Pepper is hugging a human.

mainly about the actions of the robot. However, most mention interaction and comforting, which is our main goal during this scenario. Still, the quotes *"Emotional support? it feels very staged though - even more than the last examples"* and *"To pretend to be interested in the human's emotional state."* show that the naturalness of the interaction is lacking a lot. Another important aspect is that not everyone may like the behaviour of the robot. This is expressed by *"gettin its hug fetish on"*, which displays that a hugging robot may seem creepy. Thus, we think for actual usage of this scenario it is important to gather information about possible interaction partner beforehand and to integrate an asking step for the hug, which we did.

7.3.2.2. Feedback about the behaviour

The previously reviewed question about the goal of the robot already contains some feedback and opinions about the interactions. Further individual but similar feedback is expected when asking only about the behaviour of the robot.

Scenario 1 We have tried to make the robot act human-like in some respect but were advised not to do so and instead use its abstraction from a human being (*"the robot shouldn't pretend to act like a human, but rather, like a cartoon character – it will make it much easier for people to accept it as a companion"*). With regard to the design of the robot, this makes sense as it is human-like but still not totally human. Some of the used gesture was decided by the autonomous mode of the robot. The quote *"the robot stretches out it's arm which makes it(??) seem a little bit "bitchy""* refers to one of those movements, reminding that you can also use too much fidgeting.

Although, the movements could be faster (*"The reaction time is a bit slow, which makes it feel a bit arkward"*) participants suggest more actions by the robot as well as more talking, also including changes of voice and face. We have to add that the used robot cannot generate facial expression but has a rigid face. Instead, we were trying to use changes of the colour of its eyes to convey emotions. However, it seems this was not perceptible in the video and someone even suggests doing this (*"Blinking and eye movement convey a big chunk of human emotions in interactions,"*). Still, this indicates the naturalness of using eyes and colours to convey emotional information.

Scenario 2 We got the proposal to time speech and movement better and so some of it simultaneously. This seems like a good idea to make the interaction faster and therefore seem less staged. Further, we should use the contextual information of distance to vary the behaviour of the robot. Like *"don't wave at people who stand directly in front of you no one does that"* states some behaviour is only making sense when having a farther distance. A possibility to avoid such situations would be to mirror the interaction partner like *"Moore copy the user to get more trustment"* suggests. In this case the robot would only do what the interaction partner would do and thus adapt to his level of acceptance.

Scenario 3 Scenario 3 is a bit longer than the previous scenarios and presents more of the autonomous mode. This may be a reason why *"There is too much movement of the robot. After every sentence he is using all of his body this feels kinda unnatural. It's very uncommon."* states too much fidgeting of the robot. Still, we can conclude that autonomous mode may be better turned off and abstaining from too much movement may be desirable. In contrast to this a *"More specific respone (answer seemed really general)"* is requested to make the dialogue more natural. Currently, this would mean to design more specific scenarios. However, a scenario is limited by contextual knowledge of the robot. We think this can be avoided by using general answers as well as by using person identification combined with a user profile to adapt possible interactions. Another aspect about the design of the dialogue is that some felt domineered by the robot as his language is too commanding, for example, *"tell me"* could be politer or friendlier. The language the robot should use depends on the kind of image we want the robot to have. Our reasoning was making it seem like a companion but this could be irritating for strangers and outsiders.

Scenario 4 Regarding scenario 4 we got similar feedback as for the other scenarios but also specific one about its hugging. The reaction time of the robot was thought of as lacking and the hugging could be more realistic while others think *"[...] it's appropriate, and even cute."* or *"I don't think hugging a robot will have the same psychological effect as hugging a fellow human."*. We think the perception depends a lot on the person and in some cases a hug may still help, though not being the same as a human's. This is also the reason why the robot first offers a hug so it does not impose itself.

Regarding the dialogue we got the feedback *The chosen words do align with the situation but the voice could be more resonating with the human's situation"*, which leads to the suggestion of using a different robot or using your own recordings because the voice of Pepper can hardly be changed.

7.3.2.3. General feedback

We gathered general feedback per scenario and regarding all scenarios at a time. The feedback is quite similar and thus grouped together. It contains complaints about the quality of video and sound, we even got a suggestion for a cheap but good microphone (at this point we would like to thank the recommender). Besides, we got the suggestion to show an interaction in a natural environment, which could be on the street, in a shop or at home. We have not done this to refrain from biasing the participants and instead generate as general feedback as possible for future interactions of all kind. Another remark mentions the voice being to high pitched and thus artificial as well as uncomfortable. We would like to combine this with previous feedback about the robot being more cartoon-like. Depending on the image the robot should have (more human-like or more cartoon-like) the voice has to be adapted. In our case, we did not want a serious robot but rather a playful one and therefore prefer the cartoon-like image.

Feedback which focuses on the interaction is mainly about the hugging. A good suggestion was cushioning the robot so the hug would be more comfortable. Another comment ,*"The robot's height is not so suitable for a hug actually. :D"*, reminded us of the sometimes big difference in height. The robot is only about 1,21m tall while the actor is significantly taller and has to bow down for hugging. This should be considered for future interactions.

We did not receive much feedback about the changing eye colour. A reason may be that it was not very well perceivable as a comment states *"please don't rate the answers I gave on color changing of the eye of the robot because I just didn't see it"*. This is supported by the previously mentioned complaints about the image quality. Therefore, during future video taping we should pay attention to lightning conditions and zooming.

7.3.2.4. Most liked

As the question states, we were asking for the most liked part of the interactions. Surprisingly, answers often focused on the hug although we received a lot of negative comments concerning it. A reason may be that the feedback questions were voluntary while this question was mandatory and thus received more answers. Another reason may be that the hug was simply the best part but not necessarily a good part.

Comments mention *"movement of arms and head"* and *"showing emotion through gesture"* is welcomed. Another comment is describing this metaphorically as *"he cares more about me than the KIT ever did"* which lets us conclude that the basic intention we had during designing, for the robot to seem empathetic, was perceivable, though some people were unaware of it.

7.3.2.5. Least liked

Regarding the answers, we expect somewhat negative ones. Again, the hug is mention, which proves it to be ambivalent. Our guess is that information about the interaction partner has to be gathered beforehand and thus enabling the decision of offering him a hug or not. However being empathetic itself proves to be ambivalent, too (*"It tries to show emotions, when it's just a machine. :-)"*). Therefore, it is important to guarantee a way to keep the robot at distance if wished. Besides, *"It does seem staged and the movements are slow. We live in a hectic world. Maybe people with more time don't mind.."* expresses concern about the acceptance of the robot. This is more related to the speed of the interaction and pushes us to think about what would be acceptable for whom. Some may think the robot acts to slow while others may feel comfortable or relaxed.

Further, the wish for more complex scenarios has been expressed. *"The limits of the interaction."* can refer to the same. Another possibility is the limitation caused by the robot as some participants do not like the voice or not showing facial expressions. In addition, the emotional reactions can be seen as useless and thus not well liked.

7.3.2.6. Wishes

We received quite different wishes, *"Faster movements"*, *"The ones the robot had were fine. However, they were stiff and could make a person feel uncomfortable."* and *"All in all I liked the robot's reactions.*

They just need to be developed a steep further in order to feel more human-like." to mention some. What they have in common is the need to further enhance the scenarios and movements to make them more appealing, especially the movements as some perceive them as unpleasant.

The wishes *"I think since I am not used to robots I wouldn't want the reactions them to be too human like."* and *"Natural, friendly, and intention/emotion-aware"* demonstrate an opposing view as some do not want the robot to be too human while other would prefer it to be more human. During actual usage it depends on the user group. Here, we want to point out again that it is important to provide a possibility to make the robot stop. Further, we received some suggestions for additional features. The robot could provide some sort of information or be like a secretary managing appointments. Another wish was having a personality. This could be done by word choice and gesture.

7.3.3. Quantitative

The *quantitative* analysis examines the ratings. First, we analyse per question. Afterwards, we describe some limitations.

7.3.3.1. General

Our online survey was conducted over the period of 3 weeks and has 72 participants. Table A.1 contains the result of the quantitative part. Additionally, we calculated *weighted sum, mean, variance* and *standard deviation*.

7.3.3.2. All scenarios

This section contains questions which were asked for each scenario.

Staged or natural All in all, our goal was to design the interactions as natural as possible. Despite this each scenario shows a tendency to be *a little staged* and scenario 2 even seems to be *staged*. When we think about the feedback from the qualitative analysis, for example, *"annoying people directly in front of it with waving"*, it seems likely that the waving of the robot is responsible for this. The overall estimation of the interaction with the robot is *a little staged* as well and shows further need of enhancement. In addition, the variance and standard deviation support this.

For interaction 1,2,3 and the overall rating the maximum value represents the tendency. However, this is not the case for scenario 4. Here the maximum is at *a little staged* and *a little natural*. This may reflect the torn opinions about the hug.

The robot is creepy On one hand, this question relates to the design of the robot. On the other hand, it relates to its behaviour. For scenario 1 and 3 the average opinion is *somewhat disagree* while the maximum is reached for *disagree* and the minimum for *agree* or *strongly agree*. It indicates a tendency, while scenario 2 and 4 get *neither agree nor disagree* as average opinion. Further, the maximum is *somewhat agree* for scenario 2 but *disagree* for scenario 4, which has a standard deviation from about 2 and is relatively uniform (see figure 7.5).

Scenario 2 and 4 contain more active body usage of the robot, which is not liked by everyone. This could explain why the robot is seen as creepier than during the other interactions. Still, in average it is not seen as *creepy* but also not as clearly not creepy.

The robot would make a good companion With respect to the results of *the robot is creepy* we expect a lightly positive tendency. However, the average is *neither agree nor disagree*. As the question refers more to the actions and the behaviour of the robot and the opinions about these are torn the results makes more sense. Nonetheless, the maximum is voted at *somewhat agree* for each scenario, which shows the general potential. At the same time this shows the lacking of our interactions as the result is a *neutral* opinion.

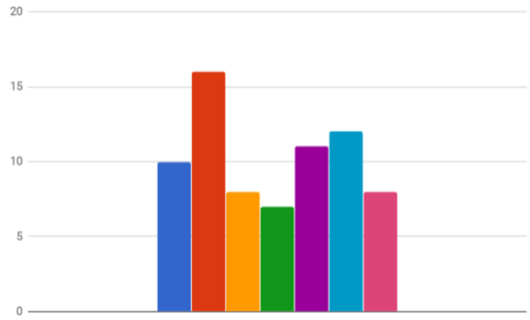
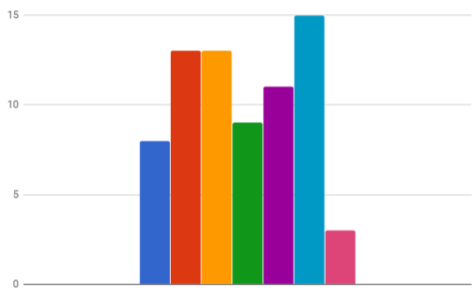


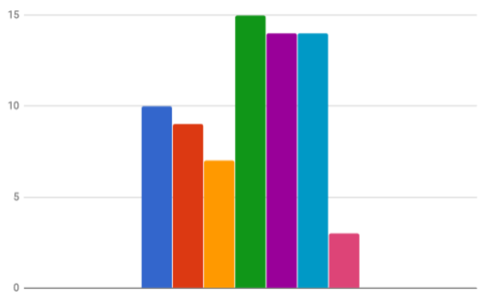
Figure 7.5.: The counts for the answer to *The robot is creepy* regarding scenario 4, from *strongly disagree* to *strongly agree*.

The voice of the robot is suitable In general, the voice of the robot seems suitable for all interactions as the average voted *somewhat agree* and further the maximum supports this as well. This is surprising because we got the feedback that *more expression in the voice* is wished. An explanation could be that because of the lack of expression the voice is not unsuitable either. Therefore, we conclude that depending on the situation changes of the voice would be preferable but the voice in general is fine as it is. However, we want to note that some participants noted that the voice is too artificial.

I would feel like the robot controls the situation In scenario 1 and 2 a human is initiating the interactions while a robot initiates them in scenario 3 and 4 and keeps them going. Thus, we guess the robot has higher control during interaction 3 and 4. This guess is supported by the maximum value but not by the average. Only for scenario 2 the average is *somewhat disagree* while it is *neither agree nor disagree* for the others. Regarding scenario 3 and 4 the standard deviation is higher (about 1,8) showing a more uniform result (see figures 7.6(a) and 7.6(b)). Those scenarios are a little bit more complex and



(a) The counts for the answer to *I would feel like the robot controls the situation* regarding scenario 3, from *strongly disagree* to *strongly agree*.



(b) The counts for the answer to *I would feel like the robot controls the situation* regarding scenario 4, from *strongly disagree* to *strongly agree*.

although the robot is asking questions the human interaction partner can still decide what and how much to answer. This could explain the higher deviation. The general opinion about all interaction together reflects *somewhat disagree*.

The answers of the robot make sense This question mainly refers to the dialogue and its adequacy. The maximum value is *agree* for all scenarios and is supported by the average with *agree* and *somewhat agree*. Further, the standard deviation for scenario 1 and 4 is about 1, resulting in a clearer opinion. Combined with the wish for more complex interactions the dialogue will have to be adapted but should still orientate by our basic scenarios.

The robot shows emotions Especially in scenario 3 and 4 we wanted the robot to express emotions (happiness and sadness). The average *somewhat agrees* for scenario 3 but *neither agrees nor disagrees* for scenario 4 and 2. However, the maximum value is *somewhat agree* in both cases. Thus, regarding scenario 4 more emotion could be expressed if wished. Since scenario 2 is a short but simple greeting we are not surprised about the result as it is so short and therefore not much time to express and perceive emotions. It is important to note that everyone would want the robot to express emotions as one participant stated (*"It tries to show emotions, when it's just a machine. :-)"*).

The eye colour of the robot changed We received a lot of feedback mentioning bad quality of our videos. Further, the colours and their changes are faint. Thus, we expect no clear opinion and a high standard deviation. For all scenarios the maximum value supports this with *neither agree nor disagree*, probably because the participants could not perceive any changes. The average states the same for scenario 3 and 4 while it states *somewhat disagree* for the others. Further, standard deviation ranges from 1,6 to 2. Based on the qualitative analysis we suggest to examine this again by filming with higher quality.

The robot reacts according to the other person's emotions In scenario 1 and 2 the actors do not express very specific emotions. In addition the interactions are very short. Therefore, we expect a neutral attitude. On the other hand, we expect clearer opinions for scenario 3 and 4. Our expectations for scenario 1 and 2 are affirmed with by the maximum value and the average (*neither agree nor disagree*). Similarly, the average *somewhat agrees* for scenario 4 with a standard deviation of 1,2. However, only the maximum supports our expectation for scenario 3 with *somehwat agree* while the average is *neither agree nor disagree*. For both cases it is important to note, that some people had difficulties to perceive the actors' emotions and stated insufficient expression of these.

The eye colours of the robot show emotions Based on the comments about bad quality of the videos and the results of *the eye colour of the robot changed* we expect neutral or disagreeing opinions. If eye colour or its changes are not perceived then they cannot express anything. This is supported by the maximum value at *neither agree nor disagree* and the average with *somewhat disagree*.

The posture and gesture of the robot show emotions We expect a positive tendency as neither posture nor gesture have to be liked to express emotions. Additionally, the question is very general by not specifying the perceived emotions so the perceived ones do not necessarily have to be the intended ones. This is supported for all scenarios by the maximum value for *somewhat agree* but only for scenario 3 and 4 by the average with *somewhat agree*. Scenario 1 and 2 have an average of *neither agree nor disagree*. All standard deviations are in the range from 1,4 to 1,7.

In regards to the greeting (scenario 2) it is possible that it is not seen as something expressing emotions and thus rated as more neutral. Nonetheless, we are surprised about scenario 1 as the robot seems rather cheerful in our opinion.

Goal accomplishment In context of scenario 1 we expect agreement because the goal was often specified as *interaction* or *wishing a happy birthday*, which is what the robot is doing. This is supported by the average opinion of *somewhat agreeing* as well as the maximum for *agreeing* and a standard deviation of 1,2.

Regarding scenario 2 *greeting* is referred to the most often. Therefore, we expect it to be seen as an accomplished goal, which is validated by the average *somewhat agreeing* and the maximum for *agreeing*.

The goal of scenario 3 is the most often referred to as *smalltalk* or *interaction* but also as other things. As the range of goals is quite various we expect less agreement but still a tendency towards it. The average is *somewhat agreeing* while the maximum value is at *agree*. However, the standard deviation (1,5) indicates that it is only a tendency and not guaranteed.

Scenario 4 is often seen as *comforting*. Hence, we would expect strong approval. However, since our actors were often commented as not very expressive it could be hard to determine whether the robot has fulfilled its goal or not. Thus, we expect mixed answers. Nonetheless, the maximum value and the average are *somewhat agree* and the standard deviation is 1,4.

Do you think people will like/dislike such a robot? [You] This question refers to the participant and to his liking of the robot all over, meaning behaviour and design. We expect slight differences per scenario but a similar tendency. In total the standard deviation for each scenario ranges from 1,8 to 1,9 and scenario 2,3 and 4 are *neither liked nor disliked* by the average. This means a wide variance. However the maximum value is *like*, which shows a positive tendency. Overall, this represents the opinion of mainly students and people from technical background.

Do you think people will like/dislike such a robot? [Elderly] This question refers to the participant's opinion about elderly liking/disliking the robot. We expect slight differences per scenario but a similar tendency. As our standard participant is between 18 and 34 the closeness to reality depends on their experience and empathy.

The standard deviation ranges from 1,7 to 1,8 and the average opinion for all scenarios is *neither like nor dislike*. The maximum value indicates a tendency to *slightly like* and *like* whereby scenario 1 and 4 are preferred. A reason in regards to scenario 2 could be that the waving is disliked.

Do you think people will like/dislike such a robot? [Children] This question refers to the participant's opinion about children liking/disliking the robot. We expect slight differences per scenario but a similar tendency to liking. As our standard participant is between 18 and 34 the closeness to reality depends on their experience and empathy.

The standard deviation is between 0,8 and 1,2 with an average of *like*, which shows a relatively uniform opinion. This may be based on the "cartoon-like" design of the robot and the image people have of children.

Do you think people will like/dislike such a robot? [Tech-savvy] This question refers to the participant's opinion about tech-savvy people liking/disliking the robot. We expect slight differences per scenario but a similar tendency to liking. As our user group is mainly perceiving themselves as tech-savvy this should be an accurate representation. Still, the standard deviation ranges from 1,4 to 1,5. The expected tendency towards liking is confirmed by an average of *somewhat liking*.

Do you think people will like/dislike such a robot? [Tech-averse] This question refers to the participant's opinion about tech-averse people liking/disliking the robot. We expect slight differences per scenario but a similar tendency to disliking. As our user group is mainly perceiving themselves as tech-savvy this is probably not representative.

The standard deviation ranges from 1,6 to 1,8 and the average is *somewhat dislike* as expected. In addition the maximum value for scenario 4 is *strongly dislike* and thus more negative than for the other scenarios. This could be caused by the dislike of the hugging. At the same time the standard deviation is the highest (1,8), which could represent the torn opinions about it.

Do you think people will like/dislike such a robot? [Others apart from you] This question refers to the participant's opinion about others liking/disliking the robot. We expect slight differences per scenario but a similar tendency. The closeness to reality depends on the participants' environment as well as their experience and empathy. We expect a tendency towards liking based on the participants estimation of their surrounding as affectionate towards technical developments.

The standard deviation is between 1,1 and 1,3. In contrast to our expectation the average opinion is *neither like nor dislike*. The same goes for the maximum value in exception for scenario 1 where it is *slightly like*.

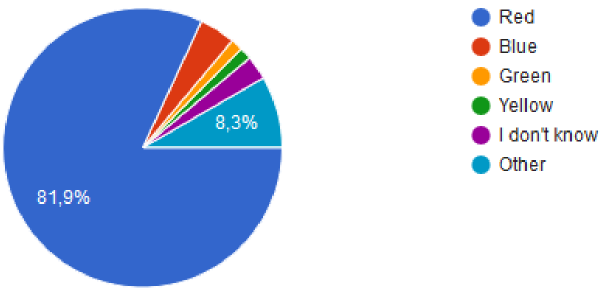


Figure 7.6.: The results for the question *"Which colour is closest to the woman's shirt?"*.

7.3.3.3. Only scenario 1

This section contains questions which are only in section 1.

I could hear a song We expect clear disagreement because of insufficient quality. Surprisingly, the maximum value *agrees* and further the average *somewhat agrees* with a standard deviation of 1,7. It is possible that a song was recognisable despite the quality but no details were understood, which would result in agreement.

I could see balloons on the tablet This question is asking about a very specific detail, not just about seeing something but balloons. Thus, and because of reflections we expect disagreement and a high standard deviation. As expected, the maximum value *strongly disagrees* and the average *disagrees*. Further, the standard deviation is 2,2, which is rather surprising because of many commenting about the insufficient quality of the video. In contrast to other questions these answers are stronger, not only showing a tendency. Therefore, in the future the video taping should be done more careful about reflections.

The robot is wishing a happy birthday We expect high agreement and low standard deviation. This is based on the the comments about the goal of the interaction, which show that it was often perceived as *wishing a happy birthday*. The maximum value *strongly agrees* while the average *agrees*. Despite this and our guess the standard deviation is about 1,9. A reason could be lack of attention or bad quality of the audio.

The robot seems happy During the design of this interaction we intended to make the robot appear happy. Thus, we expect agreement and a low standard deviation. This is validated by the maximum value as well as by the average with *somewhat agreeing*. Further, the standard deviation is 1,3. Therefore, we conclude that the robot is clearly perceived as happy.

Playing a song is a good idea At the beginning of the interaction the robot is playing a birthday song. We gathered opinions about this action. However, the maximum value is *neither agree nor disagree*. Nonetheless, the average somewhat agrees but the standard deviation is 1,7. Therefore, we can only conclude a slight tendency that using a song is a good idea. Still, it may depend on the song.

7.3.3.4. Only scenario 2

This section contains a question which is only used regarding scenario 2.

Which colour is closest to the woman's shirt? The correct answer would be something like berry, which is not available. As most similar colours we expect *red* and *blue* or alternatively *other* and *I don't know* as most selected. The received answers are depicted in figure 7.6. Almost 82% selected *red*. Surprisingly, *yellow* and *green* were each selected by 1,4% of the participants. We guess that some people do not focus on things like the colour of a shirt or did just guess as the lighting conditions should have been sufficient.

7.3.3.5. Only scenario 3

This section contains questions only used for scenario 3. They refer to its behaviour and its change dependent on the interaction partner's mood. Our intention was to make the robot appear happy.

The robot seems sad after hearing that the other person is happy We expect a tendency towards disagreement as the statement is the opposite of our intended goal. This is confirmed by the maximum value for *disagreeing* as well as by the average *somewhat disagreeing* and the standard deviation (1,3).

The robot seems happy after hearing that the other person is happy We expect a positive tendency. Both the maximum and the average *somewhat agree* with a standard deviation of 1,3, confirming the expected tendency. Combined with the previous question we can conclude, that the robot appears happy.

7.3.3.6. Only scenario 4

This section contains questions only referring to scenario 4.

I would hug the robot in this situation Based on the torn opinions about the hug, described in the qualitative analysis (section 7.3.2) we expect a similar result with high standard deviation. This is confirmed by a score of 2,2 and the average of *neither agree nor disagree* while the maximum *strongly disagrees*.

The robot seems sad after hearing that the other person is sad We intended for the robot to appear sad and therefore expect a tendency for *agreement*. In fact, a maximum at *agree* confirms this as well as the average (*agree*). Thus, we can conclude that the robot expresses sadness. Additionally, combined with the result that the posture and gesture expresses emotions we can further conclude that it is also accountable for this.

The robot seems happy after hearing that the other person is sad As we intended to make the robot appear sad we expect a tendency for disagreement. The maximum and the average for *disagree* support our thesis and further show that our implemented behaviour is not being perceived as happiness.

7.3.3.7. Overall

This section contains questions regarding all scenarios together.

Would you feel comfortable using the robot? The average is *neither comfortable nor uncomfortable* and the standard deviation 1,9, which results in a uniform distribution (see figure 7.7) with the maximum value for *comfortable*.

I would feel watched by the robot The average *neither agrees nor disagrees*. With a standard deviation of about 1,9 and the maximum *somewhat agree* a wide variance in opinion can be concluded.

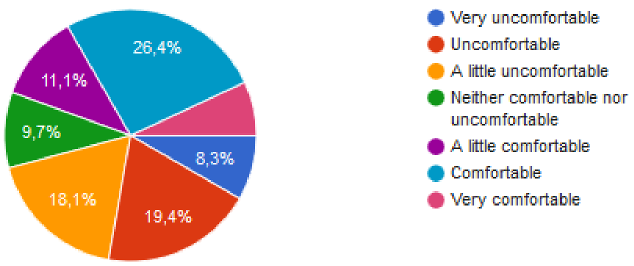


Figure 7.7.: The results for the question *"Would you feel comfortable using the robot?"*.

I would feel like having control over the robot With a standard deviation of 1,5 and the average *somewhat agreeing* we conclude a positive tendency. This confirms the results of the previous questions *I would feel like the robot controls the situation*, which was not comfirmed.

The interaction works fine for short people In two of our videos a short person interacted with the robot while it was a tall person for the other two videos. This probably influences the opinions. The average is *somewhat agree* with a standard deviation of 1,1.

I would feel comfortable in the robot's company We are expecting a similar result as for the question *Would you feel comfortable using the robot?* because it implies the robot's company. This is confirmed by an average of *neither agree nor disagree* and a standard deviation of 1,7 as well as the maximum for *somewhat agree*. Thus, we conclude that the company of the robot is at least not annoying.

The interaction works fine for tall people In two of our videos a short person interacted with the robot while it was a tall person for the other two videos. This probably influences the opinions.

The average *neither agrees nor disagrees* with a standard deviation of 1,5. The maximum is *somewhat disagree*, which shows a negative tendency. This could be caused because we use a tall actor to hug a tiny robot, which might seem inappropriate.

Do you think a robot should react towards your current emotion/mood? This question is quite general as no special reactions were specified. The average is *yes* with a standard deviation of 0,4, which indicates the wish for a reaction in general.

I think that I would like to use this robot frequently Based on the questions regarding the comfortability of the robot's attendance we expect a neutral or slightly positive tendency. However, the opposite is the case. The average is *somewhat agree* with a standard deviation of about 2.

I found the robot unnecessarily complex We expect a tendency towards disagreeing based on the simple interface of a dialogue. An average of *somewhat disagree* and a standard deviation of 1,4 confirm this.

I thought the robot was easy to use We expect a tendency towards agreeing based on the simple interface of a dialogue. The average of *somewhat agree* and the standard deviation of 1,1 validate this.

I think that I would need the support of a technical person to be able to use this robot In regards to the previous questions about the complexity of the robot we expect a tendency towards disagreeing. In fact the average and the maximum are *disagree* with a standard deviation of 1,3.

I found the various function of this robot very well integrated The average is *somewhat agreeing* with a standard deviation of 1,1.

I thought there was too much inconsistency in this robot The average is *somewhat disagree* with a standard deviation of 1,4.

I would imagine that most people would learn to use this robot quickly We expect a tendency towards agreement because the interface seems intuitive. The average *agrees* with a standard deviation of 1,1.

I found the robot very cumbersome to use We expect a negative tendency because of the intuitive interface. However, the average is *neither agree nor disagree* with standard deviation of 1,5.

I felt very confident using the robot We expect a neutral tendency. Our reasoning is that none of the participants has actually used the robot and is just guessing. This is supported by the average of *neither agree nor disagree* and a standard deviation of 1,3.

7.3.3.8. Limitations

In general our survey is limited by the quality of the videos and their sound. As the suggestion of one participant for a better microphone and further comments state both are insufficient and therefore limiting the precision of the feedback. Further, the same goes for our actors as they have no experience in this field and could be more expressive to ease the task of the participants.

Originally, our use cases were meant to take place in the context of elderly and care homes. However, our group of participants is mainly represented by students (between 18 am 34), who are commonly not living in care homes. Therefore, the survey reflects their opinion and provides guidance about general acceptance of the interactions and the robot. Nonetheless, is important and gives advice for the future when the participants will become elderly. Besides, it is not only important for the interaction partners to accept the robot but for their surroundings, friends and family, too. Another restriction is that most participants represent the technical field and perceive themselves as well as their surrounding as attracted to technical developments, which could be biased caused by the group of participants.

Our interactions themselves are very limited because of their shortness and simplicity. They only provide rough guidance like a pre-study.

8. Conclusion and future work

We first have presented some background information about emotion recognition, human-computer interaction and especially about emotional feedback. Furthermore, we have designed some interactions and implemented them. To analyse them we have conducted a survey and evaluated them.

We have examined the Affectiva SDK version 3.4.1 in various aspects and provide a baseline for comparison with other software for emotion recognition. Hereby, we have described our framework and a possible way of connecting it. The evaluation includes the detection of rotated faces, using the *Head Pose Image Database*, which works well for about +- 30 degrees of tilt and pan. Further, this includes emotion recognition. By using the *Karolinska Directed Emotional Faces* we were able to roughly conclude the performance of the emotion recognition for rotated faces. A rotation of 45 degrees is already problematic even if a face is detected.

By using the *extended Cohn-Kanade* dataset and the *Japanese Female Face Expression* database we have analysed the emotion recognition in general. For both of them, the emotion *happiness* is recognised quite well. Further, *surprise* and *disgust* are decently recognised. In contrast, *anger* and *fear* are poorly recognised.

When directly comparing the recognition of western and Japanese faces we can see that in both cases happiness is well recognised and in general the emotions of western faces are better recognised. However, our database for Japanese faces is limited by a small sample size as well as by all of them being female. Therefore, it is only an indication that the emotions of western faces are better recognised and the same experiment should be done again with a database containing a broader variety. In fact, all experiments should be repeated by using other software and thus provide guidelines easing the choice which to use. Thereby, our experiments could be complemented by other experiments, for example, an evaluation of processing time.

Besides the evaluation, we have identified aspects which should be considered for the generation of empathetic feedback. Additionally, we have composed possible use cases taking place in the context of elderly and care homes. Some basic use cases were implemented and evaluated through an online survey.

Through the analysis of our survey we can conclude that people wish a robot to react according to their emotions. However, the degree of reaction differs and some definitely do not like the idea of an empathetic robot while others do. In this case, a functional robot might be sufficient. Still, further enhanced scenarios should be designed and evaluated to decide about the usage of an empathetic robot. In addition, especially the naturalness of the robot should be increased.

During our interactions, emotional expression by the robot can be perceived although some of our used methods could be perceived only badly. We have used posture, gesture, colours and behaviour. Posture and gesture, as well as behaviour. could be partly confirmed.

Although, our intention was to compose interactions which could be used in a care home most of our participants are between 18 and 34. Thus, the acceptance and opinions do not reflect our original target group. Still, our user group does not believe elderly to dislike the robot but also not to like it. Further, the participants estimate the attitude of their environment similarly although it was estimated as accepting of technical developments. On the other hand, children are generally seen as accepting of the robot. Overall, our survey equals a pre-study and provides guidelines for further research about empathetic robots and we suggest to do so for multiple user groups.

Bibliography

[1] Non-verbal communication. URL https://digwork.com/index.php/resources/124-non-verbal-communicationl. (visited on 2018-01-16). 10

[2] Asimo the world's most advanced humanoid robot. URL http://asimo.honda.com/Inside-ASIMO/. (visited on 2018-04-10). 8

[3] Fanuc 6-axis welding robots. URL https://en.wikipedia.org/w/index.php?curid=23045059. (visited on 2018-01-16). 7

[4] Seneca lucius annaeus-grundzüge seiner philosophie. URL http://www.pinselpark.org/philosophie/s/seneca/stoa/philo_grund.html. (visited on 2018-01-16). 11

[5] boost. URL http://www.boost.org/. (visited on 2018-01-24). 29

[6] Beschäftigte in der Pflege. URL https://www.bundesgesundheitsministerium.de/themen/pflege/pflegekraefte/beschaeftigte/?L=0. (visited on 2018-05-04). 4

[7] . URL https://buy.jibo.com/products/jibo. (visited on 2018-01-18). 11

[8] . URL https://www.jibo.com/. (visited on 2018-01-18). 11

[9] Technical overview — Aldebaran 2.5.9.8-r3 documentation. URL http://doc.aldebaran.com/2-5/family/pepper_technical/index_pep.html. (visited on 2018-04-24). 25, 26

[10] Opencv about. URL https://opencv.org/about.html. (visited on 2018-01-23). 29

[11] Rapidjson documentation. URL https://miloyip.github.io/rapidjson/. (visited on 2018-01-24). 30

[12] Andere Länder, andere Gesten, February 2010. URL http://www.sueddeutsche.de/karriere/bedeutung-von-handzeichen-andere-laender-andere-gesten-1.54682. (visited on 2018-02-02). 2

[13] Robert plutchik's wheel of emotions, February 2011. URL https://commons.wikimedia.org/w/index.php?curid=13285286. (visited on 2018-01-15). 9

[14] 6 Ways Your Brain Transforms Sound Into Emotion, Aug. 2015. URL https://www.brightaudiology.com/6-ways-your-brain-transforms-sound-into-emotion/. (visited on 2018-02-09). 20

[15] affdex namespace reference, March 2017. URL https://knowledge.affectiva.com/v4.0.0/docs/getting-started-with-the-emotion-sdk-for-windows. (visited on 2018-01-22). 30

[16] Oxford dictionary, 2018. URL https://en.oxforddictionaries.com. (visited on 2018-01-15). 8

[17] F. M. Adams and C. E. Osgood. A cross-cultural study of the affective meanings of color. *Journal of Cross-Cultural Psychology*, 4, June 1973. 18, 19

[18] Affectiva. Deep learning, . URL https://www.affectiva.com/how/deep-learning-at-affectiva/. (visited on 2018-01-23). 15, 30

[19] Affectiva. Metrics, . URL https://developer.affectiva.com/metrics/. (visited on 2018-01-26). 15

[20] Affectiva. Sdk & api, . URL https://www.affectiva.com/product/emotion-sdk/. (visited on 2018-01-23). 30

[21] Affectiva. Windows sdk developer guide release 3.4.1, . URL https://knowledge.affectiva.com/v4.0.0/docs/getting-started-with-the-emotion-sdk-for-windows. (visited on 2018-01-22). 29, 30

[22] Affectiva. Affectiva, 2018. URL https://developer.affectiva.com. (visited on 2018-01-02). 51

[23] J. An, T. Li, Y. Teng, and P. Zhang. Factors Influencing Emoji Usage in Smartphone Mediated Communications. In *Transforming Digital Worlds*, Lecture Notes in Computer Science, pages 423–428. Springer, Cham, Mar. 2018. ISBN 978-3-319-78104-4 978-3-319-78105-1. doi: 10.1007/978-3-319-78105-1_46. URL https://link.springer.com/chapter/10.1007/978-3-319-78105-1_46. 17

[24] R. C. Arkin, M. Fujita, T. Takagi, and R. Hasegawa. An ethological and emotional basis for human-robot interaction, 2003. 12

[25] J.-A. Bachorowski. Vocal expression and perception of emotion. *Current Directions in Psychological Science*, 8(2):53–57, 1999. doi: 10.1111/1467-8721.00013. 17, 18

[26] M. Bayer, W. Sommer, and A. Schacht. Font Size Matters—Emotion and Attention in Cortical Responses to Written Words. *PLOS ONE*, 7(5):e36042, Sept. 2012. ISSN 1932-6203. doi: 10.1371/journal.pone.0036042. (visited on 2018-02-08). 19

[27] J. Brooke. Sus: A retrospective. *Journal of usability studies*, 8(2):29–40, february 2013. 46

[28] S. J. Burton, A.-A. Samadani, R. Gorbet, and D. Kulic. *Laban Movement Analysis and Affective Movement Generation for Robots and Other Near-Living Creatures*, pages 25–48. Springer, Cham, 2016. ISBN 978-3-319-25737-2 978-3-319-25739-6. 16

[29] I. Cohen, R. Looije, and M. A. Neerincx. Child's recognition of emotions in robot's face and body. pages 123–124, March 2011. 12, 16, 17

[30] M. Coulson. Attributing Emotion to Static Body Postures: Recognition Accuracy, Confusions, and Viewpoint Dependence. *Journal of Nonverbal Behavior*, 28(2):117–139, June 2004. ISSN 0191-5886, 1573-3653. doi: 10.1023/B:JONB.0000023655.25550.be. 16, 17

[31] R. W. Cowdry, D. L. Gardner, K. M. O'Leary, E. Leibenluft, and D. R. Rubunow. Mood variability: a study of four groups. *American Journal of Psychiatry*, 148(11):1505–1511, Nov. 1991. ISSN 0002-953X. doi: 10.1176/ajp.148.11.1505. URL https://ajp.psychiatryonline.org/doi/abs/10.1176/ajp.148.11.1505. (visited on 2018-02-10). 21

[32] D. Cyr, M. Head, and H. Larios. Colour appeal in website design within and across cultures: A multi-method evaluation. *International Journal of Human-Computer Studies*, 68(1):1 – 21, 2010. ISSN 1071-5819. doi: https://doi.org/10.1016/j.ijhcs.2009.08.005. URL http://www.sciencedirect.com/science/article/pii/S1071581909001116. 18

[33] G. Dalakov. Jacques de vaucanson. URL http://history-computer.com/Dreamers/Vaucanson.htmll. (visited on 2018-01-17). 7

[34] M. M. de Graaf and S. B. Allouch. Exploring influencing variables for the acceptance of social robots. *Robotics and Autonomous Systems*, 61:1476–1486, 2013. 12

[35] K. Dorscheid. Gestik: Gesten aus aller Welt. URL http://www.geo.de/GEOlino/mensch/gestik-kultur-mal-anders-gesten-aus-aller-welt-59416.html. (visited on 2018-02-02). 2

[36] P. Ekman. An argument for basic emotions. pages 169–200, 1992. 2, 8, 17

[37] P. Ekman and W. V. Friesen. Facial action coding system facs, 1977. 9, 17

[38] P. Ekman, W. V. Friesen, and P. Ellsworth. *Emotion in the Human Face: Guidelines for Research and an Integration of Findings*. Elsevier, Oct. 2013. ISBN 978-1-4831-4763-5. Google-Books-ID: dOFFBQAAQBAJ. 2, 17

[39] K. El-Maleh, M. Klein, G. Petrucci, and P. Kabal. Speech/music discrimination for multimedia applications. In *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, volume 6, pages 2445–2448 vol.4, 2000. doi: 10.1109/ICASSP.2000.859336. 20

[40] T. Fong, I. Nourbakhsh, and K. Dautenhahn. A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42:143–166, 2003. 12

[41] J. Goetz, S. Kiesler, and A. Powers. Matching robot appearance and behavior to tasks to improve human-robot cooperation. In *The 12th IEEE International Workshop on Robot and Human Interactive Communication, 2003. Proceedings. ROMAN 2003.*, pages 55–60, October 2003. 18, 26

[42] N. Gourier, D. Hall, and J. L. Crowley. Estimating face orientation from robust detection of salient facial strucutres. *Proceedings of Pointing, ICPR, International Workshop on Visual Observation of Deictic Gestures*, 2004. 49

[43] R. Gray. Pepper grows up! 'emotional' humanoid becomes the first robot to enrol at a school in japan, April 2016. URL http://www.dailymail.co.uk/sciencetech/article-3540307/Pepper-grows-Emotional-humanoid-robot-enroll-SCHOOL-Japan.html. (visited on 2018-01-05). 25, 26

[44] E. Griffin. *A first Look at Communication Theory*, chapter 6 Proxemic Theory. McGraw-Hill. 19

[45] M. M. Gross, E. A. Crane, and B. L. Frederickson. Methodology for assessing bodily expression of emotion. *Journal of Nonverbal Behavior*, pages 223–248, 2010. doi: 10.1007/s10919-010-0094-x. 16

[46] A. S. Group. *Beginner's Guide Pepper*. March 2015. 25, 26

[47] S. A. A. Hakami. The Importance of Understanding Emoji: An Investigative Study. page 20, 2017. 17

[48] E. T. Hall. *The Silent Language*. 1959. 19

[49] J. M. Iverson, O. Capirci, V. Volterra, and S. Goldin-Meadow. Learning to talk in a gesture-rich world: Early communication in Italian vs. American children. *First Language*, 28(2):164–181, May 2008. ISSN 0142-7237. 2, 16

[50] E. S. Jee, S. Y. Park, C. H. Kim, and H. Kobayashi. Composition of musical sound to express robot's emotion with intensity and synchronized expression with robot's behavior. In *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, pages 369–374, Sept. 2009. doi: 10.1109/ROMAN.2009.5326258. 20

[51] N. Kaya and H. H. Epps. Relationship between color and emotion: A study of college students. *College Student Journal*, September 2004. 18, 19

[52] C. D. Kidd, W. Taggart, and S. Turkle. A sociable robot to encourage social interaction among the elderly. *Proceedings of the 2006 IEEE International Conference on Robotics and Automation*, May 2006. 12

[53] R. Kirby, J. Forlizzi, and R. Simmons. Affective social robots. *Robotics and Autonomous Systems*, 58:322–332, 2010. 12

[54] M. L. Knapp, J. A. Hall, and T. G. Horgan. *Nonverbal Communication in Human Interaction*. Monica Eckman, 8 edition, 2014, 2010, 2007. 2, 18

[55] K. Larson, R. Kevin Picard, L. , and R. Picard. The aesthetics of reading. 02 2018. 19

[56] A. Lichtenstein, A. Oehme, S. Kupschick, and T. Jürgensohn. Comparing two emotion models for deriving affective states from physiological datacomparing two emotion models for deriving affective states from physiological data. pages 35–50. 2

[57] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews. the extended cohn-kanade dataset (ck+): A complete dataset for aktion unit and emotion-specified expression. 50

[58] D. Lundqvist, A. Flykt, and A. Öhman. The karolinska directed emotional faces - kdef. 1998. 49

[59] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba. Coding facial expressions with gabor wavelets. *Third IEEE International Conference on Automatic Face and Gesture Recognition*, pages 200–205, April 1998. 50

[60] J. Malek, J. Zdansky, and P. Cerva. Robust automatic recognition of speech with background music. In *2017 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5210–5214, March 2017. doi: 10.1109/ICASSP.2017.7953150. 20

[61] B. Manav. Color-emotion associations and color preferences: A case study for residences. *Color Research & Application*, 32(2):144–150, Apr. 2007. ISSN 1520-6378. doi: 10.1002/col.20294. (visited on 2018-02-08). 19

[62] A. Mehrabian. *Silent Messages*. Wadsworth Publishing Company, 1971. 3, 8, 16

[63] I. R. Murray and J. L. Arnott. Toward the simulation of emotion in synthetic speech: A review of the literature on human vocal emotion. *The Journal of the Acoustical Society of America*, 93(2): 1097–1108, 1993. doi: 10.1121/1.405558. URL `https://doi.org/10.1121/1.405558`. 17, 18

[64] Y. Nishimura. A SOCIOLINGUISTIC ANALYSIS OF EMOTICON USAGE IN JAPANESE BLOGS: VARIATION BY AGE, GENDER, AND TOPIC. *AoIR Selected Papers of Internet Research*, 5(0), Dec. 2016. ISSN 2162-3317. URL `https://spir.aoir.org/index.php/spir/article/view/1137`. (visited on 2018-04-15). 17

[65] N. Novielli, F. de Rosis, and I. Mazzotta. User attitude towards an embodied conversational agent: Effects of the interaction mode. *Journal of Pragmatics*, 42:2385–2397, 2010. 1

[66] OECD. Elderly population (indicator. URL `https://data.oecd.org/pop/elderly-population.htmr`. (visited on 2018-01-31). 4

[67] R. O´Hare. Pepper the robot begins work in belgian hospitals: Friendly droid is being used to greet patients and reception, June 2016. URL `http://www.dailymail.co.uk/sciencetech/article-3641468/Pepper-robot-finds-job-healthcare-friendly-droid-trialled-two-hospitals-Belgium.html`. (visited on 2018-01-05). 25

[68] A. K. Pandey, R. Gelin, R. Alami, R. Viry, A. Buendia, R. Meertens, M. Chetouani, L. Devillers, M. Tahon, D. Filliat, Y. Grenier, M. Maazaoui, F. L. Abderrahmane Kheddar, and L. F. Duval. Romeo2 project: Humanoid robot assistant and companion for everyday life: I. situation assesment for social intelligence. pages 140–147, November 2014. 12

[69] R. W. Picard. Affective computing for hci. *Proceeding of HCI*, August 1999. 1

[70] R. W. Picard, E. Vyzas, and J. Healey. Toward machine emotional intelligence: Analysis of affective physiological state. *IEEE Transactions on Patten Analysis and Machine Intelligence*, 23, October 2001. 2

[71] B. Reeves and C. Nass. *The Media Equation*. The Center for the Study of Language and Information Publications, September 1996. 1

[72] C. Reynolds and R. W. Picard. Designing for affective interactions. 2

[73] J. A. Russell. A circumplex model of affect. *Journal of Personality and Social Psychology*, 39: 1161–1178, 1980. 8

[74] A. Shaikh, B. Chaparro, and D. Fox. Perception of Fonts: Perceived Personality Traits and Uses, Feb. 2006. URL `http://usabilitynews.org/perception-of-fonts-perceived-personality-traits-and-uses/`. (visited on 2018-02-08). 19

[75] SoftBankRobotics. Gallery, . URL `https://www.ald.softbankrobotics.com/en/press/gallery/pepper`. (visited on 2018-01-29). 25

[76] SoftBankRobotics. Find out more about pepper, . URL `https://www.ald.softbankrobotics.com/en/robots/pepper/find-out-more-about-pepper`. (visited on 2018-01-29). 25

[77] TakanishiLaboratory. Anthropomorphic flutist robot wf-4riv. URL `http://www.takanishi.mech.waseda.ac.jp/top/research/music/flute/wf_4riv/index.htm`. (visited on 2018-01-17). 7, 8

[78] T. Vogt, E. André, and J. Wagner. Automatic recognition of emotions from speech: A review of the literature and recommendations for practical realisation. pages 75–91. 2

[79] K. Wada, T. Shibata, T. Saito, and K. Tanie. Analysis of factors that bring mental effects to elderly people in robot assisted activity. *Proceedings of the 2002 IEEE//RSJ International Conference on Intelligent Robots and Systems*, October 2002. 12

[80] H. G. Wallbott. Bodily expression of emotion. *European Journal of Social Psychology*, 28 (6):879–896, 11 1998. ISSN 1099-0992. doi: 10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W. 16, 17

[81] P. Weingartner. Normative Principles of Rational Communication. *Erkenntnis (1975-)*, 19:405–416, 1983. ISSN 0165-0106. 1

[82] what-when how. Facial expression recognition (face recognition techniques) part 1. URL `http://what-when-how.com/face-recognition/facial-expression-recognition-face-recognition-techniques-part-1/`. (visited on 2018-01-16). 9, 10

[83] S. K. Whitbourne. 4 Ways to Improve Your Emotional Communication, December 2014. URL `http://www.psychologytoday.com/blog/fulfillment-any-age/201412/4-ways-improve-your-emotional-communication`. (visited on 2018-02-03). 1, 3

[84] B. Wrede, S. Kopp, K. Rohlfing, M. Lohse, and C. Muhl. Appropriate feedback in asymmetric interactions. *Journal of Pragmatics*, 42:2369–2384, 2010. 21

[85] A. Yamazaki, K. Yamazaki, M. Burdelski, Y. Kuno, and M. Fukushima. Coordination of verbal and non-verbal actions in human–robot interaction at museums and exhibitions. *Journal of Pragmatics*, 42:2398–2414, 2010. 16

A. Appendix

Table A.1.: The results of our online survey. Each section is spilt by a line. To calculate weighted sum, mean, variance and standard deviation the most left column is mapped to one. The following columns complete the series of numbers. If the are assigned a different number then the number is written in brackets. Age has the ranges *under 18*, *18-24*, *25-34*, *35-44*, *45-54*, *55-64* and *65+*. Regarding yes-no questions the first answer is no. For gender the answers are *male*, *female* and *other*. The education levels are *high school graduate*, *bachelors*, *masters*, *PhD*, *MBA*, *other*. The fields of highest degree are *business*, *computing*, *engineering*, *healthcare*, *mathematics*, *medicine*, *science*, *technology* and *other*. The current deployment can be *student*, *employed for wages*, *self-employed*, *out of work*, *unable to work* and *retired*. Further details about the questions and answers can be found in section 6.2.

										SUM	weighted sum	mean	variance	standard deviation
Do you think the robot acts staged or natural?	1	18	22	5	18	7	1			72	262	3,639	2,064	1,437
[The robot is creepy.]	9	27	8	8	17	3	0			72	222	3,083	2,299	1,516
[The robot would make a good companion.]	4	12	13	10	21	11	1			72	285	3,958	2,373	1,541
[The voice of the robot is suitable.]	4	7	9	6	21	22	3			72	327	4,542	2,609	1,615
[I would feel like the robot controls the situation.]	12	25	10	8	10	7	0			72	216	3,000	2,556	1,599
[The answers of the robot make sense.]	0	1	1	6	15	38	11			72	409	5,681	0,940	0,969
[The robot shows emotions.]	9	17	15	10	15	5	1			72	240	3,333	2,444	1,563
[The eye colour of the robot changed]	15	13	3	31	5	3	2			72	231	3,208	2,498	1,581
[The robot reacts according to the other person's emotions.]	1	12	13	20	19	4	3			72	284	3,944	1,830	1,353
[The eye colours of the robot show emotions.]	13	15	8	27	7	1	1			72	223	3,097	2,032	1,426
[The posture and gesture of the robot show emotions.]	4	7	9	5	31	11	5			72	321	4,458	2,470	1,572
[I could hear a song.]	4	4	3	5	14	23	19			72	382	5,306	2,907	1,705
[I could see balloons on the tablet.]	16	6	8	9	10	11	12			72	288	4,000	4,694	2,167
[The robot is wishing a happy birthday.]	6	4	1	3	8	16	34			72	403	5,597	3,685	1,920
[The robot seems happy.]	1	5	2	15	28	14	7			72	350	4,861	1,731	1,316
[Playing a song is a good idea.]	3	6	8	18	10	14	13			72	336	4,667	2,917	1,708
Do you think the robot accomplishs its goal?	0	5	4	0	25	34	4			72	379	5,264	1,472	1,213

Table A.1 continued from previous page

Do you think people will like/dislike such a robot? Please rate the following groups. [You]	4	7	14	2	15	22	8			72	331	4,597	3,157	1,777
Do you think people will like/dislike such a robot? Please rate the following groups. [Elderly]	5	12	14	9	12	16	4			72	291	4,042	3,040	1,744
Do you think people will like/dislike such a robot? Please rate the following groups. [Children]	0	1	0	1	4	39	27			72	449	6,236	0,653	0,808
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-savvy]	1	3	5	9	19	22	13			72	376	5,222	1,978	1,407
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-averse]	11	23	9	11	7	9	2			72	231	3,208	2,998	1,732
Do you think people will like/dislike such a robot? Please rate the following groups. [Others apart from you]	0	4	8	24	25	10	1			72	320	4,444	1,164	1,079
Do you think the robot acts staged or natural?	19	22	16	0	10	4	1			72	192	2,667	2,528	1,590
Please read the following statements and rate them. [The robot is creepy.]	6	14	9	7	24	8	4			72	285	3,958	2,929	1,711
Please read the following statements and rate them. [The robot would make a good companion.]	2	12	14	14	19	8	3			72	288	4,000	2,194	1,481
Please read the following statements and rate them. [The voice of the robot is suitable.]	1	4	9	4	24	23	7			72	359	4,986	1,986	1,409
Please read the following statements and rate them. [I would feel like the robot controls the situation.]	16	29	10	9	4	3	1			72	185	2,569	2,051	1,432
Please read the following statements and rate them. [The answers of the robot make sense.]	1	3	4	3	15	24	22			72	404	5,611	2,043	1,429
Please read the following statements and rate them. [The robot shows emotions.]	4	11	10	11	26	8	2			72	292	4,056	2,302	1,517

Table A.1 continued from previous page

Please read the following statements and rate them. [The eye colour of the robot changed]	14	14	3	27	7	4	3			72	239	3,319	2,801	1,674
Please read the following statements and rate them. [The robot reacts according to the other person's emotions.]	7	12	11	22	12	7	1			72	261	3,625	2,234	1,495
Please read the following statements and rate them. [The eye colours of the robot show emotions.]	17	14	5	28	8	0	0			72	212	2,944	1,969	1,403
Please read the following statements and rate them. [The posture and gesture of the robot show emotions.]	6	5	9	8	25	13	6			72	320	4,444	2,775	1,666
Do you think the robot accomplishs its goal?	1	6	4	5	20	24	12			72	373	5,181	2,231	1,494
Which colour is closest to the woman's shirt?	59	3	1	1	2	6				72	118	1,639	2,342	1,530
Do you think people will like/dislike such a robot? Please rate the following groups. [You]	4	14	9	9	14	15	7			72	304	4,222	3,256	1,804
Do you think people will like/dislike such a robot? Please rate the following groups. [Elderly]	6	12	11	6	18	16	3			72	294	4,083	3,076	1,754
Do you think people will like/dislike such a robot? Please rate the following groups. [Children]	0	1	4	4	7	29	27			72	428	5,944	1,414	1,189
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-savvy]	1	5	6	7	23	21	9			72	361	5,014	2,069	1,438
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-averse]	12	18	14	11	10	4	3			72	229	3,181	2,759	1,661
Do you think people will like/dislike such a robot? Please rate the following groups. [Others apart from you]	1	4	4	33	20	9	1			72	314	4,361	1,175	1,084
Do you think the robot acts staged or natural?	14	12	18	3	15	9	1			72	240	3,333	3,056	1,748
Please read the following statements and rate them. [The robot is creepy.]	6	24	13	9	9	9	2			72	242	3,361	2,731	1,652

Table A.1 continued from previous page

Please read the following statements and rate them. [The robot would make a good companion.]	6	9	13	5	19	15	5			72	303	4,208	3,109	1,763
Please read the following statements and rate them. [The voice of the robot is suitable.]	2	5	8	6	23	24	4			72	347	4,819	2,120	1,456
Please read the following statements and rate them. [I would feel like the robot controls the situation.]	8	13	13	9	11	15	3			72	275	3,819	3,231	1,798
Please read the following statements and rate them. [The answers of the robot make sense.]	1	1	10	4	14	34	8			72	379	5,264	1,805	1,344
Please read the following statements and rate them. [The robot seems sad after hearing that the other person is happy.]	12	32	12	11	3	1	1			72	184	2,556	1,580	1,257
Please read the following statements and rate them. [The robot seems happy after hearing that the other person is happy]	2	4	6	16	22	20	2			72	336	4,667	1,778	1,333
Please read the following statements and rate them. [The robot shows emotions.]	3	6	9	10	26	17	1			72	321	4,458	2,054	1,433
Please read the following statements and rate them. [The eye colour of the robot changed]	16	7	3	18	9	15	4			72	274	3,806	3,851	1,962
Please read the following statements and rate them. [The robot reacts according to the other person's emotions.]	4	9	6	12	26	13	2			72	310	4,306	2,323	1,524
Please read the following statements and rate them. [The eye colours of the robot show emotions.]	20	7	8	22	10	5	0			72	226	3,139	2,675	1,636
Please read the following statements and rate them. [The posture and gesture of the robot show emotions.]	4	6	7	10	23	17	5			72	329	4,569	2,495	1,580
Do you think the robot accomplishs its goal?	3	7	5	4	20	27	6			72	352	4,889	2,571	1,603
Do you think people will like/dislike such a robot? Please rate the following groups. [You]	8	9	6	5	18	21	5			72	315	4,375	3,484	1,867

Table A.1 continued from previous page

Do you think people will like/dislike such a robot? Please rate the following groups. [Elderly]	6	9	12	6	20	17	2			72	300	4,167	2,861	1,691
Do you think people will like/dislike such a robot? Please rate the following groups. [Children]	0	2	4	2	8	31	25			72	425	5,903	1,504	1,227
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-savvy]	2	5	3	7	19	23	13			72	373	5,181	2,342	1,530
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-averse]	12	19	12	12	9	7	1			72	228	3,167	2,667	1,633
Do you think people will like/dislike such a robot? Please rate the following groups. [Others apart from you]	1	5	5	29	22	10	0			72	312	4,333	1,222	1,106
Do you think the robot acts staged or natural?	14	10	16	2	16	12	2			72	256	3,556	3,469	1,863
Please read the following statements and rate them. [The robot is creepy.]	10	16	8	7	11	12	8			72	277	3,847	4,046	2,011
Please read the following statements and rate them. [The robot would make a good companion.]	7	10	8	7	19	16	5			72	305	4,236	3,264	1,807
Please read the following statements and rate them. [The voice of the robot is suitable.]	2	5	8	4	23	24	6			72	353	4,903	2,227	1,492
Please read the following statements and rate them. [I would feel like the robot controls the situation.]	10	9	7	15	14	14	3			72	284	3,944	3,191	1,786
Please read the following statements and rate them. [I would hug the robot in this situation.]	19	13	8	4	9	8	11			72	255	3,542	4,943	2,223
Please read the following statements and rate them. [The answers of the robot make sense.]	0	1	3	7	13	40	8			72	400	5,556	1,080	1,039
Please read the following statements and rate them. [The robot seems sad after hearing that the other person is sad.]	2	3	16	7	19	21	4			72	333	4,625	2,179	1,476

Table A.1 continued from previous page

Please read the following statements and rate them. [The robot seems happy after hearing that the other person is sad]	21	26	16	5	2	1	1			72	164	2,278	1,534	1,239
Please read the following statements and rate them. [The robot shows emotions.]	3	10	8	8	29	13	1			72	309	4,292	2,207	1,485
Please read the following statements and rate them. [The eye colour of the robot changed]	14	12	1	23	9	11	2			72	258	3,583	3,243	1,801
Please read the following statements and rate them. [The robot reacts according to the other person's emotions.]	3	0	1	7	24	29	8			72	384	5,333	1,556	1,247
Please read the following statements and rate them. [The posture and gesture of the robot show emotions.]	2	4	7	8	26	18	7			72	350	4,861	2,064	1,437
Please read the following statements and rate them. [The eye colours of the robot show emotions.]	15	13	7	23	6	7	1			72	233	3,236	2,708	1,646
Do you think the robot accomplishs its goal?	1	6	12	3	29	16	5			72	337	4,681	2,079	1,442
Do you think people will like/dislike such a robot? Please rate the following groups. [You]	9	9	11	7	12	17	7			72	299	4,153	3,741	1,934
Do you think people will like/dislike such a robot? Please rate the following groups. [Elderly]	5	13	9	9	10	22	4			72	304	4,222	3,284	1,812
Do you think people will like/dislike such a robot? Please rate the following groups. [Children]	1	2	1	2	9	33	24			72	427	5,931	1,481	1,217
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-savvy]	2	3	9	6	17	25	10			72	364	5,056	2,302	1,517
Do you think people will like/dislike such a robot? Please rate the following groups. [Tech-averse]	20	12	11	10	9	10	0			72	222	3,083	3,160	1,778

Table A.1 continued from previous page

Do you think people will like/dislike such a robot? Please rate the following groups. [Others apart from you]	1	7	12	22	17	13	0			72	302	4,194	1,601	1,265
Do you think the robot acts staged or natural?	5	16	21	5	16	8	1			72	255	3,542	2,387	1,545
Would you feel comfortable using the robot?	6	14	13	7	8	19	5			72	290	4,028	3,499	1,871
Please read the following statements and rate them. [I would feel watched by the robot]	6	13	8	5	21	11	8			72	303	4,208	3,443	1,855
Please read the following statements and rate them. [I would feel like having control over the robot.]	2	7	9	9	21	21	3			72	331	4,597	2,241	1,497
Please read the following statements and rate them. [The interaction works fine for short people.]	1	1	2	9	25	27	7			72	381	5,292	1,262	1,123
Please read the following statements and rate them. [I would feel comfortable in the robot's company.]	5	10	13	11	16	10	7			72	297	4,125	2,998	1,732
Please read the following statements and rate them. [I would feel like the robot controls the situation.]	9	15	21	9	10	5	3			72	239	3,319	2,579	1,606
Please read the following statements and rate them. [The interaction works fine for tall people.]	5	10	23	9	16	6	3			72	267	3,708	2,345	1,531
Do you think a robot should react towards your current emotion/mood?	15	57								72	129	1,792	0,165	0,406
Please read the following statements and rate them. [I think that I would like to use this robot frequently.]	17	13	8	6	12	14	2			72	249	3,458	3,859	1,965
Please read the following statements and rate them. [I found the robot unnecessarily complex.]	5	32	17	6	8	3	1			72	209	2,903	1,838	1,356
Please read the following statements and rate them. [I thought the robot was easy to use.]	1	3	1	9	24	32	2			72	372	5,167	1,306	1,143

Table A.1 continued from previous page

Please read the following statements and rate them. [I think that I would need the support of a technical person to be able to use this robot.]	21	31	9	4	5	2	0			72	163	2,264	1,639	1,280
Please read the following statements and rate them. [I found the various function of this robot very well integrated.]	1	3	4	22	26	16	0			72	333	4,625	1,207	1,098
Please read the following statements and rate them. [I thought there was too much inconsistency in this robot.]	4	26	16	12	9	5	0			72	227	3,153	1,852	1,361
Please read the following statements and rate them. [I would imagine that most people would learn to use this robot quickly.]	0	1	5	2	19	35	10			72	400	5,556	1,164	1,079
Please read the following statements and rate them. [I found the robot very cumbersome to use.]	3	18	16	18	10	3	4			72	255	3,542	2,193	1,481
Please read the following statements and rate them. [I felt very confident using the robot.]	3	6	7	23	23	9	1			72	304	4,222	1,701	1,304
Your age?	0	35	35	0	1	1	0			72	186	2,583	0,493	0,702
Your gender?	45	26	1	0	0	0	0			72	100	1,389	0,265	0,515
What is the highest level of education you have completed? Please choose the most fitting.	20	39	9	2	0	2				72	145	2,014	0,958	0,979
What describes your current employment best?	59	11	1	0	1	0				72	89	1,236	0,375	0,612
Which of the following best describes the field in which you received your highest degree?	0	35	10	2	3	1	4	6	11	72	304	4,222	7,784	2,790
How many people currently live in your household? Please write digits.	15 (1)	20 (2)	14 (3)	9 (4)	7 (5)	3 (6)	1 (10)	2 (15)	1 (30)	72	302	4,194	16,905	4,112
Please read the following statements and rate them. [I am tech-savvy.]	1	0	2	8	17	28	16			72	404	5,611	1,349	1,161
Please read the following statements and rate them. [I like to use new technical developements.]	0	2	3	5	21	25	16			72	400	5,556	1,414	1,189

Table A.1 continued from previous page

Please read the following statements and rate them. [People in my milieu tend to use new technical developements.]	0	1	6	1	21	29	14			72	401	5,569	1,329	1,153