Universidade do Minho
Escola de Engenharia

Vinicius Corrêa Alves da Silva

Mirroring and recognizing emotions through facial expressions for a Robokind platform

julho de 2015

UMinho | 2016
Vinicius Corrêa Alves da Silva
Mirroring and recognizing emotions through facial expressions for a Robokind platform

Universidade do Minho
Escola de Engenharia

Vinicius Corrêa Alves da Silva

Mirroring and recognizing emotions through facial expressions for a Robokind platform

Dissertação de Mestrado
Ciclo de Estudos Integrados Conducentes ao Grau de
Mestre em Engenharia Eletrónica Industrial e Computadores

Trabalho efetuado sob a orientação da
Professora Doutora Filomena Soares

julho de 2016

DECLARAÇÃO

Nome: Vinícius Corrêa Alves da Silva

Endereço eletrónico: a65312@alunos.uminho.pt Telefone: 917672420 (alternativo)

Bilhete de Identidade/Cartão do Cidadão: 14374652

Título da dissertação: Mirroring and recognizing emotions through facial expressions for a Robokind platform

Orientadora: Professora Doutora Filomena Maria da Rocha Menezes de Oliveira Soares

Ano de conclusão: 2016

Mestrado Integrado em Engenharia Eletrónica Industrial e Computadores

Universidade do Minho, 11/07/2016

Assinatura:

ACKNOWLEDGEMENTS

I am using this opportunity to express my gratitude to everyone who supported me throughout the development and writing of the present work. I am thankful for their apiring guidance, invaluably constructive criticism and friendly advice during the project work.

In particular, I would like to give my thanks to:

- My supervisor Dr. Filomena Soares for her support, guidance, and freedom to pursue my own ideas. She consistently allowed this dissertation to be my own work, but steered me in the right direction whenever she thought I needed it. All advice she gave were invaluable and I thank her sincerely;
- My co-supervisor Dr. João Sena Esteves for his assistance, advices, supervision, and excellent feedback.
- My colleague Joana Figueiredo for her knowledge on machine learning algorithms. I am gratefully indebted for her very valuable comments on this dissertation;
- Dr. Cristina Santos for her assistance, insight, and excellent feedback;
- Dr. Ana Paula Pereira for her insight, enthusiasm, and for sharing her vast knowledge on Autism Spectrum disorders and especial education.
- Dr. Celina Leão for her insight, enthusiasm, and for sharing her vast knowledge in statistics analysis.
- The professionals (therapists, psychologists, and teachers), parents, and the children who participated in the studies presented in this dissertation.
- My family, especially my mother for providing me with unfailing support and continuous encouragement throughout my years of study and through the process of researching and writing this dissertation. This accomplishment would not have been possible without them.

AGRADECIMENTOS

Estou usando esta oportunidade para expressar a minha gratidão a todos que me apoiaram ao longo do desenvolvimento e escrita do presente trabalho. Eu sou grato pela orientação, pelas criticas construtivas indispensáveis e conselhos durante a dissertação.

Em particular, gostaria de agradecer:

- A minha supervisora Dra. Filomena Soares, pelo seu apoio incondicional, orientação e liberdade para perseguir as minhas ideias. Ela constantemente permitiu que esta dissertação fosse o meu próprio trabalho, mas me guiou na direção certa sempre que fosse necessário. Todos os conselhos que me deu forma inestimáveis e agradeço-lhe sinceramente.
- Ao meu coorientador Dr. João Sena Esteves, pela sua ajuda, conselhos, supervisão e um excelente *feedback*.
- A minha colega Joana Figueiredo pelo seu valioso conhecimento em *machine learning*. Estou eternamente agradecido pelas suas valiosas observações em relação ao trabalho desenvolvido.
- A Dra. Cristina Santos por sua assistência e excelente *feedback*.
- A Dra. Ana Paula Pereira pela sua visão e entusiasmo, e por compartilhar o seu vasto conhecimento sobre Perturbações do Espectro do Autismo (PEA).
- A Dra. Celina pela sua visão e entusiasmo, e por compartilhar o seu vasto conhecimento em análise estatística.
- Aos profissionais (terapeutas, psicólogos e professores), pais e as crianças que participaram dos estudos apresentados nesta dissertação.
- A minha família, especialmente a minha mãe por fornecer apoio incondicional e incentivo contínuo ao longo dos meus anos de estudo e também no processo de pesquisa e escrita da dissertação. Esta conquista não teria sido possível sem eles.

Abstract

Facial expressions play an important role during human social interaction, enabling communicative cues, ascertaining the level of interest or signalling the desire to take a speaking turn. They also give continuous feedback indicating that the information conveyed has been understood. However, certain individuals have difficulties in social interaction in particular verbal and non-verbal communication (e.g. emotions and gestures). Autism Spectrum Disorders (ASD) are a special case of social impairments. Individuals that are affected with ASD are characterized by repetitive patterns of behaviour, restricted activities or interests, and impairments in social communication. The use of robots had already been proven to encourage the promotion of social interaction and skills in children with ASD.

Following this trend, in this work a robotic platform is used as a mediator in the social interaction activities with children with special needs. The main purpose of this dissertation is to develop a system capable of automatic detecting emotions through facial expressions and interfacing it with a robotic platform in order to allow social interaction with children with special needs.

The proposed experimental setup uses the Intel RealSense 3D camera and the Zeno R50 Robokind robotic platform. This layout has two subsystems, a Mirroring Emotion System (MES) and an Emotion Recognition System (ERS). The first subsystem (MES) is capable of synthetizing human emotions through facial expressions, on-line. The other subsystem (ERS) is able to recognize human emotions through facial features in real time. MES extracts the user facial Action Units (AUs), sends the data to the robot allowing on-line imitation. ERS uses Support Vector Machine (SVM) technique to automatic classify the emotion expressed by the User in real time.

Finally, the proposed subsystems, MES and ERS, were evaluated in a laboratorial and controlled environment in order to check the integration and operation of the systems. Then, both subsystems were tested in a school environment in different configurations. The results of these preliminary tests allowed to detect some constraints of the system, as well as validate its adequacy in an intervention setting.

KEYWORDS: Human Computer Interaction (HCI), Zeno R50 Robot, Facial Expressions, Emotions, Autism Spectrum Disorders (ASD)

Resumo

As expressões faciais desempenham um papel importante na interação social, permitindo fornecer pistas comunicativas, conhecer o nível de interesse ou sinalizar o desejo de falar. No entanto, algumas pessoas têm dificuldades na interação social, em particular, na comunicação verbal e não-verbal (por exemplo, emoções e gestos). As Perturbações do Espectro do Autismo (PEA) são um caso especial de transtorno e dificuldades sociais. Os indivíduos que são afetados com PEA são caracterizados por padrões repetitivos de comportamento, atividades e interesses restritos e possuem deficiências na comunicação social. A utilização de robôs para incentivar a promoção da interação social e habilidades em crianças com PEA tem sido apresentada na literatura.

Seguindo essa tendência, neste trabalho uma plataforma robótica é utilizada como um mediador nas atividades de interação social com crianças com necessidades especiais. O objetivo principal desta dissertação é desenvolver um sistema capaz de detetar automaticamente emoções através de expressões faciais e fazer interface com uma plataforma robótica, a fim de permitir uma interação social com crianças com necessidades especiais.

O trabalho experimental proposto utiliza a câmara Intel RealSense 3D e a plataforma robótica Zeno R50 Robokind. Este esquema possui dois subsistemas, um sistema de imitação de expressões faciais (MES) e um sistema de reconhecimentos de emoções (ERS). O primeiro subsistema (MES) é capaz de sintetizar *on-line* as emoções humanas através de expressões faciais. O subsistema ERS é capaz de reconhecer em tempo-real emoções humanas através de características faciais. O MES extrai as Unidades de Ação faciais do utilizador (UAs), envia os dados para o robô permitindo imitação *on-line*. O ERS utiliza Support Vector Machine (SVM) para automaticamente classificar a emoção exibida pelo utilizador.

Finalmente, os subsistemas propostos, MES e ERS, foram avaliados num ambiente laboratorial e controlado, a fim de verificar a integração e a operação de ambos. Em seguida, os subsistemas foram testados num ambiente escolar em diferentes configurações. Os resultados destes testes preliminares permitiram detetar algumas limitações do sistema, bem como validar a sua adequação na intervenção com crianças com necessidades especiais.

Palavras Chave: Interação Humano Robô (IHR), Robô Zeno R50, Expressões Faciais, Emoções, Perturbações do Espectro do Autismo (PEA).

CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

ANN	**A**rtificial **N**eural **N**etwork
API	**A**pplication **P**rogramming **I**nterface
ASD	**A**utism **S**pectrum **D**isorders
AU	**A**ction **U**nit
AUC	**A**rea **U**nder the **C**urve
CV	**C**ross **V**alidation
DoF	**D**egrees **o**f **F**reedom
FACS	**F**acial **A**ction **C**odding **S**ystem
FPP	**F**eature **P**oint **P**ositions
ERS	**E**motion **R**ecognition **S**ystem
GUI	**G**raphical **U**ser **I**nterface
HCI	**H**uman **C**omputer **I**nteraction
HRI	**H**uman **R**obot **I**nteraction
IDE	**I**ntegrated **D**evelopment **E**nvironment
k-NN	**k** – **N**earest **N**eighbours
LDA	**L**inear **D**iscriminant **A**nalysis
MCC	**M**athews **C**orrelation **C**oefficient
MES	**M**irroring **E**motion **S**ystem
OAA	**O**ne-**A**gainst-**A**ll
OAO	**O**ne-**A**gainst-**O**ne
OSH	**O**ptimal **S**eparating **H**yperplane
RBF	**R**adial **B**asis **F**unction
ROC	**R**eceiver **O**perating **C**haracteristic
SDK	**S**oftware **D**evelopment **K**it
SVM	**S**upport **V**ector **M**achines
SVs	**S**upport **V**ectors
ZECA	**Z**eno **E**ngaging **C**hildren with **A**utism

1 INTRODUCTION

Summary

In this chapter the problem statement, motivations, and scope of the work are presented. It starts with a brief reference to the concept of affective computing and the importance of social robots. Then, the objectives and ethical considerations of the present work are defined. Finally, the results of the developed scientific activity as well as the structure of the dissertation are presented.

Faces contain a large portion of the human emotionally expressive behaviour. Facial expressions are used to display emotional states and to manage interactions. Furthermore, humans can express and read emotions in faces effortlessly. On the other hand, automatic understanding of facial expressions is a very difficult task for some people, especially, for children with Autism Spectrum Disorders (ASD).

Nowadays, assistive robotics focus to help Users with special needs in their daily activities. Assistive robots are designed to identify, measure, and react to social behaviours, being repeatable and objective offering an exceptional occasion for quantifying social behaviour (Tapus, Member, & Scassellati, 2007). They can be a social support to motivate children, socially educate them and beyond that help transferring knowledge. According to studies, it was observed that children with autism disorders can exhibit certain positive social behaviours when interacting with robots in contrast to what is perceived when interacting with their peers, caregivers, and therapists (Ricks & Colton, 2010). Furthermore, a few projects world-wide pursue to include robots as part of the intervention program for individuals with autism (Dautenhahn, 2000). These studies have demonstrated that robots can promote a high degree of motivation and engagement in subjects, including subjects who are improbable or reluctant to interact socially with human therapists (Scassellati, 2007).

The robot's physical appearance plays an important role in the interaction process with a person. The physical appearance of robots used for autism intervention vary greatly from simple designs, e.g. four-wheeled mobile robots, to many levels of anthropomorphic forms, including humanoid (Costa, 2014), animal-like (Kim *et al.*, 2013), and machine-like systems (Michaud *et al.*, 2005). Consequently, it is important to discuss the appropriate features a robot must have, in order to be used in intervention processes with children with ASD. A humanoid robot can be a useful tool to develop social-emotional skills in the intervention of children with ASD, due to the engagement and positive learning outcome (Costa, 2014).

However, robotic systems are emotionally blind. Conversely, successful human-human communication relies on the ability to read affective and emotional signals. Following this trend, there are many studies about affective computing and how it can be beneficial to their Users (Picard & Klein, 2002; Robinson & el Kaliouby, 2009). Moreover, affective computing can illuminate early emotional dynamics and provide tools for intervention in disordered emotional functioning. It can be used to understand emotional communication in typically developing children and children with ASD (Robinson & el Kaliouby, 2009; S.Messinger, 2014).

In order to make future progress in Human Computer Interaction (HCI), it is thought that it is necessary for the machine to recognise Users' affect and adapt its behaviour to them, giving an appropriate response for those emotions. Affective computing tries to build a connection between the emotionally expressive human and the emotionally lacking computer (D'Mello & Calvo, 2013).

There are many possible application areas for affective computing such as marketing systems that evaluate the User's reaction to a product, online learning systems that adapts the teaching method according to the student's emotions, and video games that can adapt their level of difficulty based on the player engagement. Additionally, another application could be a robotic system that can detect the engagement level of an operator and modify its behaviour if required (Rani & Sarkar, 2007); this can be applied in autism intervention.

The affective behaviour displayed by humans is multi-modal, subtle and complex. Humans use affective information such as facial expressions, eye gaze, various hand gestures, head motion, and posture to deduce the emotional state of each other (Ambady & Rosenthal, 1992). In addition, the face has received the most attention since it is central to the communication of emotions since the early stages until the last stages of the human life. In fact, faces are the most noticeable social part of the human body. The face is crucial for human identity, it can reveal emotions (Ekman & Rosenberg, 2005), communicate intent and help regulate social interactions. As well as the face, head gestures are also important in human communication, they can offer several key conversational foundation cues and are used in face-to-face interaction among people (Morency, Sidner, Lee, & Darrell, 2006).

1.1 Problem Statement, Motivations and Scope

Sometimes the face reveals what the words hide. This means that non-verbal information can prevail over the verbal information. The understanding of this non-verbal information such as gestures, facial expressions, head poses, plays an important role on human interaction and communication since these movements convey the emotional state of an individual. This capacity to understand or experience another being, as well to respond with an appropriate emotion to other person's mental states (thoughts and emotions) is called empathy (Baron-Cohen & Wheelwright, 2004).

Considering emotions in the perspective of evolution, Darwin argued in his book *The Expressions of the Emotions in Man and Animals* (Darwin, 1872) that emotions actually is useful for humans, in communication and also in sustaining their survival. Furthermore, Darwin claimed that emotions evolved via natural selection and therefore have cross-cultural counterparts. Ekman (Ekman *et al.*, 1987) claims that certain emotions appeared to be universally recognized which led him to classify six “basic” emotions: anger, disgust, fear, happiness, sadness and surprise.

This lack of empathy is present generally in individuals with ASD since they have difficulty interpreting the emotional states of others as well expressing themselves. The empathy deficits may have a crucial influence on their social behaviour. Systems that can synthesize affects, for example virtual characters (avatars), and robotic platforms can be applied in autism intervention (Costa, 2014).

Following this idea, one of the motivations for the research presented in this dissertation is to endow a robotic humanoid platform of emotional knowledge, “humanizing” the interaction between humans and robots. In addition, this humanoid robotic platform is to be used as a mediator in social interaction activities. In this dissertation, a humanoid robotic platform capable of express emotions is used as a mediator in social interaction with children with ASD.

1.2 Objectives

The focus of the present work is in the development and application of interactive and assistive technologies to support and promote new adaptive teaching/learning approaches for children with ASD. Following this idea, the main goals of this dissertation are:

1) developing a system capable of automatically detecting emotions through facial expressions;
2) interfacing the described system with a robotic platform in order to allow social interaction with children with ASD.

For achieving the mentioned goals, the proposed experimental setup uses the Intel RealSense 3D camera and the Zeno R50 Robokind robotic platform. This layout has two subsystems, a Mirroring Emotion System (MES) and an Emotion Recognition System (ERS). The first subsystem (MES) is capable of on-line synthetizing human emotions through facial expressions. The other subsystem (ERS) is able to recognize human emotions through facial features. MES extracts the User facial Action Units (AUs), and sends the data to the robot allowing on-line imitation. ERS uses Support Vector Machine (SVM) technique to automatic classify the emotion expressed by the User. In a first stage, by using facial landmarks and AUs, the robotic platform is able to mimic in on-line time the emotional state of the User. Finally, the humanoid robot has the capacity to detect five of the six “basic” emotions (anger, fear, happiness, sadness and surprise) plus neutral displayed by the User.

1.3 Ethical Considerations

The work presents studies involving children with ASD. Thus, the following issues were ensured to meet the ethical concerns:

- Protocols: The school which participated in the studies established a protocol with the University where the research was developed. Prior to the experiments, a meeting took place in the school to clarify any questions from the professionals who interact daily with the children.
- Parents' consent: The children's parents/tutors signed an informed consent, available in the Appendix A.1, in which they allowed the participation of their children in the research. This consent was accompanied by a document clarifying the objectives, risks, and benefits of the research, as well as the full freedom to accept participating in the study and withdraw their child at any time.
- Privacy: The personal data of the participants in the research is enclosed and all private information collected during the study is confidential and dealt according to the rules on data and private life. Anonymity is guaranteed at any time of the study; only the researcher and the professionals who follow the children on a daily basis have knowledge of this data.

1.4 Results of the developed scientific activity

Part of the work developed was submitted and approved for oral presentation in an international conference:

- Vinícius Silva, Filomena Soares, João S. Esteves, Joana Figueiredo, Cristina Santos, Ana Paula Pereira, Happiness and sadness recognition system – Preliminary results with an Intel RealSense 3D Sensor, CONTROLO16, 12th Portuguese Conference on Automatic Control, Guimarães, Portugal, 14-16 September, 2016.

Another two papers were already submitted to an international conference.

As part of this research work, the project was presented in three lectures:

- "Algumas das coisas que se podem fazer num Mestrado em Engenharia..." (In Portuguese), Project "Robótica e Autismo", EB 2/3 Júlio Brandão, Vila Nova de Famalicão, 23 March, 2016.

- “Demonstração do Robô ZECA” (In Portuguese) at the conference “Conferência 2016: Dia Mundial da Consciencialização do Autismo” at Faculty of Psychology of University of Porto, Porto 2 April 2016.
- “Demonstração do Robô ZECA” (In Portuguese), Invited Talk at Hospital Magalhães Lemos, Porto, 13 Maio, 2016.

1.5 Dissertation Structure

The dissertation is organized as follows:

- Chapter 2 provides background knowledge in the areas of emotional processes, affective computing, and the use of social robots with children with ASD. This chapter starts with an overview of the concept of emotions and the theories of emotional processes. Then, it presents an overview of the field of affective computing with emphasis on affective sensing from facial expressions. Posteriorly, the problematic related to ASD is highlighted, focusing on the difficulty in social interaction, imitation, and emotion recognition. Consequently, the last section concerns studies using robots to interact with children with ASD discussing their application to promote social interaction, and emotion recognition. The information from this chapter was used as basis for the developed work.
- Chapter 3 presents the methodologies used in the present work. It starts by presenting the methods used for extracting the facial features. Then, it introduces an overview of machine learning methods with more emphasis on SVMs and how to assess the performance of a classifier. Finally, it presents the methods generally used for validating systems that are capable of synthesizing facial expressions and recognize emotions and a methodology usually used for creating a database of facial characteristics.
- Chapter 4 focus on describing the general implementation of the system, particularizing the two subsystems, the MES and ERS subsystems. The chapter starts by presenting the hardware and software used for implementing the system developed in the present work. Then, it describes the overall system, emphasizing the implementation of the two subsystems. Additionally, the User Interface and the software architecture are presented.
- Chapter 5 presents the results obtained from the present work. It starts by showing the results from the MES subsystem. Then, it presents the results obtained from evaluation of the ERS subsystem. Finally, a preliminary study involving children with ASD is presented.

- Chapter 6 draws the conclusion of the work described in the dissertation and provides some outlook for the future use of robotics in intervention with ASD. The research presented in this dissertation highlights systems that are capable of synthesizing and recognizing facial expressions and how they can be applied in ASD intervention.

References

Ambady, N., & Rosenthal, R. (1992). Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis.

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. http://doi.org/10.1023/B:JADD.0000022607.19833.00

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders*. University of Minho.

D'Mello, S., & Calvo, R. (2013). Beyond the basic emotions: what should affective computing compute? *... Abstracts on Human Factors in Computing ...*, 2287–2294. http://doi.org/10.1145/2468356.2468751

Darwin, C. (1872). *The expression of emotions in man and animals*. *Library* (Vol. 3). http://doi.org/10.1097/00000441-195610000-00024

Dautenhahn, K. (2000). Design issues on interactive environments for children with autism. *3Th International Conference on Disability ,Virtual Reality and Associated Technologies*, 153–162. http://doi.org/10.1.1.33.4997

Ekman, P., Friesen, W. V, O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ... Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*. US: American Psychological Association. http://doi.org/10.1037/0022-3514.53.4.712

Ekman, P., & Rosenberg, E. (2005). *What the face reveals*. *What the Face Reveals*.

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038–1049. http://doi.org/10.1007/s10803-012-1645-2

Michaud, F., Laplante, J.-F., Larouche, H., Duquette, a., Caron, S., Letourneau, D., & Masson, P. (2005).

Autonomous spherical mobile robot for child-development studies. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *35*(4), 471–480. http://doi.org/10.1109/TSMCA.2005.850596

Morency, L., Sidner, C., Lee, C., & Darrell, T. (2006). The Role of Context in Head Gesture Recognition. *The Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, 1650–1653.

Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, *14*(Spitz 1945), 141–169.

Rani, P., & Sarkar, N. (2007). Operator engagement detection for robot behavior adaptation. *International Journal of Advanced Robotic Systems*, *4*, 1–12. http://doi.org/10.1109/ROBOT.2005.1570415

Ricks, D. J., & Colton, M. B. (2010). Trends and considerations in robot-assisted autism therapy. *Proceedings - IEEE International Conference on Robotics and Automation*, 4354–4359. http://doi.org/10.1109/ROBOT.2010.5509327

Robinson, P., & el Kaliouby, R. (2009). Computation of emotions in man and machines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3441–3447. http://doi.org/10.1098/rstb.2009.0198

S.Messinger, D. (2014). Affective Computing, Emotional Development, and Autism, 516–536.

Scassellati, B. (2007). How social robots will help us to diagnose , treat , and understand autism. *Robotics Research*, 552–563. http://doi.org/10.1007/978-3-540-48113-3_47

Tapus, A., Member, S., & Scassellati, B. (2007). The Grand Challenges in Socially Assistive Robotics. *IEEE Robotics and Automation Magazine*, *14*, 1–7. http://doi.org/10.1109/MRA.2010.940150

2 LITERATURE REVIEW

Summary

This chapter presents an overview of the field of affective computing with emphasis on affect sensing from facial expressions, since this field is too broad to be described in detail in this dissertation. Then, the emotional process and Autism Spectrum Disorders (ASD) are characterized. Finally, the pair Human Robot Interaction (HRI) and ASD are referred, showing some research projects involving social robots and individuals with ASD.

Emotion is often defined as a complex state of feeling that results in physical and psychological changes that influence thought and behaviour. Michel Cabanac, (Cabanac, 2002), proposes that (1) "emotion is any mental experience with high intensity and high hedonic content (pleasure/displeasure)". (2) "An emotion is a mental state, even when somatic signals participate in this mental experience". (3) "Any mental experience with high intensity and hedonism should be accompanied with the usual objective somatic signs of emotions, for instance tachycardia".

As mentioned above, Charles Darwin proposed that emotions developed as an evolutionary advantage via natural selection, as they are useful for humans, in communication and also in sustaining their survival, for instance feelings of fear compel people to either fight or flee the source of danger (Darwin, 1872). In addition, Darwin considered that emotional facial expressions are innate and by being able to interpret them correctly, a person can respond and avoid danger. Darwin's main goal was to demonstrate how emotions in humans were comparable to those in animals, supporting his hypothesis of the existence of a common ancestor for man and animals.

According to the Schachter-Singer theory, the process of experiencing an emotion evolves first a physiological arousal, then the identification of the reason for this arousal, searching the immediate environment for emotional cues (Schachter & Singer, 1962). Furthermore, this theory suggests that similar physiological responses can produce varying emotions, depending on the environment that the person is at the moment.

Ekman's neuro-cultural theory of emotions, suggests the existence of universal emotions, i.e., the significance of an emotion is the same for all people in cultures, and therefore everyone expresses in the same manner (Ekman *et al.*, 1987). In order to sustain his theory, Ekman carried out an experiment where he showed photographs of different emotions expressions to observers in the United States, Japan, Chile, Argentina, and Brazil. The observers in these different cultures had to match one of six emotion words to each photograph they saw. If the meaning of a facial expression differs from culture to culture, then a facial expression said to be anger by Americans might be labelled as disgust or fear by people in Brazil. However, the findings suggest otherwise, i.e. the same facial expressions were judged as showing the same emotions in all these countries, regardless of language or culture. Despite these findings lead to the conclusion that some facial expressions are universal, a loophole remained. The people used in this study had some shared visual contact, usually not directly but through the mass media. Considering this gap, another series of experiments were conducted in the Southeast highlands of New Guinea, where people were in no way used to taking psychological tests or participating in experiments. In these experiments, an emotion story was read while three photographs were showed at once to the person. Then, the observer had to point to the photograph that fit the story. The results, showed that these people selected the same face for the same emotion as did people in all other cultures.

Despite, the stated theories about emotions and the emotional process, a consensus on a definition or interpretation of emotional processes has not yet been reached. However, most research agrees that the recognition of emotions is largely universal, with the implication that this skill is not learned, but rather has an evolutionary and thus biological basis (Elfenbein & Ambady, 2003).

2.1 Affective Computing

Affective computing was first introduced by Rosalind Picard in her book, *"Affective Computing"*. This interdisciplinary field, embraces computer science, psychology, and cognitive science and consists in the study and development of systems and devices that can recognize, interpret, process, and simulate human affects (Picard, 1997). A motivation for the research is the ability to simulate empathy. Moreover, affect sensing, affect synthesis and the design of emotionally intelligent interfaces are fields encompassed in affective computing (Baltrusaitis, 2014). A computing device with this capacity could gather cues, from a variety of sources, and use it to respond to its User's emotions and other stimuli. Facial expressions, posture, speech, gestures, force, heart rate, and body temperature can all potentially signify emotional changes. Technologies such as speech, gesture, and facial expressions recognition are being explored and employed in affective computing applications.

Following this trend, there are many benefits that affective computing can provide in an almost limitless range of applications. Detection and tracking of stress and tiredness could be highly valuable in safety critical systems. For example, a car might automatically monitor the emotions of all occupants and invoke safety measures, potentially recommending the driver to safely stop the vehicle if it detects any signs of tiredness, or alerting others around him if it detects any negative affective states, thus preventing accidents (Baltrusaitis, 2014).

Affective computing could also be applied in systems that monitors patients in hospitals or medical facilities. One promising application under development is the automatic detection of pain as proposed by (Ashraf *et al.*, 2007). Pain is typically assessed by patient self-report, which sometimes can be difficult

to interpret. This approach can automatically recognize acute pain from a video through facial indicators of pain. These findings can be useful to reach a diagnosis.
Another application of affective computing is on mobile apps. One example, already on the market, is the "*Musicovery*" app that selects the type of music that is playing based on the User emotional state ("Musicovery | Mobile Aplication," 2015). The app has a feature called the "mood pad" which lets the User change the type of music by tapping on the area that fits the User's mood at the moment.
Companies such *Emotient* ("Emotient | A Leader in Emotion Measurement," 2015) and *Noldus* ("Noldus | Innovative solutions for behavioral research," 2015) deliver cloud-based services capable of emotion detection and sentiment analysis based on facial expressions. This can be used as a marketing strategy to infer whether a product will be well-received by the market by detecting facial or speech changes in potential customers when they read an ad or first use the product.
Another area that affective computing is being use is on entertainment industry, specifically in affective gaming (Hudlicka, 2008). This field has received much attention lately, most from the gaming community, as they recognize the importance of emotions in the development of engaging games. The present main focus in affective gaming is on the sensing and recognition of the player's emotions modifying the game responses to these emotions. A parallel effort is also being applied to generating "affective behaviours" in the game characters (avatars), enhancing their realism and believability. In addition, synthesis of "affective behaviours" is also being applied in robotic platforms (Costa, 2014). These systems capable of processing and displaying affective information can offer more functionality alongside human workers in uncertain or complex environments.
Affective computing is also starting to be applied to the development of communicative technologies for use in autism intervention. A personal mobile robot designed to both embody and elicit reflection on shyness behaviours, Shybot (Lee, Kim, Breazeal, & Picard, 2008), is being used as a new direction in the intervention process in children with autism. Shybot is capable to detect human presence and familiarity from face detection and proximity sensing, categorising people as friend or strangers to interact with. Systems as mentioned could be helpful in making progress in improving the capacity for complex social interaction.

2.1.1 Affect sensing – Facial Expressions

Faces embody a large portion of the human emotionally expressive behaviour. Moreover, facial expressions are used to display emotional states and to manage interactions, being one of the most important channels of non-verbal communication. The first steps on research on facial expression began with Darwin's "*The Expression of the Emotions in Man and Animals*" (Darwin, 1872). Darwin proposed that emotions are universal and evolved via natural selection. Furthermore, Darwin claimed that emotions are important for communication and also to sustain human survival. In order to substantiate his theory, Darwin obtained data from informants in different countries and analysed observers' responses to different expressions. However, a more important step in the research on facial expression came from Paul Ekman, (Ekman & Rosenberg, 2005), with his work on the basic emotions (happiness, sadness, anger, surprise, disgust, and fear), and the Facial Action Coding System (FACS) (Ekman & Friesen, 1978). FACS is a comprehensive, anatomically based system for measuring all visually noticeable facial movement. FACS describes all visually distinguishable facial activity, as well as several categories of head and eye positions and movements. It associates the action of the muscles to the changes in facial appearance. The measurements of the FACS are called Action Units (AUs) which are actions performed by a muscle or a group of muscles, and each AU has a numeric code. This system made it possible for researches to analyse and classify facial expressions in a standardised framework. As referenced by (Cohn, 2007), FACS is the most comprehensive psychometrically rigorous, and widely used method for analysing facial expressions.

Head pose and eye gaze together with facial expressions, are very important to convey emotional states. Head nods helps emphasizing an idea during a conversation. In addition, it also helps to agree or disagree with a point of view through nodding signals such as signal 'yes' or 'no' head movements, synchronizing the interactional rhythm of the conversation (Hadar, Steiner, & Clifford Rose, 1985). Eye gaze is important for analysing attentiveness, competence, as well as intensity of emotions.

Facial expressions, head pose and eye gaze play an important role in expressing affect and communicating social signals. They are processed together, when analysing human emotional states, from a computational point of view (Baltrusaitis, 2014).

2.1.2 Affect sensing – Systems and devices

Systems and devices that can recognize, interpret, process and simulate human affects have been in development for a considerable time. Recognizing emotional information requires the extraction of meaningful patterns from the gathered data. Some researchers are using machine learning techniques to detect such patterns. Following this idea, there has been an extensive research focusing on automatic facial expression recognition.

(Michel & Kaliouby, 2000) presented an approach to emotion recognition in live video. Initially, the system extracted 22 facial features from the live video stream of a camera. Then, the displacements for each feature between a neutral and a representative frame of an emotion were calculated. These data, as well as the correspondent labels, were used as an input to the training stage of a Support Vector Machine (SVM) classifier. Posteriorly, the trained SVM model was used to classify an emotion between the six pre-defined emotions (six basic emotions): anger, disgust, fear, happiness, sadness, and surprise. Finally, the system was evaluated in a first stage by a professional User, familiar with the approach and aware of how the basic emotions are typically expressed, with an overall accuracy of 87.5%. The accuracy results, in this first stage, for each emotion were the following: anger – 83.3%, disgust – 100%, fear - 83.3%, happiness – 75.0%, sadness - 83.3% and surprise - 100.0%. Then, the system was evaluated in a more challenging scenario, where inexperienced Users were asked to naturally express the emotions. The overall accuracy achieved in this last evaluation was 60.7% and the accuracy results for each emotion were the following: anger – 59.4%, disgust – 58.1%, fear – 51.7%, happiness – 63.6%, sadness – 54.4% and surprise – 75.8%.

The system proposed by (Youssef, Aly, Ibrahim, & Abbott, 2013) attempts to recognize the six basic emotions using a Microsoft Kinect sensor. At first, the system extracts 4D facial points (dynamic 3D facial points). Then, these facial points are partitioned into two sets according to their locations on the face, where the first set represents the upper part of the face (containing the eyes and eyebrows) and the second set represents the lower part of the face (containing the mouth). SVM was used to classify the six basic emotions, where each set was fed to a separate SVM module, and then a third SVM performed the final decision based on the output values from the first two SVMs. The database used for training contained 4D data of 14 different persons performing the six basic emotions. For the purpose of comparison, the authors used also the k-Nearest Neighbour (k-NN) classifier. For individuals who did not participate in the training of the classifiers, the overall accuracy results were 38.8% (with SVM) and 34.0% (with k-NN). However, the accuracy levels raised to 78.6% (SVM) and 81.8% (k-NN), when considering only the individuals that participated in the training.

The work developed by (Alabbasi, Moldoveanu, & Moldoveanu, 2015) consisted in a system that uses a Microsoft Kinect for Windows sensor V2 and the face tracking SDK to recognize eight emotions (anger, disgust, fear, happiness, sadness, surprise, neutral and contempt). The implementation of the emotion recognition application was developed with Visual Studio 2013 (C++) and Matlab 2014. The Kinect sensor and the face tracking SDK were used to obtain the facial Action Units (AUs). Firstly, on the application interface, the face features values were saved into a *.mat* file. Then, using Matlab, it was implemented an Artificial Neural Network (ANN) that classified the emotions based on the facial features. The training set contained the facial AUs from 12 persons for each of the eight emotions. Finally, the system was tested, in a first stage, with persons that participated in the creation of the database obtaining an overall accuracy of 96%. Then, the system was tested with people that did not participate in the database creation, obtaining an overall accuracy of 92%.

(Silva, Sobral, & Vieira, 2014), proposed an automatic human-face expression recognition frame-based system that classifies seven different facial expressions: happiness, anger, sadness, surprise, disgust, fear, and neutral. The proposed framework first detects the face, extracting and normalizing the features building a vector of features that is used as an input for the classifiers. In order to choose the best classifier, three different classifiers (Artificial Neural Network (ANN), Linear Discriminant Analysis (LDA) and k-Nearest Neighbour (k-NN)), were chosen for evaluating the system performance. Each classifier was trained and tested with two different databases (MUG and FEEDTUM databases). The overall accuracy for each classifier was the following: ANN – 99.28%, LDA – 99.71% and k-NN – 76.42%, for the MUG database. Using the FEEDTUM database, the overall accuracy was: ANN – 97.90%, LDA – 99.55% and k-NN – 54.70%. The classifier with the highest accuracies, LDA, was chosen.

(Mao, Pan, Zhan, & Shen, 2015) proposed a real-time emotion recognition approach based on both 2D and 3D facial expression features captured by Kinect sensors. The facial features selected were the facial Action Units (AUs) and Feature Point Positions (FPPs) or facial landmarks both obtained from the Microsoft Kinect Facial Tracking SDK. The proposed approach first acquires the facial features from each frame. Then, in separate channels, seven sub SVM classifiers (1-vs-1) with Radial Basis Function kernel, RBF, are trained. The seven way 1-vs-1 classification establishes an independent sub-classifier per emotion (six basic emotions and the neutral), and seven output labels representing the membership in the class (denoted as '+1') or out of the class (denoted as '-1') are created to weight the confidence of each emotion. After that, a fusion algorithm based on Improved Emotional Profiles (IEPs) is proposed to obtain the emotion of a video sequence. The pre-recognition results of the latest 30 frames in AU and FPP channels are combined, i.e., the output labels of each 7-way 1-vs-1 classifier are produced and input into two memory buffers: a memory buffer of AUs recognition results and a memory buffer of FPPs recognition results. Finally, the confidence of each emotion will be achieved, and the emotion with the maximum confidence will be the estimated emotion in each channel. The authors created their own database which was used for training and testing the system. The experiments were conducted on a computer with an Intel dual-core, 2.8 GHz CPU and 4GB of RAM. The following average accuracies for each emotion were obtained: anger – 86.14%, disgust – 83.28%, fear – 81.88%, happiness – 80.67%, neutral – 87.86%, sadness – 80.48% and surprise – 88.49%. The authors concluded that the system can meet the real-time requirement.

(Ghimire *et al.*, 2015) presented a method for fully automatic facial expression recognition in facial image sequences using features extracted from tracking of facial landmarks. Fifty-two facial landmarks were tracked in the consecutive video frames over time. Initially, the locations of these facial landmarks were used to extract the shape of facial features, in this case triangular geometric features, and the movements of facial features. Then, this geometric shapes composed of facial landmarks are selected using the multi-class AdaBoost with Extreme Learning Machine (ELM) classifier. In the proposed approach, the authors used AdaBoost as a feature selection method, selecting the triangles that carried the discriminative information required for efficiently recognizing the facial expressions. Finally, the facial expressions are recognized using SVMs classification. The proposed approach was evaluated using the k-fold cross validation approach (k=5) and two databases, the extended Cohn-Kanade (CK+) and Multimedia Understanding Group (MUG) databases. The results obtained with the CK+ dataset were the follow: anger – 97.5%, disgust – 96.67%, fear – 96%, happiness – 100%, sadness – 96.67% and surprise – 100%, achieving an overall accuracy of 97.80%. The results obtained with the MUG dataset were the follow: anger – 100%, disgust – 100%, fear – 85%, happiness – 100%, sadness – 90% and surprise – 98%, obtaining an overall accuracy of 95.50%.

(Zhang, Zhang, & Hossain, 2015) proposed a real-time 3D emotions recognition system that used the Microsoft Kinect sensor. Firstly, the system continuously tracks in real-time 121 3D facial landmarks. Then, motion-based facial features were extracted, by calculating displacements between couple of landmarks, and selected by using minimal-redundancy-maximal-relevance (mRMR) based automatic feature selection method. These facial geometric features were subsequently employed as inputs to an array of Neural Networks and Support Vector Regressors (SVRs) respectively to estimate the intensities of the AUs. Finally, the derived AUs intensities were used as input to a multiclass SVM with RBF kernel with the purpose of recognizing the six basic emotions. Two types of evaluations were performed in order to assess the system performance: static off-line and real-time on-line evaluations. In the offline evaluation, the system was trained an tested using the Bosphorus database ("The Bosphorus," 2009) achieving the following accuracies for each emotion: anger – 91.3%, disgust – 85.6%, fear – 91.1%, happiness – 95.6%, sadness – 82.7% and surprise – 96.5%, with an average accuracy of 90.5%. In the online evaluation, the system was tested by a set of 11 participants achieving an overall accuracy of 84%. The accuracy for each emotion were the follow: anger – 79.4%, disgust – 83.7%, fear – 81.6%%, happiness – 88.1%, sadness – 77.9% and surprise – 93.2%. The computational cost of the system in the real-time testing was about 3-5*ms* with a frame rate of 25-30 fps on an i7 quad-core CPUs with 8GB of RAM.

One thing in common with all of the above outlined systems is that they use machine learning methods for classifying emotions. Some use as input for the classifiers facial landmarks or feature point positions (FPPs) and other use action units (AUs). In general, some systems use 2D RGB cameras, and other systems uses 3D sensors, usually the Kinect sensor from Microsoft. Considering all the machine learning methods that are employed in emotions recognitions, the SVM is the most used.

Table 2.1.2-1 summarizes and compares the facial recognition accuracies for all the projects presented above.

Table 2.1.2-1 Emotion Recognition - Facial recognition accuracies for the mentioned projects.

Author	**Method**	**Implementation goal**	**Results**
(Michel & El Kaliouby, 2003)	SVM	Recognize the six basic emotions using live video images.	First evaluated by a professional User: anger – 83.3%, disgust – 100%, fear - 83.3%, happiness – 75.0%, sadness - 83.3% and surprise - 100.0%. Overall accuracy of 87.5%. Then by inexperienced Users: anger – 59.4%, disgust – 58.1%, fear – 51.7%, happiness – 63.6%, sadness – 54.4% and surprise – 75.8%. Overall accuracy of 60.7%.
(Youssef *et al.*, 2013)	SVM and k-NN	Recognize the six basic emotions using a Microsoft Kinect sensor and the face tracking SDK.	Two tests were conduct in order to validate the system. The first one with individuals that participated in the training of the classifiers with the accuracies of 78.6% (SVM) and 81.8% (k-NN). The second test with individuals who did not participate in the training of the classifiers, with accuracies of 38.8% and 34.0%.
(Alabbasi *et al.*, 2015)	ANN	Recognize eight emotions (the six basic emotions + neutral and contempt), using the Microsoft Kinect for Windows sensor V2 and the face tracking SDK.	Two tests were conduct in order to validate the system. The first one with individuals that participated in the creation of the database achieving an accuracy of 96%. Then with individuals that not participated in the construction of the database with an accuracy of 92%.
(Silva *et al.*, 2014)	ANN, LDA and k-NN	Recognize seven facial expressions (the six basic emotions + neutral) using a frame-based system.	Each classifier was trained and tested with two different databases. The overall accuracy for each classifier were the following: ANN – 99.28%, LDA – 99.71% and k-NN – 76.42%.
(Mao *et al.*, 2015)	SVM	Real-time emotions recognition system based on both 2D and 3D facial expression features captured by Kinect sensors.	The results obtained, after testing the system, were: anger – 81.14%, disgust – 83.28%, fear – 81.88%, happiness – 80.67%, neutral – 87.86%, sadness – 80.48% and surprise – 88.49%.

Author	**Method**	**Implementation goal**	**Results**
(Ghimire *et al.*, 2015)	SVM	A method for fully automatic facial expression recognition in facial image sequences using geometric features.	The system was tested with two databases. The results obtained with the CK+ dataset were the follow: anger – 97.5%, disgust – 96.67%, fear – 96%, happiness – 100%, sadness – 96.67% and surprise – 100%, achieving an overall accuracy of 97.80%. The results obtained with the MUG dataset were the follow: anger – 100%, disgust – 100%, fear – 85%, happiness – 100%, sadness – 90% and surprise – 98%, obtaining an overall accuracy of 95.50%.
(Zhang *et al.*, 2015)	SVM	Real-time 3D emotions recognition system that used the Microsoft's Kinect sensor and the face tracking SDK.	Two types of evaluation were performed: static off-line and real-time on-line evaluations. The results obtained in the off-line evaluation were the following: anger – 91.3%, disgust – 85.6%, fear – 91.1%, happiness – 95.6%, sadness – 82.7% and surprise – 96.5%, with an average accuracy of 90.5%. The results obtained in the on-line evaluation were the follow: anger – 79.4%, disgust – 83.7%, fear – 81.6%%, happiness – 88.1%, sadness – 77.9% and surprise – 93.2%, achieving an overall accuracy of 84%.

2.2 Emotional Process and Autism Spectrum Disorders (ASD)

Although the process of recognition and displaying emotions could be an easy task for the majority of humans, it is a very difficult task for individuals with ASD.

Individuals with ASD are characterized by displaying repetitive patterns of behaviour, for having restricted activities or interests, and impairments in social communication. Furthermore, these individuals have difficulties recognizing body language, making eye contact, understanding other people's emotions, and the lack of social or emotional reciprocity (Scassellati, Henny Admoni, & Matarić, 2012). These difficulties in interpreting social situations in general, causes children with ASD lose or miss information on what is happening or happened during the social exchange (Happé, Briskman, Frith, Happé, & Frith, 2001).

Baron-Cohen refers that children with ASD have difficulty in understanding the perspective of others and predict behaviours, i.e., children with ASD have difficulty in sharing and showing empathy.

Children with ASD presents imitations impairments. In fact, typically developing children begin to imitate observed actions in the early stages of their lives, which demonstrates that learning by imitation is a fundamental tool to help in the learning process and to improve social skills (Piaget, 1976). Some studies have shown that children with ASD have imitation impairments, probably due to a malfunction of the mirror neuron system (Williams, Whiten, Suddendorf, & Perrett, 2001). Besides the imitation impairments, children with ASD have difficulty in recognizing and understanding mental states in themselves and in others. Usually, children with ASD find it difficult to identify facial expressions and the emotions that they represent, to understand and control their own emotions, and also to interpret emotions. For example, a study conducted by (Hobson, 1986), in which children with ASD and typically developing children had to match drawings of gestures to videotaped vocalizations and facial expressions representative of four emotional states, showed that children with ASD had significantly more difficulty performing the task in comparison with typically developing children.

There is a big contrast in the development of children with ASD when compared with typically developing children, and for that reason they need timely and suitable intervention that responds to all of their impaired areas of development. Therefore, different approaches are used in intervention process with children with ASD. More recently, the use of social robots has been explored in order to facilitate intervention processes of children with ASD.

2.3 Human Robot Interaction (HRI) – Social Robots and ASD

Human-Robot Interaction (HRI) is an important topic in the scientific community. HRI considers the research where “intelligent” robots are used to improve the quality of life of a human being. The general idea lies on robots collaborating with humans, autonomously performing tasks and effectively communicating their “intentions”. One of the applications of HRI is on Autism Spectrum Disorders (ASD) research, where social robots help Users with special needs in their daily activities (Tapus, Member, & Scassellati, 2007). As referenced above, individuals with ASD have difficulties in social interaction, in particular in verbal and non-verbal communication (e.g. emotions and gestures expressions). Moreover, they have difficulties in putting themselves into someone else’s perspective and being aware of how to react to another’s feelings, in real time, i.e. they have difficulty to show empathy (Baron-Cohen & Wheelwright, 2004). In addition, children with ASD have deficits in imitation behaviours (Ingersoll, 2008). When face-to-face, people use a wide variety of sensory and motor modalities to communicate. Faces are the most noticeable social part of the human body. Following this trend, the face has received the most attention in HRI research since it is a key point in the communication of emotions from the early stages until the last stages of human life.

Researchers have used a variety of facially expressive robots in their work, including the research focusing on children with ASD. These robots offer a wide range in their expressivity, facial degrees-of-freedom (DoF), and visual appearance.

The work developed by (Mazzei *et al.*, 2011) consisted of a first stage in developing the humanoid robot FACE to allow children with ASD to deal with expressive and emotional information. FACE is a female android that is actuated by 32 servo motors moving the artificial skin. This allows human facial expressions to be re-created. Posteriorly, the system was tested with five children with ASD and fifteen typically developing children. The evaluated emotions were the six basic emotions. The results demonstrated that happiness, sadness and anger were correctly labelled with high accuracy for both children with ASD and typically developing children. Conversely, fear disgust, and surprise had not been labelled correctly, particularly by participants with ASD. The overall recognition rate for FACE with children with ASD was 60.0%, and the recognition results for each emotion were the following: anger – 100%, disgust – 20%, fear – 0%, happiness – 100%, sadness 100%, surprise – 40%. The results for FACE's recognition rates with typically developing children were: anger – 93%, disgust – 20%, fear – 46.7%, happiness – 93.3%, sadness – 86.7%, surprise – 40%, and the average of all emotions was 61.1%.

In (Sosnowski, Kuehnlenz, & Buss, 2006) it is presented EDDIE, a robotic head with 23 DoF, where actuators are assigned to the particular action units of the Facial Action Coding System (FACS). The system was evaluated by twenty-four participants: eight children from five to eight years old and sixteen adults from twenty-five to forty-eight years old. The study consisted of a multiple-choice test in which people should build a correspondence between six shown facial expressions to ten given answers. The average recognition rate for all emotions was 57.0%, and the recognition rates for each emotion were the following: anger – 54.0%, disgust – 58.0%, fear – 42.0%, happiness – 58.0%, sadness – 58.0%, surprise – 75.0%.

Similar to EDDIE, Kismet (Breazeal, 2000) is able to engage humans in expressive social interaction. It has 15 DoF and it was designed with the possibility to process a variety of social signals from visual and audio channels, and deliver social signals to the human with whom it was interacting. The values of the recognition rate were higher than those obtained with EDDIE. The average value for all six basic emotions was 73.0%, and the recognition rates for each emotion were the following: anger – 76.0%, disgust – 71.0%, fear – 47.0%, happiness – 82.0%, sadness – 82.0%, surprise – 82.0%.

SAYA (Hashimoto, Kobayashi, & Kato, 2011), a tele-operated android robot used in the role of a teacher, is capable of express human-like facial expression and perform some communicative functions with its head and eye movements. The face has 19 DoF for generating facial expressions. In order to evaluate the designed facial expressions, the system was initially tested by twenty adults that watched videos of SAYA performing each of the six basic emotions. The results showed a high recognition rate, 97.3%, for all the six basic emotions, and the recognition rates for each emotion were the following: anger – 92.0%, disgust – 92.0%, fear – 100.0%, happiness – 100.0%, sadness – 100.0%, surprise – 100.0%.

ZECA (Costa, Soares, & Santos, 2013), Zeno Engaging Children with Autism, is a humanoid robot that is used in a project called *Robótica Autismo* at University of Minho. This project, *Robótica Autismo*, emerged from a partnership between University of Minho and APPACDM association in Braga (*Associação Portuguesa de Pais e Amigos do Cidadão Deficiente*), and seeks to use robotic platforms to improve social skills of individuals with ASD. ZECA was used in a study with the purpose to analyse the use of a humanoid robot as a tool to teach recognition and labelling of emotions. This humanoid robot developed by *Robokind* has the ability to simulate facial expressions, since the face has 11 Degrees of Freedom (DoF) and it is covered with a polymeric material called *Frubber*. In order to evaluate the designed facial expressions two experiments were conducted. In the first one, the system was tested by forty-two typically developing children aged between 8 and 10 years old, group A, that watched videos of ZECA performing the following facial expressions: neutral, surprise, sadness, happiness, fear, and anger. Then sixty-one adults aged between 18 and 59 years old, group B, watched the same videos. Both groups completed a questionnaire that consisted in selecting the most appropriate correspondence for each video. The recognition rates of the facial expressions for group A were the following: anger – 26.191%, fear – 45.238%, happiness – 83.333%, neutral – 85.714%, sadness – 97.619%, surprise – 76.190%, and the average of all emotions was 69.048%. The recognition rates of the facial expressions for group B were the following: anger – 24.59%, fear – 77.049%, happiness – 91.803%, neutral – 90.164%, sadness – 91.803%, surprise – 86.607%, and the average of all emotions was 77.0%. The second experiment, consisted in showing similar videos of ZECA performing the same facial expression, but now with gestures. The recognition rates of the facial expressions improved in general, but with more impact for the emotions: fear (73.81%) and anger (47.619%). Similar to group A, the recognition rates in group B, in general, also improved. The recognition rates of the facial expressions, adding gestures, for group B were the following: anger – 70.492%, fear – 93.442%, happiness – 98.361%, neutral – 91.803%, sadness – 88.525%, surprise – 83.607%, and the average of all emotions was 77.0%.

Table 2.3-2 compares the facial expressions' recognition rates of all works presented above, where Sub – Subjects, TdC – Typically developing children, ASD_C – children with ASD, Ad – Adults, H – Happiness, Sa – Sadness, Su – Surprise, D – Disgust, F – Fear, A – Anger, N – Neutral, and Avg – Average.

(Costa, 2014), presented a child-robot interaction study, where children had to recognize and label emotions displayed by ZECA in two different game scenarios, involving imitation and storytelling activities. The goal of these scenarios was to help a child with ASD to acquire knowledge about different emotions and to improve their recognition skills. The Imitate Me game scenario consisted in ZECA displaying a facial expression, and then asking the child for showing that same facial expression. The experimenter used a wireless keypad to classify the answer as either successful or unsuccessful. In the Storytelling game scenario, social stories were told by ZECA, and the child had to identify the emotion felt by the character, in this case ZECA, in each story. Table 2.3-1 presents the mean and the corresponding standard deviation (SD) of the answers ("successful", "unsuccessful" and "unanswered") given by the children during three sessions.

Table 2.3-1 Percentage mean and the standard deviation (SD) of the answers ("successful", "unsuccessful" and "unanswered") given by the children during three sessions.

Session Number	**Successful Answers**	**Unsuccessful Answers**	**Unanswered Prompts**
Imitate Me 1	79.3 (20.0)	7.0 (6.1)	13.7 (15.2)
Storytelling 1	75.4 (29.4)	9.5 (16.5)	15.1 (14.4)
Imitate Me 2	100.0 (0.0)	0.0 (0.0)	0.0 (0.0)
Storytelling 2	61.7 (37.5)	19.2 (18.8)	19.2 (18.8)
Storytelling 3	70.6 (37.4)	19.1 (33.0)	10.3 (9.0)

In general, the results show that the children accomplished a good performance. In addition, and as expected, the storytelling game scenario provided more difficulties to the children as they have difficulties in recognizing emotional states in the others.

Table 2.3-2 Summary of the facial recognition rates for all the referred projects.

Robot	Sub	H	Sa	Su	D	F	A	N	Avg
FACE (Mazzei, Lazzeri, Hanson, & De Rossi, 2012)	TdC	93.3%	86.7%	40%	20%	46.7%	93%	-	61.1%
	ASD_C	100%	100%	40%	20%	0%	100%	-	60.1%
EDDIE (Sosnowski *et al.*, 2006) 	TdC and Ad	58%	58%	75%	58%	42%	54%	-	57%
Kismet (Breazeal, 2000) 	TdC and Ad	82%	82%	82%	71%	47%	76%	-	73%
SAYA (Hashimoto *et al.*, 2011) 	Ad	100%	100.0%	100.0%	92%	100%	92%	-	97.3%
ZECA (Costa *et al.*, 2013)	TdC	83.3%	97.6%	76.2%	-	45.2%	26.2%	85.7%	69%
	Ad	91.8%	91.8%	86.6%	-	77.1%	24.6%	90.2%	77%
	TdC	81%	83.3%	69%	-	73.8%	47.6%	78.6%	72.2%
	Ad	98.4%	88.5%	83.6%	-	93.4%	70.5%	91.8%	77%

*Sub – Subjects, TdC – Typically developing children, ASD_C – children with ASD, Ad – Adults, H – Happiness, Sa – Sadness, Su – Surprise, D – Disgust, F – Fear, A – Anger, N – Neutral, and Avg – Average

References

Alabbasi, H. A., Moldoveanu, P., & Moldoveanu, A. (2015). Real Time Facial Emotion Recognition using Kinect V2 Sensor. *IOSR Journal of Computer Engineering Ver. II*, *17*(3), 2278–661. http://doi.org/10.9790/0661-17326168

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Ambadar, Z., Prkachin, K., ... Theobald, B.-J. (2007). The Painful Face - Pain Expression Recognition Using Active Appearance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, *27*(12), 9–14. http://doi.org/10.1016/j.imavis.2009.05.007

Baltrusaitis, T. (2014). *Automatic facial expression analysis*. *Pattern recognition*. University of Cambridge. Retrieved from http://www.sciencedirect.com/science/article/pii/S0031320302000523

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. http://doi.org/10.1023/B:JADD.0000022607.19833.00

Breazeal, C. (2000). *Sociable machines: Expressive social exchange between humans and robots*. *Expressive Social Exchange Between Humans and Robots*. Retrieved from http://groups.csail.mit.edu/lbr/mars/pubs/phd.pdf

Cabanac, M. (2002). What is emotion? *Behavioural Processes*, *60*(2), 69–83. http://doi.org/10.1016/S0376-6357(02)00078-5

Cohn, J. (2007). Foundations of human-centered computing: Facial expression and emotion. *International Joint Conference on Artificial Intelligence*, 233–238. http://doi.org/10.1145/1180995.1181043

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders*. University of Minho.

Costa, S., Soares, F., & Santos, C. (2013). Facial expressions and gestures to convey emotions with a humanoid robot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8239 LNAI*, 542–551. http://doi.org/10.1007/978-3-319-02675-6_54

Darwin, C. (1872). *The expression of emotions in man and animals*. *Library* (Vol. 3). http://doi.org/10.1097/00000441-195610000-00024

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. *Consulting Psychologists Press*.

Ekman, P., Friesen, W. V, O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ... Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion. *Journal of Personality and Social Psychology*. US: American Psychological Association.

http://doi.org/10.1037/0022-3514.53.4.712

Ekman, P., & Rosenberg, E. (2005). *What the face reveals*. *What the Face Reveals*.

Elfenbein, H. A., & Ambady, N. (2003). Universals and cultural differences in recognizing emotions. *Current Directions in Psychological Science*, *12*(5), 159–164. http://doi.org/10.1111/1467-8721.01252

Emotient | A Leader in Emotion Measurement. (2015). Retrieved May 1, 2016, from http://www.emotient.com/

Ghimire, D., Lee, J., Li, Z. N., Jeong, S., Park, S. H., & Choi, H. S. (2015). Recognition of facial expressions based on tracking and selection of discriminative geometric features. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(3), 35–44. http://doi.org/10.14257/ijmue.2015.10.3.04

Hadar, U., Steiner, T. J., & Clifford Rose, F. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, *9*(4), 214–228. http://doi.org/10.1007/BF00986881

Happé, F., Briskman, J., Frith, U., Happé, F., & Frith, U. (2001). Exploring the cognitive phenotype of autism: weak "central coherence" in parents and siblings of children with autism: II. Real-life skills and preferences. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(3), 309–16. http://doi.org/10.1017/S0021963001006916

Hashimoto, T., Kobayashi, H., & Kato, N. (2011). Educational system with the android robot SAYA and field trial. *IEEE International Conference on Fuzzy Systems*, *8*(3), 766–771. http://doi.org/10.1109/FUZZY.2011.6007430

Hobson, R. P. (1986). The autistic child's appraisal of expressions of emotion: a further study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(February), 671–680. http://doi.org/10.1111/j.1469-7610.1986.tb01836.x

Hudlicka, E. (2008). Affective Computing for Game Design. *Scientist*, 5–12. Retrieved from http://www.imgd.wpi.edu/speakers/0910/Hudlicka_GAMEON_TUT_01.pdf

Ingersoll, B. (2008). The Social Role of Imitation in Autism. *Infants & Young Children*, *21*(2), 107–119. http://doi.org/10.1097/01.IYC.0000314482.24087.14

Lee, C.-H. J., Kim, K., Breazeal, C., & Picard, R. (2008). Shybot: friend-stranger interaction for children living with autism. *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, *2*, 3375–3380. http://doi.org/http://doi.acm.org/10.1145/1358628.1358860

Mao, Q.-R., Pan, X.-Y., Zhan, Y.-Z., & Shen, X.-J. (2015). Using Kinect for real-time emotion recognition via facial expressions *. *Front Inform Technol Electron Eng*, *16*(4), 272–282. http://doi.org/10.1631/FITEE.1400209

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., … De Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2011*, 4515–8. http://doi.org/10.1109/IEMBS.2011.6091119

Mazzei, D., Lazzeri, N., Hanson, D., & De Rossi, D. (2012). HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, 195–200. http://doi.org/10.1109/BioRob.2012.6290687

Michel, P., & El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. *Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI '03*, 258. http://doi.org/10.1145/958468.958479

Michel, P., & Kaliouby, R. El. (2000). Facial Expression Recognition Using Support Vector Machines. Retrieved from http://www.cs.cmu.edu/~pmichel/publications/Michel-FacExpRecSVMAbstract.pdf

Musicovery | Mobile Aplication. (2015). Retrieved May 1, 2016, from http://musicovery.com/

Noldus | Innovative solutions for behavioral research. (2015). Retrieved May 1, 2016, from http://www.noldus.com/

Piaget. (1976). *Piaget and His School*. (B. Inhelder, H. H. Chipman, & C. Zwingmann, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-46323-5

Picard, R. W. (1997). *Affective Computing*. *Rosalind W. Picard*. Retrieved from http://www.amazon.com/exec/obidos/ASIN/0262161702/acmorg-20

Scassellati, B., Henny Admoni, & Matarić, M. (2012). Robots for Use in Autism Research. *Annual Review of Biomedical Engineering*, *14*(1), 275–294. http://doi.org/10.1146/annurev-bioeng-071811-150036

Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. *Psychological Review.*, *69(5)*, 379–399. http://doi.org/10.1037/h0046234

Silva, C., Sobral, A., & Vieira, R. T. (2014). An automatic facial expression recognition system evaluated with different classifiers. *X Workshop de Visão Computacional (WVC'2014)*, 208–212. Retrieved from http://www.researchgate.net/publication/264295959_An_automatic_facial_expression_recognition_system_evaluated_with_different_classifiers

Sosnowski, S., Kuehnlenz, K., & Buss, M. (2006). EDDIE - An Emotion Display with Dynamic Intuitive Expressions. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3113–3118. http://doi.org/10.1109/IROS.2006.282208

Tapus, A., Member, S., & Scassellati, B. (2007). The Grand Challenges in Socially Assistive Robotics. *IEEE Robotics and Automation Magazine*, *14*, 1–7. http://doi.org/10.1109/MRA.2010.940150

The Bosphorus. (2009). Retrieved May 10, 2016, from http://bosphorus.ee.boun.edu.tr/Home.aspx

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews*. http://doi.org/10.1016/S0149-7634(01)00014-8

Youssef, A. E., Aly, S. F., Ibrahim, A. S., & Abbott, a. L. (2013). Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid. *International Journal of Modeling and Optimization*, *3*(2), 112–115. http://doi.org/10.7763/IJMO.2013.V3.247

Zhang, Y., Zhang, L., & Hossain, M. A. (2015). Adaptive 3D facial action intensity estimation and emotion recognition. *Expert Systems with Applications*, *42*(3), 1446–1464. http://doi.org/10.1016/j.eswa.2014.08.042

3 METHODOLOGIES

Summary

In this chapter are presented the methodologies used in the present work. It starts by presenting the methods used for extracting the facial features. Then, it introduces some machine learning methods, with emphasis on Support Vector Machines and in how to assess the performance of a classifier. Finally, it presents the methods used for validating systems that are capable of synthesizing facial expressions and recognizing emotions and also the methodology used for creating the database.

3.1 Facial Features Extraction

Facial expressions are innate in any communication and interaction between humans. They can transmit emotions, opinions, and clues regarding cognitive states. Several psychological studies have been conducted in order to decode the information contained in a facial expression. For example, the system developed by Ekman and Friesen (Ekman & Friesen, 1978), the Facial Action Codding System (FACS), allowed researches to analyse and classify facial expressions in a standardised framework. The process of extracting facial features has a strong influence on the accuracy and the computational cost of the overall emotions recognition system (Jamshidnezhad & Jan Nordin, 2012). Furthermore, the type of feature to be extracted and the corresponding methods are fundamental for the overall performance.

The present work uses geometrical methods combined with Action Units to extract the desired facial features.

3.1.1 Geometrical Methods

Geometric features are retrieved from selected landmarks positions of essential parts of the face (e.g. eyes, eyebrows, nose, and mouth) that are obtained from face features recognition techniques. These extraction methods, geometric feature extraction, are characterized by their simplicity and low computational cost, but their accuracy is extremely dependent on the face recognition performances. However, high accuracies on emotion detection usually requires a calibration with a neutral face or a manual grid nodes positioning (Ghimire *et al.*, 2015; Kotsia & Pitas, 2007).

The landmarks positions in the image space are used in different ways to extract the shape of facial features, and movements of facial features. The geometric features can be extracted on the variation in shape of the triangles, or ellipses (eccentricity features), (Loconsole, Miranda, Augusto, & Frisoli, 2004). Additionally, it is also possible to extract linear geometric features. In the current work, geometric features are extracted using a derive series of 2D distance features between key facial points, i.e., the geometric features are extracted by calculating linear distances between a couple of 2D points in the image space. Equation 1 defines the calculation of any motion-based facial feature in the 2D Euclidean space, where x and y are the coordinates of the i and j points.

$$d_{ij} = \sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \qquad (1)$$

3.1.2 Action Units

Generally, an emotion is expressed by a person when reacting to a certain event. Emotions are normally characterized as negative (sadness, anger or fear), positive (happiness or surprise) or neutral. Developed by Paul Ekman and Wallace Friesen in 1978, the earliest method for characterizing the physical expression of emotions is the Facial Action Coding System (FACS), (Ekman & Friesen, 1978). This system associates the action of the muscles to the changes in facial appearance. The measurements of the FACS are called Action Units (AUs) which are actions performed by a muscle or a group of muscles. There are a total of 46 AUs from which 12 are for the upper face and 18 are for the lower face. The AUs 1 through 7 refer to brows, forehead or eyelids (Ekman & Friesen, 1978; "FACS (Facial Action Coding System)," 2002). Each emotion is then characterized by a combination of the action of specific muscles.

The six emotions considered in this work are characterized by the following facial features, (Sacavém, António; Wezowski, Kasia; Wezowski, 2014):

- Happiness – muscle around the eyes are tightened, wrinkles appears around the eyes, cheeks are raised and lip corners are raised diagonally;
- Sadness – inner corner of the eyebrows is raised, eyelids are loose and lip corners are pulled down.
- Anger – eyebrows are pulled down, upper lids are pulled up, lower lids are pulled up and lips may be tightened.
- Surprise – entire eyebrows are pulled up, eyelids are also pulled up and mouth are widely open.
- Fear – eyebrows are pulled up and together, upper eyelids are pulled up and mouth is stretched.
- Neutral – face muscles are naturally relaxed, no micro expressions are detected.

FACS is an index of facial expressions, and the AUs are the fundamental actions of individual muscles or group of muscles. Table 3.1.2-1, shows the list of Action Units used in the present work (with underlying facial muscles). In Appendix A.2, it is available a more complete table of the Action Units.

Table 3.1.2-2, shows the six emotions considered in this work, as well as, the respective Action Units for each emotion.

Table 3.1.2-1 - List of Action Units (with underlying facial muscles) (Ekman & Friesen, 1978; "FACS (Facial Action Coding System)," 2002

AU#	FACS Name	Muscular Basis
0	Neutral Face	
1	Inner Brow Raiser	*Frontalis, pars medialis*
2	Outer Brow Raiser	*Frontalis, pars lateralis*
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*
5	Upper Lid Raiser	*Levator palpebrae superioris*
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*
12	Lip Corner Puller	*Zygomaticus major*
15	Lip Corner Depressor	*Depressor anguli oris (a.k.a. Triangularis)*
16	Lower Lip Depressor	*Depressor labii inferioris*
20	Lip Stretcher	*Risorius w/ platysma*
23	Lip Tightener	*Orbicularis oris*
26	Jaw Drop	*Masseter, relaxed Temporalis and internal Pterygoid*
51	Head turn left	
52	Head turn right	
53	Head up	
54	Head down	
55	Head tilt right	
56	Head tilt right	
57	Head forward	
58	Head back	
61	Eyes turn left	
62	Eyes turn right	
63	Eyes up	
64	Eyes down	

Table 3.1.2-2 - Facial Expressions and corresponding Action Units (AUs).

Facial Expression	Action Units (AUs)
Happiness	6 + 12
Sadness	1 + 4 + 15
Surprise	1 + 2 + 5 + 26
Fear	1 + 2 + 4 + 5 + 7 + 20 +26
Anger	4 + 5 + 7 + 23
Neutral	0

3.2 Classification methods

A classification problem determines if an object is (or is not) a member of a set, or to which set it belongs. An often approach to solve a classification problem is to use machine learning algorithms. Instead of following strictly static program instructions, machine learning algorithms operate by building a model from example inputs in order to make data-driven predictions or decisions expressed as outputs. Moreover, machine learning algorithms are suitable to work with multidimensional data, with the benefit of easily incorporating newly available data to improve prediction performance (Chih-Wei Hsu, Chih-Chung Chang, 2008).

Machine learning algorithms are typically applied under two major scenarios, the supervised learning, in which the desired outputs are known and are given to the learn algorithm for training, and the unsupervised learning, where the desired outputs are unknown, leaving the learn algorithm on its own to find the structure in the input data. The present work uses Support Vector Machine, SVM, classifier which is a supervised machine learning method, for classifying user's facial expressions. As referred in the literature (Michel & Kaliouby, 2000), SVM is a more suitable classifier to a dynamic, interactive approach to face expression recognition.

3.2.1 Support Vector Machine – SVM

Initially introduced by Vapnik (Burges, 1998), SVM is a supervised learning method capable of analysing data for classification and regression analysis. Usually implemented in binary classification, the classification is performed in the feature space through the construction of a linear separating hyperplane (Burges, 1998). Therefore, the goal of the SVM is to find an Optimal Separating Hyperplane (OSH) that divides all the data points from one class to the other class. This is achieved by finding the largest margin between two classes, which is the OSH (Figure 3.2.1-1). In order to obtain the maximum width of the margin, an optimization problem is computed. This problem is controlled by a parameter C, a trade-off between the maximum width of margin and minimum classification error (Burges, 1998). The data points closest to the separating hyperplane are the support vectors (SVs). However, some binary classification problems do not have a simple hyperplane as a useful separating criterion. Thus, SVM employs kernel methods to map data to a higher dimensional feature space, allowing nonlinear classification. The commonly used kernels functions of the SVM, that define the nature of the decision surface, are: linear (equation 2), polynomial (equation 3), radial basis function (RBF) (equation 4), and sigmoid (equation 5), where x means the data matrix (observations vs. features).

$$linear: K(x_i, x_j) = x_i^T x_j \quad (2)$$
$$polynomial: K(x_i, x_j) = (\Upsilon x_i^T x_j + r)^d, \Upsilon > 0 \quad (3)$$
$$radial\ basis\ function\ (RBF): K(x_i, x_j) = exp\left(-\Upsilon \|x_i - x_j\|^2\right), \Upsilon > 0 \quad (4)$$
$$sigmoid: K(x_i, x_j) = tanh(\Upsilon x_i^T x_j + r) \quad (5)$$

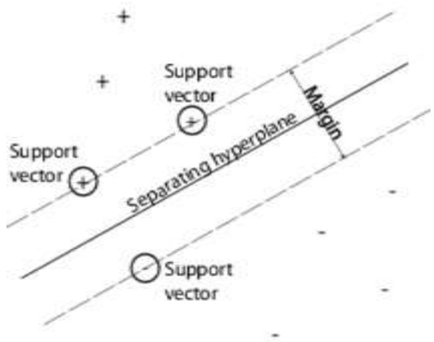

Figure 3.2.1-1 Optimal Separating Hyperplane ("Support Vector Machines (SVM) - MATLAB & Simulink," 2015).

Accordingly to (Michel & Kaliouby, 2000), SVM show high classification accuracy even when a small amount of training data is available, making them particularly suitable to a dynamic, interactive approach to face expression recognition. The trained classifier provides a model that can be used to predict the emotion corresponding to a set of features.

The accuracy of the SVM's classification depends on the chosen value for C parameter, and also the gamma (Υ) value when RBF kernel is used. Frequently, it is employed a grid-search (exhaustive search by means of optimization problems) to find the best C and gamma (Υ) values that minimize the classification error. It is also recommended to combine the grid-search with cross-validation in order to find the best values of C and gamma (Υ) (Chih-Wei Hsu, Chih-Chung Chang, 2008).

The C parameter trades off misclassification of training examples against simplicity of the decision surface. Small values of C (Figure 3.2.1-2, on the right) will cause the optimizer to look for a larger-margin separating hyperplane, even if that hyperplane misclassifies more points. Conversely, a large value of C (Figure 3.2.1-2, on the left) the optimizer will only consider points close to line of separation (this can be applied when the data are less reliable), i.e., the optimization will choose a smaller-margin hyperplane if that hyperplane does a better job of getting all the training points classified correctly.

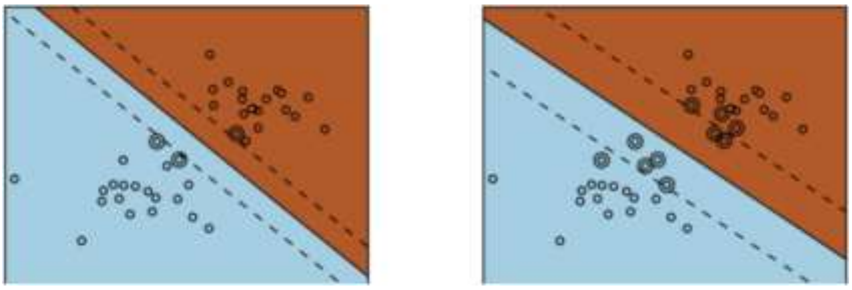

Figure 3.2.1-2 The influence of the parameter *C*: on the left for a higher value of *C*, on the right for a lower value of *C* ("SVM Margins Example," 2014).

For the RBF kernel, the gamma (Υ) parameter defines the extent of the influence of a single training example, with low values meaning 'far' and high values meaning 'close'. For large gamma (Υ) values the radius of the area of influence of the support vectors only includes the support vector itself. Conversely, when gamma (Υ) is very small, the model is too constrained and cannot capture the complexity or "shape" of the data, i.e., the region of influence of any selected support vector would include the whole training set, the resulting model would present a linear behaviour ("RBF SVM parameters — scikit-learn 0.17.1 documentation," 2015).

Although SVM, by default, is a binary classifier it has been proposed with success for multiclass classification with "one-against-one" (OAO) and "one-against-all" (OAA) approaches (Baltrusaitis, 2014; Kotsia & Pitas, 2007; Loconsole *et al.*, 2004; Mao, Pan, Zhan, & Shen, 2015; Michel & El Kaliouby, 2003). The OAA approach constructs one SVM model or binary learner per class, distinguishing the samples of one class from all remaining classes, i.e., for each binary learner one class is positive and the rest are negative. In this case, with the OAA approach are created K binary learners or SVM models, where K is the number of distinct classes (Hsu & Lin, 2002; Milgram, Cheriet, & Sabourin, 2006). Conversely, the OAO approach constructs one SVM model or binary learner for each pair of classes. This means that in the OAO for each SVM model or binary learner, one class is positive another is negative, and the rest ignored, being created $K\frac{K-1}{2}$ SVM models or binary learners, as result of all possible combinations between K classes (Milgram *et al.*, 2006).

There are two stages involved in any binary and multiclass classification, the training stage and the test stage. In the training stage, the SVM takes a set of input data and classifies it within a set of different possible classes, and by means of optimization it finds the OSH. Then, it builds a model that assigns data into each class. In the test stage, the SVM refers to the calculated classification outputs against the known class labels for data that were not used in training. Lastly, the SVM works as a non-probabilistic linear classifier since it predicts the set of classes based on an optimization problem; it is not based on a probabilistic model. The primary advantage of SVM is the convergence to a global optimal, avoiding the local minima and over-fitting in the training process (Wu, Wang, & Liu, 2007). Thus, even with a limited training data set, SVM has the ability to minimize both structural and empirical risk leading to a better generalization for new data classification, producing stable and reproducible results (Michel & Kaliouby, 2000). One disadvantage of the SVM classifier lies in the performance dependence on internal learning parameters (e.g. regularization parameter for SVM), which can be difficult to interpret (Chih-Wei Hsu, Chih-Chung Chang, 2008).

3.3 Classifier Performance

Usually validation methods are used to assess the performance of a machine learning algorithm, with intend to evaluate their generalization, mainly when the data sets are limited, and to test their ability to classify new instances. Training a model and testing it with the same data is a methodological mistake: a model would achieve a perfect score, but would fail to predict new data. This situation is called overfitting. In order to avoid it, it is common practice to partition the available data into two sets: a training set and a test set. An automatic and most common method of preventing overfitting and validating the model is the k-Fold cross-validation (CV) method ("Cross-validation: evaluating estimator performance," 2015).

The k-Fold CV method divides the total data set in k subsets, and one subset is retained as the validation data for testing the model and the remaining subsets (k-1) are used as training data. Then the cross-validation process is repeated k times (the folds), with each of the k subsets used once as the validation data ("Cross-Validation - MATLAB," 2015). The k results from the folds can be averaged (or otherwise combined) to produce a single estimation. The main advantage of this method is that all observations are used for both training and validation, and each observation is used for validation exactly once. CV contributes for the generalization of the classifier and also it avoids overfitting.

3.3.1 Performance Metrics

There are many ways to quantify/measure the classifier performance. The most popular metrics are: accuracy, the confusion matrix, sensitivity, specificity, and the Area under the Curve (AUC) metric.

Accuracy is the most common and it simply measures how often the classifier makes the correct prediction, where TP, FP TN, and FN correspond to true positive, false positive, true negative, and false positive, respectively (Zheng, 2015). It is the ratio between the number of correct predictions (TP and TN) and the total number of predictions (TP, TN, FP, and FN), equation 6.

$$accuracy\ (\%) = \frac{TN+TP}{TP+TN+FP+FN} \times 100\% \tag{6}$$

However, when a data set is unbalanced (where the number of samples in each class is significantly different) the accuracy evaluated of a classifier is not representative of the true performance of the classifier, i.e., the accuracy will give a distorted picture, because the class with more examples will dominate the statistic. In this case, it is advised to look at per-class accuracy, both the average and the individual per-class accuracy numbers (Zheng, 2015). Accuracy is a simple metric, but it makes no distinction between classes, which a confusion matrix does.

A confusion matrix is a specific table layout, Figure 3.3.1-1, which allows visualization of the performance of a learning algorithm, usually a supervised one (Zheng, 2015). The confusion matrix shows in more detail a breakdown of correct and incorrect classification for each class. The rows of the matrix correspond to ground truth labels (the actual class), and the columns represent the prediction (predicted class).

	p' (Predicted)	n' (Predicted)
p (Actual)	True Positive	False Negative
n (Actual)	False Positive	True Negative

Figure 3.3.1-1 Confusion matrix, where each row correspond to the actual class and each column to the predict class ("Confusion matrix.png," 2015).

It is possible to acquire the following measures from a confusion matrix: TP (true positive), TN (true negative), FP (false positive), and FN (false negative). From these measures, it is possible to obtain two more metrics – sensitivity and specificity.

Sensitivity, equation 7, measures the proportion of actual positives which are correctly identified as such (Zheng, 2015).

$$sensitivity\ (\%) = \frac{TP}{TP+FN} \times 100\% \qquad (7)$$

Specificity, equation 8, measures the proportion of negatives which are correctly identified as such (Zheng, 2015).

$$specificity\ (\%) = \frac{TN}{TN+FP} \times 100\% \qquad (8)$$

Another metric for assess the classifier performance is the Area Under the Curve (AUC). This metric is obtained through integration by the trapezoid method with base in the Receiver Operating Characteristic curve (Figure 3.3.1-2), or ROC curve (Zheng, 2015). The ROC curve allows visualizing the trade-off between the TP rate and FP rate, i.e., it shows how many correct positive classifications can be gained as more and more false positives are allowed. The AUC value of a perfect classifier, that makes no mistakes, would be 1. The AUC represents the average sensitivity across all possible specificities.

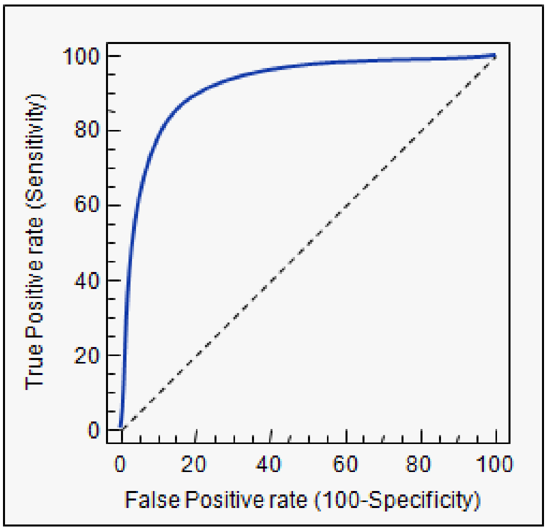

Figure 3.3.1-2 ROC curve - trade-off between the TP rate and FP rate (“roc_intro3.png,” 2015).

An extra metric to assess the performance of a classifier is the Matthews Correlation Coefficient (MCC). The MCC is a correlation coefficient between the actual classes and the predict, that takes into account the true positives (TP), true negatives (TN), false positives (FP), and false negatives (FN), and it is generally regarded as a balanced measure, which can be used even if the classes are unbalanced (with different sizes), (Jurman & Furlanello, 2010). It returns a value between -1 and +1, where +1 represents a perfect prediction, 0 no better than a random prediction, and -1 indicates total disagreement between prediction and observation. The MCC is generally regarded as being a measure that can summarize/describe a confusion matrix by a single number. The MCC can be calculated directly from the confusion matrix using equation 9, for binary classification.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \quad (9)$$

The quality of the overall performance of a multi-class classifier, is usually assessed in two ways: a measure is the average of the same measures calculated for the classes C_1,..., C_i, using one-against-all approach (*macro-averaging*), or the sum of counts to obtain cumulative TP, TN, FP, FN and then calculating a performance metric (*micro-averaging*) (Sokolova & Lapalme, 2009). Macro-averaging treats all classes equally while micro-averaging benefits bigger classes (Sokolova & Lapalme, 2009). For multiclass classification the MCC metric is usually calculated by using the method proposed by (Jurman & Furlanello, 2010).

3.4 Systems Validation

In the literature are presented some standard methods for validating systems that can synthesize and recognize facial expressions. In the next subchapters are presented the most common procedures to validate these systems.

3.4.1 Facial Expressions Imitation

Usually, two approaches are used in order to quantify the performance of a system that can synthesize facial expressions. The first approach consists in using computational based evaluation, i.e., use a software for automatically analyse synthesized facial expressions by comparing each synthesized expression with a predefined computer model. Thus, when designing such evaluation method, it is necessary to develop an accurate computer model of facial parts for each facial expressions (Moosaei & Riek, 2013). Researches have worked on developing a computational model of facial parts such as muscular and skin models or lip shapes (Lee, Terzopoulos, & Walters, 1995; Moosaei, Hayes, & Riek, 2015; Ochs, Niewiadomski, Pelachaud, & Sadek, 2005). The computational evaluation aims to provide a fast and fully automatic evaluation method which does not require human judges. However, designing an accurate model for each facial part is very complicated.

The second approach consists in performing a subjective evaluation, i.e., subjects judge the synthesized expressions. This is the most commonly approach used in literature (Becker-Asano & Ishiguro, 2011; Costa, Soares, & Santos, 2013; Moosaei & Riek, 2013), and consists in subjects observing synthesized facial expressions and then answer a predefined questionnaire. By analysing collected answers, researches evaluate the expressions of their robot or virtual avatar. Even though user studies are costly and time consuming, they can provide valuable information about the acceptability and believability of a robot (Moosaei & Riek, 2013). When designing a subjective evaluation, it is necessary to take into account methodological issues when choosing the participants, such as: their average age, their educational level, their cultural background, their native language, and the gender of participant. Another experimental design concern is how to select subjects to participate in the evaluation. Christoph, (Christoph, 2005), suggests that an ideal group of subjects are those most likely to use the robot in the future. For example, a robotic platform that is going to be used for interactions with children, should be evaluated by them since the robot is intended to be useful for them. There are several examples in literature of strong experimental design for subjective synthesis evaluation. One example is the study conducted by Becker-Asano and Ishiguro, (Becker-Asano & Ishiguro, 2011), in order to compare expressivity of Geminoid F's six facial expressions (anger, fear, happiness, sadness, surprise and neutral) with expressions of the real model person. The perceptual study consisted in subjects matching one of the six labels (anger, fear, happiness, sadness, surprise or neutral) to each different facial expression expressed by the robot. The researchers also performed the experiment in German, Japanese and English language with the objective to study intercultural differences.

Another type of subjective evaluation is side-by-side comparison in which subjects have to visually compare the robot's synthesized facial expression with the performer's facial expression, and answer a set of questions (Moosaei & Riek, 2013).

3.4.2 Emotion recognition

Generally, there are two types of evaluation for assessing the performance of an emotion recognition system: the static off-line and the real-time evaluations. In the off-line evaluation the machine learning model is trained and tested and the performance is usually assessed by using the performance metrics described in the previous section.

For the real-time evaluation, the system in a first stage is trained and the training computational cost is measured. Then a set of participants, which did not participate in the database creation process used for training, are recruited. The usual experimental set-up consists in each participant performing the emotion requested by the researcher in front of a sensor. Then this data is used to quantify the performance, usually in terms of accuracy per class and average accuracy. It is also assessed the real-time computational cost of the system. The real-time evaluation, generally, focuses more on the assessment of the system's real-time performance.

3.5 Database Creation

Generally, the methodology employed for creating the database consists firstly in recruiting a set of participants. Then, the usual experimental set-up for extracting the facial features consists in each participant performing the emotion requested by the researcher in front of a sensor. Usually, in order to serve as a clue, the researcher shows a photograph corresponding to the emotion requested. The acquired data are then saved into a file.

Since SVM algorithms are not scale invariant (Burges, 1998), it is recommended to scale the extracted data to a normalized range, by applying the equation 10, where *N* is the normalized facial feature value, *W* is the facial feature value to normalize, min_v is the minimum value from the normalized set, max_v is the maximum value from the normalized set, and *[A, B]* is the range for the value (*W*) after normalization.

$$N = \frac{W - min_v}{max_v - min_v}(B - A) + A \tag{10}$$

References

Baltrusaitis, T. (2014). *Automatic facial expression analysis*. *Pattern recognition*. University of Cambridge. Retrieved from http://www.sciencedirect.com/science/article/pii/S0031320302000523

Becker-Asano, C., & Ishiguro, H. (2011). Evaluating facial displays of emotion for the android robot Geminoid F. *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - WACI 2011: 2011 Workshop on Affective Computational Intelligence*, (20220002), 22–29. http://doi.org/10.1109/WACI.2011.5953147

Burges, C. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and*

Knowledge Discovery, *2*(2), 121–167. http://doi.org/10.1023/A:1009715923555

Chih-Wei Hsu, Chih-Chung Chang, and C.-J. L. (2008). A Practical Guide to Support Vector Classification. *BJU International*, *101*(1), 1396–400. http://doi.org/10.1177/02632760022050997

Christoph, N. (2005). Empirical Evaluation Methodology for Embodied Conversational Agents. In *From Brows to Trust SE - 3* (Vol. 7, pp. 67–90). http://doi.org/10.1007/1-4020-2730-3_3

Confusion matrix.png. (2015). Retrieved May 22, 2016, from http://3.bp.blogspot.com/_txFWHHNYMJQ/THyADzbutYI/AAAAAAAAAf8/TAXL7IySrko/s1600/Picture+8.png

Costa, S., Soares, F., & Santos, C. (2013). Facial expressions and gestures to convey emotions with a humanoid robot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8239 LNAI*, 542–551. http://doi.org/10.1007/978-3-319-02675-6_54

Cross-Validation - MATLAB. (2015). Retrieved January 20, 2016, from http://www.mathworks.com/discovery/cross-validation.html

Cross-validation: evaluating estimator performance. (2015). Retrieved May 22, 2016, from http://scikit-learn.org/stable/modules/cross_validation.html

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. *Consulting Psychologists Press*.

FACS (Facial Action Coding System). (2002). Retrieved January 25, 2016, from http://www.cs.cmu.edu/~face/facs.htm

Ghimire, D., Lee, J., Li, Z. N., Jeong, S., Park, S. H., & Choi, H. S. (2015). Recognition of facial expressions based on tracking and selection of discriminative geometric features. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(3), 35–44. http://doi.org/10.14257/ijmue.2015.10.3.04

Hsu, C., & Lin, C. (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, *13*(2), 415–425. http://doi.org/10.1109/TNN.2002.1021904

Jamshidnezhad, A., & Jan Nordin, M. D. (2012). Challenging of facial expressions classification systems: Survey, critical considerations and direction of future work. *Research Journal of Applied Sciences, Engineering and Technology*, *4*(9), 1155–1165.

Jurman, G., & Furlanello, C. (2010). A unifying view for performance measures in multi-class prediction. *PLoS ONE*, *7*(8). http://doi.org/10.1371/journal.pone.0041882

Kotsia, I., & Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, *16*(1), 172–187. http://doi.org/10.1109/TIP.2006.884954

Lee, Y., Terzopoulos, D., & Walters, K. (1995). Realistic modeling for facial animation. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH 95*, *95*, 55–62. http://doi.org/10.1145/218380.218407

Loconsole, C., Miranda, C. R., Augusto, G., & Frisoli, A. (2004). Real-Time Emotion Recognition: a Novel Method for Geometrical Facial Features Extraction.

Mao, Q.-R., Pan, X.-Y., Zhan, Y.-Z., & Shen, X.-J. (2015). Using Kinect for real-time emotion recognition via facial expressions *. *Front Inform Technol Electron Eng*, *16*(4), 272–282. http://doi.org/10.1631/FITEE.1400209

Michel, P., & El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. *Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI '03*, 258. http://doi.org/10.1145/958468.958479

Michel, P., & Kaliouby, R. El. (2000). Facial Expression Recognition Using Support Vector Machines. Retrieved from http://www.cs.cmu.edu/~pmichel/publications/Michel-FacExpRecSVMAbstract.pdf

Milgram, J., Cheriet, M., & Sabourin, R. (2006). "One Against One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs? *Tenth International Workshop on Frontiers in Handwriting Recognition*, 1–6. Retrieved from http://hal.inria.fr/inria-00103955

Moosaei, M., Hayes, C. J., & Riek, L. D. (2015). Performing Facial Expression Synthesis on Robot Faces: A Real-time Software System. *New Frontiers in Human-Robot Interaction*, 67. Retrieved from http://www.mahasalem.net/AISB2015/NF-HRI-2015-full_proceedings.pdf#page=68

Moosaei, M., & Riek, L. D. (2013). Evaluating Facial Expression Synthesis on Robots. *HRI Workshop on Applications for Emotional Robots at the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI).*, 1–6.

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2005). Intelligent expressions of emotions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3784 LNCS*, 707–714. http://doi.org/10.1007/11573548_91

RBF SVM parameters — scikit-learn 0.17.1 documentation. (2015). Retrieved May 20, 2016, from http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html#example-svm-plot-rbf-parameters-py

roc_intro3.png. (2015). Retrieved May 26, 2016, from https://www.medcalc.org/manual/_help/images/roc_intro3.png

Sacavém, António; Wezowski, Kasia; Wezowski, P. (2014). *A Linguagem Corporal Revela o que as Palavras Escondem*. Top Books. Retrieved from http://www.wook.pt/ficha/a-linguagem-corporal-revela-o-que-as-palavras-escondem/a/id/15612673

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. http://doi.org/10.1016/j.ipm.2009.03.002

Support Vector Machines (SVM) - MATLAB & Simulink. (2015). Retrieved January 20, 2016, from http://www.mathworks.com/help/stats/support-vector-machines-svm.html

SVM Margins Example. (2014). Retrieved from http://scikit-learn.org/stable/auto_examples/svm/plot_svm_margin.html

Wu, J., Wang, J., & Liu, L. (2007). Feature extraction via KPCA for classification of gait patterns. *Human Movement Science*, *26*(3), 393–411. http://doi.org/10.1016/j.humov.2007.01.015

Zheng, A. (2015). How to Evaluate Machine Learning Models: Classification Metrics. Retrieved May 22, 2016, from http://blog.dato.com/how-to-evaluate-machine-learning-models-part-2a-classification-metrics

4 DEVELOPMENT

Summary

The following chapter describes the adopted materials and general procedures used in the present work. It starts by detailing the hardware and software used in the work. Then, it presents a global overview of the system, detailing the implementation of the two subsystems. Finally, the Graphical User Interface and the software architecture are presented.

4.1 Hardware and Software description

The hardware adopted in the present work consists in a robotic platform, the Zeno R-50 from *Robokind* (“Robokind | Advanced Social Robots,” 2015), and a 3D sensor, the Intel RealSense 3D sensor from Intel.

4.1.1 Robotic platform

The robotic platform used in the present work is a humanoid child-like robot, Figure 4.1.1-1, with 34 degrees of freedom: 4 are located in each arm, 6 in each leg, 11 in the head, and 1 in the waist (“R50 Documentation -,” 2014). The major feature that distinguishes this robot from the others is the ability to express emotions thanks to servo motors mounted on its face and a special material, *Frubber*, which looks and feels like human skin. Table 4.1.1-1 summarizes the principal characteristics of the Zeno R-50 robot.

Figure 4.1.1-1 Zeno R-50.

The *Robokind* software performs animation and motion control functions and it includes an Application Programming Interface (API) for rapid integration of the other components, distributed computation and shared control. Additionally, it is also available a Workshop (Figure 4.1.1-2), a program where it is possible to create animations by controlling each servo motor characteristic curve, which can subsequently be used in the robot programming. Technical drawings of the robot can be found in the Appendix A.3. Henceforth, the robot is going to be referred as ZECA (Zeno Engaging Children with Autism).

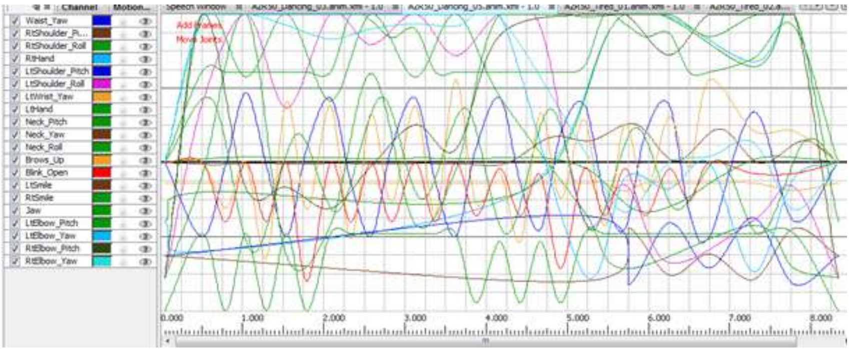

Figure 4.1.1-2 Workshop, where it is possible to develop an Animation, by controlling every DoF of the robot (Meah, 2014).

Table 4.1.1-1 Principal characteristics of the Zeno R-50 robot from Robokind.

Physical Characteristics		**Multimedia**	1 Loud Speaker
Height	69 cm		3 Microphones
Weight	5,7 kg		2 CMOS Digital Hi-Def Cameras (720p)
Degrees of Freedom (DOF)		**Network Access**	
Total	36	Wi-Fi (IEEE 802.11a/b/g/n)	
Head	11	Ethernet Connection (Gigabit/100/10)	
Arm	4	**Motherboard**	
Waist	1	*Intel Atom* x86 Z530 1,6 GHz processor	
Leg	6	RAM 1GB DDR2	
Actuators		*Flash* memory 4 GB + 16 GB micro SD	
Dynamixel	(x10) RX – 64	**Available ports**	
	(x11) RX – 28	Audio: *Stereo* Line (*in* and *out*)	
PWM Servo	(x11)	(2x) USB 2.0	
Sensors		Video: (1x) HDMI (*out*)	
Available Types	1x gyro meter 3 axes	**Software Compatibilities**	
	1x accelerometer 3 axis	OS	Ubuntu Linux (32 bit x86)
	1x compass 3 axes		
	2x bumpers	Programing Language	Java
	2x ground contact		
	2x cliff	Software	Windows and Linux compatible
	1x IR proximity		
	21x Potentiometer		

4.1.2 Intel RealSense 3D Camera

Intel RealSense is a platform for implementing gesture-based Human Computer Interaction (HCI) techniques (“Intel® RealSense™ Technology,” 2015). It contains a conventional RGB camera, an infrared laser projector, an infrared camera, and a microphone array (Figure 4.1.2-1). A grid is projected onto the scene by the infrared projector and the infrared camera records it, computing the depth information. The microphone array allows localizing sound sources in space and performing background noise cancellation. Intel announced two models with distinct specifications: F200, a front facing sensor, and R200, a rear facing sensor. The present work uses the Intel RealSense model F200, which is a front facing device with a host of capabilities including, but not limited to, gesture interaction, facial recognition, and voice command. This device, along with the required Windows software, Intel RealSense SDK, was used to obtain the face data from the user.

Figure 4.1.2-1 Intel RealSense 3D Camera, model F200 (“realsense_intel,” 2015).

Along with the camera, Intel provides a Software Development Kit (SDK) that possesses many capabilities, such as, facial recognition, hand gesture, background removal, depth enabled phot, scene perception, 3D scanning, and much more. The Integrated Development Environment (IDE) is the Microsoft’s Visual Studio, where it is possible to use a set of programming languages (C++, C# Processing, and Java) to develop an application.

4.2 System Overview

The system implemented in this work (Figure 4.2-1) consists of an Intel RealSense camera, a computer, and ZECA.

Figure 4.2-1 . Experimental setup. Starting from the left: Intel RealSense (model F200), computer and ZECA robot.

Figure 4.2-2 shows the block diagram of the developed system. The face data (the user facial Action Units, AUs) obtained from the Intel RealSense sensor is filtered. Then, the data is normalized and send to one of the two subsystems: Mirroring Emotions System (MES) or Emotions Recognition System (ERS). An application developed in C# runs on the computer, computing the user face data and establishing the communication with ZECA. A Java application runs in the robot in order to execute the designed program.

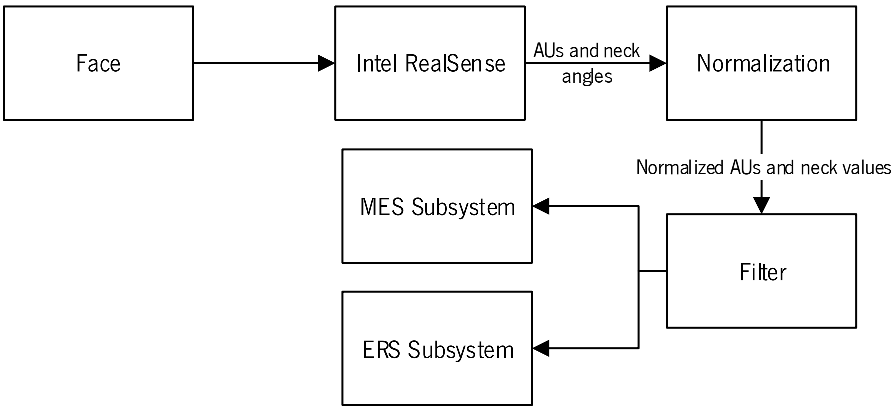

Figure 4.2-2 Overall system block diagram. First the AUs and neck angles are obtained and normalized. Finally, the values are filtered and send to one of the two subsystems.

4.2.1 Obtaining the AUs and Head motion

Intel RealSense is able to calculate the scores for a few supported facial expressions as well as detecting up to 78 facial landmarks using the depth information. The user must be in front of the Intel RealSense as shown in Figure 4.2.1-1. The effective range for face tracking is 20 cm to 120 cm.

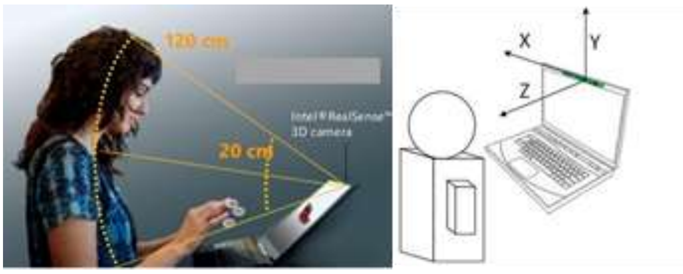

Figure 4.2.1-1 On the left: the recommended user position. On the right: the Intel RealSense coordinate system (“Camera range,” 2015).

However, as Intel RealSense is still under development, some facial expressions are detected with low accuracy and some expressions are not yet available. In order to solve this problem, in this work, certain facial expressions were acquired through facial landmarks. The landmarks position in the image space can be used in different ways in order to extract the shape of facial features, and the movements of facial features, also called geometric features.

Table 4.2.1-1 lists the facial AUs from Intel RealSense that were used in this work, differentiating those provided by the Intel RealSense SDK from the ones obtained through facial landmarks (geometric features).

In order to obtain the geometric features from landmarks, ten facial landmarks were selected. Figure 4.2.1-2 shows the selected and labelled facial landmarks for geometric features extraction. Table 4.2.1-2 lists the significance of the selected 10 facial landmarks. These geometric features were determined by calculating linear distances between couples of landmarks (Eq. 11, Eq. 12, Eq. 13, Eq. 14, and Eq. 15), using the method in 3.1.1 described in the previous chapter.

Table 4.2.1-1 Selected Facial AUs.

Facial AUs	**Intel RealSense**	**Facial landmarks**
Eye brow raiser (left and right)	✔	
Eye brow lower (left and right)	✔	
Eyelids		✔
Eyes up	✔	
Eyes down	✔	
Eyes left	✔	
Eyes right	✔	
Mouth open		✔
Lip stretcher	✔	
Lip depressor (left and right)		✔

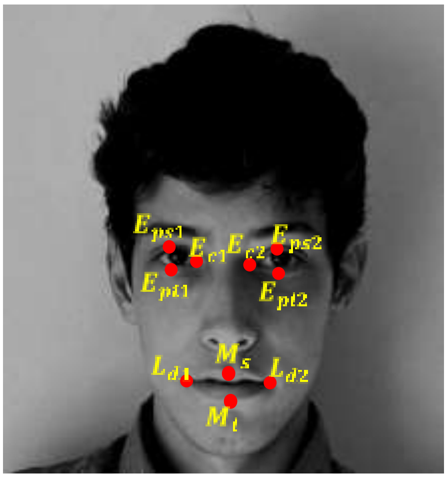

Figure 4.2.1-2 Selected facial landmarks.

$$Lip\ depressor_{right} = \sqrt{\left(E_{c1_x} - L_{d1_x}\right)^2 + \left(E_{c1_y} - L_{d1_y}\right)^2} \tag{11}$$

$$Lip\ depressor_{left} = \sqrt{\left(E_{c2_x} - L_{d2_x}\right)^2 + \left(E_{c2_y} - L_{d2_y}\right)^2} \tag{12}$$

$$Eye_{right} = E_{ps1_y} - E_{pi1_y} \tag{13}$$

$$Eye_{right} = E_{ps1_y} - E_{pi1_y} \tag{14}$$

$$MouthOpen = M_{s_y} - M_{i_y} \tag{15}$$

Table 4.2.1-2 Facial landmarks significance.

Facial Landmark	**Significance**
E_{c1}	Eye corner
E_{c2}	Eye corner
L_{d1}	Lip depressor
L_{d2}	Lip depressor
E_{ps1}	Eye palpebrale superius
E_{ps2}	Eye palpebrale superius
E_{pi1}	Eye palpebrale inferius
E_{pi2}	Eye palpebrale inferius
M_S	Mouth superius
M_I	Mouth inferius

The facial AUs obtained from Intel RealSense are normalized in a 0 to 100 intensity scale, where 0 means that the AU is not present and 100 means that the AU is definitely present.

The geometric features that were obtained through facial landmarks are normalized, using equation 10 from the previous chapter, to the scale 0 to 100. To find the minimum and maximum values for each of these geometric features, a performer was asked to execute a wide range of extreme facial movements while seated in front of the Intel RealSense sensor. For example, in order to find the maximum value of the geometric feature Eyelid, the person opened his eyes as much as he could. The minimum value of the geometric feature is zero, corresponding to the eyes closed. This way, the minimum and maximum values for each geometric feature were experimentally found.

The Intel RealSense SDK can return the user head angles, Euler angles (“Face Pose Data [F200,SR300],” 2015). The Euler angles indicate where the face pose is in terms of rotation along the three axes, as illustrated in Figure 4.2.1-3. This allows obtaining the User’s head motion, which is important in the emotion communication process.

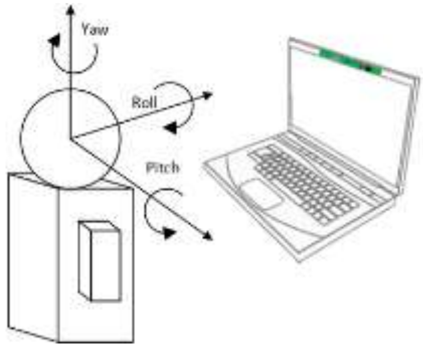

Figure 4.2.1-3 Face pose angles (pitch, roll, and yaw) according to the Intel RealSense coordinate system ("Face Pose Data [F200,SR300]," 2015).

4.2.2 Filter and Communication Protocol

In order to smooth out short-term fluctuations due to rapid facial movements, a moving average filter (N=10) was applied for each AU value. In Figure 4.2.2-1 it is possible to see the effect of this filter in one AU value of a facial expression, in this case the eyebrow movement AU value.

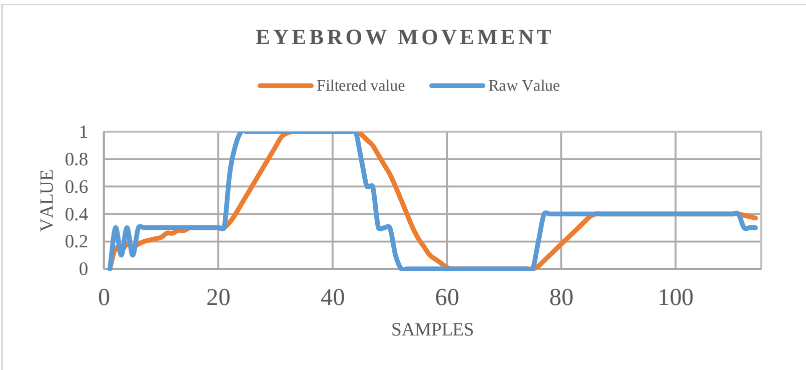

Figure 4.2.2-1 Comparison between the raw eyebrow value with the normalized eyebrow value.

The Intel RealSense subsystem is connected to the robot ZECA through a local network, communicating via a TCP/IP socket. A client application that runs on the computer transmits the data to the applications that runs on the robot.

4.2.3 Designing the MES subsystem

The MES subsystem is capable of on-line synthetizing human emotions through facial expressions. This subsystem after extracting the user's facial AUs and head motion data, maps each facial AU to a servo motor or group of servo motors. An AU is linked to actions of an individual muscle or group of muscles.

Therefore, each AU was associated to a servo motor or a group of servo motors responsible for a facial movement on the face of ZECA (Figure 4.2.3-1). This association can be seen in Table 4.2.3-1.

Figure 4.2.3-1 Mapping of the servo motors on the face and the neck of ZECA (Costa, 2014).

Table 4.2.3-1 Matching of the AUs and the servo motors of ZECA.

Facial AUs	**Servo motors of ZECA**
Eye brow raiser	Servo Motor 4
Eye brow lower	Servo Motor 4
Eyelids	Servo Motor 3
Eyes up	Servo Motor 5
Eyes down	Servo Motor 5
Eyes left	Servo Motor 8 and Servo Motor 10
Eyes right	Servo Motor 8 and Servo Motor 10
Mouth open	Servo Motor 6
Lip stretcher	Servo Motor 11 and Servo Motor 9
Lip depressor	Servo Motor 11 and Servo Motor 9

Due to the robot mobility limitations, it is necessary to delimit the movements of the neck. For example, the maximum range for the servos 0, 1, and 2, which are responsible for the yaw, roll, and pitch angles, respectively, is limited as can be seen in the Figure 4.2.3-2 (Costa, 2014). This was taken into consideration. Posteriorly, all facial AUs were converted to a scale 0-1. The conversion to this scale is made according to the API (Application Programming Interface) of the robot which only accepts numbers between 0 and 1 for each servomotor.

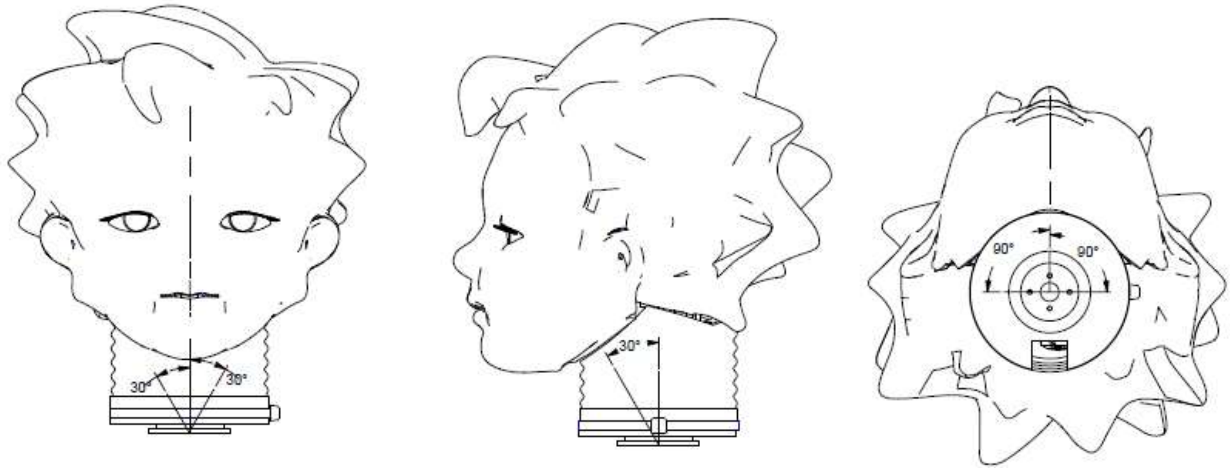

Figure 4.2.3-2 The maximum ranges for the servos placed in the neck. Starting from the left, the roll angle with a maximum range of 60^{0} degrees. Then, in the middle the pitch angle with a maximum range of 30^{0} degrees. Finally, the yaw angle with a maximum range of 180^{0} ("R50 Documentation -," 2014).

Finally, the data package containing the eight AUs values and the three neck angles values (Figure 4.2.3-3) is created and sent to the robot allowing it to actuate.

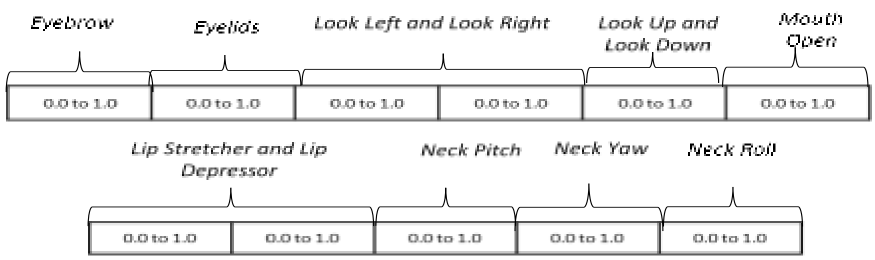

Figure 4.2.3-3 MES data package.

Figure 4.2.3-4, shows a system process diagram that summarizes the various processes that are involved in the MES system. First, the User is in front of the Intel RealSense camera. Then, the system extracts the facial AUs and neck angles, and normalizes according to the robot API. Finally, a data package is constructed and sent to the robot allowing it to actuate.

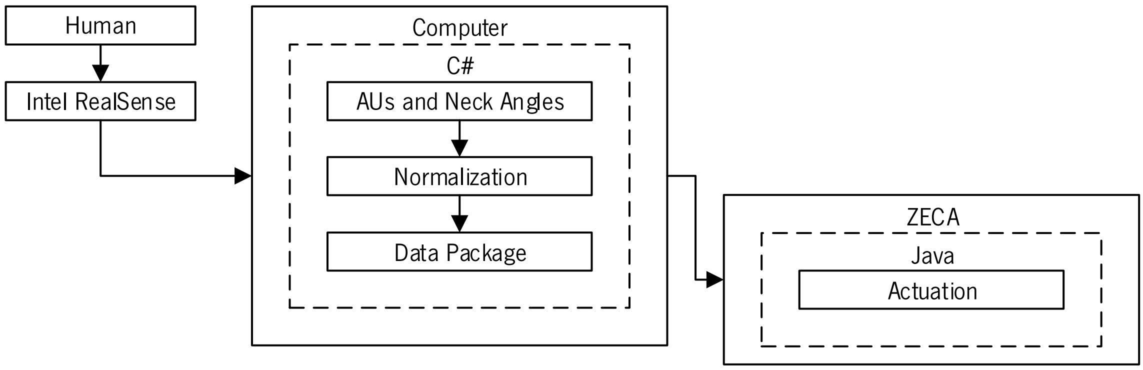

Figure 4.2.3-4 MES subsystem processes diagram.

4.2.4 Designing the ERS subsystem

The ERS subsystem, Emotions Recognition System, is able to recognize human facial expression (anger, fear, happiness, sadness, surprised, and neutral) through facial features in real-time. In this subsystem each extracted facial AU and head motion data is normalized, in a scale 0 to 100, and then filtered. Posteriorly, a multiclass SVM model (with six classes) is trained using the Accord Machine Learning C# library (Souza, 2014), which allows multiclass classification by using the One-Against-One (OAO) approach for SVM. The database employed in the SVM training process was constructed, by applying the methodology 3.5, using the 16 head features (face and neck) mentioned above, acquired from the Intel RealSense 3D sensor, and corresponding to the six emotions considered in the present work.

The participants considered for the database construction were:

- 32 typically developing children from the 1st cycle with 6 to 9 years old. The tests were conducted in the school environment in a closed room and performed once;
- 11 adults with 18 to 30 years old. The tests were conducted in a laboratorial environment in a closed room and repeated 10 times for each emotion.

It is worth mention that 'Sad' was performed in two different ways, with and without head bowed, categorized both as being sad. The acquired data, facial AUs and geometric features were saved into a file.

Finally, after the training process of the SVM model, the model predicts the User's emotional state and sends to ZECA a data package containing the User's current emotion. ZECA uses this information to autonomously detect the User's facial expression, helping the fluidity of the activity. Two activities were developed: IMITATION and EMOTIONS. The IMITATION activity consists in the robot performing a facial expression and the User displaying the same facial expression. On the other hand, in the activity

EMOTIONS the robot asks the User to perform a facial expression. In both activities ZECA verifies the correctness of the prompts. The activities are explained in more detail in section 4.4.2.

Figure 4.2.4-1, shows a system process diagram that summarizes the various processes that are involved in the ERS system. First, the User is in front of the Intel RealSense camera. Then, the system extracts and normalizes the facial AUs and neck angles. Additionally, a SVM multiclass classifier is trained, and the model estimates the User facial expression. Finally, a data package is constructed and sent to ZECA, where activities are performed.

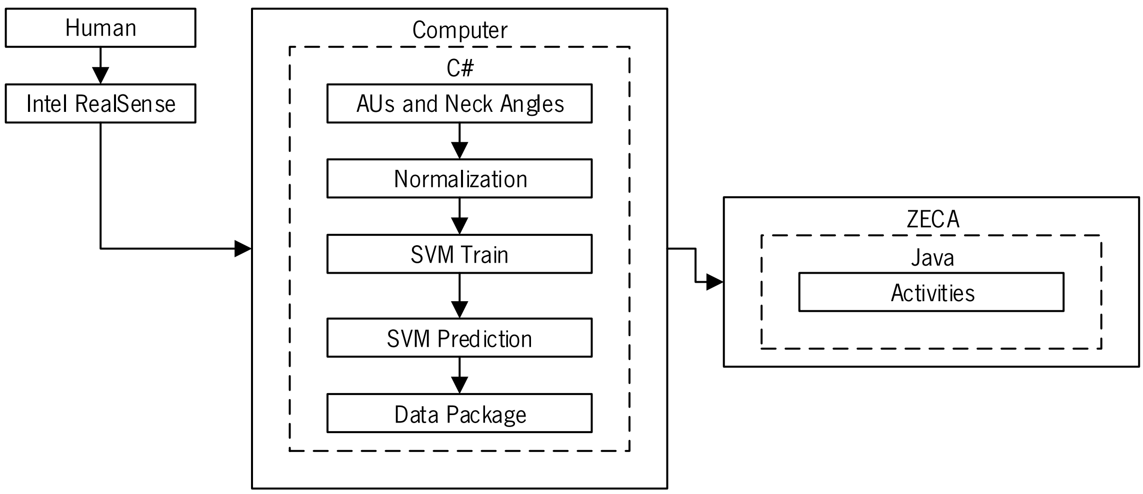

Figure 4.2.4-1 ERS subsystem processes diagram.

4.3 Graphical User Interface (GUI)

In order to control the whole system, monitor the recognition of the emotions and the child evolution a Graphical User Interface (GUI) was created. This interface has two windows: the *'MES and ERS GUI'* and a *'Debug Window'*. The *'MES and ERS GUI'* window, Figure 4.3-1, possesses three tabs: the *'Main'* tab, the *'Machine'* tab, and the *'Child'* tab. In the *'Main'* tab, Figure 4.3-1, it is possible to see the User's face and the emotion estimated (yellow rectangle). This tab also presents the User's heart beat value and chart (red rectangles), acquired by the sensor. Through this window, *'MES and ERS GUI'*, the User has total control of the system, He/she can control and establish a connection between the computer and ZECA, by pressing the buttons: *'Connect'*, *'Close'*, *'Start'*, and *'Pause'* (black rectangle).

The checkboxes in the orange rectangle controls the MES subsystem. When the User selects the *'RT Imitation App'* checkbox, the MES subsystem is initiated. The checkboxes *'Block neck'* and *'Block eye'* (orange rectangle) allows the User to block the ZECA's neck and eyes movements, respectively, actuating as a safety measure.

The checkboxes in the blue and pink rectangles controls the ERS subsystem. When the User selects the *'Imitation App'* checkbox, the Imitation activity/game initiates, i.e., it sends a start command to the robot for initiate the respective activity. Conversely, by selecting the *'Emotions App'*, the Emotion activity starts. In the pink rectangle (General Controls) the User can extend the session time in seconds, by sliding the time bar, if needed. At any time, the User can stop both subsystems, by unselecting the checkbox that corresponds to each subsystem. Finally, by pressing the *'Debug Open'* button, the Debug Window opens. The controls in the dark green rectangle are available in all the tabs of the *'MES and ERS GUI'* window.

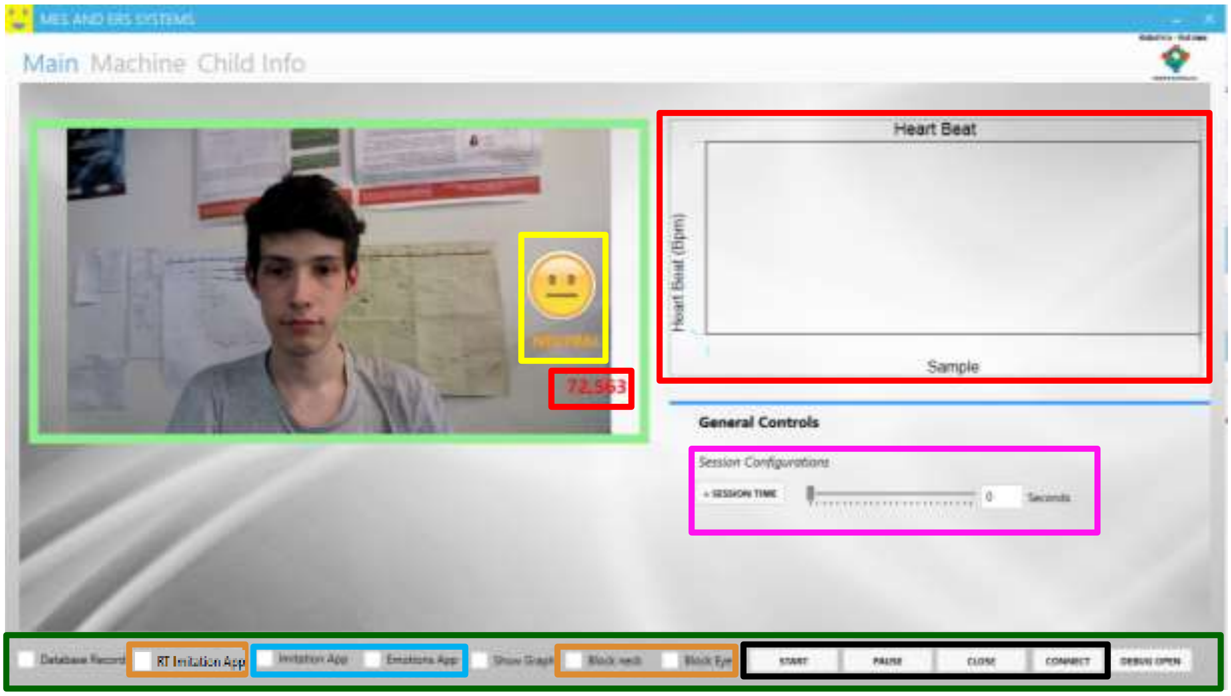

Figure 4.3-1 The 'MES and ERS GUI' window showing the 'Main' tab.

In the *'Debug Window'*, Figure 4.3-2, it is possible to see each facial AU and neck value. This window provides the control of the communication, since (at the bottom) it is possible to insert the ZECA IP address, as well as, the port for establishing the connection.

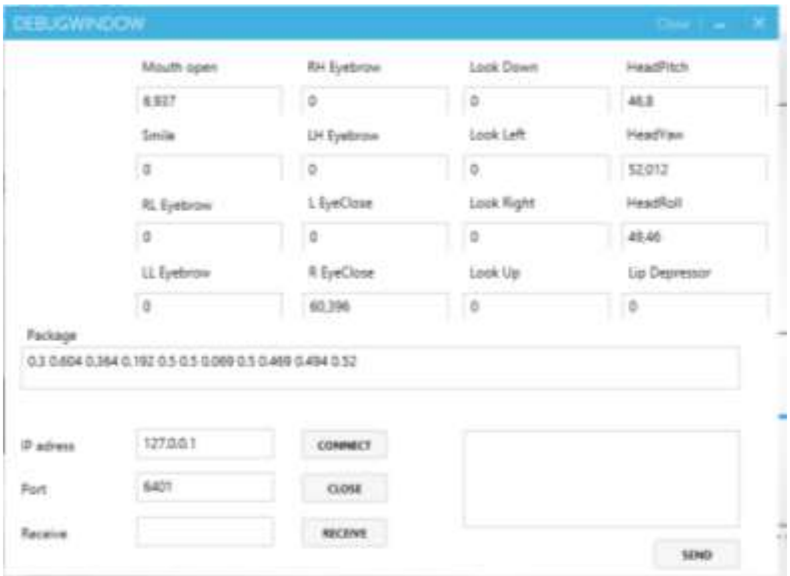

Figure 4.3-2 The 'Debug Window', where it is possible to have more control in the communication protocol and see each AU value.

In the *'Machine'* tab, Figure 4.3-3, the User can configure the multiclass SVM model by choosing or estimating the *C* and gamma (*Y*) parameters. It is also possible to see the database (blue rectangle).

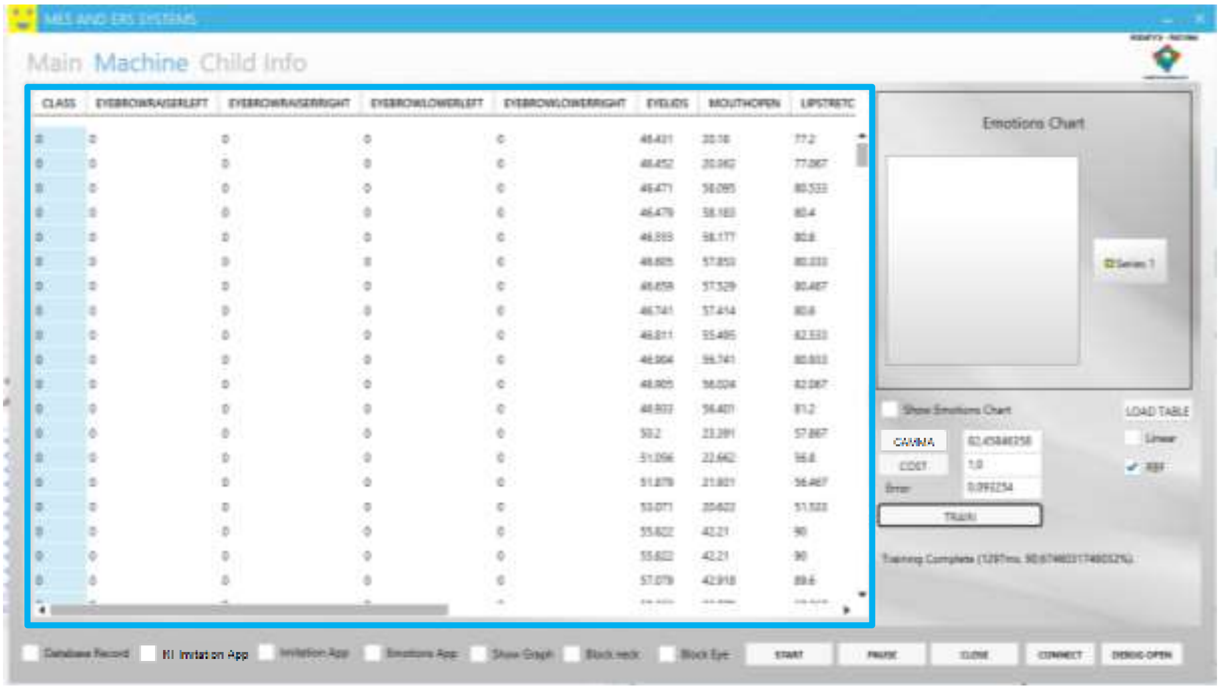

Figure 4.3-3 The 'MES and ERS GUI' window showing the 'Machine' tab, where it is possible to configure and train the multiclass SVM model.

The *'Child'* tab presents the child information, Figure 4.3-4. The User have access to the children database, which is organized by *'code'*, *'name'*, *'age'*, *'number of sessions'* of the child (blue square). By double clicking in a child in the children's table (blue square), the selected child information is displayed (yellow square). It is possible to add a new child to the database (red rectangle). At the bottom, it is possible to see the performance of a child in a session, showing the session time duration and the number of correct and wrong answers (green rectangle), as well as, the child evolution per session (black rectangle).

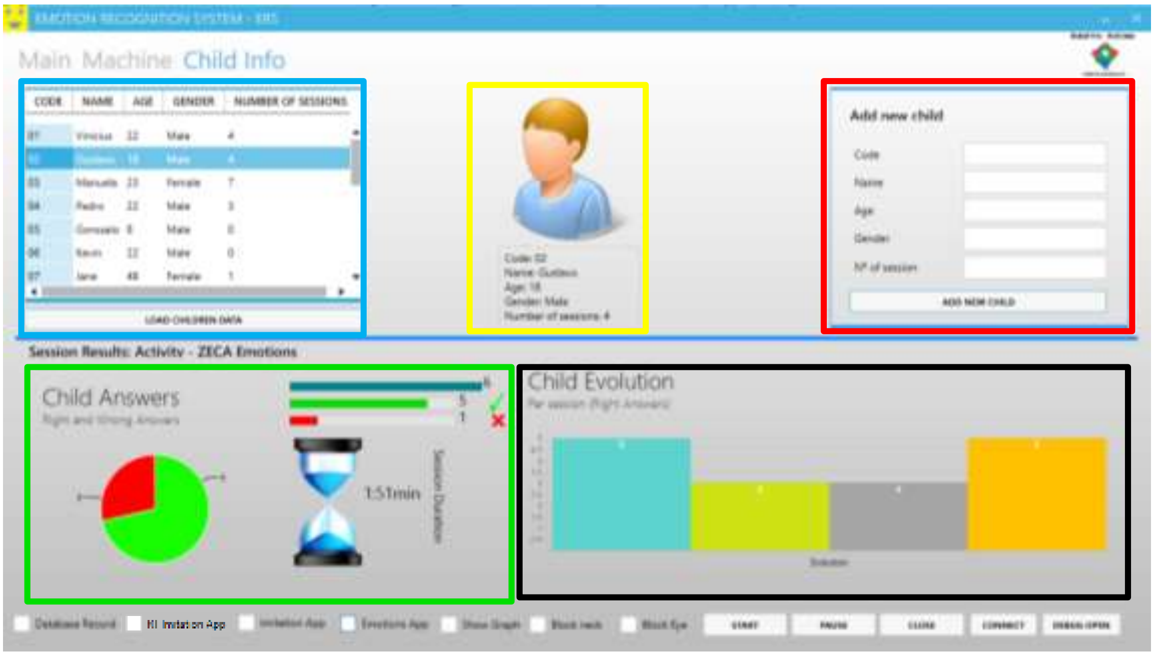

Figure 4.3-4 The 'MES and ERS GUI' window showing the 'Child' tab, where it is possible to see the child's information, performance, and evolution.

4.4 Software Architecture

This section presents the main procedures of the software produced included in the MES and ERS subsystems.

4.4.1 MES Architecture

Figure 4.4.1-1 presents the MES subsystem flow chart with the main procedure of the software in the GUI. After detecting, extracting, normalizing, and filtering each AU and neck value, a data package is created containing this information. Then, this data package is sent to ZECA, allowing it to actuate. The same process repeats until the User unchecks the *'RT Imitation App'* checkbox in the GUI, sending the "END" command to ZECA.

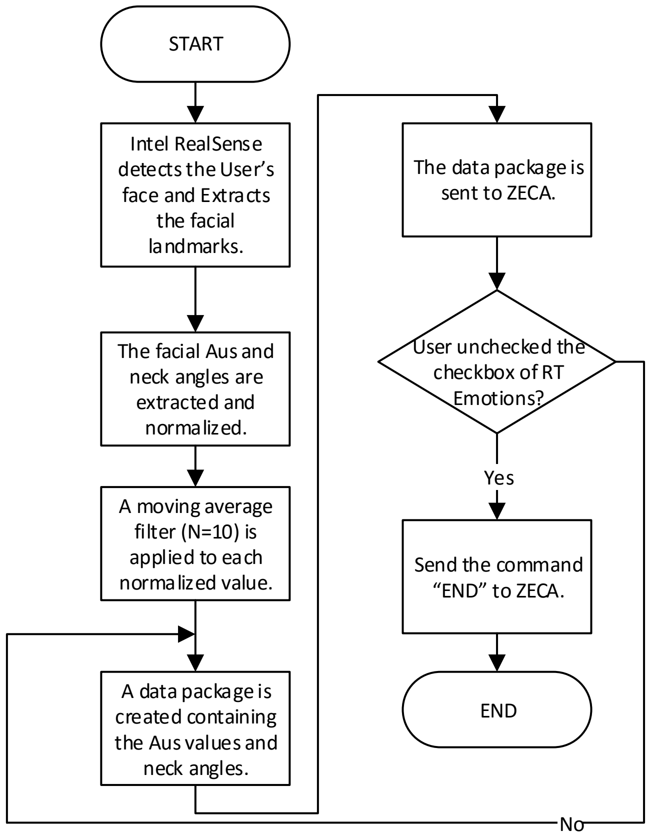

Figure 4.4.1-1 GUI – MES subsystem flow chart.

Figure 4.4.1-2 shows the flow chart with the main procedure of the software in ZECA. First, if a message is available, ZECA verifies if this message is an "END" command. Then, if it is not an "END" command it verifies if the command received is valid, and then update the servo motors values. When ZECA receives the "END" command, from the GUI, the program ends.

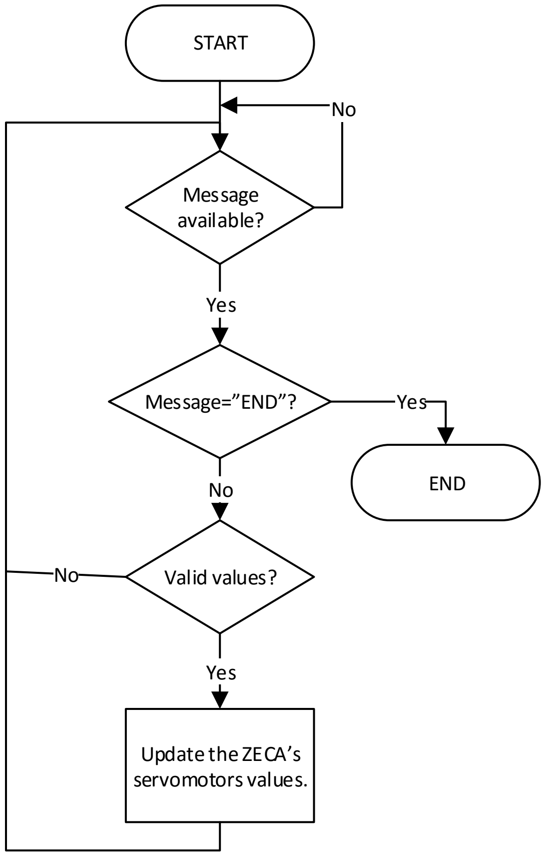

Figure 4.4.1-2 ZECA – MES subsystem flow chart.

4.4.2 ERS Architecture

Figure 4.4.2-1 presents the ERS subsystem flow chart with the main procedure of the software in the GUI. After detecting, extracting, normalizing, and filtering each AU and neck value, a multiclass SVM model is trained. Then, the User selects on the GUI one of the ERS activities available, by selecting one of the follow checkboxes: *'Imitation App'* or *'Emotions App'*. Then, the GUI sends to ZECA the activity code. The SVM model starts estimating the User's facial expressions and the GUI sends the prediction to ZECA. This process is repeated until the User unchecks the activity checkbox that is running. The GUI sends to ZECA the "END" command, finishing the activity.

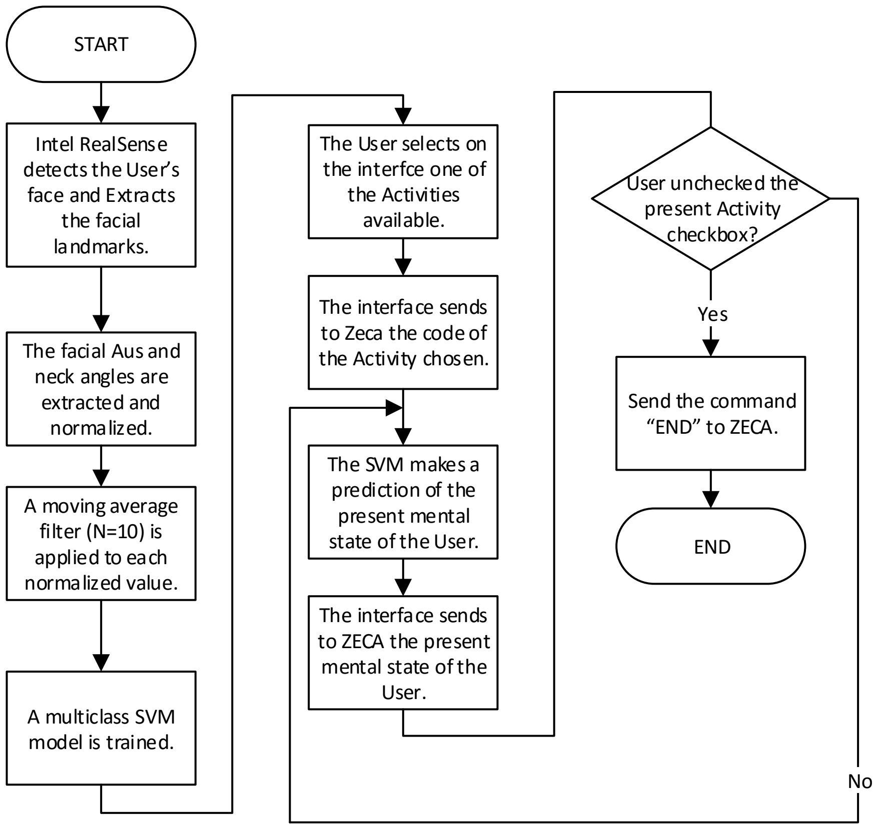

Figure 4.4.2-1 GUI – ERS subsystem flow chart.

Figure 4.4.2-2 show the ERS subsystem flow chart with the main software procedure to include in ZECA. After ZECA receiving the child's information, it says the name of the child and prompts the experimenter to select the activity that is going to be performed. After the experimenter selects the activity on the GUI, ZECA greets the child and gives the instruction to the chosen game. The number of sessions of the child is updated and the folder to save the files of the session is created, recording the date, time, performance, and answers in that session.

The ACTIVITY subroutine, Figure 4.4.2-3, describes the common procedures to the two game scenarios, the IMITATE and EMOTIONS. After getting the information from the child and the session, the classes related to the answers, performance, and the timer are instantiated. The activity timer is then started. In the IMITATE game ZECA prompts an emotion and asks for the child to imitate its facial expression. Conversely, in the EMOTIONS game ZECA asks for the Child to display a facial expression. After, in both game scenarios, ZECA verifies if the answer is correct and prompts a positive or negative reinforcement accordingly to the correctness of the answer. When the time is up, ZECA asks if the experimenter wants to continue. The experimenter can extend the time of the session by sliding the time bar in the GUI. But if the experimenter decides to stop the activity, by unchecking on the GUI the checkbox of the activity that is running, the activity finish, with a robot's farewell, and ZECA sends to the GUI the performance data of the child in that session.

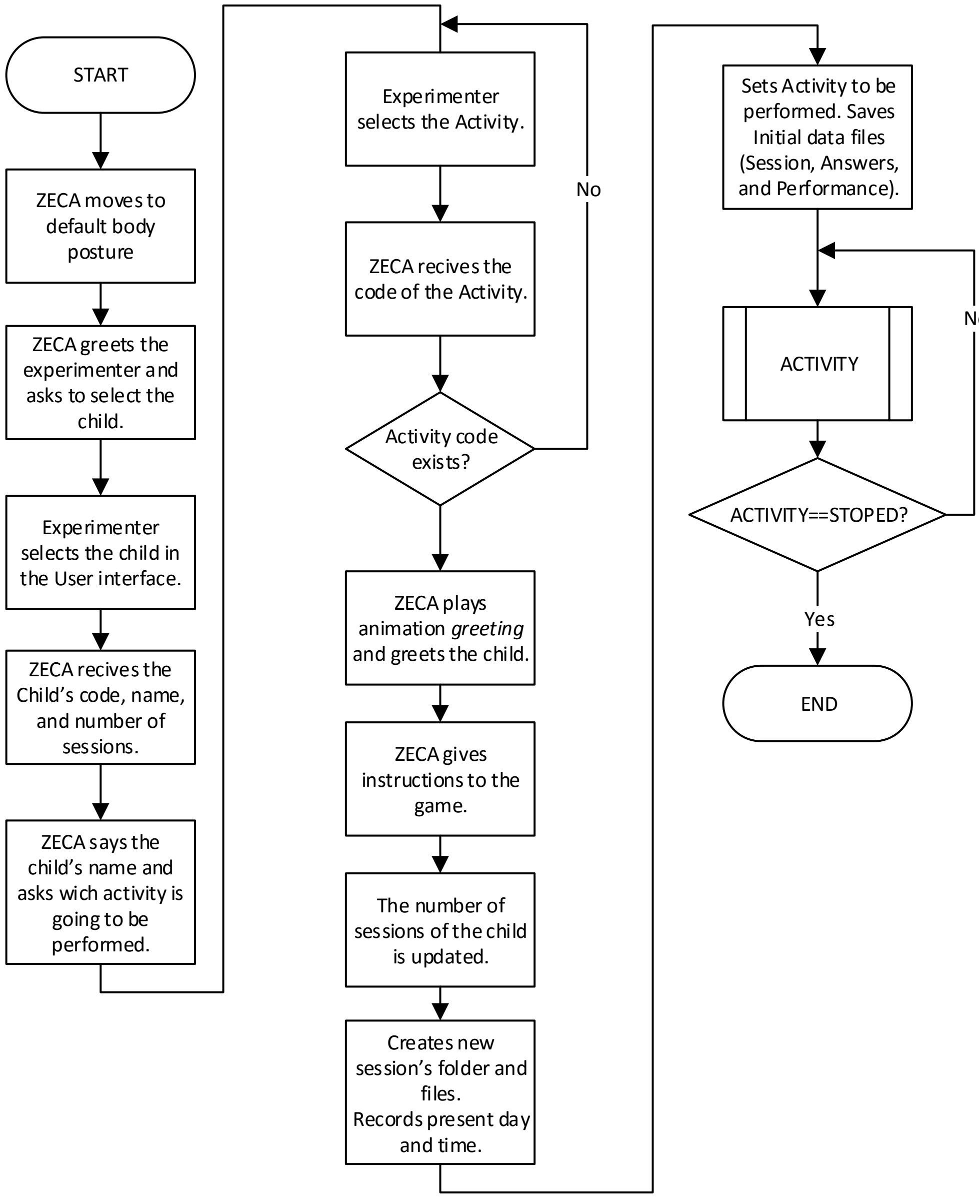

Figure 4.4.2-2 Flow chart with the main procedure of the software on ZECA, calling the subroutine ACTIVITY.

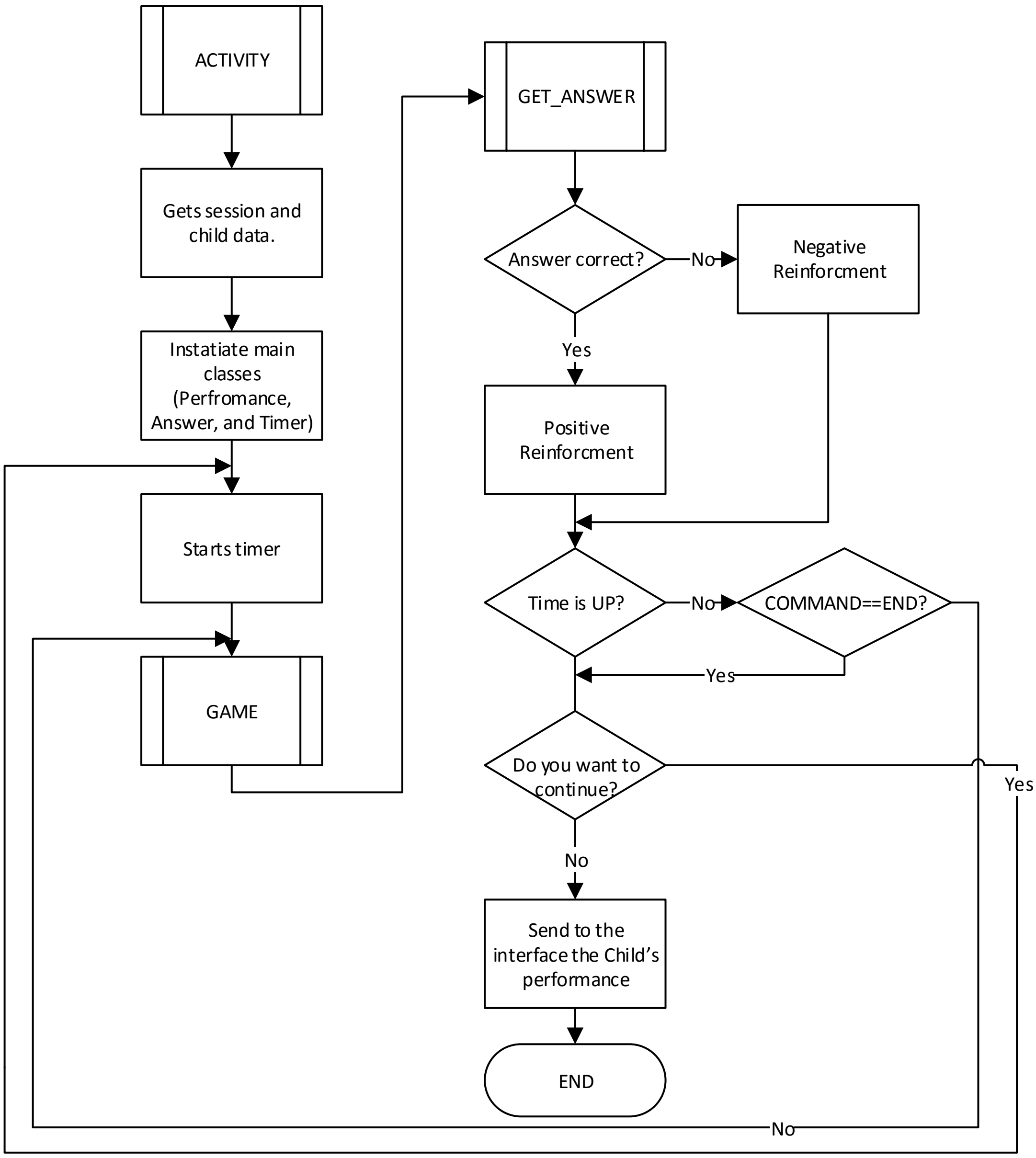

Figure 4.4.2-3 Subroutine showing the general process for all game scenarios.

References

Camera range. (2015). Retrieved February 4, 2016, from https://software.intel.com/sites/default/files/managed/99/95/camera2.png

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders*. University of Minho.

Face Pose Data [F200,SR300]. (2015). Retrieved January 16, 2016, from https://software.intel.com/sites/landingpage/realsense/camera-sdk/v1.1/documentation/html/index.html?doc_face_face_pose_data.html

Intel® RealSense™ Technology. (2015). Retrieved January 15, 2016, from http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html

Meah, L. (2014). R50 “Zeno” User Guide. Sheffield. Retrieved from http://www.robokindrobots.com/wp-content/uploads/2014/11/UnivOfSheffield.pdf

R50 Documentation -. (2014). Retrieved July 12, 2015, from http://www.robokindrobots.com/support-documentation/r50/

realsense_intel. (2015). Retrieved from http://ade2.nhtv.nl/~dpj/sensorsatplay/images/realsense_intel.png

Robokind | Advanced Social Robots. (2015). Retrieved May 1, 2016, from http://www.robokindrobots.com/

5 RESULTS

Summary

This chapter starts by presenting the evaluation of the performance of the two subsystems developed, MES and ERS, for mimicking and recognizing emotions. Then, it shows the preliminary results of a study involving children with ASD. This preliminary study was conducted in order to assess the performance of the ERS subsystem in a real world environment.

5.1 MES Subsystem Evaluation

The performance of the MES subsystem was evaluated using the methodology – MES validation, presented in the chapter 3 section 3.4.1, for validating systems that are capable of synthesizing facial expressions.

The overall subsystem was first tested in a laboratory environment, to verify its correct functioning using the software *FaceReader*. Then, a perceptual study was conducted in order to test the similarity between expressions of a performer and the robot. Finally, the system was evaluated in a school environment with children in the 1st cycle, where the robot mimicked the child emotional facial expressions (happiness, sadness, and neutral).

In https://www.youtube.com/watch?v=vXYJ7szyY8s it is possible to see a video of the robot imitation performance.

5.1.1 System Evaluation using *FaceReader* – a computational based evaluation

In a first stage, the software *FaceReader* was used to automatically analyse the synthesized facial expressions. *FaceReader* is a professional software from *Noldus* ("Noldus | Innovative solutions for behavioral research," 2015) for automatic analysis of the six basic emotions (Happiness, sadness, anger, fear, disgust, and surprise), as well as classifying neutral and contempt states. It also provides gaze direction, head orientation, person characteristics (e.g. gender and age), and a detailed analysis of the facial AUs. This software can analyse from live video, recorded video, or images. Therefore, a video of ZECA mirroring the emotions of a performer was used. This video was subjected to the *FaceReader* software analysis.

In Figures 8 to 10 are presented the examples of the facial expressions associated with happiness, surprise, and fear, respectively, registered for the robot and the performers. The *FaceReader* results are presented in each legend in terms of performance matching. A video presenting some of the performed tests may be seen in the following link: https://youtu.be/qt_4nJpXPz8.

Figure 5.1.1-1 Happiness facial expression: match of 54% for the robot and 85% for the performer.

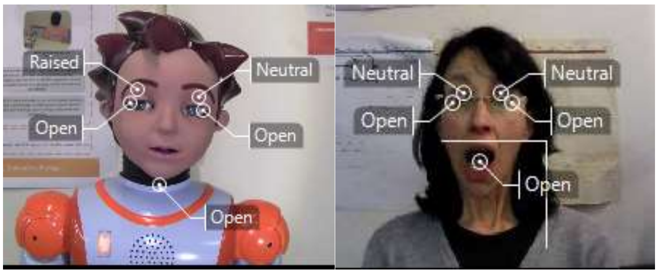

Figure 5.1.1-2 Surprise facial expression: match of 77% for the robot and 87% for the performer.

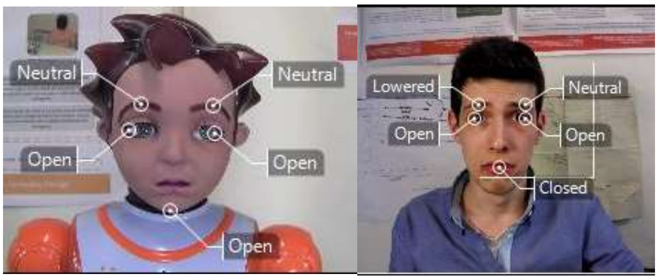

Figure 5.1.1-3 Fear facial expression: match of 69% for the robot and 62% for the performer.

Despite of the high recognition rates for the above facial expression, anger was really hard to be recognized using this software, even though most of the AUs necessary to represent that expression were present. Most probably, they were not marked enough for the software to recognize them. Moreover, the software *FaceReader* is prepared to recognize facial expressions in human faces. In this case, the software was run on a non-human face, a robot face, which may be harder for the software to recognize.

5.1.2 Perceptual study (Quiz) – a subjective evaluation

In order to test the similarity between expressions of a performer and the robot, a perceptual study was conducted using side-by-side comparison or "copy synthesis" (chapter 3, section 3.4.1). Thirty-one children between 6-and 9 years old and 28 adults between the ages 18-52 participated in the study (total sample: 59 participants). The synthesized expressions on a physical face (ZECA's face) were shown side-by-side with the performer's face to the participants. In this perceptual study, the participants had to select the robot facial expression more similar to the performer's facial expression. Figures 5.1.2-1 and 5.1.2-2 present the results obtained with children and adults, respectively.

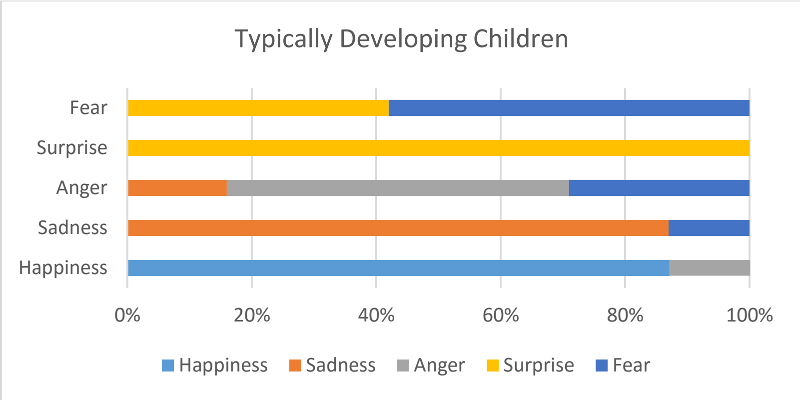

Figure 5.1.2-1 Matching Results – Children.

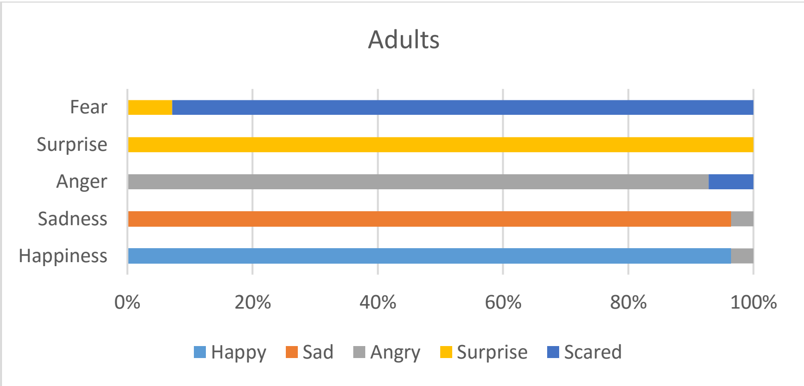

Figure 5.1.2-2 Matching Results – Adults.

The facial expression with the highest similarity is surprise, with a score of 100% in adults and children. Happiness and sadness have similar matching scores in children and adults: happiness (children – 87%, adults – 96%) and sadness (children – 87%, adults – 96%). The facial expressions with the lowest matching score for children are anger and fear, with 55% and 58%, respectively. The matching scores of these facial expressions in adults were higher, with 93% for both expressions.

5.1.3 Perceptual Experiment

A perceptual experiment was conducted to evaluate the proposed system. This is a common method for evaluating synthesized facial expressions (Becker-Asano & Ishiguro, 2011; Mazzei, Lazzeri, Hanson, & De Rossi, 2012).

In this perceptual experiment, three emotions were considered: happiness, sadness (since they are opposite emotions, they are more easily identifiable) and the neutral state. It was also considered the head motion as it gives more emphasis to the emotional states, making it easier to distinguish the facial expression.

The tests for the perceptual experiment considered 32 typically developing children from the 1st cycle with 6 to 9 years old. The tests were performed in school environment in a closed room.

The experimental setup may be seen in Figure 5.1.3-1, with the camera placed in the chest of the robot. The child sat in front of the robot, looking at ZECA and performed the emotions happiness, sadness and neutral, requested by the researcher. The researcher showed a photograph corresponding to each emotion in order to serve as a clue. Sadness was performed in two different ways, with and without the head bowed. A supervisor person reported the results for each of the expressed emotions, Figure 5.1.3-2.

Figure 5.1.3-1 Experimental setup: ZECA and Intel RealSense.

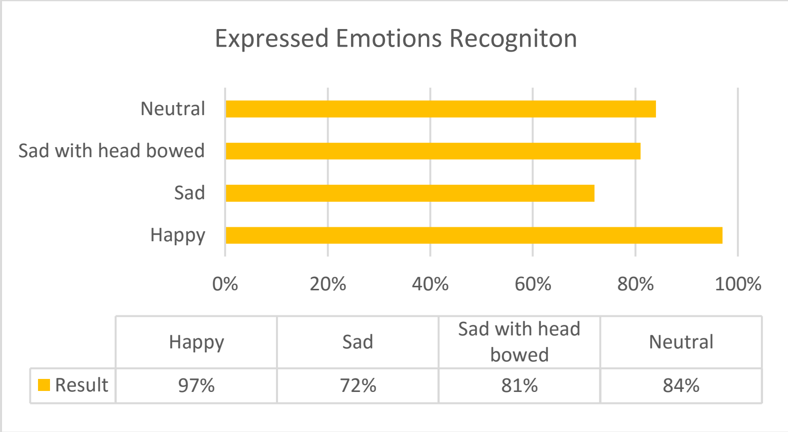

Figure 5.1.3-2 Recognition of the expresses emotions.

The Happiness emotion had the highest matching score (97%) and the neutral state obtained a mark of 84%. Regarding sadness, it was better determined considering head bowing (81%) compared to the expression without head movement (72%).

The relatively high overall scores of the expressed emotions suggest that the present system can accurately online map facial expressions of a user onto a robot.

5.2 ERS Subsystem Evaluation

The performance of the ERS subsystem was evaluated using the methodology presented in the chapter 3 section 3.4.2, for validating experimental systems for recognizing emotions.

Two evaluations were conducted in order to quantify the ERS subsystem performance: a static off-line and real-time evaluations. The follow section presents the results obtained with the proposed subsystem in the recognition of the six emotions considered in this work: 'Happiness', 'Sadness', 'Anger', 'Surprise', 'Fear', and 'Neutral'.

5.2.1 SVM Model – Simulation Results

The SVM model was simulated in Matlab using the *fitcecoc* function, which is the multiclass SVM function. In order to choose the best kernel function, the model was first implemented using the Linear kernel and later the RBF kernel.

5.2.1.1 Using a Linear Kernel

The SVM model was trained and validated using the linear kernel. The results of the full analysis – confusion matrix, accuracy, sensitivity, specificity, AUC, and MCC metrics are presented below.

Table 5.2.1.1-1 shows the recognition accuracy confusion matrix for the six emotions obtained by SVMs with Linear kernel, achieving an overall accuracy of 88.2%. More specifically, the best performances are achieved for the recognition of 'Happiness' and 'Surprise', with recognitions accuracies of 99% and 94% respectively. For 'Fear, a significantly lower accuracy of 66% is observed. For the remaining classes ('Sadness', 'Anger', and 'Neutral') are observed slightly similar accuracies.

Table 5.2.1.1-1 Confusion Matrix - Linear SVM

	Happiness	**Sadness**	**Anger**	**Surprise**	**Fear**	**Neutral**
Happiness	**99%**	0%	0%	1%	1%	0%
Sadness	0%	**89%**	4%	0%	1%	6%
Anger	1%	3%	**89%**	0%	7%	0%
Surprise	0%	1%	0%	**94%**	0%	5%
Fear	3%	14%	0%	5%	**66%**	12%
Neutral	0%	5%	0%	1%	3%	**91%**

Table 5.2.1.1-2 compares the sensitivity, specificity, and AUC metrics for each class. 'Fear is the class with an overall lower performance when compared to the other classes, with the lowest sensitivity, meaning that 65% of the actual positives (TP) were identified as such, and in consequence has the lowest AUC of 93%. This can indicate that the data has a nonlinear behaviour, being more difficult for a linear kernel to separate classes that have a similar distribution of the feature values.

Table 5.2.1.1-3 shows the overall performance of the SVM with Linear kernel, with an accuracy of 88.2%, an average sensitivity and specificity of 87.2% and 97.3% respectively, an average AUC of 97.2%, and MCC of 84.3%.

Table 5.2.1.1-2 Per-Class performance - Linear SVM

	Sensitivity	**Specificity**	**AUC**
Happiness	99%	99%	99.2%
Sadness	88%	96%	96.5%
Anger	89%	98%	98%
Surprise	94%	99%	99.5%
Fear	65%	97%	93%
Neutral	88%	95%	97.1%

Table 5.2.1.1-3 Overall Linear SVM performance.

Metric	**Value**
Accuracy	88.6%
Average Sensitivity	87.2%
Average Specificity	97.3%
Average AUC	97.2%
MCC	84.3%

5.2.1.2 Using RBF Kernel

The SVM model was trained and validated using the RBF kernel. The results of the full analysis – confusion matrix, accuracy, sensitivity, specificity, AUC, and MCC metrics – are presented below.

Table 5.2.1.2-1 shows the recognition accuracy confusion matrix for the six emotions obtained by SVMs with RBF kernel, achieving an overall accuracy of 93.6%. More specifically, the best performances are achieved for the recognition of 'Happiness' and 'Surprise', with recognitions accuracies of 98% and 98% respectively. For the 'Fear' and 'Neutral' classes, slightly lower accuracies of 89% and 90%, respectively, are observed. For the remaining classes ('Sadness' and 'Anger') are observed slightly similar accuracies (94% and 95%, respectively).

Table 5.2.1.2-1 Confusion matrix - RBF SVM

	Happiness	Sadness	Anger	Surprise	Fear	Neutral
Happiness	**98%**	1%	0%	0%	1%	0%
Sadness	0%	**94%**	1%	0%	<1%	4%
Anger	0%	5%	**95%**	0%	0%	0%
Surprise	0%	1%	0%	**98%**	1%	0%
Fear	2%	6%	0%	3%	**89%**	1%
Neutral	0%	6%	0%	0%	4%	**90%**

Table 5.2.1.2-2 compares the sensitivity, specificity, and AUC metrics for each class. 'Fear' and 'Neutral' are the classes with an overall lower performance when comparing with the other classes, with similar lower sensitivities (88% and 89% respectively) and AUCs (97.5% and 98.4%).

Table 5.2.1.2-2 Per-class performance - RBF SVM

	Sensitivity	Specificity	AUC
Happiness	97%	100%	99.7%
Sadness	94%	96%	99.3%
Anger	95%	100%	99.4%
Surprise	98%	99%	99.8%
Fear	88%	98%	97.5%
Neutral	89%	98%	98.4%

Table 5.2.1.2-3 shows the overall performance of the SVM with RBF kernel, with an accuracy of 93.6%, an average sensitivity and specificity of 93.5% and 98.5% respectively, an average AUC of 99%, and MCC of 92.2%.

Table 5.2.1.2-3 Overall RBF SVM performance.

Metric	Value
Accuracy	93.6%
Average Sensitivity	93.5%
Average Specificity	98.5%
Average AUC	99%
MCC	92.2%

5.2.1.3 Choosing the best Kernel

By comparing the results obtained in the previous section, it is possible to conclude that the SVM model with the RBF kernel presents an overall better performance than the SVM model with the linear kernel. The accuracies per class increased in the RBF SVM model, especially the accuracy of the class 'Fear' (66% to 89%). In consequence, the overall accuracy of the SVM model with the RBF kernel also increased from 88.15% to 93.63%, which can indicate that the relation between class labels and attributes is nonlinear. Unlike the linear kernel, the RBF kernel can handle the case when the relation between class labels and attributes is nonlinear. Moreover, RBF has fewer number of hyper-parameters than other nonlinear kernels (e.g. polynomial kernel), which may decrease the complexity of model selection (Hsu, Chang, & Lin, 2010). Also, RBF usually has lower computational complexity, which in turn improves real-time computational performance (Zhang, Zhang, & Hossain, 2015).

It is worth notice that the RBF kernel is widely used as a kernel function in emotion classification (Michel & El Kaliouby, 2003).

5.2.2 SVM Model – Experimental results

As mentioned in chapter 4, the SVM model was trained using the Accord Machine Learning C# library (Souza, 2014), which allows multiclass classification by using the OAO approach for SVM. The kernel used for the SVM model was the RBF kernel, since it performed significantly better than the Linear kernel, as showed in the previous section. The gamma (Υ) and C parameters were estimated by using built in functions from the Accord Machine Learning library.

The proposed system was implemented and evaluated in a laboratorial environment with 14 adults (18-49). The participant sat in front of the camera, looked at the Intel RealSense and performed the emotion requested by the researcher.

Table 5.2.2-1 shows the recognition accuracy confusion matrix for the six emotions, with an overall accuracy of 88.3%. In general, the on-line system yields comparable results to that obtained in off-line evaluation. 'Happiness' and 'Sadness' emotional states have accuracies over 90% and the other four facial expressions are consistently beyond 85%.

Table 5.2.2-1 Confusion matrix - Real-time system

	Happiness	**Sadness**	**Anger**	**Surprise**	**Fear**	**Neutral**
Happiness	**93%**	0%	0%	0%	7%	0%
Sadness	0%	**93%**	0%	0%	0%	7%
Anger	7%	7%	**86%**	0%	0%	0%
Surprise	0%	7%	0%	**86%**	7%	0%
Fear	0%	14%	0%	0%	**86%**	0%
Neutral	0%	0%	7%	0%	7%	**86%**

Finally, regarding real-time performance, the emotion recognition system was tested in a frame rate of 30 fps on i5 quad-core CPUs with 16 GB RAM. The required time for the system to perform efficiently facial expression recognition is 1-3*ms*. This means that the working frequencies achievable for sampling and processing are very high and do not compromise the real-time feature of the interaction process. The training computational cost of the system is approximately 1-2*s* for the multi-class SVM classifier.

Although there is not yet available in the literature emotions recognition systems based on Intel RealSense, the performance of the proposed system was compared to the results presented in (Zhang *et al.*, 2015). This system was based on a Kinect sensor and used the *Bosphorus* database and SVM for facial expressions classification. The overall accuracy of the proposed system is 88% compared to 84% in (Zhang *et al.*, 2015). Regarding the required time for the proposed system to perform facial emotion recognition is 1-3*ms* compared to 3-5*ms* in (Zhang *et al.*, 2015).

5.3 Experimental study with children with ASD

The ERS subsystem was tested in a school environment in order to evaluate the two game scenarios that are part of the ERS subsystem: the IMITATE, where the child has to mimic the ZECA's facial expression and EMOTIONS, where the child has to perform the facial expression asked by ZECA, both described in chapter 4 section 4.2.4. This experimental phase had a twofold goal: to verify if the system can implement a procedure that makes the children able to interact in a comfortable and natural way and, on the other side, to evaluate the appropriateness of the system's (in this case the ERS system) in a real environment with children with ASD. The main question was: can a humanoid robot be used as an eligible mediator in emotions recognition activities with children with ASD? Part of the result analysis was qualitative. Further tests are mandatory in order to conclude the suitability of the proposed system to be used as a complement to the traditional interventions. The validation of the system for clinical interventions must be performed with a larger sample, with a quantified recognition success rate and a quantified children evolution in terms of predefined behaviour indicators.

Following this trend, a set of preliminary experiments were carried out involving six children with ASD (high-functioning autism or Asperger's syndrome) aged between eight and nine, four boys and two girls. As recommended by the therapists, the original group of six was uniformly divided into two subsets of three children (one and two). In the subset one, three different facial expressions were investigated (anger, happiness, and sadness). On the other hand, in the subset two, five facial expressions were investigated (anger, fear, happiness, sadness, and surprised). The facial expressions were asked or performed randomly by the robot. The experiments were performed individually in activities involving the therapist, the robot, and the child. The robot had the role of mediator in the process of imitation and recognition of facial expressions. The therapist only intervened if necessary in order to "regulate" the child's behaviour. The researcher supervises the progress of the activity, and monitors the system. Each child was placed in front of the robot. Figure 5.3-1 shows the experimental configuration used. Seven sessions of two-three minutes each were performed.

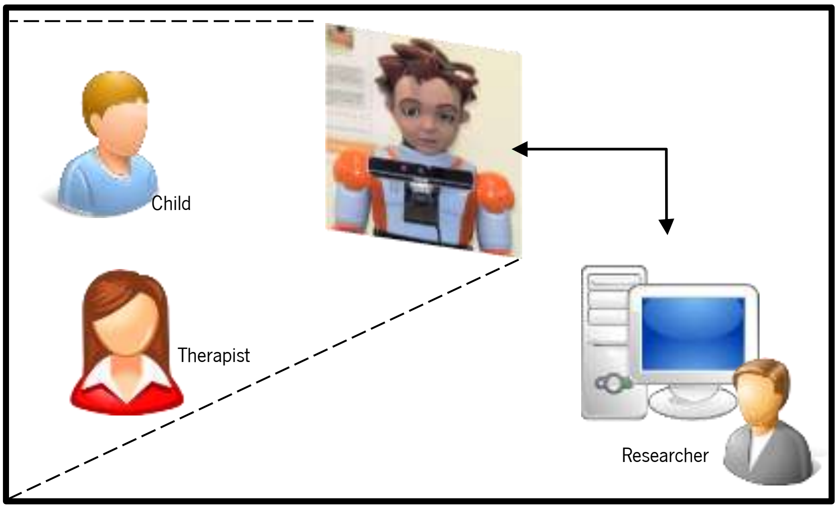

Figure 5.3-1 Experimental configuration: Child-ZECA.

The quantitative behaviours analysed were the following: answers (wrong and correct), and the child's response time per session. The response time was counted from the time the robot gave the prompt to the time the child performed the correspondent facial expression.

5.3.1 Results from the subset one

Figures 5.3.1-1 and 5.3.1-2 shows the results of the two game scenarios obtained with two of the three children (A and B) from the subset one. The results of the child C were inconclusive as most sessions were unsuccessfully ended. The child did not perform as expected, since he/she was more attracted by the robot components or he/she was tired/annoyed, and consequently he/she was not focused on the activity.

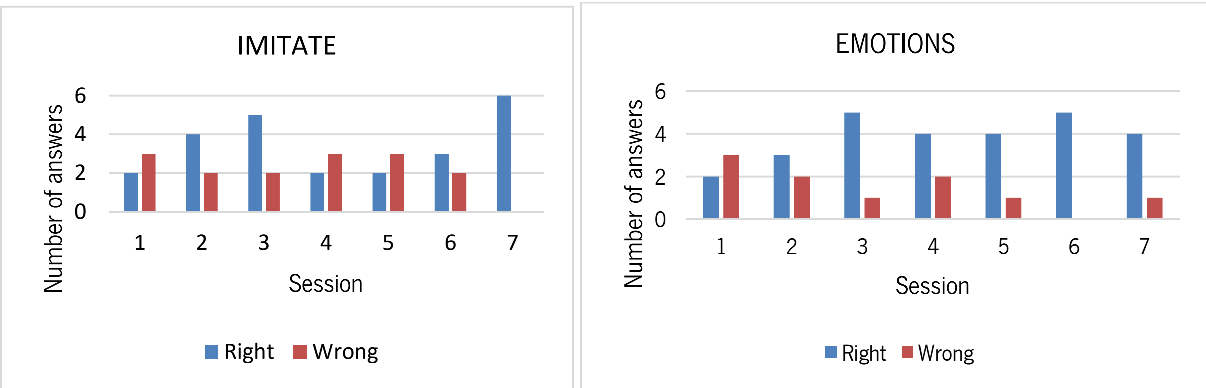

Figure 5.3.1-1 Child A results - On the left the progress over seven sessions with the IMITATE activity. On the right the progress over seven sessions with the EMOTIONS activity.

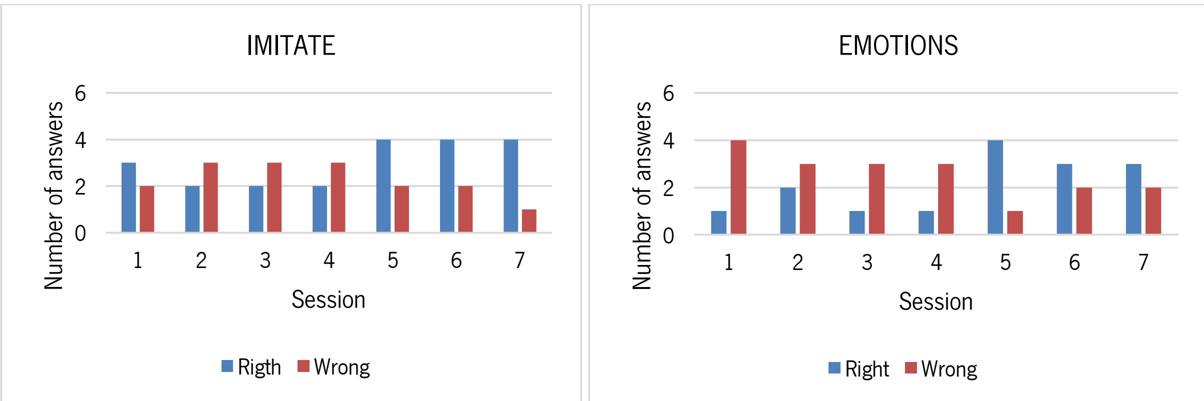

Figure 5.3.1-2 Child B results - On the left the progress over seven sessions with the IMITATE activity. On the right the progress over seven sessions with the EMOTIONS activity.

The results in Figure 5.3.1-1, on the left, shows that in the first session the child A gave more incorrect answers than correct ones, whereas child B gave slightly more correct answers in the same session (Figure 5.3.1-2 on the left). Then, in the following sessions the performance of the child A slightly improved, by having more correct answers than incorrect. However, the performance of the child B slightly worsened, improving only in the last three sessions. Conversely, in the session 4 and 5 the progress of the child A slightly worsened, by giving more incorrect answers. In the last session, both children had a good performance. It is possible to conclude that the overall performance of the Child A in the IMITATE activity fluctuated with a good performance in the last session, whereas by analysing the results of the same activity for the child B there was a positive evolution. In the EMOTIONS activity, Figures 5.3.1-1 and

5.3.1-2 on the right, child A had a distinctly better performance than child B in the first sessions. Considering the last three sessions, the performance of the child B improved, equalling up to the performance of the child A. It is possible to conclude that both children had a positive evolution in the EMOTIONS activity.

Tables 5.3.1-1 and 5.3.1-2, presents the children's mean response time, and standard deviation (SD) of successful answers in the activities IMITATE and EMOTIONS, given in the corresponding session, respectively. Both participants took more time answering to the prompt in the last session, Session 7. Participant 1 was usually faster to answer the prompt from the robot than participant 2. Additionally, it is possible to conclude that in the last three sessions on the EMOTIONS activity, where the participants took more time at performing the asked facial expression, the performance improved in both cases, having more correct answers than incorrect ones (Figure 5.3.1-1 and Figure 5.3.1-2).

Table 5.3.1-1 Children's in the subset one mean response time in seconds for successful answers (SD) in the IMITATE activity. In general, the response time increased in the last session.

Session number	**Child A**	**Child B**
1	16.54 (0.02)	19.64 (1.5)
2	18.16 (1.95)	18.09 (4.23)
3	17.88 (1.27)	17.45 (0.44)
4	19.66 (0.99)	16.54 (0.44)
5	18.53 (1.98)	17.06 (0.18)
6	18.65 (1.21)	18.79 (0.54)
7	19.63 (2.08)	18.99 (3.08)

Table 5.3.1-2 Children's in the subset one mean response time in seconds for successful answers (SD) in the EMOTIONS activity. In general, the response time increased in the last sessions.

Session number	**Child A**	**Child B**
1	16.02 (0)	16.90 (0)
2	16.57 (0.64)	18.21 (2.17)
3	17.47 (1.62)	18.26 (0)
4	16.82 (1.0)	24.82 (0)
5	20.49 (4.27)	20.68 (2.67)
6	21.63 (1.68)	21.07 (2.06)
7	20.77 (1.99)	20.00 (1.33)

5.3.2 Results from the subset two

Figures 5.3.2-1, 5.3.2-2, and 5.3.2-3 shows the results of the two game scenarios obtained with the three children (D, E, and F) from the subset two.

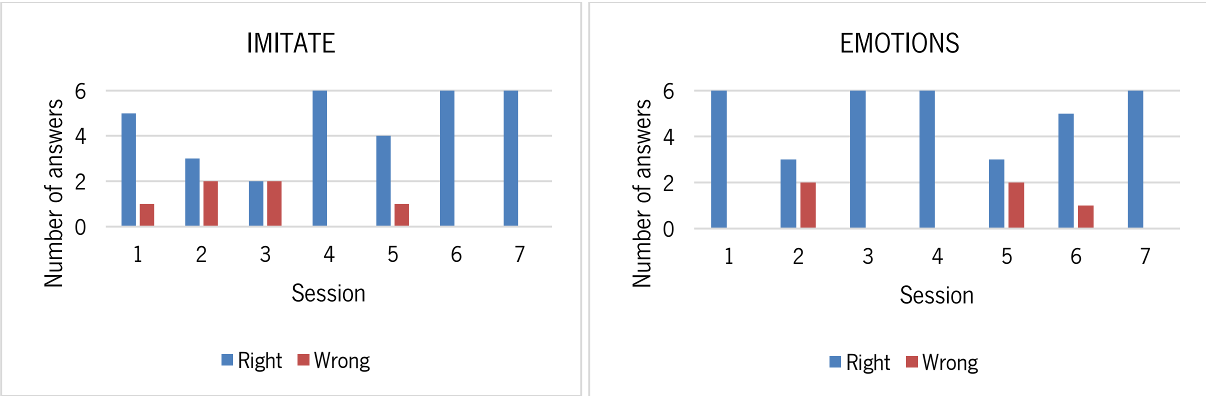

Figure 5.3.2-1 Child D results - On the left the progress over seven sessions with the IMITATE activity. On the right the progress over seven sessions with the EMOTIONS activity.

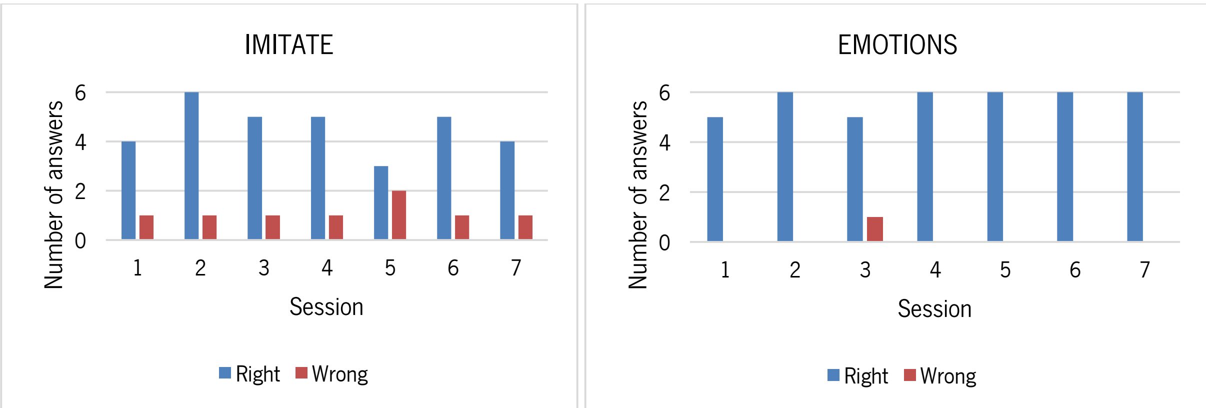

Figure 5.3.2-2 Child E results - On the left the progress over seven sessions with the IMITATE activity. On the right the progress over seven sessions with the EMOTIONS activity.

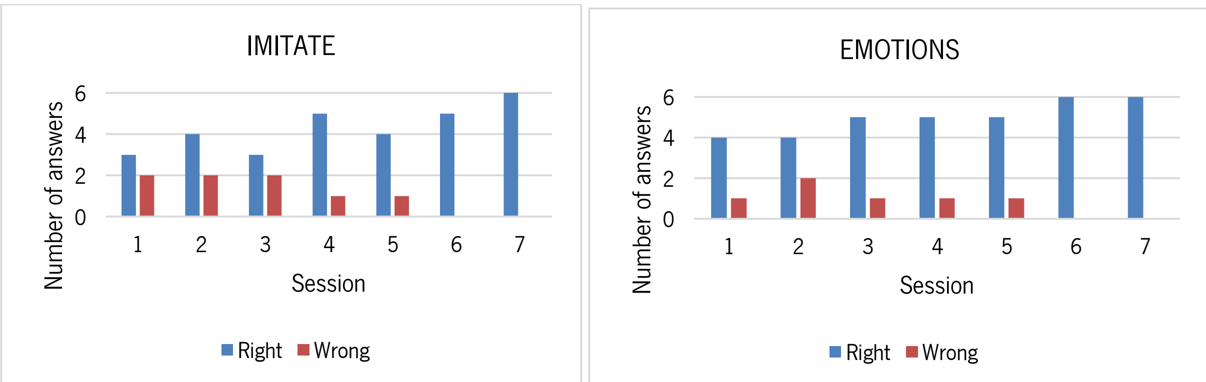

Figure 5.3.2-3 Child F results - On the left the progress over seven sessions with the IMITATE activity. On the right the progress over seven sessions with the EMOTIONS activity.

By analysing the performance of the three children in the activity IMITATE, the children D and F had a positive evolution, whereas the performance of the child E fluctuated, having an overall good performance. In the EMOTIONS activity the three children had, in general, a good performance over the sessions. In particular, child F had a more notable positive evolution. This child improved his performance in displaying the anger facial expression, since until the las two sessions He/she would did not perform correctly the anger expression.

Tables 5.3.2-1 and 5.3.2-2, present the children's mean response time, and standard deviation (SD), in the activities IMITATE and EMOTIONS, of successful answers given in the corresponding session, respectively. All children took more time answering to the prompt in the last session, Session 7. Generally, child D was usually faster to answer the prompt from the robot than the rest of the children.

Table 5.3.2-1 Children's in the subset two mean response time in seconds for successful answers (SD) in the IMITATE activity. In general, the response time increased in the last session.

Session number	**Child D**	**Child E**	**Child F**
1	17.68 (1.74)	19.14 (1.5)	20.37 (2.59)
2	16.54 (0.02)	21.41 (4.23)	19.05 (1.54)
3	17.64 (1.75)	16.88 (0.44)	22.20 (4.31)
4	17.89 (2.68)	17.18 (0.44)	16.54 (0.02)
5	18.43 (2.74)	17.31 (0.18)	17.89 (1.45)
6	18.17 (2.34)	17.04 (0.54)	18.73 (2.77)
7	20.47 (3.27)	18.45 (3.08)	20.01 (3.01)

Table 5.3.2-2 Children's in the subset two mean response time in seconds for successful answers (SD) in the EMOTIONS activity. In general, the response time increased in the last session.

Session number	**Child D**	**Child E**	**Child F**
1	18.27 (2.6)	19.92 (4.42)	21.58 (2.78)
2	19.00 (3.57)	18.80 (3.21)	18.75 (2.05)
3	19.50 (3.33)	18.63 (2.88)	17.08 (0.74)
4	17.89 (0.72)	17.79 (1.83)	18.29 (1.2)
5	16.63 (0.68)	17.35 (1.7)	17.73 (1.68)
6	16.75 (1.0)	17.56 (1.75)	17.84 (1.49)
7	18.27 (2.65)	18.45 (3.16)	18.82 (3.11)

Regarding the qualitative analysis in the first reaction of the child to the robot in the first session, all participants from the two subsets were specifically interested in the face of the robot, touching it repeatedly and always in a gentle way. None of the children abandoned the room. Moreover, with exception of the child C from the first subset, none of the participants got up of the chair during the sessions indicating, in general, that they were interested in the activity.

The results obtained allowed to conclude that the proposed system is able to interact with children with ASD in a comfortable and natural way, giving a strong indication about the use of this particular system in the context of emotion recognitions and imitation skills. Although the sample is small (and further tests are mandatory), the results point out that a humanoid robot can be used as an eligible mediator in emotions recognition activities with children with ASD.

References

Becker-Asano, C., & Ishiguro, H. (2011). Evaluating facial displays of emotion for the android robot Geminoid F. *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - WACI 2011: 2011 Workshop on Affective Computational Intelligence*, (20220002), 22–29. http://doi.org/10.1109/WACI.2011.5953147

Hsu, C. W., Chang, C. C., & Lin, C. J. (2010). A practical guide to support vector classication. *National Taiwan University*. Retrieved from http://www.csie.ntu.edu.tw/{~}cjlin/papers/guide/guide.pdf

Mazzei, D., Lazzeri, N., Hanson, D., & De Rossi, D. (2012). HEFES: An Hybrid Engine for Facial Expressions Synthesis to control human-like androids and avatars. *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, 195–200. http://doi.org/10.1109/BioRob.2012.6290687

Michel, P., & El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. *Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI '03*, 258. http://doi.org/10.1145/958468.958479

Noldus | Innovative solutions for behavioral research. (2015). Retrieved May 1, 2016, from http://www.noldus.com/

Souza, C. R. (2014). The Accord.NET Framework. Retrieved May 22, 2016, from http://accord-framework.net

Zhang, Y., Zhang, L., & Hossain, M. A. (2015). Adaptive 3D facial action intensity estimation and emotion recognition. *Expert Systems with Applications*, *42*(3), 1446–1464. http://doi.org/10.1016/j.eswa.2014.08.042

6 CONCLUSIONS AND FUTURE WORK

Summary

The following chapter draws the conclusion of the work described in the dissertation and provides some outlook for the future use of robotics in intervention with ASD.

The present work concerns the development and application of interactive and assistive technologies to support and promote new adaptive teaching/learning approaches for children with ASD. Nowadays, assistive robotics focus on helping Users with special needs in their daily activities. Assistive robots can be a social support to motivate children, socially educate them and beyond that help transferring knowledge. Moreover, they can be a useful tool to develop social-emotional skills in the intervention process of children with ASD.

However, robotic systems are emotionally blind. Conversely, successful human-human communication relies on the ability to read affective and emotional signals. Currently, assistive robots are getting "more emotional intelligent", since affective computing has been employed allowing to build a connection between the emotionally expressive human and the emotionally lacking computer.

Following this idea, the main goals of the present dissertation were to develop a system capable of automatically detecting facial expressions through facial cues and to interface the described system with a robotic platform in order to allow social interaction with children with ASD.

In order to achieve the proposed main goals, an experimental setup that uses the recent Intel RealSense 3D camera and the Zeno R50 Robokind platform (ZECA robot) was developed. This layout has two subsystems, a Mirroring Emotion System (MES) and an Emotion Recognition System (ERS). The first subsystem (MES) is capable of on-line synthetizing human emotions through facial expressions. The other subsystem (ERS) is able to recognize human emotions through facial features. MES extracts the User facial Action Units (AUs), and sends the data to the robot allowing on-line imitation. ERS uses Support Vector Machine (SVM) technique to automatic classify in real-time the emotion expressed by the User. In the works presented in the literature, the recognition of facial expressions is mostly performed by using SVMs but none uses the Intel RealSense to obtain the face data from the user. Therefore, the present work proposed a system that uses the recent Intel RealSense 3D to promote imitation and recognition of facial expressions, using a robot (ZECA) as a mediator in social activities.

The developed system was tested in different configurations, in order to assess the performance of each subsystem, the MES and ERS subsystems.

First, the MES subsystem was evaluated. In a first stage, the software FaceReader was used to automatically analyse the synthesized facial expressions (anger, fear, happiness, sadness, surprise, and neutral) on the robot when mimicking the performer. All except anger had a match higher than 50%. Even though most of the AUs necessary to represent anger were present, the emotional state was not correctly recognized. Most probably, in the robot face the AUs were not marked enough for the software to recognize them. Additionally, it may be difficult for the software to recognize facial expressions in a non-human face as it is prepared to work with human faces.

The perceptual study, performed with adults and children using a questionnaire, allowed to conclude that surprise obtained the highest similarity, with a score of 100% in both groups. Happiness and sadness have similar matching scores, both in children (87%) and adults (96%). Anger and fear were the facial expressions with the lowest performance. In children the matching score was 55% for anger and 58% for fear. However, these facial expressions had a better matching rate in adults, with 93% for both expressions.

A perceptual experiment was conducted with children from 6 to 9 years old to evaluate the proposed system. The overall recognition scores of the expressed emotions were: 97% for happiness; 72% for sadness without head movement; 81% for sadness with head bowing; and 84% for the neutral state. These results indicate that the first subsystem (MES) based on the Intel RealSense 3D sensor can, on-line and accurately, map facial expressions of a user onto a robot.

Finally, the performance of the ERS subsystem was assessed. The subsystem was first tested in simulation using Matlab and the performance of the two kernels was compared. RBF presented the best results, as the relation between class labels and attributes is nonlinear, with an average accuracy of 93.6%. Then, the real-time subsystem was tested in a laboratorial environment with a set of 14 participants, obtaining an overall accuracy of 88%. The required time for the system to efficiently perform facial expression recognition is 1-3ms at frame rate of 30 fps on an i5 quad-core CPUs with 16 GB RAM. Then, the proposed subsystem was compared to other state-of-the-art 3D facial expression recognition development in terms of overall accuracy, obtaining a performance of 88% against 84%, respectively. Finally, an experimental study, involving six children with ASD aged between eight and nine, was conducted in a school environment in order to evaluate the two game scenarios that are part of the ERS subsystem: the IMITATE, where the child has to mimic the ZECA's facial expression and EMOTIONS, where the child has to perform the facial expression asked by ZECA. The original group of six was uniformly divided into two subsets of three children (one and two). In the subset one, three different facial expressions were investigated (anger, happiness, and sadness). On the other hand, in the subset two, five facial expressions were investigated (anger, fear, happiness, sadness, and surprised). By analysing the results from both subsets, it is possible to conclude that in general the children had a positive evolution over the sessions. In general, all children took more time answering to the prompt in the last session. The increase of the response time over the sessions might be related to the children thinking and considering all options they have available. These results give a strong indication about the use of this particular system in the context of emotion recognitions and imitation skills.

The results obtained allowed to conclude that the proposed system is able to interact with children with ASD in a comfortable and natural way, giving a strong indication about the use of this particular system in the context of emotion recognitions and imitation skills. Although the sample is small (and further tests are mandatory), the results point out that a humanoid robot can be used as an eligible mediator in emotions recognition activities with children with ASD.

Future work developments could be divided into short and long term research. Regarding the short term research, further experiments could be conducted to conclude the suitability of the proposed system to be used as a complement to the traditional interventions.

Long term research can employ a system which modifies the robot behaviour to the child's actions during an intervention session. The adaptation can be based on a predictive model using a database of non-verbal behaviours, eye movements' analysis and the child's performance.

Additionally, objects based on playware technology can be used in order to interact with the children. Playware is defined as intelligent technology for children's play and playful experiences for the User. This technology emphasizes the role of interplay between morphology and control using processing, input and output.

This would lead to a hybrid approach composed by robots and playware technology to interact with children with ASD.

REFERENCES

Alabbasi, H. A., Moldoveanu, P., & Moldoveanu, A. (2015). Real Time Facial Emotion Recognition using Kinect V2 Sensor. *IOSR Journal of Computer Engineering Ver. II*, *17*(3), 2278–661. http://doi.org/10.9790/0661-17326168

Ambady, N., & Rosenthal, R. (1992). Thin Slices of Expressive Behavior as Predictors of Interpersonal Consequences: A Meta-Analysis.

Ashraf, A. B., Lucey, S., Cohn, J. F., Chen, T., Ambadar, Z., Prkachin, K., ... Theobald, B.-J. (2007). The Painful Face - Pain Expression Recognition Using Active Appearance Models. *IEEE Transactions on Systems, Man, and Cybernetics*, *27*(12), 9–14. http://doi.org/10.1016/j.imavis.2009.05.007

Baltrusaitis, T. (2014). *Automatic facial expression analysis*. *Pattern recognition*. University of Cambridge. Retrieved from http://www.sciencedirect.com/science/article/pii/S0031320302000523

Baron-Cohen, S., & Wheelwright, S. (2004). The empathy quotient: An investigation of adults with asperger syndrome or high functioning autism, and normal sex differences. *Journal of Autism and Developmental Disorders*, *34*(2), 163–175. http://doi.org/10.1023/B:JADD.0000022607.19833.00

Becker-Asano, C., & Ishiguro, H. (2011). Evaluating facial displays of emotion for the android robot Geminoid F. *IEEE SSCI 2011 - Symposium Series on Computational Intelligence - WACI 2011: 2011 Workshop on Affective Computational Intelligence*, (20220002), 22–29. http://doi.org/10.1109/WACI.2011.5953147

Breazeal, C. (2000). *Sociable machines: Expressive social exchange between humans and robots*. *Expressive Social Exchange Between Humans and Robots*. Retrieved from http://groups.csail.mit.edu/lbr/mars/pubs/phd.pdf

Burges, C. C. (1998). A Tutorial on Support Vector Machines for Pattern Recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167. http://doi.org/10.1023/A:1009715923555

Camera range. (2015). Retrieved February 4, 2016, from https://software.intel.com/sites/default/files/managed/99/95/camera2.png

Chih-Wei Hsu, Chih-Chung Chang, and C.-J. L. (2008). A Practical Guide to Support Vector Classification. *BJU International*, *101*(1), 1396–400. http://doi.org/10.1177/02632760022050997

Christoph, N. (2005). Empirical Evaluation Methodology for Embodied Conversational Agents. In *From Brows to Trust SE - 3* (Vol. 7, pp. 67–90). http://doi.org/10.1007/1-4020-2730-3_3

Cohn, J. (2007). Foundations of human-centered computing: Facial expression and emotion. *International Joint Conference on Artificial Intelligence*, 233–238. http://doi.org/10.1145/1180995.1181043

Confusion matrix.png. (2015). Retrieved May 22, 2016, from http://3.bp.blogspot.com/_txFWHHNYMJQ/THyADzbutYI/AAAAAAAAAf8/TAXL7IySrko/s1600/Picture+8.png

Costa, S. (2014). *Affective Robotics for Socio-Emotional Skills Development in Children with Autism Spectrum Disorders*. University of Minho.

Costa, S., Soares, F., & Santos, C. (2013). Facial expressions and gestures to convey emotions with a humanoid robot. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *8239 LNAI*, 542–551. http://doi.org/10.1007/978-3-319-02675-6_54

Cross-Validation - MATLAB. (2015). Retrieved January 20, 2016, from http://www.mathworks.com/discovery/cross-validation.html

Cross-validation: evaluating estimator performance. (2015). Retrieved May 22, 2016, from http://scikit-learn.org/stable/modules/cross_validation.html

D'Mello, S., & Calvo, R. (2013). Beyond the basic emotions: what should affective computing compute? *... Abstracts on Human Factors in Computing ...*, 2287–2294. http://doi.org/10.1145/2468356.2468751

Darwin, C. (1872). *The expression of emotions in man and animals*. *Library* (Vol. 3). http://doi.org/10.1097/00000441-195610000-00024

Dautenhahn, K. (2000). Design issues on interactive environments for children with autism. *3Th International Conference on Disability ,Virtual Reality and Associated Technologies*, 153–162. http://doi.org/10.1.1.33.4997

Ekman, P., & Friesen, W. V. (1978). *Facial Action Coding System: A Technique for the Measurement of Facial Movement*. *Consulting Psychologists Press*.

Ekman, P., Friesen, W. V, O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ... Tzavaras, A. (1987). Universals and cultural differences in the judgments of facial expressions of emotion.

Journal of Personality and Social Psychology. US: American Psychological Association. http://doi.org/10.1037/0022-3514.53.4.712

Ekman, P., & Rosenberg, E. (2005). *What the face reveals*. *What the Face Reveals*.

Emotient | A Leader in Emotion Measurement. (2015). Retrieved May 1, 2016, from http://www.emotient.com/

Face Pose Data [F200,SR300]. (2015). Retrieved January 16, 2016, from https://software.intel.com/sites/landingpage/realsense/camera-sdk/v1.1/documentation/html/index.html?doc_face_face_pose_data.html

FACS (Facial Action Coding System). (2002). Retrieved January 25, 2016, from http://www.cs.cmu.edu/~face/facs.htm

Ghimire, D., Lee, J., Li, Z. N., Jeong, S., Park, S. H., & Choi, H. S. (2015). Recognition of facial expressions based on tracking and selection of discriminative geometric features. *International Journal of Multimedia and Ubiquitous Engineering*, *10*(3), 35–44. http://doi.org/10.14257/ijmue.2015.10.3.04

Hadar, U., Steiner, T. J., & Clifford Rose, F. (1985). Head movement during listening turns in conversation. *Journal of Nonverbal Behavior*, *9*(4), 214–228. http://doi.org/10.1007/BF00986881

Happé, F., Briskman, J., Frith, U., Happé, F., & Frith, U. (2001). Exploring the cognitive phenotype of autism: weak “central coherence” in parents and siblings of children with autism: II. Real-life skills and preferences. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *42*(3), 309–16. http://doi.org/10.1017/S0021963001006916

Hashimoto, T., Kobayashi, H., & Kato, N. (2011). Educational system with the android robot SAYA and field trial. *IEEE International Conference on Fuzzy Systems*, *8*(3), 766–771. http://doi.org/10.1109/FUZZY.2011.6007430

Hobson, R. P. (1986). The autistic child’s appraisal of expressions of emotion: a further study. *Journal of Child Psychology and Psychiatry, and Allied Disciplines*, *27*(February), 671–680. http://doi.org/10.1111/j.1469-7610.1986.tb01836.x

Hsu, C., & Lin, C. (2002). A comparison of methods for multiclass support vector machines. *Neural Networks, IEEE Transactions on*, *13*(2), 415–425. http://doi.org/10.1109/TNN.2002.1021904

Hsu, C. W., Chang, C. C., & Lin, C. J. (2010). A practical guide to support vector classication. *National Taiwan University*. Retrieved from http://www.csie.ntu.edu.tw/{~}cjlin/papers/guide/guide.pdf

Hudlicka, E. (2008). Affective Computing for Game Design. *Scientist*, 5–12. Retrieved from http://www.imgd.wpi.edu/speakers/0910/Hudlicka_GAMEON_TUT_01.pdf

Ingersoll, B. (2008). The Social Role of Imitation in Autism. *Infants & Young Children*, *21*(2), 107–119. http://doi.org/10.1097/01.IYC.0000314482.24087.14

Intel® RealSense™ Technology. (2015). Retrieved January 15, 2016, from http://www.intel.com/content/www/us/en/architecture-and-technology/realsense-overview.html

Jamshidnezhad, A., & Jan Nordin, M. D. (2012). Challenging of facial expressions classification systems: Survey, critical considerations and direction of future work. *Research Journal of Applied Sciences, Engineering and Technology*, *4*(9), 1155–1165.

Jurman, G., & Furlanello, C. (2010). A unifying view for performance measures in multi-class prediction. *PLoS ONE*, *7*(8). http://doi.org/10.1371/journal.pone.0041882

Kim, E. S., Berkovits, L. D., Bernier, E. P., Leyzberg, D., Shic, F., Paul, R., & Scassellati, B. (2013). Social robots as embedded reinforcers of social behavior in children with autism. *Journal of Autism and Developmental Disorders*, *43*(5), 1038–1049. http://doi.org/10.1007/s10803-012-1645-2

Kotsia, I., & Pitas, I. (2007). Facial expression recognition in image sequences using geometric deformation features and support vector machines. *IEEE Transactions on Image Processing*, *16*(1), 172–187. http://doi.org/10.1109/TIP.2006.884954

Lee, C.-H. J., Kim, K., Breazeal, C., & Picard, R. (2008). Shybot: friend-stranger interaction for children living with autism. *Proceedings of ACM CHI 2008 Conference on Human Factors in Computing Systems*, *2*, 3375–3380. http://doi.org/http://doi.acm.org/10.1145/1358628.1358860

Lee, Y., Terzopoulos, D., & Walters, K. (1995). Realistic modeling for facial animation. *Proceedings of the 22nd Annual Conference on Computer Graphics and Interactive Techniques SIGGRAPH 95*, *95*, 55–62. http://doi.org/10.1145/218380.218407

Loconsole, C., Miranda, C. R., Augusto, G., & Frisoli, A. (2004). Real-Time Emotion Recognition: a Novel Method for Geometrical Facial Features Extraction.

Mao, Q.-R., Pan, X.-Y., Zhan, Y.-Z., & Shen, X.-J. (2015). Using Kinect for real-time emotion recognition via facial expressions *. *Front Inform Technol Electron Eng*, *16*(4), 272–282. http://doi.org/10.1631/FITEE.1400209

Mazzei, D., Lazzeri, N., Billeci, L., Igliozzi, R., Mancini, A., Ahluwalia, A., ... De Rossi, D. (2011). Development and evaluation of a social robot platform for therapy in autism. *Conference Proceedings : ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual Conference*, *2011*, 4515–8. http://doi.org/10.1109/IEMBS.2011.6091119

Mazzei, D., Lazzeri, N., Hanson, D., & De Rossi, D. (2012). HEFES: An Hybrid Engine for Facial

Expressions Synthesis to control human-like androids and avatars. *Proceedings of the IEEE RAS and EMBS International Conference on Biomedical Robotics and Biomechatronics*, 195–200. http://doi.org/10.1109/BioRob.2012.6290687

Meah, L. (2014). R50 "Zeno" User Guide. Sheffield. Retrieved from http://www.robokindrobots.com/wp-content/uploads/2014/11/UnivOfSheffield.pdf

Michaud, F., Laplante, J.-F., Larouche, H., Duquette, a., Caron, S., Letourneau, D., & Masson, P. (2005). Autonomous spherical mobile robot for child-development studies. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans*, *35*(4), 471–480. http://doi.org/10.1109/TSMCA.2005.850596

Michel, P., & El Kaliouby, R. (2003). Real time facial expression recognition in video using support vector machines. *Proceedings of the 5th International Conference on Multimodal Interfaces - ICMI '03*, 258. http://doi.org/10.1145/958468.958479

Michel, P., & Kaliouby, R. El. (2000). Facial Expression Recognition Using Support Vector Machines. Retrieved from http://www.cs.cmu.edu/~pmichel/publications/Michel-FacExpRecSVMAbstract.pdf

Milgram, J., Cheriet, M., & Sabourin, R. (2006). "One Against One" or "One Against All": Which One is Better for Handwriting Recognition with SVMs? *Tenth International Workshop on Frontiers in Handwriting Recognition*, 1–6. Retrieved from http://hal.inria.fr/inria-00103955

Moosaei, M., Hayes, C. J., & Riek, L. D. (2015). Performing Facial Expression Synthesis on Robot Faces: A Real-time Software System. *New Frontiers in Human-Robot Interaction*, 67. Retrieved from http://www.mahasalem.net/AISB2015/NF-HRI-2015-full_proceedings.pdf#page=68

Moosaei, M., & Riek, L. D. (2013). Evaluating Facial Expression Synthesis on Robots. *HRI Workshop on Applications for Emotional Robots at the 8th ACM/IEEE International Conference on Human-Robot Interaction (HRI).*, 1–6.

Morency, L., Sidner, C., Lee, C., & Darrell, T. (2006). The Role of Context in Head Gesture Recognition. *The Twenty-First National Conference on Artificial Intelligence (AAAI-06)*, 1650–1653.

Musicovery | Mobile Aplication. (2015). Retrieved May 1, 2016, from http://musicovery.com/

Noldus | Innovative solutions for behavioral research. (2015). Retrieved May 1, 2016, from http://www.noldus.com/

Ochs, M., Niewiadomski, R., Pelachaud, C., & Sadek, D. (2005). Intelligent expressions of emotions. *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, *3784 LNCS*, 707–714. http://doi.org/10.1007/11573548_91

Piaget. (1976). *Piaget and His School*. (B. Inhelder, H. H. Chipman, & C. Zwingmann, Eds.). Berlin, Heidelberg: Springer Berlin Heidelberg. http://doi.org/10.1007/978-3-642-46323-5

Picard, R., & Klein, J. (2002). Computers that recognize and respond to user emotion: theoretical and practical implications. *Interacting with Computers*, *14*(Spitz 1945), 141–169.

Picard, R. W. (1997). *Affective Computing*. *Rosalind W. Picard*. Retrieved from http://www.amazon.com/exec/obidos/ASIN/0262161702/acmorg-20

R50 Documentation -. (2014). Retrieved July 12, 2015, from http://www.robokindrobots.com/support-documentation/r50/

Rani, P., & Sarkar, N. (2007). Operator engagement detection for robot behavior adaptation. *International Journal of Advanced Robotic Systems*, *4*, 1–12. http://doi.org/10.1109/ROBOT.2005.1570415

RBF SVM parameters — scikit-learn 0.17.1 documentation. (2015). Retrieved May 20, 2016, from http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html#example-svm-plot-rbf-parameters-pyrealsense_intel. (2015). Retrieved from http://ade2.nhtv.nl/~dpj/sensorsatplay/images/realsense_intel.png

Ricks, D. J., & Colton, M. B. (2010). Trends and considerations in robot-assisted autism therapy. *Proceedings - IEEE International Conference on Robotics and Automation*, 4354–4359. http://doi.org/10.1109/ROBOT.2010.5509327

Robinson, P., & el Kaliouby, R. (2009). Computation of emotions in man and machines. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *364*(1535), 3441–3447. http://doi.org/10.1098/rstb.2009.0198

Robokind | Advanced Social Robots. (2015). Retrieved May 1, 2016, from http://www.robokindrobots.com/

roc_intro3.png. (2015). Retrieved May 26, 2016, from https://www.medcalc.org/manual/_help/images/roc_intro3.png

S.Messinger, D. (2014). Affective Computing, Emotional Development, and Autism, 516–536.

Sacavém, António; Wezowski, Kasia; Wezowski, P. (2014). *A Linguagem Corporal Revela o que as Palavras Escondem*. Top Books. Retrieved from http://www.wook.pt/ficha/a-linguagem-corporal-revela-o-que-as-palavras-escondem/a/id/15612673

Scassellati, B. (2007). How social robots will help us to diagnose , treat , and understand autism. *Robotics Research*, 552–563. http://doi.org/10.1007/978-3-540-48113-3_47

Scassellati, B., Henny Admoni, & Matarić, M. (2012). Robots for Use in Autism Research. *Annual Review of Biomedical Engineering*, *14*(1), 275–294. http://doi.org/10.1146/annurev-bioeng-071811-

150036

Silva, C., Sobral, A., & Vieira, R. T. (2014). An automatic facial expression recognition system evaluated with different classifiers. *X Workshop de Visão Computacional (WVC'2014)*, 208–212. Retrieved from http://www.researchgate.net/publication/264295959_An_automatic_facial_expression_recognition_system_evaluated_with_different_classifiers

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. http://doi.org/10.1016/j.ipm.2009.03.002

Sosnowski, S., Kuehnlenz, K., & Buss, M. (2006). EDDIE - An Emotion Display with Dynamic Intuitive Expressions. *2006 IEEE/RSJ International Conference on Intelligent Robots and Systems*, 3113–3118. http://doi.org/10.1109/IROS.2006.282208

Souza, C. R. (2014). The Accord.NET Framework. Retrieved May 22, 2016, from http://accord-framework.net

Support Vector Machines (SVM) - MATLAB & Simulink. (2015). Retrieved January 20, 2016, from http://www.mathworks.com/help/stats/support-vector-machines-svm.html

SVM Margins Example. (2014). Retrieved from http://scikit-learn.org/stable/auto_examples/svm/plot_svm_margin.html

Tapus, A., Member, S., & Scassellati, B. (2007). The Grand Challenges in Socially Assistive Robotics. *IEEE Robotics and Automation Magazine*, *14*, 1–7. http://doi.org/10.1109/MRA.2010.940150

The Bosphorus. (2009). Retrieved May 10, 2016, from http://bosphorus.ee.boun.edu.tr/Home.aspx

Williams, J. H. G., Whiten, A., Suddendorf, T., & Perrett, D. I. (2001). Imitation, mirror neurons and autism. *Neuroscience and Biobehavioral Reviews*. http://doi.org/10.1016/S0149-7634(01)00014-8

Wu, J., Wang, J., & Liu, L. (2007). Feature extraction via KPCA for classification of gait patterns. *Human Movement Science*, *26*(3), 393–411. http://doi.org/10.1016/j.humov.2007.01.015

Youssef, A. E., Aly, S. F., Ibrahim, A. S., & Abbott, a. L. (2013). Auto-Optimized Multimodal Expression Recognition Framework Using 3D Kinect Data for ASD Therapeutic Aid. *International Journal of Modeling and Optimization*, *3*(2), 112–115. http://doi.org/10.7763/IJMO.2013.V3.247

Zhang, Y., Zhang, L., & Hossain, M. A. (2015). Adaptive 3D facial action intensity estimation and emotion recognition. *Expert Systems with Applications*, *42*(3), 1446–1464. http://doi.org/10.1016/j.eswa.2014.08.042

Zheng, A. (2015). How to Evaluate Machine Learning Models: Classification Metrics. Retrieved May 22, 2016, from http://blog.dato.com/how-to-evaluate-machine-learning-models-part-2a-classification-metrics

APPENDIXES

Summary

Materials used in the present dissertation.

Appendix

A.1 Consent form delivered to the children's parents (In Portuguese)

---*Front page*---
----

7 de Janeiro de 2016

Exmo.(a). Senhor(a) Encarregado(a) de Educação/Tutor(a),

Vários investigadores por todo o mundo têm-se dedicado ao estudo da **influência da utilização de robôs no desenvolvimento de competências cognitivas e comportamentais em crianças com Perturbações do Espectro do Autismo (PEA)**.

O **projeto Robótica-Autismo** (http://robotica-autismo.com/) visa a aplicação de ferramentas robóticas como forma de melhorar a vida social de jovens com problemas intelectuais e jovens com PEA. Em particular, o objetivo é **melhorar as habilidades de interação e comunicação com o meio-ambiente e com outras pessoas**.

No âmbito de uma Dissertação de Mestrado em Engenharia Eletrónica Industrial e Computadores da Universidade do Minho, estamos a desenvolver um sistema de imitação e reconhecimento de emoções em que o robô Zeca é o mediador da interação. Numa primeira fase, e de forma a validar o sistema desenvolvido, gostaríamos de trabalhar com crianças do 1º ciclo sem PEA.

Assim, gostaríamos de convidar o seu educando a participar nas sessões de teste: a criança deve exprimir uma emoção (contente, triste, medo, zangado, assustado ou neutro) de cada vez e o robô deverá imitá-la. Estas sessões têm uma duração de cerca de 15 minutos, são realizadas durante o tempo letivo, mas sem prejuízo do normal funcionamento das aulas, NÃO são gravadas e as respostas são anónimas. Garantimos, desde já, que os dados recolhidos serão apenas utilizados nesta investigação e divulgação científica da mesma, nunca explorando os mesmos noutras situações que não estejam relacionadas com o projeto. Solicitamos, assim, a sua colaboração dando o seu consentimento através da devolução do anexo devidamente preenchido e assinado.

Com os melhores cumprimentos,

Filomena Oliveira Soares
Coordenadora Científica do Projeto Robótica-Autismo
Professora Associada do Departamento de Electrónica Industrial
Universidade do Minho

Vinícius Silva
Responsável pelo estudo
Aluno finalista do Mestrado Integrado em Engenharia Electrónica Industrial e Computadores
Departamento de Electrónica Industrial
Universidade do Minho

---*Back page*---

Eu____encarregado(a) de Educação do(a)/tutor(a) do(a) ____ declaro ter compreendido os objetivos do estudo, ter-me sido dada a oportunidade de fazer todas as perguntas sobre o assunto e para todas elas ter obtido resposta esclarecedora, ter-me sido garantido que não haverá prejuízo para os direitos assistenciais se eu recusar esta solicitação, e ter-me sido dado tempo suficiente para refletir sobre esta proposta.

Declaro também que autorizo o meu (a minha) educando(a) a participar no Projeto de Investigação Robótica-Autismo, em particular na interação com o robô Zeca.

Fui informado(a) que:

- Os resultados decorrentes desta investigação serão utilizados única e exclusivamente na divulgação científica do projeto.
- Os dados pessoais e os dados obtidos na investigação não serão divulgados e serão mantidos por um período de dez anos, ao fim do qual serão destruídos.
- Todas as informações de caráter pessoal recolhidas no decurso da investigação serão consideradas confidenciais e tratadas de acordo com as regras relativas à proteção de dados e da vida privada.
- Se o encarregado(a) de educação/tutor(a) o entender, o aluno (a aluna) pode abandonar o projeto em qualquer altura.
- A participação, a recusa na participação ou o posterior abandono do(a) encarregado(a) de educação/tutor(a), e/ou a do seu (da sua) dependente não prejudicarão a relação com a equipa de investigadores.
- Não se preveem quaisquer riscos para os participantes durante as sessões. Caso a criança demonstre desconforto, a sessão será terminada.

____ de_____________de 2016

____

Assinatura Completa do(a) Encarregado(a) de Educação e/ou tutor(a)

Este documento é composto de 2 páginas e feito em duplicado: uma via para os investigadores e outra para o(a) encarregado(a) de educação/tutor(a).

A.2 List of facial Action Units and Action Descriptors

Table A.2-1 List of Action Units and Action Descriptors (with underlying facial muscles) (Ekman & Friesen, 1978; “FACS (Facial Action Coding System),” 2002

AU#	FACS Name	Muscular Basis
0	Neutral Face	
1	Inner Brow Raiser	*Frontalis, pars medialis*
2	Outer Brow Raiser	*Frontalis, pars lateralis*
4	Brow Lowerer	*Corrugator supercilii, Depressor supercilii*
5	Upper Lid Raiser	*Levator palpebrae superioris*
6	Cheek Raiser	*Orbicularis oculi, pars orbitalis*
7	Lid Tightener	*Orbicularis oculi, pars palpebralis*
8	Lips Toward Each Other	*Orbicularis oris*
9	Nose Wrinkler	*Levator labii superioris alaquae nasi*
10	Upper Lid Raiser	*Levator labii superioris*
11	Nasolabial Deepener	*Zygomaticus minor*
12	Lip Corner Puller	*Zygomaticus major*
13	Sharp Lip Puller	*Levator anguli oris (also known as caninus)*
14	Dimpler	*Buccinator*
15	Lip Corner Depressor	*Depressor anguli oris (a.k.a. Triangularis)*
16	Lower Lip Depressor	*Depressor labii inferioris*
17	Chin Raiser	*Mentalis*
18	Lip Pucker	*Incisivii labii superioris and Incisivii labii inferioris*
19	Tongue Show	
20	Lip Stretcher	*Risorius w/ platysma*
21	Neck Tightener	*Platysma*
22	Lip Funneler	*Orbicularis oris*
23	Lip Tightener	*Orbicularis oris*
24	Lip Pressor	*Orbicularis oris*
25	Lips part	*Depressor labii inferioris or relaxation of Mentalis, or Orbicularis oris*
26	Jaw Drop	*Masseter, relaxed Temporalis and internal Pterygoid*
51	Head turn left	
52	Head turn right	

AU#	FACS Name	Muscular Basis
53	Head up	
54	Head down	
55	Head tilt right	
56	Head tilt right	
57	Head forward	
58	Head back	
61	Eyes turn left	
62	Eyes turn right	
63	Eyes up	
64	Eyes down	

A.3 Technical Drawings of the Robot

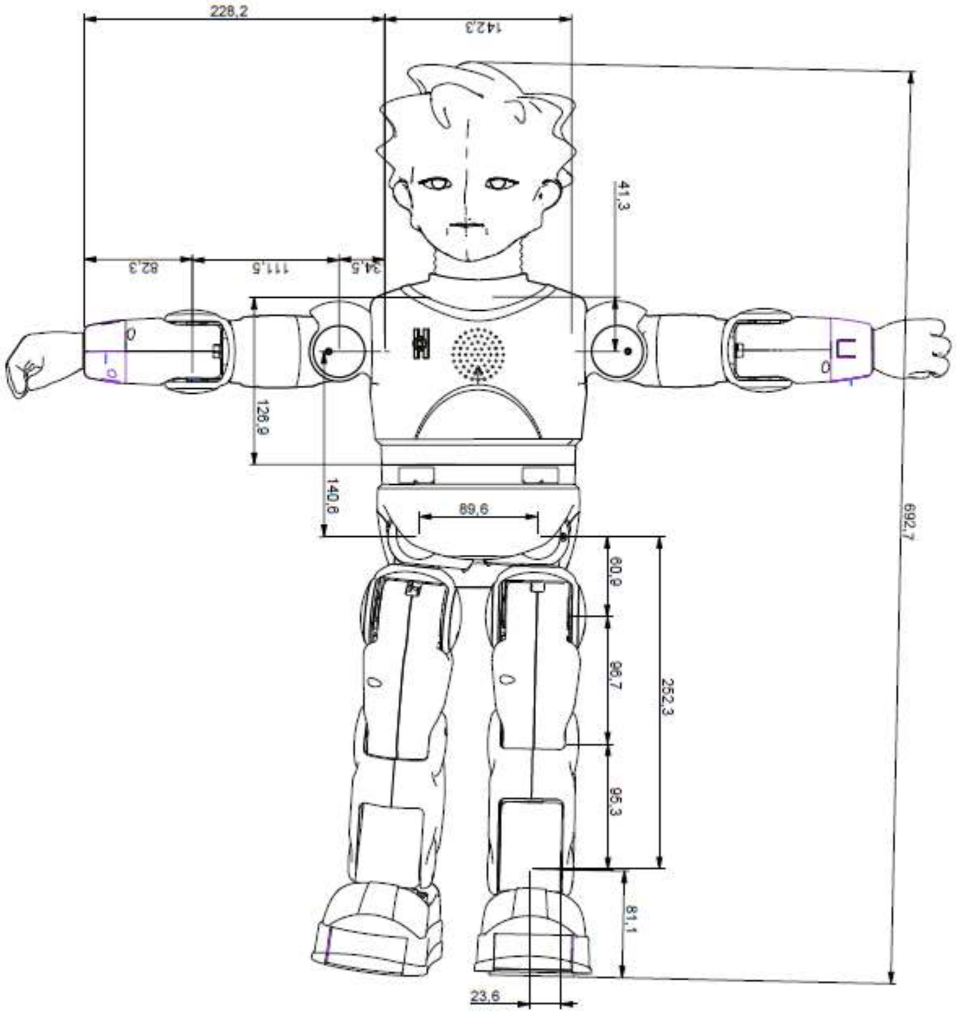

Figure A.3-1 Robot technical drawing showing its entire body in a front view (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

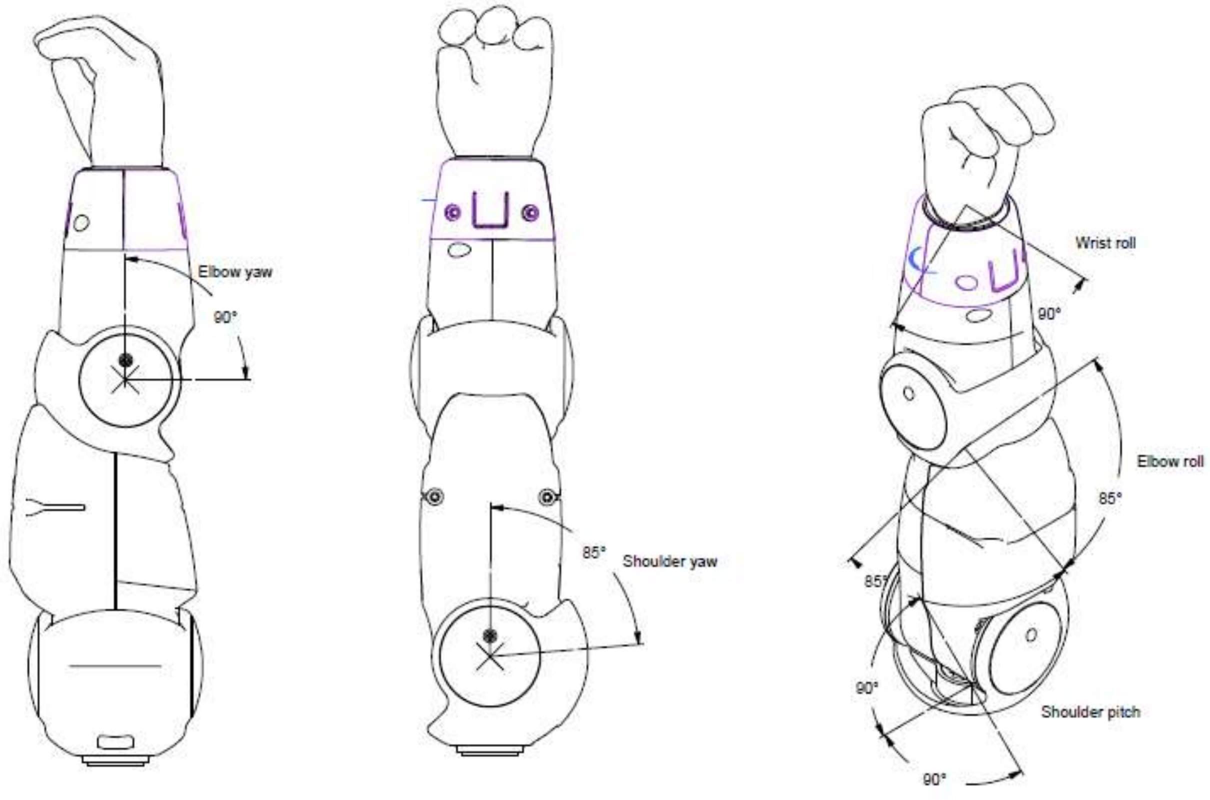

Figure A.3-2 Technical drawing of the robot showing its arms (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

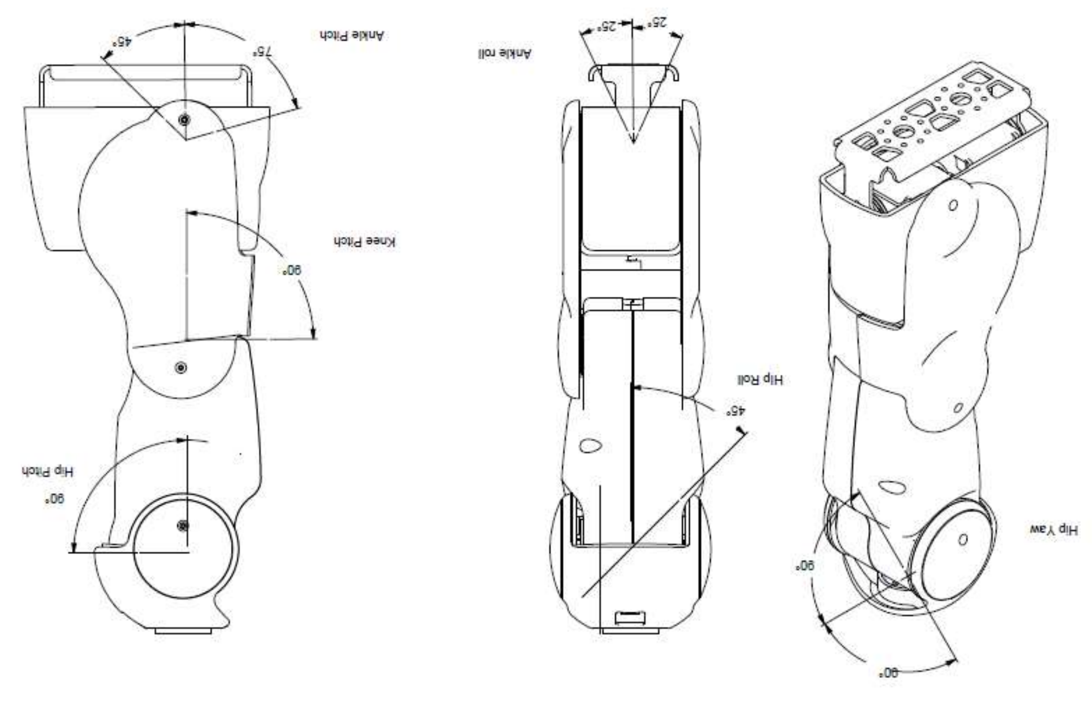

Figure A.3-3 Technical drawing of the robot showing its legs (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

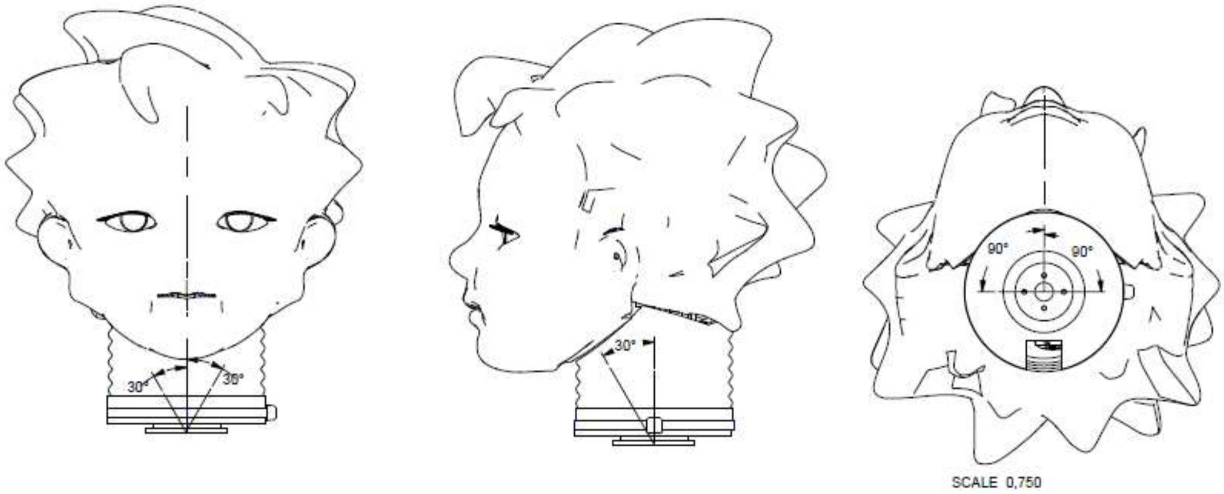

Figure A.3-4 Robot technical drawing showing specifically its head (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).

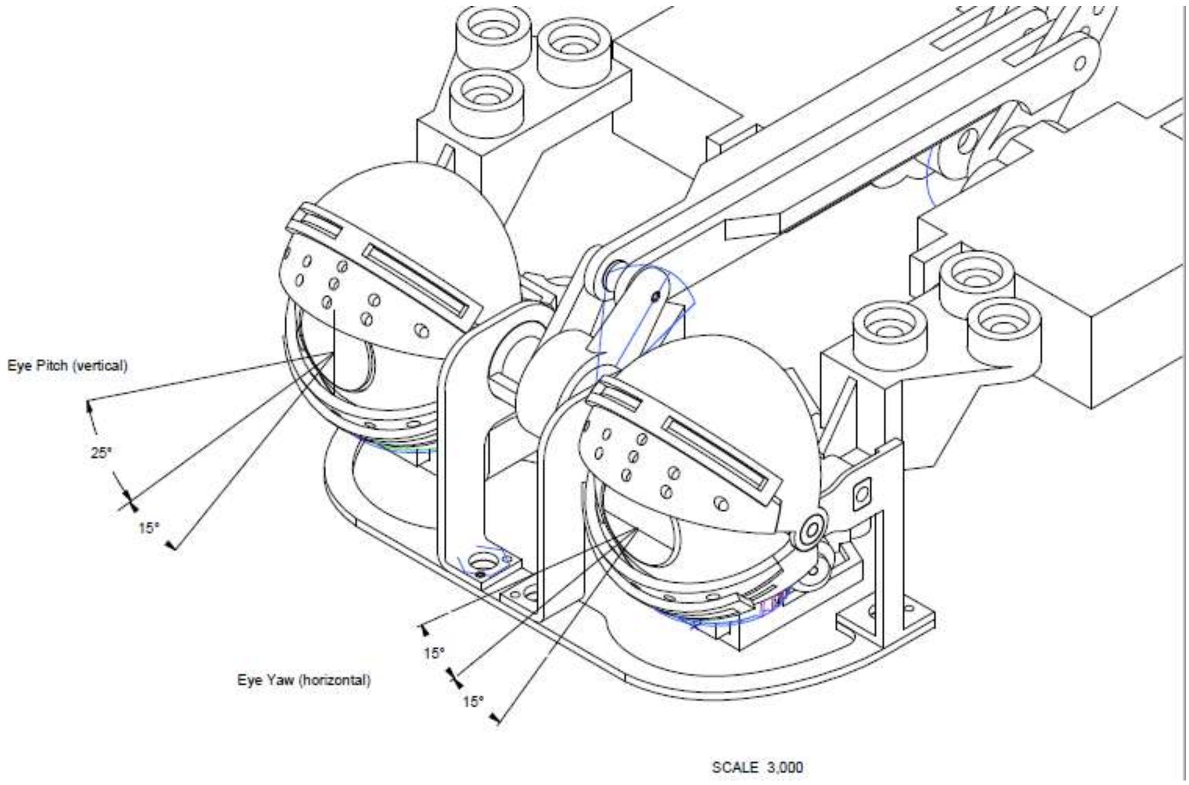

Figure A.3-5 Technical drawing of the robot showing the detailing the eyes (retrieved from: http://www.robokindrobots.com/support-documentation/r50/).