Universidade do Minho
Escola de Engenharia

Bruno Filipe Viana Amaro

# Behavioural attentiveness patterns analysis – detecting distraction behaviours

Dissertação de Mestrado

Mestrado Integrado em Engenharia Eletrónica Industrial e Computadores

Trabalho efetuado sob a orientação da

Professora Doutora Filomena Soares

Outubro de 2018

# ACKNOWLEDGEMENTS

Finished this work, that represents the closing of a cycle so important in my life, I can not fail to thank all those who, more or less actively, but not least, had a decisive role in keeping me motivated and determined to achieve my goals.

So I would like to thank:

- To my supervisor Dr. Filomena Soares, who was in a friendly and extremely professional way able to broaden my point of view and to arouse in me a strong interest in the area in which I focused my dissertation. It was the person who, in the difficult phase of choosing a dissertation theme, was kind enough to introduce me and invite me to integrate a project of great interest and application at a professional and social level. I can not fail to mention how she constantly allowed me to use all the means at her disposal to ensure that nothing was lacking in the proper development of the work.
- To my co-supervisor and colleague Eng. Vinícius Silva for his strong professionalism, motivation and example of dedication to the project in particular in the area of machine learning. Without his collaboration, I would not have reached the proposed objectives so quickly. His collaboration was fundamental at the technical and research level.
- To Dr. João Sena Esteves, for his help and interest expressed in the support and analysis of the project.
- To my colleagues Eng. Pedro Leite, Eng. Alexandre Calado, Huang Hao and André Teixeira, for their collaboration whenever necessary, and for help in obtaining data for the development of a machine learning algorithm.
- The basic school in Gualtar, especially the therapists and children involved in the special education unit.
- To my family, especially to my father, for the support and encouragement given, during these 5 years of study, without their support none of this would be possible.

## AGRADECIMENTOS

Ao terminar este trabalho que representa o fecho de um ciclo tão importante na minha vida, não poderei deixar de agradecer a todos aqueles que de forma mais ou menos ativa, mas não menos importante, tiveram um papel determinante para me manter motivado e determinado a atingir os meus objetivos. Assim sendo gostaria de agradecer:

- Á minha orientadora Dra. Filomena Soares, que foi quem de forma simpática e extremamente profissional conseguiu alargar o meu ponto de vista e despertar em mim um forte interesse pela área em que centrei a minha dissertação. Foi a pessoa que na difícil fase de optar por um tema de dissertação, teve a amabilidade de me apresentar e convidar para integrar um projeto de grande interesse e aplicação ao nível profissional e social. Não poderei deixar de referir a forma como constantemente me permitiu a utilização de todos os meios ao seu alcance para que nada faltasse ao bom desenvolvimento do trabalho.
- Ao meu coorientador e colega Eng. Vinícius Silva pelo forte profissionalismo, motivação e exemplo de dedicação ao projeto em particular na área de *"machine learning".* Sem a sua colaboração não teria chegado de forma tão rápida aos objetivos propostos. A sua colaboração foi fundamental ao nível técnico e de pesquisa.
- Ao Dr. João Sena Esteves, pela sua ajuda e interesse manifestado no apoio e análise do projeto.
- Aos meus colegas Eng. Pedro Leite, Eng. Alexandre Calado, Huang Hao e André Teixeira, pela sua colaboração sempre que necessário, e pela ajuda na obtenção de dados para o desenvolvimento de um algoritmo de *"machine learning".*
- Á escola básica de Gualtar, em especial às terapeutas e crianças envolvidas na unidade de ensino especial.
- Á minha família, especialmente ao meu pai, pelo apoio e incentivo dado, ao longo destes 5 anos de estudo, sem o seu apoio nada disto seria possível.

# ABSTRACT

The capacity of remaining focused on a task can be crucial in some circumstances. In general, this ability is intrinsic in a human social interaction and it is naturally used in any social context. Nevertheless, some individuals have difficulties in remaining concentrated in an activity, resulting in a short attention span. Children with Autism Spectrum Disorder (ASD) are a special example of such individuals. ASD is a group of complex developmental disorders of the brain. Individuals affected by this disorder are characterized by repetitive patterns of behaviour, restricted activities or interests, and impairments in social communication. The use of robots has already proved to encourage the developing of social interaction skills lacking in children with ASD. However, most of these systems are controlled remotely and cannot adapt automatically to the situation, and even those who are more autonomous still cannot perceive whether or not the user is paying attention to the instructions and actions of the robot.

Following this trend, this dissertation is part of a research project that has been under development for some years. In this project, the Robot ZECA (Zeno Engaging Children with Autism) from Hanson Robotics is used to promote the interaction with children with ASD helping them to recognize emotions, and to acquire new knowledge in order to promote social interaction and communication with the others.

The main purpose of this dissertation is to know whether the user is distracted during an activity. In the future, the objective is to interface this system with ZECA to consequently adapt its behaviour taking into account the individual affective state during an emotion imitation activity. In order to recognize human distraction behaviours and capture the user attention, several patterns of distraction, as well as systems to automatically detect them, have been developed. One of the most used distraction patterns detection methods is based on the measurement of the head pose and eye gaze. The present dissertation proposes a system based on a Red Green Blue (RGB) camera, capable of detecting the distraction patterns, head pose, eye gaze, blinks frequency, and the user to position towards the camera, during an activity, and then classify the user's state using a machine learning algorithm.

Finally, the proposed system is evaluated in a laboratorial and controlled environment in order to verify if it is capable to detect the patterns of distraction. The results of these preliminary tests allowed to detect some system constraints, as well as to validate its adequacy to later use it in an intervention setting.

KEYWORDS: HUMAN-ROBOT INTERACTION, ZECA ROBOT, DISTRACTION PATTERNS, EMOTIONAL STATES, MACHINE LEARNING

# RESUMO

A capacidade de permanecer focado numa tarefa pode ser crucial em algumas circunstâncias. No geral, essa capacidade é intrínseca numa interação social humana e é naturalmente usada em qualquer contexto social. No entanto, alguns indivíduos têm dificuldades em permanecer concentrados numa atividade, resultando num curto período de atenção. Crianças com Perturbações do Espectro do Autismo (PEA) são um exemplo especial de tais indivíduos. PEA é um grupo de perturbações complexas do desenvolvimento do cérebro. Os indivíduos afetados por estas perturbações são caracterizados por padrões repetitivos de comportamento, atividades ou interesses restritos e deficiências na comunicação social. O uso de robôs já provaram encorajar a promoção da interação social e ajudaram no desenvolvimento de competências deficitárias nas crianças com PEA. No entanto, a maioria desses sistemas é controlada remotamente e não consegue-se adaptar automaticamente à situação, e mesmo aqueles que são mais autônomos ainda não conseguem perceber se o utilizador está ou não atento às instruções e ações do robô. Seguindo esta tendência, esta dissertação é parte de um projeto de pesquisa que vem sendo desenvolvido há alguns anos, onde o robô ZECA (Zeno Envolvendo Crianças com Autismo) da *Hanson Robotics* é usado para promover a interação com crianças com PEA, ajudando-as a reconhecer emoções, adquirir novos conhecimentos para promover a interação social e comunicação com os pares. O principal objetivo desta dissertação é saber se o utilizador está distraído durante uma atividade. No futuro, o objetivo é fazer a interface deste sistema com o ZECA para, consequentemente, adaptar o seu comportamento tendo em conta o estado afetivo do utilizador durante uma atividade de imitação de emoções. A fim de reconhecer os comportamentos de distração humana e captar a atenção do utilizador, vários padrões de distração, bem como sistemas para detetá-los automaticamente, foram desenvolvidos. Um dos métodos de deteção de padrões de distração mais utilizados baseia-se na medição da orientação da cabeça e da orientação do olhar. A presente dissertação propõe um sistema baseado numa câmera *Red Green Blue* (RGB), capaz de detetar os padrões de distração, orientação da cabeça, orientação do olhar, frequência do piscar de olhos e a posição do utilizador em frente da câmera, durante uma atividade, e então classificar o estado do utilizador usando um algoritmo de "*machine learning*". Por fim, o sistema proposto é avaliado num ambiente laboratorial, a fim de verificar se é capaz de detetar os padrões de distração. Os resultados destes testes preliminares permitiram detetar algumas restrições do sistema, bem como validar a sua adequação para posteriormente utilizá-lo num ambiente de intervenção.

PALAVRAS CHAVE: INTERAÇÃO HUMANO-ROBÔ, ROBÔ ZECA, PADRÕES DE DISTRAÇÃO, ESTADO EMOCIONAL, APRENDIZAGEM DA MÁQUINA.

## CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# ABBREVIATIONS

**ADHD A**ttention **D**eficit **H**yperactivity **D**isorder

**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

**BPAS B**ehavioural **A**ttentiveness patterns **A**nalysis **S**ystem

**CV C**ross **V**alidation

**DoF D**egrees of **F**reedom

**ECG E**lectro**c**ardio**g**ram

**FACS F**acial **A**ction **C**odding **S**ystem

**GUI G**raphical **U**ser **I**nterface

**HCI H**uman **C**omputer **I**nteraction

**HRI H**uman **R**obot **I**nteraction

**IMU I**nertial **M**easurement **U**nit

**k-NN k** – **N**earest **N**eighbours

**MCC M**athews **C**orrelation **C**oefficient

**OSH O**ptimal **S**eparating **H**yperplane

**RBF R**adial **B**asis **F**unction

**RGB R**ed **G**reen **B**lue

**ROC R**eceiver **O**perating **C**haracteristic

**SVM S**upport **V**ector **M**achines

**SVs S**upport **V**ectors

**ZECA Z**eca **E**ngaging **C**hildren with **A**utism

# 1. INTRODUCTION

**Summary**

In this chapter the motivations and problem statement of the work are presented. It starts with a brief reference to the difficulties of being attentive during a particular task, especially considering children with autism; a brief description regarding autism spectrum disorder is also given, as well as the works presented in the literature focused on the use of robots as mediators in the interaction with these children. 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.

Understanding human behaviour is a paramount subject. Paying attention uses specific networks in the brain. It is a skill that develops over time. I order to pay attention, we need to be alert. This allows us to sort out the right information from our surroundings and to put this information together. Children may find it hard to filter out distractions, which makes it difficult for them to pay attention (Kasari, Brady, Lord, & Tager-Flusberg, 2013), in particular children with special education needs, as it is the case of children with Autism Spectrum Disorder (ASD). These children are characterized by social-interaction difficulties, communication challenges and a tendency to engage in repetitive behaviours. The difficulties in social interaction begin to manifest between 8 and 10 months of age. These signs include not responding to their own name, lack of interest in other people and difficulty in expressing themselves verbally. During childhood, many children with autism prefer to play and have fun alone, instead of playing in groups. They also do not seek comfort or respond to expressions of anger or affection from parents. In accordance to the literature (Dawson, Webb, & Mcpartland, 2005; Klin, Jones, Schultz, Volkmar, & Cohen, 2002), authors suggest that individuals with ASD pay less attention to faces than typically developing individuals. Another symptom shown by this type of children is related to how they express themselves: instead of words, they tend to use their eyes, when someone calls them, and gestures to show what they want. Unusual repetitive behaviours are another symptom. These repetitive behaviours can be hand movements, swinging, jumping and spinning, arranging and rearranging objects, and repeating sounds, words, or phrases. Sometimes, repetitive behaviours such as wiggling fingers in front of eyes, are self-stimulating (“Symptoms | What is Autism? | Autism Speaks,” n.d.).

In an attempt to increase and captivate these individuals’ interest, authors have been proposing new technological tools in the field of assistive robotics to help users with special needs in their daily activities. Assistive robots are designed to identify, measure, and react to social behaviours by, being repetitive and objective, offering exceptional conditions for quantifying social behaviour (Tapus, Matari, Member, & Scassellati, n.d.).

They can be a social support to motivate children, socially educate them and beyond that to help transferring knowledge. Furthermore, research with assistive robots have showed that, in general, individuals with ASD express elevated interest while interacting with robots: increase attention (Kim, Paul, Shic, & Scassellati, 2012), recognition and imitation abilities (Fujimoto et al., 2011), verbal utterances (Kim, Paul, Shic, & Scassellati, 2012), among others. According to studies, it was observed that children with ASD 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). However, few projects worldwide pursue to include robots as part of the intervention program for individuals with autism (Chevalier et al., 2017; Dautenhahn, n.d.).

These robots are presented with different embodiments, varying their physical appearance from simple designs, e.g., four-wheeled mobile robots, to many levels of anthropomorphic forms, including humanoid (S. C. C. Costa, 2014), animal-like (Kim et al., 2013), and machine-like systems (F. Michaud et al., 2005). Recently, the research area of assistive robotics has moved to using robots with a humanoid design, since it can promise a great potential for generalisation, especially in tasks of imitation and emotion recognition which can be more complicated if the robot does not present a human form (Begum et al., 2015; Chevalier et al., 2017; S. C. C. Costa, 2014).

The majority of the systems proposed in the literature are controlled using the Wizard-of-Oz (WOZ) setup, meaning that the robot does not adapt its behaviour to the children's actions as it does not perceive them (Pennisi et al., 2016). Additionally, there have been various studies that used assistive robots with the goal of measuring the children eye gaze duration (S. Costa, Lehmann, Dautenhahn, Robins, & Soares, 2015; Robins, Dautenhahn, & Dickerson, 2009b) and direction. However, this analysis is not performed in real-time by the robot, meaning that, usually, the sessions are recorded, and the metrics are manually quantified during a post-analysis of the videos.

Taking this into account, it would be particularly relevant to detect patterns of distraction in real time in order to adapt the behaviour of the robot and to make the communication between the robot and the child more fluid.

## 1.1 Motivations and Problem Statement

As observed by various authors (Flanagan & Johansson, 2003; Gredebäck & Falck-Ytter, 2015), when an individual tries to reach a particular object, the observer's gaze reaches the target before the action is completed. Additionally, the predictive gaze provides the time for the observer to plan and execute an action towards a goal.

Following this idea, the attention span and the eye gaze can be very important elements in social interaction, as they can help an individual to perceive the intention of others. However, some individual's present a low attention span, especially children with ASD (Chevallier et al., 2015) when, for example, focusing on things that do not interest them, i.e., activities that involve shared attention.

Nowadays, robotic systems are used in social contexts in order to teach children new skills. For example, there are robotic platforms that, through games, can help children to interact with other people, by teaching them how to communicate, recognize emotions, and how to express themselves, among other things, thus helping them in their social integration (Ferrari et al., 2009; Robins et al., 2005).

In general, these systems lack the ability of understanding the child's level of interest during a certain activity, which compromises the robot-child interaction if no proper action is taken by the system when the child's attention diminishes. Consequently, the latter may not take as much advantage of the activity as he/she could. Taking this problem into account, one of the motivations for the research presented in this dissertation is the development of an algorithm capable of recognizing attention/distraction patterns of a user/child during an activity (e.g., detecting if he/she is engaged/interested in the activity by tracking his/her eye gaze and head motion) for later use with ZECA (Zeno Engaging Children with Autism), a RoboKind Zeno R50 robotic platform from Hanson Robotics. This will allow the system to detect if the user is not paying proper attention to a certain task and adapt its behaviour accordingly.

## 1.2 Objectives

The focus of the present work is the development of a framework to detect patterns of distraction during an activity. Generally, in order to track the user's distraction patterns, the literature approaches combine the use of several wearable sensors (which can be invasive) with non-wearable sensors. The proposed Behaviour Patterns Analysis System (BPAS) allows inferring children's distraction patterns (if any) when performing a task. In the future, this system may adapt the behaviour of the robot accordingly.

The final goal of the present work is to collect the selected patterns (head pose, eye gaze, blink frequency, and the user to position towards the camera), and based on these patterns classify the user state, attentive or distracted, during an activity.

## 1.3 Ethical Considerations

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

- Protocols: The school that 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, 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 two international conferences:

- Bruno Amaro, Vinícius Silva, Filomena Soares, and João S. Esteves – Building a behaviour architecture: an approach for promoting Human-Robot Interaction, Regional HELIX 2018, International Conference on Innovation, Engineering and Entrepreneurship, Guimarães, Portugal, 27-29 June, 2018;
- Bruno Amaro, Vinícius Silva, Filomena Soares, and João S. Esteves – An Approach to Behavioural Distraction Patterns Detection and Classification in a Human-Robot Interaction, SENSORDEVICES 2018, The Ninth International Conference on Sensor Device Technologies and Applications, Venice, Italy, 16-20 September, 2018.

## 1.5 Dissertation Structure

The dissertation is organized as follows:

- Chapter 2 provides background knowledge in the areas of determining distraction patterns and the importance of detecting these patterns in situations such as driving, as well as the use of social robots with children with ASD. This chapter starts with an overview of the concept of attention, and the difficulty that exists for a person to be attentive, especially in children with special needs. Then, studies using robots to interact with children with ASD are presented, including a discussion of their application to promote social interaction. Subsequently, in the last section, some studies focused on distraction patterns detection for automobile accident prevention are examined. 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 emphasis on Support Vector Machines (SVMs) and k-Nearest Neighbours (k-NNs), and how to evaluate the performance of each of these classifiers. Finally, it presents the methods generally used for validating systems that are capable of detecting distraction patterns and a methodology usually used for creating a database of facial characteristics.
- Chapter 4 focuses on describing the design and implementation of the system. The chapter starts by presenting the hardware and software used for implementing the system developed in the present work. Then, the software architecture and the User Interface are presented.
- Chapter 5 presents the results obtained from the present work. It starts by showing the Inertial Measurement Unit (IMU) results. Afterwards, the laboratorial results of each machine learning model used to detect the distraction patterns are presented, followed by the results from the children with ASD. Finally, the results using a database with adults and children with ASD and a discussion of this results are presented.
- Chapter 6 draws the conclusions from the work described in the dissertation and provides some insights regarding the future use of detecting distraction patterns in ASD intervention. The research presented in this dissertation highlights systems that are able to detect and classify patterns of distraction and how they can be applied in Human-Robot interaction.

# 2. LITERATURE REVIEW

**Summary**

In general, social behaviours imply being attentive towards something. These states are characterised by attention patterns, which are missing or disguised in some impairments. This section addresses this topic as well as some of the systems developed to detect and, possibly, to overcome this problem. Additionally, there are presented some robotic platforms that are used with children with ASD, in order to help them to integrate into society, as well as transmitting them new (social) knowledge.

## 2.1 Attention

The human behaviour has been the object of study in several areas, for example in the medical area where electronic products were developed in order to study and to help improving these behaviours. Robots have been increasingly used as mediators between a psychologist (or a doctor) and a patient because they are able to identify, measure and react to something, with the ability to be repetitive and objective. An example of a robotic application in this area is the possibility that they may offer to the children with ASD by helping them to better integrate into the society.

Among the human behaviours studied in the literature, the patterns of attention/distraction play a relevant role. Since attention is a cognitive process, it is linked to the learning and assimilation of new concepts, both at school and at work. Lack of attention may be seen as a problem in some activities such as receptionist or when one is reading a book, studying or driving a car (De Engenharia, João, & Pacheco, 2014).

Many of the attention problems are related to psychiatric disorders, as is the case of Attention Deficit Hyperactivity Disorder (ADHD).

According to José Pacheco (De Engenharia, João, & Pacheco, 2014), attention may be defined as a) “a concentration of the mind on a single object or thought, especially one preferentially selected from a complex one, with a view to limiting or clarifying receptivity by narrowing the range of stimuli.”, b) “a state of consciousness characterized by such concentration.”, or c) “a capacity to maintain selective or sustained concentration”. In short, attention may be seen as what keeps us focused on accomplishing a task, in order to complete it successfully. An example of a task is listening to a lecture and assimilating it. In this case, attention plays an important role as it is used to allocate human perceptual and cognitive abilities resources most effectively.

In order to pay attention, it is necessary to use specific networks of the brain, which is a skill developed over time that requires the user to be attentive. All this allows the user to collect correct information about the surrounding environment and to gather it.

Children tend to find it more difficult to filter out the distractions that are around them, especially children with special education needs, which makes their attention more difficult (Kasari et al., 2013).

Nowadays, there are systems that are used in social contexts to help children acquire new knowledge and integrate them more easily into society. However, some of these systems fail to capture the child's attention continuously. Taking this into account systems are being developed in order to monitor the user attentiveness. Generally these devices fall in two different categories: non wearable (e.g. Red Green Blue

(RGB) camera or a depth sensor) (E. Bekele, Crittendon, Swanson, Sarkar, & Warren, 2014; Fanelli, Gall, & Van Gool, 2011; Kondori, Yousefi, Li, Sonning, & Sonning, 2011; Liang, Reyes, & Lee, 2007), which is less invasive, and wearable devices which can be more accurate but more invasive (Lee & Chung, 2012). The wearable devices normally have an Inertial Measurement Unit (IMU) that can be used to track the user movements.

In most systems developed that aim to detect if the user is distracted, for example, in a human-robot interaction or when a person is driving, the patterns normally used are the head pose and the eye gaze (E. Bekele, Crittendon, Swanson, Sarkar, & Warren, 2014; Fanelli, Gall, & Van Gool, 2011; Kondori, Yousefi, Li, Sonning, & Sonning, 2011; Liang, Reyes, & Lee, 2007). For the detection of drowsiness, the main pattern is the blink frequency (Caffier, Erdmann, & Ullsperger, 2003; Danisman, Bilasco, Djeraba, & Ihaddadene, 2010).

## 2.2 Human–Computer Interaction – Detection of Patterns of Distraction

Being attentive is a key factor in our lives. Driving is an example where distractions may be fatal. There has been an investment in this application area. Some systems have already been developed in order to help the driver detecting whether he/she is sleepy or if he/she is distracted by something.

Arun Sahayadhas (Sahayadhas, Sundaraj, & Murugappan, 2012) performed a study on the various methods available to determine the state of drowsiness of a driver. In his article, he also discusses the various ways in which drowsiness can be manipulated in a simulated environment. The techniques used to detect these patterns may be subjective, vehicle-based, physiological, and behavioural, each having advantages and disadvantages. Although the accuracy rate of using physiological measures to detect drowsiness is high, they are highly intrusive. Despite this, it may be solved using non-contact electrodes. The authors conclude that it is worthwhile to merge physiological measures, such as the Electrocardiogram (ECG), with behavioural and vehicle-based measures in a drowsiness pattern, but nevertheless, it is important to consider the driving environment for optimal results (Sahayadhas, Sundaraj, & Murugappan, 2012).

Another project carried out by Yulan Liang (Liang, 2009) was developed with the aim of detecting patterns of distraction. Two hypotheses were tested: 1) when visual and cognitive distraction occurs in combination, the effects of combined distraction are dominated by visual distraction and with more severe impairments due to the occurrence of cognitive distraction, and 2) quantitative methods, especially data mining methods, may be used to construct models to detect visual, cognitive, and combined distractions.

The three aims were 1) developing a layered algorithm that integrates two data mining methods in order to improve the detection of cognitive distraction, 2) developing algorithms to estimate visual distraction, and to demonstrate the strong relationship between visual distraction and collision risk, and 3) developing an effective strategy to detect the combined distraction. After completing the project, the author concludes that visual distraction is the primary deficiency associated with secondary tasks and may represent a large proportion of the variance in the risk of collision. The author points out that it is increasingly important to design systems in the vehicle to limit the visual demands that are imposed on drivers when they use the system and offer an alternative mode of operation that requires small or no visual demand such as auditory systems and voice recognition that can replace manual operation in order to prevent off-road occurrences (Liang, 2009).

There are also other studies that in addition to detecting the patterns also use classifiers to categorize the individual as attentive or distracted. These studies typically use a computer vision system with machine learning algorithms (e.g. Support Vector Machines (SVMs) and Hidden Markov Models (HMMs), among others) (Liang et al., 2007). In one of these studies where SVM is used, a method of real-time detection of cognitive distraction and degraded driving performance has been developed (Liang et al., 2007). For the development of this project, the data was collected in a simulator experiment in which the participants interacted with an in-vehicle information system (IVIS) during the driving (Liang et al., 2007). The obtained results showed that the SVM models were able to detect the driver's distraction with an average accuracy of 81.1%. The conclusion of this study is that SVM provides a viable mean of detecting cognitive distraction in real-time, overcoming the traditional approach of logistic regression. Additionally, this study demonstrated that the eye movements and simple measurements of the driver's performance can be used to detect, in real time, the driver's distraction.

Another study presenting some early assessment results also uses an SVM method to monitor the distraction of a driver. The module is able to detect the driver's visual and cognitive workload by merging stereo vision and tracking data by running SVM-based machine classification methods. The results showed more than 80% of success in the detection of visual distraction and a success of 68 to 86% in the detection of cognitive distraction. In the authors' opinion the results were satisfactory. In the course of the study, a comparison is made with the HMM method. It is concluded that the advantage of HMM is that it takes into account the transitions from one state to another. However, SVM can adapt better to momentary changes. Bearing this in mind, it is concluded that HMM may be better at detecting drowsiness, as this is a process that passes through some states ("Hidden Markov models for time series classification," n.d.). SVM could obtain better results in detecting a distraction caused by the phone ringing, because it is a momentary distraction. In conclusion, the classification algorithms obtained satisfactory results, and it is still necessary to reduce the cost of the system (Kutila, Jokela, Markkula, & Rue, 2007).

These are some examples of studies conducted in the field of distraction patterns detection. Some companies are already starting to appear in this area as is the case of Tobii Tech ("Integrating Tobii Eye Tracking," 2015), which already has an eye tracking system to detect these patterns and thus helping the driver to have a safer driving.

Although driving is the core application area, the detection/characterization of attentiveness behaviours may be also relevant in the promotion of social behaviours. In fact, Autism Spectrum Disorder may be an important application field.

## 2.3 Human-Robot Interaction

Currently, one of the goals of assistive robotics is to help a user with special needs in his/her daily tasks. The robotic platforms used are designed to react to social behaviours, with the ability of being repetitive and objective. The robot can provide social support to motivate children, educate them socially, and also helping them to transfer the knowledge gained through interaction with other partners. However, most of these systems are controlled remotely, failing to automatically adapt to the situation.

The robot's physical appearance plays an important role in the interaction process with a person. For example, in autism intervention, the physical appearance of robots varies greatly from simple designs, e.g. four-wheeled mobile robots, to many levels of anthropomorphic forms, including humanoid (S. C. C. Costa, 2014), animal-like (Kim et al., 2013), and machine-like systems (François Michaud et al., 2005). Some examples of such systems are: IROMEC, KEEPON, KASPAR and ZECA (S. C. C. Costa, 2014; Ferrari et al., 2009; Kozima et al., 2009; Wainer, Robins, Amirabdollahian, & Dautenhahn, 2014), all of them focused on promoting social interaction and transmitting knowledge in a simple and interactive way.

IROMEC (Figure 1) is a robot designed to support a game for children with special education needs. This robot demonstrates that it is possible to learn and play at the same time, offering games in which the robot may have different behaviours and different scenarios, helping children in the integration and communication with other people.

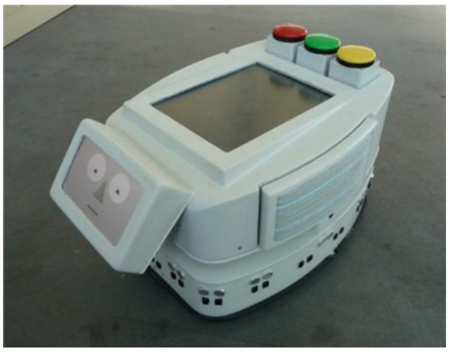


Figure 1 - Robot IROMEC (“IROMEC,” n.d.).

IROMEC goals are focused on harnessing the strengths of children, thus reducing their limitations. To help the child interact with the robotic platform, different sensors were developed with visual and tactile perception and spatial awareness (Ferrari, Robins, & Dautenhahn, 2009). In order to perform several tests, ten scenarios were developed and implemented, with the collaboration of specialists. Each scenario had a set of objectives, so that it would be possible to work in an iterative process with therapists and teachers. The robot works autonomously when a game scenario is selected, taking into account the needs of the children, being able to promote different learning abilities (Ferrari, Robins, & Dautenhahn, 2009).

KEEPON (Figure 2) is a small yellow robot designed to study and test the psychological models of social intelligence development when interacting with children through nonverbal interactions. Its simple appearance and behaviour is intended to help children, even those with developmental disorders, such as autism, to understand their attentive and emotional actions.

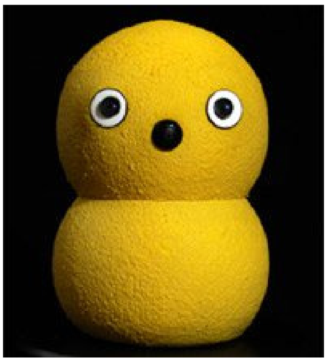


Figure 2 - Robot KEEPON (“KEEPON,” n.d.).

The intended results in the use of this robot are to confirm the effectiveness of a minimal design project in the exchange of emotions and visual contact between the child with ASD and the robot, so that it is possible to analyse how the interactions change according to the age of the child (Kozima, Michalowski, & Nakagawa, 2009).

In the experiments carried out, KEEPON alternated between making eye contact and looking at an object, moving its body, reacting whenever the child showed some behaviour that corresponded to a meaningful interaction.

After some tests the authors found that KEEPON simple appearance and predictable responses provided spontaneous and engaging interaction. As a result, children were able to expand their interaction in interpersonal communication, where KEEPON was the bridge between the child and the adults/colleagues.

The conclusion that is mentioned by the authors is that simple robot with minimal expressiveness can facilitate the natural exchanges of mental states of children with ASD.

KASPAR (Figure 3) is a humanoid robot with sensors on various parts of the body. This robot also discourages reprehensible interactions, using a variety of facial and body expressions to help breaking social isolation, stimulating reactions and helping parents and teachers to deal with children who suffer from ASD. The robot also makes didactic games, of imitation, among other activities.

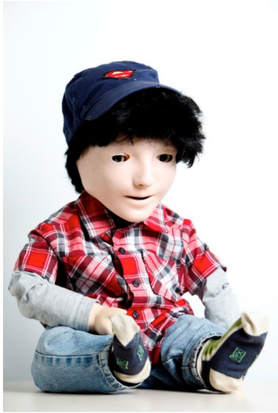


Figure 3 - Robot Kaspar (“Kaspar,” n.d.).

This robot has been used in several studies with children with ASD (S. Costa et al., 2015; Robins et al., 2005; Robins, Dautenhahn, & Dickerson, 2009a; Robins, Ferrari, & Dautenhahn, 2008; Wainer, Dautenhahn, Robins, & Amirabdollahian, 2010), and it has also been employed in other studies with typically developing children (Kose-Bagci, Dautenhahn, Syrdal, & Nehaniv, 2010; Kose-Bagci, Ferrari, Dautenhahn, Syrdal, & Nehaniv, 2009; L. J. Wood et al., 2013). One of the qualities that stands out in this robot is that it can adapt to different degrees of the spectrum, providing multimodal interaction depending on the child and with a variable difficulty, depending, for example, on the child's tactile interaction. In studies with children with ASD, the authors found that the alternating gaze between a dyadic collaborative game and the other players was significantly greater when playing with KASPAR. The children had more fun, seemed more invested in the game and collaborated better with their partners

(Wainer et al., 2010). In conclusion, the researchers showed that the use of KASPAR not only could demonstrate important competencies of social interaction, but also showed a level of direct, physical engagement. It was also verified that children appeared to generalize this behaviour at least to the experimenter.

The robot ZECA (Figure 4) has been used to help children with autism disorder. One of these projects aims to help children to recognize expressions so that they can interact and express themselves with their partners (S. C. C. Costa, 2014). In this study, a humanoid robot is used to provide socio-emotional development in children with ASD. In the project the chosen skills were eye reflexes, tactile interaction, verbal and non-verbal communication and recognition of emotions. The triadic interactions aimed to establish between the child with ASD and the experimenter with the robot as a common object of attention.

The main studies were:

- Checking if a humanoid robot can help children with ASD to learn appropriate physical social involvement, facilitating the ability to acquire knowledge about parts of the human body;
- Create a set of game scenarios using a humanoid robot as the main tool to develop socio-emotional skills in children with ASD;
- Evaluate the use of a humanoid robot, as a tool to teach recognition and labelling of emotions;
- Understand if and how a humanoid robot could promote triadic interactions between a child with ASD and another person.

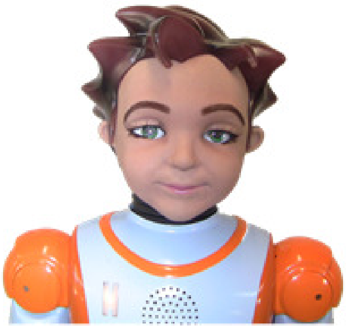


Figure 4 - Robot ZECA (“HomePage - Robótica Autismo,” n.d.).

In each of these studies, it was possible to infer that the use of a robotic platform such as ZECA has good results in helping children to develop new skills and interact socially.

More recently, there has been a concern in developing more adaptive approaches to interact with children with ASD. These recent approaches usually use wearable and non-wearable technologies in order to measure the children affective states.

The work developed by (Mazzei et al., 2011) consisted in controlling the robot reactions and responses, by using a combination of hardware, wearable devices, and software algorithms to measure the affective states (e.g. eye gaze attention, facial expressions, vital signs, skin temperature, and skin conductance signals) of children with ASD. The wearable devices that the authors used were a sensorized t-shirt to acquire the subject's physiological signals (ECG and respiration rate), wireless electrode bands to collect the user's skin temperature and Electro Dermal Activity (EDA) and the HATCAM, a system composed of a hat with markers and a grid of cameras to estimate the user's gaze. The developed system, FACET, includes a multisensory room in which a psychologist drives a stepwise protocol involving the android FACE and the autistic subject. This interaction between the robot and the subject is tailored by the therapist. A preliminary test was conducted with six male subjects: four individuals with ASD aged between 15 and 22 years old and two typically developing individuals aged between 15 and 17 years old. By analysing the results, the authors conclude that the subjects were calm during the activity and responded well to the robot. Additionally, the results confirmed that the system can be used as an innovative tool during the intervention sessions with subjects with ASD.

Bekele and colleagues (E. Bekele et al., 2014; E. T. Bekele et al., 2013) developed and later evaluated a humanoid robotic system capable of intelligently managing joint attention prompts and adaptively respond based on gaze and attention measurements. The system is composed of a humanoid robot with augmented vision by using a network of cameras for real-time head tracking using a distributed architecture. In order to track the child's head motion, a hat with markers was used. Thus, based on the cues from the child's head motion, the robot adapts its behaviour to generate prompts and reinforcements. A pilot usability study was conducted with six children with ASD. The results allowed to conclude that the children directed their gaze towards the robot when it prompted them with a question. The authors suggested that robotic systems, endowed with enhancements for successfully captivating the child's attention, might be capable to meaningfully enhance skills related to coordinated attention.

It is worth pointing out that, most of these systems are operated through a Wizard-of-Oz and in case the child is not attentive to the task, the therapist tries to motivate him/her directly or through robot prompts. So, it is a manual detection of attention/distraction patterns and appropriate reaction. An automatic adaptive robot behaviour is seldom implemented in the referred works.

In fact, the automatic detection of patterns of distraction may be included in robotic platforms, allowing an adaptive behaviour of the robot and improving the intervention process with the children.

# 3. METHODOLOGIES

**Summary**

This chapter presents the methodologies used in the present work. It starts by explaining the methods used for extracting the facial features. Then, it introduces two machine learning methods, the Support Vector Machine and the k-Nearest Neighbours, as well as how to assess the performance of the classifier. It presents the methods used for validating the system to recognize the human affective state and finally the methodology used for creating the database.

## 3.1 Facial Features Extraction

Facial expressions are innate in humans, through them it is possible to perceive how a person feels in different social situations. They can transmit emotions, opinions and clues about cognitive states. Nowadays, several psychological studies have been carried out with the purpose of decoding the information contained in a facial expression. For example, the system developed by Ekman and Friesen (Ekman & Friesen, 1978), the Facial Action Coding System (FACS), allows surveys to analyse and classify facial expressions in a standardized framework. Ekman also proposed the six basic emotions (Ekman & Friesen, 1978), also considered the six universal emotions – happiness, sadness, anger, surprise, fear, and disgust. But it should be noted that in systems where facial features need to be extracted, the type of feature to be extracted and the corresponding methods are critical to the overall performance.

The FACS system, developed by Ekman and Friesen (Ekman & Friesen, 1978), allows researchers to analyse, and classify facial expressions in a standardized framework.

### 3.1.1 Action Units

The FACS associates the action of muscles with changes in facial appearance. The basic metric of the FACS system are the Action Units (AU) which are actions, contraction or relaxation, performed by a muscle or a group of muscles. Of the main AU, 12 are for the upper face and 18 are for the lower face. AU 1 to 7 refer to eyebrows, forehead or eyelids (Ekman & Friesen, 1978; “FACS (Facial Action Coding System),” n.d.).

In this work for the detection of distraction patterns, the most relevant AU, because they are the ones that are most related for the detection of distraction patterns used in this work, were:

- 45 (blink), to calculate the blink frequency;
- From 51 to 58 (different head movements) for the calculation of the head pose;
- From 61 to 64 (different eyes movements) for the calculation of the eye gaze.

Table 1 shows the list of Action Units used in the present work (with underlying facial muscles). In Appendix B, it is available a more complete table of the Action Units.

Table 1 - List of Action Units (Ekman & Friesen, 1978; "FACS (Facial Action Coding System)," n.d.)

| AU# | FACS Name |
|---|---|
| **45** | Blink |
| **51** | Head turn left |
| **52** | Head turn right |
| **53** | Head up |
| **54** | Head down |
| **55** | Head tilt left |
| **56** | Head tilt right |
| **57** | Head forward |
| **58** | Head back |
| **61** | Eyes turn left |
| **62** | Eyes turn right |
| **63** | Eyes up |
| **64** | Eyes down |

### 3.1.2 Methods

For the feature (head pose, eye gaze, blink frequency, and the user to position towards the camera) extraction an algorithm based on the OpenFace (Baltrusaitis, Mahmoud, & Robinson, 2015; Baltrušaitis, Robinson, & Morency, n.d.-a, n.d.-b; E. Wood et al., 2015) library is used to track the face and eyes of the user, as well as the user's Action Units.

The OpenFace is an open source library that makes available a collection of facial landmarks (a total of 68 facial points) and AU based on the FACS which are used for the tracking of the face and the eyes, as well as other useful functions.

Additionally, OpenCV is also used due to its suitability and applicability in computer vision solutions.

To train and test the machine learning models, an algorithm based on Accord ("Accord.NET Machine Learning Framework," n.d.) was used. The Accord .NET Framework is a .NET machine learning framework combined with audio and image processing libraries developed in C #. It is a complete and useful tool for the creation of production level computer vision, computer hearing, signal processing and statistical applications.

## 3.2 Classification Methods

When in a given situation, it is necessary to know whether or not an object belongs to a set, or to which set it belongs, there is a classification problem. To solve such problems, machine learning algorithms are generally used. Using a machine learning algorithm, the program, instead of following the instructions, becomes more autonomous because, based on a model formed from sample inputs, it can predict or make decisions based on expressed data as outputs. Another advantage is that these algorithms are able to work with multidimensional data, with the benefit of easily incorporating new available data to improve prediction performance (Fowler, 2000).

Machine learning algorithms have two main scenarios where they can be applied, supervised learning in which the desired results are known and given to the learning algorithm for training and unsupervised learning, where the desired results are unknown, leaving the learning algorithm to find the structure in the input data.

The present work uses two supervised machine learning algorithms, the Support Vector Machine, SVM and k nearest neighbours, k-NN, both to classify the attention status of the user in an activity.

### 3.2.1 Support Vector Machines – SVM

The supervised learning method called Support Vector Machines was initially introduced by Vapnik (Burges, 1998); this method is capable of analysing data for classification and regression analysis. SVM is usually implemented in binary classification, in which the objective is to separate two classes, as shown in Figure 5. In this type of problem the classification is performed in the feature space through the construction of a linear separating hyperplane (Burges, 1998). At first sight, the separation of these classes can be done by any line dividing the regions containing only the "plus" and only the "minus". But intuitively it is noticed that the dark line seems to provide a decision boundary better than the dashed line because it seems to have a greater margin of safety. Therefore, the goal of SVM is to find an Optimal Separating Hyperplane (OSH) that divides all data points from one class to the other. This is achieved by finding the largest margin between two classes, which is OSH. To get the maximum width of the margin, an optimization problem is calculated. This problem is controlled by a parameter C, a trade-off between the maximum width of the margin and the minimum classification error (Burges, 1998). The C parameter allows a trade-off between training set accuracy and expected generalization capability. A large value of C (Figure 6, left) will give rise to a hyperplane that commits fewer errors in the training data, but will have a smaller margin of safety and therefore less expected generalization. On the other hand, a small value of C (Figure 6, right) will allow more errors in the training data, but the margin of safety will be wider.

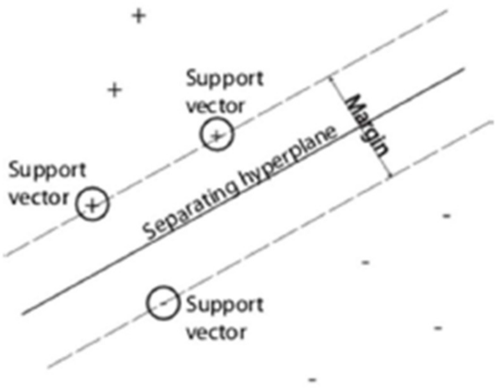


Figure 5 - Optimal Separating Hyperplane ("Statistics and Machine Learning Toolbox Documentation," n.d.).

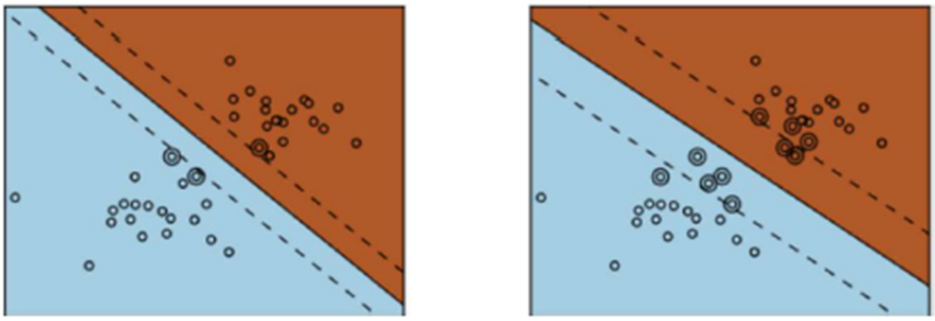


Figure 6 - The influence of the parameter C: on the left for a higher value of C, on the right for a lower value of C ("SVM Margins Example," n.d.).

The points closest to the hyperplane are called Support Vectors (SV), on which the weight vector depends. However, some binary classification problems do not have a simple hyperplane as a useful separation criterion. In order to solve these problems, called nonlinear classification, the SVM employs kernel methods to map data to a space of larger dimensional characteristics. The commonly used SVM kernel functions that define the nature of the decision surface are: linear (equation 1), polynomial (equation 2), radial basis function (RBF) (equation 3) and sigmoid equation (4) where $x$ stands for the data matrix (observations vs. characteristics).

$$linear: k(x_i\ , x_j) = x_i^T x_j \qquad \text{(eq.1)}$$

$$polynomial: k(x_i\ , x_j) = (\gamma x_i^T x_j + r)^d\ , \gamma > 0 \qquad \text{(eq.2)}$$

$$radial\ basis\ function\ (RBF): k(x_i\ , x_j) = \exp\ \ \gamma\|x_i - x_j\|^2, \gamma > 0 \qquad \text{(eq.3)}$$

$$sigmoid: k(x_i\ , x_j) = \tan(\gamma x_i^T x_j + r) \qquad \text{(eq.4)}$$

SVM shows high classification accuracy even when the size of the database to train the model is small, thus making the model suitable for dynamic and interactive approach to face the recognition of distraction patterns or for example for expression recognition (Michel & Kaliouby, n.d.).

The accuracy of the SVM classification depends on the value chosen for the C parameter and also for the gamma ($\gamma$) value, when the RBF kernel is used. In order to obtain the best values of C and gamma ($\gamma$), which minimize classification error, a grid search (exhaustive search through optimization problems) is often used, but it is also recommended to combine the grid search with cross-validation (Fowler, 2000).

The gamma ($\Upsilon$) parameter of RBF kernel defines the extent of the influence of a single training example. For large gamma ($\Upsilon$) values the radius of the area of influence of the support vectors only includes the support vector itself. On the other hand, small values of gamma ($\Upsilon$) means that the region of influence of any selected support vector would include the whole training set and the resulting model would present a linear behaviour ("RBF SVM parameters," n.d.).

There are two stages involved in any binary classification, the training stage and the testing stage. In the training stage, a set of input data is provided to the SVM and it sorts them into a set of different possible classes and, through optimization, it finds the OSH. It then creates a template that assigns data to each class. In the test stage, data that was not used in the training is used and the SVM refers to the results of the calculated classification in relation to the known class labels. Finally, the SVM functions is a non-probabilistic linear classifier because it predicts the set of classes based on an optimization problem.

The main advantage of SVM is the convergence to a global optimum, avoiding local minima and excessive adjustment in the training process (Wu, Wang, & Liu, 2007). Thus, the SVM has the ability to minimize structural and empirical risks, even with a limited set of training data, the SVM has the capacity to minimize structural and empirical risks, leading to a better generalization to the new classification of data, producing results stable and reproducible (Michel & Kaliouby, n.d.). A disadvantage of the SVM classifier lies in the dependence of performance on internal learning parameters (eg SVM regularization parameter), which may be difficult to interpret (Fowler, 2000).

### 3.2.2 k-Nearest Neighbours – k-NN

The K-nearest neighbour algorithm (k-NN) is a type of supervised machine learning algorithm. This is a relatively simple algorithm to understand and implement. The k-NN consists of assigning a feature vector to a class, according to its closest neighbours. In a data set $S$, the nearest neighbour of a data object $q$ is the data object $S_i$, which minimizes $d(q, S_i)$. The function $d$ represents a measure of distance defined for the object in question (Sammut & Webb, n.d.). This learning algorithm is lazy because it does not have a specialized training phase. Instead, it uses all the data for training when sorting a new data point or instance. k-NN is a nonparametric learning algorithm, which means that it assumes nothing about the underlying data ("K-Nearest Neighbors Algorithm in Python," n.d.). The k-NN aims to assign the dominant class to an object among its closest k neighbours within the training set. If the selected k value is high enough and there are enough training samples, the k-NN can approximate any function, allowing this classifier to generate non-linear decision limits (Lotte, Congedo, Lécuyer, Lamarche, & Arnaldi, 2007). Like any other machine learning algorithm this method also has advantages and disadvantages. Some of the advantages is that it is easy to implement, as said before, it is a lazy learning algorithm and therefore it does not require training before making real-time predictions. This makes the k-NN algorithm much faster than other algorithms that require training, for example SVM, linear regression. New data can be added without problems, since the algorithm does not require training before making predictions, and lastly there are only two parameters required to implement k-NN, that is, the K value and the distance function (eg Euclidean or Manhattan). As disadvantages, k-NN algorithm does not work well with high dimensional data because it becomes difficult for the algorithm to calculate the distance in each dimension. The k-NN algorithm has a high predictive cost for large data sets. This is because, in large datasets, the cost of calculating the distance between the new point and each existing point becomes larger. Finally, the k-NN algorithm does not work well with categorical features, because it is difficult to find the distance between dimensions with categorical features.

## 3.3 Classifier Performance

Generally, validation methods are used to evaluate the performance of a machine learning algorithm, with the intention of evaluating their generalization, especially when data sets are limited, and test their ability to classify new instances. Training a model and testing it with the same data is a methodological error: a model would achieve a perfect score but could not predict new data. This situation is called overfitting. To avoid this, it is a common practice to divide the available data into two sets: a training set and a set of tests. An automatic and more common method of preventing overfitting and validating the model is the k-Fold cross-validation (CV) method ("Cross-validation: evaluating estimator performance," n.d.). The CV k-Fold method divides the total data set into k subsets and a subset is retained as the validation data to test 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 once used as the validation data ("Cross-Validation - MATLAB &amp; Simulink," n.d.). The k fold results can be calculated (or combined) to produce a single estimate. The main advantage of this method is that all observations are used for training and validation, and each observation is used for validation once. The CV contributes to the generalization of the classifier and also avoids overfitting.

### 3.3.1 Performance Metrics

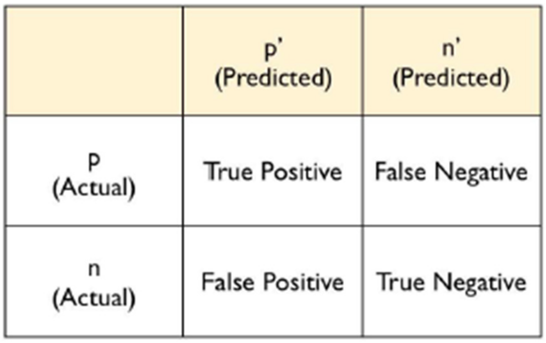
There are many ways to quantify/measure the classifier performance. The most popular metrics are: accuracy, the confusion matrix, sensitivity, specificity, and the Area under the Curve (AUC) metric.

Accuracy (equation 5 where TP, FP TN, and FN correspond to true positive, false positive, true negative, and false positive, respectively) is the most common and it simply measures how often the classifier makes the correct prediction, (Zheng, n.d.). It is the ratio between the number of correct predictions (TP and TN) and the total number of predictions (TP, TN, FP, and FN).

$$accuracy\ (\%) = \frac{TN+TP}{TP+TN+FP+FN} \times 100\% \qquad \text{(eq.5)}$$

However, one of the problems that does not make the evaluated accuracy of a classifier representative of the actual performance of the classifier is if there is an unbalanced data set (where the number of samples in each class is significantly different). The accuracy will give a distorted image, because the class with more examples will dominate the statistic. Therefore, it is advisable to observe accuracy by class, both the mean and the individual precision numbers by class (Zheng, n.d.).

Accuracy is a simple metric, but does not distinguish between classes, which a confusion matrix does. A confusion matrix is a specific table layout, Figure 7, which allows the visualization of the performance of a learning algorithm, usually supervised (Zheng, n.d.). This matrix shows in more detail an unfolding of the correct and incorrect classification for each class. The rows of the matrix correspond to the truth labels of the terrain (the real class) and the columns represent the prediction (predicted class).



| | p'<br>(Predicted) | n'<br>(Predicted) |
|---|---|---|
| p<br>(Actual) | True Positive | False Negative |
| n<br>(Actual) | False Positive | True Negative |

Figure 7 - Confusion matrix, where each row correspond to the actual class and each column to the predict class (“Confusion metrics,” n.d.).

From the confusion matrix it is possible to acquire the following measures: TP (true positive), TN (true negative), FP (false positive) and FN (false negative). From these measurements, it is possible to obtain two more metrics - sensitivity and specificity. The sensitivity, equation 6, measures the proportion of real positives that are correctly identified as such (Zheng, n.d.).

$$sensitivity\ (\%) = \frac{TP}{TP+FN} \times 100\% \qquad \text{(eq.6)}$$

Specificity, equation 7, measures the proportion of true negatives which are correctly identified as such (Zheng, n.d.).

$$specificity(\%) = \frac{TN}{TN+FP} \times 100\% \qquad \text{(eq.7)}$$

Another metric to evaluate the classifier performance is the area under the curve (AUC). This metric is obtained through the integration by the trapezoidal method based on the Receiver Operating Characteristic curve (Figure 8), or ROC curve (Zheng, n.d.). From this curve it is possible to visualize the trade-off between the TP rate and the FP rate, that is, it shows how many correct positive classifications can be obtained as more and more false positives are allowed. The AUC value of a perfect, error-free classifier would be 1. The AUC represents the mean sensitivity in all possible specificities.

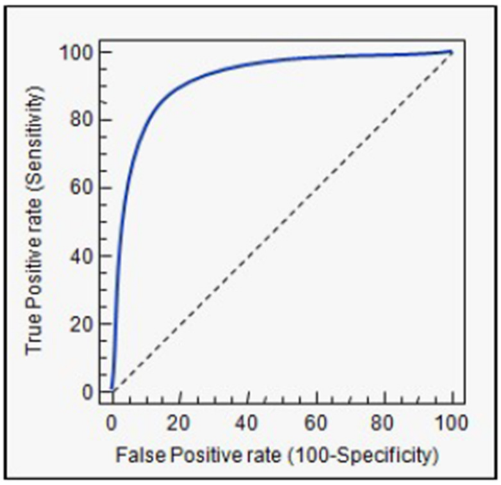


Figure 8 - ROC curve - trade-off between the TP rate and FP rate (“ROC,” n.d.).

Matthews Correlation Coefficient (MCC) is a correlation coefficient between real classes and prediction, which takes into account true positives (TP), true negatives (TN), false positives (FP) and false negatives (FN), and it is usually considered as a balanced measure, which can be used even if classes are unbalanced (with different sizes) (Jurman & Furlanello, 2010). It returns a value between -1 and +1, where +1 represents a perfect forecast, 0 is not better than a random forecast, and -1 indicates total mismatch between forecast and observation. This metric can summarize/describe a matrix of confusion by a single number. The MCC can be calculated directly from the confusion matrix using equation 8, for binary classification.

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP+FP)(TP+FN)(TN+FP)(TN+FN)}} \qquad \text{(eq.8)}$$

### 3.3.2 Permutation Importance

After training the model and assembling the performance of a classifier one of the doubts that may arise is: which features have the greatest impact on the predictions? There are several ways to know this, one of which is the use of the permutation importance. Compared to other approaches, the permutation importance is faster to compute and easier to understand (Altmann, Toloşi, Sander, & Lengauer, 2010). The permutation importance is calculated after a model has been fitted. Therefore, it does not change the model or change the predictions.

Having a certain database, what would happen if the column of validation data was randomly scrambled, leaving the target and all other columns in place, how would that affect the accuracy of the predictions in these now shuffled data?

Randomly reordering a single column should cause less accurate predictions, since the resulting data no longer corresponds to anything observed in the real world. The accuracy of the model suffers especially if a column is scrambled on which the model relied heavily on predictions.

The process is as follows: after training a model select a single column and shuffle the values, then make predictions using the resulting dataset. Taking into account these predictions, compute the accuracy in order to see how much the loss function has suffered from the shuffle. This change in the accuracy measures the importance of the variable that was scrambled. Then, reset the data to the original order. Finally, repaeat the process for the next column of the dataset until the importance of each column is calculated (Altmann, Toloşi, Sander, & Lengauer, 2010).

In this work, this method is used to assess the importance of each standard for the classification of the user state, as well as to know if any of the patterns are harming the model instead of helping.

## 3.4 System Validation - Recognition of distraction patterns

In the literature, some standard methods can be found for validating systems that can recognize Human distraction behaviours. The most common procedures to validate this type of methods are detailed in the following.

Generally, two types of evaluation are used, an offline evaluation and a real-time evaluation. In offline evaluation, the machine learning model created is trained and tested and the performance is usually evaluated using the performance metrics described in section 3.3. With this evaluation, besides evaluating a model of machine learning, it is also possible to compare two models used for the same purpose in a simpler way (comparing to the real-time evaluation) and thus to see which one allows better results.

For real-time evaluation, the system in a first stage is trained, then, a set of participants, who have not participated in the database creation process, are recruited. The usual experimental setup consists of each participant pretending to be attentive to an activity or distracted from it. This data is then used to quantify the performance, usually in terms of accuracy per class and average accuracy. Real-time evaluation usually focuses more on assessing the system's real-time performance.

## 3.5 Database Creation

Generally, the methodology used to create the database consists first on recruiting a group of participants. So, the usual experimental setup to extract the facial features (head pose, eye gaze, blinks frequency, and the user to position towards the camera) is for each participant to pretend to be attentive and distracted when requested by the researcher in front of a sensor.

The acquired data is saved to a file. Since SVM algorithms are not invariant to scale (Burges, 1998), it is recommended to scale the extracted data to a normalized range by applying equation 9, where *N* is the value of the normalized facial trait, *W* is the value of the facial trait for normalize, $min_v$ is the minimum value of the normalized set, $max_v$ is the maximum value of the normalized set and *[A, B]* is the interval for the value *(W)* after normalization.

$$N = \frac{W - min_v}{max_v - min_v}(B - A) + A \qquad \text{(eq.9)}$$

# 4. DEVELOPMENT

**Summary**

This chapter describes the adopted materials and general procedures used in the present work. It starts by presenting a global overview of the system. Then, it details the hardware used, followed by an explanation of how the patterns of distraction are obtained and how the models used for classification are created. Finally, the software architecture and the Graphical User Interface are discussed.

## 4.1 System Overview

The BPAS, Behaviour Patterns Analysis System, implemented in this work (Figure 9) consists of a RGB camera to detect patterns of user distraction in order to, posteriorly, adapt the behaviour of the robot during the activity. In addition to the RGB camera, the experimental setup uses a computer (where data is processed) and the ZECA humanoid robot (the mediator partner in the activity).

The design of the emotion imitation activity, where these patterns are going to be detected, consists in ZECA displaying a facial expression and asking the child to imitate it. Then, the robot automatically verifies if the answer is correct and responds accordingly. Meanwhile, if the system detects any distraction pattern in the child, the robot adapts its behaviour, encouraging the child to return and participate in the activity. This adaptation of the robot behaviour, after the automatic classification of the child engagement/distraction in the activity, will allow a more fluidic and active interaction.

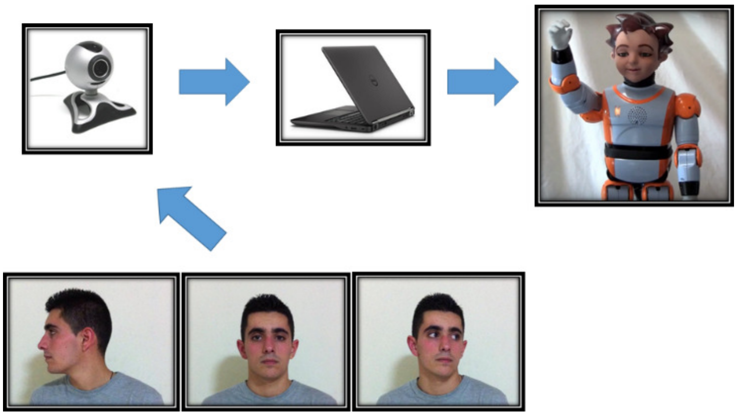


Figure 9 - Experimental setup: starting from the left, the RGB camera; in the centre, the computer; and on the right, the humanoid robot ZECA.

In addition to developing this non-invasive solution, a wearable solution based on an Inertial Measurement Unit (IMU) was designed to extract the head pose. This solution had the goal to make the system more robust in case of failure of the camera face detection.

The system with the inclusion of the IMU of 10DOF (placed in the child's hat) is shown in Figure 10.

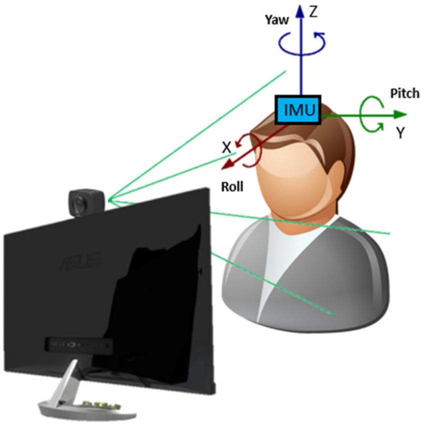


Figure 10 - Wearable system with an IMU

## 4.2 Hardware Description

In the following, the hardware used in the system is detailed.

### 4.2.1 Robotic Platform

The Zeno R50 RoboKind humanoid child-like robot ZECA (Figure 11) is a robotic platform that has 32 degrees of freedom: 4 are located in each arm, 6 in each leg, 11 in the head, and 1 in the waist. The robot is capable of expressing facial cues thanks to the servo motors mounted on its face and a special material, Frubber, which looks and feels like human skin, being the major feature that distinguishes Zeno R50 from other humanoid robots. Table 2 summarizes the principal characteristics of the Zeno R 50 robot.

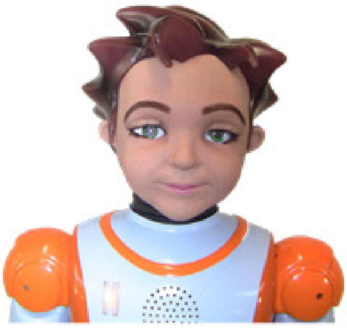


Figure 11 - Zeno R-50 (“HomePage - Robótica Autismo,” n.d.).

Robokind software includes an application programming interface (API) for fast integration of other components, distributed computing and shared control; this software is also capable of performing functions of animation and motion control. The robot's technical drawings can be found in Appendix A. From now on, the robot will be referred to as ZECA (Zeno Engaging Children with Autism).

Table 2 - Principal characteristics of the Zeno R-50 robot from Robokind (Corrêa & Da Silva, 2015).

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

### 4.2.2 RGB camera

The camera used in the present work is a RGB camera, more specifically the Microsoft VX-1000 camera (Figure 12). This device has a minimum resolution of 320 by 240 pixels and a maximum resolution of 640 by 480 pixels. The camera dimensions are 5.5 cm of width, 6.8 cm of height, and 5.3 cm of depth. For a better detection and feature extraction, the camera is placed on the robot chest, being at the same level as the user's face.

Figure 12 - Microsoft VX-1000 camera ("Camera," n.d.).

### 4.2.3 Wearable box

The IMU used in the proposed framework, REF, it is the fusion of MPU9255 and BMP180 (barometric pressure sensor), it has 10 Degrees of Freedom (DoF) – 3 axis gyroscope, 3 axis accelerometer, and 3 axis compass/magnetometer – and through the use of sensor fusion algorithms, it can output the Euler angles (pitch, roll, and yaw).

Comparing with MPU6050, the MPU9255 is lower power consumption, and more suitable for wearable devices.

The device presented in Figure 13 includes an IMU, a small development board (ESP32) that already has built-in Bluetooth and Wi-Fi communication, and a Li-Po battery.

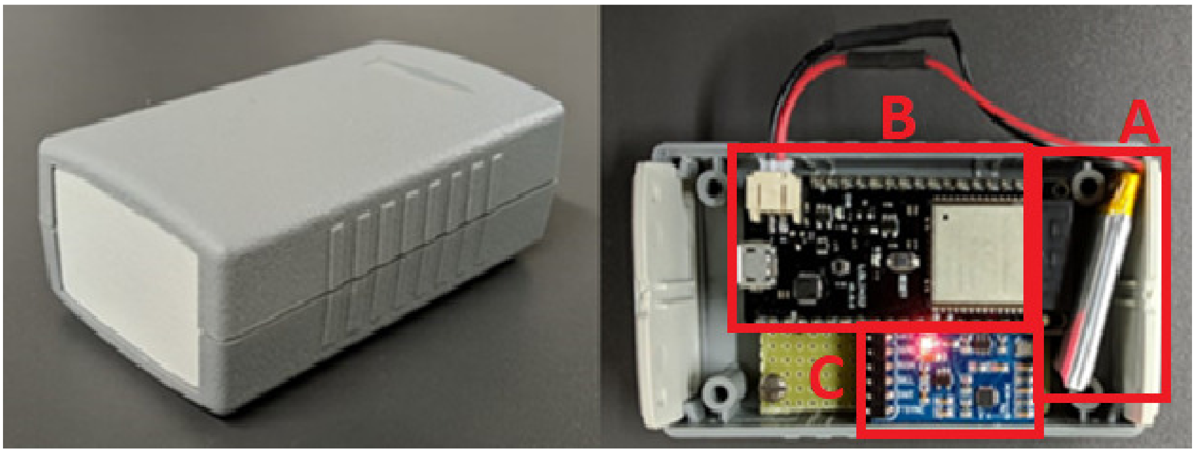


Figure 13 - Wearable device used for extracting the user's head motion. It includes a Li-Po battery (A), a small development board (ESP32) that already has built-in Bluetooth and Wi-Fi communication (B), and an IMU (C).

## 4.3 Obtaining the distraction patterns

Since the purpose of this work is to know whether the user is distracted, a literature review was conducted to determine which patterns best fit in this evaluation. The following patterns were identified: the eye gaze, the head orientation, the eyes blinking frequency, and the user to position towards the camera (this last is important to know if the user is at a safe distance from the robot due to its movements, and in case the user moves away from the robot, which is a distraction pattern, since it is no longer performing the activity). These patterns will be detected during an emotion imitation activity.

The user must be in front of the RGB camera, as shown in Figure 10. The effective range for the tracking of faces depends on the camera used and its quality; in the camera used in this work the range is between 15 cm and 260 cm.

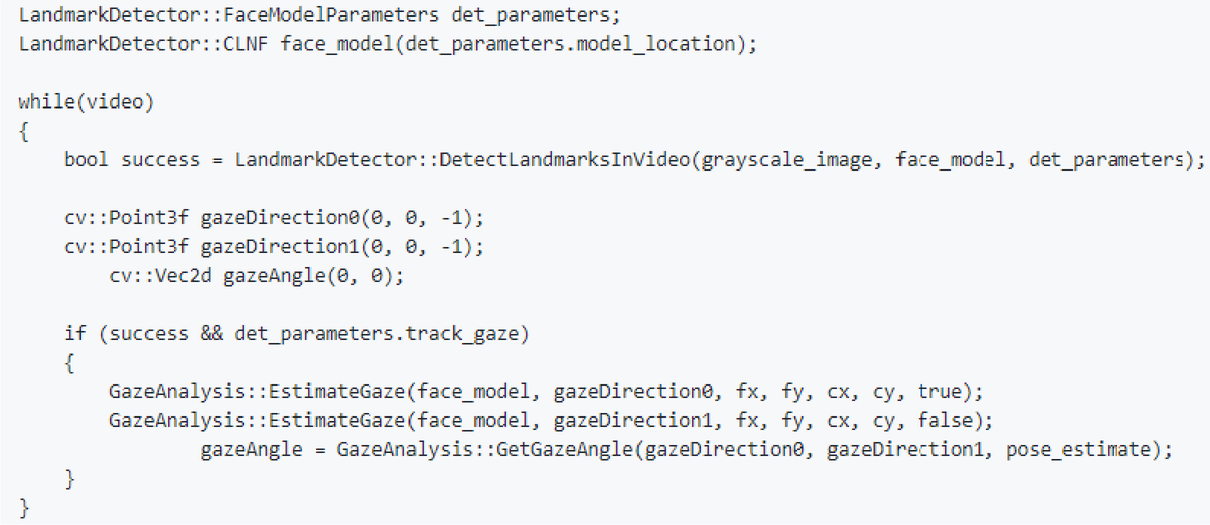
Using the OpenFace library it is possible to calculate the scores of some supported facial expressions, as well as to detect up to 68 facial landmarks.

Facial AUs obtained through the OpenFace (Baltrušaitis, Robinson, & Morency, n.d.) library are normalized to an intensity scale of 0 to 1, where 0 means that AU is not present and 1 means that AU is definitely present. To find the minimum and maximum values for each of these characteristics, a performer was asked to perform a wide range of extreme facial movements while sitting in front of the camera. In this way, the minimum and maximum values for each feature were found experimentally.

The head pose is stored in the following format (X, Y, Z, rot_x, roty_y, rot_z), translation is in millimetres with respect to camera centre (positize Z away from camera), rotation is in radians around X,Y,Z axes with the convention R = Rx * Ry * Rz, left-handed positive sign. To get the head pose this function is used: "cv::Vec6d head_pose = LandmarkDetector::GetPose(face_model, fx, fy, cx, cy);",

fx,fy,cx,cy are camera calibration parameters needed to infer the 3D position of the head.

To extract eye gaze (Figure 14), it is used the facial landmarks detected using a LandmarkDetector::CLNF model that contains eye region landmark detectors.



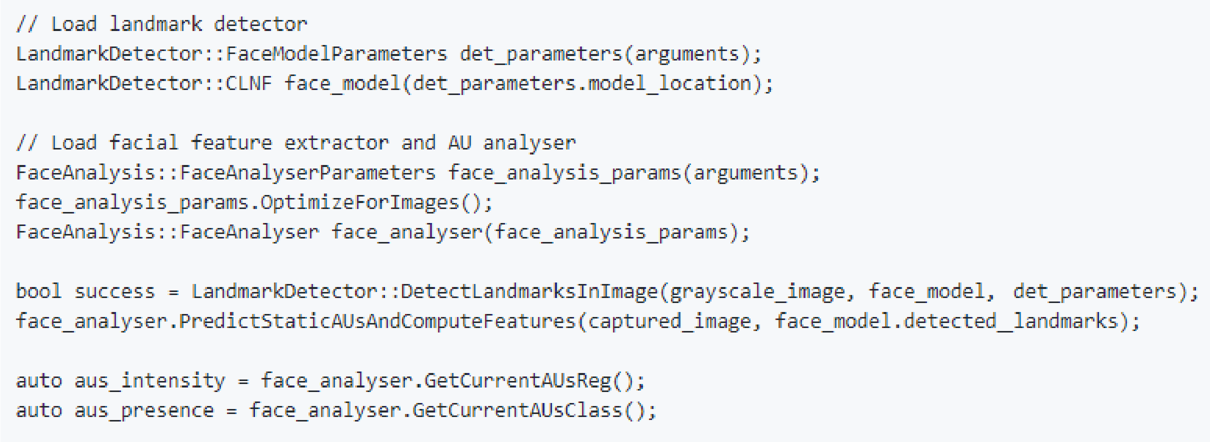
```
LandmarkDetector::FaceModelParameters det_parameters;
LandmarkDetector::CLNF face_model(det_parameters.model_location);

while(video)
{
    bool success = LandmarkDetector::DetectLandmarksInVideo(grayscale_image, face_model, det_parameters);

    cv::Point3f gazeDirection0(0, 0, -1);
    cv::Point3f gazeDirection1(0, 0, -1);
        cv::Vec2d gazeAngle(0, 0);

    if (success && det_parameters.track_gaze)
    {
        GazeAnalysis::EstimateGaze(face_model, gazeDirection0, fx, fy, cx, cy, true);
        GazeAnalysis::EstimateGaze(face_model, gazeDirection1, fx, fy, cx, cy, false);
                gazeAngle = GazeAnalysis::GetGazeAngle(gazeDirection0, gazeDirection1, pose_estimate);
    }
}
```

Figure 14 - Eye gaze extraction code (Baltrušaitis, Robinson, & Morency, n.d.).

The blinks frequency is calculated based on the AU corresponding to the blink of an eye, when it is activated, a variable is incremented.

For the detection of drowsiness, it is used the same AU, but when the eye blink takes more than 3 seconds, a warning is given in the interface.

The Facial Action Units can be extracted (Figure 15) in each image in a static manner or extracted from a video in a dynamic manner. The dynamic model is more accurate if enough video data is available for a person (roughly more than 300 frames that contain a number of non-expressive/neutral frames).



```
// Load landmark detector
LandmarkDetector::FaceModelParameters det_parameters(arguments);
LandmarkDetector::CLNF face_model(det_parameters.model_location);

// Load facial feature extractor and AU analyser
FaceAnalysis::FaceAnalyserParameters face_analysis_params(arguments);
face_analysis_params.OptimizeForImages();
FaceAnalysis::FaceAnalyser face_analyser(face_analysis_params);

bool success = LandmarkDetector::DetectLandmarksInImage(grayscale_image, face_model, det_parameters);
face_analyser.PredictStaticAUsAndComputeFeatures(captured_image, face_model.detected_landmarks);

auto aus_intensity = face_analyser.GetCurrentAUsReg();
auto aus_presence = face_analyser.GetCurrentAUsClass();
```

Figure 15 - Facial Action Units extraction code (Baltrušaitis, Robinson, & Morency, n.d.).

The variables aus_intensity and aus_presence will contain a std::vector<std::pair<string, double>>, each vector contains the AU name and presence or intensity value.

For the calculation of the user to position towards the camera, it is used OpenFace and OpenCV functions.

For the calculation of the head pose using the IMU, an algorithm to extract and calculate the yaw, pith and roll is developed in the ESP32 board, and the data is sending to the interface via Bluetooth.

## 4.4 Classification Models

The BPAS is able to detect when a person is distracted, based on 4 patterns (head pose, eye gaze, blinks frequency and the user to position towards the camera), through facial features in real time. In this system each of the data extracted from the AU and head movement is normalized, on a scale of 0 to 1.
Subsequently, two models, the SVM with the Gaussian kernel and the k-NN are trained using Accord Machine Learning C # software.

The three databases used in the training process of the implemented models were constructed, applying the methodology referenced in section 3.5, using the features acquired through OpenFace (as referred in section 4.3).

It was decided to use three databases in such a way as to be able to have: a database with only adult data (1000 samples), another database with children with ASD (3000 samples), and finally another database with all the data (adults and children with ASD), in order to compare the results of each of them.

The participants considered for the construction of the aforementioned databases used were:

- 4 children with ASD with 6 to 9 years of age. The tests were conducted in a controlled environment during an activity, with the therapist always present.
- 10 adults with 18 to 50 years of age. The tests were performed in a laboratory environment.

The acquired data, facial AUs and features were saved in a file.

Finally, after the training process of both models, the result of the user state is placed in the interface, being updated in real time. It is worth mentioning that although two models are used, they have not been trained and tested simultaneously, only one of them is used at a time.

## 4.5 BPAS Architecture

This section presents the main procedures of the software produced included in the BPAS.

### 4.5.1 Affective state detection and classification

After the extraction of the chosen patterns (head pose, eye gaze, blinks frequency, and the user to position towards the camera), the classification is done using an algorithm based on machine learning, thus classifying the user as attentive or distracted.

Figure 16 presents the block diagram of the overall system considering the detection of distraction patterns, as well as in parallel the classification of the emotional state (the classification model of the emotional state was not the focus of the present work) of the chid during an activity.

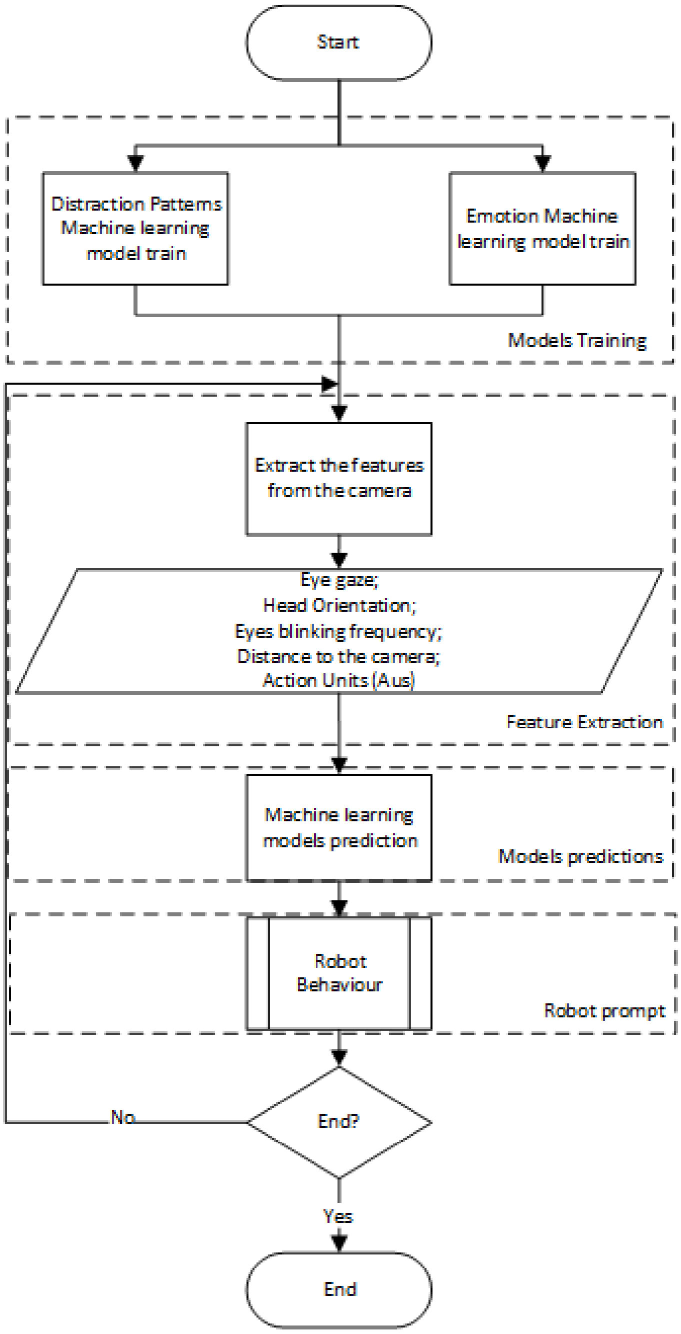


Figure 16 - System flowchart highlighting the defined modules for detecting distraction patterns and emotional states: Models training, Feature extraction, Models predictions, and Robot prompt.

The first module, “Models Training”, consists in training each machine learning model with a previously defined database.

Three other modules are defined: “Feature Extraction”, “Models predictions”, and “Robot prompt”.

For the feature extraction, an algorithm based on the OpenFace library is used to extract the user data. Generally, some researchers use machine learning techniques in order to infer the user emotional states (Silva et al., 2016). Then, by using machine learning methods, such as Support Vector Machines (SVM) or k-Nearest Neighbours (k-NN), the distraction patterns and the user facial expression will be recognized. Once the patterns corresponding to distraction and emotional states are detected, the robot should trigger the corresponding action (robot prompt module) to acknowledge the emotion, to give reinforcement and, if necessary, to capture the user's attention again.

### 4.5.2 Robot Behaviour

The general procedure during an activity is:

1) ZECA greets the researcher and the user;
2) ZECA asks which activity shall be played;
3) The selected activity starts and continues until the experimenter decides to end it.

In the emotions recognition activity, the robot prompts a different behaviour accordingly to the results, the child attentiveness, and response. Thus, according to the classifier output, four conditions may occur:

- the user is attentive and answer to the robot prompt;
- the user is attentive but does not answer to the robot prompt;
- the user is distracted and does not answer to the robot prompt;
- the user is distracted but answer to the robot prompt.

In order to adapt the robot behaviour accordingly to the four different conditions mentioned, a state machine model is proposed.

A State Machine is a simple model to track the events triggered by external inputs. This is done by assigning intermediate states to decide what happens when a specific input comes in and which event is triggered. In this case, the events are being attentive or not being attentive and answering or not answering the prompted question.

The working of the State Machine (Figure 17) to be used in the present work is described in the following points:

- The time between a stage and an inter-stage is 5s and the time between stages is 10s;

- The time without a response until the robot takes an action is 10s;
- The user unanswered timeout until the robot moves to the next stage of the activity is 30s.

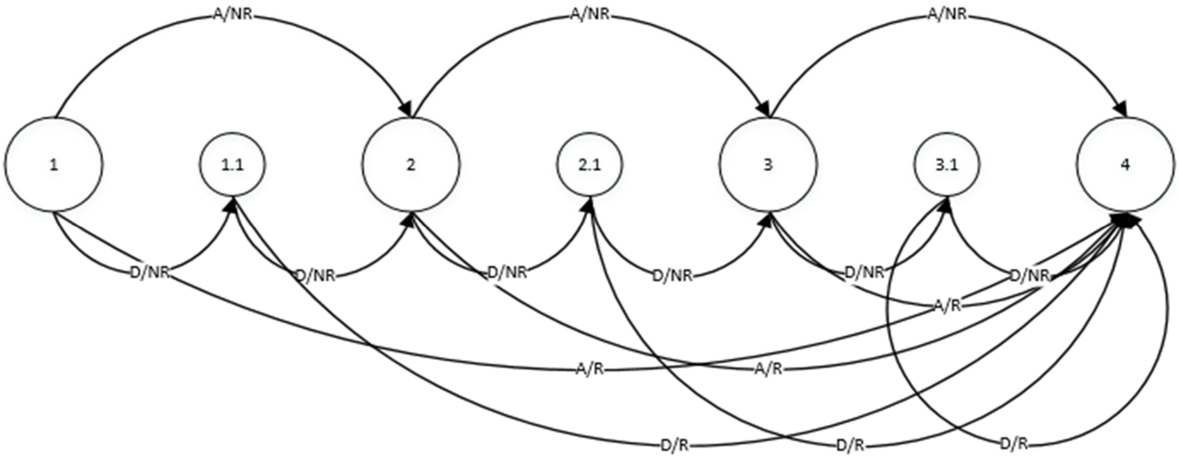


Figure 17 - State Machine with the workflow for the behaviour architecture of the robot (in this State Machine, A=Attentive, R=Responds, D=Distracted, and NR=does Not Respond).

The explanation for the above diagram is given below:

- The inputs of the State Machine are whether or not the user is attentive and whether or not the user answers the question. The output is accordingly to the correctness of the user's response and the detected distraction behaviour. The bubbles represent the states.
- The initial state is 1. This state corresponds to the moment when the robot asks the question and goes directly to stage 4, if the user is attentive and responds to the question, or goes through all the different stages, if the user does not respond, or if he/she responds in one of the following stages, he then moves on to stage 4 as well.
- In the stages 2 and 3, if the user is attentive but does not respond, the robot repeats the question, motivating the user to answer.
- The sub-stages between each of these steps are considered for the cases when the user is inattentive and does not respond, causing the robot to try to regain his/her attention. These also serve to detect false positives that could happen if the user is classified as inattentive, but answers the question, which causes the robot to go directly to stage 4.
- In a stage 4, the robot verifies if the user response was correct, taking action accordingly. Then, it moves on to the next phase/question of the activity.
- If the user does not respond in 30s and he/she is considered attentive, it means that the user does not know the answer, despite having been attentive to the robot.

According to the correctness of the user's answer, ZECA prompts a positive or negative reinforcement through sounds, gestures, and verbal communication (S. C. C. Costa, 2014).

After classification, it is necessary to take an action, that is, if the robot classified the user as attentive, then it will continue the activity. Then, ZECA revises the patterns, so as to always know if the user is attentive. If it has previously considered the user inattentive, then it will trigger an action in order to capture the user's attention again.

At the end of the cycle, it is always checked if the activity is complete; if it is accomplished, the robot does not need to revise the defaults; if it is not, then the robot will have to continue to analyse the user defaults.

Figure 18 depicts the general procedure that happens during an activity.

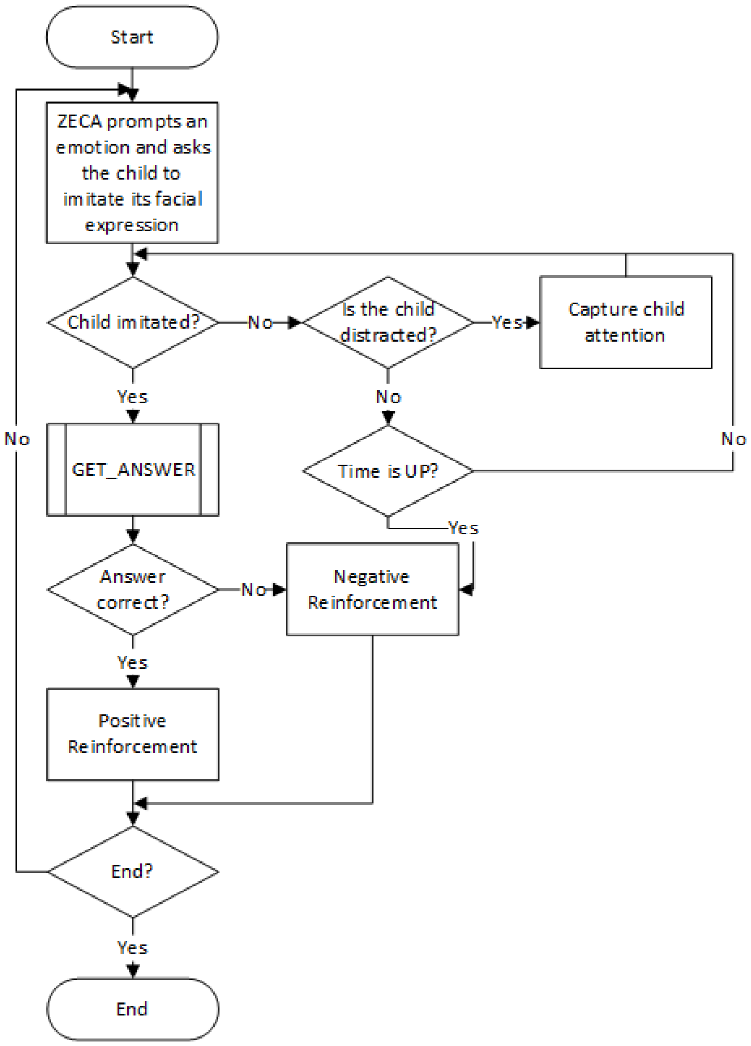
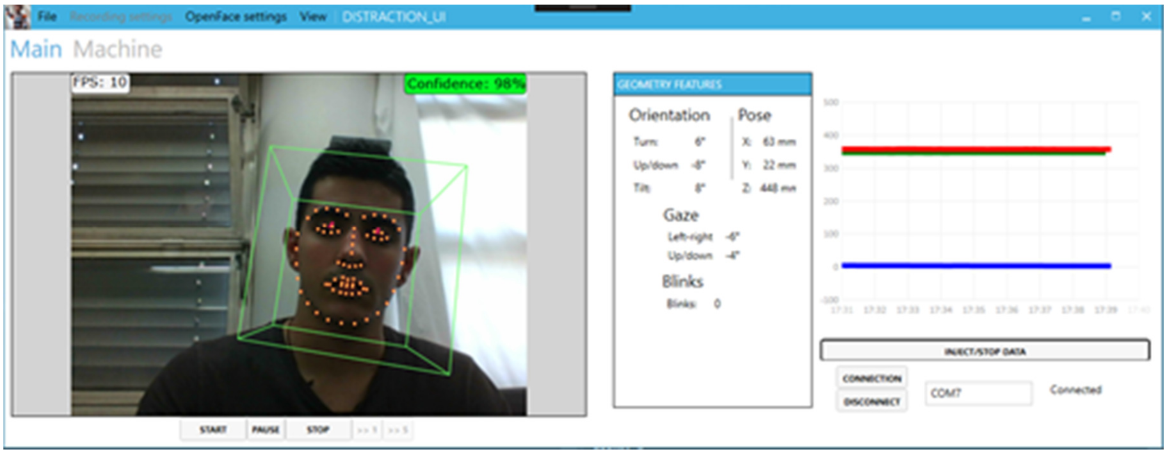


Figure 18 - Robot behaviour flowchart.

## 4.6 Graphical User Interface (GUI)

In order to display the user's information and to monitor the activity, a Graphical User Interface (GUI) was developed. Figure 19 shows the first GUI with the IMU collection data, and Figure 20 shows the final GUI where the head pose angles (pitch, roll, and yaw), the number of blinks, the gaze estimation, the user to position towards the camera, and the prediction of the classification model are displayed.



FPS: 10

Confidence: 98%

Orientation

Turn: 6°

Up/down -8°

Tilt: 8°

Pose

X: 63 mm

Y: 22 mm

Z: 448 mm

Gaze

Left-right -6°

Up/down -4°

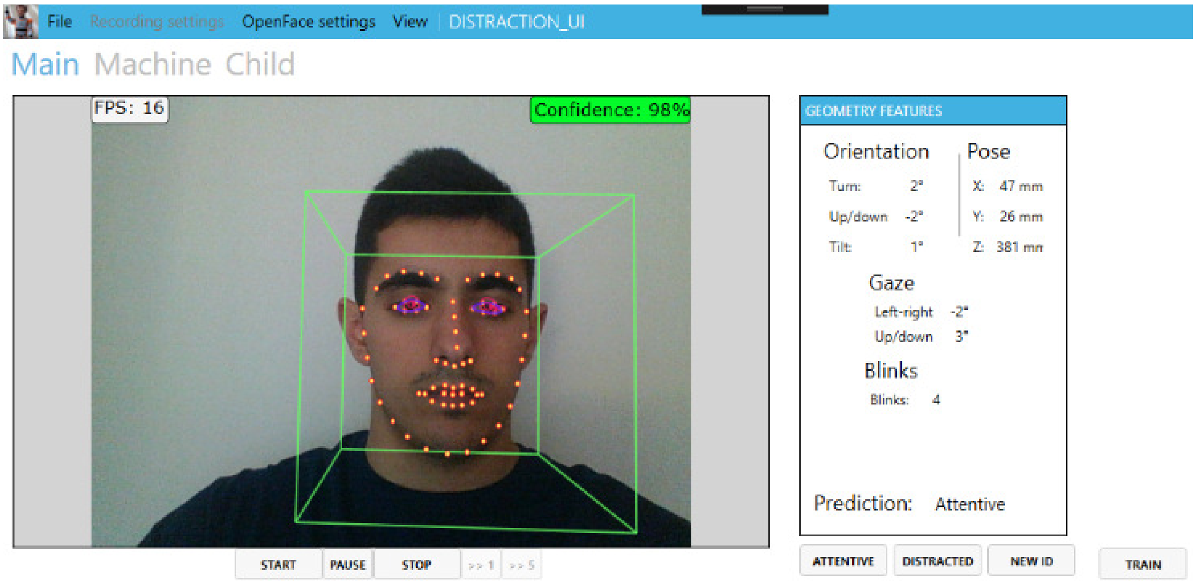
Blinks

Blinks: 0

COM7

Connected

Figure 19 - The GUI displaying the webcam feed, the user's eye gaze and head pose (left), and the IMU values on a chart (right).



FPS: 16

Confidence: 98%

Orientation

Turn: 2°

Up/down -2°

Tilt: 1°

Pose

X: 47 mm

Y: 26 mm

Z: 381 mm

Gaze

Left-right -2°

Up/down 3°

Blinks

Blinks: 4

Prediction: Attentive

Figure 20 - The GUI displaying the webcam feed, the user's eye gaze, the head pose, the number of blinks, the user to position towards the camera, and the prediction of the classification model.

In the "Main" tab, it is possible to see the user's face with the rendering of selected facial landmarks of the OpenFace. This tab also presents others metrics such as the number of blinks, the head orientation values (pitch, yaw, and roll), the direction of the pupil, the distance that the user is from the robot, and the prediction of the machine learning model. The distance between the user and the robot, is also an important parameter, because it can be used to detect if the user is close or far from the robot and, according to that, maintain an appropriate and safe position during the interaction. If the user is starting to be drowsy, an alert will appear on the interface.

The experimenter can initiate or pause or stop the system by pressing the buttons, ‘START’, ‘PAUSE’, AND ‘STOP’, and train the model with new features by pressing the buttons, ‘TRAIN’, ‘ATTENTIVE’ (for attentive patterns), ‘DISTRACTED’ (for distraction patterns), and ‘NEW ID’ (for another subject).

In the “Machine” tab, the experimenter have access to the configuration of the connection between the system and the robot.

The 'Child' tab will display the user’s information. The operator will have access to the database for the users, which will be organized by 'code', 'name', ‘age’, and ‘gender’.

# 5. RESULTS

**Summary**

This chapter starts by presenting the results with the IMU, and the results of the drowsiness detection. Then, it is showed the evaluation of the performance of the two machine learning models used, SVM and k-NN, for the user state classification, with the three databases created. Finally, a discussion of the results is presented.

5.1 IMU Results

5.2 Drowsiness detection

5.3 SVM and k-NN models results

5.4 Laboratorial Results

5.5 Results with children with ASD

5.6 Results with the merged database (children with ASD and adults)

5.7 Discussion of the results

The performance of the BPAS was evaluated using the methodology – System validation, presented in the chapter 3 section 3.4.

The overall system was first tested in a laboratory environment, to verify manually whether the system is working properly. Then, features of children with the autism spectrum disorders were used to train and test the developed model, using offline classification. Finally, the system was evaluated using data from children with ASD and data from typically developing individuals.

## 5.1 IMU Results

As a first approach, the present work proposes the development of a framework that combines the use of two sensors: a RGB camera and an IMU. These two devices are used to collect data, it is used two different sensors to collect the same data, in order to make the system more robust, to avoid failures when, for example, the system cannot detect the user's face.

The IMU used has 10 DOF (3-axis gyroscope, 3-axis accelerometer, and 3-axis compass / magnetometer), that through the use of sensor fusion algorithms, can output the Euler angles (pitch, roll, and yaw), giving more robustness to the system. The data is transferred via Bluetooth to the application, which then displays it in the interface, enabling the user to compare the results obtained by the camera and the IMU. The preliminary results of the IMU data extraction are represented in Figure 21, where it is possible to see the pitch (red), roll (green) and yaw (blue) representation of a user's head pose. During the experiments with the children with ASD, the IMU was placed on a hat.

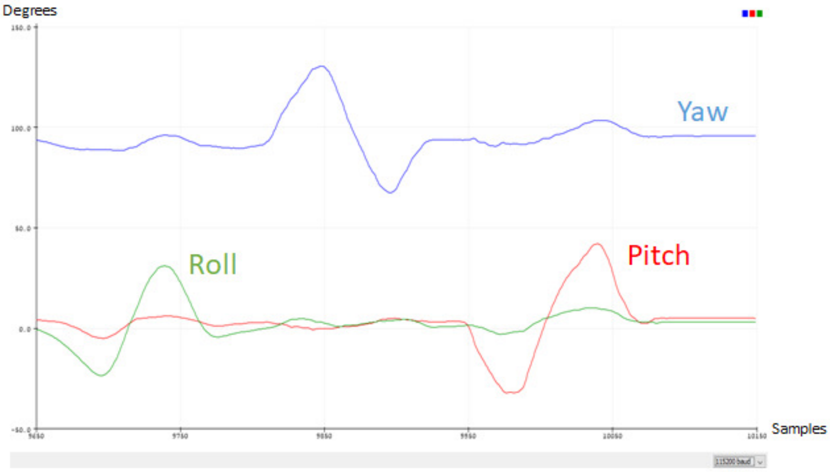
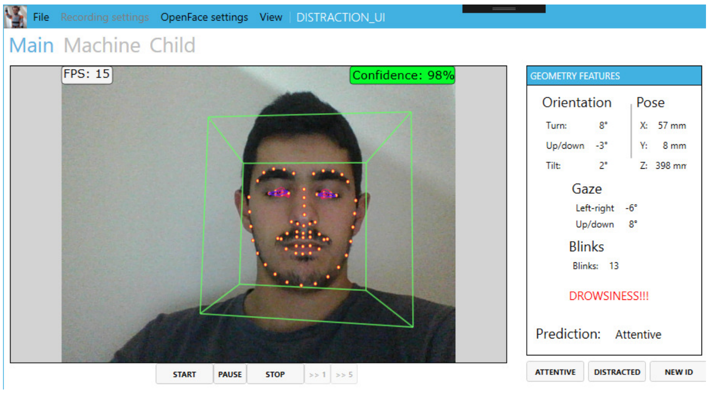


Figure 21 - Preliminary results of the IMU data extraction

However, in general, most of the children with ASD that participated in the study did not adhere to the hat. This was outcome was expected because, in general, children with ASD do not adhere well to using wearable devices without a special reason from their point of view (Mikkelsen, Wodka, Mostofsky, & Puts, 2018). Therefore, the use of the IMU was withdrawn from the study, thus continuing the work with only one device (non-wearable), the RGB camera.

## 5.2 Drowsiness detection

Although it is not usual for a child to exhibit drowsiness behaviour during an activity, it should not be a discarded as a distraction pattern, since in longer activities this can happen. Taking this into account, an algorithm capable of detecting drowsiness was developed, based on the action unit corresponding to the blink. If a user stays for more than 3 seconds with their eyes closed an alert appears on the interface, Figure 22. After this detection in the future, the robot may be able to catch the user's attention during the activity.



FPS: 15

Confidence: 98%

Orientation

Turn: 8°

Up/down -3°

Tilt: 2°

Pose

X: 57 mm

Y: 8 mm

Z: 398 mm

Gaze

Left-right -6°

Up/down 8°

Blinks

Blinks: 13

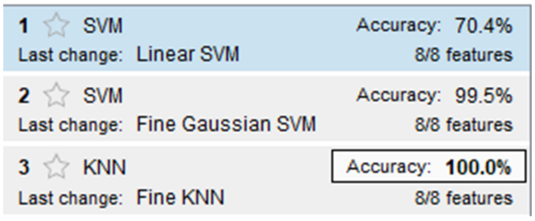
DROWSINESS!!!

Prediction: Attentive

Figure 22 - The GUI displaying the drowsiness alert. The alert output is presented in the interface with red letters.

## 5.3 SVM and k-NN models results

The SVM and k-NN models were simulated in Matlab, with the first database created in a laboratorial environment, using the Machine Learning Toolbox. In order to choose the best kernel function in the case of the SVM, the model was tested first using the Linear kernel and later the Gaussian kernel, where better results were obtained, as can be seen in the Figures 23, 24, and 25 (accuracy, Confusion Matrix, and ROC curve). By analysing the accuracy, confusion matrix, and roc curve obtained for each of the models it is possible to verify that the SVM model with the linear kernel had the worst performance. Since k-NN and the SVM with the Gaussian kernel presented the best performance, these two models were implemented by using the Accord library ("Accord.NET Machine Learning Framework," n.d.) to test, in real time, the performance of the two models. In order to choose the best value for *C* parameter of the SVM, a grid search algorithm was implemented, returning the of 0.5 as being the best for the *C* parameter. In the algorithm used for the k-NN model the value of k used was 5.



1 ☆ SVM Accuracy: 70.4%
Last change: Linear SVM 8/8 features

2 ☆ SVM Accuracy: 99.5%
Last change: Fine Gaussian SVM 8/8 features

3 ☆ KNN Accuracy: **100.0%**
Last change: Fine KNN 8/8 features

Figure 23 - Accuracy of the three models tested.

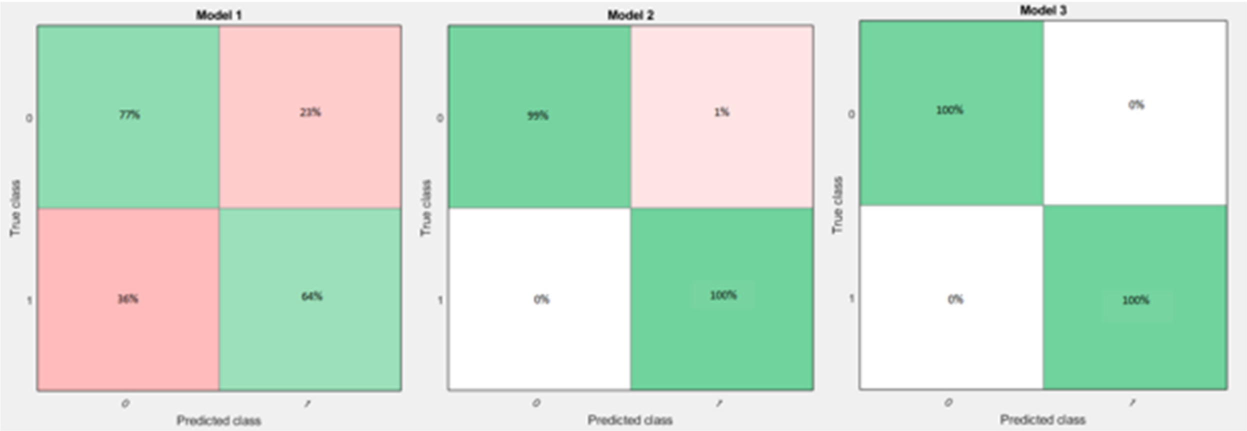


Figure 24 - Confusion matrix of the three models tested (SVM with the linear kernel in the left, SVM with Gaussian kernel in the middle, and k-NN in the right.

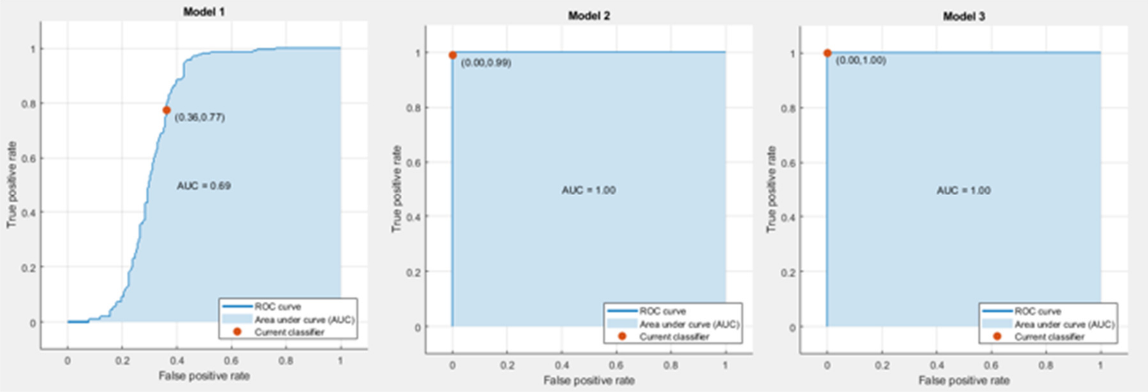


Figure 25 - ROC curve of the three models tested (SVM with the linear kernel in the left, SVM with Gaussian kernel in the middle, and k-NN in the right.

## 5.4 Laboratorial Results

Several distraction and attention behaviours were simulated in a laboratorial environment, where the user sat in front of the camera. The corresponding patterns were stored in a training database (head pose, eye gaze, and user to position towards the camera).

To test the robustness of the system a test database was created, which corresponds to 20% of the total database created (80% for the training database and 20% for the test database), and tested with the two models developed, the SVM with Gaussian kernel and the k-NN, in order to find out the best model for this type of classification.

Tables 3 and 4 represent the confusion matrix for the Gaussian SVM method with the two databases created.

Table 3 - Confusion matrix for Gaussian SVM method with training database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 100% | 0% |
| **Distracted** | 0% | 100% |

Table 4 - Confusion matrix for Gaussian SVM method with test database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 80% | 3% |
| **Distracted** | 20% | 97% |

By analysing Table 3, it is possible to see that the results are perfect, since the accuracy of the attentive and distracted classes are 100%. Regarding the confusion matrix with the test database (Table 4), it is possible to see that the accuracy of the distracted class slightly decreased to 97%. Conversely, the accuracy of the attentive class decreased to 80%.

In Tables 5 and 6, it is presented the confusion matrix for the k-NN method with the two databases created.

Table 5 - Confusion matrix for k-NN method with training database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 100% | 0% |
| **Distracted** | 0% | 100% |

Table 6 - Confusion matrix for k-NN method with test database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 64% | 3% |
| **Distracted** | 36% | 97% |

Analysing Table 5, the k-NN had a performance equal to the Gaussian SVM method during the training phase. However, by analysing the performance of the k-NN method with the test database (Table 6), although the accuracy of the distraction class is the same as that obtained with SVM, with the attentive class the accuracy is very low compared to that obtained with SVM.

Tables 7 and 8 present the results for the accuracy, Matthews Correlation Coefficient (MCC), the sensitivity, the specificity, the precision, and the Area Under the Curve (AUC) obtained with the Gaussian SVM and the k-NN methods for the training and test databases, respectively.

Table 7 - Metrics obtained with the Gaussian SVM method for the training and test databases.

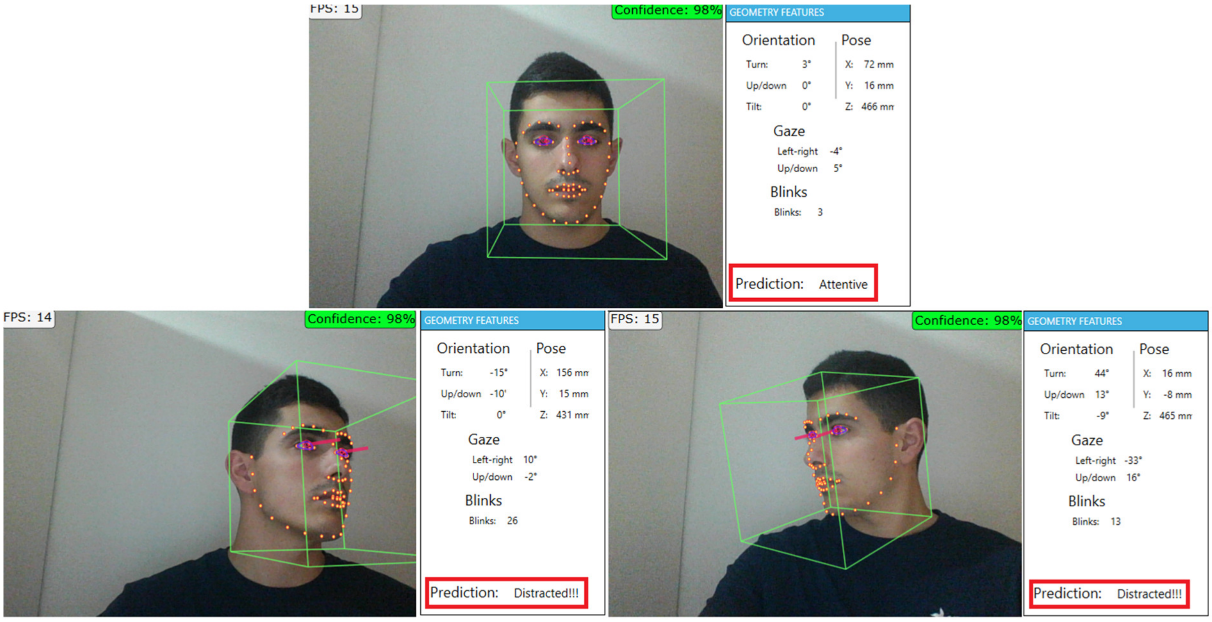
| **Metrics** | **Database** | |
|---|---|---|
| | **Training** | **Test** |
| **Accuracy** | 100% | 88% |
| **MCC** | 100% | 77% |
| **Sensitivity** | 100% | 80% |
| **Specificity** | 100% | 97% |
| **Precision** | 100% | 97% |
| **AUC** | 100% | 89% |

Table 8 - Metrics obtained with the k-NN method for the training and test databases.

| **Metrics** | **Database** | |
|---|---|---|
| | **Training** | **Test** |
| **Accuracy** | 100% | 79% |
| **MCC** | 100% | 64% |
| **Sensitivity** | 100% | 64% |
| **Specificity** | 100% | 97% |
| **Precision** | 100% | 97% |
| **AUC** | 100% | 81% |

Analysing both Tables, it is possible to conclude that, in general, the SVM with the Gaussian kernel achieved better results with an accuracy of 88% when compared with the k-NN method (accuracy: 79%) with the test database. Although the k-NN method obtains excellent results with the training database, the same does not happen with the test database, having the lowest results in some of the metrics, more concretely in the MCC and sensitivity metrics.

As the Gaussian SVM model showed better results, it was used in a real-time laboratorial environment evaluation. The user sat in front of the camera and performed simulated behaviours, attentive and distracted. The system automatically classified the user state. Some of the results for different positions obtained using this method are shown in Figure 26.



FPS: 15
Confidence: 98%
GEOMETRY FEATURES
Orientation
Turn: 3°
Up/down 0°
Tilt: 0°
Pose
X: 72 mm
Y: 16 mm
Z: 466 mm
Gaze
Left-right -4°
Up/down 5°
Blinks
Blinks: 3
Prediction: Attentive

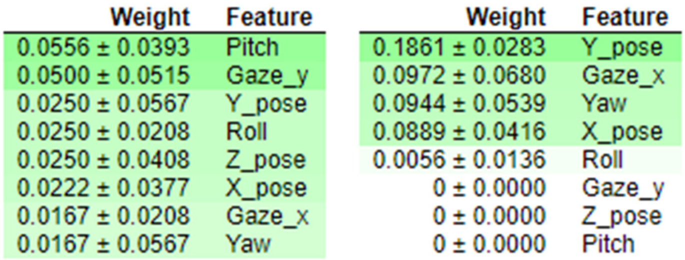
FPS: 14
Confidence: 98%
GEOMETRY FEATURES
Orientation
Turn: -15°
Up/down -10°
Tilt: 0°
Pose
X: 156 mm
Y: 15 mm
Z: 431 mm
Gaze
Left-right 10°
Up/down -2°
Blinks
Blinks: 26
Prediction: Distracted!!!

FPS: 15
Confidence: 98%
GEOMETRY FEATURES
Orientation
Turn: 44°
Up/down 13°
Tilt: -9°
Pose
X: 16 mm
Y: -8 mm
Z: 465 mm
Gaze
Left-right -33°
Up/down 16°
Blinks
Blinks: 13
Prediction: Distracted!!!

Figure 26 - The GUI displaying the classification considering different poses (results using Gaussian SVM method). The classifier output is presented in the interface (red rectangles).

### 5.4.1 Permutation Importance in the laboratorial database

After training and testing the models, one way to make sure that all the analysed features are really contributing for classification the permutation importance method was used. By analysing Figure 27, it is possible to conclude that, in the case of SVM all the features are important, the two most important being the pitch and the gaze_y. This means that adults tend to become more distracted by lowering or raising their heads (not facing the activity) or looking away. In the case of k-NN only 4 features have some importance to the model (y_pose, gaze_x, yaw, x_pose), with y_pose being the one with the highest decision weight in the classification. Since the SVM obtained better results in the classification the patterns that must be more important and describe better the patterns of distraction for the adults is the pitch and gaze_y.



| Weight | Feature |
|---|---|
| 0.0556 ± 0.0393 | Pitch |
| 0.0500 ± 0.0515 | Gaze_y |
| 0.0250 ± 0.0567 | Y_pose |
| 0.0250 ± 0.0208 | Roll |
| 0.0250 ± 0.0408 | Z_pose |
| 0.0222 ± 0.0377 | X_pose |
| 0.0167 ± 0.0208 | Gaze_x |
| 0.0167 ± 0.0567 | Yaw |

| Weight | Feature |
|---|---|
| 0.1861 ± 0.0283 | Y_pose |
| 0.0972 ± 0.0680 | Gaze_x |
| 0.0944 ± 0.0539 | Yaw |
| 0.0889 ± 0.0416 | X_pose |
| 0.0056 ± 0.0136 | Roll |
| 0 ± 0.0000 | Gaze_y |
| 0 ± 0.0000 | Z_pose |
| 0 ± 0.0000 | Pitch |

Figure 27 - SVM (left), and k-NN (right) permutation importance results for the laboratorial database.

## 5.5 Results with children with ASD

In the tests with the children with ASD, the data was extracted during 4 sessions where the children participated in an activity that consisted in recognize the facial expression displayed by the robot. The features extracted are the same as the one extracted in the laboratorial tests.

To test the robustness of the system a test database was created, which corresponds to 20% of the total database created, as used in laboratory tests. This database was tested with the two models (Gaussian SVM and k-NN), in order to find out the best model for this type of classification.

Tables 9 and 10 represent the confusion matrix for the Gaussian SVM method with the two databases created.

Table 9 - Confusion matrix for Gaussian SVM method with training database.

| **Predicted** | **Actual Class** | |
|---|---|---|
| **Class** | **Attentive** | **Distracted** |
| **Attentive** | 83% | 2% |
| **Distracted** | 17% | 98% |

Table 10 - Confusion matrix for Gaussian SVM method with test database.

| **Predicted** | **Actual Class** | |
|---|---|---|
| **Class** | **Attentive** | **Distracted** |
| **Attentive** | 76% | 18% |
| **Distracted** | 24% | 82% |

By analysing Table 9, it is possible to see that the accuracy of both classes stayed above 80%. Regarding the confusion matrix with the test database (Table 10), it is possible to see that the accuracy of both classes have decrease, but both staying above 75%.

In Tables 11 and 12, it is presented the confusion matrix for the k-NN method with the two databases created.

Table 11 - Confusion matrix for k-NN method with training database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 90% | 1% |
| **Distracted** | 10% | 99% |

Table 12 - Confusion matrix for k-NN method with test database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 49% | 18% |
| **Distracted** | 51% | 82% |

Analysing Table 11, the k-NN had a better performance than Gaussian SVM method during the training phase. However, by analysing the performance of the k-NN method in with the test database (Table 12), its performance decreased, achieving accuracy values less than 50% for one of the classes.

Tables 13 and 14 present the values of accuracy, the Matthews Correlation Coefficient (MCC), the sensitivity, the specificity, the precision, and the Area Under the Curve (AUC) obtained with the Gaussian SVM and the k-NN methods for the training and test databases, respectively.

Table 13 - Metrics obtained with the Gaussian SVM method for the training and test databases.

| **Metrics** | **Database** | |
|---|---|---|
| | **Training** | **Test** |
| **Accuracy** | 96% | 80% |
| **MCC** | 85% | 57% |
| **Sensitivity** | 83% | 75% |
| **Specificity** | 98% | 82% |
| **Precision** | 91% | 71% |
| **AUC** | 91% | 79% |

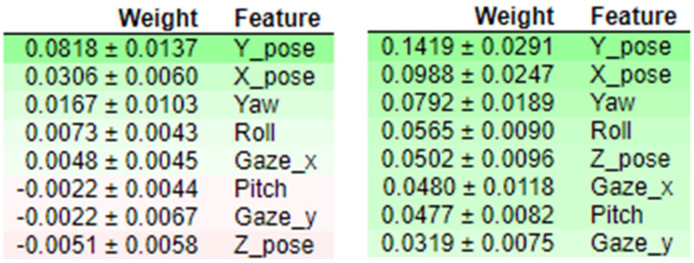
Table 14 - Metrics obtained with the k-NN method for the training and test databases.

| **Metrics** | **Database** | |
|---|---|---|
| | **Training** | **Test** |
| **Accuracy** | 98% | 70% |
| **MCC** | 91% | 33% |
| **Sensitivity** | 90% | 49% |
| **Specificity** | 99% | 82% |
| **Precision** | 96% | 62% |
| **AUC** | 95% | 66% |

Analysing the two Tables, it is possible to conclude that, in general, the SVM had more acceptable results with the test database, despite having a very low MCC, but still above 55%. Additionally, it is also possible to conclude that the k-NN model, as expected after analysing the table 12, presented the worst performance.

### 5.5.1 Permutation Importance in the children with ASD database

After analysing the obtained results of the permutation importance with the database with children with ASD (Figure 28), it is possible to conclude that, in the case of SVM, only 5 features contribute to the classification. It is also possible to see that 3 of the features don't contribute to the model but given the low significance of these values, the features were maintained. In the case of k-NN all the features are important, and the three with the highest decision weight are the same as those obtained with the SVM.



| Weight | Feature |
|---|---|
| 0.0818 ± 0.0137 | Y_pose |
| 0.0306 ± 0.0060 | X_pose |
| 0.0167 ± 0.0103 | Yaw |
| 0.0073 ± 0.0043 | Roll |
| 0.0048 ± 0.0045 | Gaze_x |
| -0.0022 ± 0.0044 | Pitch |
| -0.0022 ± 0.0067 | Gaze_y |
| -0.0051 ± 0.0058 | Z_pose |

| Weight | Feature |
|---|---|
| 0.1419 ± 0.0291 | Y_pose |
| 0.0988 ± 0.0247 | X_pose |
| 0.0792 ± 0.0189 | Yaw |
| 0.0565 ± 0.0090 | Roll |
| 0.0502 ± 0.0096 | Z_pose |
| 0.0480 ± 0.0118 | Gaze_x |
| 0.0477 ± 0.0082 | Pitch |
| 0.0319 ± 0.0075 | Gaze_y |

Figure 28 - SVM (left), and k-NN (right) permutation importance results for the children with autism database.

### 5.5.2 Experimental study with children with ASD

This subchapter presents the analysis of the 3 children (child A, B, and C) that participated in the intervention activity with the robot. In this activity the robot was placed on a table and its face was above the eye horizon of the child. The camera was placed on the robot chest.

The study was conducted by analysing the first and last session in order to classify the state of the child in each activity by registering the profiles of the following features: head pose, eye gaze, and the user position towards the camera. In all graphics, the x axis corresponds to the acquired samples. For the head position feature, the y axis registers the angles of the head (in degrees), being the values near zero the vertical position of the head. For the eye gaze feature, y axis registers the angles of the eye gaze (in degrees): the x_gaze corresponds to horizontal ocular position and the y_gaze corresponds to the vertical ocular position. Since the robot was placed above the eye horizon of the child, the y_gaze has a positive shift. Considering the user position feature, the y axis registers the user position towards the camera (in mm).

It is worth to mention that the data is not normalized, to be easier to understand the child's actions.

The results obtained in the first section and last session with child A are presented in figures 29, 30, and 31, and in tables 15 and 16, respectively.

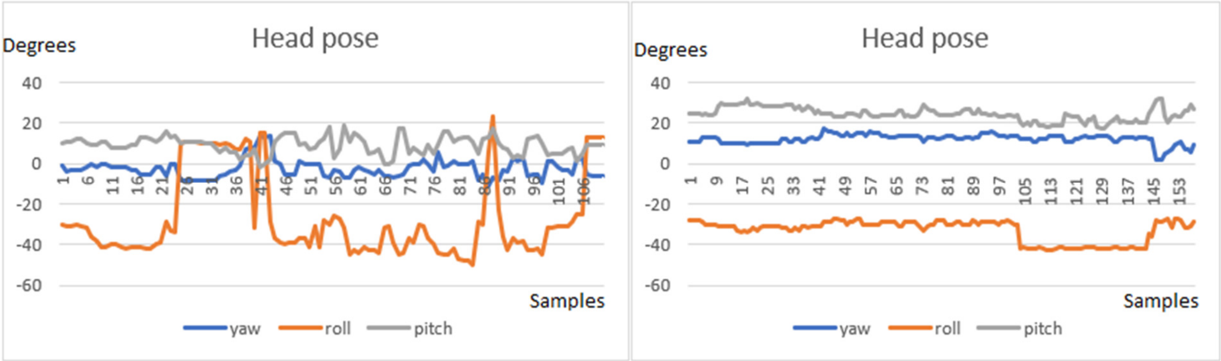


Figure 29 - Head pose extraction of the child A in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the head (in degrees), being the values near zero the vertical position of the head.

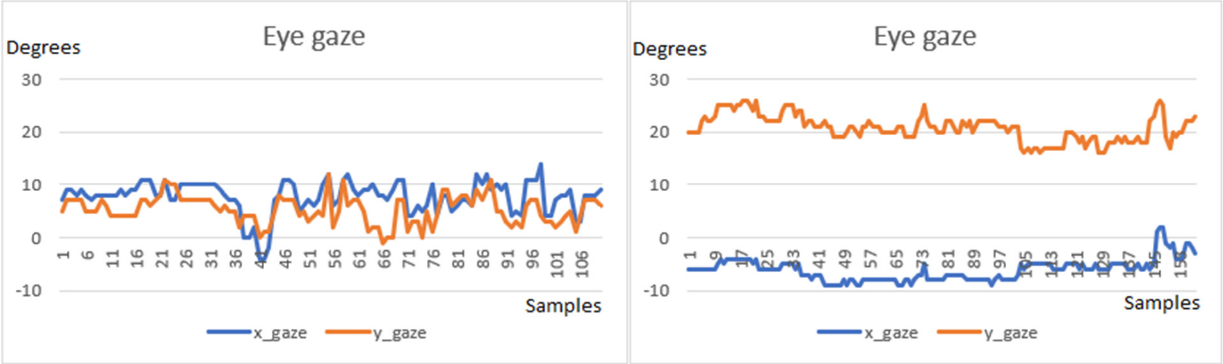


Figure 30 – Eye gaze extraction of the child A in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the eye gaze (in degrees): x_gaze corresponds to horizontal ocular position and y_gaze corresponds to the vertical ocular position. Since the robot was placed above the eye horizon of the child, the y_gaze has a positive shift.

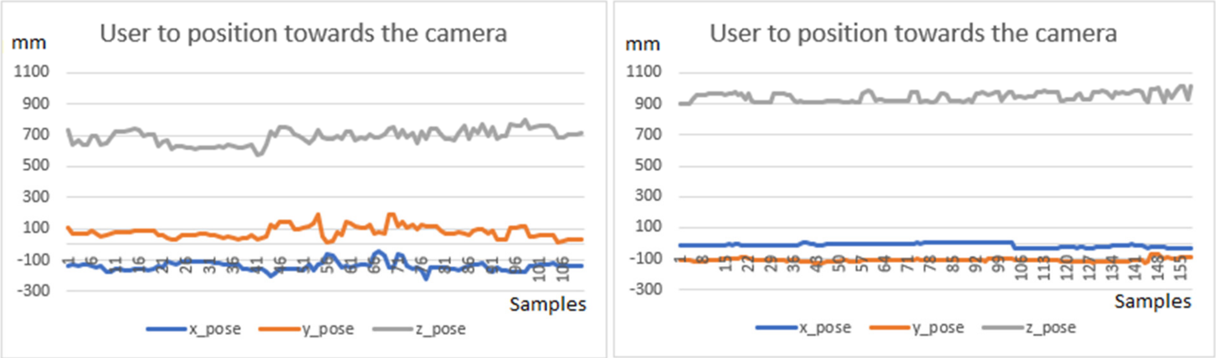


Figure 31 – User to position towards the camera extraction of the child A in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the user position towards the camera (in cm).

Table 15 - Average and Standard deviation of child A in session 1.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | -2.41 | -27.03 | 9.16 | 7.7 | 5.25 | -139.86 | 78.17 | 691.03 |
| **Standard deviation** | 4.55 | 20.60 | 4.33 | 3.13 | 2.66 | 28.36 | 37.29 | 47.57 |

Table 16 - Average and Standard deviation of child A in session 4.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 12.20 | -33.03 | 24.41 | -6.12 | 20.73 | -11.79 | -108.27 | 946.78 |
| **Standard deviation** | 2.50 | 5.35 | 3.29 | 2.09 | 2.50 | 11.59 | 9.43 | 29.15 |

Analysing the results obtained in the first session, it is possible to conclude that child A has some high standard deviation values for some features, namely the roll, x_pose, y_pose and z_pose (-27.03, 28.3, 37.29, and 47.57, respectively). The high value of the roll standard deviation may be due to the fact that this child is not tolerant to noise and therefore placed the head over the shoulder. The high standard deviation values of x_pose, y_pose and z_pose may indicate some distraction throughout the activity (the child is not focused on the activity and she is moving the head and the eye gaze).

After four sessions it is possible to conclude that there was a change in behaviour, since in general, in all the features, the standard deviation decreased, mainly in the roll, where after the help of the therapist and the activity, the child more tolerant to the noise, not putting the head over her shoulder. In the case of x_pose, y_pose, and z_pose, there was also an improvement, and the value of the standard deviation of these features also decreased which might indicate that in the last session child A was more focused on the activity.

The results obtained in the first and last session with child B are presented in figures 32, 33, and 34, and in tables 17 and 18, respectively.

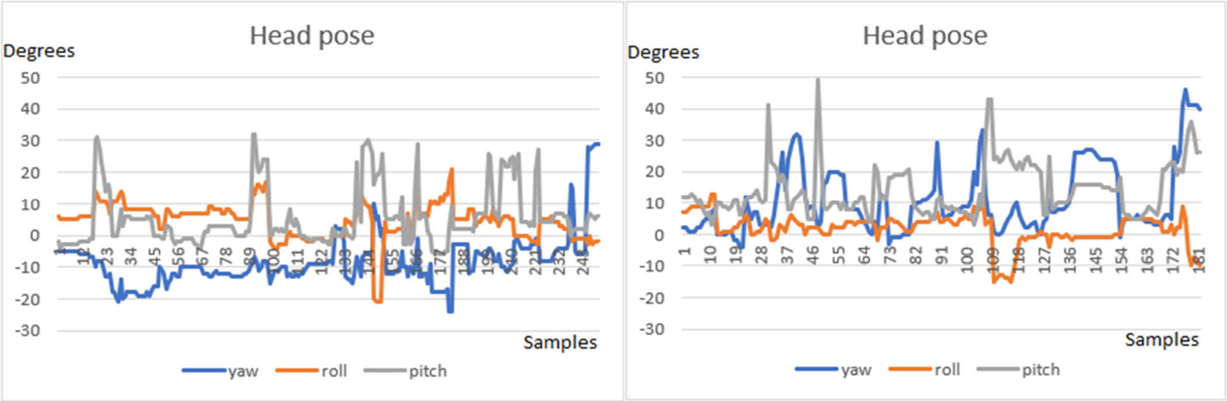


Figure 32 - Head pose extraction of the child B in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the head (in degrees), being the values near zero the vertical position of the head.

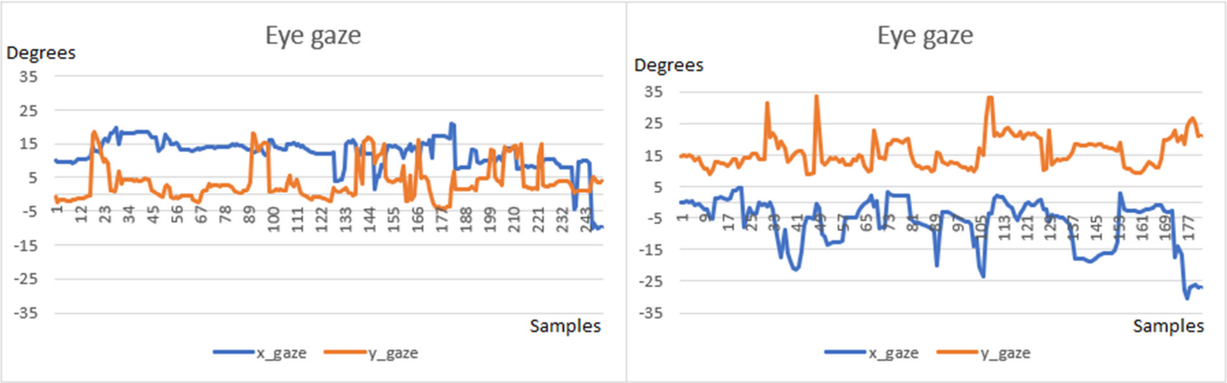


Figure 33 – Eye gaze extraction of the child B in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the eye gaze (in degrees): x_gaze corresponds to horizontal ocular position and y_gaze corresponds to the vertical ocular position. Since the robot was placed above the eye horizon of the child, the y_gaze has a positive shift.

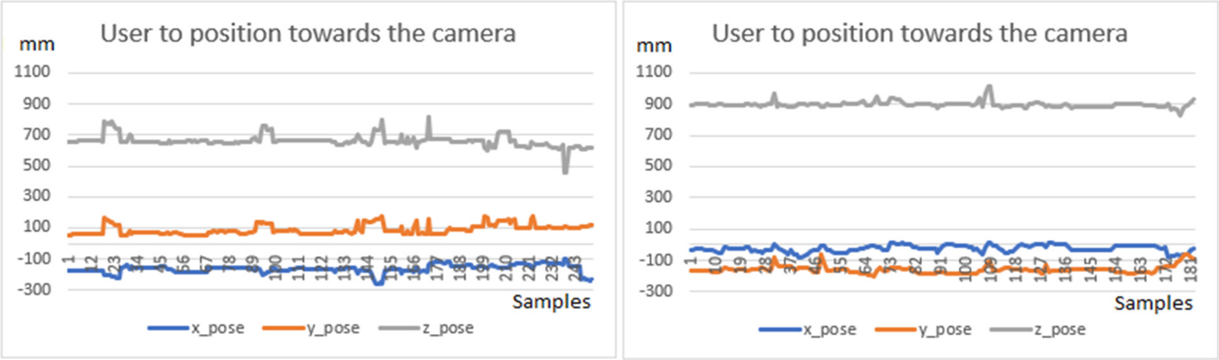


Figure 34 – User to position towards the camera extraction of the child B in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the user position towards the camera (in mm).

Table 17 - Average and Standard deviation of child B in session 1.

| | **Yaw** | **Roll** | **Pitch** | **X_gaze** | **Y_gaze** | **X_pose** | **Y_pose** | **Z_pose** |
|---|---|---|---|---|---|---|---|---|
| **Average** | -8.50 | 4.49 | 5.43 | 12.01 | 3.56 | -162.91 | 88.04 | 661.63 |
| **Standard deviation** | 8.12 | 5.77 | 8.97 | 4.97 | 5.28 | 26.39 | 31.17 | 38.99 |

Table 18 - Average and Standard deviation of child B in session 4.

| | **Yaw** | **Roll** | **Pitch** | **X_gaze** | **Y_gaze** | **X_pose** | **Y_pose** | **Z_pose** |
|---|---|---|---|---|---|---|---|---|
| **Average** | 10.80 | 1.75 | 14.03 | -6.57 | 15.88 | -25.36 | -159.13 | 898.43 |
| **Standard deviation** | 10.84 | 4.93 | 8.19 | 7.74 | 4.95 | 21.42 | 24.32 | 19.76 |

Analysing the results obtained in the first session, it is possible to conclude that, in general, child B is more active (not so focused) than child A, having in almost all the features, a standard deviation greater than the values obtained with child A. The features that present a greater standard deviation are: yaw, pitch, x_pose, y_pose, z_pose (8.12, 8.97, 26.39, 31.17, and 38.99). In the case of pitch, this value may not correspond to distractions of the child because the child had to look at the robot that was above the level of the child's head and then look at the object that was in his/her hands, to respond to the robot. In the remaining features the high standard deviation value may be an indicator of some distractions throughout the activity.

After four sessions it is possible to conclude that there was an improvement, since in general, the values of standard deviation decreased, and in y_pose and z_pose there was a decrease in the most significant standard deviation values (24.22, and 19.76). It can be concluded that although child B is more active than child A, there was also a changing in behaviour during the four sessions, Additionally, by analysing the results of the last session it might indicate that child was more focused on the activity.

The results obtained in the first section and last session with child C are presented in figures 35, 36, and 37, and in tables 19 and 20.

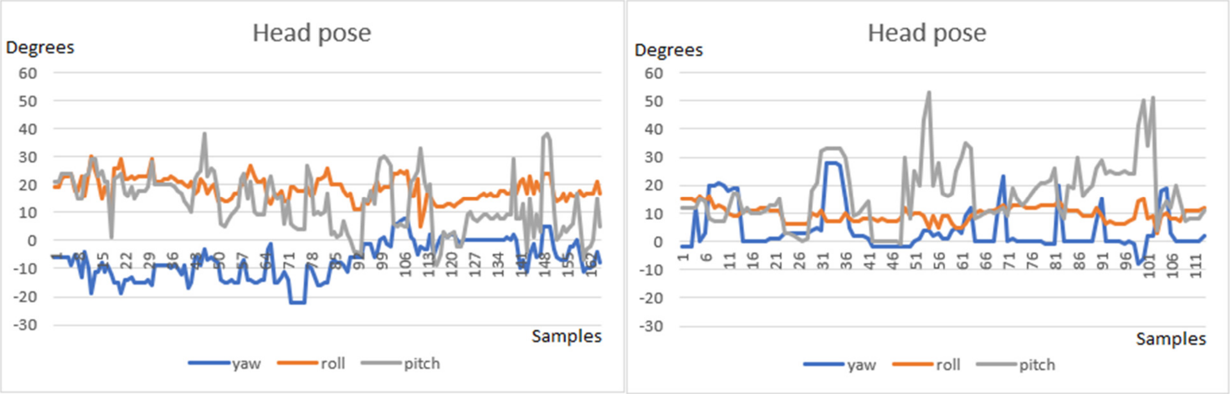


Figure 35 - Head pose extraction of the child C in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the head (in degrees), being the values near zero the vertical position of the head.

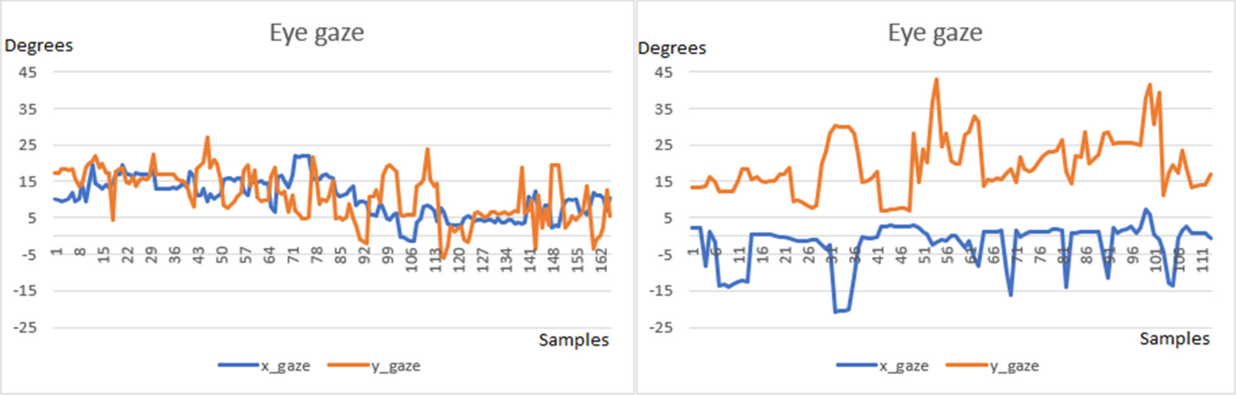


Figure 36 – Eye gaze extraction of the child C in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the eye gaze (in degrees): x_gaze corresponds to horizontal ocular position and y_gaze corresponds to the vertical ocular position. Since the robot was placed above the eye horizon of the child, the y_gaze has a positive shift.

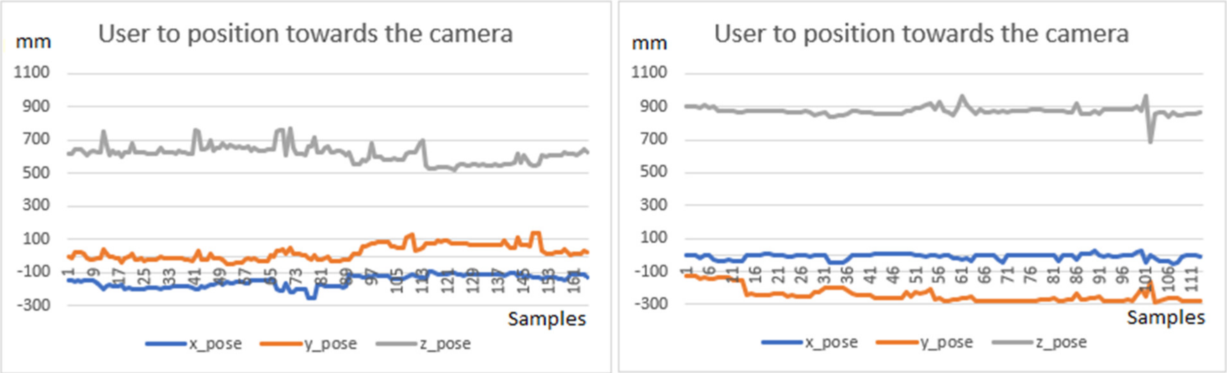


Figure 37 – User to position towards the camera extraction of the child C in session 1(left) and session 4(right). The x axis corresponds to the acquired samples. In y axis are registered the user position towards the camera (in mm).

Table 19 - Average and Standard deviation of child C in session 1.

| | **Yaw** | **Roll** | **Pitch** | **X_gaze** | **Y_gaze** | **X_pose** | **Y_pose** | **Z_pose** |
|---|---|---|---|---|---|---|---|---|
| **Average** | -7.06 | 18.63 | 13.25 | 10.42 | 10.93 | -150.56 | 22.93 | 617.53 |
| **Standard deviation** | 6.70 | 4.13 | 10.08 | 5.32 | 6.90 | 35.95 | 45.13 | 50.85 |

Table 20 - Average and Standard deviation of child C in session 4.

| | **Yaw** | **Roll** | **Pitch** | **X_gaze** | **Y_gaze** | **X_pose** | **Y_pose** | **Z_pose** |
|---|---|---|---|---|---|---|---|---|
| **Average** | 4.12 | 9.73 | 16.73 | -2.16 | 19.58 | -9.97 | -242.30 | 874.05 |
| **Standard deviation** | 7.94 | 2.82 | 11.32 | 6.03 | 7.79 | 16.28 | 43.98 | 27.90 |

Analysing the results obtained in the first session, it is possible to conclude that, as child B, this child is also more active than child A, having in almost all features a standard deviation greater than the values obtained with child A. In this case the features that have a high standard deviation value are: pitch, x_pose, y_pose, and z_pose (10.08, 35.95, 45.13, and 50.85). Just as with child B, the high value for the standard deviation in pitch may be due to the fact that the child has to look at the robot that was above the level of the child's head and then look at the object that was in his/her hands to respond to the robot, thus it might not be an indicator of distraction.

After four sessions it is possible to conclude that in general, the child improved, being more attentive to the activity. The features where the decrease of the standard deviation was most pronounced was in x_pose and z_pose (16.28 and 27.90). Although the change of behaviour of child C is slightly lower than children A and B, it seemed that he/she was also focused on the activity during the last session.

As expected, the head pose and eye gaze features vary synchronously, meaning that the head pose values follow the eye gaze values. It is interesting to notice that for all the children, in the first session, the x_gaze and y_ gaze vary randomly, which migt indicate that the child was not focused neither on the activity nor on the robot. However, in the last session, the x_gaze keeps near zero, indicating the child is looking forward and might be attentive to the activity; and since the y_gaze has a positive shift (the robot was placed above the eye horizon of the child) it might indicate that the child was also more attentive to the robot.

The user position feature is important to infer if the child is present or not in the field of view of the camera which can help to adapt the robot behaviour.

Finally, in general, this study may indicate that all children improve their attention in the activity during the four sessions, suggesting that the use of robots in the intervention sessions may help children with ASD to improve their social interaction.

Also, the proposed system for detecting the distraction patterns may be relevant to allow automatic and adaptive robot behaviours in the intervention sessions.

## 5.6 Results with the merged database (children with ASD and adults)

In order to verify whether it is suitable to have a system capable of detecting patterns of distraction in adults and children with ASD, the two databases were merged. Additionally, to evaluate the robustness of the system a test database was created, which corresponds to 20% of the total database created. This database was tested with the two models (SVM with Gaussian kernel and k-NN), in order to find out the best method for this type of classification.

Tables 21 and 22 represent the confusion matrix for the Gaussian SVM method with the two databases created.

Table 21 - Confusion matrix for Gaussian SVM method with training database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 81% | 1% |
| **Distracted** | 19% | 99% |

Table 22 - Confusion matrix for Gaussian SVM method with test database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 98% | 61% |
| **Distracted** | 2% | 39% |

Regarding the confusion matrix with the test database (Table 22), it is possible to see that the accuracy of the attention class in comparison to the results obtained during the training phase (Table 21). However, in the case of the distraction class, the accuracy decreased to 39%.

In Tables 23 and 24, it is presented the confusion matrix for the k-NN method with the two databases created.

Table 23 - Confusion matrix for k-NN method with training database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 92% | 1% |
| **Distracted** | 8% | 99% |

Table 24 - Confusion matrix for k-NN method with test database.

| **Predicted Class** | **Actual Class** | |
|---|---|---|
| | **Attentive** | **Distracted** |
| **Attentive** | 76% | 25% |
| **Distracted** | 24% | 75% |

Analysing Table 23, and 24 the k-NN had a better performance than the Gaussian SVM method. Even with the test database, achieving an accuracy over 75% for both classes.

Tables 25 and 26 present the values of accuracy, the Matthews Correlation Coefficient (MCC), the sensitivity, the specificity, the precision, and the Area Under the Curve (AUC) obtained with the Gaussian SVM and the k-NN methods for the training and test databases, respectively.

Table 25 - Metrics obtained with the Gaussian SVM method for the training and test databases.

| **Metrics** | **Database** | |
|---|---|---|
| | **Training** | **Test** |
| **Accuracy** | 95% | 69% |
| **MCC** | 83% | 46% |
| **Sensitivity** | 81% | 98% |
| **Specificity** | 98% | 39% |
| **Precision** | 93% | 63% |
| **AUC** | 90% | 68% |

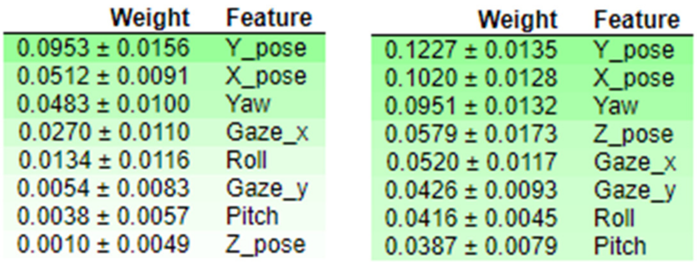
Table 26 - Metrics obtained with the k-NN method for the training and test databases.

| **Metrics** | **Database** | |
|---|---|---|
| | **Training** | **Test** |
| **Accuracy** | 98% | 76% |
| **MCC** | 93% | 52% |
| **Sensitivity** | 92% | 76% |
| **Specificity** | 99% | 75% |
| **Precision** | 96% | 76% |
| **AUC** | 95% | 76% |

Analysing the Table 25 and 26, it is possible to conclude that the SVM model has the lowest results in contrast to the k-NN model. Therefore, the k-NN model achieved the best results with an accuracy of 76% and a MCC of 52%.

### 5.6.1 Permutation Importance in the children with Autism and adults database

After analysing the obtained results of the permutation importance in the children with ASD, and adult's database (Figure 38), it is possible to conclude that, since there are more data with children with ASD (since they are the target of this research), the three characteristics with the greatest weight of decision in both models are the same as those obtained with the database of children with ASD (y_pose, x_pose, and yaw). In both models all features have decision weight, although in the case of SVM, three of them (gaze_y, pitch, z_pose), are not significant.



| Weight | Feature |
|---|---|
| 0.0953 ± 0.0156 | Y_pose |
| 0.0512 ± 0.0091 | X_pose |
| 0.0483 ± 0.0100 | Yaw |
| 0.0270 ± 0.0110 | Gaze_x |
| 0.0134 ± 0.0116 | Roll |
| 0.0054 ± 0.0083 | Gaze_y |
| 0.0038 ± 0.0057 | Pitch |
| 0.0010 ± 0.0049 | Z_pose |

| Weight | Feature |
|---|---|
| 0.1227 ± 0.0135 | Y_pose |
| 0.1020 ± 0.0128 | X_pose |
| 0.0951 ± 0.0132 | Yaw |
| 0.0579 ± 0.0173 | Z_pose |
| 0.0520 ± 0.0117 | Gaze_x |
| 0.0426 ± 0.0093 | Gaze_y |
| 0.0416 ± 0.0045 | Roll |
| 0.0387 ± 0.0079 | Pitch |

Figure 38 - SVM (left), and K-NN (right) permutation importance results for the children with ASD, and adult's database.

## 5.7 Discussion of the results

Analysing the results obtained in the three databases, it is possible to conclude that in the first two, the SVM model had good results, being able to classify with an acceptable accuracy the state of the user. The k-NN model in these two databases did not achieve better results as the Gaussian SVM, especially in the database of children with ASD, where the results were in some of the metrics below 50%. In the results with the merged database the k-NN achieved better results that the Gaussian SVM model. However, both models with the merged database presented lower results in comparison to the other models trained with the children and adult data separately. This difference in the results of the first two databases to the third database (the merged database) can infer that these children might exhibit different patterns of distraction in comparison to typically developing individuals, so it is convenient to have a model dedicated to analysing them. As the objective of the work was to develop a model capable of detecting the distraction patterns of children with ASD, analysing the results it is possible to conclude that this is accomplished with the Gaussian SVM model. The permutation importance also helps explain the fact that adult distraction patterns are different from those of children with ASD, since the features that contribute for the adult model differ from those of children model.

# 6. Conclusion and Future Work

**Summary**

This chapter draws the conclusion of the work described in the dissertation and provides some outlook for the future use of the detection of distraction patterns in a Human-Robot Interaction, especially in the intervention sessions with children with ASD.

The present dissertation concerns the development of a framework to estimate the user/child attentive states. The ability of concentrating on a task for an extended period of time is a paramount skill to develop. In general, when a person tries to reach a particular object, the observer's gaze reaches the target before the action is completed. Thus, the predictive gaze provides the time for the observer to plan and execute an action towards a goal. During a social interaction, the predictive gaze and the attention span can be crucial elements, as they can help an individual to perceive the intentions of the others. However, some individuals present a low attention span, especially children with Autism Spectrum Disorder (ASD), and in general they attend less to faces than typically developing individuals.

Researches have been using robotic platforms for promoting social interaction with individuals with ASD. Furthermore, it has already been proven that the use of robots encourages the development of social interaction skills lacking in children with ASD. However, most of these systems are controlled remotely and cannot adapt automatically to the situation. Even those who are more autonomous still cannot perceive whether or not the user is paying attention to the instructions and the actions of the robot. Additionally, some of these systems use an array of cameras and a hat with markers in order to infer the user gaze.

Taking this into account, the main objective of this dissertation is to detect distraction patterns during an activity, in order to classify the user's state (attentive or distracted) in a task. In the future this system can be used in a human-robot interaction with children with ASD.

The system is based on a camera to detect and follow the face and contours, and extract the following features: the user head orientation angles (yaw, pitch, and roll), eye gaze, action units, blinking frequency, and the user to position towards the camera. It applies an algorithm based on OpenCV functions and OpenFace library. Using the features extracted from the user, machine learning models (Gaussian SVM or k-NN) were trained in order to recognize these patterns and classify the user as distracted or attentive. At a first stage, it was proposed a state machine algorithm to adapt the robot behaviour to the child's state.

In order to evaluate the developed system, three types of tests were performed with three different databases, one with data from adults, another with data from children with ASD and finally another one by merging the two previous databases. The results with the first two databases generally had a better accuracy using the SVM model. Moreover, in the database with only adults, the Gaussian SVM model achieved an accuracy of 88%, against 79 % for the k-NN. Concerning the database with only children with ASD, the SVM model had an accuracy of 80% outperforming the k-NN model that obtained an accuracy of 70%. In the results with the database with data from children with ASD and adults, the model that achieved the best results was the k-NN with an accuracy of 76% against 69% of for the Gaussian SVM. Despite these results, by analysing the permutation importance for each database (adults and children) it is possible to observe that the features that contribute for the adult model differ from those of the children model. Consequently, this difference in the results of the first two databases to the third database (the merged database) can infer that these children might exhibit different patterns of distraction in comparison to typically developing individuals, so it is convenient to have a model dedicated to analysing them.

Since the models trained with the child and adult data separately achieved a better performance and the aim was to create a system capable of detecting patterns of distraction in children with ASD, by analysing the results it is possible to conclude that this is achieved with the Gaussian SVM model with the children database.

In order to test the adequacy of the proposed system in the intervention sessions with children with ASD, an analysis was also made considering each feature (head pose, eye gaze and distance to the camera) registered in the first and last session for each of the three children. These data allow to perceive an evolution in their state of attention/distraction in a task indicating that the proposed system, through the automatic classification as attentive or distraction in an activity, may be a relevant tool in the triadic intervention sessions (child, therapist, and robot ZECA) with this target group.

A test was also performed using a wearable device (an IMU placed in a hat) to determine the head pose of the child. In spite of being a potential suitable solution, it is not tolerable by these children and so it was discarded.

As future work, it is necessary to acquire the patterns associated to distraction and emotional states and classify the corresponding attentive and affective states during a triadic emotion recognition activity with children with ASD (child, researcher and robot ZECA), where the child facial expression is recognized through facial features in real-time. Robot behaviour will be constantly adapted taking into account child attentive state. Through the use of a friendly interface, the teacher/therapist will be able to access the child's performance as well as to monitor the running intervention activity.

# REFERENCES

Accord.NET Machine Learning Framework. (n.d.). Retrieved December 13, 2017, from http://accord-framework.net/

Altmann, A., Toloşi, L., Sander, O., & Lengauer, T. (2010). Permutation importance: a corrected feature importance measure. *Bioinformatics*, *26*(10), 1340–1347. https://doi.org/10.1093/bioinformatics/btq134

Baltrusaitis, T., Mahmoud, M., & Robinson, P. (2015). Cross-dataset learning and person-specific normalisation for automatic Action Unit detection. In *2015 11th IEEE International Conference and Workshops on Automatic Face and Gesture Recognition (FG)* (pp. 1–6). IEEE. https://doi.org/10.1109/FG.2015.7284869

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (n.d.). Constrained Local Neural Fields for robust facial landmark detection in the wild. Retrieved from https://www.cl.cam.ac.uk/research/rainbow/projects/ccnf/files/iccv2014.pdf

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (n.d.). OpenFace: an open source facial behavior analysis toolkit. Retrieved from https://www.cl.cam.ac.uk/research/rainbow/projects/openface/wacv2016.pdf

Baltrušaitis, T., Robinson, P., & Morency, L.-P. (n.d.). *OpenFace: an open source facial behavior analysis toolkit*. Retrieved from https://www.omron.com/ecb/products/mobile/

Begum, M., Serna, R. W., Kontak, D., Allspaw, J., Kuczynski, J., Yanco, H. A., & Suarez, J. (2015). Measuring the Efficacy of Robots in Autism Therapy. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction - HRI '15* (pp. 335–342). New York, New York, USA: ACM Press. https://doi.org/10.1145/2696454.2696480

Bekele, E., Crittendon, J. A., Swanson, A., Sarkar, N., & Warren, Z. E. (2014). Pilot clinical application of an adaptive robotic system for young children with autism. *Autism : The International Journal of Research and Practice*, *18*(5), 598–608. https://doi.org/10.1177/1362361313479454

Bekele, E. T., Lahiri, U., Swanson, A. R., Crittendon, J. A., Warren, Z. E., & Sarkar, N. (2013). A Step Towards Developing Adaptive Robot-Mediated Intervention Architecture (ARIA) for Children With Autism. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, *21*(2), 289–299. https://doi.org/10.1109/TNSRE.2012.2230188

Burges, C. J. C. (1998). A tutorial on support vector machines for pattern recognition. *Data Mining and Knowledge Discovery*, *2*(2), 121–167. https://doi.org/10.1023/A:1009715923555

Caffier, P. P., Erdmann, U., & Ullsperger, P. (2003). Experimental evaluation of eye-blink parameters as a drowsiness measure. *European Journal of Applied Physiology*, *89*(3), 319–325. https://doi.org/10.1007/s00421-003-0807-5

Camera. (n.d.). Retrieved October 19, 2018, from https://i.ebayimg.com/14/!!d5hLSg!mM~$(KGrHgoOKicEjlLmZB)tBKI8w6SUh!~~_35.JPG?set_id=89040003C1

Chevalier, P., Li, J. J., Ainger, E., Alcorn, A. M., Babovic, S., Charisi, V., ... Evers, V. (2017). Dialogue Design for a Robot-Based Face-Mirroring Game to Engage Autistic Children with Emotional Expressions (pp. 546–555). https://doi.org/10.1007/978-3-319-70022-9_54

Chevallier, C., Parish-Morris, J., McVey, A., Rump, K. M., Sasson, N. J., Herrington, J. D., & Schultz, R. T. (2015). Measuring social attention and motivation in autism spectrum disorder using eye-tracking: Stimulus type matters. *Autism Research : Official Journal of the International Society for Autism Research*, *8*(5), 620–8. https://doi.org/10.1002/aur.1479

Confusion metrics. (n.d.). Retrieved October 19, 2018, from http://3.bp.blogspot.com/_txFWHHNYMJQ/THyADzbutYI/AAAAAAAAAf8/TAXL7lySrko/s1600/Picture+8.png

Corrêa, V., & Da Silva, A. (2015). Vinicius Corrêa Alves da Silva Mirroring and recognizing emotions through facial expressions for a Robokind platform Universidade do Minho Escola de Engenharia. Retrieved from http://repositorium.sdum.uminho.pt/bitstream/1822/46634/1/Dissertation_Vinicius%2BCorrêa%2BAlves%2Bda%2BSilva_2016 %281%29.pdf

Costa, S. C. C. (2014). Affective robotics for socio-emotional development in children with autism spectrum disorders. Retrieved from http://hdl.handle.net/1822/35675

Costa, S., Lehmann, H., Dautenhahn, K., Robins, B., & Soares, F. (2015). Using a Humanoid Robot to Elicit Body Awareness and Appropriate Physical Interaction in Children with Autism. *International Journal of Social Robotics*, *7*(2), 265–278. https://doi.org/10.1007/s12369-014-0250-2

Cross-validation: evaluating estimator performance. (n.d.). Retrieved October 19, 2018, from http://scikit-learn.org/stable/modules/cross_validation.html

Cross-Validation - MATLAB &amp; Simulink. (n.d.). Retrieved October 19, 2018, from https://www.mathworks.com/discovery/cross-validation.html

Danisman, T., Bilasco, I. M., Djeraba, C., & Ihaddadene, N. (2010). Drowsy driver detection system using eye blink patterns. In *2010 International Conference on Machine and Web Intelligence* (pp. 230–233). IEEE. https://doi.org/10.1109/ICMWI.2010.5648121

Dautenhahn, K. (n.d.). Design Issues on Interactive Environments for Children with Autism. Retrieved from http://homepages.feis.herts.ac.uk/~comqkd/

Dawson, G., Webb, S. J., & Mcpartland, J. (2005). Understanding the Nature of Face Processing Impairment in Autism: Insights From Behavioral and Electrophysiological Studies. *DEVELOPMENTAL NEUROPSYCHOLOGY*, *27*(3), 403–424. Retrieved from https://pdfs.semanticscholar.org/7d4f/5c9402b225e4cd3bdf6a7019fc183b9de81d.pdf

De Engenharia, E., João, J., & Pacheco, F. (2014). Universidade do Minho Analysis of interaction patterns -Attention. Retrieved from http://islab.di.uminho.pt/camcof/docs/dissertação_José Pacheco.pdf

Ekman, P., & Friesen, W. V. (1978). Facial Action Coding System: A Technique for the Measurement of Facial Movement. Retrieved from www.paulekman.com/wp-content/uploads/2013/07/Measuring-Facial-Movement.pdf

FACS (Facial Action Coding System). (n.d.). Retrieved May 9, 2018, from https://www.cs.cmu.edu/~face/facs.htm

Fanelli, G., Gall, J., & Van Gool, L. (2011). Real time head pose estimation with random regression forests. In *CVPR 2011* (pp. 617–624). IEEE. https://doi.org/10.1109/CVPR.2011.5995458

Ferrari, E., Robins, B., & Dautenhahn, K. (2009). Therapeutic and educational objectives in robot assisted play for children with autism. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (pp. 108–114). https://doi.org/10.1109/ROMAN.2009.5326251

Flanagan, J. R., & Johansson, R. S. (2003). Action plans used in action observation. *Nature*, *424*(6950), 769–771. https://doi.org/10.1038/nature01861

Fowler, B. (2000). A sociological analysis of the satanic verses affair. *Theory, Culture and Society*, *17*(1), 39–61. https://doi.org/10.1177/02632760022050997

Fujimoto, I., Matsumoto, T., Ravindra, · P, De Silva, S., Kobayashi, M., & Higashi, M. (2011). Mimicking and Evaluating Human Motion to Improve the Imitation Skill of Children with Autism Through a Robot. *Int J Soc Robot*, *3*, 349–357. https://doi.org/10.1007/s12369-011-0116-9

Gredebäck, G., & Falck-Ytter, T. (2015). Eye Movements During Action Observation. *Perspectives on Psychological Science : A Journal of the Association for Psychological Science*, *10*(5), 591–8. https://doi.org/10.1177/1745691615589103

Hidden Markov models for time series classification. (n.d.). Retrieved October 20, 2018, from https://towardsdatascience.com/hidden-markov-models-for-time-series-classification-basic-overview-a59b74e5e65b

HomePage - Robótica Autismo. (n.d.). Retrieved December 14, 2017, from http://roboticaautismo.com/

Integrating Tobii Eye Tracking. (2015). Retrieved from https://www.tobii.com/tech/

IROMEC. (n.d.). Retrieved October 19, 2018, from https://content.iospress.com/media/tad/2017/29-3/tad-29-3-tad160166/tad-29-tad160166-g001.jpg?width=755

Jurman, G., & Furlanello, C. (2010). *A unifying view for performance measures in multi-class prediction*. Retrieved from https://arxiv.org/pdf/1008.2908.pdf

K-Nearest Neighbors Algorithm in Python. (n.d.). Retrieved October 17, 2018, from https://stackabuse.com/k-nearest-neighbors-algorithm-in-python-and-scikit-learn/

Kasari, C., Brady, N., Lord, C., & Tager-Flusberg, H. (2013). Assessing the minimally verbal school-aged child with autism spectrum disorder. *Autism Research*. https://doi.org/10.1002/aur.1334

Kaspar. (n.d.). Retrieved October 19, 2018, from http://3.bp.blogspot.com/-t5XWPQRauRs/VhPZWBL-CdI/AAAAAAAAADo/fRwkPpf387c/s1600/Kaspar-002.jpg

KEEPON. (n.d.). Retrieved October 19, 2018, from https://www.jameco.com/Jameco/workshop/inthenews/therapyrobot-fig1.jpg

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. https://doi.org/10.1007/s10803-012-1645-2

Kim, E. S., Paul, R., Shic, F., & Scassellati, B. (2012). Bridging the Research Gap: Making HRI Useful to Individuals with Autism. *Journal of Human-Robot Interaction*, *1*(1), 26–54. https://doi.org/10.5898/JHRI.1.1.Kim

Klin, A., Jones, W., Schultz, R., Volkmar, F., & Cohen, D. (2002). Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism. *Archives of General Psychiatry*, *59*(9), 809–816. https://doi.org/10.1001/archpsyc.59.9.809

Kondori, F. A., Yousefi, S., Li, H., Sonning, S., & Sonning, S. (2011). 3D head pose estimation using the Kinect. In *2011 International Conference on Wireless Communications and Signal Processing (WCSP)* (pp. 1–4). IEEE. https://doi.org/10.1109/WCSP.2011.6096866

Kose-Bagci, H., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2010). Drum-mate: Interaction dynamics and gestures in human-humanoid drumming experiments. *Connection Science*, *22*(2), 103–134. https://doi.org/10.1080/09540090903383189

Kose-Bagci, H., Ferrari, E., Dautenhahn, K., Syrdal, D. S., & Nehaniv, C. L. (2009). Effects of embodiment and gestures on social interaction in drumming games with a humanoid robot. *Advanced Robotics*, *23*(14), 1951–1996. https://doi.org/10.1163/016918609X12518783330360

Kozima, H., Michalowski, M. P., & Nakagawa, C. (2009). Keepon. *International Journal of Social Robotics*, *1*(1), 3–18. https://doi.org/10.1007/s12369-008-0009-8

Kutila, M., Jokela, M., Markkula, G., & Rue, M. R. (2007). Driver Distraction Detection with a Camera Vision System. In *2007 IEEE International Conference on Image Processing* (p. VI-201-VI-204). IEEE. https://doi.org/10.1109/ICIP.2007.4379556

Lee, B.-G., & Chung, W.-Y. (2012). A Smartphone-Based Driver Safety Monitoring System Using Data Fusion. *Sensors 2012, Vol. 12, Pages 17536-17552*, *12*(12), 17536–17552. https://doi.org/10.3390/S121217536

Liang, Y. (2009). Detecting driver distraction. *ProQuest Dissertations and Theses*, *3356220*, 152. Retrieved from http://ezproxy.net.ucf.edu/login?url=http://search.proquest.com/docview/304900929?accountid=10003%5Cnhttp://sfx.fcla.edu/ucf?url_ver=Z39.88-2004&rft_val_fmt=info:ofi/fmt:kev:mtx:dissertation&genre=dissertations+&+theses&sid=ProQ:ProQuest+Dissertations+&+T

Liang, Y., Reyes, M. L., & Lee, J. D. (2007). Real-Time Detection of Driver Cognitive Distraction Using Support Vector Machines. *IEEE Transactions on Intelligent Transportation Systems*, *8*(2), 340–350. https://doi.org/10.1109/TITS.2007.895298

Lotte, F., Congedo, M., Lécuyer, A., Lamarche, F., & Arnaldi, B. (2007). A review of classification algorithms for EEG-based computer interfaces. *Journal of Neural Engineering*, *4*, R1–R13. https://doi.org/10.1088/1741-2560/4/2/R01

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. In *2011 Annual International Conference of the IEEE Engineering in Medicine and Biology Society* (pp. 4515–4518). IEEE. https://doi.org/10.1109/IEMBS.2011.6091119

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. https://doi.org/10.1109/TSMCA.2005.850596

Michaud, F., Laplante, J. F., Larouche, H., Duquette, A., Caron, S., Létourneau, 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. https://doi.org/10.1109/TSMCA.2005.850596

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

Mikkelsen, M., Wodka, E. L., Mostofsky, S. H., & Puts, N. A. J. (2018). Autism spectrum disorder in the scope of tactile processing. *Developmental Cognitive Neuroscience*, *29*, 140–150. https://doi.org/10.1016/J.DCN.2016.12.005

Pennisi, P., Tonacci, A., Tartarisco, G., Billeci, L., Ruta, L., Gangemi, S., & Pioggia, G. (2016). Autism and social robotics: A systematic review. *Autism Research*, *9*(2), 165–183. https://doi.org/10.1002/aur.1527

RBF SVM parameters. (n.d.). Retrieved October 19, 2018, from http://scikit-learn.org/stable/auto_examples/svm/plot_rbf_parameters.html#example-svm-plotrbf-parameters-py

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

Robins, B., Dautenhahn, K., & Dickerson, P. (2009a). From isolation to communication: A case study evaluation of robot assisted play for children with autism with a minimally expressive humanoid robot. In *Proceedings of the 2nd International Conferences on Advances in Computer-Human Interactions, ACHI 2009* (pp. 205–211). https://doi.org/10.1109/ACHI.2009.32

Robins, B., Dautenhahn, K., & Dickerson, P. (2009b). From Isolation to Communication: A Case Study Evaluation of Robot Assisted Play for Children with Autism with a Minimally Expressive Humanoid Robot. In *2009 Second International Conferences on Advances in Computer-Human Interactions* (pp. 205–211). IEEE. https://doi.org/10.1109/ACHI.2009.32

Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., & Olsson, L. (2005). Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: Lessons learnt from an exploratory study. In *Proceedings - IEEE International Workshop on Robot and Human Interactive Communication* (Vol. 2005, pp. 716–722). https://doi.org/10.1109/ROMAN.2005.1513864

Robins, B., Ferrari, E., & Dautenhahn, K. (2008). Developing scenarios for robot assisted play. In *Proceedings of the 17th IEEE International Symposium on Robot and Human Interactive Communication, RO-MAN* (pp. 180–186). https://doi.org/10.1109/ROMAN.2008.4600663

ROC. (n.d.). Retrieved October 19, 2018, from https://www.medcalc.org/manual/_help/images/roc_intro3.png

Sahayadhas, A., Sundaraj, K., & Murugappan, M. (2012). Detecting driver drowsiness based on sensors: a review. *Sensors (Basel, Switzerland)*, *12*(12), 16937–53. https://doi.org/10.3390/s121216937

Sammut, C., & Webb, G. I. (n.d.). *Encyclopedia of Machine Learning and Data Mining Second Edition*. Retrieved from https://universalflowuniversity.com/Books/Computer Programming/Machine Learning and Deep Learning/Encyclopedia of Machine Learning and Data Mining 2nd Edition.pdf

Silva, V., Soares, F., Esteves, J. S., Figueiredo, J., Leao, C. P., Santos, C., & Pereira, A. P. (2016). Real-time emotions recognition system. In *2016 8th International Congress on Ultra Modern Telecommunications and Control Systems and Workshops (ICUMT)* (pp. 201–206). IEEE. https://doi.org/10.1109/ICUMT.2016.7765357

Statistics and Machine Learning Toolbox Documentation. (n.d.). Retrieved October 19, 2018, from http://www.mathworks.com/help/stats/support-vector-machines-svm.html

SVM Margins Example. (n.d.). Retrieved October 19, 2018, from http://scikit-learn.org/stable/auto_examples/svm/plot_svm_margin.html

Symptoms | What is Autism? | Autism Speaks. (n.d.). Retrieved December 12, 2017, from https://www.autismspeaks.org/what-autism/symptoms

Tapus, A., Matari, M. J., Member, S., & Scassellati, B. (n.d.). The Grand Challenges in Socially Assistive Robotics. *IEEE ROBOTICS AND AUTOMATION MAGAZINE SPECIAL ISSUE ON GRAND CHALLENGES IN ROBOTICS*, *1*. Retrieved from https://scazlab.yale.edu/sites/default/files/files/Tapus-RAM2007.pdf

Wainer, J., Dautenhahn, K., Robins, B., & Amirabdollahian, F. (2010). Collaborating with Kaspar: Using an autonomous humanoid robot to foster cooperative dyadic play among children with autism. In *2010 10th IEEE-RAS International Conference on Humanoid Robots, Humanoids 2010* (pp. 631–638). https://doi.org/10.1109/ICHR.2010.5686346

Wainer, J., Robins, B., Amirabdollahian, F., & Dautenhahn, K. (2014). Using the Humanoid Robot KASPAR to Autonomously Play Triadic Games and Facilitate Collaborative Play Among Children With Autism. *IEEE Transactions on Autonomous Mental Development*, *6*(3), 183–199. https://doi.org/10.1109/TAMD.2014.2303116

Wood, E., Baltruaitis, T., Zhang, X., Sugano, Y., Robinson, P., & Bulling, A. (2015). Rendering of Eyes for Eye-Shape Registration and Gaze Estimation. In *2015 IEEE International Conference on Computer Vision (ICCV)* (pp. 3756–3764). IEEE. https://doi.org/10.1109/ICCV.2015.428

Wood, L. J., Dautenhahn, K., Rainer, A., Robins, B., Lehmann, H., & Syrdal, D. S. (2013). Robot-Mediated Interviews - How Effective Is a Humanoid Robot as a Tool for Interviewing Young Children? *PLoS ONE*, *8*(3). https://doi.org/10.1371/journal.pone.0059448

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

Zheng, A. (n.d.). *Evaluating Machine Learning Models*. Retrieved from https://pindex.com/uploads/post_docs/evaluating-machine-learning-models(PINDEX-DOC-6950).pdf

# APPRENDIXES

**Summary**

Materials used in the present dissertation.

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

B. List of facial Action Units and Action Descriptors

C. Data of the 4 sessions used in the experimental study with children with ASD

## A – Consent form delivered to the children’s parents (In Portuguese)

Robotica - Autismo
robotica-autismo.com
Universidade do Minho

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

O **projeto Robótica-Autismo** (http://roboticaautismo.com) visa a aplicação de ferramentas robóticas como forma de melhorar as habilidades de interação e comunicação das crianças com Perturbações do Espectro do Autismo (PEA). No âmbito de teses de Mestrado e de Doutoramento em Engenharia Eletrónica Industrial e Computadores da Universidade do Minho estamos a desenvolver um sistema de imitação e reconhecimento de emoções em que um jogo de computador e o robô Zeca são os mediadores da interação. Assim, gostaríamos de convidar o seu educando a participar nas sessões de teste: a criança deve exprimir ou adivinhar uma emoção (contente, triste, medo, zangado, assustado ou neutro). Estas sessões têm uma duração de cerca de 10 minutos, são realizadas durante o tempo letivo, mas sem prejuízo do normal funcionamento das aulas. As sessões serão gravadas em vídeo e as respostas são anónimas. Em caso de divulgação científica dos vídeos, a cara da criança será desfocada. Solicitamos, assim, a sua colaboração dando o seu consentimento através da devolução do anexo devidamente preenchido e assinado.

7 de Maio de 2018

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

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. Em caso de divulgação científica dos vídeos, a cara da criança será desfocada.
- 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.

________________________________________________

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

## B – List of facial Action Units and Action Descriptors

Table 27 - List of Action Units (with underlying facial muscles) (Ekman & Friesen, 1978; "FACS (Facial Action Coding System)," 2002

| AU# | FACS Name | Muscular Basis |
|---|---|---|
| **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 |
| **9** | Nose Wrinkler | Levator labii superioris alaquae nasi |
| **10** | Upper Lip Raiser | Levator labii superioris |
| **11** | Nasolabial Deepener | Zygomaticus minor |
| **12** | Lip Corner Puller | Zygomaticus major |
| **13** | Cheek Puffer | Levator anguli oris (a.k.a. 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 Puckerer | Incisivii labii superioris and Incisivii labii inferioris |
| **20** | Lip stretcher | Risorius w/ 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 |
| **27** | Mouth Stretch | Pterygoids, Digastric |
| **28** | Lip Suck | Orbicularis oris |
| **41** | Lid droop | Relaxation of Levator palpebrae superioris |

| | | |
|---|---|---|
| **42** | Slit | Orbicularis oculi |
| **43** | Eyes Closed | Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis |
| **44** | Squint | Orbicularis oculi, pars palpebralis |
| **45** | Blink | Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis |
| **46** | Wink | Relaxation of Levator palpebrae superioris; Orbicularis oculi, pars palpebralis |
| **51** | Head turn left | |
| **52** | Head turn right | |
| **53** | Head up | |
| **54** | Head down | |
| **55** | Head tilt left | |
| **56** | Head tilt right | |
| **57** | Head forward | |
| **58** | Head back | |
| **61** | Eyes turn left | |
| **62** | Eyes turn right | |
| **63** | Eyes up | |
| **64** | Eyes down | |

## C - Data of the 4 sessions used in the experimental study with children with ASD

Child A:

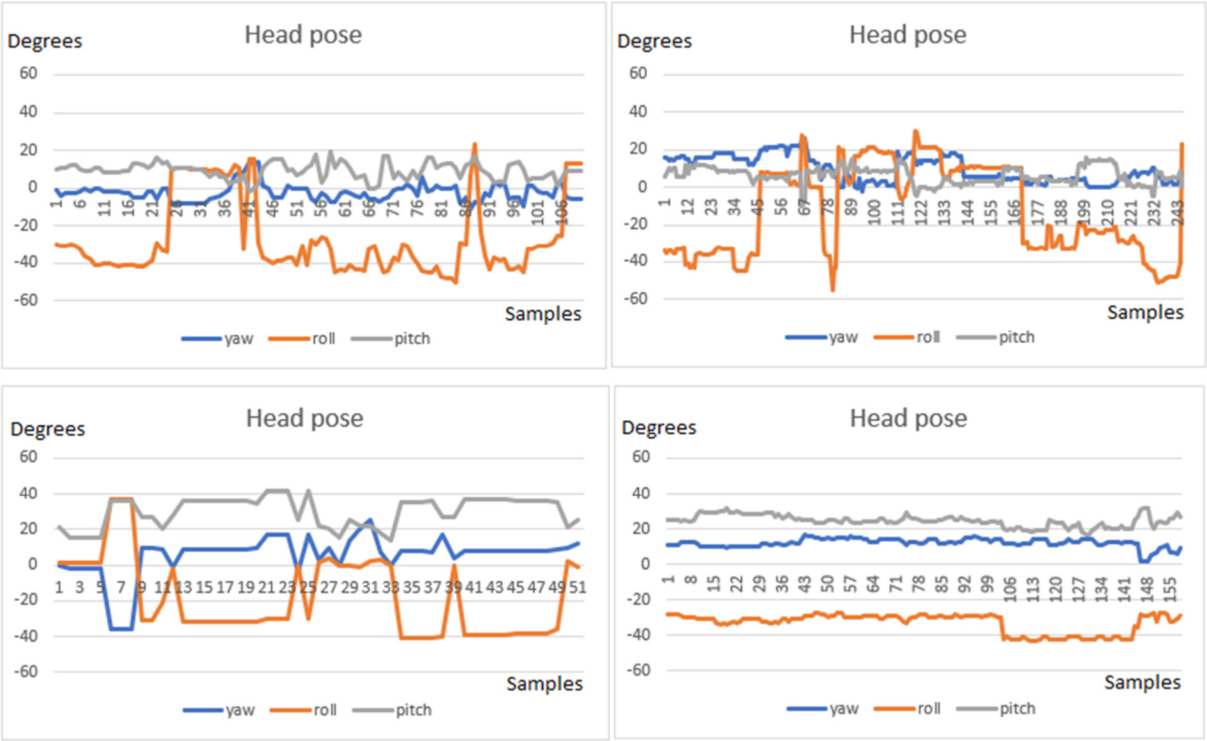


Figure 39 - Head pose extraction of the child A in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right).
The x axis corresponds to the acquired samples. In y axis are registered the angles of the head (in degrees), being the values near zero the vertical position of the head.

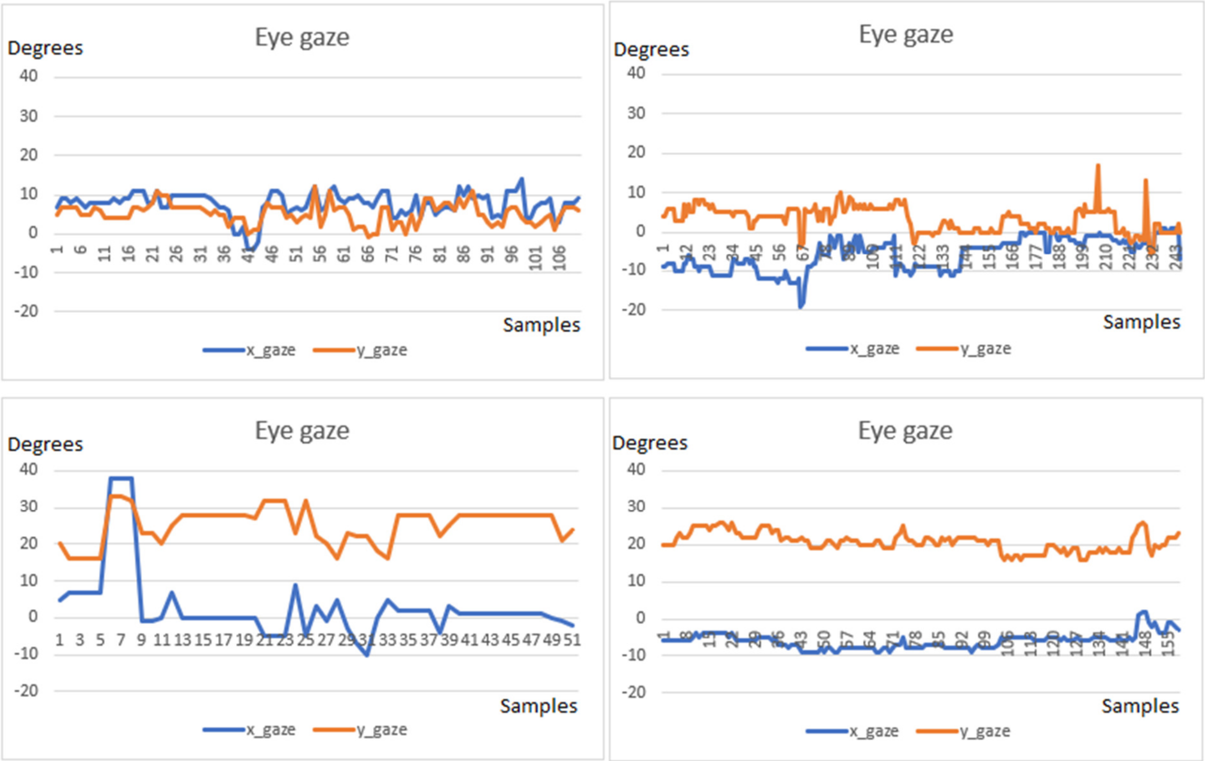


Figure 40 - Eye gaze extraction of the child A in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the eye gaze (in degrees): x_gaze corresponds to horizontal ocular position and y_gaze corresponds to the vertical ocular position.

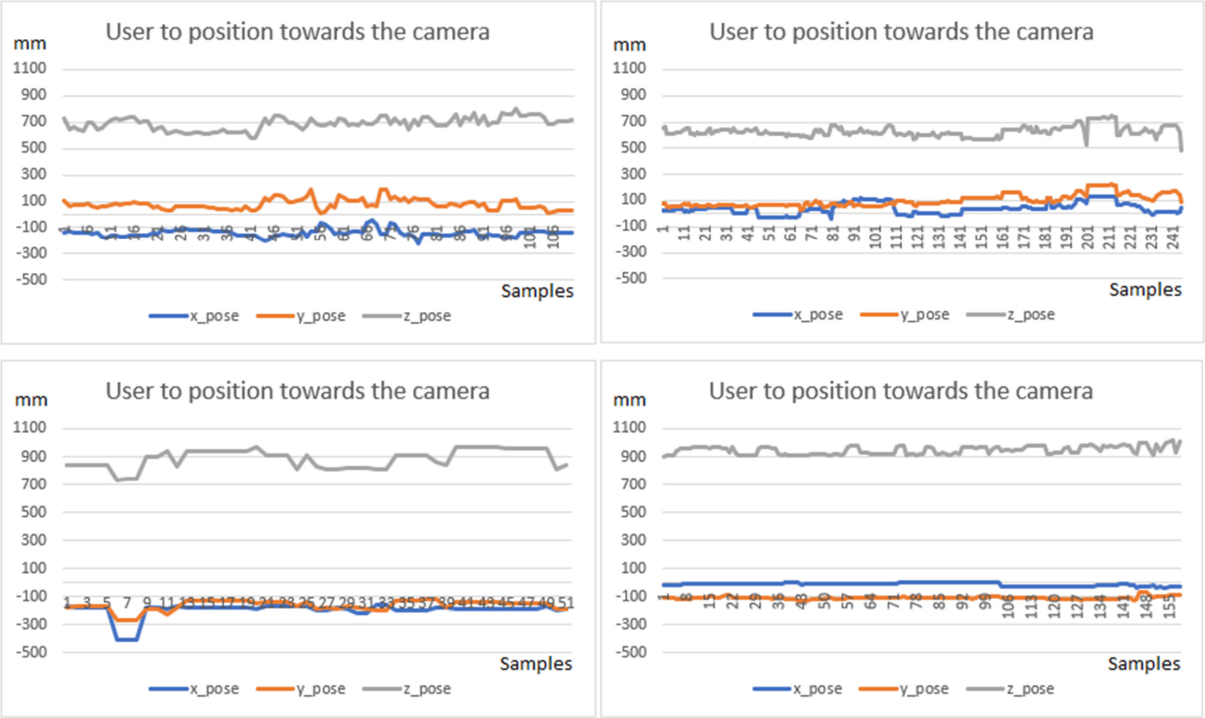


Figure 41 - User to position towards the camera extraction of the child A in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right). The x axis corresponds to the acquired samples. In y axis are registered the user position towards the camera (in mm).

Table 28 - Average and Standard deviation of child A in sesssion 1.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | -2.41 | -27.03 | 9.16 | 7.7 | 5.25 | -139.86 | 78.17 | 691.03 |
| **Standard deviation** | 4.55 | 20.60 | 4.33 | 3.13 | 2.66 | 28.36 | 37.29 | 47.57 |

Table 29 - Average and Standard deviation of child A in sesssion 2.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 9.56 | -12.39 | 6.42 | 5.69 | 3.26 | 40.09 | 100.93 | 628.14 |
| **Standard deviation** | 6.93 | 24.06 | 4.51 | 4.18 | 3.17 | 43.43 | 45.23 | 39.69 |

Table 30 - Average and Standard deviation of child A in sesssion 3.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 5.55 | -18.08 | 30.06 | 2.86 | 25.29 | -195.43 | -162.57 | 885.37 |
| **Standard deviation** | 12.07 | 22.20 | 8.57 | 9.63 | 4.98 | 55.31 | 35.96 | 68.97 |

Table 31 - Average and Standard deviation of child A in sesssion 4.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 12.20 | -33.03 | 24.41 | -6.12 | 20.73 | -11.79 | -108.27 | 946.78 |
| **Standard deviation** | 2.50 | 5.35 | 3.29 | 2.09 | 2.50 | 11.59 | 9.43 | 29.15 |

Child B:

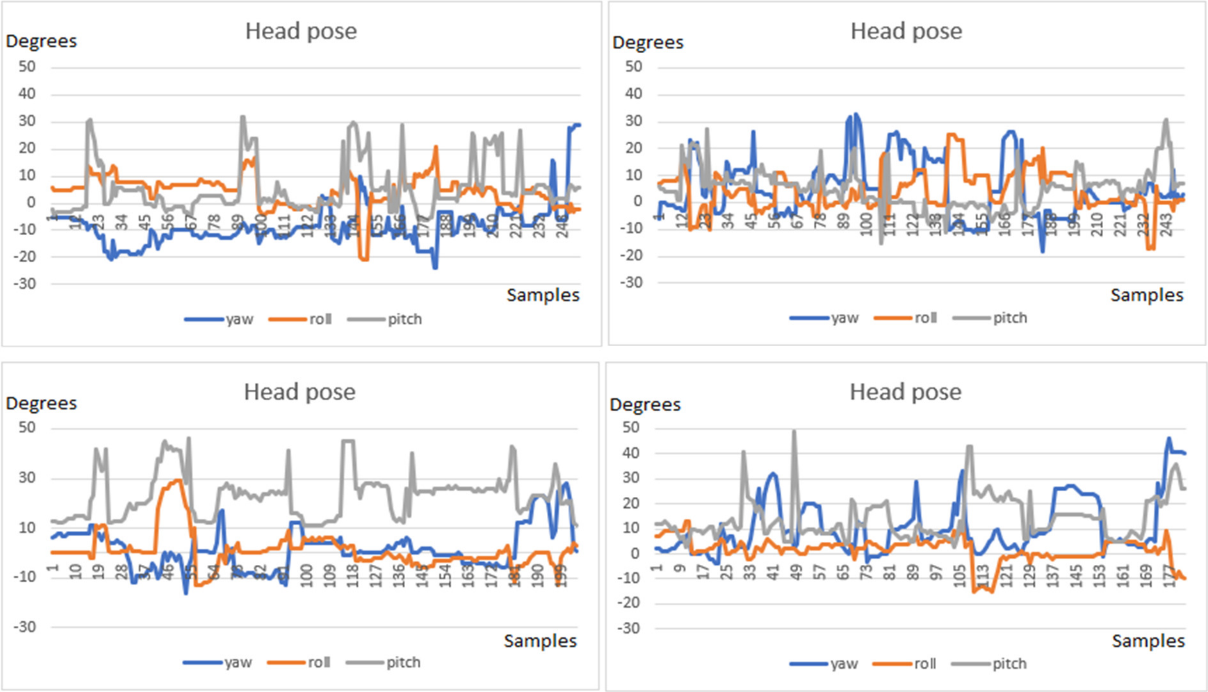


Figure 42 - Head pose extraction of the child B in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right).
The x axis corresponds to the acquired samples. In y axis are registered the angles of the head (in degrees), being the values near zero the vertical position of the head.

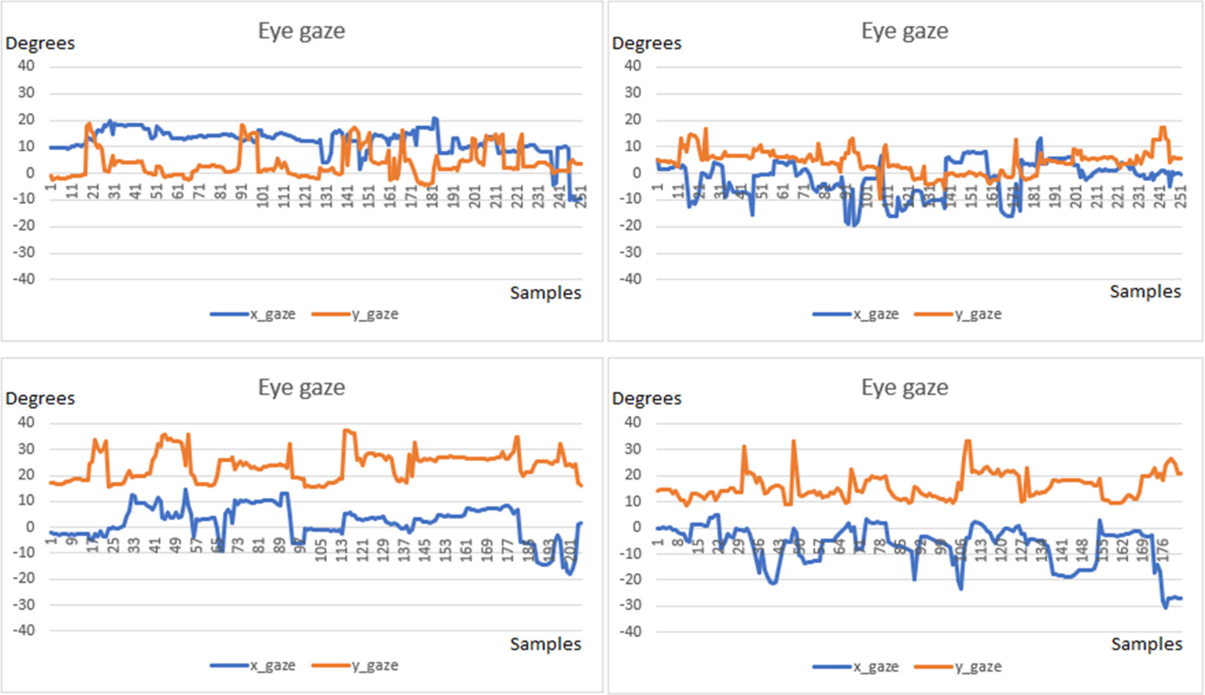


Figure 43 - Eye gaze extraction of the child B in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the eye gaze (in degrees): x_gaze corresponds to horizontal ocular position and y_gaze corresponds to the vertical ocular position.

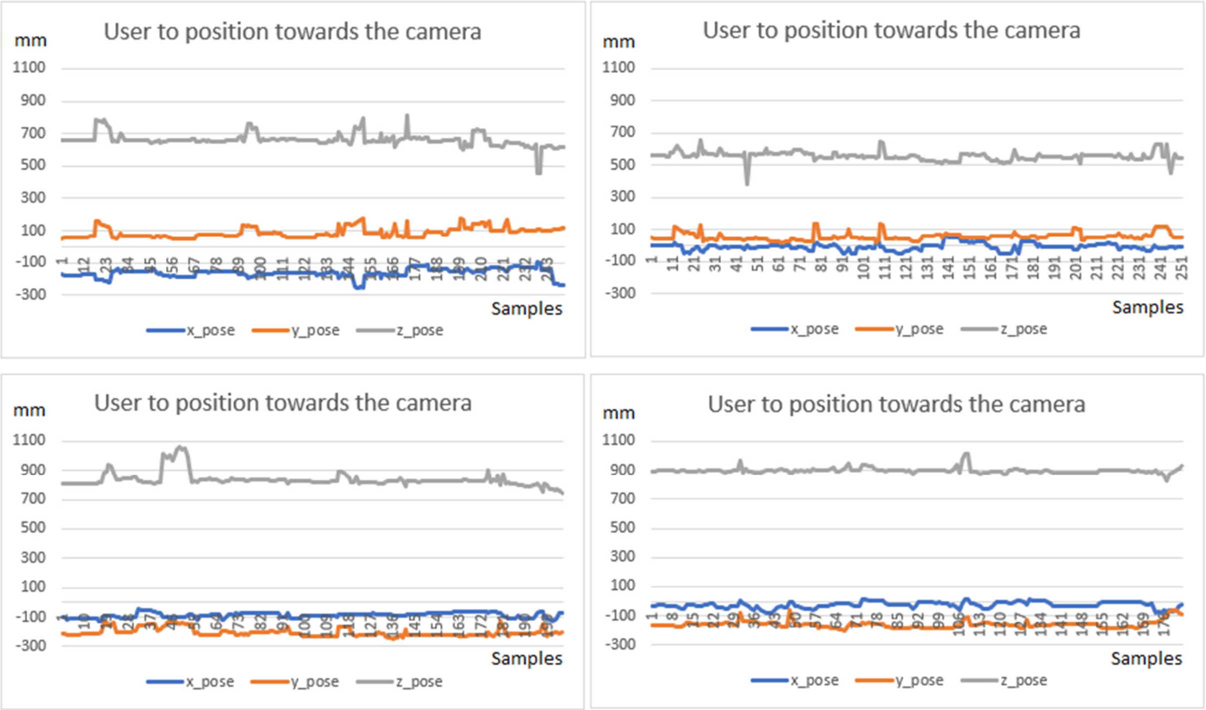


Figure 44 - User to position towards the camera extraction of the child B in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right). The x axis corresponds to the acquired samples. In y axis are registered the user position towards the camera (in mm).

Table 32 - Average and Standard deviation of child B in sesssion 1.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | -8.50 | 4.49 | 5.43 | 12.01 | 3.56 | -162.91 | 88.04 | 661.63 |
| **Standard deviation** | 8.12 | 5.77 | 8.97 | 4.97 | 5.28 | 26.39 | 31.17 | 38.99 |

Table 33 - Average and Standard deviation of child B in sesssion 2.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 4.82 | 4.08 | 4.99 | -1.94 | 4.30 | -7.56 | 56.23 | 557.50 |
| **Standard deviation** | 10.19 | 7.16 | 7.11 | 6.79 | 4.44 | 21.57 | 22.88 | 26.70 |

Table 34 - Average and Standard deviation of child B in sesssion 3.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 1.85 | 1.13 | 23.21 | 1.85 | 23.92 | -86.72 | -204.639 | 836.98 |
| **Standard deviation** | 8.39 | 7.53 | 8.95 | 6.58 | 5.51 | 17.81 | 27.46 | 50.48 |

Table 35 - Average and Standard deviation of child B in sesssion 4.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 10.80 | 1.75 | 14.03 | -6.57 | 15.88 | -25.36 | -159.13 | 898.43 |
| **Standard deviation** | 10.84 | 4.93 | 8.19 | 7.74 | 4.95 | 21.42 | 24.32 | 19.76 |

Child C:

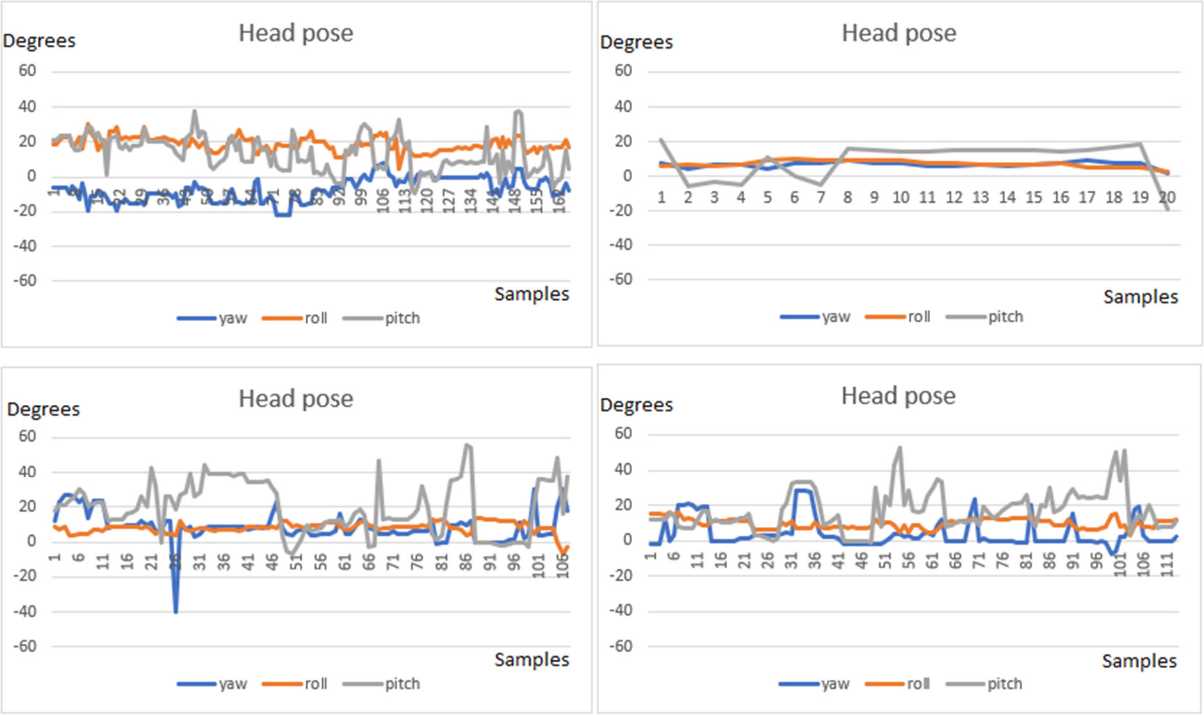


Figure 45 - Head pose extraction of the child C in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right).
The x axis corresponds to the acquired samples. In y axis are registered the angles of the head (in degrees), being the values near zero the vertical position of the head.

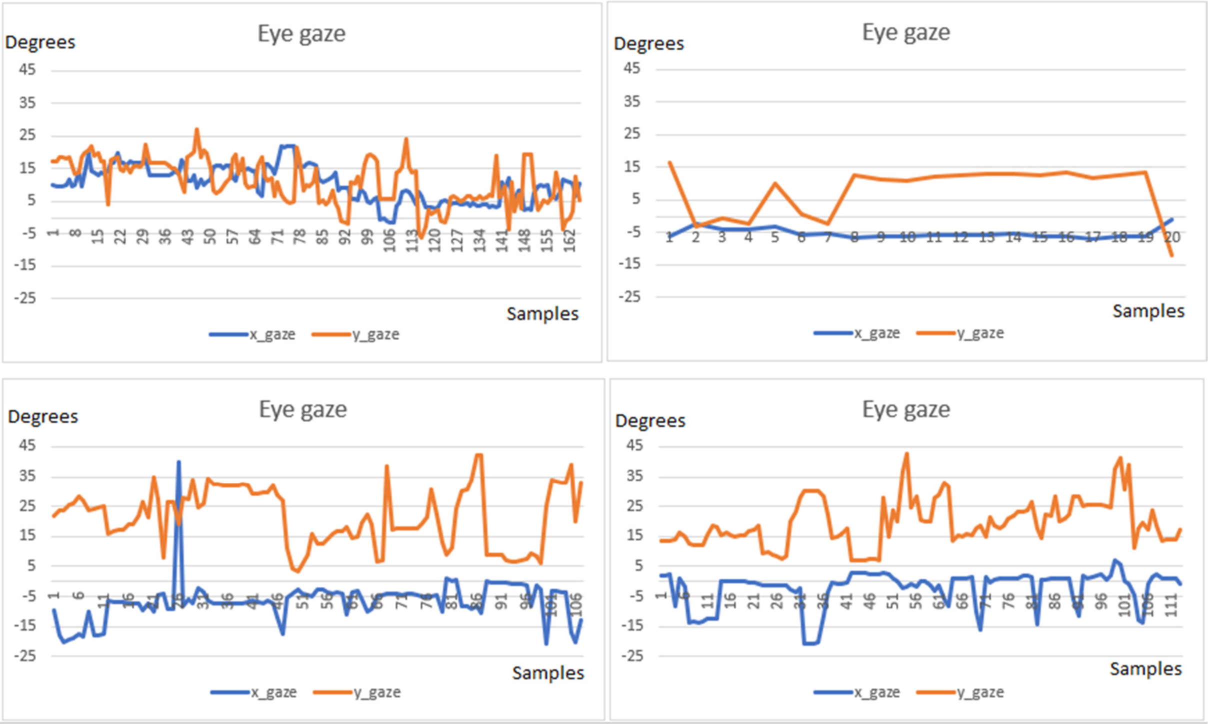


Figure 46 - Eye gaze extraction of the child C in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right). The x axis corresponds to the acquired samples. In y axis are registered the angles of the eye gaze (in degrees): x_gaze corresponds to horizontal ocular position and y_gaze corresponds to the vertical ocular position.

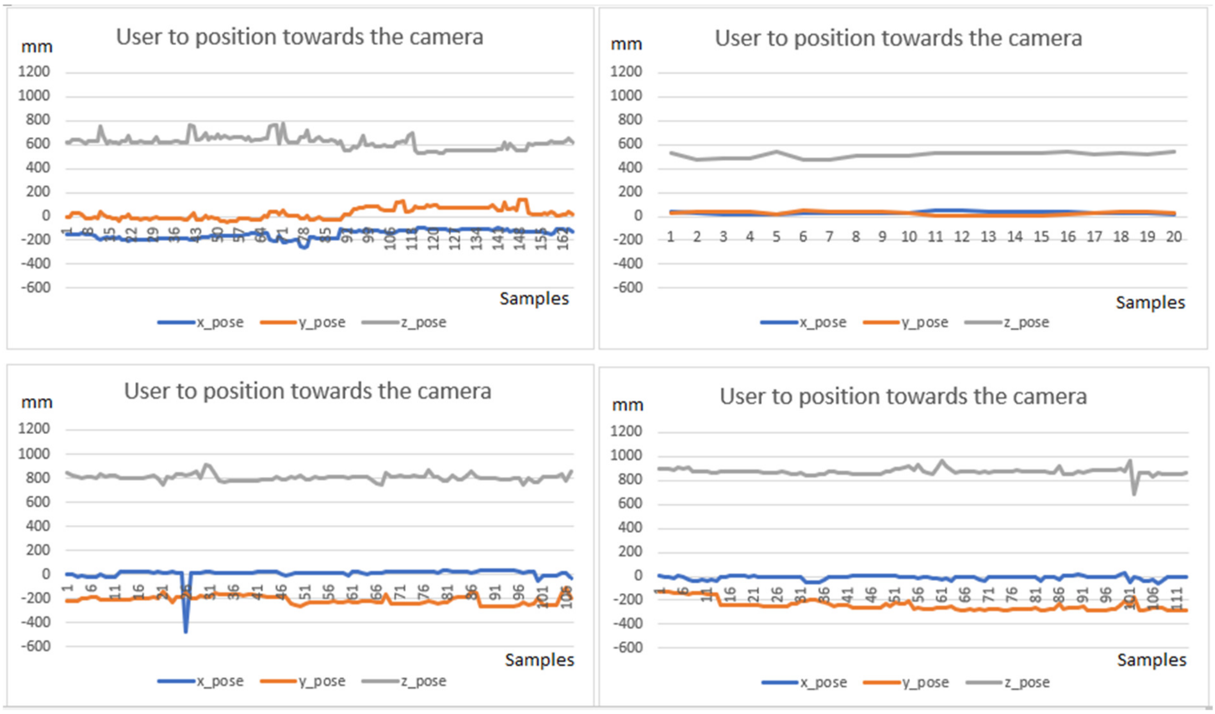


Figure 47 - User to position towards the camera extraction of the child C in session 1(upper left), session 2 (upper right), session 3 (down left), and session 4 (down right). The x axis corresponds to the acquired samples. In y axis are registered the user position towards the camera (in mm).

Table 36 - Average and Standard deviation of child C in sesssion 1.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | -7.06 | 18.63 | 13.25 | 10.42 | 10.93 | -150.56 | 22.93 | 617.53 |
| **Standard deviation** | 6.70 | 4.13 | 10.08 | 5.32 | 6.90 | 35.95 | 45.13 | 50.85 |

Table 37 - Average and Standard deviation of child C in sesssion 2.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 6.90 | 7.20 | 8.85 | -5.19 | 7.82 | 31.95 | 26.4 | 513.85 |
| **Standard deviation** | 1.80 | 1.79 | 10.90 | 1.53 | 7.92 | 9.54 | 14.85 | 22.68 |

Table 38 - Average and Standard deviation of child C in sesssion 3.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 8.76 | 8.30 | 20.01 | -6.56 | 21.58 | 9.93 | -210.12 | 805.81 |
| **Standard deviation** | 8.52 | 3.11 | 15.07 | 6.90 | 9.58 | 50.21 | 33.51 | 26.26 |

Table 39 - Average and Standard deviation of child C in sesssion 4.

| | Yaw | Roll | Pitch | X_gaze | Y_gaze | X_pose | Y_pose | Z_pose |
|---|---|---|---|---|---|---|---|---|
| **Average** | 4.12 | 9.73 | 16.73 | -2.16 | 19.58 | -9.97 | -242.30 | 874.05 |
| **Standard deviation** | 7.94 | 2.82 | 11.32 | 6.03 | 7.79 | 16.28 | 43.98 | 27.90 |