ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

SOCIAL BEHAVIOR LEARNING FOR AN ASSISTIVE COMPANION ROBOT

Ph.D. THESIS

Pınar ULUER

Department of Mechatronics Engineering

Mechatronics Engineering Doctorate Program

JANUARY 2023

ISTANBUL TECHNICAL UNIVERSITY ★ GRADUATE SCHOOL

SOCIAL BEHAVIOR LEARNING FOR AN ASSISTIVE COMPANION ROBOT

Ph.D. THESIS

Pınar ULUER
(518132005)

Department of Mechatronics Engineering

Mechatronics Engineering Doctorate Program

Thesis Advisor: Prof. Dr. Hatice KÖSE

JANUARY 2023

İSTANBUL TEKNİK ÜNİVERSİTESİ ★ LİSANSÜSTÜ EĞİTİM ENSTİTÜSÜ

YARDIMCI ROBOTLAR İÇİN SOSYAL DAVRANIŞ ÖĞRENİMİ

DOKTORA TEZİ

Pınar ULUER
(518132005)

Mekatronik Mühendisliği Anabilim Dalı

Mekatronik Mühendisliği Doktora Programı

Tez Danışmanı: Prof. Dr. Hatice KÖSE

OCAK 2023

Pınar ULUER, a Ph.D. student of ITU Graduate School student ID 518132005 successfully defended the thesis entitled “SOCIAL BEHAVIOR LEARNING FOR AN ASSISTIVE COMPANION ROBOT”, which he/she prepared after fulfilling the requirements specified in the associated legislations, before the jury whose signatures are below.

Thesis Advisor :	**Prof. Dr. Hatice KÖSE** Istanbul Technical University	....
Jury Members :	**Prof. Dr. Duygun EROL BARKANA** Yeditepe University	....
	Assoc. Prof. Dr. Gülay ÖKE GÜNEL Istanbul Technical University	....
	Prof. Dr. Sırma YAVUZ Yıldız Technical University	....
	Assist. Prof. Dr. S. Murat YEŞİLOĞLU Istanbul Technical University	....

Date of Submission : 26 December 2022
Date of Defense : 26 January 2023

To mom and dad,

FOREWORD

First, I hold the utmost respect and thanks to my advisor Prof.Dr. Hatice KÖSE for her supervision, support and guidance throughout this entire journey. She was present, honest and genuine with her support and feedback, and always made sure that we crafted the best plans to keep me on track when I was out of my mind and off balance. I also wish to express my sincere gratitude for her boundless patience in the development of this study as well as my professional and personal character.

I wish to express my sincere thanks to all the jury members for listening to me patiently through long meetings and presentations, and for their criticisms and insightful contributions to this study.

I am really grateful to all my friends for caring for me and adding more color, breath, optimism and energy to my life. But I owe so much to Nazlı Bülay DOĞAN a.k.a my sister from another mister, Günce Keziban ORMAN and Ozan ÇAĞLAYAN, for being there when I needed them the most, for listening to me as I rave and cry, for supporting me patiently, for tolerating all my obsessions and especially for making me laugh until I am out of breath.

I can't thank enough to my professors, colleagues, friends and fellow students at the university for their support and contributions throughout my academic journey.

Finally, but most importantly, I'd like to thank my parents for their unconditional support, encouragement, patience and boundless love, even if I am moody, grumpy and bossy all the time. I can never thank and love them enough for being there for me without a doubt.

January 2023 Pınar ULUER

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SOCIAL BEHAVIOR LEARNING FOR AN ASSISTIVE COMPANION ROBOT

SUMMARY

Designing robots having the ability to build and maintain interpersonal relationship with humans by exchanging social and emotional cues has attracted much attention recently because robots are vastly in use in a wide variety of places with a diverse range of tasks from healthcare to edutainment industry. In order to interact with humans in a natural way, emotion recognition and expressive behavior generation are crucial for these robots to act in a socially aware manner. This requires the ability to recognize the affective states, actions and intentions of their human companions. The robot should be able to interpret the social cues and its human companion's behavioral patterns to be able to generate assistive social behaviors in return.

There are several popularly known robotic platforms such as Kismet, Nao and Pepper which are able to recognize its human partner's emotional state based on facial expressions, vocal cues and body postures and express simple human emotions. However these robots do not have an emotional character or an affective state, they are just capable to mirror the human's expressions independently from the social context of the interaction. For the robots to be accepted as social entities, they should be endowed with the capability to interpret the human's mood as well as his/her emotions and the social context of the interaction. In order to achieve this purpose, it is not enough to treat expressive behaviors of humans as only a mirror of their internal state. Therefore, it is crucial to incorporate a generative account of expressive behavior into an emotional architecture. This requires perceiving and understanding others' affective states and behaving accordingly, it corresponds to the most generic definition of empathy.

Empathy is a focal point in the study of interpersonal relationships therefore it should also be considered as one of the major elements in the social interaction between the robots and humans. Humans have the ability to feel empathy not only for other humans but also for fictional characters and robots. But the robots are not able yet to display any empathic behavior in return.

The motivation of this thesis study is to design and implement a computational emotion and behavior model for a companion robot. The proposed model is designed to process multi-channel stimuli in order to interpret the affective state of its human companion and to generate in-situ affective social behaviors based on the processed information coming from the human companion and the environment, that is the social context of the interaction. This dissertation attempts to explore the following research objectives:

- Would the robot be able to display basic emotions? Could the human companion identify correctly the emotions displayed by the robot?

- How could a social robot infer and interpret its companion's emotional states?
- How can we computationally model an artificial emotional mechanism and implement it on a social robot to provide a natural social interaction?
- Which learning techniques should we use for the affective robot assistant to learn which emotional behavior to express during the interaction?
- Could a social robot designed with emotional understanding and expression foster the interaction gain in a HRI scenario based on assistance?

In order to explore the answers of these research questions, user studies with children having different developmental profiles and two affective robots, Pepper and Kaspar, were conducted in coordination with two research projects.

The first project, titled as RoboRehab, we aimed to use an affective social robot to support the auditory tests of children with hearing disabilities in clinical settings. During their hospital visits, children take several tests to determine the level of their hearing and to adjust their hearing aid or cochlear implant, if necessary. The audiologists mention that the children usually get stressed and tend to be in a negative mood, when they are in the hospital during their consultation. This affects children's' performance in the auditory perception tests negatively, and their cooperation decrease significantly.

In Roborehab, we used machine learning-based emotion recognition approaches to detect the children's emotional state and adjust the robot's actions accordingly. We designed a feedback mechanism to reduce the stress of children and help the children to improve their mood during the audiometry tests. We used a socially assistive humanoid robot Pepper, enhanced with emotion recognition, and a tablet interface, to support children in these tests. We investigated the quantitative and qualitative effects of the different test setups involving a robot, a tablet app and the conventional method. We employed traditional machine learning techniques and deep learning approaches to analyze and classify the physiological data (blood volume pulse, skin temperature, and skin conductance) of children collected by E4 smart wristband.

The second project, entitled as "Affective loop in Socially Assistive Robotics as an intervention tool for children with autism (EMBOA)", was a research project with the aim of combining affective computing technologies with the social robot intervention in children with autism spectrum disorder (ASD). Children with ASD are known to display limited social and emotional skills in their routine interactions. Inspired by the promising results presented in the social robotics field, we aimed to investigate affective technologies with a social robot, Kaspar, to assist children with ASD in the development of their social and emotional skills, help them to overcome social obstacles and make the children more involved in their interactions. Interaction studies with Kaspar were conducted in 4 collaborating country; Poland, North Macedonia, United Kingdom and Turkey; with more than 65 children with ASD and interaction data collection by different sensor modalities (visual, audio, physiological and interaction-specific data) was performed within a longitudinal study design.

In this dissertation, a computational model for emotional understanding and emotional behavior generation and modulation was designed and implemented for a companion robot based on the collected data and findings through The RoboRehab and EMBOA projects. The presented models were designed: (1) to process multi-channel stimuli, i.e. vision-based facial landmarks, physiological data-based signals, in order to detect the affective states of the human companion; (2) to generate in-situ affective social behaviors for the robot based on the interaction context; (3) to adapt the intensity of the robot's emotional expressions based on the preferences of the human companion.

Despite challenging situations, with Covid-19 outbreak at the top of the list, and computational limitations, the previously mentioned research objectives were investigated in detail and the results were presented in this dissertation. We were able to answer the five research questions on the affective social behavior of a companion robot.

The user studies conducted with hearing-impaired children in RoboRehab showed that Pepper robot was able to display emotional behaviors and the children could correctly identify and interpret them. Moreover, the RoboRehab findings revealed that the affective robot was able to trigger some emotional changes in children and cause difference in their physiological signals. These signals were used to distinguish the emotions of children with machine learning and deep learning approaches in different setups. On the other hand, the results from the EMBOA studies with children with ASD showed that a multimodal approach based on behavioral metrics, i.e. facial action units, physiological signals, interaction-specific metrics, can be used to understand the emotional state and infer the emotional model of the human companion.

Different frameworks were presented in order to model an artificial emotional mechanism for a social robot. The first one was a simple linear affect model based on the two-dimensional valence and arousal representation. The second one, a vision-based model where the robot detects the affective state of its companion and develops a behavioral strategy to adapt its emotional behaviors and improve the negative mood of its companion. And finally, a behavioral model were presented for the robot to predict the preferences of its companion and regulate the intensity of its emotional behaviors accordingly. The results of the user studies and the findings revealed that an emotional model can be computationally modeled and implemented for a social robot to adapt its emotional behaviors to the preferences and needs of its companion, consequently, to make the interaction between the human companion-robot pair more natural.

The RoboRehab and EMBOA user studies with different groups of children (with typical development, with hearing impairment, with autism) showed that independently from their profile the children were more involved and paid attention to the social robot. Even though the objective evaluation metrics did not point a significant difference, the subjective evaluation of the audiologists, therapists, pedagogues complied with the presented hypothesis. Furthermore, the self-report surveys with children and their parents showed that the children accepted the affective robot as an intelligent and funny social being. These results demonstrate that an affective social robot with personalized emotional behaviors based on the preferences of its companion can foster the interaction gain in an assistive human-robot interaction scenario.

YARDIMCI ROBOTLAR İÇİN SOSYAL DAVRANIŞ ÖĞRENİMİ

ÖZET

Robotların sağlık, eğitim, eğlence ve kamusal alanlar gibi insan merkezli ortamlarda varlık göstermeye başlamasıyla beraber, insan-robot etkileşimi çalışmaları çevreleriyle doğal bir şekilde etkileşim kurabilecek robotların tasarımına odaklanmıştır. Bu alanda yapılan araştırmalar sonucunda, robotların insanlarla etkileşimlerinin daha doğal bir akışta gerçekleşebilmesi için bu etkileşime sosyal ve duygusal nitelikler katılması gerektiği ortaya çıkmıştır.

Sosyal ve duygusal durumlarını gösteren ipuçları kullanarak insanlarla birebir ilişki kurabilen ve zaman içerisinde bu ilişkiyi sürdürebilen robotlar tasarlamak oldukça zorlu bir süreçtir. Bu durum robotların bir nevi sosyal farkındalık göstermesini gerektirir, yani, robotun karşısındaki kişinin duygudurumunu tespit edip onun tutum ve davranışlarını algılaması ve kendi duygudurumunu buna göre ayarlayıp buna uygun duygusal davranışları sergilemesi anlamına gelir. Robot, karşısındaki insanın duygudurumunu yüz ifadelerinden, kullandığı sesli ipuçlarından ve vücut duruşundan tespit edip onun davranış kalıplarını yorumlayabilmeli ve gerektiği durumlarda bu kişiye destek olurken sosyal ve duygusal ipuçları kullanabilmelidir.

Kismet, Nao ve Pepper gibi bazı robotlar karşısındaki insanın duygudurumunu tespit ederek karşılığında bazı basit duyguları temsil edebilmektedir. Ancak bu robotlar duygusal bir karaktere veya belli bir duygudurum modeline sahip olmaksızın, etkileşimin sosyal bağlamından bağımsız olarak insanın ifadelerini birebir taklit ederler. İnsanların davranışlarını içsel durumlarının bir aynası olarak ele alıp bunu taklit etmek yeterli değildir, robotların sosyal varlıklar olarak kabul edilebilmesi için, insanın ruh halini, duygularını ve etkileşimin sosyal bağlamını yorumlayabilme yeteneğine sahip olmaları gerekir. Bunun için de robotun belli ölçüde duygusal bir mimariye sahip olması ve duygudurumlarını buna göre dışavurması gerekir. Bu da robotun, insanın duygudurumu algılaması, anlaması, yorumlaması ve buna dayanarak kendi duygudurum modelini oluşturması ve bu modele bağlı olarak duygusal davranışlar üretmesi demektedir ki, bu da empatinin en genel tanımlarından biridir.

Empati, bireyler arası karşılıklı kişisel ilişkilerin incelenmesinde önemli bir odak noktasıdır, bu nedenle robotlar ve insanlar arasındaki sosyal etkileşimin ana unsurlarından biri olarak ele alınması gerekmektedir. İnsanlar sadece diğer insanlar için değil, aynı zamanda kurgusal karakterler ve robotlar için de empati hissetme yeteneğine sahiptir. Ancak buna karşılık robotlar, henüz herhangi bir empatik davranış sergileme becerisine sahip değildir.

Bu tez çalışmasının amacı, yardımcı sosyal robotlar için bir duygudurum ve davranış modelinin tasarlanması ve bu modelin robotik bir platform üzerinde uygulanmasıdır.

Önerilen model, robotun farklı sensörlerden gelen çok kipli verileri işleyerek karşısındaki insanın duygusal durumunu yorumlayabilmesi ve bu insanın duygudurumundan ve sosyal etkileşim bağlamından gelen verileri sentezleyerek tutarlı ve yerinde duygusal davranışlar üretebilmesi için tasarlanmıştır. Bu tez çalışması aşağıdaki soruları cevaplandırabilmek amacıyla kurgulanmıştır:

- Bir robot mutluluk, üzüntü, vb. temel duyguları doğru bir şekilde temsil edebilir mi? Etkileşimde bulunduğu insan, robotun sergilediği bu duyguları doğru bir şekilde tanımlayıp yorumlayabilir mi?
- Sosyal bir robotun, karşısındaki kişinin farklı duygusal durumlarını ve davranışlarını tespit edip yorumlayabilmesi için nasıl bir yol izlenmelidir?
- Yardımcı robotlar için yapay bir duygu mekanizması nasıl modellenmelidir? Robot ile insan arasında doğal bir sosyal etkileşim kurulmasını sağlamak için bu mekanizma bir sosyal robota nasıl entegre edilebilir?
- Duygusal ve sosyal bir robotun etkileşim esnasında hangi duygusal davranışı ifade edeceğini öğrenmesi için hangi öğrenme tekniklerini kullanmak gerekir?
- Yapay bir duygu mekanizmasına ve temel duyguları ifade edebilme yeteneğine sahip sosyal bir robot, yardım üzerine kurgulanmış bir etkileşim senaryosunda kişinin etkileşimden sağladığı kazancı artırabilir mi?

Bu soruların cevaplarını araştırmak ve yardımcı sosyal robotun duygudurum modelini tasarlamak amacıyla, iki araştırma projesi kapsamında farklı gelişim profillerinden çocuklarla ve iki robotla, Pepper ve Kaspar, kullanıcı çalışmaları yapılmıştır. Bu çalışmalar esnasında farklı sensörler kullanılarak robotla etkileşime giren kişilerden yüz verisi, fizyolojik veri ve etkileşime özel veriler toplanmıştır.

İlk proje olan RoboRehab'in amacı klinik ortamda işitme engelli çocuklara işitsel algı testlerinde destek olmak için kullanılacak duygusal ve sosyal bir robotik platform tasarlamaktır. Odyoloji kliniklerinde, cocukların işitme düzeylerini tespit etmek ve kullandıkları işitme cihazlarını / koklear implantlarını buna göre ayarlamak için çeşitli testler uygulanmaktadır. Önceki çalışmalarda odyologlar, hastaneye konsültasyona gelen çocukların genellikle stresli ve olumsuz bir ruh halinde olduğunu gözlemlemişlerdir. Bu durumun çocukların işitsel algı testindeki performanslarını olumsuz etkilediğini ve test esnasında çocukların işbirliğini önemli ölçüde azalttığını tespit etmişlerdir. Bu proje kapsamında çocukların duygularını tespit edebilmek için makine öğrenmesi tabanlı duygu tanıma yaklaşımları kullanılmıştır. Ayrıca odyoloji testleri sırasında çocuğun stresini azaltmak ve olumsuz ruh halinden kurtulmasını sağlamak adına robot için duygusal davranışa dayalı bir geri bildirim mekanizması tasarlanmıştır. Bu sayede çocukların test performanslarının iyileştirilmesi ve odyologlarla yaptıkları klinik çalışmalarda motivasyonlarının yükselmesi hedeflenmiştir.

RoboRehab projesi kapsamında yürütülen etkileşim çalışmalarında, işitsel algı testlerinde çocuklara yardımcı olması için bir tablet oyunu ve sosyal robot Pepper kullanılmıştır. Robot, tablet uygulaması ve standart yönteme dayalı farklı test kurulumlarının etkileri nicel ve nitel olarak incelenmiştir. Çocukların duygudurumlarının yüz ifadelerine ve fizyolojik sinyallerine (kan hacmi nabzı, cilt sıcaklığı ve cilt iletkenliği) yansımasını analiz etmek ve sınıflandırmak için geleneksel makine öğrenimi teknikleri ve derin öğrenme yaklaşımları test edilmiştir.

“Otizmli çocuklar için bir müdahale aracı olarak sosyal yardımcı robotlarda duygusal döngü (EMBOA)” başlıklı ikinci araştırma projesinin amacı, otizmli çocukların rehabilitasyon sürecinde kullanılan sosyal robotları ve duyguya dayalı bilişim teknolojilerini birleştirmektir. Otizm spektrum bozukluğu (OSB) olan çocukların, rutin etkileşimlerinde sergiledikleri sosyal ve duygusal becerilerin sınırlı olduğu bilinmektedir. Alanyazında yer alan çalışmalar sosyal robotların, otizmli çocukların sosyal ve duygusal becerilerini geliştirmelerine ve sosyal engellerin üstesinden gelebilmelerine yardımcı olmak, çocukları etraflarındaki bireylerle etkileşimde bulunmaya teşvik etmek için kullanıldığı ve bu çalışmalardan umut verici sonuçlar elde edildiği görülmüştür. Bu sonuçlardan esinlenilerek, proje ortağı olan 4 ülkede (Polonya, Makedonya, İngiltere ve Türkiye) 65’ten fazla OSB’li çocuk ve Kaspar robot ile kullanıcı çalışmaları gerçekleştirilmiştir. Zaman içine yayılan ve birden fazla oturumda tekrarlanan bu çalışmalar esnasında çocuklardan çok kipli etkileşim verileri (yüz verisi, fizyolojik sinyaller ve etkileşime özel ölçümler) toplanmıştır.

Bu tez çalışmasında, RoboRehab ve EMBOA projeleri sayesinde toplanan verilere ve bulgulara dayalı olarak yardımcı bir robot için duygusal anlayış mekanizması modellenmiş, ve robotun hangi duygusal davranışı ne zaman ne şiddette sergilemesi gerektiği belirlenmiştir. Önerilen duygusal davranış modeli: (1) Etkileşimdeki insanın duygusal durumlarını tespit etmek için çok kanallı uyaranları, yani kamera ile tespit edilen yüz verilerini, fizyolojik bileklik ile alınan sinyalleri işler; (2) etkileşim bağlamına uygun ve yerinde duygusal sosyal davranışlar oluşturur; (3) robotun duygusal ifadelerinin yoğunluğunu, etkileşimde olduğu insanın tercihlerine göre kişiselleştirir ve ayarlar.

Çalışma esnasında karşılaşılan zorluklara, en başta Covid-19 pandemisi olmak üzere, ve kabul edilen kısıtlamalara rağmen, yukarıda bahsi geçen araştırma soruları detaylı olarak incelenmiş ve cevapları aşağıda açıklanmıştır.

RoboRehab projesi kapsamında yürütülen kullanıcı çalışmalarında, Pepper robotun duygusal davranışlar sergileyebildiği ve işitme engelli çocukların bunları doğru bir şekilde tanımlayıp yorumlayabildiği görülmüştür. Ayrıca, RoboRehab projesinde elde edilen sonuçlar, duygusal davranışlar sergileyen robotun çocuklarda bir takım duygusal değişiklikler tetiklediğini ve fizyolojik sinyallerinde farklılığa neden olduğunu göstermiştir. Buna bağlı olarak etkileşim esnasında çocuklardan toplanan bu sinyaller, makine öğrenmesi ve derin öğrenme yaklaşımları ile çocukların duygularını ayırt etmek için kullanılmıştır. Öte yandan, EMBOA projesi kapsamında otizmli çocuklarla yapılan çalışmalarda elde edilen sonuçlar, çok kipli davranışsal ölçümlere; yani yüz eylem birimlerine, fizyolojik sinyallere, etkileşime özgü ölçümlere; dayalı bir yaklaşımın çocukların duygusal durumu anlamak ve modellemek için kullanılabileceğini göstermiştir.

Bu bulgulara dayanarak tez çalışmasında sosyal bir robot için yapay duygudurum mekanizması modellenmiş ve farklı modelleme yaklaşımları sunulmuştur. Bunlardan ilki, duyguların iki boyutlu değerlik ve uyarılma gösterimine dayanan basit bir doğrusal duygulanım modelidir. İkincisi, robotun karşındaki insanın yüz verilerine dayanarak duygusal durumunu tespit ettiği ve kendi duygusal davranışlarını buna göre uyarlayarak arkadaşının olumsuz ruh halini iyileştirmek için davranışsal bir strateji geliştirdiği bir modeldir. Son olarak robotun, karşısındaki insanın tercihlerini tahmin etmesi ve buna göre duygusal davranışlarının yoğunluğunu düzenlemesi için bir davranış modeli sunulmuştur. Sosyal robotun hangi durumda nasıl davranması gerektiği üzerine 500 kullanıcı ile yapılan bir anket çalışması sonucunda toplanan veriler incelenmiştir. Deneysel sonuçlar, tasarlanan duygu modeli ile bir sosyal robotun duygusal davranışlarını karşısındaki insanın tercihlerine ve ihtiyaçlarına göre uyarlamasının mümkün olduğunu ve bu sayede aralarındaki etkileşimin daha doğal ilerlediğini göstermiştir.

Gelişimsel olarak farklılık gösteren (tipik gelişim gösteren, işitme engelli, otizmli) çocuklar ile yapılan RoboRehab ve EMBOA çalışmaları sonucunda, çocukların profillerinden bağımsız olarak sosyal robota daha fazla ilgi gösterdikleri ve onunla daha fazla ilgilendikleri görülmüştür. Objektif ölçümlere dayalı değerlendirmeler her ne kadar istatistiksel açıdan anlamlı bir farklılığa işaret etmese de odyologların, terapistlerin, ve pedagogların sübjektif değerlendirmeleri sunulan hipotezle uyumluluk göstermektedir. Ayrıca, çocuklar ve aileleri ile yapılan kişisel bildirime dayalı anketlerde, çocukların duygusal robotu zeki ve komik bir sosyal varlık olarak kabul ettikleri görülmüştür. Bu sonuçlar, etkileşimde olduğu insanın tercihlerine göre kişiselleştirilmiş duygusal davranışlara sahip bir sosyal robotun, yardıma dayalı bir etkileşim senaryosunda etkileşim kazancını iyileştireceğini kanıtlamaktadır.

1. INTRODUCTION

As robots make their presence known in human interactive environments, such as health, education, entertainment and social domains, we require methods to design robots that can interact with their environment in a social way. This shift in the robotics studies bring forward the necessity to design robots using empathy as a social cue. Designing robots having the ability to build and maintain interpersonal relationship with their human companions has recently received considerable attention. Because robots, especially social ones, have become more present and apparent in a wide variety of places with a diverse range of tasks. In order to interact with humans in a natural way by exchanging social and emotional cues, emotion recognition and expressive behavior generation are crucial for these robots to be able to act in a socially aware manner. Therefore robots should be able to interpret their human companion's social cues and behavioral patterns to be able to generate assistive social behaviors in return. This requires the recognition of the affective states, actions and intentions of the human companion.

There are several popularly known robotic platforms such as Kismet, Nao and Pepper which are able to recognize its human partner's emotional state based on facial expressions, vocal cues and body postures and express simple human emotions. However these robots do not have an emotional character or an affective state, they just mirrors the human's expressions independently form the social context of the interaction. If we introduce this emotional transparency into the interaction without implementing any emotional understanding on the robot, then the interaction gain of the human companion would be at risk depending on the context. For example, let's consider an interaction scenario between a deaf student and a social robotic companion assisting him/her in sign language practice. After a rough day in school, the student comes home with a negative mood and starts to practice sign language with the robot companion. The student makes a lot of mistakes through the study, and the robot companion without any affective understanding displays its standard behavior, mirrors the mood of the student and imposes the student for correcting his/her mistakes. Consequently, the perception of the student about the robot would be impaired.

For the robots to be accepted as social entities, they should be endowed with the ability to interpret the human companion's mood as well as his/her emotions and the social context. In order to achieve this purpose, it is not enough to treat expressive behaviors of humans as only a mirror of their internal state. Therefore, it is crucial to incorporate a generative account of expressive behavior into an emotional architecture. This requires perceiving and understanding others' affective states and behaving accordingly, it corresponds to the most generic definition of empathy. Let's go back to the same scenario between the student and the robot, but this time the robot has the ability of empathic understanding. The robot is able to detect the negative mood of its human companion and perceives that he/she is having a difficult day. The robot adapts its responses and feedback accordingly and motivates the student to increase the educational gain of the interaction, thus helping the student to "lift his/her mood" through the interaction.

Empathy is a focal point in the study of interpersonal relationships therefore it should also be considered as one of the major elements in the social interaction between the robots and humans. Humans have the ability to feel empathy not only for other humans but also for fictional characters and robots. But the robots are not able yet to display any empathic behavior in return. For the robots to be accepted as empathic entities, they should have the ability to build and maintain interpersonal relationship with humans by exchanging social and emotional cues and the ability and to interpret the human's affective states as well as his/her moods, emotions based on the interaction context. Following this point of view, the researchers have demonstrated the benefits of empathy in the design of robots' behaviors.

The next section incorporates the related studies on emotional theories, computational models of emotions/empathy, emotional behavior learning and control, and evaluation metrics to asses the interaction between an affective robot and its human companion. And then the motivation and objectives of this study are detailed in the next sections.

1.1 Literature Review

Human-robot interaction (HRI) and collaboration have been popular research fields since the robots have started to be a part of our daily life. As they moved out from the controlled laboratory environments towards human-centric everyday life scenes, the need to establish an interaction in a more 'natural' way has increased. In a detailed overview of human-robot interaction studies, Fong et al. [1] attempt to explore what makes the interaction between a robot and a human desirable and they conclude that the robot should obtain social skills similar to humans' skills. In another study, Goodrich and Schultz [2] also remark that the robots should display and generate more human-like motions to improve the interaction quality and make the interaction establishment easier between the communicators. While the concept of socially interactive robots has spread in the literature, the idea of the collaboration with a robot partner has arisen. For a detailed work on the characteristics of a possible collaboration between a human and a robot, one may refer to the study of Bauer et al. [3]. In their study, the authors describe the state of the art of the human-robot collaboration and discussed the limitations of it. Apart from collaboration, the social robotics field incorporates another component, socially assistive robotics (SAR). SAR, being an interdisciplinary research area, focuses on helping human users through social interaction. Tapus et al. [4] review the principle application domains of SAR, as well as the key concept of SAR studies: embodiment, personality, empathy, engagement and adaptation, and defines the field's grand challenges.

Recent studies in socially assistive robotic systems enable the use of robots in many application areas such as healthcare, education, and elderly care, as a part of our daily lives. When people see robots in human-robot interactive social environments, they expect a level of socially intelligent and socially aware behavior based on the robot's capabilities and human-robot interaction settings.

In this sense, Cifuentes et al. [5] explore the implications of social robots in healthcare scenarios, and they conduct a review study on the applications of social robots, how they were perceived and accepted by children and adults. Their review reveal that adults and children who were exposed to an intervention with social robots, showed signs of improvements in social connection and communication and their current mood, and they displayed less signs of depression, anxiety and fear. Similarly, Kabacinska et al. [6] present a literature survey on how social robots were used as means to support mental health in children. The study outcomes suggest that interventions with social robots had positive impact on stress relief and improved positive affect level.

Ferrante et al. [7] explore how social robots can be used to increase therapeutic adherence in pediatric asthma based on the assumption that the robots increased the motivation of children and their engagement to treatment. Although they present a conceptual study, the authors investigate the available robotic platforms, the nature of interaction and the intervention, and they discuss the feasibility, acceptability and efficacy of socially assistive robots in clinical practice.

Logan et al. [8] propose to use social robots as engaging tools to address the emotional needs of hospitalized children. They introduced an interactive teddy bear robot, Huggable, into the inpatient setting of a pediatric clinic to improve engagement of hospitalized children and to reduce their stress during their intervention. They conduct a pilot study with 54 children (3 to 10 years old), the children were exposed to an intervention with a teddy bear social robot, a tablet-based avatar of it and a plush teddy bear accompanied by a human presence. The pilot study results demonstrate that the children exposed to the social robot intervention showed more positive affect and they expressed greater levels of joyfulness and agreeableness than the other conditions.

In another study, Moerman and Jansens [9] use a baby dinosaur pet robot with children in a hospitalized setting to distract them and to improve their well-being through their interactions with the robot. The authors conduct a preliminary study with 9 children (aged 4-13 years old) who received personalized therapeutic sessions with the robot. The results show that the children expressed less boredom, anxiety and stress and displayed more active and playful behavior based on the behavioral analysis of experts and the interviews with the parents.

Socially assistive robots can be personalized and adapted to the specific requirements to increase the user performance and to improve the interaction gain with robots. Besides, most of the users are inexperienced in terms of robots, and might have special needs such as elderly, people with hearing disabilities or children with Autism spectrum disorder (ASD) [10]–[12].

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder that is associated with persistent deficits in the socio-communicative skills. The mean age of diagnosis of ASD is around 4 to 5 years of age [13]. Limited socio-communicative skills and deficits in the recognition and expression of emotions influence the ability of children with ASD to interact and communicate with others, affecting their relationships with family members, peers, and therapists.

There are promising results in the use of robots in supporting the social and emotional development of children with autism. Using robots as social mediators to engage children in tasks, allows for predictable and reliable environment, which is an important prerequisite in promoting prosocial behaviors. Social robots proved to be a way to get through the social obstacles of the children with ASD and make them involved in the interaction. Once the interaction is established, it presents a unique opportunity to engage the children in building and practicing social and emotional skills.

Clabaugh et al. present a 30-day long in-home pilot study with a fully autonomous SAR and 2 young male siblings [14]. The authors report their observations from the month-long interaction between the robot SPRITE and 2 children, a 6-year-old diagnosed with ASD and a 5-year-old child with typical development. The preliminary results show that both children interacted with the robot voluntarily (15 times in 30-days for average 10-min periods) and they reported that they would like to have the robot longer, showing that the setup is promising for long-term engagement.

In another study, Javed and Park propose a framework for an emotionally expressive agent (Romo robot with a penguin character interface) to implement empathy-based emotion regulation in order to engage the children with ASD in turn-based emotional interactions [15]. They conduct a user study with 17 children (4-10 year-olds): 11 TD children (4-10 year-olds) and 6 children with ASD (7-10 year-olds). The authors report that the results confirmed the feasibility of the proposed framework for both groups of children and the longer interactions in the emotion game triggered higher engagement levels in other activities based on sensory skills.

Li et al. propose to use a teacher-teleoperated Zeno robot and tablet system combined with various learning activities, to help teach facial emotions to minimally verbal autistic children [16]. The results show that the children were highly engaged and they payed more attention to the robot's behaviors/expressions than its verbal utterances but although they were highly interested with the robot, they were not interested with the learning activities. Hence, the importance of the interaction scenario design.

In brief, the interaction with socially assistive robots can be personalized and adapted to the needs and requirements to increase the children's engagement and to improve its quality and gain.

Socially assistive robots are designed to provide active assistance to their human companions. In order for this assistance to be effective, these socially assistive robots should be able to display appropriate emotions/behaviors and in-situ motions and to adapt/adjust these abilities with respect to the human companion's needs. Hence, the inspiration to design computational model of emotions. Emotions have an important part in social interaction. Gratch and Marsella [17] suggest that models of emotion may give insight into building models of intelligent behavior for social robots and the effort to accurately model emotional mechanisms may have a major contribution on the human understanding of emotions.

1.1.1 Emotional theories

Human Theory of Mind (ToM) is defined as the ability to infer the full range of mental states (i.e. emotions, intents, beliefs, drives, etc.) of oneself and others [18]. It is a mechanism helping us to understand the other individuals having different mental states and make sense of their behavior in specific context to predict their next behavior. Hence, it is crucial in human social interactions [19].

Aylett et al. [20] point that ToM provides the researchers a starting point to model the mental states of an embodied agent about another and it is considered to be the minimum requirement to consider the impact of one's own expressive behaviors on someone else.

Inspired by this idea, the researchers aiming to represent the mental states of embodied agents propose different methodologies to model and understand the ToM of both sides taking part in HRI scenarios. These methodologies are mostly based on emotional understanding or empathy because it is deemed as indispensable for communication [21].

1.1.1.1 Basic emotion theory

The basic emotion theory is built on the idea that there is a basic/primary set of emotions and these emotions are evolved through generations because they make the adaptation easier [22]. These basic emotions are labeled as “anger", “fear", “disgust", “joy", “sorrow" and “surprise". From an evolutionary-developmental perspective, Ekman and Oster [23] indicate that each emotion in the set has a particular function in a given context: providing adaptivity to the environment. The authors remark that the basic emotions also serve as behavior learning reinforcers, i.e. they play an important role in acquiring new behaviors.

Although theorists studying on basic emotions report that social experience has an important role in emotional development of an individual, the authors take an evolutionary perspective and mostly ignore the cognitive aspects of emotions in their emotion generation theories.

1.1.1.2 Cognitive appraisal theory

Several theorists argue that emotions have evolved just as an adaptation mechanism based on environment. Consequently, they claim that there is an appraisal system assessing the perceived environmental factors with respect to the individual's well-being, its plans and goals [24,25]. Scherer [26] suggests that this appraisal system requires a high level of cognition because people tend to appraise events with respect to a large variety of factors, such as novelty, pleasantness, goal significance, coping mechanism, etc.

These appraisals evoke a particular emotional state that requires some sort of adjustment such as physiological changes, behavioral responses and expressions. These adjustments are done with the aim of reaching to a stabilized condition. Plutchik [27] names this stabilizing condition as “behavioral homeostasis". Through this process, emotions establish a link between the individual and the environmental factors, pulling the individual toward a certain stimuli or pushing it from others.

Following the cognitive appraisal theory, Ortony et al. [28] make a connection between emotions and the individual’s evaluation of the environment. The authors define emotions as reactions to consequences of events, actions of agents, and aspects of objects. The authors propose that an individual judges the desirability of an event (pleasantness), the approbation of an action (conformity) and the attraction of an object (liking). These factors are called appraisal variables shaping the generated emotions.

On the other hand, Lazarus [29] presents a motivational cognitive theory of emotion, he proposes that emotions emerge from the appraisal of the interaction between the individual and environment. According to him, there are two processes to reach the behavioral homeostasis for the individual: 1) a cognitive evaluation (appraisal) process permitting to conserve or modify the relationship between the agents and their goals, 2) an adaptation (coping) process including cognitive and behavioral efforts to manage internal or external demands with respect to the resources of an individual.

Another theorist, Scherer [30] defines emotions as a multi-component process having:

- a cognitive component, processing and evaluating the stimuli
- a neurophysiological component, assessing the change in the internal state
- a motivational component, providing action preparation
- a motor component, expressing the behavior and emotion
- subjective feeling

He posits an appraisal process based on the evaluation of the stimuli. This process evaluates the novelty, pleasantness and compatibility of the stimuli as well as its congruence with the individual’s goals and the coping potential.

1.1.1.3 Dimensional theories

Dimensional theories argue that emotion and affective states may be conceptualized and represented in a continuous dimensional space. They provide a numerical representation of emotions in a space defined by 2 or 3 dimensional axes.

The most popular dimensional theory is the one proposed by Mehrabian and Russell [31], their “PAD" model is constructed based on three dimensions corresponding to:

- Pleasure, representing the valence of the affect,

- Arousal, representing the level of affect activation,

- Dominance, representing the notion of power or control.

Computational dimensional models are mostly used for behavior generation, because they can represent emotions in a restricted number of dimensions which can be mapped to some behavioral features. They are used as the fundamental framework for systems designed to recognize human emotional behavior. Therefore, most studies on the emotion modeling are based on the PAD model variations. Although the 3D model do not directly provide any basis to generate expressive behaviors, it is possible to link the mathematical representations of emotions on PAD space and some action units, such as the Facial Action Coding System (FACS) proposed by Ekman [32], to generate an affective response in an embodied agent.

Example studies includes the affective robot Kismet [33] using a 3D emotional model based on arousal, valence and stance. Although named differently, the dimensions represent the same notions and make easier the human emotion recognition and robotic expression generation. In another study, Lim and Aylett [34] use only a 2D model based on pleasure and arousal to control the generation of emotional face in a virtual agent. Another study using the two dimensions of arousal-pleasure axis, which is part of the circumplex model of affect proposed by Russell [35], is the study of Watanabe et al. [36]. But in this study, the authors follow another method to model the emotional development of the robotic agent based on a fundamentally different approach: the cognitive developmental robotics [37].

1.1.1.4 Cognitive developmental paradigm

The developmental paradigm proposes that one physical entity is able to develop its own emotions and behaviors through interactions with its environment and other agents [21]. Inspired by the role played by caregivers during the infant's cognitive, social and skill development, the researchers working on this field propose to use intuitive parenting and mimicry to model and generate emotional mechanisms for social humanoid robots.

One of the first studies based on the developmental paradigm propose to use intuitive parenting [36]. Intuitive parenting may be defined as a scaffolding mechanism where the infant empirically learns to associate their internal states to the expressions of their caregiver who is mimicking the infant's expressions. Watanabe et al. [36] propose to use the changes in the sensory information to model the internal state of the robot: When the robot displays a facial expression and the caregiver imitates it, the robot detects the change in the caregiver's expression and associates this modifications with the one in its internal state. The authors report that the robot was able to learn how to modify its internal state based on the caregiver's facial expressions.

In another study, Boucenna et al. [38] propose an online facial expression learning model based on imitation of the caregiver's expressions. The authors explore the non-verbal interaction between the infant and the caregiver and its influence on the link between the infant's emotions and the facial expressions of the caregiver. Accordingly to the interaction scheme designed by Boucenna et al. [38], first, the robot produces facial expressions based on its internal emotional state then the caregiver imitate the robot's facial expression, the robot associates this facial expression with its internal emotional state. In short, the communication is established via imitation. They use a simple neural network architecture for providing the ability of adaptation and online learning to the robot. They report that the robot is able to learn the facial expressions without any prior knowledge about what is a face or an emotional expression. They use four basic emotions: happiness, sadness, anger and surprise and a neutral face. In a recent study, the authors [39] use again the same imitation principle to teach the robot different postures in a similar interaction scheme: First the robot observes its caregiver's imitation of its poses then it learns the correspondence between its body/pose and the human body/posture using a “sensory-motor architecture", i.e. a simple neural network.

Following the developmental psychology approach, Lim and Okuno [40] explore the idea that emotional behaviors are developed through experience. The authors present a robot called MEI (Multimodal Emotional Intelligence) with an implemented emotion development system. They propose to use robot-directed baby talk to help the robot associate its inner feeling (or “gut feeling" as the authors refer) with its interactions with the caregiver based on voice modality. They propose to use a neural architecture mirroring the human development process: Their model is composed of 3 modules, a mirror system to represent the action of the caregiver, a gut feeling module to represent the physical inner feeling of the robot based on battery level or motor heat, and an associative module to match the caregiver's expression with the corresponding inner feeling.

1.1.2 Computational models of emotion in HRI scenarios

Marsella [41] provide a detailed review of the work on computational models of emotion and the underlying methodologies and assumptions on which these models are built. The following studies explore the different aspects of computational modeling of emotions.

El-Nasr et al. [42] propose a computational model of emotions based on the event appraisal model of Ortony et al. [28]. The authors propose to use fuzzy logic based approach to represent emotions by intensity, and to map events and observations to emotional states and behaviors. The model also incorporates learning algorithms (rule-based decisions, reinforcement learning, a probabilistic and a heuristic approaches) to learn the environment, event patterns, object associations and expectations of the user, thus providing adaptive responses to the interaction partner. In order to evaluate the abilities of their model, the authors implement a case study based on a simulation of a pet dog. The users virtually interacting with the dog report that the adaptive component has an important role on the believability of the agent.

Another agent incorporating an emotional model based on the theory of appraisal proposed by Ortony et al. [28] is ParleE [43]. ParleE evaluates the events based on a probabilistic planning algorithm. The model also incorporates motivational states influencing the emotional state of the agent and personality having effect on the intensity of emotions.

In another study, Marsella and Gratch [44] propose to use an emotional architecture (EMA - Emotion and Adaptation) based on cognitive appraisal theory and Lazarus' theory on coping [25]. In EMA, the authors implement problem-focused coping as a behavioral strategy to select appropriate actions against the threats for goal achievement. They implement also emotion-focused coping as a cognitive strategy to change the mental state of the agent when it is faced with similar threats.

Another emotion model is Grace (Generic Robotic Architecture to Create Emotions), which is proposed by Dang et al. [45], where the authors model the emotional process as a response to an internal or external event. The model incorporates an appraisal [28] and a coping mechanism [25], and it includes Scherer's theory [30] and considers also the personality theories in the composition of emotion. The novelty of the approach comes from the intuition module integrated to the system. The proposed module processes the information coming from the internal cognitive state which is computed based on the mood and feelings of the subject.

In a related study, Saint-Aimé et al. [46] propose an emotional interaction model iGrace, based on the emotional model Grace, to be integrated into a companion robot (EmI - Emotional model of Interaction). The proposed model iGrace is designed to allow the robot to interact emotionally based on the speech characteristics of its human companion. They validate their model and the expressiveness of EmI through an interaction study with 52 school children aged 7 to 9 years [47].

In a recent study, Leite et al. [48] propose a children-robot interaction chess scenario where the robot, iCat, infers the children's affective states using a SVM-based affect recognition module based on the children's valence. The authors also use the facial expressions to infer the affective state of the child. The contribution of the study comes from the determination of the child's affective state: If it is deemed to be negative and below a specified threshold then the robot displays supportive social behaviors. The authors also remark that the empathic model has a positive impact in a 5-week long-term interaction scenario. They report that the children felt supported by the empathic robot in a similar way when they feel supported by their peers. Leite et al. [49] propose another variation of the same study where the emotional robot iCat use perspective-taking to capture the emotions of human players in a chess game. The robot uses also the game context and assesses the game scenario from the perspective of one of the players and modulates its empathic responses for the favor of this player (companion), while it annotates the other player as the opponent. Based on the perspective-taking, it reacts in a more empathic way towards its companion and acts in neutral way towards the opponent. The authors report that the companion perceived the robot as friendlier, the results indicate the empathic behavior has a positive effect in the human-robot relation establishment.

As Paiva et al. [50] mention in their detailed survey on empathy and emotional modeling in virtual agents and robots, there are various studies on empathy-based social behavior generation focusing on different aspects of human-robot interaction. The authors remark that modeling the emotional architecture has some similarities for virtual and physical agents but the physical embodiment influence the perception of users. The study of Seo et al. [51] investigate the difference between a physical Nao robot and its simulated representation in terms of empathic responses of the human users. The interaction scenario is based on the misfortune of a Nao robot with a viral infection with a potential memory loss effect. The results indicate that the empathic responses of human users are higher with the physical robot and the physical sad Nao robot evokes more empathy than its simulated sad representation.

Tsiourti et al. [52] had previously investigated how the humans perceived the multimodal emotional behaviors of humanoid robots in social interactions. It had been explored if people are able to recognize the emotions expressed by humanoid robots based on visual and auditory modalities in a congruent and incongruent social context, and how humans perceived the robot in terms of believability, perceived intelligence, and likability [52]. It had been noted that the contextual incongruence had a drastic effect on the perception of the humanoid robots and provided guidelines to design emotional expressions for a robot to be accurately recognized to have a positive effect on the people's impressions and attitudes towards the robot.

1.1.3 Control & learning of emotional behavior representation

In terms of control and generation of artificial emotional expression and behaviors, significant work on the literature focuses on creating agents that can display emotional responses but a large corpora of them takes a user-centered view. Instead of focusing on the underlying mechanisms to generate emotional behavior, these studies focus mostly on the evaluation process of the impact of empathic or emotional interactions and humans' reactions towards affective robots. Therefore researchers rather prefer to use Wizard-of-Oz (WoZ)-based studies [53] and restrain the autonomy of robotic agents in controlling their behaviors and actions. In their survey, Paiva et al. [50] remark that this approach is widely adopted for many interaction studies with social agents. In a similar way, the emotions displayed by the agents are generated using either a rule-based approach or carefully scripted set of actions such as in the case of dimensional theories and symbolic action units display systems.

In terms of learning, the symbolic representations of actions/emotions lead to a symbol grounding problem, i.e. a human expert must provide knowledge to the system and check for the accuracy of it. And manual preprogramming of robot behaviors requires endless annotations and restricts the robot's set of actions/behaviors. In order to avoid this problem, researchers propose to use machine learning-based approaches. The most popular learning scheme in the human-robot interaction studies is imitation learning. Current work in the literature on imitation learning mostly focus on the articulated motor coordination and movement generation for humanoid robots but Breazeal [33] suggests that imitation-based learning has a crucial part in building robots that can learn social and emotional aspects of interaction from their human companions.

For a detailed review on imitation learning and proposed approaches, one may refer to the study of Hussein et al. [54]. The authors investigate on one hand the use of conventional machine learning methods such as classification, regression and hierarchical methods in direct learning mechanisms, on the other hand they explore the indirect learning mechanisms based on reinforcement learning, transfer learning and online learning.

In the socially assistive robot studies based on imitation learning, the multimodal robotic behaviors are mostly designed either using online learning settings [38,39] or learning from demonstration [55,56] or reinforcement learning [57,58]. Or a combination of the mentioned learning strategies as in the study of Moro et al. [59] who propose to combine learning form demonstration and reinforcement learning in order to teach social assistive robots personalized behaviors.

The aforementioned studies do not place the human expert into the learning cycle, whereas in their study, Senft et al. [60] propose the concept of supervised autonomy during the online learning process where the robot is guided by a human expert for the purpose of reducing the cost of exploration of possible behaviors. Their approach combines reinforcement learning and Supervised Progressively Autonomous Robot Competencies (SPARC) to assist the robot in learning an action policy and corresponding behavior while being supervised by a human expert. The reported results show that the proposed system allows for safer and faster learning by the robot with a moderate workload on the human teacher.

1.1.4 Evaluation metrics

Developing emotional and empathic abilities for social robots is essential to ensure a natural interaction between the human companion and the robot. All of the reported metrics have been individually adapted by the researchers for the human-robot interaction scenarios. Even though there are recent studies to unify and standardize the measurement of emotional or behavioral response [61], there is no consensus on how to evaluate the human-robot interaction experience, the empathic abilities displayed by robots and the human empathy towards them. Additionally, the lack of a common method to measure the emotional behavior prevents also the comparison of the relevant studies, as remarked by Nomura [62].

Present studies use various metrics from self-report questionnaires to brain activity measurement by medical imaging or to hand-crafted metrics. These metrics are categorized by their evaluation type and detailed in the following subsections.

1.1.4.1 Subjective evaluation metrics

The subjective evaluation metrics consists of self-report questionnaires frequently used in psychology. They are derived from the similar metrics in the psychology literature aiming to assess the different aspects of human personality and interpersonal relationships. The self-report evaluation metrics may be investigated in three different perspectives based on:

(1) Attitudes towards robots

It is important to generate psychological scales to measure the mental states of human users in order to explore the internal factors corresponding to underlying mechanism of communication with the robots, and to evaluate their attitudes and anxiety in their interaction with the robots.

- Negative Attitude towards Robots Scale (NARS):

  NARS is a lexical questionnaire developed to measure and evaluate the anxieties of human participants towards robots [63]. Previously being designed as a free-form response questionnaire, it was later standardized as a 5-point semantic scale metric which is frequently used as a preliminary measure to test the attitudes of test participants taking part in HRI experiments and explain the differences in their behaviors.

  The questionnaire consists of items addressing the attitudes of participants towards either interacting with robots or their social influence or emotions arisen by the robot during the interaction.

- Robot Anxiety Scale (RAS):

  It is a psychological scale developed for measuring the anxiety of human participants working with robots in HRI studies [64]. The questionnaire is similarly designed as the NARS questionnaire but with a focus on the state-like anxiety evoked by the robot in real and imaginary HRI scenarios, preventing the human partner from a natural interaction with the robotic companion.

The scale consists of 11 questions classified into 3 major categories measuring the anxiety toward the communication capability of the robot (such as whether the robot will understand or not the context and content of the dialog), behavioral characteristics of the robot (such as brusque movement or strength of the robot) and discourse ability (such as what to say to the robot when it is talking with the participant).

(2) The emotional experience on behalf of the human

- Barret-Lennard Relationship Inventory (BLRI):

  Barrett-Lennard [65] proposed an instrument for measuring the quality of interpersonal relationships based on the widely accepted indicators of human relationship quality: empathic understanding, level of regard and congruence. Despite the fact that the inventory was developed to explore patient-therapist relationship dynamics and quality, it has been adapted for use in various relationship scenarios such as marital and/or family, child-adult relationships or meta-perception (predicting the other's view) applications [66]. Another adaptation of the metric includes HRI studies, Tapus and Mataric [67] propose to use BLRI, an index to assess human-human empathy, to evaluate the robot's empathic abilities.

- Interpersonal Reactivity Index (IRI):

  IRI was developed by Davis [68] as a multidimensional assessment tool for measuring the individual differences in empathy built on four separate aspects of it: empathic concern (others-oriented feelings), perspective taking, fantasy (imagination-oriented feelings) and personal distress (self-oriented feelings). It is a self-report questionnaire based on a 5-point Likert scale. The metric provides a tool to measure individual differences in empathy and offer an insight into the relationship between empathy and social functioning, self-esteem, emotionality and sensitivity to others.

- Big Five Personality Traits:

  The “Big Five" model is frequently used in psychological research to measure and assess personality based on five basic personality factors [69] defined as:

  – Openness (or intellect): Tendency to be inventive, curious, imaginative, original, etc.

  – Conscientiousness: Tendency to adhere to norms and rules, to be organized, reliable, responsible, etc.

  – Extroversion: Tendency to be active, assertive, energetic, outgoing, etc.

  – Agreeableness: Tendency to be appreciative, kind, compliant, friendly, etc.

  – Neuroticism (or emotional stability): Tendency to be nervous, sensitive, anxious, worrying, etc.

  The Big Five is widely used in capturing the personality structure of humans because it provides extensive validity and reliability when compared to other models in literature [70]. In addition to its use to evaluate human personality structure, it has been used in several human-robot interaction studies [71,72]. In their study, Santamaria and Nathan-Roberts [73] remark that the Big Five personality traits are commonly used in HRI studies to measure human and robot personality, especially the dimension of Extroversion, and the Big Five may be helpful for providing consistency as a tool in the measurement of personality in human-robot interaction.

- Mini International Personality Item Pool-Five-Factor (Mini-IPIP):

  The Mini-IPIP is a short version of International Personality Item Pool-Five-Factor Model [74] derived from the Big Five personality traits. Donnellan et al. [75] state that the Mini-IPIP scales are consistent and have similar coverage with Big Five measures, their results indicate that Mini-IPIP scale is a useful and short measure of the Big Five factors of personality.

(3) Perception / impressions about robots / agents based on the interaction experience

- Godspeed Questionnaires:

It is a series of questionnaires proposed by Bartneck et al. [76] to measure the users' perception of robots in Human-Robot Interaction (HRI) studies using a 5-point semantic differential scale. The authors propose to use this set of questionnaires to standardize the measurement methodology and provide the opportunity to compare the results of the current studies based on five key concepts in HRI: anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety.

Godspeed questionnaires intend to measure rather the perception of the robot and its acceptance by the human companion than the empathic response of the latter, but it is commonly used in emotional modeling studies as a metric, such as in the study of Charrier et al. [77].

- Bickmore's Questionnaires:

  It is a set of questionnaires designed to evaluate the interaction between a relational agent built to maintain long-term, social-emotional relationships with a human user and the user's impressions about the agent [78]. The questionnaires consist of questions about the aspects of interaction (such as comfort, familiarity, engagement, fun, naturality, etc.) and about the impressions of the human user on the relational agent (such as competence, friendliness, intelligence, reliability, likability, etc.).

- Self-Assessment Manikin (SAM):

  SAM scale is an assessment technique to measure a person's affective experience to a large variety of stimuli, directly measuring the pleasure, arousal and dominance associated with the given stimuli [79]. The advantage of the SAM scale comes from the fact that it is a non-verbal iconic assessment tool preventing the language barrier. It is widely used in human-robot interaction scenarios because of its pictorial representation of emotions. For example, Saerbeck and Bartneck [80] explore the relation between robot motion and the perceived affective state of the robot and select to use SAM scale in their study. They report that the SAM is appropriate to use in similar studies because all participants were able to express their impressions based on the pictographic representation of the emotions.

1.1.4.2 Objective evaluation metrics

Even though the subjective evaluation reports are popularly used in interaction studies, they have their limitations. Humans are often biased when they report on their own experiences and they may not have introspective ability to accurately evaluate themselves about their feelings and experience. Therefore, the researchers prefer to use the following objective evaluation metrics alongside the subjective ones:

- Behavioral analysis:

  The behavioral analysis is a popular metric used by the researchers because the analysis is performed by the specialized experts from the related research areas, e.g. therapists, psychologists, tutors, etc. The analysis process consists of the annotation of audiovisual recordings captured during the interaction scenarios based on a predetermined set of behaviors. The advantage of the behavioral analysis comes from the individual differences of participants captured by the experts.

  Boucenna et al. [39] and Alves-Oliveira [81] use behavioral analysis in their studies to evaluate the test participants' responses to the robot. In another study, the long term engagement with the robot is also analyzed over time based on the behavioral variables in video recordings and interviews [48].

  There is also a specialized case of behavioral analysis, that is Accurate Empathy Scale (AES). The scale was developed by Truax [82] with the purpose of rating the quality of the therapist empathy towards the patient during their counseling sessions. The evaluation is done based on short excerpts of recorded interviews by trained judges on a nine-point scale. The judges evaluate the subjects by the presence or absence of certain set of predetermined and ad-hoc behaviors (verbal or non-verbal).

- Physiological measures:

  The researchers studying the neuroscientific aspect of emotional modeling prefer rather to use medical imaging or physiological measures to accurately observe if the artificial entities (whether it is a robot or virtual agent) triggers also the same response with natural interaction partners.

In their study, Miura et al. [83] and Rosenthal-Von Der Pütten et al. [84] use fMRI to evaluate the brain activity triggered by the empathic robot, where Suzuki et al. [85] prefer to use electroencephalography (EEG) to examine the same phenomenon.

- Interaction-specific measures:

  Apart from the aforementioned methods, researchers prefer to use interaction specific metrics to evaluate the participant's emotional state. These metrics are mostly hand-crafted measures adapted to evaluate the human-robot interaction quality based on the interaction type.

  In their study to explore empathic capability of Keepon and Infanoid robots, Kozima et al. [86] uses eye-contact and joint attention to evaluate the interaction between the robots having different embodiment and toddlers interacting with them.

  In an online learning task of human postures based on a simple neural network architecture, Boucenna et al. [39] propose to use the number of neuron recruited for the learning task as an evaluation metrics for their study.

  In their study, Charrier et al. [77] propose an "attention-based empathy module" to improve the quality of interaction between the human user and Pepper robot during a quiz-based entertainment game they play. The authors propose to use the number of disengagement, distance between the robot and the participant, interaction duration and the number of correct answers to assess the interaction.

  In another study, Alves-Oliveira [81] investigate a collaborative learning scenario with an autonomous and empathic robot tutoring on sustainability. The authors use writing assignment tasks to evaluate the educational impact of the empathic robot in addition to behavior analysis.

1.1.5 Limitations of the recent studies

Although the increasing popularity of affective robotics make way to a large number of academic studies, the proposed studies have major issues and present new challenges:

− The proposed computational emotion frameworks are mostly based on theoretic studies, and they do not provide any guideline to implement the proposed frameworks into a real-world HRI scenario. Thus, they do not present any user studies that will support their hypothesis or assumptions on the presented framework.

− The studies mostly focus on emotion recognition more than emotion generation. Although the emotion recognition task is crucial to infer and interpret the human companion's emotions, the generation of expressive behaviors by the robot is also vital to improve the natural aspect of HRI.

− The presented emotion recognition studies rely on complex models requiring high processing power and memory to obtain high levels of accuracy. But, the robots working in the real-world situations have more cost-effective designs and do not have the required infrastructure to use these proposed methods for recognition.

− The affective robots presented in the literature have mostly fixed emotions with a restricted numbers of emotions and animations. The most frequently used emotions are basic emotions, consisting of “happy", “sad" and/or “neutral".

− Even though the presented studies use affective robots to foster the interaction gain, these robots do not any adaptation or moderation mechanism to adjust the intensity of their emotional behaviors or animations.

1.2 Motivation and Objectives

Inspired by the role of emotional understanding in human-human communication and social relationships, we can say that empathy is a major element in human–robot interaction studies. Humans have the ability to feel empathy not only for other humans but also for fictional characters and robots. But the robots are not able to display any empathic behavior in return. As robots make their presence known in human interactive environments, such as health, education, entertainment and social domains, we require methods to design robots that can interact with their environment in a social way. This shift in the robotics studies bring forward the necessity to design robots having emotional understanding and using empathy as a social cue.

The recent literature studies show that developing emotional and empathic abilities for social robots is essential to ensure a natural interaction between the human companion and the robot, therefore we can ask this question: Why not use the robots with emotional and empathic abilities to improve the robot's assistance and foster the interaction gain of the human companion?

The motivation of this study is to design and implement a computational emotion and behavior model for a companion robot based on the data from the human companion and the social interaction context. The proposed framework is designed to process multi-channel stimuli, to interpret the affective state of its human companion and generate in-situ affective social behaviors.

1.2.1 Research questions

The proposed work has the aim to explore the impact of an assistive robot having empathic abilities and thus which is able to regulate its behaviors with respect to the affective states of its human companion.

We will take the following research questions into consideration and attempt to answer them throughout this thesis study:

RQ1. Would the robot be able to display basic emotions? Could the human companion identify correctly the emotions displayed by the robot?

RQ2. How could a social robot infer and interpret its human companion's emotional states?

RQ3. How can we computationally model an artificial emotional mechanism and implement it on a social robot to provide a natural social interaction?

RQ4. Which learning techniques should we use for the affective robot assistant to learn which emotional behavior to express during the interaction?

RQ5. Could a social robot designed with emotional understanding and expression foster the interaction gain in a HRI scenario based on assistance?

1.2.2 Challenges

The proposed study incorporates some challenges and open issues because emotional agents and social behavior generation are relatively new research domains. Some of the challenges are addressed below with possible suggestions to solve them:

- Human affective behavior:

  Automated analysis of human affective behavior is a popular and multidisciplinary research field. However, in their survey, Zeng et al. [87] remark that the existing methods typically handle only deliberately displayed and exaggerated expressions of prototypical emotions without considering the spontaneous structure of human behavior. In their paper, they explore and investigate the recent advances in multi-modal fusion for human affect analysis and discuss important challenges like the data collection, training and testing the available data.

- Data collection & training:

  Learning from demonstration and reinforcement learning allows the robotic agent to learn behaviors without any preprogramming but they require a wide and populated data set of interactions with demonstrators for training purposes. The data collection challenge may be investigated in sub-problems:

  - How to collect interaction data?

    Natural HRI data may be collected on the fly with sensor networks in target environments but the collected data should be checked in term of interpretability. Therefore it should be collected in a controlled environment within a specific interaction scheme under the supervision of the experimenter. For the supervised learning tasks, the collected data should be labeled and annotated by the experts.

  - How to decide how much data is enough to train the system?

    One of the major challenges of this study is how to decide if the collected data is enough. Therefore the data collection and training of the system will be maintained in a cycle and the quantity will be checked continuously.

Apart form the real-world interaction scenarios, the collected may be populated by existing data sets (a detailed list is presented by Zeng et al. [87] in their survey) and simulation studies in virtual environment.

- Embodiment:

  Assessing the effects of robots with emotional and empathic behavior in physically embodied social robots has just recently become popular if compared to the research in the virtual agents community [50]. Because it is easier to infer the human user's affective states in interaction scenarios with virtual agents than it is in physical robot scenarios. In virtual interaction scenarios, the human companion is usually situated in a predefined and controlled environment, and mostly inform the agent of his/her emotional state, as in self-reported studies. However the interaction with physical robots tends to be more complex, and thus it is not always possible to accurately perceive and assess the human user's affective state.

- Context of interaction:

  A large number of methods presented in the literature is based on the idea that the facial expressions mirror directly the internal affective state of the robot/human companion. However there is another factor which needs to be considered in terms of emotional transparency of the agent [87]: The context of interaction (environment, observed subject, the current task, etc) in which the affective behavior is displayed. Without any contextual information about the interaction, it is not always possible to interpret the affective state of the agent.

- Evaluation:

  As previously mentioned in Section 1.1.4, most of the evaluation metrics have been individually adapted by the researchers for the human-robot interaction scenarios. Even though there are present studies with the aim of standardizing the assessment of emotional or behavioral response, Nomura [62] remarks that there is no consensus on how to evaluate the HRI experience, the empathic abilities displayed by robots and the human empathy towards them, an no common method to compare the outcomes of presented studies.

1.2.3 Limitations of the proposed work

The limitations of the study are as follows:

- We assume that the companion robot will stay fixed to a reference point, therefore its mobility or navigation through people will be left out of scope for the study to simplify the challenges introduced by a moving robot in human-inhabited locations.
- Although one of the inputs in the proposed architecture is video recordings of human companions, the performance of vision-based emotion recognition methodologies won't be explored in the scope of this work. The proposed study will use the outputs of available open source vision modules.
- Voice characteristics of the human companion won't be incorporated as a modality for the proposed work, therefore voice / sound recognition / detection or speech recognition or generation methodologies will be left out of scope for this study. Since the speech won't be included as a modality, the semantic analysis of the expressions articulated by the human companion won't be integrated into the system.

2. THE DESIGN OF AN AFFECTIVE ASSISTIVE ROBOT

As the robots moved out from the controlled laboratory environments towards human-oriented everyday life scenes, the need to establish an interaction in a more 'natural' way has increased. In a detailed overview of HRI studies, Fong et al. [1] explore what makes the interaction between a robot and a human desirable and they conclude that the robot should obtain social skills similar to humans' skills. Similarly in another study, Goodrich and Schultz [2] remark that the robots should display and generate more human-like motions to improve the interaction quality and make the interaction establishment easier between the communicators.

The socially assistive robotics (SAR) field focuses on helping human users through social interaction. In their review of the SAR studies, Tapus et al. [4] remark on the principle application domains of SAR, as well as the key concept of SAR studies: embodiment, personality, empathy, engagement and adaptation. Socially assistive robots are designed to provide active assistance to their human companions. In order for this assistance to be effective, these socially assistive robots should be able to display appropriate emotions/behaviors and in-situ motions and to adapt/adjust these abilities with respect to the human companion's needs; hence, the inspiration to design computational model of emotions.

Gratch and Marsella [17] propose that models of emotion may give insight into building models of intelligent behavior for social robots and the effort to accurately model emotional mechanisms may have a major contribution on the human understanding of emotions. In a study based on the emotion models existing in psychology literature, Scherer [30] cites that emotions has a motivational component which is responsible in the formation of the individual's behaviors. Inspired by this principle, Saint-Aimé et al. [46] propose an emotional interaction model iGrace, based on the emotional model Grace, to be integrated into a companion robot (EmI - Emotional model of Interaction). The proposed model iGrace is designed to allow the robot to interact emotionally based on the speech characteristics of its human companion. They validate their model and the expressiveness of EmI through an interaction study with 52 school children aged 7 to 9 years [47].

In a recent study, Leite et al. [48] propose a children-robot interaction chess scenario where the robot, iCat, infers the children's affective states. The contribution of the study comes from the determination of the child's affective state: If it is deemed to be negative and below a specified threshold then the robot displays supportive social behaviors. They report that the children felt supported by the empathic robot in a similar way when they feel supported by their peers.

Inspired by the aforementioned studies, two case studies were designed to validate the accuracy of the emotional behaviors of a humanoid robot, Pepper, and to explore the impressions of test participants about an emotional robot. The design and implementation of the affective robot was coordinated in accordance with the Tübitak project, entitled “RoboRehab: Assistive Audiology Rehabilitation Robot".

2.1 RoboRehab Project

The motivation and objective of the RoboRehab project is to use assistive robots to promote rehabilitation for children with special needs via machine learning and HRI techniques. For this purpose, an auditory perception test was gamified and a Pepper humanoid robot having basic emotional behaviors is integrated into the gamified version of the test to help children and reduce their stress during their clinic visits.

The Pepper humanoid robot, the profile of test participants and the experimental design of the two case studies are explained in detail in the following subsections.

2.1.1 Pepper: "The robot built for people"

Pepper is a social robot designed for the purpose of interacting with people. It is 120 cm high and has 20 degrees of freedom (DoF) to generate more natural and expressive movements. The robot is equipped with preprogrammed perception modules to recognize and interact with people in its environment. It has touch sensors, LEDs and microphones for multimodal interactions and makes use of infrared sensors, bumpers, an inertial unit, 2D and 3D cameras and sonars for omnidirectional and autonomous navigation. It also provides additional modules for speech recognition and dialog in 15 languages.

2.1.1.1 Configuration

In order to control and implement the behaviors of the robot, the developers should access the NAOqi Framework, which is provided by the manufacturer. NAOqi is the main software that runs on the robot and controls it. It is accessible via its Application Programming Interface (API) and Software Development Kits (SDK) developed for popular programming languages.

It is a cross-platform and cross-language software providing the developers to monitor the tasks executed by the different modules. NAOqi comes with a list of modules that are always available and every module has a group of default methods. The emotional behavior of the robot is implemented based on the following modules:

- NAOqi core modules
  - ALBehaviorManager: It manages the robot's behaviors, it starts and stops predefined behaviors.
  - ALMemory: It controls and displays the data flow between available modules.
  - ALModule: It provides the ability to create new modules for the developers.
  - ALMood: It manages the emotional state of active users around the robot, it is mainly based on NAOqi vision and people perception modules, explained below.
  - ALTabletService: It manages the tablet, loads web applications and plays visual files on the tablet.
- NAOqi interaction engine modules
  - ALAutonomousLife: It maintains the robot's life cycle, i.e. it gives the robot a more 'natural' look with simple and smooth head and torso movements and tracking abilities.

  - ALBasicAwareness: It allows the robot to react to its environment, i.e. the robot has the ability to perceive if there is anyone in its environment and if it detects someone, it can make and keep eye contact with them or turn towards their voice.

- NAOqi motion modules

  - ALAnimationPlayer: It starts and stops predefined animations (available on the NAOqi libraries or programmed by the developer).
  - ALRobotPosture: It makes the robot go to a predefined posture (available on the NAOqi libraries or programmed by the developer).
  - ALNavigation: It makes the robot move among the people.
  - ALMotion: It directly controls the kinematics properties of the joints (stiffness, forward and inverse kinematics, velocity control, etc.)

- NAOqi audio modules

  - ALAudioDevice: It manages all the audio inputs and outputs.
  - ALAudioPlayer: It plays the preloaded audio files on the robot.
  - ALTextToSpeech: It makes the robot voice the given text.
  - ALAnimatedSpeech: It combines speech and gestures, provides the robot more human-like interaction properties (e.g. the robot nods its head while saying an affirmative clause).

- NAOqi video modules

  - ALPhotoCapture: It takes pictures using one of the several cameras mounted on the robot and saves them on disk.
  - ALVideoDevice: It manages video inputs.
  - ALVideoRecorder: It records video using the camera mounted on the robot.

- NAOqi people perception modules

- ALFaceCharacteristics: It provides the estimated human characteristics based on facial analysis. It detects a large number of facial points and infer basic characteristics about the person's mood and emotions.
- ALFaceDetection: It makes the robot detect and recognize human faces.
- ALGazeAnalysis: It analyzes the gaze direction of a person interacting with the robot, tracking the eye movement.
- ALPeoplePerception: It helps the robot to track the people around it.

There are 2 different ways to access the modules of the NAOqi Framework, the first one is to use Choregraphe desktop application and the second is to use the provided API and SDKs developed for other programming languages: C++, Python and Java.

Choregraphe is a cross-platform, block-based desktop application with various panels. It is a visual programming interface incorporating several predefined motions, animations, behaviors to design simple robotic applications. It provides a simple robot simulator with a virtual robot. One may also monitor the robot's present status based on its battery, the joints' stiffness, preloaded programs available on its memory, the events triggered by the execution of different modules, etc.

The NAOqi version of the Pepper robot (2.5.7.10) is accessible through its Python 2.7 SDK. NAOqi modules are called through their proxies or session services based on a TCP connection.

2.1.1.2 Verbal interaction of Pepper for RoboRehab

In human-robot interaction studies, researchers prefer to use a machine-like and genderless voice for the robots to eliminate the gender bias issued by the female or male voices among the participants. Although the audiologists requested to use a deep male voice because it is easier to perceive it for people using hearing aid or having cochlear implant, a subset of recordings in Turkish with different voice profiles (male, female, machine-like voices with varying pitch and speed properties) were prepared using open source text-to-speech software and Pepper's embedded text-to-speech module. The recordings are voted by the audiologists and a small set of children. The most liked voice [88] is used to generate the robot's verbal feedback.

The feedback sentences used to enhance the supportive behavior of the robot are also provided by the audiologists. They stated that they don't usually give negative feedback to children and they use only positive and neutral feedback mechanisms, therefore the verbal feedback of the robot are generated accordingly. Examples for the positive and neutral utterances are as follows (The complete set of the audio recordings of predefined feedback utterances, raw and edited versions included, can be found at [89]):

- “Harika(sın)!"
- “Bravo!"
- “Süper!"
- “Çok güzel!"
- “Çok iyi(sin)!"
- “Aferin (sana)!"
- “Şimdi bir daha deneyelim"
- “Dikkatli dinle"

In the case of positive feedback utterances, when the child's answer is correct a positive utterance is randomly selected from the list and voiced by the robot. The neutral feedback mechanism is controlled by the therapist and hasn't been implemented autonomously as per the request of the audiologists.

2.1.1.3 Non-verbal interaction of Pepper for RoboRehab

In order to determine the non-verbal feedback of the robot, all the animations accessible from the Choregraphe library are investigated in detail (The complete list of the videos for predefined animations available in Choregraphe can be found at [90] and a selection of the most relevant and interesting videos for the children at [91]). The animations are eliminated first with respect to the safety of children. The target group being children 5 to 7 years old, the animations with brusque and high-powered movements are discarded. The second criteria for the selection was the compatibility of the animations with the cultural background and cognitive developmental age of the children.

After the examination of the animations available in Choregraphe library, another library composed of predefined animations labeled by the manufacturer not available through Choregraphe but available via Python SDK has been found. Although there is a small intersection set between the two libraries, the Python library is richer than the

Choregraphe library. Therefore, after an extended search on the Python library, a set of animations is selected to be used in the design and implementation of the affective robot's behavior.

The selected animation set for the case studies is composed of preloaded Choregraphe animations labeled by the manufacturer as “affirmative" (e.g. head nod), “exclamatory" (e.g. lean back with open arms) and “positive moods" (e.g. happy, excited). As a summary, when the robot voices a positive utterance then it is accompanied by a positive posture and a positive affective animation.

2.1.2 Test participants

The target group of the RoboRehab consists of children aged 5 to 8 years old with a typical cognitive development and without any mental or physical disorders. The test studies are planned to be conducted with 3 groups having different profiles:

1. Children with typical development (control group)
2. Children with hearing aid
3. Children with cochlear implant

In order to test and explore beforehand the interaction setup and the effect of Pepper's emotional behaviors, a preliminary study was conducted with only 3 typical hearing children from the control group.

The main user study was conducted with 16 children with hearing aid or cochlear implant. The inclusion criteria in the study were determined by the audiologists, the children:

− with typical development based on AGTE [92], Denver [93], and WISC-R [94] scales for children,

− with typical language and speech development,

− with hearing loss, and who have been using hearing aids or cochlear implants for at least 1 year,

− who do not have any other known neurological or psychological diagnosis were included in the study.

The children in the test group were invited by the audiologists based on their prior diagnosis. The test group was composed of children with typical cognitive development, and without any mental or physical disorders, having either cochlear implants or hearing aids. The parents of the children were informed about the experimental procedure for the test, data acquisition setup and they had signed a written consent.

The tests for the 2 case studies were conducted in a laboratory setup at the Istanbul Technical University. The children came to our laboratory with their parents. 25 children were invited for the study but only 19 of them attended to play with Pepper robot. 3 of the children had multiple disabilities that hindered their physical and cognitive performance on tests, so their data were not included in the results.

The case studies were conducted with 16 children (10 girls, 6 boys) aged 5 to 8 years old (M = 6.2 and SD = 0.9). The profile of children are displayed in Table 2.1 in detail. 15 of the children had no prior experience with real-world physical robots, except one of them, an 8-year-old girl who had previously participated in our pilot study with a smartphone-powered education robot (Smart Robot Albert, a bird-like small-sized robot).

Table 2.1 : The profile of children with hearing impairment.

Gender	Hearing Device	N	Age (M, SD)
F	Hearing Aid	4	5.5, 0.6
	Cochlear Implant	6	6.8, 0.8
M	Hearing Aid	2	6.0, 0.0
	Cochlear Implant	4	6.0, 1.2
Total		16	6.2, 0.9

2.2 Case Study 1 - Validation of Robot's Emotions

As mentioned previously, Pepper robot has several predefined animations labeled by the manufacturer but after a detailed examination of the animation libraries (accessed via Choregraphe or Python SDK), we have seen that the animations and their labels are not always congruent. Therefore a preliminary study was designed to check the validity of these animations and explore if the children in the target group understand which behavior or emotion the robot tries to manifest.

In the scope of this study, two different cases are explored by asking the children a few short questions:

1. If given a related context, could the child select the video clip with the congruent behavior of the robot?
2. If given a video clip of the robot displaying an emotion, could the child annotate a congruent label with it?

A restricted set of emotions is selected for the study. The explored emotions are happiness, sadness and surprise, considered as “social emotions" which is defined as emotions having a social and interpersonal function in human interactions [95]. In addition, these three emotions belong to different quadrants of 2D valence-arousal [35] and 3D pleasure-arousal-dominance space [31], therefore it is easy to identify them computationally.

The questionnaire used for this study is designed as an online survey filled by the children via a tablet or a laptop, accompanied by an adult. The children were asked to fill the survey for the expressiveness of the robot, and one of the experimenters accompanied and assisted them during the survey.

2.2.1 Congruent emotional behavior from context

The aim in this case is to check if the child can select the video clip with the congruent behavior of the robot when a related context is given. The questions related to matching the accurate behavior with the context are composed of really short narratives. Due to the age of children, the context is narrated textually as well as visually via pictorial representation of the story. Once the narrative is stated, then the children are asked to watch the short video clips of robot behaviors and select the most congruent with the context of the narrative. An example question is displayed in Figure 2.1.

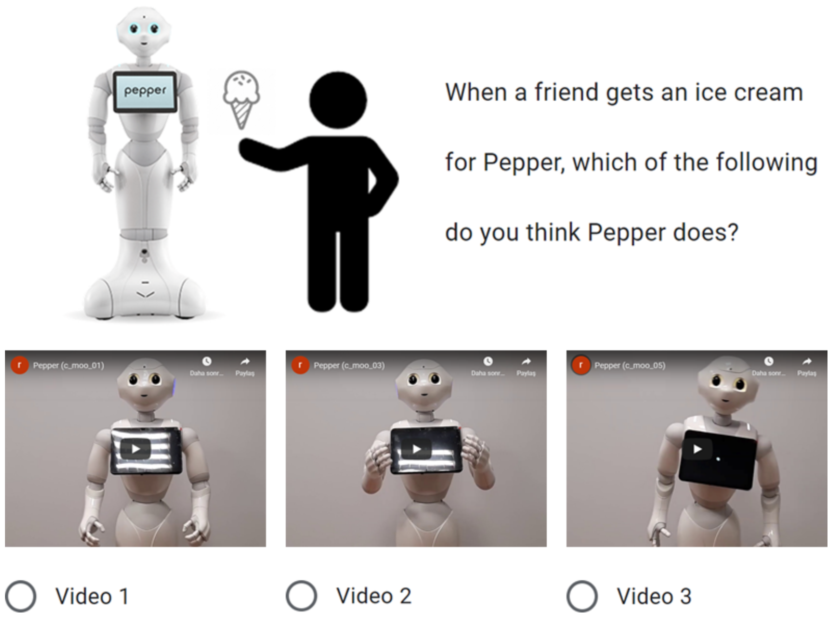


Figure 2.1 : Example question for identifying the congruent emotional behavior based on the narrative context.

2.2.2 Context from congruent emotional behavior

The aim in this case is to explore if given a video clip of the robot displaying an emotion, can the child annotate a congruent emotion with it. The questions are composed of video clips of the robot, each displaying a basic emotion. The children are asked to guess how the robot feels, as seen in Figure 2.2.

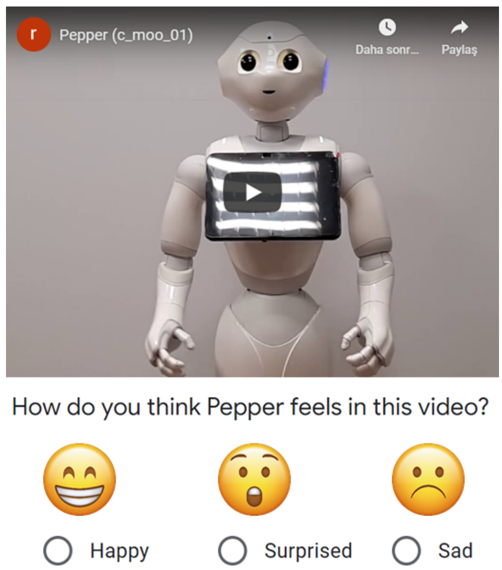


Figure 2.2 : Example question for identifying the context based on Pepper's emotional behavior.

2.2.3 Experimental results

The results of the online survey about the interpretability and validity of Pepper's emotional behaviors showed that the children succeeded in annotating the robot's emotion correctly [96].

In the first task, where the children matched the given narrative with a congruent behavior, the results showed that they were able to match a happy, sad, surprised context with a happy (10 out of 16, $\chi^2(2, N = 16)$ = 6.125, p = 0.047), sad (13/16, $\chi^2(2, N = 16)$ = 26.375, $p < 0.001$), surprised (9/15, $\chi^2(2, N = 15)$ = 6.4, p = 0.041) expressive behavior. Although the last narrative about a surprise birthday party organized by its friends for Pepper was a trick question where all the options were congruent (surprised, happy and shy emotional behaviors), the χ^2 test showed that the children's preference was not equally distributed and most of them selected the behavior with the "surprised" label.

In the second task, where the children asked to guess the emotion displayed by Pepper, the children managed to annotate accurately the emotions of the robot, as seen in Table 2.2. The chi-square test of goodness-of-fit showed that the children's answers for the questions were not equally distributed; "happy", "sad" and "surprised" with $\chi^2(2, N = 16)$ = 21.125, 17.375 and 11.375 respectively and for all corresponding p-values < 0.01.

Table 2.2 : The confusion matrix for emotion labeling based on behavior.

		Actual Behavior Labels		
	Emotion	Happy	Sad	Surprised
Labeled	Happy	0.88	0.19	0.31
by	Sad	0.06	0.81	0
Children	Surprised	0.06	0	0.69

These findings are promising for the next step of the project, where the robot will navigate in the hospital corridors accompanying the children to their test rooms. The affective animations of Pepper will be integrated in to the interaction games in between the auditory test sessions to increase the motivation of children for their audiology examination.

2.3 Case Study 2 - Auditory Perception Test with an Emotional Robot

The second case study is based on the gamification of an auditory perception test for the target group of children with different setup conditions: a tablet, a tablet and a robot together, and the conventional test setup.

The experimental design of the study, i.e. audiology test questions, physical setup, and the interaction interface are specified by the audiologists and pedagogue working on the project and explained in detail in the following subsections.

2.3.1 Experimental setup

The experimental setup for this case study consists of a conventional auditory perception test, different interaction interfaces for the gamified version of the test and a set of questionnaire to evaluate the impressions of children about the gamification of the audiometry test.

2.3.1.1 Auditory perception test

The auditory test used in the study is a Turkish-translated version of the Developmental Test of Auditory Perception (DTAP). The test is designed for assessing the auditory perception of children without invoking higher order verbal reasoning [97]. DTAP is composed of different item sets that measures various aspects of auditory perception. Each set is composed of 30 questions and each question composed of a pair of auditory stimuli. The children are asked to indicate if the pairs are the same or different. Only the non-language related item sets, which are:

- tonal patterns, i.e. sequence of beeps with varying frequency and length
- environmental sounds, i.e. crying babies, barking dogs, traffic noises, etc.

are selected for this study and they are adapted for the gamification process.

In the conventional test setup, an audiologist accompanies the child taking the test and marks the child's answers on a paper-based questionnaire. In the proposed gamification-based setup, the audiologist preserves his/her place along with the child but the child answers the test questions using the tablet. If the recordings are the same, then the answer with the picture of 2 identical muffins should be selected, if not, then the answer with the picture of 1 muffin and 1 ice-cream cone should be selected by the child (Figure 2.3a).

2.3.1.2 Child robot/tablet interaction interface

In the conventional DTAP setup, an audiologist plays a pair of recordings using a computer or music player and asks the child if the recordings are the same or not. The child taking the test, answers verbally and the audiologist marks the child's answers on a paper-based questionnaire. In this study, two gamification-based setups are developed to digitalize the auditory test: tablet and robot+tablet.

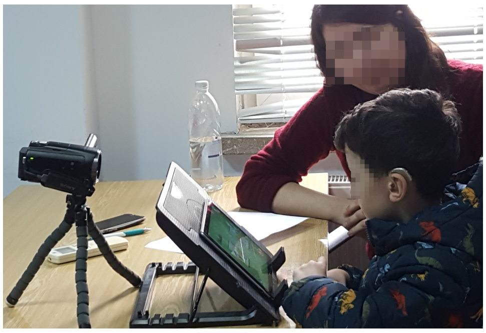
The audiologist preserves his/her place along with the child but the child answers the test questions using a tablet in both setups. If the pairs of auditory stimuli are the same, then the child should select the answer with the picture of 2 identical muffins, if not, then the answer with the picture of 1 muffin and 1 ice-cream cone should be selected. Both game setups have a short tutorial phase for conditioning the children and demonstrating them the concepts “same" and “different" represented with pictures.

The setups differ in the presence of the robot and the feedback mechanism. The first setup consists of a tablet-based game (Figure 2.3b). The same game is played with the tablet and Pepper robot in the second setup (Figure 2.3c).

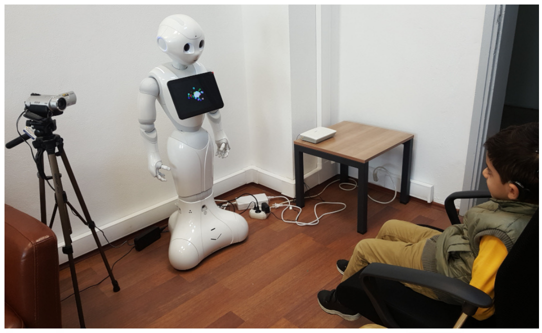
The different physical setups for the auditory perception study is schematized in Figure 2.4. In all the setups, the children are equipped with an Empatica E4 wristband, capturing real-time physiological data during the test. The whole session is recorded with a supplementary video camera. Children interact with the tablet accompanied by an audiologist. In the robotic setup, the tablet communicates with the robot via an intermediary device monitored by the researcher, all the components are connected via a wireless network.

The feedback is given audio-visually in the tablet case and the feedback is given by the robot's supportive behaviors in the robot+tablet case. The verbal and non-verbal cues of the robot are designed and implemented considering the specifications provided by

(a) Visual representations for the concepts “same" (left) and “different" (right).



(b) Tablet setup.



(c) Robot+tablet setup.

Figure 2.3 : Test setups for auditory perception game.

the audiologists and pedagogue, as detailed in Section 2.1.1.2 & 2.1.1.3. The robot’s verbal and non-verbal behaviors are triggered by the correct answers of the children.

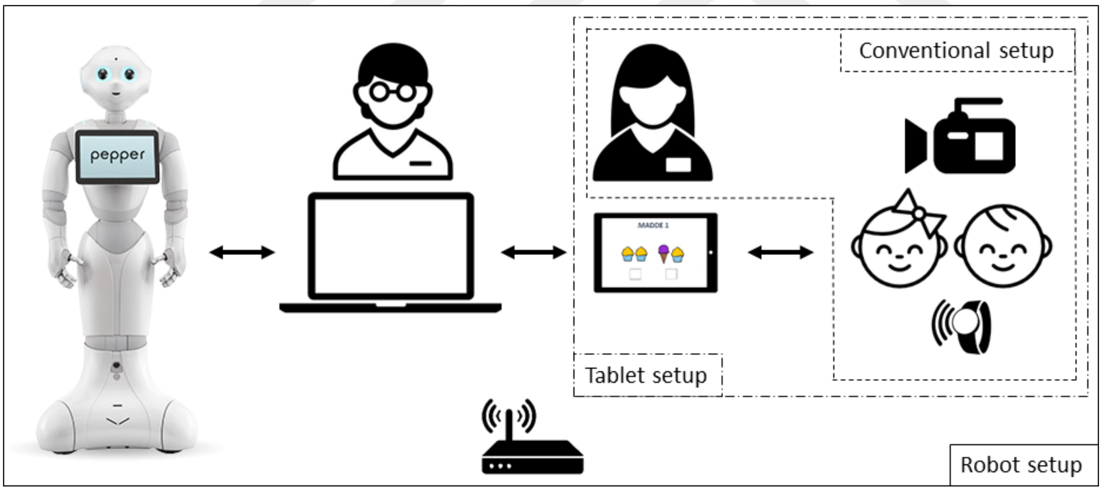


Figure 2.4 : Different test setups for the auditory perception game.

2.3.1.3 Pre-test and post-test questionnaires

In order to measure and evaluate the emotional responses of the children toward the robot and their impressions about its behaviors, 2 questionnaires were prepared to be filled, the first one before (pre-test) and the second one after the interaction with the robot (post-test).

The pre-test questionnaire is composed of simple questions to get the demographic data of the children. Due to the age of children, the well-known questionnaires to measure attitudes towards robots such as NARS (Negative Attitude towards Robots Scale, [63]) or RAS (Robot Anxiety Scale, [64]) cannot be used since both questionnaires contains questions that require a higher linguistic and cognitive understanding. Therefore, the pre-test questionnaire includes only a single question about the prior experience of the child with a robot (if any, 5W1H questions).

The post-test questionnaire is designed to measure and assess the interaction quality and the impressions of the children about the robot displaying basic emotions. Similar to the pre-test case, the questions in the post-test are reorganized for the children, so that they are easier to follow up and understand. The questionnaire is mostly composed of yes-no questions, as seen in Table 2.3.

2.3.2 Experimental results

The experimental results for different setups and the impressions of children about the robot are detailed in the following subsections.

2.3.2.1 Auditory perception game: Robot+tablet versus tablet

All the children taking part in the auditory perception game were tested in both tablet-based gamification with or without robot setups. They were tested in alternating order and with different item sets (tonal patterns or environmental sounds).

The following test metrics were extracted from the data collected during interaction sessions for each child to explore the effect of the gamified setup and the robot's presence in the setup:

– *Test score*: The percentage of the correctly answered questions by the child,

– *Tutorial time*: The duration in seconds of the familiarization session with the gamified setups and the test questions,

– *Total test time*: The duration in seconds of the auditory perception test composed of 30 questions,

– *Response time*: The time difference in seconds between the two answers of the child.

Two-tailed paired t-tests were performed on the test metrics to investigate if the test setup with and without Pepper had any significant effect. The results showed that the setup did not have any significant effect on the test scores of the children but it caused a significant difference on the timing characteristics (Table 2.4).

Table 2.3 : The post-test questionnaire to explore children's impressions.

No.	Question
1	Did you get excited when you saw the robot? *(Robotu görmek seni heyecanlandırdı mı?)*
2	Did you like the robot? *(Robotu sevdin mi?)*
3	What did you like the most about the robot? *(Robotun en çok hangi özelliğini beğendin?)*
4	What did you like least about the robot? *(Robotun hangi özelliğini beğenmedin?)*
5	Did you like to play with the robot? *(Robotla oynamayı sevdin mi?)*
6	Would you like to play again with the robot? *(Robotla tekrar oynamak ister misin?)*
7	Would you like to be friends with the robot? *(Robotla arkadaş olmak ister misin?)*
8	Would you like to have one of these robots at home? *(Evinde bu robottan olsun ister misin?)*
9a	Would you like to have one of these robots in your school? *(Okulda bu robottan olsun ister misin?)*
9b	Would you like to have one of these robots in your hospital? *(Hastanede bu robottan olsun ister misin?)*
10	Do you think that the robot is intelligent? *(Sence bu robot akıllı mı?)*
11	Do you think that the robot looks like a human? *(Sence bu robot insana benziyor mu?)*
12	Does the robot make you laugh? *(Robot seni güldürüyor mu?)*
13	Do you think that the robot behaves like a friend? Or like a teacher? *(Sence robot arkadaş gibi mi yoksa öğretmen/anne-baba gibi mi davranıyor?)*
14	If it was you who programmed/taught the robot, what would you like it to do? *(Eğer robotu sen programlasaydın, robota ne öğretirdin?)*

One-tailed paired t-tests were performed for each significant item to check if the corresponding means of time-related metrics for the setup with the robot are smaller or greater than the other setup. The results showed that the tutorial duration of the robotic

Table 2.4 : The results of paired t-test based on test setup.

	Mean (SD)		$t(15)$	p
	Robot	Tablet		
Test Score (%)	0.64 (0.18)	0.67 (0.18)	-0.595	0.560
Tutorial (s)	190 (69)	298 (256)	-1.807	0.091
Total Test (s)	568 (75)	492 (83)	3.658	0.002*
Response (s)	239 (87)	178 (84)	2.831	0.013*

setup is significantly less than the other ($t(15) = -1.807$, $p = 0.045$). On the other hand in the setup with the robot, the total test duration and response times of children are significantly greater than the other ($t(15) = 3.658$, $p = 0.001$ for total test duration, $t(15) = 2.831$, $p = 0.006$ for response time). A two-way ANOVA test is performed to determine if there is any statistically significant interaction between the significant metrics and the test order. The results showed that there was no significant interaction between the timings of children and the test order.

Previous analysis showed that the test setup do not have any significant effect on test scores, therefore further analysis was conducted to see if the profile of children or test-related factors had a significant effect on the test scores, as seen in Table 2.5. The results demonstrated that only the item set caused a significant interaction on the children scores in both setup (Table 2.5).

Table 2.5 : The results of ANOVA based on test setup.

Factor	Robot		Tablet	
	$F(1,14)$	p	$F(1,14)$	p
Gender	0.950	0.346	1.196	0.292
Hearing Device	0.382	0.547	0.328	0.576
Age	1.609	0.225	1.005	0.333
Test Order	2.222	0.158	1.603	0.226
Item Set	11.12	0.005*	4.772	0.046*

One-tailed paired t-tests showed that the children's scores in tests with environmental sounds item set were significantly higher than tonal patterns for both test setup. The mean scores and number of test participants for both setup and item set are summarized in Table 2.6.

Table 2.6 : The results of t-test based on item set.

Setup	Item Set	N	Score (*M*, *SD*)	t(14)	p
Robot	Env	10	0.73, 0.16	3.335	0.002*
	Ton	6	0.48, 0.10		
Tablet	Env	11	0.72, 0.17	2.185	0.023*
	Ton	5	0.54, 0.10		

4 of the children were tested with the same item set (environmental sounds) in both setup to examine the effect of the physical presence of the robot. The results of the paired t-test, displayed in Table 2.7, showed that there is no significant difference between the both setup and no significant interaction with the test order.

Table 2.7 : The results of paired t-test based on test setup with the same item set.

	Mean (SD)		$t(3)$	p
	Robot	Tablet		
Test Score (%)	0.82 (0.18)	0.83 (0.21)	-0.157	0.885
Tutorial (s)	145 (20)	164 (42)	-1.104	0.350
Total Test (s)	514 (72)	492 (101)	0.914	0.428
Response (s)	184 (71)	162 (98)	1.070	0.363

As a summary, the statistical analysis showed no significant difference on the test scores for both setup, but the timing of the children differed significantly. The longer test duration in the robotic setup is expected since when the children answers correctly, the robot displays happy behaviors. The tutorial session in the robotic setup is shorter which may result from lower attention of the children, when the robot explains them the test procedure. Furthermore, the longer response time may also be explained by the higher attention and care of the children to the questions asked by the robot instead of the tablet, since they take their time to answer correctly. A behavioral annotation study was conducted on the video recordings of the 4 children tested in both setups with the same item set, to answer the 3 questions mentioned above.

The collected videos were coded by an annotator using Behavioral Observation Research Interactive Software (BORIS) [98], an open-source annotation software for video-coding and live observations. The annotator coded the start and end times of each child's utterance (whether the child is speaking with the experimenter (audiologist) or with/about the robot/tablet), gaze direction (whether the child is looking at the robot, tablet or experimenter), gestures (smiling, mimicking the robot) and boredom signs (sighing, yawning, fidgeting, etc.).

The number of occurrences, total duration and mean duration of each behaviors were extracted and analyzed to examine if there is any significant differences between the 2 setup. The paired t-test conducted on the coded features showed that the null hypothesis was not rejected, i.e. there was no significant difference between both setup in terms of children's speaking with the experimenter, their gaze direction and boredom signs.

On the other hand, the annotations revealed that children smiled and talked to the robot and mimicked it whereas there were no coded occurrence of these behaviors in the simple tablet setup without the robot. The children's interactive engagement with the robot may be interpreted as a higher involvement with the game due to the presence of the robot, however the results are not significantly representative because of the small sample size.

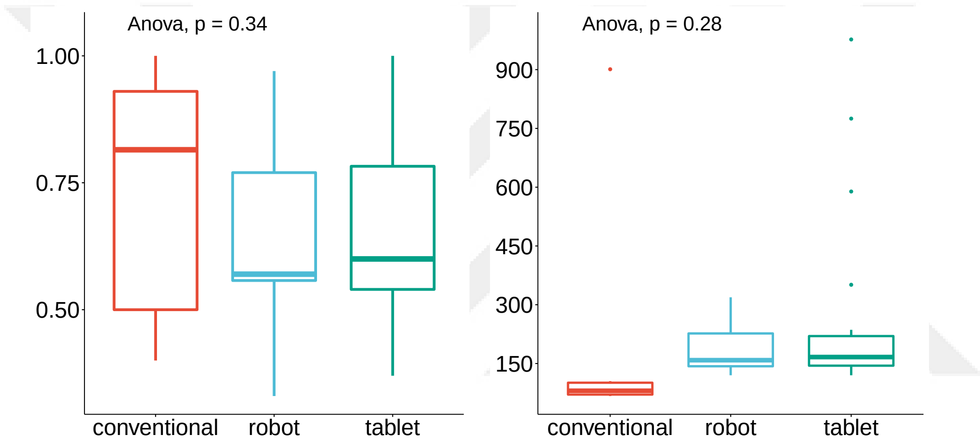
2.3.2.2 Auditory perception game: Robot+tablet versus tablet versus conventional setup

In order to explore in detail the effect of the gamified setup over the conventional setup, the hypothesis testing were repeated to compare the three setups.

A one-way ANOVA test was performed on the collected test metrics (Table 2.8), and Tukey and t-tests were computed to perform pairwise-comparison between the means of the 3 groups (Figure 2.5). The results showed that the setup did not have any significant effect on the test scores of the children and the tutorial duration but it caused a significant difference on the total test time and response time of the children. As displayed in Figure 2.5c and 2.5d, the group means were significantly different between the conventional-robot setup and robot-tablet setup. The significance of the time difference is expected because the robot interacts with the children using verbal and non-verbal cues to maintain their attention, therefore the interaction session takes longer.

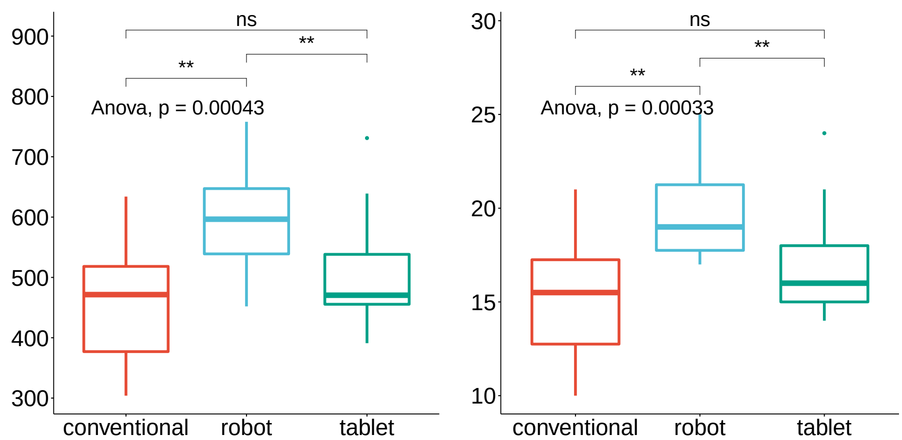
Table 2.8 : The results of ANOVA for test metrics across test setups.

	Mean (SD)			$F(2,47)$	p
	Conventional	Robot	Tablet		
Test Score (%)	0.74 (0.22)	0.63 (0.18)	0.66 (0.19)	1.101	0.341
Tutorial (s)	150 (237)	190 (69)	258 (226)	1.319	0.277
Total Test (s)	461 (109)	596 (81)	502 (81)	9.169	< 0.001
Response (s)	15 (4)	20 (3)	17 (2)	9.541	< 0.001



(a) Test score

(b) Tutorial time



(c) Test time

(d) Response time

Figure 2.5 : The distributions of the test metrics for all the setups, and t-test significance between setup pairs: (ns) not significant, (**) are significant at $p \leq 0.01$.

The result of the one-way ANOVA demonstrated that test scores of the children are not affected by the test setups, in order to explore if the children's profile (age, gender, hearing device) or auditory perception test-related factors (test order and item set) had any significant effect on the scores, further analysis were conducted. The results of ANOVA showed that only the item set (tonal pattern/environmental test) had a significant effect on the test scores in the gamified setups. The mean scores and number of test participants for both setup and item set are summarized in Table 2.9. One-tailed paired t-tests showed that the children's scores in tests with environmental sounds item set were significantly higher than the test with tonal patterns for both test setup, as displayed in Figure 2.6.

Table 2.9 : The results of Welch two sample t-test for test setups based on item set.

Setup	Item Set	N	Score (*M*, *SD*)	t	p
C	Env	6	0.85, 0.22	1.887	0.090
	Ton	6	0.63, 0.17		
R	Env	10	0.72, 0.16	3.311	0.002
	Ton	6	0.48, 0.10		
T	Env	14	0.75, 0.18	3.417	< 0.001
	Ton	8	0.51, 0.10		

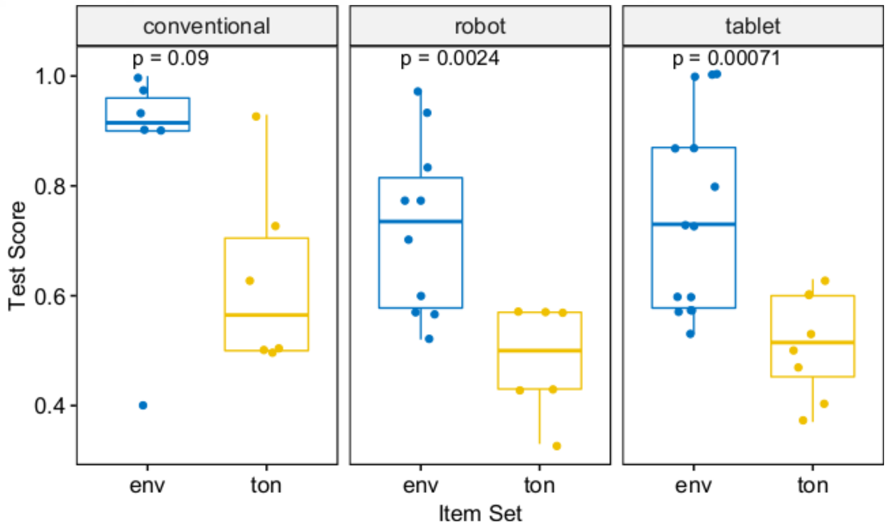


Figure 2.6 : The results of Welch two sample t-test for test setups based on item set.

The audiologists comment that this difference in the scores is due to the nature of the auditory perception test, the auditory stimuli in the tonal pattern item set is harder to perceive and identify than the stimuli in the environmental sound set [99].

In order to eliminate the effect of the test-related factors on the test setup, a group of 6 children were tested with the same item set in the same test order for all the setups. They were first tested with the robotic setup, then the tablet and finally in a conventional setup with a time difference designated by the audiologists between sessions.

One-way ANOVA test showed that the test setup did not have a significant effect on the test scores, total test duration and response time of the 6 children, it had only an effect on the tutorial duration, as displayed in Table 2.10.

Table 2.10 : The ANOVA results for test metrics of the group tested with the same item set in all the test setups.

	Mean (SD)			$F(2,16)$	p
	Conventional	Robot	Tablet		
Test Score (%)	0.75 (0.23)	0.67 (0.2)	0.72 (0.27)	0.177	0.84
Tutorial (s)	89 (15)	152 (38)	150 (23)	10.53	0.001
Total Test (s)	450 (113)	555 (47)	481 (66)	2.684	0.101
Response (s)	15 (4)	18 (2)	16 (2)	2.591	0.108

The paired t-test computed for the pairwise comparison between the test setups showed that the tutorial time differed significantly between the conventional-robot and conventional-tablet setups (Figure 2.7b). This difference is based on the fact that the tutorial of the conventional setup incorporates only instructions about the test questions and do not have any extra information about digital setup/media (i.e. robot and tablet).

The t-test results also showed that, similar to the previous analysis on the all participants, the group means of total test duration (Figure 2.7c) and response time (Figure 2.7d) were significantly different between the conventional-robot setup and robot-tablet setup. As mentioned before, the significance was issued by the verbal and non-verbal feedback of the robot during the interaction session.

As a summary, the statistical analysis showed no significant difference on the test scores for the setups, but the timing characteristics differed significantly, as expected due to the feedback mechanism of the robot. In order to investigate if the robot had any significant effect on the engagement of children despite the longer duration of the interaction session, a behavioral analysis study was planned to be conducted on the video recordings of the children tested with the same item set on all the setups. Unfortunately, due to the fact that the conventional study were performed after the COVID-19 outbreak and all the children wear facial mask and protective clothing, the behavioral analysis results were inconclusive for the comparison of the conventional setup with the gamified setups.

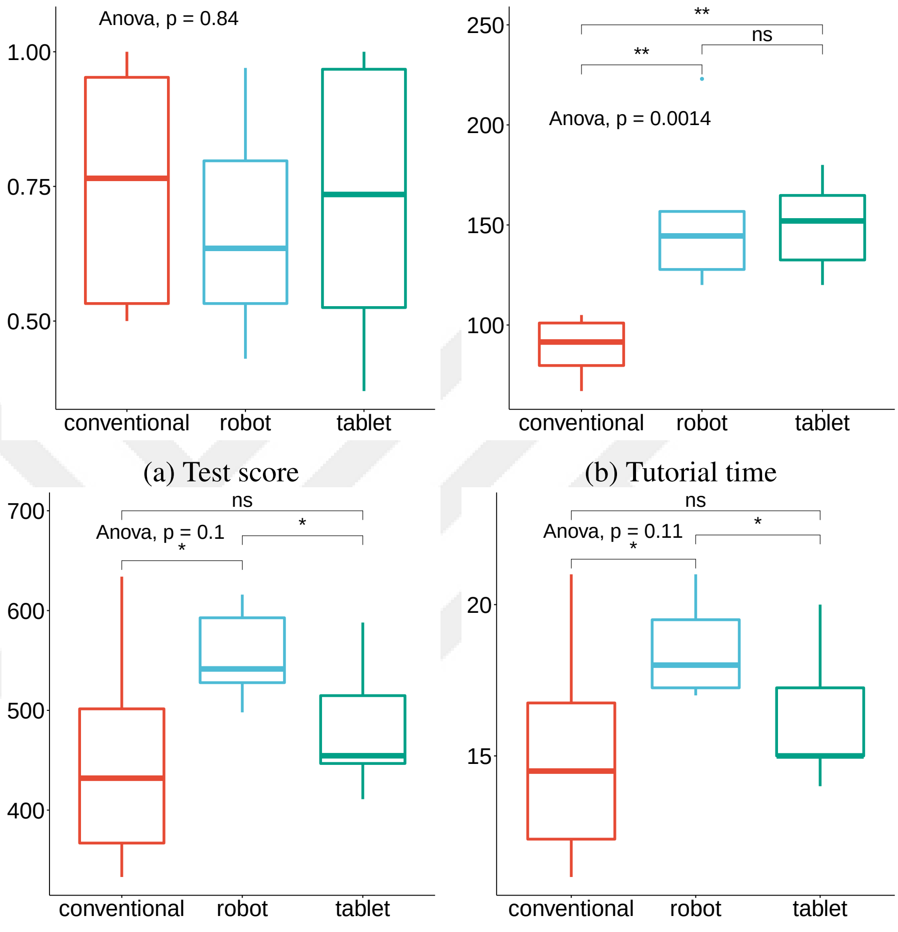


Figure 2.7 : The distributions of the test metrics of the group tested with the same item set in all the setups, and t-test significance between setup pairs: (ns) not significant, (*) and (**) are significant at $p \leq 0.05$ and $p \leq 0.01$, respectively.

2.3.2.3 Impressions of children about the affective robot

The children's answers to the post-test questionnaire reveal that they have positive impressions about the robot, its behaviours and abilities. The number of affirmative "yes" answers and the distribution of the other options are shown in Figure 2.8. The post-test on the children's impressions about the perceived character and presence of the robot showed that the children were really excited to see the robot (Q1), liked the robot (Q2) and playing with it (Q5). The results point that the children would like to play with the robot again (Q6) and they want to be friends with the robot (Q7).

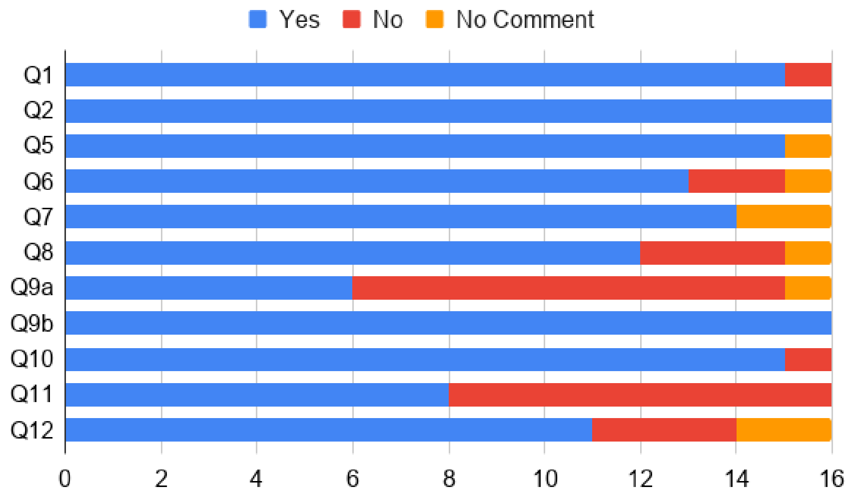


Figure 2.8 : The distribution of children's answers for yes-no questions.

Most of the children reported that they would like to have a robot at home (Q8) and in their hospital (Q9b) but they were not in agreement about the presence of the robot in their school (Q9a) and the character of the robot whether it is more like a friend or teacher (Q13). A χ^2 test was performed to determine if the presence of the robot in school was independent from its perceived character because of the disagreement between the children. The results showed that the two are not independent ($\chi^2(4, N = 16) = 9.079$, p = 0.059 and p = 0.035 with Monte Carlo simulation due to the small sample size).

On the other hand, the result showed that the children found the robot intelligent (Q10) and funny (Q12) but they were in disagreement about Pepper looking like a human (Q11). When they were asked to elaborate why the robot do not look like a human, they reported that its color was wrong, it had no hair, it was not very talkative and not moving enough.

Finally on the open-ended questions, the children reported that they liked the most (Q3) its saying "bravo" and its being a robot. The children also reported that there was nothing about it they did not like (Q4) but if they were the one "teaching the robot to do something" they would teach it to move more, to play various games and to study with them for their lessons.

The answers of the 16 children taking the auditory test in robotic setup are encoded and quantified for both survey conducted in both case studies. The online survey on the validation of the affective behavior of the robot is included as the “second survey". The answers of the children were scored out of 100, the children’s scores varying from 0.33 to 1 (M = 0.75 and SD = 0.21) were integrated into a correlation analysis to examine the correlations among different factors influencing children’s impressions about the robot. The correlated factors are displayed in Figure 2.9.

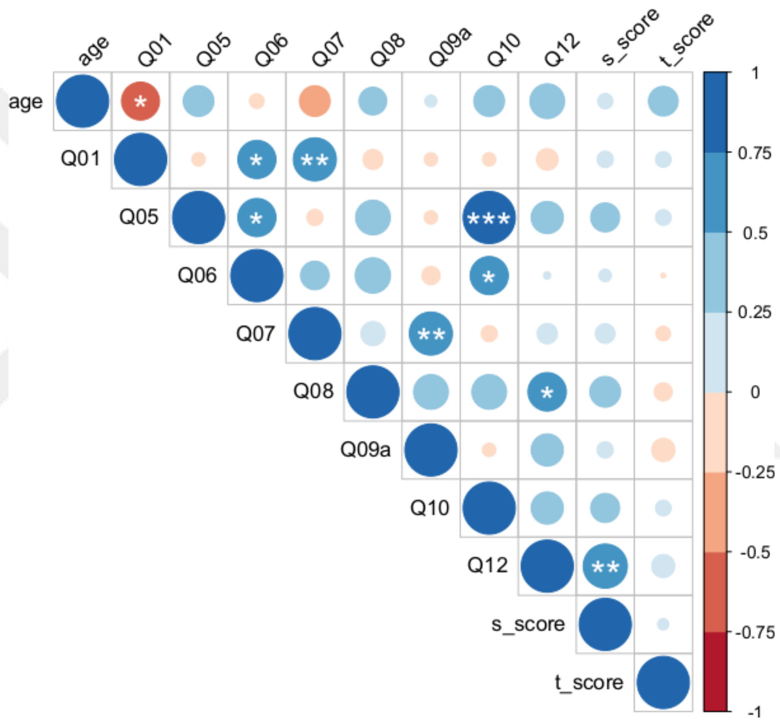


Figure 2.9 : Correlation among different factors on impressions of children: Correlations with (*), (**) and (***) are significant at $p \leq 0.05$, $p \leq 0.01$ and $p \leq 0.001$ respectively.

While the negatively correlated items showed that younger children were more excited to see the robot, the positively correlated items revealed that the children who described themselves as “excited to see the robot" would like to be friends with the robot and play again with it. The children who liked to play with the robot and would like to continue to play with it, found the robot intelligent and funny. There was also a positive correlation with the funny aspect of the robot and the children’s desire to have a similar robot in their home. Another positive correlation were found between the answers of the children who would like to be friends with the robot and the children desiring to see the robot in their schools. The results also revealed that children who correctly interpreted robot's behavior in the first case study(having a higher *s_score*) found the robot more funny than the others.

Another aspect revealed by the correlation analysis was that the auditory perception test score (*t_score*) of the children did not have any significant correlation with any of the other factors. It is a promising outcome because it reveals that even though the children do not perform well during the auditory test, they do not relate it with their experience with the robot and their impressions about it.

2.3.3 Physiological signal-based classification

In order to explore the effect of the robot and/or the gamified test setups on the physiological signals of the children, a classification study were performed on the collected data set [100]. The data collection, the preprocessing and the classification tasks are explained in the following subsections.

2.3.3.1 Physiological data collection and processing

Physiological data of children were collected using the E4 wristband during the test study. Skin conductivity (SC) data was acquired with an Electrodermal activity (EDA) sensor at 4 Hz. Blood volume pulse (BVP) data was acquired via the PPG sensor with a sampling rate of 64 Hz. Skin temperature (ST) data was acquired with infrared thermopile with a sampling rate of 4 Hz.

The data collected from the EDA and ST sensors were upsampled to 64 samples per second so that the number of samples of all sensors is equal. The data from all sensors were epoched into 1-second slices. Multisensor data fusion at the data-level was used to fuse the information from all sensors, then feature vectors (64x3) were obtained for each 1-second slice. The work flow for the physiological data-based classification tasks is schematized in Figure 2.10.

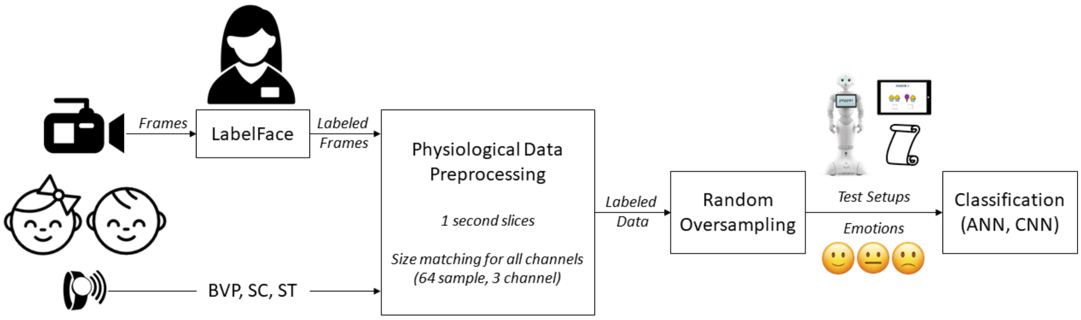


Figure 2.10 : Work flow of classification task using physiological signals.

The classifiers were trained for two classification tasks. The first one is the classification of test setups based on the captured physiological signals. And the second is the classification of emotions based on the test setups. We selected commonly used machine learning methods for the classification task based on physiological signals. The classification performance of Support Vector Machine (SVM), Random Forest (RF) and Artificial Neural Network (ANN) were analyzed and compared based on their classification performance. Based on this analysis, the model with the highest average accuracy and F1-score in all classification tasks was selected to be compared with a deep learning method: Convolutional Neural Networks (CNN).

The ANN model, which was the classifier with the highest performance metrics, was used as a classical machine learning approach. The neural network architecture consisted of only 1 hidden layer with 64 units using a sigmoid function. A fully connected dense layer, which used the softmax function, was added. The loss function was selected as binary cross-entropy for binary classifications, and categorical cross-entropy was used for the multi-class classifications.

CNN was selected as a deep learning method for the classification task. CNN model consisted of one convolutional layer with 8 filters which had kernel sizes of 8-by-2. A maximum pooling layer followed with the pool size of 64-by-1. The same arrangement as the classical neural network was set at the last layer because of the binary and multi-class classification tasks.

The classifiers and their performance for the classification of test setups (conventional, tablet, and robot) and the classification of emotions (pleasant, unpleasant, and neutral) are given in details in the following subsections.

2.3.3.2 Classification of test setups

The effects of gamified setups on children were evaluated by classifying the physiological signals captured during the three different test setup in this study. The data samples were labeled considering the tools used in the experimental setups: the robot+tablet (R), the tablet (T), and no tool (conventional (C)) for different item sets (environmental sounds, tonal patterns, and both of them).

The classifiers were trained and tested for both binary (robot - tablet (RT), robot - conventional (RC), tablet - conventional (TC)) and multi-class (robot - tablet - conventional (RTC)) classification. The training and test sets were prepared for three-fold cross-validation. Due to the unbalanced numbers of samples for each class (R, T, C) of the classification tasks (RT, RC, TC, and RTC), the training dataset was augmented using the random oversampling method. The classification accuracies that were obtained for RT, RC, TC and RTC are presented in Table 2.11.

Table 2.11 : Classification results of RT, RC,TC, RTC using ANN and CNN.

Item set	Metric	ANN				CNN			
		RT	RC	TC	RTC	RT	RC	TC	RTC
env	Accuracy	0.990	0.980	0.747	0.809	0.988	0.976	0.753	0.817
	F1-score	0.990	0.980	0.743	0.734	0.988	0.976	0.744	0.740
ton	Accuracy	0.994	0.994	0.713	0.796	0.995	0.994	0.765	0.857
	F1-score	0.994	0.994	0.712	0.718	0.995	0.994	0.763	0.798
env+ton	Accuracy	0.989	0.984	0.645	0.747	0.991	0.985	0.619	0.750
	F1-score	0.989	0.984	0.642	0.656	0.991	0.985	0.615	0.655

The results show that the classification accuracy scores of both ANN and CNN models were higher when robot was involved (Table 2.11). Both RT (environmental sounds, 99%-ANN and 98.8%–CNN and tonal patterns, 99.4%-ANN and 99.5%-CNN) and RC (environmental sounds, 98%-ANN and 97.6%–CNN and tonal patterns, 99.4%-ANN and 99.4%-CNN) were classified accurately. The results comply with the hypothesis that the robot acted as a stimulating agent and caused difference in the physiological signals of children. Since the test technology between two cases (R and T), test questions and tablet interface were identical in both cases; the interaction with the robot and the feedback from the robot, as well as the physical embodiment of the robot might cause such difference in the physiological signals compared to the other cases T and C (environmental sounds, 74.7%-ANN and 75.3%–CNN and tonal patterns, 71.3%-ANN and 76.5%-CNN). The multi-class classification (RTC) also resulted in high accuracy but with lower F1-scores. Increasing the number of children might be helpful to improve the performance of the multi-class classifiers for future studies.

2.3.3.3 Classification of emotional states

The results of the classification task for the test setups, presented in Table 2.11 in the previous subsection, comply with our assumption that the robot stimulated the children more than the other two test setups. Based on this, we decided to analyze the emotions of children during the auditory tests, especially with the robotic setup.

The emotions of the children when they were using a robot, tablet, or no tool at all (conventional setup) have been classified in this study. Data samples were labeled as pleasant (P), unpleasant (U), and neutral (N) considering the emotions assigned by the psychologist in the experiment during the auditory perception tests. The classifiers were trained and tested for both binary (pleasant - unpleasant (PU), pleasant - neutral (PN), neutral - unpleasant (NU)), and multi-class (pleasant - neutral - unpleasant (PNU)) classification. The training and test sets were again prepared for three-fold cross-validation. Because of the unbalanced numbers of samples for each class (P, N, U) for the classification tasks, the training data set was again augmented using the random oversampling method.

The emotion classification accuracy scores using ANN and CNN were presented in Table 2.12 and Table 2.13. Since there was few samples in emotion tasks, the number of samples belonging to different classes in the test sets remained unbalanced when trying to select as many samples as possible. Therefore, F1-score, recall, precision, specificity, Matthew's correlation coefficient (MCC), the geometric mean of recall and precision (GSP), the geometric mean of recall and specificity (GSS) values besides classification accuracy scores for emotion classifications are also displayed in Table 2.12 and Table 2.13.

Table 2.12 : Classification results of PU, PN, NU, PNU in robotic setup using ANN and CNN.

Item set	Metric	ANN				CNN			
		PU	PN	NU	PNU	PU	PN	NU	PNU
R_{env}	Accuracy	0.771	0.562	0.628	0.486	0.792	0.521	0.654	0.445
	Precision	0.843	0.565	0.646	0.507	0.809	0.544	0.657	0.476
	Recall	0.771	0.562	0.628	0.328	0.792	0.521	0.654	0.285
	F1-score	0.757	0.560	0.622	0.373	0.783	0.482	0.652	0.333
	Specificity	0.771	0.562	0.628	0.671	0.792	0.521	0.654	0.645
	GSP	0.782	0.562	0.630	0.394	0.792	0.506	0.654	0.355
	GSS	0.735	0.557	0.614	0.448	0.773	0.438	0.650	0.410
	MCC	0.610	0.127	0.273	0.001	0.600	0.062	0.311	-0.071
R_{ton}	Accuracy	0.521	0.542	0.682	0.472	0.458	0.604	0.636	0.389
	Precision	0.524	0.538	0.686	0.489	0.460	0.612	0.640	0.415
	Recall	0.521	0.542	0.682	0.314	0.458	0.604	0.636	0.249
	F1-score	0.514	0.523	0.680	0.368	0.455	0.593	0.631	0.287
	Specificity	0.521	0.542	0.682	0.650	0.458	0.604	0.636	0.584
	GSP	0.518	0.532	0.682	0.384	0.457	0.601	0.634	0.308
	GSS	0.507	0.500	0.677	0.440	0.450	0.580	0.624	0.354
	MCC	0.045	0.081	0.368	-0.035	-0.082	0.216	0.276	-0.179

Table 2.13 : Classification results of emotions in tablet and conventional setups using ANN and CNN.

Item set	Metric	NU		Item set	Metric	PN	
		ANN	CNN			ANN	CNN
T_{env}	Accuracy	0.960	0.980	C_{env}	Accuracy	0.762	0.774
	Precision	0.963	0.979		Precision	0.785	0.846
	Recall	0.947	0.977		Recall	0.762	0.774
	F1-score	0.954	0.977		F1-score	0.758	0.760
	Specificity	0.947	0.977		Specificity	0.762	0.774
	GSP	0.954	0.978		GSP	0.766	0.784
	GSS	0.946	0.977		GSS	0.751	0.738
	MCC	0.910	0.956		MCC	0.546	0.614
T_{ton}	Accuracy	0.841	0.952	C_{ton}	Accuracy	0.795	0.487
	Precision	0.904	0.968		Precision	0.692	0.455
	Recall	0.762	0.929		Recall	0.690	0.444
	F1-score	0.789	0.943		F1-score	0.671	0.441
	Specificity	0.762	0.929		Specificity	0.690	0.444
	GSP	0.811	0.945		GSP	0.680	0.445
	GSS	0.722	0.924		GSS	0.508	0.365
	MCC	0.650	0.895		MCC	0.426	-0.100

The results showed that ANN and CNN classifiers performed best in the classification of PU in the environmental sounds with the robot (77.1%-ANN and 79.2%–CNN). However the highest classification accuracy results (68.2%-ANN and 63.6%–CNN) in the tonal patterns were obtained in NU classification for the robotic setup. Since the audiologists define tonal pattern tests as harder than the one with the environmental sounds, the children display less emotions during the tonal pattern test because they are more focused on the auditory task. Therefore, the highest accuracy in NU classification might be due to the higher number of samples in neutral and unpleasant labels than pleasant ones during the tests with the tonal patterns.

After the annotation task was finished by the experts, we also noticed that the number of samples in T and C test setups were very low for classification of emotions. Thus, we only run the NU classification for T (environmental sounds, 96%-ANN and 98%–CNN and tonal patterns, 84.1%-ANN and 95.2%-CNN) and PN classification for C setup (environmental sounds, 76.2%-ANN and 77.4%–CNN and tonal patterns, 79.5%-ANN and 48.7%-CNN) (Table 2.13). The reason behind the low number of labeled samples for the other emotional cases might be due to the fact that the children were less emotionally stimulated in these tests. When they used tablet, they showed only neutral or unpleasant emotions, which might be due to the fact that the children were not motivated. On the other hand, in C test setup, PN could be classified because of the high number of positive and neutral labeled samples and no unpleasant samples. This might be due to the individual motivation of the audiologist, who know the children for a long time (familiarity effect).

Note that if the audiologist is inexperienced or unfamiliar with the children in the test, this will effect the motivation and performances of the children in the conventional test setup. Thus, use of socially assistive robots and tablets may standardize the tests, and the feedback mechanism, and avoid the familiarity bias.

2.4 Chapter Summary

RoboRehab is an assistive robotic system enhanced with an affective module for children with hearing disabilities. The system is designed and developed for the audiometry tests and rehabilitation of children in clinic settings. The system involves a social robot Pepper, tablet, an interface specially designed for the verbal and nonverbal interaction of children with the robot, gamification of the tests, sensory setup and a machine learning-based emotion recognition module to achieve this goal.

This chapter presents a user study involving 16 children with hearing disabilities (using either cochlear implant or hearing aid). Two case studies with children were presented.

The results of the first case study show that the robot was able to display affective behavior and the children could correctly identify and interpret these behaviors. These findings answers the first research question (RQ1).

In the second case study, the conventional setup, setup with robot+tablet and setup with only tablet conditions were tested, and the results were reported based on the detailed analysis of physiological signals captured by E4, performances of children during the auditory perception tests, and self-report questionnaire. The results reveal that the robotic setup is well separated with high classification accuracy from the other test setups for the auditory tests. The physiological signals were also used with ANN and CNN models to recognize the emotions of children (pleasant, unpleasant or neutral) for the test setups. The results show that the children's positive and negative emotions can be distinguished more accurately when they interact with the robot, than the other two cases in the environmental sounds test. The overall results show that the robot can stimulate the children's emotions and cause difference in their physiological signals and machine learning approaches can be used to understand the emotions of children during their interaction with a socially assistive robot. This findings show that using physiological signals to differentiate the emotions of children may be an efficient solution to answer the second research question (RQ2).

The evaluation based on objective test metrics show that the test scores of children in tablet and robot setups were as good as the conventional tests with human audiologist. Moreover, the results show that the time spent with the robot is significantly higher than the conventional test but it is observed by the audiologists and experimenters that in the robotic setup, the children were more involved and showed attention to the test than the other setups. The subjective evaluation of the audiologist comply with the findings of previous studies: In robotic studies, even though the test performances of participants do not differ significantly between the robotic and conventional setups and the test duration increases when the robot is present, the results report higher engagement and higher cooperation times by the participants [101,102]. Therefore, we expect that the proposed method will facilitate the children's experience during their auditory tests, increase the interaction and cooperation time and improve the interaction quality, especially for individuals with additional disabilities, who have short attention and cooperation time.

The self-report surveys conducted after auditory tests with the robotic setup show that the robot is accepted as an intelligent and funny social entity by the children. Although there was no significant difference of using a robot in terms of test scores, the subjective evaluation of the robotic setup reveal that the children were excited to see the robot and happy to play with it. Additionally, the audiologists reported that the children were more motivated and engaged in the digitalized setups than the conventional setup. This is a motivating preliminary step to use these affective robots as a part of audiometry testing. These findings answers the fifth research question on the improvement of the interaction gain with an affective robot (RQ5).

3. A MULTIMODAL APPROACH TO DETECT THE EMOTIONAL STATE OF THE HUMAN COMPANION

To design and model an emotional understanding mechanism for an assistive robot, first of all the robot should be able to identify and interpret its human companion's affective states accurately. For the assistive robot to be able to recognize the emotions of its human companion, it should process multi-channel stimuli coming from different sensors.

In order to evaluate available emotion recognition technologies and to explore the different modalities for automatic recognition of emotions, a set of user studies were conducted in accordance with the Erasmus+ project, entitled “Affective loop in Socially Assistive Robotics as an intervention tool for children with autism", abbreviated as EMBOA.

3.1 EMBOA Project

EMBOA is a research project with the aim of combining affective computing technologies with the social robot intervention in children with autism. Children with autism spectrum disorder (ASD) are known to display limited social and emotional skills in their routine interactions. The social robotics field provides promising results in the use of robots to assist children with ASD in the development of their social and emotional skills, help them to overcome social obstacles and make the children more involved in their interactions.

The EMBOA project is an international research project combining three domains of research; autism therapy, social robots and automatic emotion recognition, in order to develop guidelines and practical evaluation of applying emotion recognition technologies in robot-supported intervention in children with autism. The motivation is to investigate the feasibility of the available practices and find a novel approach for creating an affective loop in child-robot interaction that would enhance the intervention regarding emotional intelligence building in children with autism [103]. The future aim of this approach is that a social robot can automatically recognize the emotions of a child and adapt its reactions according to the child's current needs and affective states.

3.2 Experimental Setup

The ethical board of the Gdansk University of Technology, Poland approved the EMBOA user studies. The parents of the children provided written informed consent, in which they declared: (1) They are informed about the purpose, the experimental procedure of the study, and the data acquisition setup. (2) They understand why their and their children's consent is needed. (3) They may refuse their children's participation in the study at any time during the project. (4) They agree that their children can be recorded during the study. (5) They consent to process their children's data within the General Data Protection Regulation (GDPR).

The profile of children participating in the study, social robot Kaspar, interaction scenarios for EMBOA project and interaction setup were explained in details in the following subsections.

3.2.1 Profile of children with ASD

The inclusion criteria in the study were determined by the EMBOA project consortium as children diagnosed officially with ASD, under treatment or not, without any other known neurological or psychological diagnosis. In addition, the age range of the study participants is selected to be between 2 and 12 years-old.

The interaction studies were conducted in 4 collaborating countries (the institutional abbreviations for the partners are reported between parentheses following the name of the country) : Macedonia (MAAP), Turkey (ITU-YU), United Kingdom (UH) and Poland (GUT).

The user studies for the EMBOA Project were conducted in two rounds with a one-year-gap between the rounds of studies. The studies were conducted with 65 children in total. The demographic information, diagnosis, education and/or rehabilitation history, and siblings information were collected. But there is an ongoing study on the second round of studies in the international level, therefore only the demographic data of children participating in the first round of studies, as well as the repeated session numbers were reported and displayed in Table 3.1. For the two rounds of user studies conducted in Turkey, please refer the Section 3.3.

Table 3.1 : Profile of children for the first round of studies.

Institutional Abbreviation	No. of Children (Gender)	Age Range	No. of Sessions (Min,Max)
MAAP	11 (9 M, 2 F)	2-6	(2,11)
ITU-YU	8 (7 M, 1 F)	6-10	(0,1)
UH	7 (6 M, 1 F)	10-12	(2,3)
GUT	3 (M)	6-6	(1,4)

3.2.2 Kaspar: The social robot

Kaspar is a child-sized humanoid robot, it is 55cm high and 45cm wide and interacts with people, mostly children, in a seated posture, as seen in Figure 3.1. It has embedded 2D cameras, touch sensors and microphones to perceive its environment and interact with it.

Figure 3.1 : Kaspar, curtsy of the Adaptive Systems Research group, University of Hertfordshire, UK.

The robot, having a silicone face, is capable to have simple expressive features and show different expressions. It's been used for robot-assisted therapy for children with autism and as a research platform to study social robotics. Kaspar can be used as a social mediator and help children to better interact and communicate with their environment, to explore and recognize basic emotions, to learn about socially acceptable tactile interaction and learn fundamental social skills such as imitation, turn-taking and collaboration skills.

Kaspar has three modes of operation, the first one is the remote control. In the study with children it is remotely controlled via a keypad communicating with the Kaspar GUI. The second mode is the autonomous response to touch and the third one is a hybrid control mode which is the combination of both previously mentioned modalities.

The Kaspar GUI provides the developers to control the motions and expressions of the robot, implement and design new ones using poses composed of joint positions, sequences composed of poses and keymaps to control sequences. The robot is equipped with 13 predefined interactive game scenarios based on imitation, turn-taking, cause and effect relation, and collaboration.

3.2.3 Interaction scenarios

A set of interaction scenarios were implemented with the Kaspar robot with the aim to evaluate the feasibility of available automatic emotion recognition technologies in robot-assisted intervention settings.

Each interaction scenario contains positive and negative verbal and behavioral feedback:

- The robot sings the song “If you’re happy and you know it" and claps its hands as part of a short familiarization process, the same song is also used for engaging the children when they get distracted during the interaction games,
- The robot uses short and clear sentences such as “What’s your name?", “Can you show me [...] ?", “What does the [animal] say?, “Thank you", etc. for reciprocal communication with the children,
- Kaspar mimics a specific set of animals: monkey, cat, dog, cow, duck and pig to engage and entertain children,

- Kaspar uses the sentences “This is nice, well done" or “Well done"+trumpet sound to reward and engage the children when they correctly perform the actions prompted by the robot,
- The robot uses the verbal feedback “Auch, you’re hurting me!" for the times the children get violent toward itself, and also hides its face and torso as if it really hurts.
- When the robot asks the children to mirror its movements and the children do not comply or can not perform correctly the requested action, then it uses a neutral feedback as “Not quite, try again" to encourage the children.

Each interaction session starts with Kaspar introducing itself and inviting the children to play with it. And, each session ends with Kaspar saying “Bye bye, see you next time". All scenarios contain positive and neutral feedback sentences, as well as the "Auch, you’re hurting me!" reaction.

Each scenario with Kaspar is based on the principle of turn-taking, imitation, and role changing (i.e. Kaspar can ask the children to do something or the children can ask Kaspar to do it). These interaction scenarios require basic receptive and expressive language skills related to emotions, animals, and body parts. There are 4 interaction scenarios:

1. Standard: Kaspar introduces itself, and plays an imitation game with the child. The scenario does not require high communication skills, once Kaspar introduces itself, the therapist prompts the robot to perform some upper body gestures using only its arms, such as raising up its right hand, pointing left with its left hand, etc. and asks the children if they can copy Kaspar’s movements. If the child performs correctly then a positive feedback is given, if not then the robot uses the neutral feedback and asks the child to repeat again. The song “If you’re happy and you know it" is included in this scenario, and used for familiarization with the robot or get the attention of the child back to the robot.

2. Emotions: In this scenario, Kaspar asks to see the happy/sad/surprised and "hiding" expressions of the child. The robot itself does not display the corresponding emotions but gives feedback when the child performs the facial expressions.

3. Animals: This scenario incorporates animal sounds and mimics. Kaspar asks the child what does the dog, cow, cat, duck, pig, monkey say and requests the child to imitate these animals both verbally and behaviorally. The robot also uses the animal sounds to engage and amuse the children.

4. Body parts: This scenarios consists of a pointing game where Kaspar asks the children to show their head, nose, mouth, eyes, ears, toes, hands. When it is the robot's turn, it points out and says the name of the corresponding body part as well.

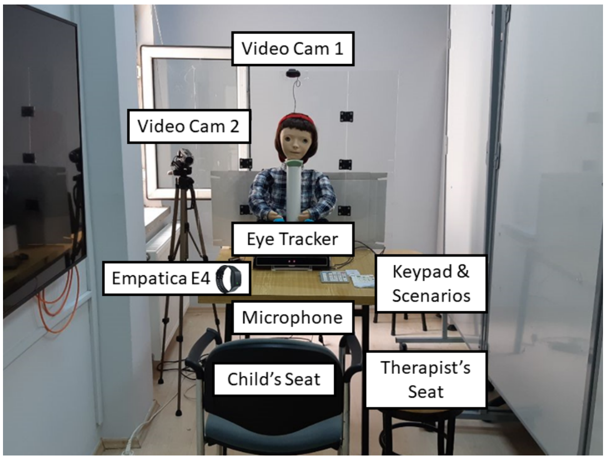
The flow of the entire interaction session is determined by the therapist or experimenter according to the child's profile or reactions. The scenarios are not ordered, they can be repeated multiple times or some scenarios can be skipped if the child gets bored.

3.2.4 Interaction setup & data collection

The setup for the interaction and data collection were implemented in an allocated room in the university laboratory in Istanbul Technical University (Figure 3.2a). The children were equipped with the Empatica E4 smart wristband during the interaction session (Figure 3.2b). The real-time physiological signals, i.e. BVP, ST, SC, were collected with E4 wristband. The facial expressions of children were collected by two video cameras, one placed on the transparent screen above Kaspar (cam 1), and the second one placed on the right side of the robot (cam 2). In addition, the whole session was recorded with a supplementary video camera positioned to monitor Kaspar's movement.

3.3 User Studies in Turkey

The user studies in Turkey were conducted in two rounds. Before the first round, a preliminary study in Istanbul Technical University (ITU) were performed with a small



(a) Data collection setup for user studies.

(b) A child mirrors Kaspar.

Figure 3.2 : Data collection and interaction setup for EMBOA case studies.

set of children with ASD and children with typical development for testing the setup and the sensors to be used.

3.3.1 Preliminary study in Istanbul Technical University

The interaction studies were conducted at the ITU laboratory (Figure 3.2b). A total of 14 children with ASD were invited for the study but the drop-out was approximately 50% and only 5 of 8 children who showed up completed the study with the robot. The mean age of children interacting with Kaspar was 6 years-old. They were all high functioning boys with high scores on language understanding, verbal skills, animal and body parts knowledge. They interacted and played with Kaspar at least 10 minutes and all the scenarios were used during the interaction session.

The experimental setup was also tested with 3 typically developing (TD) children: two male siblings (3.5 and 6.5 years old) and one 9-year-old girl. Despite the simple content and basic knowledge of interaction scenarios, the older children were cooperative and they played along with the robot. But the youngest boy did not cooperate and did not want to play with the robot, he accepted to sit in front of the robot but did not react to its requests or animations.

3.3.2 User studies in a special education center

The first and second rounds of the interaction studies were conducted in a special education and rehabilitation center in Mecidiyeköy with a one-year-gap between the studies.

In the first round of studies displayed in Figure 3.3a, a total of 10 children with ASD were selected and invited by the therapists of the center following the inclusion criteria determined by the EMBOA consortium. The drop-out for the study was fortunately 20% and 8 children who showed up completed the study with Kaspar.

In the second round shown in Figure 3.3b, a different set of 10 children were invited by the therapist, even though there were no-show participants there were also walk-ins therefore there was no drop-out.

(a) First round of studies. (b) Second round of studies.

Figure 3.3 : EMBOA case studies with Kaspar and children with ASD.

The user studies in Turkey were conducted with 21 children with ASD in total. The ages of children (2 girls, 19 boys) interacting with Kaspar ranged from 5 to 10 years-old, with mean M = 7.13 and standard deviation SD = 1.55. The children were diagnosed with either atypical autism or pervasive developmental disorder, and they had at least 1 year therapy background. They were all high functioning children with high scores on language understanding, verbal skills, knowledge about animals and body parts.

They interacted and played with Kaspar at least 10 minutes and all the scenarios were used during the interaction session. The interaction study were performed in only one session, and it couldn't be repeated first due to the holiday season, and then due to the Covid outbreak.

3.3.3 Highlights from the experimenter's point of view

Standard subjective evaluation questionnaires, such as NARS [64] or Godspeed [76] questionnaires, were not used for the evaluation of the interaction studies conducted for the EMBOA project. In order to compensate the absence of subjective evaluations of children and their impression about the system, the comments of the experimenter working with the children were summarized below:

- Reluctance to play with Kaspar: Some of the children did not comply to play with Kaspar, they refused to get closer to the robot or sit in front of it, and one of them tried to hit the robot and harm the keypad and its accessories. They got stressed and irritated, and angry, they started to display negative behaviors and wanted to leave the allocated room.

- Non-responsiveness to Kaspar: At first, some of the children were non-responsive towards the robot and the experimenter. They were out of focus, displayed some repetitive behaviors, repeated out of context concepts and required additional help from their family members or therapist accompanying them.

- Physical discomfort due to E4 wristband: Some of the children had low physical intolerance for the wristband, they did not want to put the E4 on their wrist. They got irritated and did not cooperate with the robot as long as the wristband were on their wrist. Therefore we had to remove it to let them continue their interaction with the robot. As an alternative, we discussed that the wristband could be put loosely but then the data quality would be questionable. Therefore to minimize the discomfort of the children and distract them from the wristband, we decided to use colored bandannas to cover the wristband as an accessory, to make the E4 something more bearable and attractive.

- Noise sensitivity: Some of the children were highly sensitive to noise, they got irritated with the standard scenario having the children song. Even though we adjusted the sound level of Kaspar, some children plugged their ears during the interaction and they refused to cooperate.

On the other hand, the families of children reported that their child were happy to see the robot and play with it, and eager to take part in the interaction especially after they got used to Kaspar. They pointed that standard scenario with the song was the most popular among the children. Even the children who did not sing along with Kaspar, clapped synchronously during the song as a part of the interaction emerging between the child and the robot.

3.4 Multimodal Aspect of Emotional Expression

The behavioral and emotional annotation of child-robot interaction scenarios is a challenging task, as well as the evaluation of the interaction. Automated annotation of the collected data is crucial in order to reduce the time complexity of the manual annotation task conducted by experts, and to infer quantitative and objective evaluation metrics directly from the interaction data.

Two different approach is proposed for the automatic annotation of the multimodal data collected during the EMBOA user studies. The first approach is based on the use of facial expressions and focuses on the annotation task. The second one is based on the combination of various interaction-specific metrics and sensory data focusing on the annotation as well as the evaluation of the interaction.

3.4.1 Vision-based emotional metrics for data annotation

The first approach for the annotation rely on the principle of using the facial cues to infer the emotional state of children, and use the inferred emotion label to annotate the synchronized data coming from the other sensor modalities.

The vision-based emotional metrics are used in order to detect and recognize the emotional state of children. These metrics are the facial action units (AUs) representing the subtle changes in the facial muscles, i.e. the modular components of emotional expressions.

The video-based feature extraction was performed by the use of an open source toolkit for facial landmark detection, head pose estimation, facial action unit recognition, and eye-gaze estimation: OpenFace [104].

OpenFace provides:

– the intensity (from 0 to 5) of 17 AUs: AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU45

– the presence (0 absent, 1 present) of 18 AUs: Previous 17 AUs + AU28

In the annotation study, the intensity information is discarded for further use, and the annotation is performed based on the presence of the selected AUs. The set of AUs used for emotional annotation and inference are presented in Table 3.2 based on the Facial Action Coding System (FACS) guidelines provided by Ekman [32] (for a good visualization of the descriptions please visit iMotions FACS – A Visual Guidebook [105]).

Table 3.2 : Action units for emotional expressions.

AU number	AU description	Emotional mapping
AU01	Inner Brow Raiser	Sad, Surprised, Afraid
AU02	Outer Brow Raiser (unilateral, right side)	Surprised, Afraid
AU04	Brow Lowerer	Sad, Afraid, Angry
AU05	Upper Lid Raiser	Surprised, Afraid, Angry
AU06	Cheek Raiser	Happy
AU07	Lid Tightener	Afraid, Angry
AU09	Nose Wrinkler	Disgusted
AU12	Lip Corner Puller	Happy
AU15	Lip Corner Depressor	Sad, Disgusted
AU20	Lip stretcher	Afraid
AU23	Lip Tightener	Angry
AU26	Jaw Drop	Surprised, Afraid

The workflow for the AU-based annotation of children's emotional state from the video recordings can be summarized as follows:

- The video recordings are fed to OpenFace, OpenFace provides the information about the presence of 18 AU in a binary data format for each frame in the video,
- The binary data provided by OpenFace is transformed to emotion labels based on the FACS guidelines [32], e.g. if AU06 and AU12 are both present in a frame, then that frame is labeled as "Happy",
- If the frame is not labeled with a basic emotion (happy, sad, surprised, afraid, angry, disgusted), then it is labeled as "Neutral".

Once the video frames are labeled then the data from the corresponding time interval captured by the other sensor modalities are also labeled with the same emotion. Even though this approach provides a solution for the challenging task of manual annotation, the annotation performed by the experts is vital for the validation of the proposed automatic annotation system. Therefore there is an ongoing study conducted by the experts to assess the performance of the automatic approach.

3.4.2 Multimodal behavioral metrics for data annotation and evaluation of the child-robot interaction

In order to annotate and evaluate the child-robot interaction, a set of multimodal behavioral metrics are proposed based on the interaction flow. The set consist of specific features belonging to different modalities, mainly the features extracted from interaction logs, video and audio recordings, and physiological data collection. The proposed set of metrics for behavioral annotation are presented in Table 3.3.

In Chapter 2, the features extracted from the physiological data captured by the E4 smart wristband, interaction metrics derived from test logs and behavioral annotation study were used to differentiate three setups for a gamified audiology test: Setup with a tablet (*tab*), with a tablet combined with an affective robot (*rob*), and conventional setup (*con*). In the previous section, the vision-based metrics for automatic annotation were presented and examined in order to explore if they can be used for the same task.

In order to examine the feasibility of the proposed behavioral metrics, a preliminary study were conducted on the RoboRehab data collection, and once the user studies

Table 3.3 : The proposed set of multi-modal behavioral metrics for data annotation.

Modality	Feature	Definition
Interaction logs	Test score	The percentage of the correctly answered questions (RoboRehab)
		The percentage of correctly mirrored actions (EMBOA)
	Total test time	The duration in seconds of the auditory perception test composed of 30 questions (RoboRehab)
		The duration in seconds of the different interaction scenarios (EMBOA)
		The total interaction session duration (EMBOA)
Video recordings	Gaze Direction	The percentage of time the children change their gaze direction from: left→right, right→left, up→down, down→up
		The percentage of time when the children face away from the camera
	Facial action units	Intensity: The normalized intensity of the 17 facial action units (AU) detected by OpenFace
		Presence: The average occurrence of the 18 facial action units during the interaction
Audio recordings	Silence	The average percentage time of silence
	Speech activity	Children: The average percentage of time when the children verbally express themselves during the interaction
		Therapist: The average percentage of time when the therapist verbally interacts with children
Physiological data	Blood volume pulse	All the physiological markers derived from blood volume pulse
	Skin temperature	All the physiological markers derived from skin temperature
	Skin conductance	All the physiological markers derived from skin conductance

with children with ASD for the EMBOA project will be completely annotated, their interaction measures will also be analyzed and evaluated by the proposed behavioral metrics. The audio recordings and the features extracted from physiological data are left out of scope for the preliminary study because they do not meet with the assumption of normality for the hypothesis testing. The proposed metrics for the preliminary study are the combination of vision-based metrics and interaction-specific measures such as test scores, test duration, etc.

In the preliminary study, the video-based feature extraction was performed by the use of OpenFace [104]. The set of 40 video-based features is composed of:

− The percentage of time the children change their gaze direction:

(1) left to right, (2) right to left, (3) upword direction, (4) downward direction

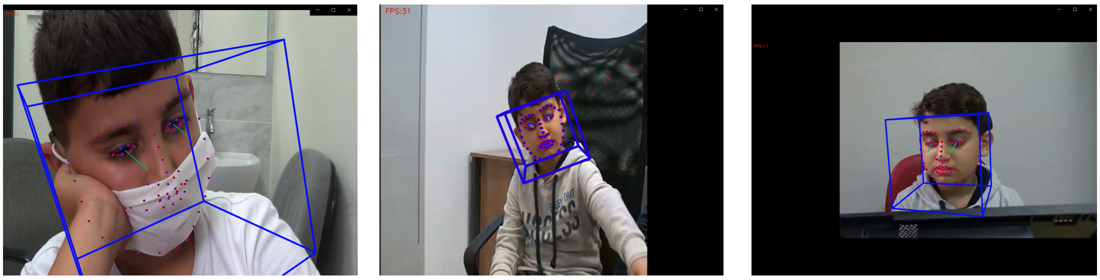
− (5) The percentage of time the children face away from the camera

− (6-22) The normalized intensity of the 17 facial action units detected by OpenFace: AU01, AU02, AU04, AU05, AU06, AU07, AU09, AU10, AU12, AU14, AU15, AU17, AU20, AU23, AU25, AU26, AU45

− (23-40) The average occurrence of the 18 facial action units detected by OpenFace: the previous 17 + AU18.

OpenFace is also capable to detect head pose and orientation but due to the different perspective and properties (image resolution and size) of the video recordings displayed in Figure 3.4, the detection performance for these specific metrics were not reliable therefore they were not included in the evaluation process. In addition to these factors, during the recent interaction studies the children were advised to wear facial masks, therefore the detection accuracy of OpenFace suffered from the partial occlusion, as seen in Figure 3.4a. The system output on the confidence of the tracker in current landmark detection were reported as 97% and 94% for tablet (*tab*) and robot (*rob*) setups but 69% for conventional setup (*con*) with the facial mask.

For the preliminary study on the efficiency of vision-based metrics, in order to eliminate the effect of the test-related factors on the test setup, a hypothesis testing study



(a) Conventional setup (b) Robot setup (c) Tablet setup

Figure 3.4 : Different recording perspective and conditions.

were conducted on the feature set extracted from the recordings of a child who was tested with the same item set in all the setups.

A one-way ANOVA test was performed on the visual features and paired t-tests were computed to perform pairwise-comparison between the means of the 3 groups (Figure 3.5). The results showed that the setup did not have any significant effect on the visually captured metrics of children. The paired t-test computed between the test setups displayed also a similar result.

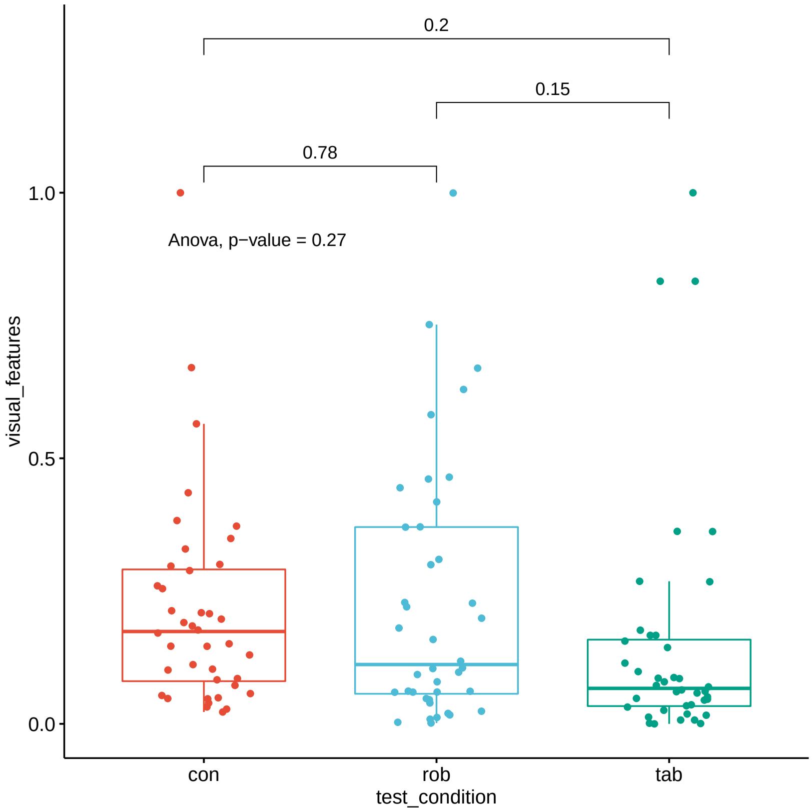


Figure 3.5 : The distributions of the visual metrics with the Anova and t-test significance between setup pairs with p-value greater than the significance level.

Although the preliminary study did not result in a significant outcome, the study will be repeated with the video recordings of the children with ASD once the annotation of the experts are finalized. As an alternative, another approach to detect the affective states of children with ASD were proposed and analyzed on the data collection for EMBOA project [106].

The proposed approach combines the use of facial action units and features obtained from physiological data collection by Empatica E4 wristband: Low frequency (LF) and high frequency (HF) power derived from BVP signal, and peak count and amplitude values derived from EDA signal. This multimodal approach incorporating the features extracted from the facial data and physiological data is displayed in Figure 3.6.

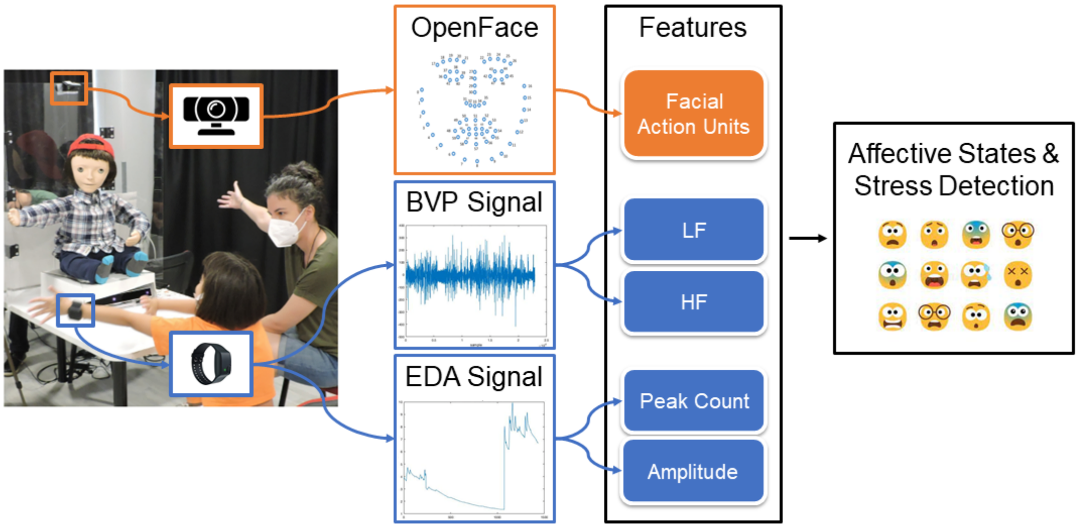


Figure 3.6 : Facial and physiological data combined for emotion detection.

The physiological features of children were used for affective state and stress detection in the study based on the previously provided reference baselines in the literature. The results were also compared with the affective labels from the automatic facial expression annotation study based on the videos recorded during the sessions synchronously, verified in a multi-modal manner. The results of the study support that the use of multimodal data is crucial, especially when the outputs of more than one signals conflict in the detection of affective states.

These findings supports the results presented in the previous chapter on the use of physiological signals to differentiate the emotions of children. They also reveal that using multimodal behavioral metrics, i.e. a set composed of facial action units, physiological features and interaction specific metrics may be an efficient solution to answer the second research question on the detection and interpretation of the human companion's emotions (RQ2). Moreover, the comments of the therapists and researchers working on the project support the hypothesis that a social robot with emotional understanding can foster the interaction gain in an assistive HRI scenario (RQ5).

4. A BEHAVIORAL MODEL FOR SOCIAL ROBOTS

For the affective robot to be able to generate social behaviors and in-situ motions, it is essential to interpret the emotional state of its human companion accurately, i.e. to process multimodal data coming form the human companion as well as the the context of the social interaction. Up to this point, the results showed that the robot is able to display basic emotional behaviors and detect the emotions of its companion using facial and physiological data. In this chapter, three emotional models are presented for the affective robots to adapt and moderate their emotional behaviors based on their human companions.

The first model is a linear model, CASPAR, incorporating an inner emotional state for the robot based on the mood and the previous emotions of its human companion. The second model is a vision-based model, where the robot detects the emotion if its companion based on FACS [32] and develops a behavioral strategy to adapt its emotional behaviors accordingly. And finally, an emotional model processing the interaction context as well as the affective states of the human companion is proposed based on the preferences and attitudes of 500 participants collected via an online study. The different emotional models are presented in the following sections respectively.

4.1 A Linear Emotion Model: CASPAR

A linear emotional model, entitled as "Computational Affect System for Personal Assistive Robots (CASPAR), has been conceptualized for an assistive robot. The proposed emotional model [107] is able to recognize its human partner's emotion via a vision-based recognition module and map the human's expression to a 3D Pleasure, Arousal, Dominance (PAD) space [31]. The model architecture for the linear emotional model CASPAR and a simulation study conducted with undergraduate participants on a desktop application with a simple graphical user interface are presented in the following subsections.

4.1.1 CASPAR model architecture

The proposed emotional model generates the next emotional state of the robot based on its mood, its inner state, its current affective state and the current affective state of the human companion interacting with it [107].

4.1.1.1 Mood calculation for the robot

The robot's mood, represented by constant m^t_{rob}, is calculated based on its emotional history. The emotional history consists of the last N emotions of the robot. The mood is computed with respect to the formula displayed in Equation 4.1:

$$m^t_{rob} = \begin{cases} 1 & \text{if } n_+ > n_- \\ 0 & \text{if } n_+ = n_- \\ -1 & \text{if } n_+ < n_- \end{cases} \tag{4.1}$$

Where n_+ represents the number of positive emotions in the past N records of the emotional history, and n_- the number of negative emotions. The emotional history of the robot is initialized with neutral emotions at $t = 0$.

4.1.1.2 Current emotional state of the robot

The current emotional state of the robot, represented by the emotional vector V^t_{rob}, is the most recent record in the emotion history. Its initial value is a 3D vector corresponding to a neutral area on the PAD space.

4.1.1.3 Inner state of the robot

The inner state of the robot, represented by the emotional vector V^t_{inr}, is modeled to represent the "gut feeling" of the robot. It is calculated based on the battery and/or motor heat level of the robot.

The inner state of the robot is initialized at full battery without any heating problem and adjusted to deteriorate gradually, i.e. the battery level decreases linearly and/or the heat level increases linearly, as time passes.

4.1.1.4 Current emotional state of the human companion

The current emotional state of the human companion, represented by the emotional vector V^t_{hum}, is computed by a vision-based emotion recognition module. The module recognizes the facial characteristics and maps it into the 3D PAD space based on a self-organizing map.

4.1.1.5 Next emotional state of the robot

The next emotional state of the robot, represented by the emotional vector V^{t+1}_{rob}, is calculated based on the current emotion of the human companion and the robot and its inner state as stated in Equation 4.2:

$$V^{t+1}_{rob} = \alpha_1 \times V^t_{hum} + \alpha_2 \times V^t_{rob} + \alpha_3 \times V^t_{inr} \tag{4.2}$$

$$\text{where} \sum_{i=1}^{3} \alpha_i = 1 \tag{4.3}$$

The α coefficients are used to prioritize the contribution of the variables in the Equation 4.2, e.g. if the model prioritizes the effect of the internal state of the robot on its next emotional state, then the α parameters can be initialized with the following set of values: $\{0.25, 0.25, 0.5\}$.

The proposed model also provides a mood correction phase for the next emotional state of the robot: If the mood is positive then the next emotional state of the robot is increased, if not it is decreased as shown in Equation 4.4:

$$V^{t+1}_{rob} = V^{t+1}_{rob} + m^t_{rob} \quad \text{where } m^t_{rob} = \begin{cases} 1 & \text{if mood is positive} \\ 0 & \text{if mood is neutral} \\ -1 & \text{if mood is negative} \end{cases} \tag{4.4}$$

The flow chart for a simulated interaction scenario is displayed in Figure 4.1. The emotion history of the robot is simulated with $N = 5$ records, the history is initialized with neutral emotions, therefore the initial mood of the robot is also neutral. The α parameters are initialized with the following set of values $\{0.5, 0.25, 0.25\}$ respectively in order to prioritize the contribution of the human companion's affective state on the next emotion of the robot. In the simulation scenario, the human companion's affective state changes from “very happy" to “very sad". The effect of the human’s emotional state on the robot’s emotions and its internal emotional process may be observed from the Figure 4.1 in details.

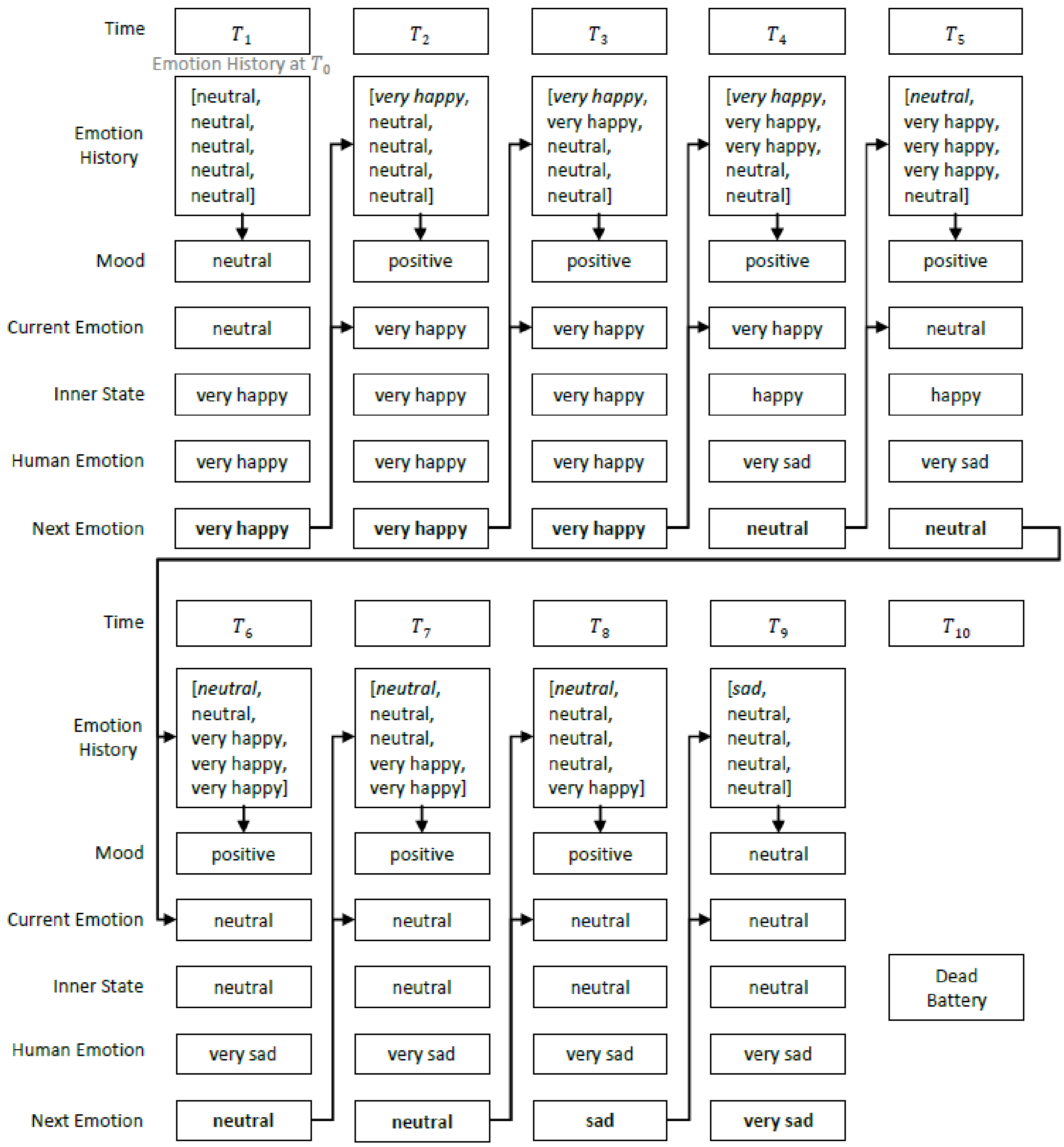


Figure 4.1 : The change in the robot’s affective state generated by CASPAR model from T_0 to T_{10} while the human companion’s affective state changes from “very happy" to “very sad".

4.1.2 Preliminary study

To explore the acceptability of the CASPAR model, a desktop application with a simple graphical user interface (GUI) was implemented to simulate the emotional changes of the robot.

4.1.2.1 Graphical user interface implementation for CASPAR

A GUI was designed and implemented in Python to evaluate the proposed emotional model and run simulations with different set of parameters. The simple GUI is displayed in Figure 4.2.

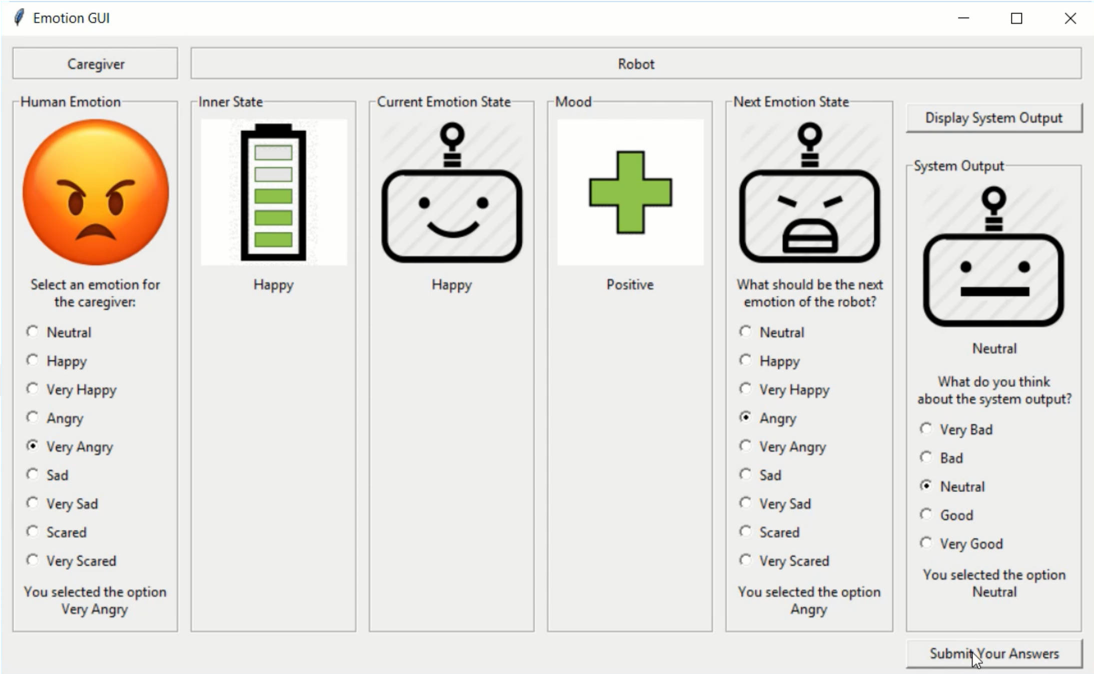


Figure 4.2 : A screenshot of the GUI implementation for CASPAR emotional model.

The first column of the GUI is reserved for the human companion's current emotional state. The user can select any of the presented emotion for the emotional state of the human companion using the radio-buttons below.

The following four column is reserved for the robot's affective state. In the implemented emotional model, the second column, the inner state of the robot is controlled by the battery level, and the battery is programmed to die out in 10 run, i.e. the inner state of the robot degrades from "happy" to "sad". The third and forth columns are dedicated for displaying the current emotion and the mood of the robot, respectively.

The last two columns in the right are reserved for the next emotional state of the robot. The first column is dedicated for the users to select the emotion that they find the most appropriate for the scenario represented by the parameters displayed on the previous columns. And then by clicking on the display system output button, the user can see what is the next emotional state of the robot proposed by the emotional model and evaluate this output using the radio-buttons below. For example, the user running the simulation displayed in Figure 4.2, expects that the robot imitates instantaneously the emotion of its human companion because he/she selected “very angry" for the human and “angry" for the robot. But according to CASPAR model, although its human companion displays a high level of anger towards the robot, the robot does not mirror directly this emotion because CASPAR takes into consideration the “happy" inner state and the current “happy" state of the robot with a positive mood. Therefore the proposed next emotional state is a neutral emotion, instead of a negative one.

4.1.2.2 GUI simulations

The preliminary study was conducted to investigate the reactions and impressions of the user about the proposed emotional model. The study was performed in 33 sessions of GUI interaction. The participants were selected from volunteer undergraduate students without any prior experience with robots. The participants freely interacted with the emotional model and evaluated the system output by using the 5-point semantic Likert scale, with 1 meaning “very bad" and 5 meaning “very good".

The preliminary study results showed that the emotional model was mostly accepted and approved by the test participants because they scored the model’s output with high marks. The average score for the system output was $M = 3.61$, and the range of the scores changed from a minimum of 2.33 to a maximum of 4.75, with a standard deviation of $SD = 0.71$.

The results are promising for incorporating an internal state mechanism for the behavioral model of the affective robot. However, the evaluation of the proposed model is based solely on simulation studies without any external input from the real-world interaction environment.

4.2 A Study in "Disgust"

In order to implement a behavioral model incorporating the external input from the human companion, a pilot study was conducted on the EMBOA video recordings. In the presented framework, the emotion detection was performed by an open source toolkit, OpenFace [104] based on facial action units.

4.2.1 Implementation of a simple emotional model based on video recordings

The annotation of the video recordings and the detection of children's emotional states are based on the visual annotation model presented in Section 3.4.1. The workflow of the model can be summarized as follows:

- The recordings are fed to OpenFace, OpenFace provides the information about the presence of 18 AUs in a binary data format for each frame in the video,
- The binary data provided by OpenFace is transformed to emotion labels based on the FACS guidelines [32]. If the frame is not labeled with a basic emotion, then it is labeled as "Neutral".
- The emotion labels are used for the discretization of children's emotions based on the 2D valence-arousal representation schematized by Toisoul et al. [108] and displayed in Figure 4.3. The discretization is done as follows:
  - Neutral emotions are placed at the origin: (0,0).
  - Emotions with a "Low" intensity are placed at 1 or -1 based on their valence and arousal components.
  - Emotions with a "Medium" intensity are placed at 3 or -3 based on their valence and arousal components.
  - Emotions with a "High" intensity are placed at5 or -5 based on their valence and arousal components.

  For example, the "Happy" emotion with a high valence but low arousal component is placed at (5,1) whereas "Afraid" having a low valence but high arousal component is placed at (-1,5).
- The discretized labels are used to represent changes in children's emotions along the valence and arousal axis, respectively. Once the child's emotion model is set, then the robot's emotional model can be adapted to the changes in the child's mood.

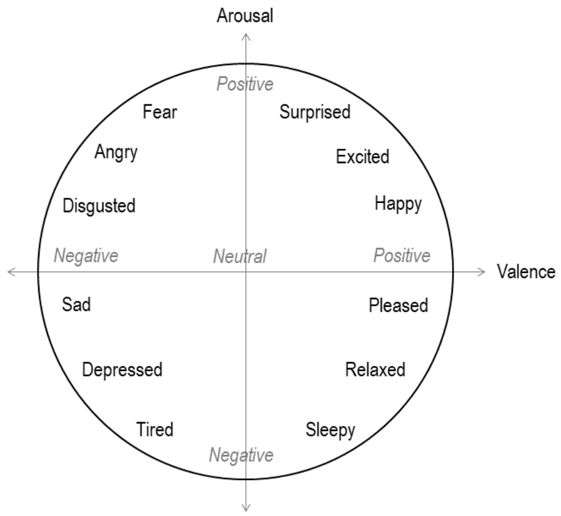


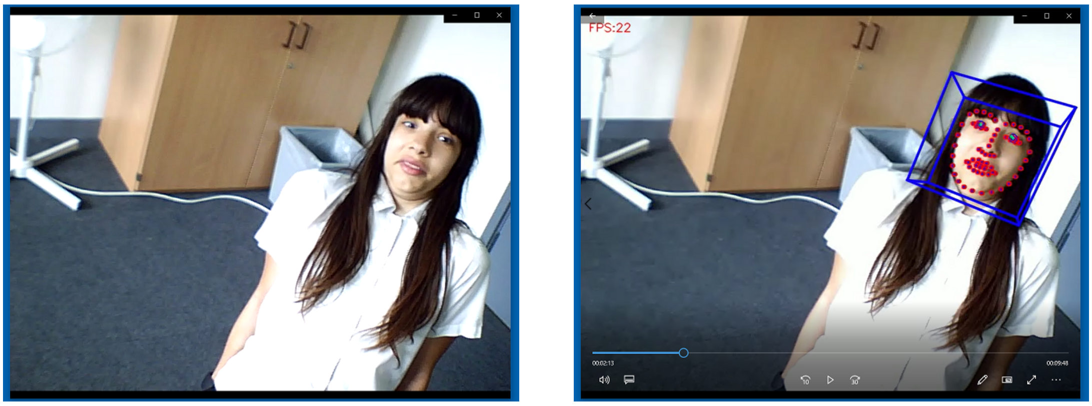
Figure 4.3 : 2D Valence-Arousal model used for the discretization of emotions.

The workflow and the proposed emotion model for the robot is conceptualized with an example and possible improvements/adaptations for the model are discussed following a pilot study with the recording of a child from United Kingdom.

As previously mentioned in Section 3.4.2, OpenFace [104] can be used for detecting head pose and orientation but due to the different resolution values of the video recordings, the detection performance were not reliable in recordings therefore these videos were not included in the pilot study. In addition, during the interaction studies in Turkey the children were advised to wear facial masks, therefore the detection accuracy of OpenFace suffered from the partial occlusion of children's faces. Consequently, the video recordings captured during the interaction studies in United Kingdom were used for the pilot study.

First, the recordings of the 7 children who participated in the interaction study were examined and analyzed, then the video of a child who displayed total reluctance towards Kaspar was selected to be an example for the emotional modeling study.

The child, coded as UH-C04, is an 11-year-old high-functioning girl who is very expressive, both vocally and physically. The moment she enters the room allocated for the study, she displays reluctant behaviors towards Kaspar and she does not want to sit in front of it, as displayed in Figure 4.4.



(a) A screen shot from the interaction. (b) OpenFace detection of the screen shot.

Figure 4.4 : UH-C04's expressions from her interaction session with the robot: "Kaspar is too creepy".

The selected recording belongs to the first session of UH-C04 and the interaction session is 12 minutes long. The recording has 14569 frames in total. When each frame was labeled and analyzed, the results revealed that UH-C04 was not comfortable during the interaction and she felt mostly disgusted (726 frames) and she was not happy to be there, because the second emotion displayed by her was sadness (444 frames), as seen in Figure 4.5. The neutral emotions having the highest number of occurrence (12831 frames, 88% of interaction) were excluded from the figure.

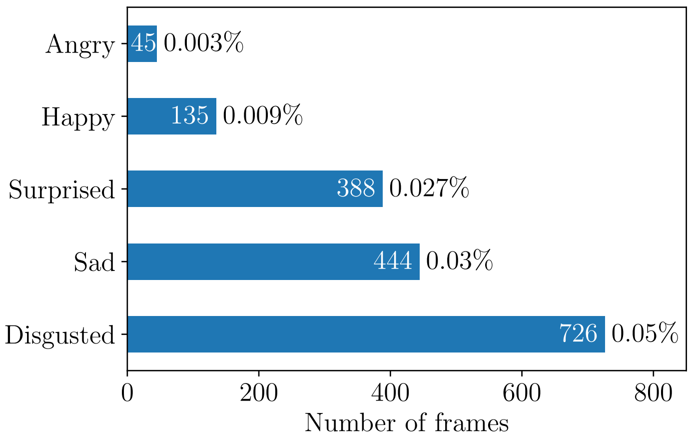


Figure 4.5 : UH-C04's emotions detected during her 12-minute (14569-frame) long interaction with Kaspar.

After the discretization of the emotion labels, UH-C04's emotions were visualized and color-coded separately in valence and arousal axes, as shown in Figure 4.6. The frequency of the emotions with negative valence combined with high-level of arousal showcase the negative state of the child during the interaction.

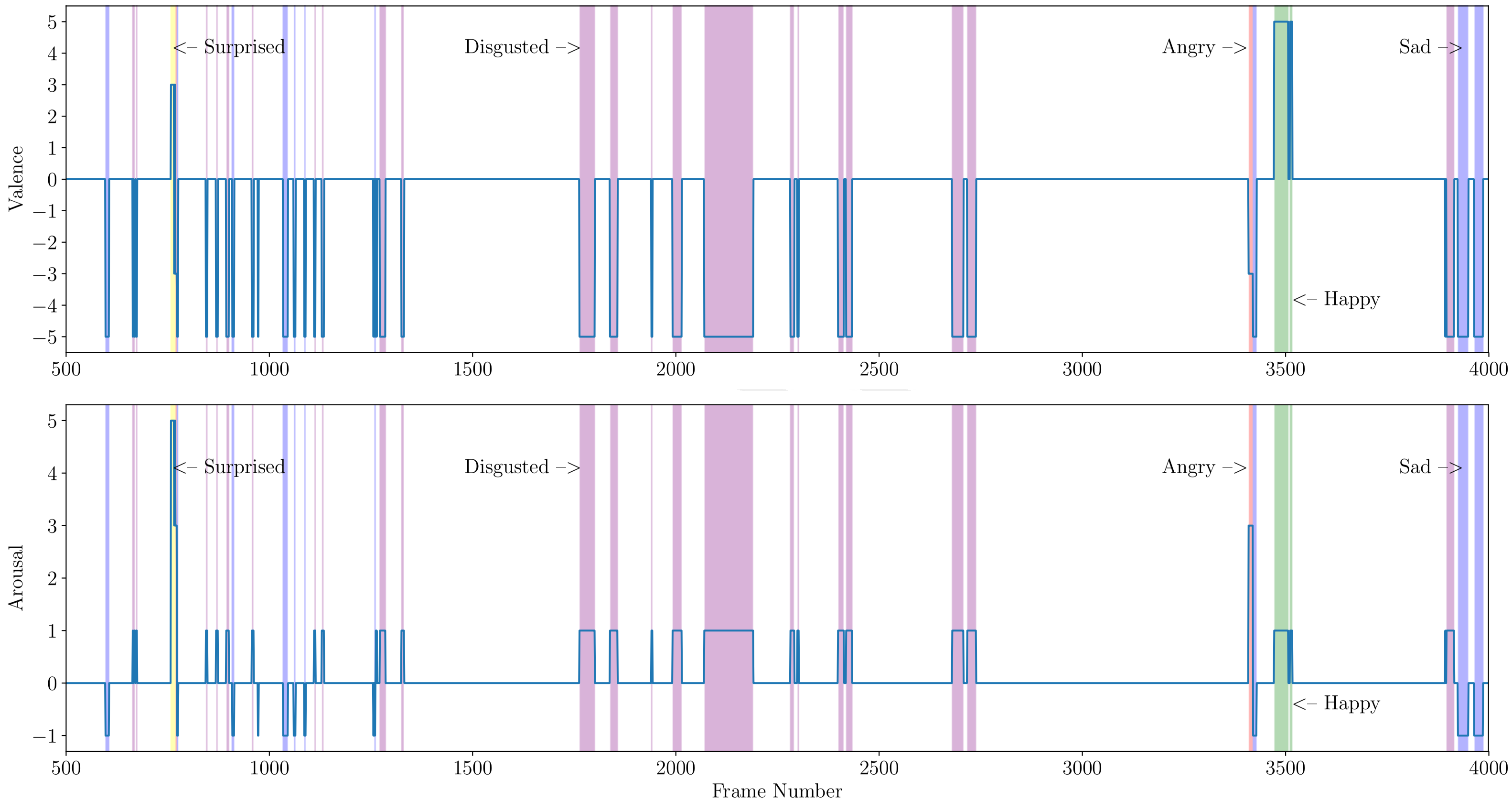


Figure 4.6 : The discretized and color-coded changes in valence and arousal axes during the first 4000 frames of interaction.

4.2.2 The robot's role and persona in the interaction

Once the companion's emotional state has been determined, this information can be transferred to the robot and used for the adaptation of robot's emotional behavior. The results of the pilot study with UH-C04 revealed that the emotional state of the child was negative when it was not neutral. Therefore to conceptualize the adaptation process in the behavioral model, the robot's main role was selected to be the "neutralizer" in this interaction. Consequently, different strategies can be modeled and implemented for the robot to help the child to neutralize her emotions:

1. Neutralize the other's emotions displaying opposing emotions:

   The robot is set to display the opposite emotions in the 2D valence-arousal model, i.e. the symmetric emotions with respect to origin. For example, if the child is labeled as "sad", then the robot will display "happy" emotions. The visual representation of the the emotional states of the child and robot is displayed in Figure 4.7.

2. Neutralize the other's emotions with opposing emotions if they have a negative valence component::

   The robot is set to display the same emotions as its companion as long as they don't have any negative valence component. Because the negative valence component is important in negative emotions, only these emotions are regulated by opposite emotions. For example, when the child is angry, the robot will display the emotion symmetrical with respect to the arousal axis, that is, surprised. The visual representation for this strategy is displayed in Figure 4.8.

3. Neutralize the other's emotions displaying the same emotions in moderate levels:

   The robot will display the same emotional state as its companion, but in moderation. If the valence and arousal components of the companion's emotions are above a specified threshold (selected as 3 in positive, -1 in negative), then the robot will reduce the intensity of its emotions, as seen in Figure 4.9.

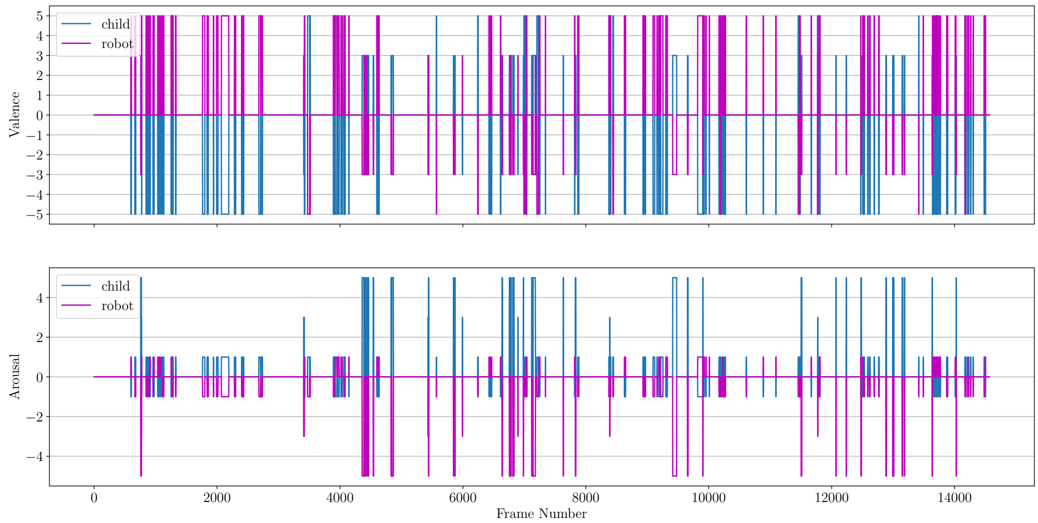


Figure 4.7 : Neutralize the child's emotions with the opposite ones.

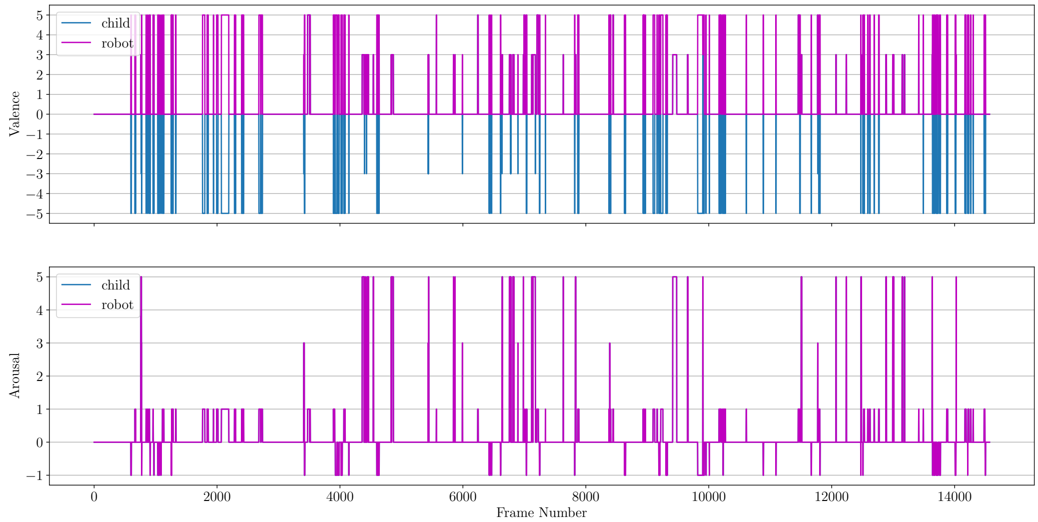


Figure 4.8 : Neutralize only the emotions with negative valence component.

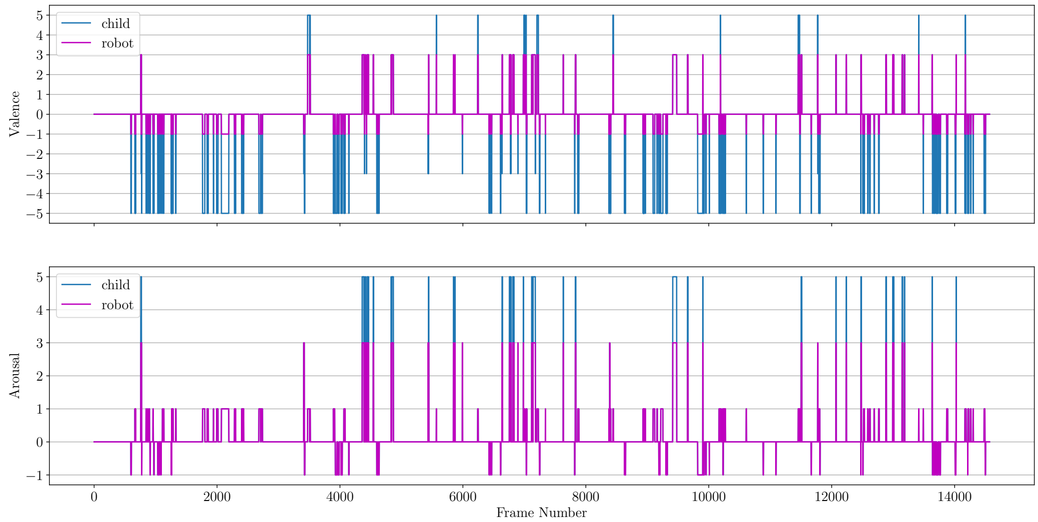


Figure 4.9 : Neutralize emotions with moderate levels.

The presented study is limited to an offline interaction scheme, instead of a live camera feed from the real world, a video recording is processed for emotional modeling. Therefore, it looks computationally cost-effective, fast, and convenient, but the implementation doesn't reflect the challenges faced when working in-the-wild. In addition, changes in the emotional states of the interaction dyad are reflected in the model simultaneously, so it is not a good representation for the emotional display in the real world.

To refine the emotional model and make it more natural and real, the discrete square signals for valence and arousal can be filtered and interpolated to produce a smoother signal in the time domain. Additionally, a delay factor can be integrated into the robot's emotional state model to increase the reliability of the displayed emotions and simulate natural changes in the human emotional state.

The proposed strategies are presented to highlight the robot's role in the interaction and the need to adapt its role in the interaction by implementing a set of different approaches. The presented study is open to improvements and requires adaptations to many interaction scenarios. The robot's main objective, set as "the neutralizer" in the presented example, may be inferred from the interaction context and different strategies may be implemented and selected for different task in order to improve interaction gain and personification of the robot.

4.3 Behavioral Analysis Study for Social Robots

In the previous section, the affective robot was given a specific role based on the emotional state of its human companion and a set of behavioral strategies related to this assigned role were presented. However, the robot's role and persona may change depending on the needs or preferences of its companion, as in the nature of interaction.

A survey study was conducted with a large group of participants in order to investigate the different roles and expectations that humans impose on robots in real-life scenarios. The survey also aims to explore the preferences of human partners about the robot's contribution to the interaction and the emotional intensity of the robot's affective responses.

4.3.1 A questionnaire for affective social robots

The online survey study was conducted based on a special questionnaire tailored to investigate people's preferences and attitudes towards robots. The questionnaire consists of 3 sections. The first section is composed of question about the demographic information of the test participants: Gender, age, education, occupation and prior experience with robots.

The second section consists of the questions on the third subscale of the NARS (Negative Attitude towards Robots Scale [63]) questionnaire on negative attitudes towards emotions in interaction with robots. These questions are accompanied by a subset of questions on interaction with robots and the social impact of robots from the other two subscales of NARS. Additionally, there is a standalone question about interacting with emotional robots. The questions in the second sections are listed as follows, and the corresponding subscale and question number in the NARS questionnaire are given in parentheses:

– NARS (S2:1) - *I would feel uneasy if robots really had emotions.*

– NARS (S3:3) - *I would feel relaxed talking with robots.*

– NARS (S3:5) - *If robots had emotions, I would be able to make friends with them.*

– NARS (S3:6) - *I feel comforted being with robots that have emotions.*

– NARS (S1:12) - *I would feel paranoid talking with a robot.*

– Standalone - *I feel that our emotions and affective states have an impact on our interactions with the robots.*

Participants were asked to answer these questions using a 5-point semantic Likert scale ranging from “I totally disagree” to “I totally agree”.

The third section of the questionnaire consists of 14 narratives, in which the human companion's emotions are conveyed by simple and illustrated narratives. The narratives are inspired by common real-life situations with a negative, neutral or positive context. The narratives, their keys, and the emotions for the human companion and the robot's responses are shown in Figure 4.10.

Keyword	Human	Narrative	Robot	Intensity	Options
Influenza	(-)	Etraftaki salgın hastalık furyasından ben de nasibimi aldım. Geçen hafta tam 3 gün 38 ateşle boğuştum. En sonunda ateşim düştü de azıcık kendime geldim. Çocukluğumdan beri böyle hasta olmamıştım.	(-)	-1	İstatistiklere göre kış mevsiminde salgın hastalıklarda artış görülür.
				-3	Geçmiş olsun.
				-5	Hastalanmana çok üzüldüm, çok geçmiş olsun.
Computer Problem	(-)	Ay geçen gün ne oldu, biliyor musun? Haftalardır üzerine çalıştığım bir proje vardı, projeyi bitirdim. Tam sisteme yükleyeceğim, bir anda bilgisayarım kapanıverdi... Ve bir daha açılmadı... İnanabiliyor musun? O kadar günün emeği resmen uçtu gitti...	(-)	-1	Anladım.
				-3	Üzüldüm, hiç iyi olmamış.
				-5	Gerçekten mi? Çok fena! Çok üzüldüm...
Broken Cup	(-)	Bulaşıkları yıkarken bir anda ne oldu anlamadım, en sevdiğim kupam elimden kaydı, yere düştü. Resmen paramparça oldu, çok üzüldüm.	(+)	1	Benzer bir şey daha önce benim de başıma gelmişti.
				3	Önemli değil, cana geleceğine mala gelsin.
				5	Olsun, iyi tarafından bak, şimdi o kupanın çok daha güzelini alabiliriz.
Movie	(-)	Geçen gün vizyona giren film var ya? En iyi arkadaşımla ona beraber gideriz diye sözleşmiştik. Sonra ne oldu biliyor musun? Filme başka biriyle gitmiş, bana haber vermek aklına bile gelmemiş...	(+)	1	Hımm, anladım.
				3	Hımm, anladım. Dert etme, olur böyle şeyler.
				5	Hımm, anladım. Üzülme, arkadaşlar arasında olur böyle şeyler, sen de başka bir arkadaşınla gidersin.
Hospital	(-)	Dün çok yakın bir arkadaşımın trafik kazası geçirdiğini öğrendim. Hastaneye kaldırmışlar, bugün çok ciddi bir ameliyat geçirmesi gerekiyormuş. Durumu kritikmiş.	(-) neutral (+)	-5	Gerçekten mi? Çok üzücü.
				0	Gerçekten mi?
				5	Gerçekten mi? Eminim kısa sürede iyileşecektir.
Birthday	(-)	Bu sene doğumgünüm bayrama denk geldiği için bir sürü arkadaşım doğumgünümü kutlamayı unuttu, inanabiliyor musun?	(-) neutral (+)	-5	İnanamıyorum, nasıl unuturlar!
				0	Olabilir, insanlık hali...
				5	Olsun, iyi tarafından bak, kendilerini affettirmek için seninle daha sık görüşürler.
Traffic	neutral	Benim evimle işim arası aslında çok yakın, aradaki mesafe 5 km ya var ya yok. Sabahları trafik olmadığı zaman 15 dakikada işyerime ulaşıyorum ama trafiğe denk gelirsem işe gitmem bazen 1 saati geçiyor.	(-) neutral (+)	-5	5 km'yi 1 saatte mi gidiyorsun? Ne kadar korkunç!
				0	Büyükşehirlerde genelde böyle oluyor.
				5	Evinle işyerin ne kadar yakınmış! Ne güzel!
Lentil Soup	neutral	Bugün yemekte mercimek çorbası vardı.	(-) neutral (+)	-5	Mercimek çorbasını pek sevmem, ayrıca yapması da hiç kolay değil.
				0	Peki, tamam.
				5	Mercimek çorbası çok güzeldir, lezzetli yemeklerden biri.
Abroad	(+)	Bir arkadaşım üniversiteyi bitirince yurtdışına gidip orada işe girdi. Uzun zaman sonra ilk defa Türkiye'ye gelmiş. Geçen gün aradı, görüştük. O kadar özlemişim ki onu, görüştüğümüz için inanılmaz mutlu oldum.	(+)	1	Peki, tamam.
				3	İyi olmuş
				5	Süper! Senin adına çok sevindim
Computer Problem Solved	(+)	Hani geçen gün sana bilgisayarım çöktü, üzerinde çalıştığım proje uçtu gitti demiştim ya? Hatırlıyor musun? Aslında öyle değilmiş, gereksiz yere panik olmuşum. Neyse ki projenin yedeğini almışım, projeyi vaktinde teslim ettim, herhangi bir sorun olmadı. Hatta projeden en yüksek notu ben aldım.	(+)	1	Gerçekten mi!
				3	Gerçekten mi! İyi bir haber.
				5	Süper! Çok güzel bir haber!
Shopping	(+)	Uzun zamandır ilk defa alışverişe çıkıp, kendime güzel bir kazak aldım. Biraz pahalıydı ama çok beğendim, dayanamadım aldım. İyi ki almışım, çok da rahatmış	(-)	-1	Belki ihtiyacın yoktu ama güle güle kullan.
				-3	Çok pahalıysa almasaydın daha iyi olurdu.
				-5	Yeni kazak almana ne gerek vardı? Dolabın kazak dolu.
Forgotten Phone	(+)	Geçen gün hava çok güzeldi, bir arkadaşımla Boğaz'a inip yürüyüş yapalım dedik. Azıcık yürüdük, sonra da deniz kenarında güzel bir bankta oturduk. Sonra ben kalktım, o oturmaya devam etti. Eve döndükten sonra bir baktım telefonum yok! Aniden kapı çaldı, açtım, arkadaşım telefonumu getirmiş, o kadar sevindim ki...	(-)	-1	Benzer bir şey daha önce benim de başıma gelmişti
				-3	Telefonunu nasıl unuttun ya?
				-5	Sana inanamıyorum telefonunu nasıl unutursun? Ya biri alıp gitseydi?
New Phone	(+)	Geçen gün teknoloji marketlerinden birinin cep telefonlarında indirim yaptığını gördüm. Benimki de çok eskimişti, şarjı 1 gün bile zor gidiyordu. İndirimi görünce hemen girdim, kendime yeni telefon aldım. Hem çok beğendim, hem de çok uygun fiyata geldi.	(-) neutral (+)	-5	Güle güle kullan ama gerçekten ihtiyacın var mıydı...
				0	Güle güle kullan.
				5	Süper! Güle güle kullan, çok iyi yapmışsın.
Second Place	(+)	Bir kaç arkadaş insan-robot etkileşimi üzerine bir proje yarışmasına katılmıştık. En sonunda jüri sonuçları açıklamış. Bizim projemiz 354 proje arasında ikinci seçilmiş!	(-) neutral (+)	-5	Güzel bir haber ama birincilik kaçmış, daha fazla çalışmanız lazım.
				0	Güzel bir haber, tebrikler
				5	Güzel bir haber, tebrikler, süpersiniz!

Figure 4.10 : Human-robot emotional state pairs for the narratives (in Turkish).

Participants were asked to select the most appropriate reaction for the robot among 3 choices with different emotional intensities: (low, medium, high) or (negative, neutral, positive). For each narrative, the robot has negative, neutral or positive reactions with varying intensity values. The emotional states of the human-robot dyad were explored in pairwise. The human companion - robot emotion pairs and corresponding keys for the narratives are listed below and one sample illustration is displayed in Figure 4.11:

- Negative-Negative Narratives: Influenza, computer problem
- Negative-Positive Narratives: Broken cup, movie
- Positive-Positive Narratives: Friend abroad, computer problem solved
- Positive-Negative Narratives: Shopping, forgotten phone
- Negative-(Negative, Neutral, Positive) Narratives: Hospital, birthday
- Neutral-(Negative, Neutral, Positive) Narratives: Traffic, lentil soup
- Positive-(Negative, Neutral, Positive) Narratives: New phone, second place

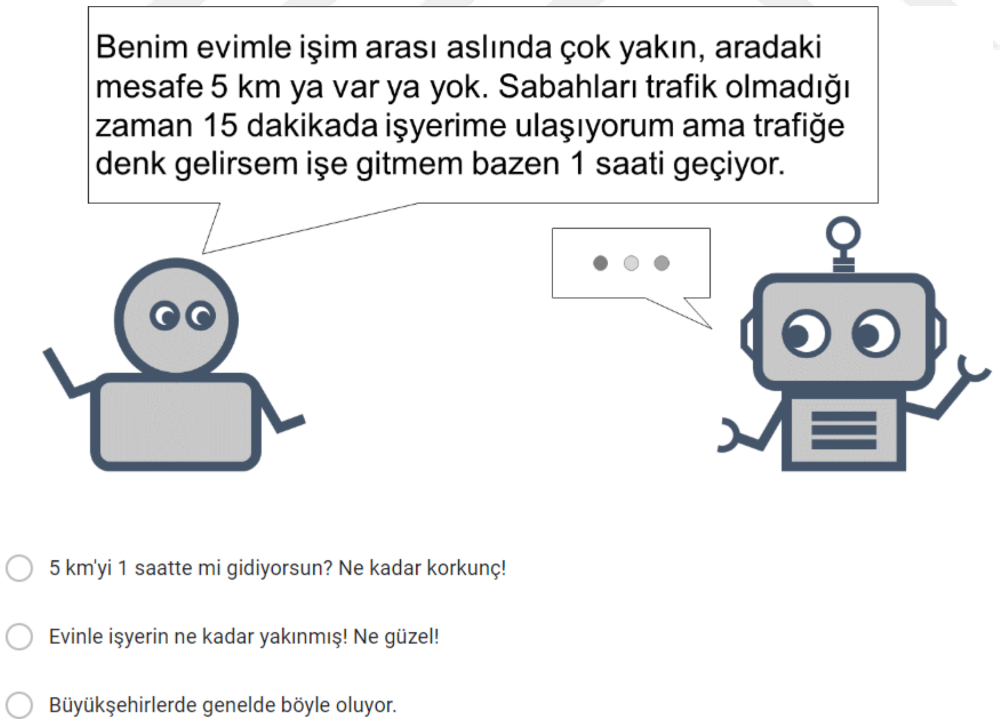
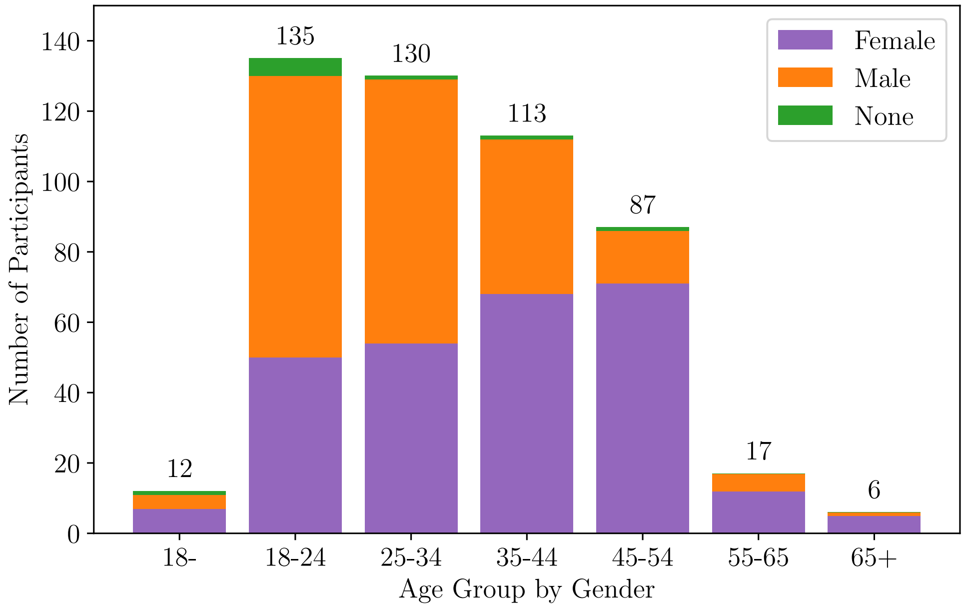


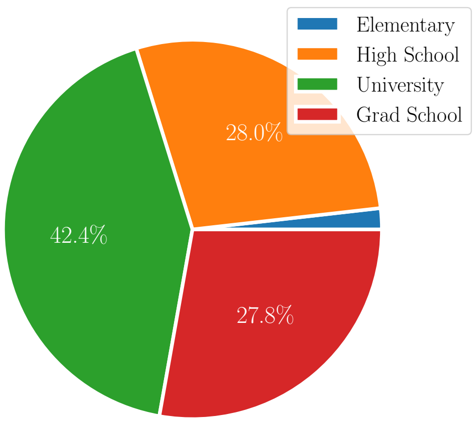
Figure 4.11 : A neutral narrative from the human perspective and the robot's negative, neutral, positive reactions to the neutral narrative "traffic".

4.3.2 Profile of survey participants

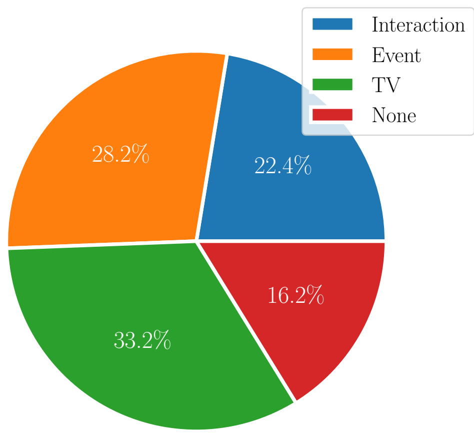
The questionnaire study was conducted with a total number of 500 participants. 53.4% of the participants (267) were identified themselves as female, 44.8% (224) as male, whereas 1.8% (9) of them would not share their gender. The majority of participants are from the age group 18-24 (27%, 135), 25-34 (26%, 130) and 35-44 (22.6%, 113), as displayed in Figure 4.12a and have at least a bachelor degree (42.4%, 212) or higher education (27.8%, 139), as shown in Figure 4.12b.



(a) Age distribution of participants color-coded by gender.



(b) Education level.



(c) Prior experience.

Figure 4.12 : Distribution of participants by gender, age groups, education level and prior experience with robots.

A large number of participants had no previous experience with robots, 33.2% (116) reported seeing robots on television or seeing them at a conference, technical fair or other organization (28.2%, 141). While 16.2% (81) of participants reported that they had never interacted with a robot before, 22.4% (112) reported that they had the opportunity to interact with a robot in real life, as shown in Figure 4.12c.

The occupation of the participants are mostly based on computer science and engineering related fields (25.4%, 127) or academy (professors as well as undergraduate students, 33.8%, 169), but there is a wide variety of participants from different professions (40.8%, 204) such as pilates instructor, actor, chef, dentist, lawyer, kindergarten teacher, etc.

4.3.3 Descriptive data analysis and hypothesis testing

The questionnaire study was conducted via an online survey, Google Forms, encouraging anonymous and volunteer participation. The responses were collected in a tabular data format. The raw data collected through the online survey tool consisted of categorical data, hence the need for the data preparation and transformation step.

4.3.3.1 Data preprocessing

The first and second sections of the questionnaire were composed of questions about the demographic information of the participants and their impressions and attitudes towards robots, the NARS questions. While the demographic information of the participants was processed as nominal features in the first section, the answers given to the NARS questions in the second section were coded as ordinal features since they were based on a 5-point semantic Likert scale.

NARS scores were computed based on the reverse item score, meaning that participants with a positive attitude towards robots got higher scores than participants with a negative attitude. The highest score for each question was 5, and the lowest score was 1, since there were 6 questions in the questionnaire, the NARS score range of the participants varied from 6 to 30, with mean $M = 19.52$ and standard deviation $SD = 4.77$.

The third section of the questionnaire consisted of 14 multiple choice questions, and the participants were asked to select the option that suited the best for the robot's reaction to the given narrative from 3 possible answers. The options selected by the participants were mapped to the discretized intensity levels displayed previously in Figure 4.10, and they were quantized between a range of -5 to 5.

The categorical variables, demographic information and NARS questions, were label encoded and also quantized for the study. The resultant data set was composed of tabular data with 500 observations and 25 features: 5 demographic, 6 NARS, 14 narratives.

4.3.3.2 Hypothesis testing

First, the reliability of NARS scores was confirmed with Cronbach's α 79.87%. Then one-way ANOVA tests were performed on NARS scores to explore its effect on the other features in the data set, demographic information and narratives. The results showed that NARS scores had a significant effect over all the other features with $p-value < 0.05$, except for one scenario ("shopping").

The Tukey post-hoc tests were performed on demographic data to compare the means of each group in a pairwise investigation. The results ($p-value < 0.01$) showed that:

- Female participants' were more positive in their attitudes towards robots than male participants.
- Participants from older generations were more positive towards robots than younger generations.
- Engineers tended to have a more positive attitude towards robots than people from other groups of profession.
- Participants with no prior experience with robots were more positive than people with different levels of experience.

Tukey multiple comparison of means tests also revealed that the mean difference of NARS scores had a significant effect on the intensity of the selected robotic behaviors in the third section of the questionnaire.

4.3.3.3 Clustering based on participants' attitudes towards robots

Inspired by the significant effect of NARS scores on the other features, a clustering study was conducted to investigate the different numbers of clusters and their characteristics. K-means clustering were used and the participants were distributed into 2, 3 and 4 clusters based on their NARS scores, their attitudes towards robots.

The group characteristics for each cluster were analyzed and each cluster were labeled based on its corresponding NARS score distribution. For example, in the case with 2 clusters, the clusters were named as “negative" and “positive" representing the attitudes of participants in the relevant cluster, whereas in the case with 4 clusters, the cluster means for the 2 “neutral " groups were much closer, therefore they were named as “neutral-" and “neutral+", as depicted in Figure 4.13.

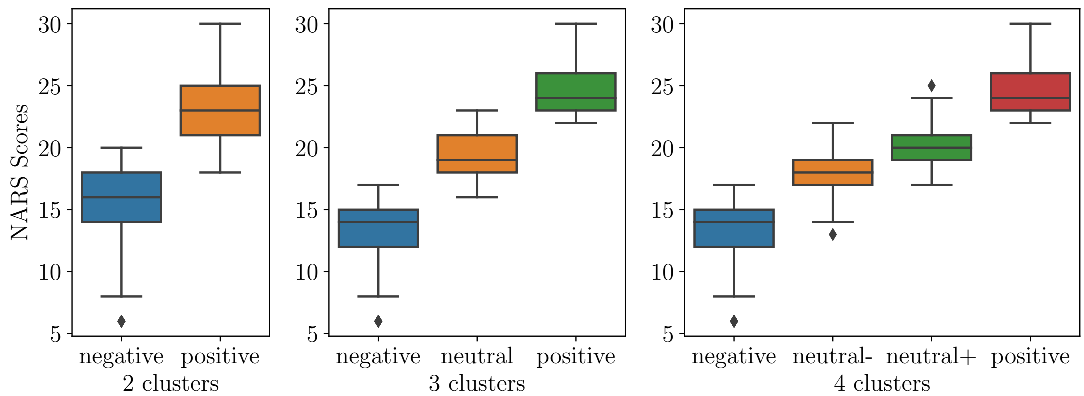


Figure 4.13 : Distribution of 2, 3 and 4 clusters based on participants' attitudes

Although the number of clusters providing the highest silhouette score were reported as 4 (with a score of 0.424), the optimal cluster number were selected as 3, since the clusters are similar and it is easier to distinguish 3 clusters and interpret the participants behavior based on their attitudes. The number of participants in each cluster, the cluster mean, standard deviation, minimum and maximum NARS scores are displayed in Table 4.1.

Table 4.1 : Descriptive statistics for the 3 attitude clusters based on NARS scores.

Cluster	Count	Mean	Std	Min	Max
Negative	126	13.33	2.48	6	17
Neutral	220	19.33	1.73	16	23
Positive	154	24.87	2.13	22	30

The attitude clusters and the choices of participants were analyzed and explored in detail for each narrative on the questionnaire. The results showed that although participants tended to differ in their attitudes towards robots, they agreed on emotional behavior and intensity based on the context of the narratives. The most agreed narratives and the distribution of participants preferences as well as their attitudes are displayed in Figure 4.14.

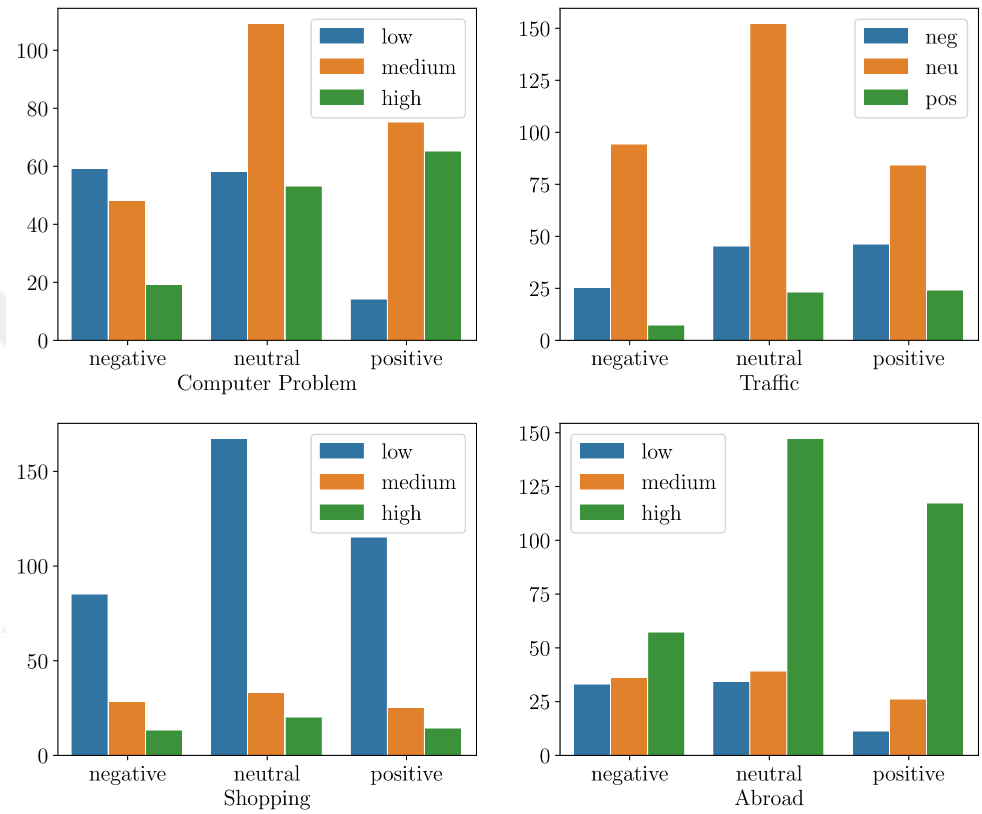


Figure 4.14 : Distribution of responses in selected narratives - computer problem, traffic, shopping, abroad.

The experimental results showed that when the human-robot emotion pair was:

- Negative-Negative: Participants tended to select the options with lower intensities, i.e. “low" or “medium", as in the computer-problem narrative. The only exception for this case was the preference of the participants from the positive attitude cluster, they had no problem selecting the most intensely negative response for the robot, they showed that they were okay with a robot displaying intense negative behaviors.

- Neutral-(Negative, Neutral, Positive): Participants tended to choose the neutral option when given a neutral narrative, regardless of their attitude towards robots, as in the traffic scenario example.

- Positive-Negative: Participants tended to select the option with the neutral intensity when faced with a positive context but a series of negative responses, as in the shopping narrative. The shopping narrative were the most agreed narrative among the participants, regardless of their attitudes.

- Positive-Positive: Participants had the tendency to select the option with the higher intensity when they were presented with a positive context of interaction and a set of positive responses with varying intensity levels, as in the abroad scenario. It is important to note that although the participants had negative attitudes, when the context was clearly positive, the majority showed that they would prefer a high-intensity positive response.

The other narratives on the questionnaire and the distribution of participants preferences based on their attitudes towards robots are shown in Figure 4.15.

4.3.4 Prediction for relevant behavior selection based on narratives

The results are consistent with the assumption that although NARS scores have a significant impact on participants' choices, they mostly agree on their preferences in given narratives based on the context of interaction. Based on the reported findings, classifiers were trained to predict the selected reaction for the robot based on the choices of participants.

A multi-class classification study were carried out, alternating between 3 intensity categories (low, moderate, high) and 3 mood categories (negative, neutral, positive). Based on the small-size of the data set and the moderate processing power of the physical robot, lightweight and standard classifiers were selected for this classification task. The selected set of classifiers and they specifications are listed as follows:

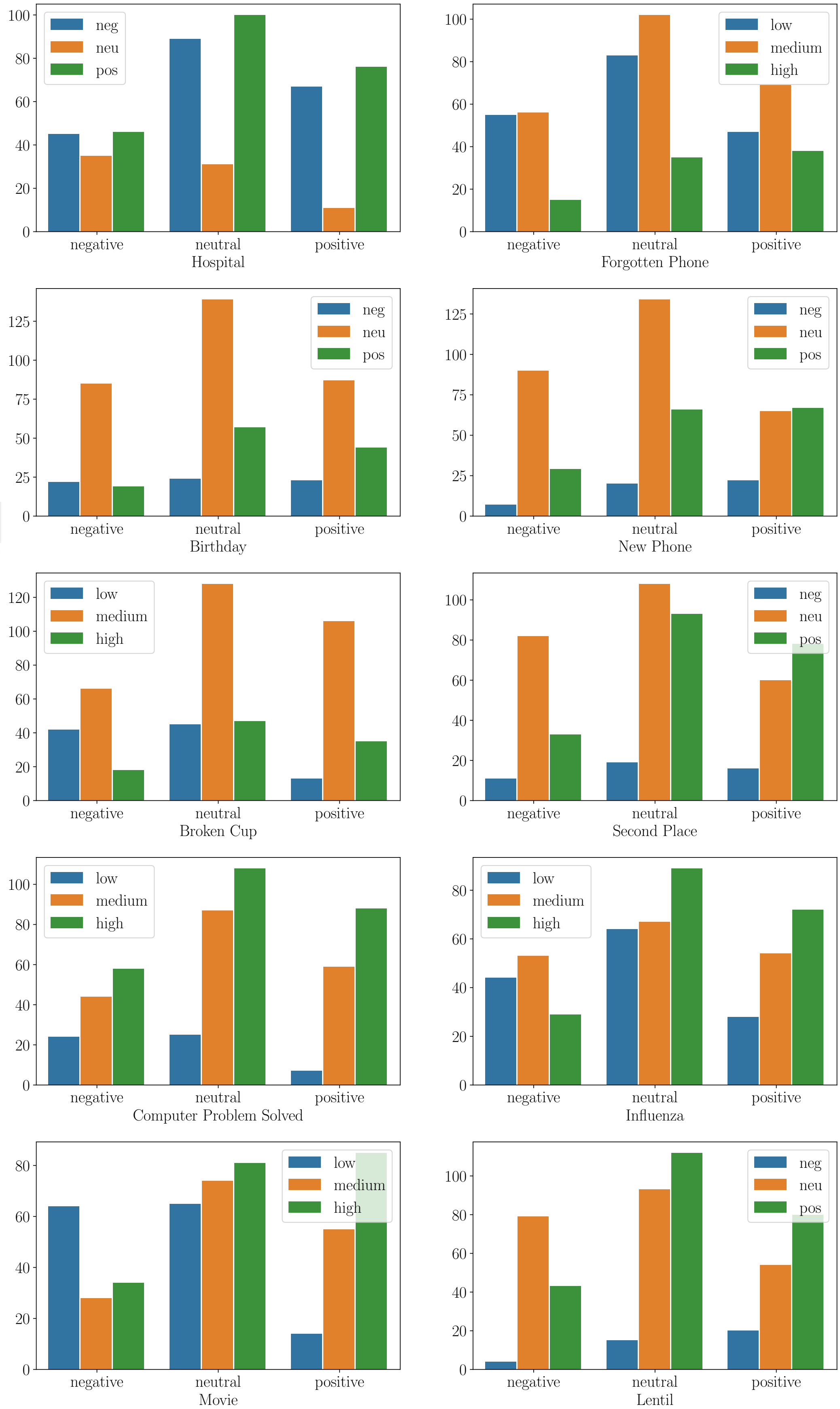


Figure 4.15 : Distribution of responses in other narratives based on the attitudes.

- K-Nearest Neighbor (KNN) - A KNN classifier with a number of neighbors set as 5 as the decision criteria, and uniform weight distribution was selected, the distance between the neighbors were computed based on Euclidean distance.
- Gaussian Naive Bayes (GNB) - A GNB classifier with no prior probabilities for classes was chosen since they did not have any dependency relation among them.
- Support Vector Machine (SVM) - A simple SVM classifier with default LIBSVM implementation and corresponding parameters was selected with an RBF kernel.
- Decision Tree (DTC) - A shallow decision tree with Gini impurity was selected based on the dimensions of the data set.
- Random Forest (RFC) - Similar to the decision tree, a shallow forest with 3 estimators and maximum depth set as 2 was selected for the classification task.
- Multilayer Perceptron (MLP) - A network with 3 hidden layers with 8 neurons on each, Relu activation with Adam optimizer, and maximum epoch number set as 500 was selected for the MLP classifier.

The classifiers were trained to predict participants' responses to a particular narrative. Therefore, they were trained with the entire data set except the responses to the particular narrative and tested with the responses to that narrative. In brief, the classification task consists of taking the responses given to one specific narrative out of the feature set and try to predict the preferred reaction for this narrative.

The classifiers were tested with the same narratives but trained on different set of features. They were trained with: (1) a set of features solely composed of the responses given to the narratives, (2) a set of features composed of the responses to the narratives and the NARS scores of the participants, (3) a set composed of the responses to the narratives, NARS scores, label encoded and discretized demographic information of the participants, and lastly, (4) the same set as the previous option but all the features were categorized and label encoded.

The performance of the classifiers trained on different set of features were analyzed and evaluated comparatively by their average accuracy value for each classification task. Even though the performance of the classifiers trained on different sets do not differ significantly based on their average accuracy, the highest value for the average accuracy was obtained with the classifiers trained with only the responses to the narratives, as seen in Table 4.2, where the different set of features and the average accuracy of the 6 classifiers trained and tested for each narratives are displayed.

Table 4.2 : Average accuracy values for the classifiers based on train set.

Train set	Average Accuracy
(1) Narratives	0.5807
(2) Narratives & NARS	0.5741
(3) Narratives & NARS & Demographic data (enc)	0.5747
(4) Narratives & NARS & Demographic data (cat)	0.5414

Since the highest average accuracy for the classifiers was obtained when they were trained solely on the responses to the narratives, the classification results for each narrative based on the performance of each classifier were reported in detail for that particular case in Table 4.3. The f1-score for the classifiers were reported as well as their accuracy to explore the narrative with responses having and imbalanced class distribution.

The results showed that the best classification accuracy values were obtained with the narratives abroad (78%, RFC), shopping (76%, RFC), and computer problem (70%, GNB). As mentioned in the previous Section 4.3.3.3, these particular narratives were the narratives most agreed upon by the participants, therefore the classifiers performed in predicting the preferred responses the best.

On the other hand, the worst accuracy values for the narratives were obtained with birthday (42%, GNB), forgotten phone (43%, RFC), computer problem solved (47%, RFC), and movie (47%, DTC) scenarios. The classifier performed the worst in birthday narrative having a negative-positive human-robot emotion pair, even though the preference of the participants mostly agreed on the response with the neutral mood, the classifier could not predict the corresponding response. However, in the case of forgotten phone narrative, the response of the participants differed significantly for each intensity level and each attitude clusters. Therefore, the classifiers were performed the worst in this prediction task.

The results based on the overall performance metrics show that the selected set of classifiers performed accurately on the prediction task for intensity and mood categories in the multi-class classification task, regardless of the attitudes of participants towards robots, as shown on Table 4.3.

Table 4.3 : Prediction accuracy values and f1-scores for each narratives.

Narrative	Metrics	KNN	GNB	SVM	DTC	RFC	MLP
comp-problem	acc	0.60	0.70	0.66	0.56	0.60	0.49
	f1-score	0.60	0.70	0.66	0.53	0.50	0.59
traffic	acc	0.60	0.56	0.66	0.51	0.66	0.63
	f1-score	0.56	0.57	0.52	0.55	0.52	0.59
hospital	acc	0.51	0.55	0.48	0.54	0.51	0.56
	f1-score	0.51	0.54	0.47	0.53	0.47	0.44
forgotten-phone	acc	0.54	0.47	0.44	0.47	0.43	0.43
	f1-score	0.52	0.47	0.38	0.46	0.39	0.46
birthday	acc	0.55	0.42	0.62	0.49	0.62	0.54
	f1-score	0.49	0.43	0.47	0.49	0.47	0.53
new-phone	acc	0.60	0.61	0.66	0.57	0.56	0.62
	f1-score	0.57	0.59	0.61	0.53	0.51	0.53
broken-cup	acc	0.62	0.61	0.61	0.48	0.54	0.62
	f1-score	0.55	0.57	0.51	0.52	0.47	0.55
shopping	acc	0.66	0.67	0.75	0.57	0.76	0.74
	f1-score	0.61	0.64	0.64	0.61	0.66	0.64
second-place	acc	0.50	0.56	0.50	0.60	0.56	0.49
	f1-score	0.48	0.55	0.48	0.56	0.53	0.52
abroad	acc	0.77	0.73	0.77	0.64	0.78	0.70
	f1-score	0.72	0.71	0.71	0.66	0.71	0.70
comp-problem	acc	0.58	0.56	0.55	0.48	0.47	0.54
-solved	f1-score	0.58	0.54	0.49	0.48	0.43	0.49
influenza	acc	0.52	0.57	0.56	0.54	0.55	0.52
	f1-score	0.51	0.55	0.54	0.51	0.49	0.52
movie	acc	0.56	0.57	0.53	0.47	0.54	0.61
	f1-score	0.57	0.57	0.53	0.47	0.45	0.54
lentil	acc	0.65	0.67	0.66	0.59	0.64	0.60
	f1-score	0.62	0.64	0.61	0.62	0.59	0.56

Inspired by the significant effect of the NARS scores on the preference of participants, another classification study was conducted with the same classifiers trained solely on the responses to the narratives to predict the preferred choice of participants based on their attitudes towards robots. Therefore, the same set of classifiers were trained and tested within the 3 attitude clusters to explore the effect of preconception of the participants on their choices about the robot's behaviors. The average accuracy and f1-score values for the classifiers are displayed in Table 4.4.

Table 4.4 : Average accuracy and f1-score metrics for classifiers trained and tested within attitude clusters.

Attitudes	Metric	KNN	GNB	SVM	DTC	RFC	MLP
All	acc	0.59	0.59	0.60	0.54	0.59	0.58
All	f1-score	0.56	0.58	0.55	0.54	0.51	0.55
Negative	acc	0.60	0.58	0.60	0.51	0.60	0.57
Negative	f1-score	0.56	0.57	0.58	0.52	0.54	0.57
Neutral*	acc	0.61	0.59	0.63	0.54	0.62	0.59
Neutral	f1-score	0.54	0.59	0.60	0.56	0.58	0.54
Positive	acc	0.54	0.54	0.56	0.53	0.54	0.54
Positive**	f1-score	0.52	0.53	0.52	0.53	0.47	0.52

*significant improvement, **significant decline

The results of the within-cluster classification showed that there was no significant change on the classification performance for the negative attitude cluster when compared with the overall results. On the other hand, a significant improvement was noted on the average accuracy of the classifiers for the neutral attitude cluster, even if there were no significant change for the f1-scores. These results support the hypothesis that people having neutral attitudes towards robots are more consistent and predictable in their preference on the robot's behavior. However, when compared with the overall results a significant decline was observed on the f1-score of the classifiers for the positive attitude cluster. Even if there was no significant difference on the average accuracy of classifiers, these results showed that the responses of participants from the positive attitude clusters did not have a balanced class distribution.

The overall results and findings reported in this chapter show that an artificial emotional model can be computationally modeled and implemented for a social robot to make the interaction between the human companion-robot pair more natural (RQ3). Since the results demonstrated that the attitudes of people towards robots had significant effect on their preferences about the robot's emotional behavior, the proposed affect model can be adapted and personalized using standard and cost-effective machine learning approaches (RQ4) according to the human companion's preference and attitude towards robot so that the social robot can learn what emotional behavior to display and adjust the intensity of its behavior.

5. CONCLUSIONS

As robots make their presence known in human interactive environments, such as health, education, entertainment and social domains, we require methods to design robots that can interact with their environment in a social way. This shift in the robotics studies bring forward the necessity to design robots having emotional understanding and using empathy as a social cue.

In this dissertation, a computational model for emotional understanding and emotional behavior generation and modulation was designed and implemented for a companion robot. The presented models were designed: (1) to process multi-channel stimuli, i.e. vision-based facial landmarks, physiological data-based signals, in order to detect the affective states of the human companion; (2) to generate in-situ affective social behaviors for the robot based on the interaction context; (3) to adapt the intensity of the robot's emotional expressions based on the preferences of the human companion.

The aim of the presented study was to conduct a detailed investigation and explore the answers for five research questions. The investigation for the answers posed many challenges throughout the research study. The most challenging situation were caused by the Covid-19 outbreak, first the psychological aspect due to isolation and the pandemic, then the computational complexity aspect caused by the facial masks, protective suits, and the restriction of the tactile interaction. Another challenge was issued by the drop-out from the longitudinal user studies resulting in a data set with a smaller size than expected. Due to not having a larger amount of multimodal interaction data, the performance of recognition and detection algorithms suffered. Another challenging situation were due to the automatic recognition algorithms having low accuracy and detection errors caused by low resolution or illumination, facial occlusion, and changes of the camera angle or the noisy physiological data captured by the sensors. These challenges brought new challenges for data collection and training as well as the evaluation of the proposed system and the quality of the interaction.

To minimize the effect of the presented challenges, we opted for some limitations on the study. First, we assumed that the companion robot would be fixed and remain stationary, without the ability to navigate between humans to simplify the challenges introduced by a moving robot. This case was only studied in a collaborative work, highlighting the social navigation in a hospital environment and gave promising results for the future studies [109]. Second, we assumed that the social robot would interact with its companion in a one-to-one interaction design. Third, although the emotional detection part is crucial for the study, the performance or improvement of the emotion recognition approaches were left out of scope. The performance improvement for vision-based recognition algorithms was only studied in a collaborative work, proposing to use a contrastive learning-based facial action unit detection system for children with hearing impairments [110]. In this study, available open-source modules were opted for the task of automatic recognition. And lastly, the audio recordings as well as the semantic analysis of the human companion's verbal expressions were left out of scope. Despite all the presented challenges and limitations, we were able to answer the five research questions on the affective social behavior of a companion robot.

The first research question was about the ability of the robot to display basic emotions and on the interpretability of these emotions by the human partner (RQ1). The presented study with hearing-impaired children and the affective social robot assisting them through their auditory perception consultation showed that the Pepper robot was able to display emotional behaviors and the children could correctly identify and interpret them.

The answer for the second research question, on the detection and recognition of the human companion's emotional states (RQ2), were explored through the user studies for the RoboRehab and EMBOA research projects. The RoboRehab results showed that the affective robot was able to trigger some emotional changes in children and cause difference in their physiological signals. And physiological signals can be used to differentiate the emotions of children with machine learning and deep learning approaches. On the other hand, the EMBOA results showed that a multimodal approach can be used for the detection of children's emotional states. They also revealed that using a combination of multimodal behavioral metrics, i.e. facial action units, physiological signals, interaction-specific metrics, provides an efficient solution to understand the emotional state and infer the emotional model of the human companion.

The answer for third research question on the modeling an artificial emotional mechanism for the social robot (RQ3) was investigated with the implementation of: (1) a linear affect model, CASPAR; (2) a vision-based model where the robot detects the affective state of its companion and develops a behavioral strategy to adapt its emotional behaviors and improve the negative mood of its companion; and (3) a behavioral model for the robot to predict the preferences of its companion and regulate the intensity of its emotional behaviors accordingly. The results of the user studies and the findings based on the online survey study revealed that an emotional model can be computationally modeled and implemented for a social robot to adapt its emotional behaviors to the preferences and needs of its companion, consequently, to make the interaction between the human companion-robot pair more natural.

The fourth research question was about on the learning techniques that will be adopted by the robot to learn which emotional behavior to express during the interaction. First, the multimodal data collected and processed through RoboRehab and EMBOA user studies showed that the emotion of the human partner can be detected by using standard machine learning approaches and deep learning-based methods. Secondly, the results of the online survey study showed that people's preferences on the robot's emotional behavior change according to their attitudes towards robots. Hence, the need to learn the human companion's preferences and attitudes. Different types of classifiers were used to predict the preferences of people using machine learning based approaches, that can be replicated and implemented on a social robot in real-world interaction scenarios. The proposed computational model can also be used for the customization of the robot based on its companion's preferences and attitude so that the social robot can learn what emotional behavior to display and adjust the intensity of its behavior.

Finally, the fifth research question was based on the hypothesis that a social robot with emotional understanding can foster the interaction gain in an assistive HRI scenario. The RoboRehab and EMBOA user studies with different groups of children (with typical development, with hearing impairment, with autism) showed that independently from their profile the children were more involved and paid attention to the social robot. Even though the objective evaluation metrics did not point a significant difference, the subjective evaluation of the audiologists, therapists, pedagogues complied with the presented hypothesis. These findings also agreed with the HRI literature, the researchers reported that human - robot interaction increases participation in a given task [111,112]. Even though the study performance of participants do not differ between the robotic setup and the conventional setup, higher engagement and higher co-operation times are reported in previous studies [101,102]. Furthermore, the self-report surveys with children in RoboRehab study showed that the children accepted the affective robot as an intelligent and funny social being. These results demonstrate that an affective social robot can foster the interaction gain in an HRI scenario based on assistance.

The overall results presented in this dissertation are promising for the affective social robots but, as mentioned previously, there remain some limitations in the presented study that we seek to address in the future. In the first step, we aim to improve the accuracy of the proposed multimodal emotion recognition approach and integrate it into the proposed affect mechanism on a physical robot. Another objective for the future studies is to populate the data set and increase the sample size and the amount of multimodal data, integrating new modalities such as gaze direction or audio features of the human companion into the presented framework in order to evaluate and also validate the performance of the robot through longitudinal user studies. And finally, we aim to integrate the proposed affect model into a mobile robot and use it for the social navigation of the affective robot.

REFERENCES

[1] **Fong, T.**, **Nourbakhsh, I. and Dautenhahn, K.** (2003). A Survey of Socially Interactive Robots, *Robotics and Autonomous Systems*, *42*, 143–166.

[2] **Goodrich, M.A. and Schultz, A.C.** (2007). Human-Robot Interaction: A Survey, *Foundations and Trends in Human-Computer Interaction*, *1*(3), 203–275.

[3] **Bauer, A.**, **Wollherr, D. and Buss, M.** (2008). Human-Robot Collaboration: A Survey, *International Journal of Humanoid Robotics*, *05*(01), 47–66.

[4] **Tapus, A.**, **Mataric, M.J. and Scassellati, B.** (2007). Socially Assistive Robotics [Grand Challenges of Robotics], *IEEE Robotics Automation Magazine*, *14*(1), 35–42.

[5] **Cifuentes, C.A.**, **Pinto, M.J.**, **Céspedes, N. and Múnera, M.** (2020). Social Robots in Therapy and Care, *Current Robotics Reports*, *1*, 57–74.

[6] **Kabacińska, K.**, **Prescott, T.J. and Robillard, J.M.** (2020). Socially Assistive Robots as Mental Health Interventions for Children: A Scoping Review, *International Journal of Social Robotics*, 1–17.

[7] **Ferrante, G.**, **Vitale, G.**, **Licari, A.**, **Montalbano, L.**, **Pilato, G.**, **Infantino, I.**, **Augello, A. and La Grutta, S.** (2021). Social Robots and Therapeutic Adherence: A New Challenge in Pediatric Asthma?, *Paediatric Respiratory Reviews*, *40*, 46–51.

[8] **Logan, D.E.**, **Breazeal, C.**, **Goodwin, M.S.**, **Jeong, S.**, **O'Connell, B.**, **Smith-Freedman, D.**, **Heathers, J. and Weinstock, P.** (2019). Social Robots for Hospitalized Children, *Pediatrics*, *144*(1).

[9] **Moerman, C.J. and Jansens, R.M.** (2021). Using Social Robot PLEO to Enhance the Well-being of Hospitalised Children, *Journal of Child Health Care*, *25*(3), 412–426.

[10] **Pennisi, P.**, **Tonacci, A.**, **Tartarisco, G.**, **Billeci, L.**, **Ruta, L.**, **Gangemi, S. and Pioggia, G.** (2016). Autism and Social Robotics: A Systematic Review, *Autism Research*, *9*(2), 165–183.

[11] **Martinez-Martin, E.**, **Escalona, F. and Cazorla, M.** (2020). Socially Assistive Robots for Older Adults and People with Autism: An Overview, *Electronics*, *9*(2).

[12] **Kose, H., Akalin, N. and Uluer, P.** (2014). Socially Interactive Robotic Platforms as Sign Language Tutors, *International Journal of Humanoid Robotics*, *11*(01), 1450003, DOI: 10.1142/S0219843614500030.

[13] **Zwaigenbaum, L. and Penner, M.** (2018). Autism Spectrum Disorder: Advances in Diagnosis and Evaluation, *Bmj*, *361*, k1674, DOI: 10.1136/bmj.k1674.

[14] **Clabaugh, C., Becerra, D., Deng, E., Ragusa, G. and Matarić, M.** (2018). Month-Long, In-Home Case Study of a Socially Assistive Robot for Children with Autism Spectrum Disorder, *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*, HRI '18, Association for Computing Machinery, New York, NY, USA, p.87–88.

[15] **Javed, H. and Park, C.H.** (2019). Interactions With an Empathetic Agent: Regulating Emotions and Improving Engagement in Autism, *IEEE Robotics Automation Magazine*, *26*(2), 40–48.

[16] **Li, J., Davison, D., Alcorn, A., Williams, A., Dimitrijevic, S.B., Petrovic, S., Chevalier, P., Schadenberg, B., Ainger, E., Pellicano, L. and Evers, V.** (2020). Non-Participatory User-Centered Design of Accessible Teacher-Teleoperated Robot and Tablets for Minimally Verbal Autistic Children, *Proceedings of the 13th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '20, Association for Computing Machinery, New York, NY, USA, pp.1–9.

[17] **Gratch, J. and Marsella, S.** (2005). Evaluating a Computational Model of Emotion, *Autonomous Agents and Multi-Agent Systems*, *11*(1), 23–43.

[18] **Premack, D. and Woodruff, G.** (1978). Does the Chimpanzee Have a Theory of Mind?, *Behavioral and Brain Sciences*, *1*(4), 515–526.

[19] **Gweon, H. and Saxe, R.** (2013). Developmental Cognitive Neuroscience of Theory of Mind, *Neural Circuit Development and Function in the Brain: Comprehensive Developmental Neuroscience*, *3*, 367–377.

[20] **Aylett, R., Ritter, C., Lim, M.Y., Broz, F., McKenna, P.E., Keller, I. and Rajendran, G.** (2021). An Architecture for Emotional Facial Expressions as Social Signals, *IEEE Transactions on Affective Computing*, *12*(2), 293–305.

[21] **Asada, M., Hosoda, K., Kuniyoshi, Y., Ishiguro, H., Inui, T., Yoshikawa, Y., Ogino, M. and Yoshida, C.** (2009). Cognitive Developmental Robotics: A Survey, *IEEE Transactions on Autonomous Mental Development*, *1*(1), 12–34.

[22] **Izard, C.E.** (1993). Four Systems for Emotion Activation: Cognitive and Noncognitive Processes, *Psychological Review*, *100*(1), 68.

[23] **Ekman, P. and Oster, H.** (1982). *Emotion in the Human Face*, Cambridge University Press.

[24] **Levenson, R.W.**, (1994). Human Emotion: A Functional View, P. Ekman and R. Davidson (Eds.), *The Nature of Emotion: Fundamental Questions*, Oxford University Press New York, NY, pp.123–126.

[25] **Lazarus, R.S.**, (1994). Universal Antecedents of the Emotions, P. Ekman and R. Davidson (Eds.), *The Nature of Emotion: Fundamental Questions*, Oxford University Press New York, NY, pp.163–171.

[26] **Scherer, K.R.**, (1994). Evidence for Both Universality and Cultural Specificity of Emotion Elicitation, P. Ekman and R. Davidson (Eds.), *The Nature of Emotion: Fundamental Questions*, Oxford University Press New York, NY, pp.172–175.

[27] **Plutchik, R.** (1991). *The Emotions*, University Press of America.

[28] **Ortony, A.**, **Clore, G.L. and Collins, A.** (1988). *The Cognitive Structure of Emotions*, Cambridge University Press.

[29] **Lazarus, R.S.** (1982). Thoughts on the Relations Between Emotion and Cognition., *American Psychologist*, *37*(9), 1019.

[30] **Scherer, K.R.** (2005). What Are Emotions? And How Can They Be Measured?, *Social Science Information*, *44*(4), 695–729.

[31] **Mehrabian, A. and Russell, J.A.** (1974). *An Approach to Environmental Psychology.*, MIT Press.

[32] **Ekman, R.** (1997). *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, Oxford University Press, USA.

[33] **Breazeal, C.** (2003). Emotion and Sociable Humanoid Robots, *International Journal of Human-Computer Studies*, *59*(1-2), 119–155.

[34] **Lim, M.Y. and Aylett, R.** (2009). An Emergent Emotion Model for an Affective Mobile Guide with Attitude, *Applied Artificial Intelligence*, *23*(9), 835–854.

[35] **Russell, J.A.** (1980). A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, *39*(6), 1161.

[36] **Watanabe, A.**, **Ogino, M. and Asada, M.** (2007). Mapping Facial Expression to Internal States Based on Intuitive Parenting, *Journal of Robotics and Mechatronics*, *19*(3), 315.

[37] **Asada, M.**, **MacDorman, K.F.**, **Ishiguro, H. and Kuniyoshi, Y.** (2001). Cognitive Developmental Robotics as a New Paradigm for the Design of Humanoid Robots, *Robotics and Autonomous Systems*, *37*(2-3), 185–193.

[38] **Boucenna, S.**, **Gaussier, P.**, **Andry, P. and Hafemeister, L.** (2010). Imitation as a Communication Tool for Online Facial Expression Learning and Recognition, *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp.5323–5328.

[39] **Boucenna, S.**, **Anzalone, S.**, **Tilmont, E.**, **Cohen, D. and Chetouani, M.** (2014). Learning of Social Signatures Through Imitation Game Between a Robot and a Human Partner, *IEEE Transactions on Autonomous Mental Development*, *6*(3), 213–225.

[40] **Lim, A. and Okuno, H.G.**, (2015). Developing Robot Emotions through Interaction with Caregivers, *Handbook of Research on Synthesizing Human Emotion in Intelligent Systems and Robotics*, IGI Global, pp.316–337.

[41] **Marsella, S.**, **Gratch, J.**, **Petta, P.** *et al.* (2010). Computational Models of Emotion, *A Blueprint for Affective Computing-A Sourcebook and Manual*, *11*(1), 21–46.

[42] **El-Nasr, M.S.**, **Yen, J. and Ioerger, T.R.** (2000). FLAME—Fuzzy Logic Adaptive Model of Emotions, *Autonomous Agents and Multi-Agent Systems*, *3*(3), 219–257.

[43] **Bui, T.D.**, **Heylen, D.**, **Poel, M. and Nijholt, A.** (2002). ParleE: An Adaptive Plan Based Event Appraisal Model of Emotions, *M. Jarke, G. Lakemeyer and J. Koehler (Eds.), KI 2002: Advances in Artificial Intelligence*, Springer Berlin Heidelberg, Berlin, Heidelberg, pp.129–143.

[44] **Marsella, S.C. and Gratch, J.** (2009). EMA: A Process Model of Appraisal Dynamics, *Cognitive Systems Research*, *10*(1), 70 – 90.

[45] **Dang, T.H.H.**, **Letellier-Zarshenas, S. and Duhaut, D.**, (2008). Grace–Generic Robotic Architecture to Create Emotions, *Advances In Mobile Robotics*, World Scientific, pp.174–181.

[46] **Saint-Aimé, S.**, **Le-Pévédic, B. and Duhaut, D.**, (2009). iGrace – Emotional Computational Model for EmI Companion Robot, V.A. Kulyukin (Ed.), *Advances in Human-Robot Interaction*, chapter 4, IntechOpen, Rijeka, pp.1–26.

[47] **Saint-Aimé, S.**, **Le Pévédic, B. and Duhaut, D.** (2011). Children Recognize Emotions of EmI Companion Robot, *2011 IEEE International Conference on Robotics and Biomimetics*, pp.1153–1158.

[48] **Leite, I.**, **Castellano, G.**, **Pereira, A.**, **Martinho, C. and Paiva, A.** (2014). Empathic Robots for Long-term Interaction, *International Journal of Social Robotics*, *6*(3), 329–341.

[49] **Leite, I.**, **Pereira, A.**, **Mascarenhas, S.**, **Martinho, C.**, **Prada, R. and Paiva, A.** (2013). The Influence of Empathy in Human–Robot Relations, *International Journal of Human-Computer Studies*, *71*(3), 250 – 260.

[50] **Paiva, A., Leite, I., Boukricha, H. and Wachsmuth, I.** (2017). Empathy in Virtual Agents and Robots: A Survey, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, *7*(3), 11.

[51] **Seo, S.H., Geiskkovitch, D., Nakane, M., King, C. and Young, J.E.** (2015). Poor Thing! Would You Feel Sorry for a Simulated Robot? A Comparison of Empathy Toward a Physical and a Simulated Robot, *2015 10th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, IEEE, pp.125–132.

[52] **Tsiourti, C., Weiss, A., Wac, K. and Vincze, M.** (2019). Multimodal Integration of Emotional Signals from Voice, Body, and Context: Effects of (In)congruence on Emotion Recognition and Attitudes Towards Robots, *International Journal of Social Robotics*, *11*(4), 555–573.

[53] **Dahlbäck, N., Jönsson, A. and Ahrenberg, L.** (1993). Wizard of Oz Studies - Why and How, *Knowledge-based systems*, *6*(4), 258–266.

[54] **Hussein, A., Gaber, M.M., Elyan, E. and Jayne, C.** (2017). Imitation Learning: A Survey of Learning Methods, *ACM Comput. Surv.*, *50*(2).

[55] **Huang, C.M. and Mutlu, B.** (2014). Learning-based Modeling of Multimodal Behaviors for Humanlike Robots, *Proceedings of the 2014 ACM/IEEE International Conference on Human-robot Interaction*, ACM, pp.57–64.

[56] **Liu, P., Glas, D.F., Kanda, T., Ishiguro, H. and Hagita, N.** (2014). How to Train Your Robot - Teaching Service Robots to Reproduce Human Social Behavior, *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, IEEE, pp.961–968.

[57] **Qureshi, A.H., Nakamura, Y., Yoshikawa, Y. and Ishiguro, H.** (2016). Robot Gains Social Intelligence Through Multimodal Deep Reinforcement Learning, *2016 IEEE-RAS 16th International Conference on Humanoid Robots (Humanoids)*, IEEE, pp.745–751.

[58] **Hemminahaus, J. and Kopp, S.** (2017). Towards Adaptive Social Behavior Generation for Assistive Robots Using Reinforcement Learning, *2017 12th ACM/IEEE International Conference on Human-Robot Interaction (HRI*, IEEE, pp.332–340.

[59] **Moro, C., Nejat, G. and Mihailidis, A.** (2018). Learning and Personalizing Socially Assistive Robot Behaviors to Aid with Activities of Daily Living, *ACM Transactions on Human-Robot Interaction (THRI)*, *7*(2), 15.

[60] **Senft, E., Baxter, P., Kennedy, J., Lemaignan, S. and Belpaeme, T.** (2017). Supervised Autonomy for Online Learning in Human-Robot Interaction, *Pattern Recognition Letters*, *99*, 77–86.

[61] **Charrier, L.**, **Rieger, A.**, **Galdeano, A.**, **Cordier, A.**, **Lefort, M. and Hassas, S.** (2019). The RoPE Scale: a Measure of How Empathic a Robot is Perceived, *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pp.656–657.

[62] **Nomura, T.** (2016). Empathy as Signalling Feedback Between Humanoid Robots and Humans, *Humanoid Robotics: A Reference*, 1–10.

[63] **Nomura, T.**, **Kanda, T.**, **Suzuki, T. and Kato, K.** (2004). Psychology in Human-Robot Communication: An Attempt Through Investigation of Negative Attitudes and Anxiety Toward Robots, *RO-MAN 2004. 13th IEEE International Workshop on Robot and Human Interactive Communication*, pp.35–40.

[64] **Nomura, T.**, **Suzuki, T.**, **Kanda, T. and Kato, K.** (2006). Measurement of Anxiety toward Robots, *ROMAN 2006 - The 15th IEEE International Symposium on Robot and Human Interactive Communication*, pp.372–377.

[65] **Barrett-Lennard, G.T.** (1962). Dimensions of Therapist Response as Causal Factors in Therapeutic Change, *Psychological Monographs: General and Applied*, *76*(43), 1.

[66] **Barrett-Lennard, G.T.**, (1986). The Relationship Inventory Now: Issues and Advances in Theory, Method, and Use, L.S. Greenberg and W.M. Pinsof (Eds.), *Guilford Clinical Psychology and Psychotherapy Series. The Psychotherapeutic Process: A Research Handbook*, Guilford Press, New York, NY, US, pp.439–476.

[67] **Tapus, A. and Mataric, M.J.** (2007). Emulating Empathy in Socially Assistive Robotics, *AAAI Spring Symposium: Multidisciplinary Collaboration for Socially Assistive Robotics*, pp.93–96.

[68] **Davis, M.H.** (1983). Measuring Individual Differences in Empathy: Evidence for a Multidimensional Approach, *Journal of Personality and Social Psychology*, *44*(1), 113.

[69] **Costa, P.T. and Mac Crae, R.R.** (1992). *Neo Personality Inventory-Revised (NEO PI-R)*, Psychological Assessment Resources Odessa, FL.

[70] **Digman, J.M.** (1990). Personality Structure: Emergence of the Five-Factor Model, *Annual Review of Psychology*, *41*(1), 417–440.

[71] **Walters, M.L.**, **Syrdal, D.S.**, **Dautenhahn, K.**, **Te Boekhorst, R. and Koay, K.L.** (2008). Avoiding the Uncanny Valley: Robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion, *Autonomous Robots*, *24*(2), 159–178.

[72] **Meerbeek, B.**, **Saerbeck, M. and Bartneck, C.** (2009). Iterative Design Process for Robots with Personality, *Proceedings of the AISB 2009 Symposium on New Frontiers in Human-Robot Interaction Edingburgh*, volume 94, p.101.

[73] **Santamaria, T. and Nathan-Roberts, D.** (2017). Personality Measurement and Design in Human-Robot Interaction: A Systematic and Critical Review, *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, *61*(1), 853–857.

[74] **Goldberg, L.R.** (1999). A Broad-bandwidth, Public Domain, Personality Inventory Measuring the Lower-level Facets of Several Five-Factor Models, *Personality psychology in Europe*, *7*(1), 7–28.

[75] **Donnellan, M.B.**, **Oswald, F.L.**, **Baird, B.M. and Lucas, R.E.** (2006). The Mini-IPIP Scales: Tiny-yet-effective measures of the Big Five factors of personality, *Psychological assessment*, *18*(2), 192.

[76] **Bartneck, C.**, **Kulić, D.**, **Croft, E. and Zoghbi, S.** (2009). Measurement Instruments for the Anthropomorphism, Animacy, Likeability, Perceived Intelligence, and Perceived Safety of Robots, *International Journal of Social Robotics*, *1*(1), 71–81.

[77] **Charrier, L.**, **Galdeano, A.**, **Cordier, A. and Lefort, M.** (2018). Empathy Display Influence on Human-Robot Interactions: A Pilot Study, *Workshop on Towards Intelligent Social Robots: From Naive Robots to Robot Sapiens at the 2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS 2018)*, Madrid, Spain, pp.1–7.

[78] **Bickmore, T.W.** (2003). *Relational Agents: Effecting Change Through Human-Computer Relationships*, (Doctoral dissertation). Retrieved from https://affect.media.mit.edu/pdfs/03.bickmore-phd.pdf.

[79] **Bradley, M.M. and Lang, P.J.** (1994). Measuring Emotion: The Self-Assessment Manikin and the Semantic Differential, *Journal of Behavior Therapy and Experimental Psychiatry*, *25*(1), 49 – 59.

[80] **Saerbeck, M. and Bartneck, C.** (2010). Perception of Affect Elicited by Robot Motion, *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*, HRI '10, IEEE Press, Piscataway, NJ, USA, pp.53–60.

[81] **Alves-Oliveira, P.**, **Sequeira, P.**, **Melo, F.S.**, **Castellano, G. and Paiva, A.** (2019). Empathic Robot for Group Learning: A Field Study, *ACM Transactions on Human-Robot Interaction*, *8*(1), 3:1–3:34.

[82] **Truax, C.B.** (1967). A Scale for the Rating of Accurate Empathy, *The Therapeutic Relationship and Its Impact*, 555–568.

[83] **Miura, N.**, **Sugiura, M.**, **Takahashi, M.**, **Moridaira, T.**, **Miyamoto, A.**, **Kuroki, Y. and Kawashima, R.** (2008). An Advantage of Bipedal Humanoid Robot on the Empathy Generation: A Neuroimaging Study, *2008 IEEE/RSJ International Conference on Intelligent Robots and Systems*, IEEE, pp.2465–2470.

[84] **Rosenthal-Von Der Pütten, A.M.**, **Schulte, F.P.**, **Eimler, S.C.**, **Sobieraj, S.**, **Hoffmann, L.**, **Maderwald, S.**, **Brand, M. and Krämer, N.C.** (2014). Investigations on Empathy Towards Humans and Robots Using fMRI, *Computers in Human Behavior*, *33*, 201–212.

[85] **Suzuki, Y.**, **Galli, L.**, **Ikeda, A.**, **Itakura, S. and Kitazaki, M.** (2015). Measuring Empathy for Human and Robot Hand Pain Using Electroencephalography, *Scientific Reports*, *5*, 15924.

[86] **Kozima, H.**, **Nakagawa, C. and Yano, H.** (2004). Can a Robot Empathize with People?, *Artificial Life and Robotics*, *8*(1), 83–88.

[87] **Zeng, Z.**, **Pantic, M.**, **Roisman, G.I. and Huang, T.S.** (2009). A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *31*(1), 39–58.

[88] **URL-1**. *Pepper's voice sample <https://youtu.be/IfgUDvbv4fs>*, date retrieved: 17.12.2022.

[89] **URL-2**. *Audio recordings of Pepper <https://drive.google.com/open?id=1xj7mx2Z6jSa527ZVOFiK_9KwvZ8sMtTe>*, date retrieved: 17.12.2022.

[90] **URL-3**. *Pepper's animations <https://drive.google.com/open?id=1ZMF2DF71OuMF6fWW2mukn7Xc2zsCJ_ZD>*, date retrieved: 18.12.2022.

[91] **URL-4**. *Different moods of Pepper <https://youtu.be/WS7D0IKTGBI>*, date retrieved: 18.12.2022.

[92] **Savaşır, I.**, **Sezgin, N. and Erol, N.** (1994). *Ankara Gelişim Tarama Envanteri El Kitabı*, Ankara Türk Psikologlar Birliği.

[93] **Frankenburg, W.K.**, **Dodds, J.**, **Archer, P.**, **Shapiro, H. and Bresnick, B.** (1992). The Denver II: A Major Revision and Restandardization of the Denver Developmental Screening Test, *Pediatrics*, *89*(1), 91–97.

[94] **Wechsler, D.** (1974). *Wechsler Intelligence Scale for Children - Revised*, Psychological Corporation.

[95] **Hareli, S. and Parkinson, B.** (2008). What's Social About Social Emotions?, *Journal for the Theory of Social Behaviour*, *38*(2), 131–156.

[96] **Uluer, P.**, **Kose, H.**, **Oz, B.K.**, **Can Aydinalev, T. and Barkana, D.E.** (2020). Towards An Affective Robot Companion for Audiology Rehabilitation: How Does Pepper Feel Today?, *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp.567–572.

[97] **Reynolds, C.R.**, **Voress, J.K. and Pearson, N.A.** (2008). *DTAP: Developmental Test of Auditory Perception*, Pro-Ed.

[98] **Friard, O. and Gamba, M.** (2016). BORIS: A Free, Versatile Open-source Event-logging Software for Video/audio Coding and Live Observations, *Methods in Ecology and Evolution*, *7*(11), 1325–1330.

[99] **Çınar, M. and Polat, Z.** (2018). *Normal İşitmeye Sahip Bireylerde İşitsel Algı Becerilerinin Değerlendirilmesi*, (Master's Thesis), Available from Yükseköğretim Kurulu Ulusal Tez Merkezi (No. 540565).

[100] **Uluer, P.**, **Kose, H. and Barkana, D.E.** (2021). Experience with an Affective Robot Assistant for Children with Hearing Disabilities, *International Journal of Social Robotics*, 1–18, DOI: 10.1007/s12369-021-00830-5.

[101] **Huijnen, C.A.**, **Lexis, M.A.**, **Jansens, R. and de Witte, L.P.** (2017). How to Implement Robots in Interventions for Children with Autism? A co-creation Study Involving People with Autism, Parents and Professionals, *Journal of Autism and Developmental Disorders*, *47*(10), 3079–3096.

[102] **Yun, S.S.**, **Choi, J.**, **Park, S.K.**, **Bong, G.Y. and Yoo, H.** (2017). Social Skills Training for Children with Autism Spectrum Disorder Using a Robotic Behavioral Intervention System, *Autism Research*, *10*(7), 1306–1323.

[103] **Bartl-Pokorny, K.D.**, **Pykała, M.**, **Uluer, P.**, **Barkana, D.E.**, **Baird, A.**, **Kose, H.**, **Zorcec, T.**, **Robins, B.**, **Schuller, B.W. and Landowska, A.** (2021). Robot-Based Intervention for Children With Autism Spectrum Disorder: A Systematic Literature Review, *IEEE Access*, *9*, 165433–165450.

[104] **Baltrusaitis, T.**, **Zadeh, A.**, **Lim, Y.C. and Morency, L.P.** (2018). OpenFace 2.0: Facial Behavior Analysis Toolkit, *2018 13th IEEE International Conference on Automatic Face Gesture Recognition (FG 2018)*, pp.59–66.

[105] **URL-5**. *iMotions Facial Action Coding System (FACS) – A Visual Guidebook* <*https://imotions.com/blog/facial-action-coding-system/*>, date retrieved: 05.12.2022.

[106] **Coşkun, B.**, **Uluer, P.**, **Toprak, E.**, **Barkana, D.E.**, **Kose, H.**, **Zorcec, T.**, **Robins, B. and Landowska, A.** (2022). Stress Detection of Children with Autism using Physiological Signals in Kaspar Robot-Based Intervention Studies, *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)*, pp.01–07.

[107] **Battini Sönmez, E.**, **Han, H.**, **Karadeniz, O.**, **Dalyan, T. and Sarıoğlu, B.** (2022). EMRES: A New EMotional RESpondent Robot, *IEEE Transactions on Cognitive and Developmental Systems*, *14*(2), 772–780.

[108] **Toisoul, A.**, **Kossaifi, J.**, **Bulat, A.**, **Tzimiropoulos, G. and Pantic, M.** (2021). Estimation of Continuous Valence and Arousal Levels from Faces in Naturalistic Conditions, *Nature Machine Intelligence*, *3*(1), 42–50.

[109] **Kivrak, H.**, **Uluer, P.**, **Kose, H.**, **Gumuslu, E.**, **Erol Barkana, D.**, **Cakmak, F. and Yavuz, S.** (2020). Physiological Data-Based Evaluation of a Social Robot Navigation System, *2020 29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, pp.994–999.

[110] **Gurpinar, C.**, **Takir, S.**, **Bicer, E.**, **Uluer, P.**, **Arica, N. and Kose, H.** (2022). Contrastive Learning based Facial Action Unit Detection in Children with Hearing Impairment for a Socially Assistive Robot Platform, *Image and Vision Computing*, *128*, 104572.

[111] **Powers, A.**, **Kiesler, S.**, **Fussell, S. and Torrey, C.** (2007). Comparing a Computer Agent with a Humanoid Robot, *Proceedings of the ACM/IEEE International Conference on Human-Robot Interaction*, pp.145–152.

[112] **Köse, H.**, **Uluer, P.**, **Akalın, N.**, **Yorgancı, R.**, **Özkul, A. and Ince, G.** (2015). The Effect of Embodiment in Sign Language Tutoring with Assistive Humanoid Robots, *International Journal of Social Robotics*, *7*(4), 537–548.

CURRICULUM VITAE

Name Surname **:** Pınar ULUER

EDUCATION **:**

- **B.Sc.** **:** 2009, Galatasaray University, Faculty of Engineering and Technology, Department of Computer Engineering
- **M.Sc.** **:** 2012, Galatasaray University, Graduate School of Science and Engineering, Department of Computer Engineering

PROFESSIONAL EXPERIENCE AND REWARDS:

- 2011- ... Teaching and Research Assistant in Galatasaray University
- 2013-2015 Boursier in “Robot Assisted Sign Language Tutoring" Project funded by Scientific and Technical Research Council of Turkey (TÜBİTAK)
- 2015-2017 Boursier in “Avatar based Machine Translation System from written Turkish to TID" funded by TÜBİTAK
- 2018-2020 Boursier in “RoboRehab: Assitive Audiology Rehabilitation Robot" funded by TÜBİTAK
- 2019-2022 Researcher in EMBOA (Affective loop in Socially Assistive Robotics as an intervention tool for children with autism) Project under the Erasmus+ programme

PUBLICATIONS, PRESENTATIONS AND PATENTS ON THE THESIS:

- Gurpinar, C., Takir, S., Bicer, E., **Uluer, P.**, Arica, N., & Kose, H. (2022). Contrastive Learning Based Facial Action Unit Detection in Children with Hearing Impairment for a Socially Assistive Robot Platform. *Image and Vision Computing*, 128, 104572. (Article Instance)
- Coşkun, B., **Uluer, P.**, Toprak, E., Barkana, D. E., Kose, H., Zorcec, T., Robins, B., & Landowska, A. (2022, August). Stress Detection of Children with Autism using Physiological Signals in Kaspar Robot-Based Intervention Studies. *2022 9th IEEE RAS/EMBS International Conference for Biomedical Robotics and Biomechatronics (BioRob)* (pp. 01-07). IEEE. (Presentation Instance)

- Biçer, E., Takır, Ş., Gürpınar, C., **Uluer, P.**, & Köse, H. (2022, May). Masking and Compression Techniques for Efficient Action Unit Detection of Children for Social Robots. *Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. (Presentation Instance)

- Aktaş, S. N. B., **Uluer, P.**, Coşkun, B., Toprak, E., Barkana, D. E., Köse, H., Zorcec, T., Robins, B., & Landowska, A. (2022, May). Stress Detection of Children With ASD Using Physiological Signals. *Signal Processing and Communications Applications Conference (SIU)* (pp. 1-4). IEEE. (Presentation Instance)

- Bartl-Pokorny, K. D., Pykala, M., **Uluer, P.**, Barkana, D. E., Baird, A., Kose, H., Zorcec, T., Robins, B., Schuller, B. W., & Landowska, A. (2021). Robot-Based Intervention for Children With Autism Spectrum Disorder: A Systematic Literature Review, *IEEE Access*, vol. 9, pp. 165433-165450. (Article Instance)

- **Uluer, P.**, Kose, H., Gumuslu, E., & Barkana, D. E. (2021). Experience with an Affective Robot Assistant for Children with Hearing Disabilities. *International Journal of Social Robotics*, 1-18. (Article Instance)

- Baglayici, E., Gurpinar, C., **Uluer, P.**, & Kose, H. (2021, January). A New Facial Expression Processing System for an Affectively Aware Robot. *International Conference on Pattern Recognition* (pp. 36-51). Springer, Cham. (Presentation Instance)

- Kivrak, H., **Uluer, P.**, Kose, H., Gumuslu, E., Barkana, D. E., Cakmak, F., & Yavuz, S. (2020, August). Physiological Data-based Evaluation of a Social Robot Navigation System. *29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 994-999). IEEE. (Presentation Instance)

- **Uluer, P.**, Kose, H., Oz, B. K., Aydinalev, T. C., & Barkana, D. E. (2020, January). Towards an Affective Robot Companion for Audiology Rehabilitation: How Does Pepper Feel Today?. *29th IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)* (pp. 567-572). IEEE. (Presentation Instance)

- Gürpınar, C., **Uluer, P.**, Akalın, N., & Köse, H. (2020). Sign Recognition System for an Assistive Robot Sign Tutor for Children. *International Journal of Social Robotics*, 12(2), 355-369. (Article Instance)

- Köse, H., **Uluer, P.**, Akalın, N., Yorgancı, R., Özkul, A., & Ince, G. (2015). The Effect of Embodiment in Sign Language Tutoring with Assistive Humanoid Robots. *International Journal of Social Robotics*, 7(4), 537-548. (Article Instance)

- **Uluer, P.**, Akalın, N., & Köse, H. (2015). A New Robotic Platform for Sign Language Tutoring. *International Journal of Social Robotics*, 7(5), 571-585. (Article Instance)

- Kose, H., Akalin, N., & **Uluer, P.** (2014). Socially Interactive Robotic Platforms as Sign Language Tutors. *International Journal of Humanoid Robotics*, 11(01), 1450003. (Article Instance)