UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas e de Computação

Robotic - Cognitive Adaptive System for Teaching and Learning (R-CASTLE)

Daniel Carnieto Tozadore

Tese de Doutorado do Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional (PPG-CCMC)

SERVIÇO DE PÓS-GRADUAÇÃO DO ICMC-USP

Data de Depósito:

Assinatura: ______________________

Daniel Carnieto Tozadore

Robotic - Cognitive Adaptive System for Teaching and Learning (R-CASTLE)

Thesis submitted to the Institute of Mathematics and Computer Sciences – ICMC-USP – in accordance with the requirements of the Computer and Mathematical Sciences Graduate Program, for the degree of Doctor in Science. *FINAL VERSION*

Concentration Area: Computer Science and Computational Mathematics

Advisor: Profa. Dra. Roseli Aparecida Francelin Romero

USP – São Carlos
July 2020

Ficha catalográfica elaborada pela Biblioteca Prof. Achille Bassi
e Seção Técnica de Informática, ICMC/USP,
com os dados inseridos pelo(a) autor(a)

C289r Carnieto Tozadore, Daniel
Robotic - Cognitive Adaptive System for Teaching and Learning (R-Castle) / Daniel Carnieto Tozadore; orientadora Roseli Aparecida Francelin Romero. -- São Carlos, 2020.
90 p.

Tese (Doutorado - Programa de Pós-Graduação em Ciências de Computação e Matemática Computacional) -- Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, 2020.

1. Artificial Intelligence for Education. 2. Human-Robot Interaction. 3. Adaptive Systems. 4. Social Robotics. I. Aparecida Francelin Romero, Roseli, orient. II. Título.

Bibliotecários responsáveis pela estrutura de catalogação da publicação de acordo com a AACR2:
Gláucia Maria Saia Cristianini - CRB - 8/4938
Juliana de Souza Moraes - CRB - 8/6176

Daniel Carnieto Tozadore

Sistema Cognitivo Adaptativo para Robótica Social na Educação (R-CASTLE)

Tese apresentada ao Instituto de Ciências Matemáticas e de Computação – ICMC-USP, como parte dos requisitos para obtenção do título de Doutor em Ciências – Ciências de Computação e Matemática Computacional. *VERSÃO REVISADA*

Área de Concentração: Ciências de Computação e Matemática Computacional

Orientadora: Profa. Dra. Roseli Aparecida Francelin Romero

USP – São Carlos
Julho de 2020

I dedicate this work to all Brazilian researchers that, in dark times like these,
fight bravely to keep the flame of knowledge alive.
Mainly, for future scientists who may arise from the sparks
of my humble yet passionate research.

ACKNOWLEDGEMENTS

First of all, I would like to thank the Great Creator and Maintainer of everything, my God. And second, the closest thing I will ever have from His love: my beloved family.

To the hero of all adventures, my father Marcos. And to love itself, my mother Herminia, my eternal gratitude for always encouraging my inner child and all the sweet things I carry in my heart.

I would also like to take the opportunity to thank all my grandparents, uncles, aunts, cousins and the "Galera" family in helping my parents in this mission.

To my legendary best friends: Juliana, Michele and Rafael — whom I have the fortune of having as siblings— no matter how much time passes, nothing is ever legendary without your presence. I will be there for you because I want in that way.

My sincere thanks to my supervisor, professor Roseli Romero, for her extraordinary work with me, for all the hours of joy and sadness together and for having matured so much and in so many aspects of my life. I will miss you very much and I know wherever I will be in my future, I will always be brilliant because you believed in me and always dreamed with me.

To my older brother Daniel Ferras, born from different parents but bonded by music, thank you for supporting me all the time with your advice and your melody.

A very special thanks to my lab-mates, of flesh or metal, from the "Laboratório de aprendizado de Robôs", who lovingly suffered this journey with me, especially my very good friend and admired researcher Dr. Adam Moreira and my favorite workmate, Teddy the robot.

Thank you to the under-grad students that I had the honor of mentoring; João Valentini, Victor Rodrigues, Marcos Camargo, Julia Pazinni, Rodrigo Zavarizz , Fernando Vendrameto and Leticia Rina. They had a fundamental role in the R-CASTLE. None of this would be possible without them. I don't need to wish them luck because I witnessed their ability to make everything come to life!

Thank you to my other lab mates from Biocom for all the jokes and professional advice. Countless times, you guys really made my day.

My deepest gratitude to all the ICMC staff, the cleaning staff, canteen attendants, other professors and researchers, and all the graduate service employees. For over 10 years— from my first days at the institute, until now— I witnessed their amazing efficiency in providing an incredible and friendly work environment. I've lost count of how many times a smile or greeting from you guys gave me the strength on my career path.

A very special thanks to the family that was formed during my time at Volei Caaso: coaches, teammates, fans and haters. Thank you for always believing in me and helping me become the man I am, for teaching me important values, including how to respect everyone unconditionally regardless of gender, colour, social status or anything else. I love you very very much, with a lovely highlight to Paula Dias, Fernando Monteiro, Ana Laura Assumpção, Alex Kraus and João Burato.

To my housemates, João Eguea, Bruno de Lima, Danilo Barbosa, Pedro Nobile and Eric Sneider and, by aggregation, Leticia Cursini, Gabriela Cassão and the "former and founders members". Thanks for really supporting me in any mood and at any cost. I will never find a place to call home that compares to being by your side. I love you all more than the USP restaurant.

I thank the amazing friends (humans and robots) I made during my time in the GAIPS, especially professor Ana Paiva. I learned a lot with you guys and wish from the bottom of my heart to see each one of you again.

At last but not least, my eternal gratitude to all my good friends that joined the R-CASTLE dream with in, by making the logo, given me advise, making videos, sharing posts and so on. Leeza Udovenko, Luiza Campos, Tales Tininis, Beatrice Bonami, Jacqueline Costa and John Maycon Alho always have a place in my prayers.

Finally, I would like to thanks the CNPq, CAPES and FAPESP for financial support.

"Valeu a pena?
Tudo vale a pena.
Se a alma não é pequena."
(Fernando Pessoa)

RESUMO

TOZADORE, D. C. **Sistema Cognitivo Adaptativo para Robótica Social na Educação (R-CASTLE)**. 2020. 115 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

A Inteligência Artificial (IA) tem assumido um papel importante na rotina das pessoas. Principalmente porque viabiliza a automação de tarefas repetitivas e a customização de serviços para cada usuário. Ambos recursos são possibilitados pelo conhecimento que se cria a partir de dados gerados por experiências passadas. Especialmente na área da educação, a IA pode ajudar os professores a otimizar seu tempo de trabalho em ações recorrentes de planejamento, execução e avaliação das atividades. Já para os alunos, a IA pode potencializar a experiência de aprendizado por meio de dispositivos interativos que, a princípio, aumentam o interesse e a motivação dos alunos por serem uma novidade e que tentam continuar produzindo esses efeitos a longo prazo por meio de técnicas de adaptação e customização. Entretanto, um dos maiores problemas é a falta de naturalidade para usar essas técnicas como aliadas. Baseado nas necessidades de professores e alunos apresentadas na literatura, este projeto buscou uma forma de atender tais necessidades em uma única abordagem, propondo uma arquitetura computacional que se comunique de uma maneira intuitiva com os professores por meio de uma interface gráfica e com os alunos por meio de um robô social. O resultado é um Sistema Cognitivo Adaptativo para Robótica Social Educacional (Robotic - Cognitive Adaptive System for Teaching an Learning - R-CASTLE). Este sistema tem como objetivo viabilizar algoritmos de IA como ferramentas de auxílio para os professores no planejamento, execução e avaliação de suas atividades educacionais sem que apresentem previamente conhecimentos técnicos desses algoritmos. Ao mesmo tempo, o R-CASTLE oferece para os alunos uma maneira tecnológica e desafiadora de realizar exercícios práticos, em um nível de dificuldade correspondente ao apresentado por cada um deles. Os algoritmos de IA permitem que o robô usar comunicação visual e verbal para coletar valores indicativos nas respostas e expressões corporais dos alunos para avaliar suas habilidades de Atenção, Comunicação e Aprendizagem. Além disso, permitem também usar esses dados na adaptação e customização do sistema a fim de manter os alunos engajados por mais tempo nas atividades. A interface gráfica também proporciona maneiras fáceis de manipular os dados gerados em atividades passadas para serem modificados e otimizados em atividades futuras. Embora seja difícil avaliar estatisticamente a eficiência deste projeto como um todo devido à grande quantidade de dados especializados para esse tipo de solução, estudos com análises dos módulos isolados e testes iniciais do sistema completo têm apontado bons indicativos sobre o potencial dessa ferramenta para colaborar de forma prática e intuitiva com alunos e professores do ensino fundamental e também para possíveis interessados em usar o R-CASTLE em outras tarefas de Interação Humano-Robô.

Palavras-chave: Inteligência Artificial para Educação, Interação Humano-Robô, Sistemas Adaptativos, Robótica Social.

ABSTRACT

TOZADORE, D. C. **Robotic - Cognitive Adaptive System for Teaching and Learning (R-CASTLE)**. 2020. 115 p. Tese (Doutorado em Ciências – Ciências de Computação e Matemática Computacional) – Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo, São Carlos – SP, 2020.

Artificial Intelligence (AI) has taken an important role in people's routine. Mainly because it enables the automation of repetitive tasks and the customization of services for each user. Both of these resources are made possible by the knowledge that is created from data generated by past experiences. Especially in the educational field, AI can help teachers to optimize their working time in recurring actions of planning, executing and evaluating their activities. For students, AI can enhance the learning experience through interactive devices that, at first, increase students' interest and motivation for being a novelty and then try to continue producing these effects in long-term interactions through techniques of adaptation and customization. However, one of the biggest problems is the lack of naturalness to use these techniques as allies. Based on the needs of teachers and students presented in literature, this project sought a way to meet these needs in a unique approach, proposing a computational architecture that communicates in an intuitive way with teachers through a graphical interface and with students through a social robot. The result is a Cognitive Adaptive System for Teaching and Learning (R-CASTLE). This system aims to enable AI algorithms as tools to assist teachers in planning, executing and evaluating their educational activities without having previously presented technical knowledge of these algorithms. At the same time, R-CASTLE offers the students a technological and challenging way to carry out practical exercises, at a level of difficulty corresponding to that presented by each of them. AI algorithms allow the robot to use visual and verbal communication to collect indicative values in students' responses and body expressions to assess their attention, communication and learning skills. Further, it allows also to use this data in adapting and customizing the system in order to maintain students engaged for longer period of time in the activities. The graphical interface also provides easy ways to manipulate data generated from past activities to be modified and optimized for future activities. Although it is difficult to statistically evaluate the efficiency of this project as a whole due to the large amount of specialized data for this type of solution, studies with analyzes of isolated modules and initial tests of the complete system have pointed out optimistic indications about the potential of this tool to collaborate in a practical and intuitive way with students and teachers of elementary schools and also for those interested in using R-CASTLE in other tasks of Human-Robot Interaction.

Keywords: Artificial Intelligence for Education, Human-Robot Interaction, Adaptive systems, Social Robotics.

CONTENTS

CHAPTER

1

INTRODUCTION

Artificial Intelligence (AI) has been playing a pivotal role in human history in the last few decades, mainly in the tasks that take advantage of automation or prediction based on previous data. The algorithms are helping people choosing movies to watch, products to buy, routes to take and even how to invest their money. Recommendation systems are a tendency since humans are achieving an amount of data big enough to identify their interests and predict their profiles. Along with the high demand for working with data in our favor, algorithms with the abilities to help manage and understand this data is crucial, once the data is an important factor in AI algorithms efficiency.

In addition, intuitive tools for data handling and analysis may offer significant contributions to people who are unfamiliar with this technology and as a result, to society in general. For example, people are not used to dealing with raw data in terminals, but they are quite comfortable using graphical interfaces to handle their data in a sheet form.

AI is constantly pointed out as the next step in the evolution of daily tasks, just as the personal computers did last century, and their contributions are countless (METROPOLIS, 2014). Computers' popularization modified a big part of jobs and had a direct or indirect impact on almost all of them. Among their principal contributions, computers allowed the emergence of Adaptive Systems. Adaptive Systems are the systems capable of storing information about the users and using it to change their own configuration to accomplish enhancement in the users' experience with the system (MCTEAR, 1993).

Nowadays, user-adaptive technology can easily be found on several devices. Systems such as smartphones and tablets, personal computers and even cable TV systems have been adopting user-adaptive methods to enhance their user experience. Cloud-based services such as Google Assistant, Apple's Siri or Amazon's Alexa, learn from their users' actions in order to improve their future interactions.

Studies regarding adaptive systems have shown an increasing - or a slower decrease - in their users' enjoyment, engagement, motivation and intention of performing the activity for a longer period of time in several tasks, in comparison to regular systems (MARTINS; SANTOS; DIAS, 2018). Adaptive parameters may assume any value that can be changed by the pursuit of the system for users' customization. Thus, their application is widespread in the rehabilitation area, in which these programs provide faster recovery time for the users.

Coming to the educational context, computers are supporting education mostly in the context of Distance Education (GUNAWARDENA; MCISAAC, 2013), in which the technology aims to break the barriers of physical presence between tutor and student. However, with regard to the traditional teaching methods of physical interaction,in most cases, technology is being used for two purposes: as a novelty to catch the students' attention and as a powerful tool to enhance the mechanical tasks done by the teachers. Results are always promising despite having a short expiration date because the students - mainly the younger ones - are expected to quickly adapt to new trends and methods, making this strategy lose the novelty as a striking factor (NAKANO *et al.*, 2009). A widely explored solution is to use Adaptive Systems to maintain the students' motivation by challenging them at a more suitable level of difficulty, breaking the boredom of a linear or monotonous difficulty (WILSON; SCOTT, 2017).

A special type of Adaptive Systems applied to education is the Intelligent Tutoring Systems (ITS) (CLANCEY, 1984). The goal of such systems is to gradually change their content's difficulty or operational parameters aiming to offer the students a suitable challenge to their knowledge. ITS present a wide application in different educational scenarios since the web-based services (BRUSILOVSKY; SCHWARZ; WEBER, 1996) to high-tech social robots for teaching (RAMACHANDRAN; SCASSELLATI, 2014). ITS studies claimed enhancements in the students' learning and engagement for many disciplines, therefore it's a powerful method that is accessible through technology.

In summary, the outcomes of using ITS are positives. They have been improving students' learning skills, providing increased attention span in grammar exercises (GHALI *et al.*, 2018), motivation in mathematical calculations (MELIS; SIEKMANN, 2004), the number of right answers and their knowledge absorption in general activities for a long time now(FUTTERSACK; LABAT, 1992).

Education is one of the biggest application fields of the adaptive systems, mainly in the hypermedia category (BRUSILOVSKY, 2000). However, adaptive systems are not the main technology spread in schools. One of the critical factors consistently being pointed out in literature about the adaptive system's unpopularity is the teachers' lack of preparation to deal with the innovation of technology (PIANFETTI, 2001; JUDSON, 2006; SHARP, 2008; NIAJ, 2019; JAKE *et al.*, 2020).

Although smartphones, tablets and computers are well received by society in this century, they faced the same lack of public acceptance in the past that AI is currently facing; and now they are not considered to be a novelty or something unknown (STRUCK, 2020). Thus, children from this century are inclined to learn new technologies faster than the older generation (KNEZEK; CHRISTENSEN; TYLER-WOOD, 2011). On the other hand, robots do not play a common role in peoples' daily lives. Nonetheless, enjoyable experiences with these devices can lead children to understand and reflect on issues that technology is currently creating in society in a more concrete way than solely theoretical classes. They could also potentially inspire them to take on careers in innovational technology (TYLER-WOOD; KNEZEK; CHRISTENSEN, 2010).

Robots emerged as mechanical devices to support repetitive physical tasks and progressively are taking place in intellectual and human-like tasks that require social and communicative skills. The robot that has the ability to perform these abilities is The Social Robot.

Social Robotics is the science of having robots exchange information with other robots— or humans— to achieve the goal of a specific task (BREAZEAL; DAUTENHAHN; KANDA, 2016). More specifically, the Human-Robot Interaction (HRI) is the Human-Robot Interaction (HRI) is the division that studies the information exchange between one or more robots with one or more humans (GOODRICH; SCHULTZ *et al.*, 2008). HRI Robots have been showing themselves useful in enhancing users experience. Another sensitive contribution of social robots is the provisioning of a lack of professionals in some sectors that have low numbers of human professionals. E.g. the medical monitoring of patients with highly contagious diseases (KRAFT, 2016), as companions for the elderly (BROEKENS *et al.*, 2009) and for young students that require personal tutoring for their specific needs. In general, the users of such studies present better outcomes when they are under social robots supporting compared to the control group.

The robots are searching for connection for empathy and customization with the user through shapes and behaviors (LI; RAU; LI, 2010), similarly to software and apps (TAPUS; MATARIC, 2008).

Their design is trying to replicate the shapes and functions of the human body in order to establish familiarity and connection with the user from the first interaction (LI; RAU; LI, 2010).

The features aforementioned show promising results in education due to the fact that the depth of learning is— in most cases — strongly related to the students' perception and trust in their tutor.

This strategy initially boosts the students' curiosity and motivation due to being a novelty (LEITE; MARTINHO; PAIVA, 2013; GLEASON; GREENHOW, 2017). However, eventually the students get used to deal with them and consequently get bored (HEERINK *et al.*, 2016). Thus the novelty factor becomes invalid. As aforesaid, adaptive systems aim to dissolve the monotony and are being used in social robots as well. Adaptive robots in education have been widely researched in the last decades with the goal of keeping the motivation, engagement and enjoyment of the students for a longer period of time. Additional outcomes from experiments with the adaptive solutions show that robots displaying personality were more effective at influencing participants to change their decisions than when the robots are displaying a smaller range of personality traits. When a robotic tutor personalizes and adaptively scaffolds self-regulated learning (SRL) behaviour, there is a greater indication of SRL behaviour and increased learning compared to control conditions where the robotic tutor does not provide SRL scaffolding (JONES; BULL; CASTELLANO, 2018). However, they are always focused on one subject and usually do not involve the teachers in their solution.

Machine Learning algorithms are crucial for the autonomous communication of the robot with the environment and users. When using supervised methods of AI (KOTSIANTIS; ZAHARAKIS; PINTELAS, 2007), they need a sufficient background of samples already known to classify the new outcomes. The success of autonomous decision-making in social robots depends on how well they can exchanging information with the environment and with the users. For instance, wrong speech recognition of the user's answer will probably lead to a wrong robot response. Similarly, mistakes in the image classification tend to generate wrong decisions in the next step of interaction by the system. Therefore, the perfect harmony between the predictions algorithms with the decision-making methods is one of the biggest current challenges of Social Robotics (BELPAEME *et al.*, 2018). Adding sensibility to these robots lies in the ML algorithms efficiency and the more values observed and assessed to achieve humans-like social skills, the more the background data is required to training (MARTINS; SANTOS; DIAS, 2018).

Finally, providing powerful and intuitive tools to the teachers is as well important as delivering an innovative, efficient and perceptive methodology to the students (KOLIKANT; MARTINOVIC; MILNER-BOLOTIN, 2020). For instance, in a study where the teachers controlled a social robot via telepresence, elementary school students showed improvements on standardized tests after the lesson. This means that the teleoperate robot system could contribute effectively to educational systems, particularly in english education (YUN; KIM; CHOI, 2013). Moreover, teachers claim to feel more useful and confident when they can control and use new methodologies in their classes or exercises (ROSSI; FEDELI, 2015), especially because to automate mechanical tasks of planning and evaluating their activities may save 3 hours of their weekly workload (JAKE *et al.*, 2020).

Taking into account the analyses presented so far, it was hypothesized that, by joining the advantages of Adaptive System, Social Robotics and accessible AI would result in a centralized platform capable of achieving an enhancement in educational processes for teachers and students, unlikely to be achieved by these approaches separately.

The R-CASTLE system that is proposed in this thesis, stands for Robotic-Cognitive Adaptive System for Teaching and Learning. It has been developed in the last six years based on studies and methods of the Robot Learning Laboratory, in ICMC, USP - São Carlos. As the name suggests, it aims to offer an autonomous adaptation of system behavior and content difficulty, focused on the learning process, from designing the activities to their evaluation. Multiple open-source algorithms of image classification and verbal recognition are at the designer's disposal to be used in the interactions, the student's database, for general knowledge and the previously performed activities and corresponding evaluations. The Python 2.7 was used in all the algorithm due to the compatibility with NAOqi [1], the operational system for Softbank Aldebaran robots. Furthermore, algorithms of Multilayer perceptron (MLP), Support Vector Machine (SVM), K-Nearest Neighbors (KNN) and an ensemble of them were studied for image classification. The Haar Cascade method from OpenCV 3 library[2] was used for face gaze detection. Hierarchical convolutional neural networks (CNN) were compared to classify face emotion expression. Ageitgey [3] face recognition was chosen for user recognition. Google Speech Recognition was considered for verbal understanding while Embeddings and Edit Distance were used for sentence matching. A Rule-Based and a Fuzzy Decision-Making Algorithm were proposed for robot behavior and content difficulty adaptation. For the purpose of this thesis, the system's interface communication with students is desired to be a robot, but it could be a computer or any other interactive device (given small hard-code changes in the output configuration). All of these features are encapsulated by a Graphical User Interface (GUI), developed with PyQT 4.8[4], that makes it easy to configure and use it in interactive educational activities. With that, the lack of preparation of the teachers in dealing with intelligent algorithm programming— which has been emphasized as one of the main limitations of making the IA popular in education— is expected to be overcome (JOHAL *et al.*, 2018).

The R-CASTLE's initial contributions were only focused on the improvement of students' skills improvements through interactive robots. However, a stronger potential in supporting teachers was noticed and, later, the obstacles faced by them, started to be also considered in the architecture's development. Therefore, the outcomes have been pointing to a two-fold contribution to the educational area: enhancing the content management for the teachers with AI techniques and, as an outcome, the students' learning experience. A video about R-CASTLE's main features can be seen at <https://www.youtube.com/watch?v=GlNj98L1Mrc>.

In summary, the technical proposal of this thesis aims to deliver a system with the aforementioned features, while its application aims to support scientific studies regarding the user experience with the system, as described below.

1.1 Motivation

The reasons that motivated this research are:

- **Addressing the issues of AI on education:**

  The teacher's lack of familiarization in dealing with intelligent algorithms is pointed out as one of the most critical barriers of applying AI in education.

[1] <http://doc.aldebaran.com/2-5/dev/programming_index.html>
[2] <https://opencv.org/opencv-3-0/>
[3] <https://github.com/ageitgey/face_recognition>
[4] <https://pypi.org/project/PyQt4/>

- **Customized and personalized learning to the students:**

  Two of the major challenges in the current education is to meet the individual needs of each students during the school time and hold their attention span for a longer period of time.

- **Supporting teachers with high-tech methodologies:**

  A few works of Social Robotics in Education allow the teachers to program their own activity and perform it autonomously by the robot. Furthermore, the majority of the ITS use multiple-choice approach in their questions, meanwhile R-CASTLE provides open answers evaluation through NLP algorithms.

1.2 Objectives

The general objective was to evaluate which computational and social methods would better fit an architecture to support teachers and students through AI techniques in educational activities in a natural way for both of them.

Nonetheless, once the R-CASTLE users are both teachers and students, most of the time the research elements are divided by their focus on each one of these users in this thesis for understanding purposes. Furthermore, this thesis analysis is divided into three aspects: its technical approaches (e.g. the analysis of the algorithms); its impact on each user (e.g. the teachers and students perceptions regarding their experience with the system); and its impact on the educational process (e.g. how the system may affect the teaching/learning factors).

Hence, the following specific objectives were set.

1.2.1 Technical approaches

Objective I *To evaluate algorithms for automatic users' measure extraction and multimodal adaptation that take into account a teacher-friendly setup and Social Robotics issues.*

Different algorithms for adaptive multimodal interaction present different advantages and disadvantages, such as accuracy, previous data dependency, training time (if needed), execution time and parametric configuration. Especially in this last one, in most cases, users cannot make changes to the algorithms parametric configuration because they do not understand its functionalities - which is the problem with the teachers that want to use such algorithms for different scenarios but have no technical knowledge about them. Similarly, automation methods for Social Robotic matters need to take into account the trade-off in their performances. When joining these two subjects in one solution it is mandatory to search for a better outcome in the specific application.

1.2.2 Users analysis

Objective II *To present an easy way for teachers to manage their educational activities through technology.*

Teachers perform practical exercises in the student books. The switch from manual methods to technological alternatives in the teaching process — apart from structure costs— is sometimes blocked by their difficulty in dealing with technical issues. Technology will be used to plan the activities, but the execution and assessment of activities will still be done manually; with pens and papers. Thus, the whole process would benefit from a system with execution and evaluation abilities. Performing all of these functions using a GUI is an easy alternative to overcome the presented issues, once people are used to exchanging information with smart devices by pressing buttons and checking charts.

Objective III *To deliver a natural user-experience with social robots that motivates the students.*

Children are usually attracted by technology, which may be used in robots as an ally to initially catch their attention. However, it is natural that they lose interest as they continue interacting with the robot. Combining audio-visual interaction with social skills techniques is an alternative to hold their motivation in their next meetings. The communication between students and robots is verbal, and this is natural to the students once they use it to communicate with other people.

1.2.3 Impact on the educational process:

Objective IV *To increase the teaching experience by providing a scalable resource that can help teachers with personalized tutoring for their students.*

K-12 teachers are used to performing practical exercises with their students through sheets and books. They do it with the whole class and try to do their best to address every student's needs. However, it is impossible to achieve personal tutoring and feedback to the whole class when evaluation is done through checking students' answers on paper afterwards. Teachers can take advantage of their students 'customized experiences with an autonomous robot that performs the exercises and presents feedback of each students' performance to both students - through voice during the interaction - and teachers - through charts and videos of the sessions.

Objective V *Increase the learning experience through Personalized Learning and Social Robotic techniques compared with their last meetings with the robot.*

Similarly to the technique of using adaptive robot behavior to hold students' attention and motivation, techniques for adapting the content difficulty also aim to provide a suitable level of content difficulty for each student. The experience would also be enhanced with new planned content, programmed by the teacher based on the data provided by the system's autonomous feedback.

1.3 Hypothesis

The initial hypothesis was that providing a smart centralized platform equipped with new technologies to the educational agents would enhance the whole pedagogical process. First, teachers would optimize their working time on mechanical tasks of planning and evaluating by making use of the AI algorithms. Then, students would present more engagement, enjoyment and learning rate, afforded by the combination of the advantages from both customized learning and social robotics.

However, searching for evidence that supports it all at once may not present the true potential of this study, which — despite its ambition— is still introductory. Following the same line of though presented in the objectives, the hypothesis was also divided into smaller hypothesis, for better comprehension analysis, that combined would support or contest it. The derived hypotheses are presented below

Technical approaches:

H1 *Algorithms without previous data dependency for multimodal adaptation present a good trade-off between intuitive setup and performance compared to supervised algorithms.*

Users:

H2 *Students would present higher scores in the personal assessment items regarding their enjoyment and rapport building in questionnaires after performing activities with systems using personalization skills compared to their last meeting with the robot.*

Impact on Education:

H3 *R-CASTLE would give more support and motivation to the teachers than traditional methods by assisting with planning, performing and evaluations.*

H4 *Students would present increasing scores in the items regarding learning statements in questionnaires after performing activities with the systems using adaptive behaviors compared to their last meeting with the robot.*

The evaluation methods and the corresponding observed variable for each hypothesis are described next.

1.4 Evaluation Methods and Observed Variables

Considering the number of observed variables in this study, this section will briefly describe the applied evaluation methods and point out their location throughout this thesis.

Algorithm evaluations were achieved in ideal conditions (laboratory) and real environments. In summary, they consisted of obtaining algorithms measures about a specific configuration and comparing them to another configuration or another algorithm.

Users' experiences with R-CASTLE were evaluated through a questionnaire of Likert Scale. Teachers answered the surveys before and after the interventions while students answered their respective survey only after the interventions.

The experiments with users were approved under the register number 72203717.9.0000.5561 of the Brazilian Ethics Board[5]. Therefore, it is worthy to highlight that every user, teacher or student, that participated in the experiments of this thesis had signed a corresponding consent form, in order to fulfill the Brazilian Ethics Board requirements. All the consent forms and evaluation questionnaires (obviously all in Portuguese-PT) are presented at the end of this thesis, in the Attachments section and in the Appendix section (A), respectively.

The observed variables to analyze the research proposals are presented in the next Subsection. Final discussions about the objective and hypotheses are presented in Section 7.1 and a summary of them is shown in Table 1.

1.4.1 Technical Evaluations

Regarding the technical objective, **Objective I**, the measures of accuracy, training time, parameter configuration, execution time and application suitability were evaluated in laboratory conditions and its results are presented in the experiments of Chapter 3. Algorithm analyses in a school environment are presented in the "Experiment" section of Chapter 5. The metric observed was the system accuracy in adapting the difficulty of the questions based on human validation.

Similarly, the **Hypothesis H1**, regarding the comparison between adaptive methods, was evaluated by analyzing their F-Measure (or F-1), the intuitiveness of their modeling and their training time. Chapter 6 is exclusively dedicated to this discussion.

1.4.2 Users Perceptions Evaluations

The evaluation of the **Objective II**, with regard to providing intuitive interactions to the teachers, consisted of analyzing their answer to the questionnaire in the items regarding the time to get used to the system, their perception of how easily they can program the content with the R-CASTLE and also their response in the questionnaire's open questions. The results of this evaluation are presented in the survey in Chapter 4. They are also mentioned in Chapter 5 in

[5] <http://plataformabrasil.saude.gov.br/>

which two teachers that participated in the presentation had two interactions with the system for programming and evaluating their contents and answered questionnaire again.

In the same way, the evaluation of **Objective III**, related to the user experience of the students with social robots, was accomplished by observing the items in the questionnaire related to their personal experience. The observed item was regarding the user's declaration about their personal enjoyment and the building of rapport with the robot.

The same metrics were analyzed to evaluate the **Hypothesis H2**, regarding user personalization in the interactions. Discussions take place in Chapter 5.

1.4.3 Educational Impact Evaluations

The assessment of the **Objective IV**, about providing a scalable resource to teachers, was based on the observations in the questionnaire items regarding their perception about how they would feel getting help by using R-CASTLE and the potential they see in its applications.

The **Hypothesis H3** that R-CASTLE would enhance teachers' motivation, workload and support was evaluated through the teachers' answers to the same items as to **Objective IV**.

At first, the enhancement in the preparation and evaluation time and workload of the teachers was expected to be measure and compare. However, the participant teachers said it would be hard to measure with few interventions. The discussion of such analysis is provided in Chapters 4 and 5 as well.

Finally, the evaluation of **Objective V**, related to the students' learning experience, was accomplished by observing the items on the questionnaire with regard to their experience as well. Observed items were regarding the users' declaration about their feeling of learning, perception about the robot's intelligence and difficulty in performing the activity. The same metrics were analyzed to evaluate **Hypothesis H4** regarding the users that may increase scores throughout their meetings with the robot.

Although this social system implementation was based on several previously performed studies (shown in Section 7.3.2), only experiments with regard to the entire R-CASTLE experience are presented in this thesis. The discussion takes place in Chapter 5.

1.5 Outline

The remainder of this thesis is organized as follows.

In Chapter 2, the issues and barriers for the popularization of AI in education, pointed out by the literature through surveys and experiments, are presented. Questions are raised regarding the challenges of efficiently fighting these problems with existing methodologies. The manuscript presented in this chapter presents the R-CASTLE scope in its higher level of the interaction between teachers, students and system, and how its design decisions aim to solve each of these questions.

In Chapter 3, a paper regarding the intelligent algorithms used to autonomously read the environment is presented. This article shows the proposed approach to assess the users' verbal answers and body language to achieve robot behavior adaptation, as well as the addressed AI methods to do so. It is the first presentation of the proposed model of Adaptive System which analyzes groups of users' skills combined in a Ruled-Based decision-making algorithm, arising from previous knowledge. The contribution of this paper is the technical deliberations to achieve autonomous natural communication between students and R-CASTLE and their tests in laboratory environments, which is considered an ideal condition.

In Chapter 4, the computational modeling of the Graphical User Interface (GUI) for the teachers is presented. The GUI is an intermediate layer that guarantees accessibility to R-CASTLE modules for teachers in a more comprehensive way than operation over technical parameters. This paper also brings the teachers' first perceptions about the efficiency of the proposed solution after they receive a presentation of the system. It's important to note that this paper was published in a special missed of the awarded papers of the Brazilian Symposium of Informatics on Education (Simpósio Brasileiro de Informática na Educação - CBIE).

In Chapter 5, a study regarding the complete workflow of the R-CASTLE is presented. This workflow started from the teachers planing and inserting the content into the R-CASTLE, passing to the students interaction with the autonomous robot performing the activities, until the feedback of the activity to the teacher was generated by the system. The final outcomes are analyses of both teachers' and students' perceptions about the system and their experience with teaching and learning with R-CASTLE. The paper also discusses the challenges of the classification algorithms in real-word conditions, once the tests were made in the school environment.

In chapter 6, the evaluation of a fuzzy decision-making algorithm is presented to adapt the content's difficult. The manuscript presented in it shows how the R-CASTLE's Adaptation module works to perform changes in the content's difficulty. The fuzzy decision-making system was compared to a simple rule-based system - presented previously - as a baseline. A dataset was created for evaluation of the proposed system, in which the measures were autonomously extracted from videos regarding student-robot interactions, during practical exercises prepared by their teacher. Beyond providing more intuitive modeling by using linguistic terms, the proposed fuzzy system presented similar performance of classification regarding the baseline method and some supervised machine learning methods, having the advantage of do not depend on previous data.

Finally, in Chapter 7, concluding remarks about this thesis as a whole, another scientific contributions of this study, and future works for improvements are presented.

CHAPTER

2

FACING THE CHALLENGES OF ARTIFICIAL INTELLIGENCE IN EDUCATION THROUGH ADAPTIVE SYSTEMS AND SOCIAL ROBOTICS

Manuscript submitted to Elsevier Journal Computers & Education.

Contribution statement

D.C. Tozadore and R. A. F. Romero conceived R-CASTLE proposal. D.C. Tozadore and R. A. F. Romero performed the literature review of Section 2. D.C. Tozadore proposed the components and the communication between them, resulted in the scheme of Figure 1. Both authors contributed on writing the manuscript. R. A. F. Romero, in addition, supervised all this research work.

Presentation

The manuscript presented in this chapter is regarding the whole R-CASTLE scope in a high-level view. Its discussion takes into account the advantages of the modules focused on teachers' and students' usage of them. It also justifies the needs of each planned part by discussing the challenges presented in literature of AI in Education and the proposed solutions that make R-CASTLE viable. Thus, it was chosen to be presented at first, in order to provide an overview. It also present the works that make part of the R-CASTLE but are not detailed throughout the remainder of the thesis. Finally, the experiments discussed herein have grounded R-CASTLE proposal implementations. They are better detailed in the cited papers throughout this chapter. An investigation regarding the terminology "User Model" or "User profile" that better fits this proposal is under course. This term is first commented in this manuscript and shown in the following chapter as well. However, this proposal follows the definition presented in (MARTINS; SANTOS; DIAS, 2018), in which a User Model in an adaptive system is a module responsible for dealing with all matters regarding the users.

Facing the challenges of Artificial Intelligence in Education through Adaptive Systems and Social Robotics ⋆

Msc Daniel C. Tozadore[a,*] (Researcher), Professor Roseli A. F. Romero[a,1] (Researcher)

[a] ICMC - Avenida Trabalhador São Sancarlense, 400

ARTICLE INFO

Keywords:
AI for Learning
Human Robot Interaction
Social Robots
Technology in Education
Adaptive Systems

ABSTRACT

Investments in AI for education had significant growth in the last few years. They also show optimist projections for the future, hitting a prediction of trillions in the next decade. This fact leads to high demands for developing new strategies for smart educational environments. Dealing with AI in its technical level in hard and fuzzy. However, using software windows to manage data and content is a method that people are used to. This strategy is an alternative to overcome the teachers' lack of technical knowledge of new technologies. In the same way, interacting with robots in human-like manners - verbal and visual communication - could make the process of adapting to new methodologies more natural to young students. Herein, the main challenges of AI in Education are analyzed from the literature and previous studies. Each one of these issues is proposed to be addressed together in an only robotic architecture, named R-CASTLE. The communication between teacher and R-CASTLE is made through a Graphical User Interface, whereas the system communicates to the students through a social robot. However, the architecture has the advantage of using other interactive devices, such as notebooks, tablets and smartphones, as the output interface with students. The main contribution of R-CASTLE is to provide AI methods alongside the advantages of Social Robotics and make them easily available to support both teachers and students in the learning process. Results showed that teachers and students noticed the potential of the system. More studies need to be done in order to properly explore all this system's advantages.

1. Introduction

Artificial Intelligence (AI) has been playing a pivotal role in human history in the last decades. Mainly in the tasks that take advantage of automation or prediction based on previous data. AI algorithms are helping people choosing movies to watch, products to buy, routes to take and even how to invest their money. However, one of the biggest challenges that the AI area faces is that most of these people do not know how AI and technology can optimize their own work and how to use them as their ally and, sometimes, this fact may even make them be afraid of such innovation.

The truth is that AI is been changing the job market because automation can support several processes that humans need to handle every day. Thus, the automation contribution goes from the physical and repetitive effort to rational and creative assignments. Additionally, a modification in the workers' formation is also required in order to prepare them to use it in their favor. The workers' formation is been raised as one of the crucial points to AI dissemination and it is reaching the majority of this century professions. The earlier people understand how to aggregate AI in their work, the earlier these results would appear in their carrier. This is also important to create an awareness that AI is no longer an exclusivity of the STEM fields but is now present in many other careers (Tomšik et al., 2016).

Smartphones, tablets, and computers had faced this phenomenon before and now they are not considered a novelty anymore. Thus, children from this century are inclined to learn new technologies due to this fact (Davis, 2010). On the other hand, enjoyable educational experiences with robots can lead children to think about the issues that technology is rising to the current society in a more concrete way than only theoretical classes - and maybe even influence them to follow innovation technological careers Rosanda and Istenic Starcic (2020); Zhao (2006).

The necessity of collecting data from the users and suggesting future recommendations to fill the users' expectations has been noticed by the Education. Once the investments in AI for education had significant growth, hitting a

⋆ This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq) and the São Paulo Research Foundation (FAPESP).

*Corresponding author

tozadore@usp.br (D.C. Tozadore); rafrance@icmc.usp.br (R.A.F. Romero)

www.danieltozadore.com,dtozadore@gmail.com (D.C. Tozadore); www.icmc.usp.br/~rafrance/ (R.A.F. Romero)

ORCID(s): 0000-0003-0744-0132 (D.C. Tozadore); 0000-0001-9366-2780 (R.A.F. Romero)

prediction of $ 15.7 trillion in the next decade (Kipouros, 2018), there is a higher demand for developing new educational methodologies and for investing in the formation of the education staff. Furthermore, it is always worthy to remind that investing in education professionals is even more powerful than only investing in last generation equipment. Conversely to the children surrounded by technological devices, their teachers are from a previous generation that is not that familiarized with high-tech innovation. It means, their generation has more difficulty getting used to these technological devices, although they know this evolution is necessary.

Based on a brief literature review presented in this paper, we raised the following questions regarding the barriers of AI in Education: (1) how AI techniques can be available as easy and accessible resources for education; (2) how the data derived from such techniques may be processed and used intuitively by the teachers; (3) how social robots could deal with adaptation and personalization educational activities; and finally, (4) how new technologies can provide fair-cost scalability and support extension for other HRI applications.

In this work, an interactive platform is designed to attend the needs pointed out. The system is named R-CASTLE, which stands for Robotic-Cognitive Adaptive System for Teaching and Learning. It is designed for supporting both teachers and students through AI techniques, combined with the advantages of Social Robotics. This kind of robots can take more attention when get to interact with humans. By calling the system cognitive, we understand that it is using the acquired information (several processed data) as long as short memories to learn how to provide personalization to the students Holland and Reitman (1978), and this knowledge can be handled by the teachers anytime in the GUI.

From the best of our knowledge, R-CASTLE is the first system to attend the needs of both teachers and students considering social robots in the learning process. Some modules of R-CASTLE were published in previous works. However, in the present paper it is presented the complete architecture in its higher layer of functionalities, taking into account their main users: the teachers and the students.

As the name suggests, the system aims to offer adaptation focused on the learning process as in relation to the content difficulty, as in its behavior itself. This adaptation is achieved in two ways: by autonomous or human evaluations. Thanks to R-CASTLE, multiples open-source algorithms of image and verbal recognition are at the disposal of the teachers to be used in the interactions. Later on the sessions, the generated data is available to be accessed at anytime. The information is regarding the students, the general knowledge, the performed activities and their corresponding evaluations are also available.

The communication of R-CASTLE with the students is established by a robot, in this case, the humanoid NAO, from Softbank. However, it could also be a computer or any other interactive device. All of these features are encapsulated by a Graphical User Interface (GUI) making the system easy to be configured and used in interactive educational activities.

It is worthy to take note that not only teachers, but other types of users can user the R-CASTLE to design their HRI activities. Examples of other profiles that used R-CASTLE to run activities are graduation and under-graduation researchers, as shown in Section 5.3. Therefore, by saying *Teacher* in this paper it is also meaning other persons that may use this system as the role of activity designer, whereas the word *Student* will be referred to those people that will use the system interacting with the robot.

This article is organized as it follows. In Section 2, a brief panorama about the challenges faced by technology and AI in education as well as possible solutions are discussed. The following sections are highlighted how Machine Learning techniques have been incorporated into R-CASTLE. In section 4.1, the architecture of the system R-CASTLE is presented. In section 4, it is explained what are the characteristics R-CASTLE offers for the designers. In section 5, it is explained what are the characteristics R-CASTLE offers for the students, highlighting the presence of a robot and evaluation module. Finally, in the section 6, in which the initial questions are answered jointly with the perspectives of future works.

2. Artificial Intelligence in Education

It is evident that the agents that are been affected the most by educational issues are the teachers and the students. Most of the actual software systems for education choose to address isolated problems due to the adversity dealing with lots of problems at once. For instance, softwares for managing teachers' content approaching are usually different from softwares to perform exercises with the students. On the other hand, systems capable of affording solutions for both teachers and students at the same time present the advantage of keeping an one and only flow throughout the entire educational process. For instance, the ones that provide autonomous customized changes for each student according to one specific need may optimize the learning experience and save teachers' time. Nonetheless, few of them take into account advantages of social robotics area, which includes all the problems faced by regular educational solutions plus the issues by the robotic area, such as the devices management Rosanda and Istenic Starcic (2020); Zhao (2006).

Furthermore, according to the teachers, some crucial factors hampering the adoption of AI in their activities are lack of time, lack of training and unfamiliarity with new technologies (Kipouros, 2018). For instance, a survey with more than 2 thousand k-12 teachers from 4 countries (Canada, Singapore, United Kingdom and the United States) by Mckinsey & Company, reported some critical problems of the current educational system that can be smoothed with efficient technological solutions Jake, Heitz, Sanghvi and Wagle (2020). Among the major findings, the study claims that areas with the biggest potential for automation are the preparation of activities, administration, evaluation, and feedback. Conversely, actual instruction, engagement, coaching, and advising are more immune to automation. The automation has great potential to save teacher´s time in repetitive tasks and use this time in more tasks where teachers are directly engaging with students, such as behavioral-, social-, and emotional-skill development.

Advances in natural-language processing make it possible for computers to assess and give detailed, formative feedback across long-form answers in all subject areas. For example, writing software can look at trends in writing across multiple essays to provide targeted student feedback that teachers can review and tailor. Combined, these technologies could save three of the current six hours a week that teachers spend on evaluation and feedback.

On the other hand, opposite to the teachers' unfamiliarity with new technologies, students are comfortable and well adapted to them. However, a controversial point is they lack methodologies provided by their own teachers. In other words, traditional methods using pens, paper sheets and books are increasingly tending to cause boredom to the young students. More than helping in theoretical exercises and their corrections, intelligent systems have been looking for modeling each student profile and adapting to their necessities in order to optimize their engagement and performance (Martins, Santos and Dias, 2018). This approach is named Personalized Learning (PL). Although PL does not have one and only correct definition, it is globally understood as "the understanding of the needs and goals of each individual student and the tailoring of instruction to address those needs and goals". Considering that a class has more than 20 students for 1 teacher, on average, it is not hard to realize how much effort from this teacher it would be necessary to provide PL for each one of these students. Moreover, alongside the adaptive and personalized learning solution, the challenge arises of providing an appropriate assessment of the results of these methodologies.

Machine Learning algorithms achieved an average accuracy of 85% in correcting exams based on human correction Nafea (2018). Most of them are more used for website content and software-based activities and they are focused on multiple choices and unique answers. Hence, one of the challenges yet to overcome is presenting tools to evaluate subjective answers for better supporting learning personalization. Once people prefer interactive systems in which they can talk with (Manning, 2018), the Natural-Language Processing (NLP) contribution is notable. Systems that hold good responses to long and robust verbal interactions with humans are desirable in such modelings as well as challenging (Mwangi, Barakova, Díaz-Boladeras, Mallofré and Rauterberg, 2018).

Although conventional technologies for education offer alternative ways for optimizing the learning process, they are limited to the virtual world. Social robots have the advantage of offering a physical and more sensitive experience in the interactions. Studies in Human-Robot Interaction (HRI) regarding physically embodied robots showed that they are capable of providing more narrative attention to the listener as compared to a virtual agents (Kano and Morita, 2019).

According to Baxter, Ashurst, Kennedy, Senft, Lemaignan and Belpaeme (2015), there are typically two main goals, and often overlapping, for social robots in classrooms. Firstly, and already discussed here, they are intended to offer teaching structures and supplementary support to children by providing an alternative and/or personalized learning experience. Secondly, they seek to examine the attitudes of the teachers and students regarding robots in the classroom and solicit their views on how applications should be implemented and used.

The teachers' engagement is essential to the development of social robots in education. They are the ones who are more sensitive to practical and ethical concerns and they know how to approach it with the students (Jones and Castellano, 2018a). Concerns about bringing the teachers as a focal point of HRI studies are recent, compared to studies focused on the students.

The social robotic approaching also takes into account the multimodal analysis of the users. It means, the search for autonomous understanding and analyzing the users information exchange. This analysis is achieved by classifying the voluntary and involuntary users responses to the interaction, such as their verbal answers and their facial emotion expression, respectively. Datasets of image and speech classification are easily found alone. However, specific datasets of these inputs combined for multimodal analysis and validation are not easy to find. Therefore, multimodal systems validation is constantly hard to achieve without data generated by those systems themselves.

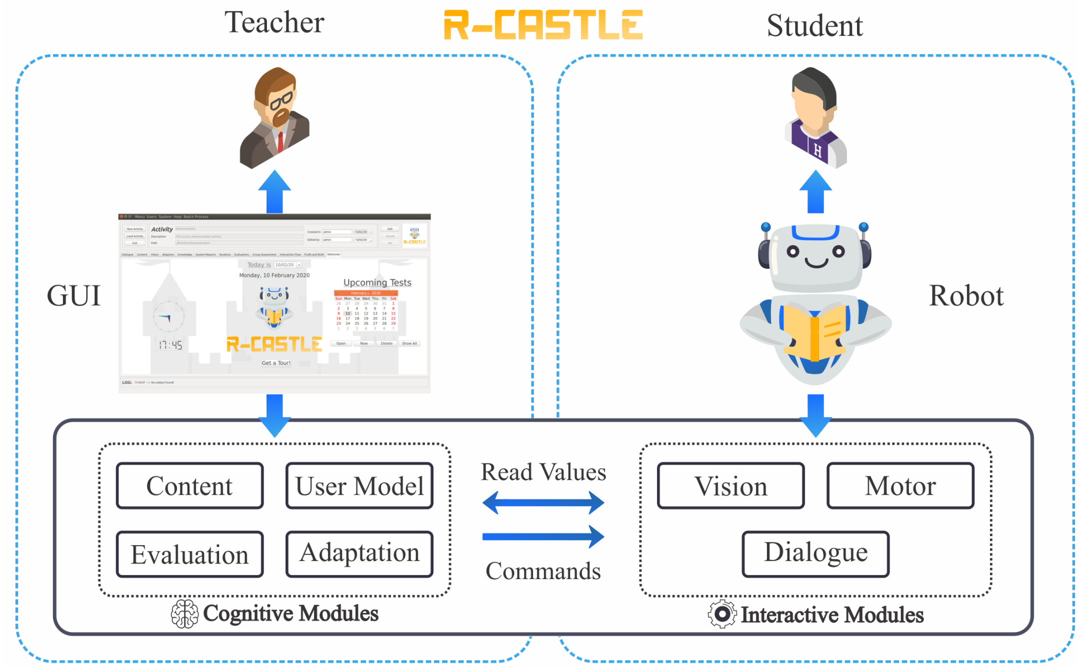

Figure 1: Architecture of R-CASTLE

3. R-CASTLE overwiew

In this section, the R-CASTLE's goal and the interaction flow of its components is presented.

Although is hard dealing with AI at its technical level, managing data and interacting with devices through software windows are more intuitive and an alternative path to overcome the teachers' lack of technical knowledge and training. Similarly, interacting with robots in human-like manners makes more natural to the students the process of adapting to new methodologies. In R-CASTLE, both of these advantages are placed together in a robotic architecture, which is capable to be used with other interactive devices as well. The Figure 1 shows the scheme of R-CASTLE and its proposed communication flow to teachers and students through social robots and AI techniques.

The communication of the teacher and the R-CASTLE is made through the Graphical User Interface whereas the R-CASTLE communicates to the students through an interactive device, usually a social robot (that is why the "R" of "Robot" in the name).

There are two groups of modules in R-CASTLE. One is the Interactive group and the other is the Cognitive. This division is only made for scope matters. The Cognitive group encloses the modules that process the outputs provided by the modules of the Interactive group, with the result of changing them into useful information to store and make decisions based on them. They are also used to store information manually inserted by the designers or autonomously acquired during the activities. They are the *Content*, the *Adaptation*, the *Student Model* and the *Evaluation* modules. The Interactive group are the modules responsible for controlling the resources of the robot and the methods to exchange information with the environment. These modules are the *Dialogue*, the*Vision* and the *Motor*. The Dialogue and the Vision modules have a fundamental role in this proposal as described in the next sections. The Motor module only controls the robot's moves, which in some cases can be dismissed. [1]

The file-based managing of data allows the information exchange between computers that use the R-CASTLE. It allows the produced data in one computer to be used in the others. Thus, training and improvement of the results produced in one computer is available in another one. E.g., exchanging evaluation and vision classification models and data to retrain the algorithms in other computers or in a server. Similarly, exchanging the created content between activities or computers it is also possible. The R-CASTLE employment in several computers and by many teachers and students as possible is an necessary achievement in order to enhance the autonomous validation on a large scale.

[1]Translating to programming, the modules - which are particular systems by themselves - are modeled in classes with variables and methods to fill the goal of the names of their corresponding modules. For instance, the Dialogue module contains all the code related to allow the robot communicates verbally with the users.

once the more specialized data the system has, the more accurate the algorithms training will be.

Experiments already performed with teachers pointed out an optimistic time to get used to the R-CASTLE's. This conclusion was stated by teachers themselves (Tozadore and Romero, 2020). Also, a high acceptance of the students to the system's performance was also reached in experiments (Tozadore, Hannauer Valentini, Rodrigues, Pazzini and Romero, 2019a).

Next, the advantages to teachers and students offered by R-CASTLE are presented.

4. R-CASTLE for Teachers

In this section, the resources available in R-CASTLE for the teachers are discussed, highlighting the AI components present on it.

In this presented proposal, teachers can use the system through the GUI and the modules they will work the most are the Cognitive modules. As shown in Figure 1, it is composed of the Content, User Model, Adaptation and Evaluation Modules. For this, we have developed an adaptation mechanism based on concepts of artificial neuron theory, assessing indicative measures of three students' cognitive skills. The content to be learned by the students can be inserted in the Content Module. Machine Learning techniques, such as Natural Language Processing (NLP) and image classification implemented in the Adaptation module have been incorporated in the Evaluation module. The Evaluation module stores an audio-visual database of all the evaluations made by the system and they are available to be analyzed and reevaluated at any time later on.

4.1. Graphical User Interface

Teachers can interact with R-CASTLE through a Graphical User Interface (GUI), which was implemented to operate over the technical implementations of the system. R-CASTLE allows to the designers, through this window, creating and managing activities as work-spaces. This is due to the fact that every activity has different setups for some specific modules, such as the Dialogue, the Content and the Vision modules. However, they can also share common data, such as the stored information of the users, definitions, evaluations and most of the things that are the same for the majority of the activities. Activities are managed in the superior part of the window. It is the fixed heading of the program. In there, the designers can create a new activity, load and edit existing ones. Modules are available for configuration in their corresponding tabs in the middle section of the window. Some animations and movements of the robot are available to be handled in the *Interaction flow* tab, in which the session sequence of interactions (contents, chats, dance, games) is set. It is believed that the features of this interface constitute a potential solution to the question (2) with regard to the teachers' access to AI techniques intuitively. The following subsections show implementation decisions to manage the modules through the GUI.

4.2. Content

The GUI makes the content insertion easy for teachers. The content is stored in the system database and can be approached in any late activity. The content is approached in Topics. Each Topic is an entry in the Content module which has a concept about it (an explanation of the topic from which the questions will be derived) and a lot of questions regarding it. For instance, if the activity is about animals, one can create a topic for each class, such as mammals, fishes, insects and so on, and their concept would be the features that make them belong to this class, followed by their corresponding questions. An activity can have as many topics as needed and the same is valid for how many questions each topic may contain. The teacher should fill the concept text box, which is the utterance the robot speaks before starting the questions regarding the current topic. After that, he/she needs to insert at least one question of each level of difficulty followed by the corresponding expected answer.

The pedagogic model of this proposal - it means the between interaction the student and the system conducted by the robot during the content approach - works like a quiz mode. It means, in each step, the robot gives explanations and makes questions about a topic that were inserted in the content module by the teachers.

Once the system has the capabilities of both speech and image recognition, the expected answers can be a sentence or an image. In the case of a sentence, the system analyzes the answer by the dialogue system, as shown in subsection 5.1 as well as in Tozadore, Valentini, Pazzini and Romero (2019c). On the other hand, the answers which are expected as images are classified by the Vision module, as shown in subsection 5.2 and also in Tozadore and Romero (2017).

Table 1
Reading values grouped by observed users' skills.

Attention (α)	Communication (β)	Learning (γ)
Face gaze (Fg)	Number of Words (nW)	Right/Wrong answers (RWa)
	Emotions (Em)	Time to answer (Tta)

For instance, in the question "What is the 3D geometric figure that has no vertices and edges?" the answer could be verbally accepted as "A ball" or by reading the camera's image in which the student would be showing a ball (as long as the designer set the question to be accepted in this way).

4.3. Adaptation

Results showed a potential perceived from teachers, students and other persons that used R-CASTLE so far. However, the lack of specific and complex data to validate the whole system leaves open studies to be explored.

The Adaptation module (or adaptive system) is responsible for adapting the robot behavior and the difficulty level of the approached content based on the information it receives from the vision and dialogue module. The indicative values autonomously took from the students are mapped into three groups with regard to the skills of the users considered by the system for adaptation optimization. The measured skill are *Attention* (α), *Communication* (β) and *Learning* (γ).

It is possible to use R-CASTLE to focus on specifics skills to be trained in the interactions by dividing these main skills. For example, giving null values to the weights of the Attention and Communication one may analyzing the Learning skill exclusively. This is an interesting alternative for tests of specifics skills in which the teacher does not want to concern with other skills. Isolating these measures is also an advantage for overcoming environmental noises. E.g., when an undesirable lightning negatively influences the cameras and, consequently, makes the emotion classification fails, the designer can lower the importance of this value because it is known that this would be an outlier. An extra caution by the person who will run the activities is necessary, once his/her role is to perceive such setbacks and use these measures balancing to overcome them.

There are two ways to achieve adaption based on the read values of the users in R-CASTLE. The first one is to use a Ruled-Based algorithm for an adaptive function to set the final value of adaptation. The second is to make the system to project the threshold values into Fuzzy Rules and use it to compute the adaptive function. The emotion recognition through the users' facial expressions and the face gaze counting are provided by the vision system. The read values of number of spoken words, answers' correctness and time to answer are provided by the dialogue system. Both of them consider the values read from the users for each skill group, as shown in Table 1. The teacher needs to set a limit value for each one of these observations and these limits which will be considered the max tolerance to switch to fits the users.

4.4. User Model

The User Model, or Student Model module, stores information about every student that interacted with the robot. The information is regarding their first name, family name, age, school year, birthday, user's pictures and eight users interest sport, dance, team, music, toy, hobby, game and food. All the data are stored in the user database and the definitions of these interests are stored in the system's knowledge database. Transactions in the database can be manually operated by the designer in the corresponding window. The students' interests are supposed to be used in small talks at the time that it was previously set or autonomously, when a high frequency of bad readings (a high disattention level or a high number of wrong answers) is detected by the system. More details about this module can be seen at Tozadore, Valentini, de Souza Rodrigues, Vendrameto, Zavarizz and Romero (2018b).

It is worth to take note that the Adaptation Module and Student Model are this proposal suggestion in contributing to the third question with regard to adaptation and personalization during the interactions with R-CASTLE.

4.5. Evaluation

The Evaluation Module reports to the teacher the students' performance during the interaction with R-CASTLE. For each activity, the teacher can access different graphs about the system assessment of each student or the whole class report. Another functionality is that the system can report to the designers performance of R-CASTLE in relation to the Machine Learning algorithms' accuracy based on the teachers' validation.

4.5.1. *Individual Reportings*

Although Machine Learning algorithms had achieved high accuracy in task corrections, they are still not efficient correcting subjective answers. Then, keeping a manual validation of the answers possible to be performed by the teachers is an alternative ways to pursuit these methods' improvement and a correct feedback to the students. Thus, in the GUI's Evaluations tab, it is possible to check and validate the classifications of all students' answers provided by the system from every previous session performed by R-CASTLE. These sessions may run by autonomous or Wizard-of-Oz[2] operation modes. A specific field in this window displays the session's operational mode. In general, all the sessions stored in the Evaluations database were assessed throughout the execution of activities with the parameters configured by the designer before the beginning of the sessions. A tab of "Topics Validation" allows human validation of each question grouped by topic approached. It also shows the question the robot asked

The Evaluations tab shows the session's information about the current evaluation. It displays: the user recognized by the system in this session; the configurations that the session was run; user's accuracy assessed by the system, the time the session started and ended, who was the designer, the robot or interactive device used; and extra observations that the person who executed the activities wanted to add. Another highlighted feature of the Evaluation module is the human validation of each system classification of the students' answers through the GUI. This validation is used to check the performance of R-CASTLE's algorithms in classifying the answers. After validating all the questions, graphs of both students' and algorithms performance are shown. The users' performance is related to how many correct answers they gave based on the expected ones, while the performance of R-CASTLE is regarding how many answers it correctly classified based on the human validation after the sessions.

The measures available for showing in these graphs are: the observed values from the users (shown in Table 1); as well as their corresponding skill values of *Attention* (α), *Communication* (β) and *Learning* (γ); the user's performance according to the system's evaluation and also according to human validation; the system's performance in evaluating the users answers and also pursuing the best adaptation.

Recorded videos made by the frontal camera of the robot are available to be watched later on. They are accessible in the "Time Mark" sub-tab of each evaluation.Every video has the time mark of the beginning of each question and then the one watching can jump to these specific moments. With the videos and the verbal answers stored, the sessions are available for as many reassessments as wanted in the tab "Off-Line Evaluation". In this way, all the algorithms of students´ readings and classification are accessible to be evaluated with different parameters.The result of each algorithm configuration with the new parameters, weights and tolerances are stored and available to be used in the adaptation methods at any time.

Therefore, beyond the responsibility of keeping the designers aware of the students' performance through evaluation graphics, the Evaluation module works as a laboratory to retest and maintain the machine learning methods in their high performance.

4.5.2. *Group Reportings*

All the graphic evaluation showed individually can be also done by groups, as long as they performed the same activity.

These features are accessible in the "Group Evaluation" tab of the R-CASTLE's modules' tabs. It also shows the students collective efficiency and flaws. Thus, teachers can understand and map the class difficulties and focusing to approach these adversities in their following classes. In this way, this resource works as a diagnostic tool for aiding teachers to better understand their students' weak points of learning on each subject.

R-CASTLE has methods to batch process the generated data by the evaluations and, with human validation support, find the best configuration of the adaptive function for each student or for a group of students. The best model of adaptive configuration can also be done for different subjects taking into account the student difficulty in a determined subject. e.g. whether a student performs well in grammar activities but not that well in math activities, its model of adaptation would be more severe to the first subject whereas it would be milder for the second.

All these resources and their tests with teachers (Tozadore et al., 2019a) lead to believes that the evaluation process can be optimized in both individual and collective ways. The fact that the processed data can be stored and easily shared may collaborate to the retraining of the algorithms and to get a more visual understanding of the generated data. Moreover, the visual features of the evaluations may support the planing of the next activities or may accuse specific students difficulties in the understanding of the content which may need closer and humanized attention. Once an efficient automation of the mechanical evaluation can save teachers' time, this module has a sensitive potential for

[2] A technique in which a hidden person teleoperated the robot, giving the impression that the robot has life itself.

helping IA in Education from a very popular perspective of pressing buttons and checking graphs, instead of analyzing raw data.

5. R-CASTLE for Students

In this section, the modules that allow the R-CASTLE autonomously interacts with the students R-CASTLE will be explained.

This interaction is possible due to the Interactive modules: Dialogue, Vision and Motor. The Dialogue and the Vision modules were developed using Machine Learning methods, such as artificial neural networks, clustering algorithms, deep learning and Natural Language Processing algorithms, as detailed next.

5.1. Dialogue

One of the most important communication forms of humans is the verbal one. Thus, the dialogue system has the biggest responsibility in keeping the communication flowing with efficiency throughout its interactions with the students. Speech recognition and Text-to-Speech are allowed through, respectively, python *SpeechRecognition* [3]) library and *Softbank Text-to-Speech API* [4]). If any NAO robot is available, the text to speech from Aldebaran is switch to the *Google Text-to-Speech* (gTTS) [5] library. Teachers can configure several dialogue settings through the dialogue window in the GUI, such as volume, speech speed, algorithms for matching the users' answers with the expected ones and their similarity threshold to match as a correct answer.

Robot's speeches are also allowed to be built in the dialogue tab. They could be any other information the system wants to exchange which is not related to the content. User's interests are also allowed to be used along with the utterance, e.g., talk about the favorite music of the current user. The speeches are saved in files and later chosen in which part of the interaction they will appear in the Interaction Flow tab.

It also takes into account the keywords that the designer inserts of affirmation, negation and doubt. The keywords of affirmation and negation are used in cases where the system asks a regular question to the students and expects a binary answer of yes or no. Then, all the words filled in the corresponding group will fit. Keywords of doubt are analyzed in every users' answers. If a high frequency of these words is detected, the system can start to repeat the questions, lower the speech speed, or send a message to the adaptive system to lower the difficulty of the questions.

The sentence matching algorithm (Tozadore et al., 2019c) is an important part of the system because it evaluates the correctness level of the users' answers, based on the expected answers registered by the teachers.

5.2. Vision

As a primary responsibility, the vision system manages the device's camera and recognition algorithms. Secondly, it also sends information to the modules which will process this information. Images of users' faces, for example, are sent to the User Model system whereas the information of the users' facial expressions such as emotion and face gaze are used by the Adaptation module.

For the image recognition in the tasks, some Machine Learning methods are available. They are the Multilayer Perceptron (MLP) Networks, K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Convolutional Neural Networks (CNN). The teacher can choose which one of them will be used in the next session and change their parameters before training in the algorithm settings section. More details about the advantages and setbacks in using these methods can be found in Tozadore and Romero (2017).

For user recognition, the system uses the Python Face Recognition [6] The Haar Cascade Viola and Jones (2001) algorithm, implemented in *OpenCV* library, is used for face gaze detection, as described in (Tozadore, Pinto, Valentini, Camargo, Zavarizz, Rodrigues, Vedrameto and Romero, 2019b).

Finally, the emotion recognition through facial expression is the result of a study where seven emotional states of Ekman's model, such as *happiness*, *anger*, *sadness*, *fear*, *disgust* and *surprise*, plus the *neutral* emotion, were trained to be detected by a CNN (Tozadore, Ranieri, Nardari, Guizillini and Romero, 2018a).

[3] `https://pypi.org/project/SpeechRecognition/` Accessed in 30/01/2020
[4] `http://doc.aldebaran.com/2-1/naoqi/audio/altexttospeech-tuto.html` Accessed in 30/01/2020
[5] `https://pypi.org/project/gTTS/` Accessed in 30/01/2020
[6] `https://github.com/ageitgey/face_recognition` Accessed in 20/01/2020.

5.3. Experiments with Interactive Devices

Social robots have beneficial outcomes when used as interfaces to connect the systems to students because they promote communication that is natural to us humans.

Studies have been conducted testing the architecture with other devices as output. In one of them, two groups of children were analyzed comparing a NAO robot with a tablet in storytelling activities for foreign language teaching (Tozadore, Pinto, Ranieri, Batista and Romero, 2017). By analyzing the eye gaze, the students showed more concentrated in the tablet condition due to the subtitles displaying in this device. However, they reported to felt more enjoyment in the robot condition. No significant difference was found in the learning rate between conditions.

In another experiment, two groups of graduate students were compared, with the difference of the system output as a NAO robot in a group and a low-cost domestic robot LARA (Pinto, Ranieri, Nardari, Tozadore and Romero, 2018) in the other group. The tasks performed in both group were daily tasks, such as organizing the shelf with groceries and playing Rock, Paper or Scissor, with the robots.

Results from both studies showed the same performance of the architecture in both conditions, as well as in the classifications performance and in the students' perceptions of the system's efficiency using interactive devices that are less expensive than the currently available social robots.

Thus, based on outcome of the presented experiments,it is expected to address the question (4), regarding the scalability and support extension of upcoming technologies. It means, R-CASTLE has potential to provide acceptable communication with users not only with high-tech robots but also with cheaper interactive devices.

6. Conclusion

In this article, system for helping teachers and students in the learning process was proposed. The main contribution of R-CASTLE is to provide AI methods and make them easily available to the teachers, alongside the advantages of a Social Robotics.

Therefore, after presented the whole proposal of R-CASTLE and its initial experiments alongside literature concerns, solutions for facing the following challenges of AI in Education were proposed: (1) how AI techniques can be available as easy and accessible resources for education: providing intuitive ways to teachers and students communicate with new technologies (2) how the data derived from such techniques may be processed and used intuitively by the teachers: using Graphical User Interface that encapsulates the AI techniques; (3) how social robots could deal with adaptation and personalization educational activities: using image and verbal classification algorithms to communicate with the students; and finally, (4) how new technologies can provide fair-cost scalability and support extension for other HRI applications: considering open-sources techniques and easy communication with any interactive robot or device.

Suggestions of studies to be performed or improved with the R-CASTLE as futures works are: empathy and trust relationships through dialogue mechanisms; personalization and machine learning teaching through vision system; adaptive and personalized learning to the students and time optimization to the teachers through content management and smart algorithms; supportive tools for both autonomous and manually evaluation; and demystification and ease dealing with AI techniques through the interactive graphical interface.

With open possibilities of building dialogues and changing the robot's behavior it is believed that R-CASTLE has the potential to play several active learning methodologies, such as Problem Based Learning, Design Thinking, Peer Learning and Constructivism, as tested in (Jones and Castellano, 2018b). This framework is also an opportunity for the teachers to explore their creativity while planning their activities, once they will need to search for manners to achieve the educational role they want, with the resources available in this tool.

The lack of specific data to train every algorithm and the integration of all the modules is an ongoing improvement, leaving open studies to be future explored. The more the system is used, the more specialized data will be produced and they can be used to retrain the system. By specialized data is understood as the combination of verbal answers and several visual measures taken from the students by the system, plus the personalized talks and content and robot's behavior adaptation that aimed to result in better engagement and performance in the activities. Another barrier is the overcoming of environmental noises that the algorithms need to handle for autonomous predictions regarding the students in every activity, since it aimed to be used in the schools.

References

Baxter, P., Ashurst, E., Kennedy, J., Senft, E., Lemaignan, S., Belpaeme, T., 2015. The wider supportive role of social robots in the classroom for teachers, in: 1st Int. Workshop on Educational Robotics at the Int. Conf. Social Robotics. Paris, France.

Davis, D.S., 2010. Genetic dilemmas: reproductive technology, parental choices, and children's futures. Oxford University Press.

Holland, J.H., Reitman, J.S., 1978. Cognitive systems based on adaptive algorithms, in: Pattern-directed inference systems. Elsevier, pp. 313–329.

Jake, B., Heitz, C., Sanghvi, S., Wagle, D., 2020. How artificial intelligence will impact k-12 teachers. URL: https://www.mckinsey.com/industries/social-sector/our-insights/how-artificial-intelligence-will-impact-k-12-teachers. accessed: 2020-01-30.

Jones, A., Castellano, G., 2018a. Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. International Journal of Social Robotics 10, 357–370. URL: https://doi.org/10.1007/s12369-017-0458-z, doi:10.1007/s12369-017-0458-z.

Jones, A., Castellano, G., 2018b. Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. International Journal of Social Robotics , 1–14.

Kano, Y., Morita, J., 2019. Factors influencing empathic behaviors for virtual agents: -examining about the effect of embodiment, in: Proceedings of the 7th International Conference on Human-Agent Interaction, pp. 236–238.

Kipouros, G., 2018. Ai will transform uk business - but it's not there yet. Raconteur Media Ltd.

Manning, E., 2018. Are emotionally intelligent bots the future of ai? https://www.raconteur.net/technology/emotionally-intelligent-ai-future. Accessed: 2020-01-30.

Martins, G.S., Santos, L., Dias, J., 2018. User-adaptive interaction in social robots: A survey focusing on non-physical interaction. International Journal of Social Robotics URL: https://doi.org/10.1007/s12369-018-0485-4, doi:10.1007/s12369-018-0485-4.

Mwangi, E., Barakova, E.I., Díaz-Boladeras, M., Mallofré, A.C., Rauterberg, M., 2018. Directing attention through gaze hints improves task solving in human–humanoid interaction. International Journal of Social Robotics 10, 343–355. URL: https://doi.org/10.1007/s12369-018-0473-8, doi:10.1007/s12369-018-0473-8.

Nafea, I.T., 2018. Machine learning in educational technology. Machine Learning-Advanced Techniques and Emerging Applications , 175–183.

Pinto, A.H.M., Ranieri, C.M., Nardari, G., Tozadore, D.C., Romero, R.A.F., 2018. Users' perception variance in emotional embodied robots for domestic tasks, in: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), pp. 476–482. doi:10.1109/LARS/SBR/WRE.2018.00090.

Rosanda, V., Istenic Starcic, A., 2020. The Robot in the Classroom: A Review of a Robot Role, in: Popescu, E., Hao, T., Hsu, T.C., Xie, H., Temperini, M., Chen, W. (Eds.), Emerging Technologies for Education, Springer International Publishing, Cham. pp. 347–357.

Tomšik, R., et al., 2016. Choosing teaching as a career: Importance of the type of motivation in career choices. Tem Journal 5, 396–400.

Tozadore, D., Hannauer Valentini, J.P., Rodrigues, V.H., Pazzini, J., Romero, R., 2019a. When social adaptive robots meet school environments doi:https://aisel.aisnet.org/amcis2019/cognitive_in_is/cognitive_in_is/4/.

Tozadore, D., Pinto, A.H., Valentini, J., Camargo, M., Zavarizz, R., Rodrigues, V., Vedrameto, F., Romero, R., 2019b. Project r-castle: Robotic-cognitive adaptive system for teaching and learning. IEEE Transactions on Cognitive and Developmental Systems 11, 581–589.

Tozadore, D., Ranieri, C., Nardari, G., Guizillini, V., Romero, R., 2018a. Effects of emotion grouping for recognition in human-robot interactions, in: Submited in 7th Brazilian Conference on Intelligent Systems (BRACIS).

Tozadore, D., Romero, R., 2017. Comparison of image recognition techniques for application in humanoid robots in interactive educational activities. from portuguese: Comparação de técnicas de reconhecimento de imagens para aplicação em robô humanoides em atividades interativas educacionais, in: XXII Conferência Internacional sobre Informática na Educação, Fortaleza - CE.

Tozadore, D., Valentini, João Rodrigues, V., Pazzini, J., Romero, R., 2019c. Matching sentences in semantic and syntax level for human-robot dialogues. To be published.

Tozadore, D.C., Pinto, A.H.M., Ranieri, C.M., Batista, M.R., Romero, R.A.F., 2017. Tablets and humanoid robots as engaging platforms for teaching languages, in: 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), pp. 1–6. doi:10.1109/SBR-LARS-R.2017.8215290.

Tozadore, D.C., Romero, R.A.F., 2020. Graphical user interface for educational content programming with social robots activities and how teachers may perceive it. Revista Brasileira de Informática na Educação 28, 191.

Tozadore, D.C., Valentini, J.P.H., de Souza Rodrigues, V.H., Vendrameto, F.M.L., Zavarizz, R.G., Romero, R.A.F., 2018b. Towards adaptation and personalization in task based on human-robot interaction, in: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), IEEE. pp. 383–389.

Viola, P., Jones, M., 2001. Rapid object detection using a boosted cascade of simple features, in: Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on, IEEE. pp. I–I.

Zhao, S., 2006. Humanoid social robots as a medium of communication. New Media & Society 8, 401–419. URL: https://doi.org/10.1177/1461444806061951, doi:10.1177/1461444806061951.

CHAPTER

3

PROJECT R-CASTLE: ROBOTIC-COGNITIVE ADAPTIVE SYSTEM FOR TEACHING AND LEARNING

Contribution statement

D.C. Tozadore and R. A. F. Romero conceived the proposal. D.C. Tozadore and R. A. F. Romero performed the literature review of Section 2. D.C. Tozadore proposed the components and the communication between them, resulted in the scheme of Figure 1 and the Figure 2. D.C. Tozadore and M. Camargo implemented the visual system and performed its tests, which resulted in the Figures 3, 4 and 5. D.C. Tozadore, J. P. H. Valentini implemented the methods of the dialogue system and their tests, resulting in the Tables II and III. D.C. Tozadore and V. H. Rodrigues implemented and tested the face deviation recognition, which resulted in Figure 6 and Table IV. All the authors contributed on writing the manuscript. R. A. F. Romero, in addition, supervised all this research work.

Presentation

The paper presented in this chapter is regarding the R-CASTLE first proposal. It presents the computational architecture built on algorithms for autonomous communication and the extraction of students' measures. It was the first paper in the chronological timeline, despite it was only published in the last year of this Ph.D. project due to the constant reviews of the published journal. The configurations of these algorithms were used in future studies with

R-CASTLE. Experiments showed herein consisted of testing the implemented algorithms in ideal conditions, it means, without environmental noises. An experiment regarding the students' perception regarding the system at this point can be seen in (TOZADORE *et al.*, 2017). Worthy to note that the facial emotion recognition algorithm was only implemented and added to R-CASTLE in 2018 (TOZADORE *et al.*, 2018), therefore, not be presented in this paper.

Project R-CASTLE: Robotic-Cognitive Adaptive System for Teaching and Learning

Daniel Tozadore, Adam H. M. Pinto, João Valentini, Marcos Camargo, Rodrigo Zavarizz, Victor Rodrigues, Fernando Vedrameto, and Roseli Romero

***Abstract*—Robots are already present in people's lives as receptionists, caregivers, and tutors. In human–robot interaction, social behavior is not only expected but often associated with users' confidence. Although several studies have been researching in this direction, the robot adaptation and the existing gap between the system and nonprogramming designers still need more effort to achieve success. In this article, a cognitive architecture is proposed and implemented into a humanoid robot. The aim is to offer a framework programmable for controlling the robot's resources, approaching previous knowledge, and new content in educational interactive activities. Furthermore, the system adapts the robot's behavior according to objective measures of users, attention and engagement during the activity. After the interactive sessions, these measures are provided in a graphical interface for students, skills evaluation. Functions of visual classification, speech processing, autonomous web search for new content, and attention detectors were tested and analyzed separately. This approach shows effectiveness in basic and medium condition levels from a set of sceneries for each module.**

***Index Terms*—Adaptive systems, educational robotics, human–robot interaction (HRI), robotic cognitive architecture.**

I. Introduction

In HUMAN–ROBOT interaction (HRI), it is well known that more adaptive techniques and human-like communication enhance the users' experience with robots. Mainly, in educational activities, enjoyable interactions of the students can increase some factors that are strongly related to learning. When a robot plays a role of a tutor, especially with young children, some behaviors such as good response time, deep knowledge, and accomplishment of the activities without mistakes are expected. This set of behaviors is often associated with the users' confidence in a new system [1].

For this reason, the majority of these studies are performed with the Wizard-of-Oz (WoZ) technique [2]. This technique consists of someone controlling the robot without the user's knowledge. Studies to develop robust systems that autonomously interact with humans (in this case, students) have increased recently, as well as the research to investigate their application impact in the final tasks [3].

However, the usage of the robot in HRI experiments by itself may produce contrary results from those initially expected. The success in such applications depends on a series of variables to be set up according to the users' expectations, culture, age, and the activity nature. Studies in robot adaptation have also shown significant importance in these applications.

Considering the points mentioned so far, in this article, the objectives and ongoing studies for developing a robotic cognitive architecture to be used as a framework in HRI designs are presented: the robotic-cognitive adaptive system for teaching and learning (R-CASTLE). R-CASTLE is being proposed to deliver a new educational tool that will allow any kind of designer to easily plan interactive activities with electronic devices. In this case, it is applied in a humanoid robot, proving an autonomous and natural communication with the robot and adaptive skills for short and long-term interaction.

The natural communication of the robot is guaranteed through modules that perform audio-visual processing and robot's gestures manipulation, while the robot's adaptation behavior relies on multiple sensors and algorithms to collect and analyze the users' perceptions.

All the variable configurations from this project are the results of several technical and interactive (user-centered) studies, focused on specific issues that, combined, produce the whole scenario of this research problem. Tests with all modules integrated or tests with users have not been performed in this article.

This article is organized as follows. In Section II, the background of educational HRI studies is presented. In Section III, the project's technical scheme is described. The results are discussed in Section V. Finally, in Section VI, the conclusions and future works are presented.

II. Background

Studies regarding the students–robot interaction have been performed to analyze the impact of the robot communication level in educational activities. A WoZ experiment was proposed to evaluate children's reaction with low and high interactive social responses of the robot NAO [4]. Participants were divided into two groups based on the interaction level that the robot would show along the sessions. The students' answers to the robot answers and in post-tests were analyzed with the result that the high-level interaction group presented

Manuscript received September 30, 2017; revised February 5, 2018, May 14, 2018, and November 27, 2018; accepted September 6, 2019. Date of publication September 12, 2019; date of current version December 9, 2019. This work was supported in part by CAPES, in part by CNPq, and in part by FAPESP. *(Corresponding author: Daniel Tozadore.)*

The authors are with the Department of Computer Science, the Institute of Mathematics and Computer Sciences, University of São Paulo, São Paulo 13566-590, Brazil (e-mail: dtozadore@gmail.com).

Color versions of one or more of the figures in this article are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TCDS.2019.2941079

more right answers compared to the low interaction level group.

Two weeks after this experiment, all the students met the robot again and performed a game about the content addressed in the first meet. The robot presented the same level of interaction for all the students in this experiment [5].

Once more, the students from the high interaction level group of the first session presented better performance in the game compared to those who were from the low interaction level group in the first experiment. Children from the high interaction group in the first experiment were, on average, 17% and 32% better in the game in answering easy and difficult questions. Also, they felt more challenged by their robot and 60% of them said that they had studied at home to prepare for the game.

To investigate whether there is a difference in users' acceptance in relation to the robot operation condition, an autonomous robot and a WoZ robot have been considered [6]. Another group of students were divided into two groups. In one group, the robot was teleoperated, and in the other group, the robot was autonomous. They had the same activity in both conditions and in the end of the session they were asked to fill out a questionnaire about their experience with the robot. After that, the condition that they were submitted was revealed and they were asked to fill out the questionnaire again. The results suggest that the students of the autonomous group presented more curiosity and motivation than the others, although their performance in providing the right answers in the activity was statistically the same.

The search for autonomous systems has highly increased in the last decades, followed by the concerns with this type of interaction design [7]. These studies range from specific functionalities to fully automation degrees and, in the majority of them, the findings encourage the HRI researchers to keep working in this direction [8], [9].

Considering technical solutions for the automation, cognitive systems have presented themselves as an acceptable option for more generic and complex automation design in robotics. A cognitive architecture consists of an intelligent system that benefits from memory, learning, and actions to solve generic problems. It constantly takes the environmental parameters to process and update internal states, learning from past experiences to deal with incoming new problems. Basically, a cognitive architecture is an intelligent system that counts on memory, learning, and decision for generic solutions [10]. Classical solutions, as the architectures Soar [11] and ACT-R [12], are constantly employed to solve generic problems. The resulting systems in studies with these architectures produce good generalizations in the tasks with social robots [13], [14]. However, for specific and focused studies, this generalization may not be the best solutions due to problems that are harder to control with classical architectures, such as encompassing the coordination of multiple sensory and motor modalities for the robot. The timing of proactive and reactive actions and the recognition of interacting human states (cognitive, affective, physical, etc.) are pointed out as the reason for the rise of new proposed architectures [15].

It is worthy to reiterate that the interactions may vary according to the users. For example, adults' and children's interactions with the robot are potentially very different due to children's neurophysical and mental development being ongoing. Furthermore, the view of cognition is being extended, suggesting that cognition is no longer an exclusive domain of the individual, but the product of a fine-grained interaction between agents of any nature, be the people or machines [16]. This is one of the main reasons for performing several studies to understand the user profile before and while developing such systems.

In HRI area, in children's educational domain using cognitive systems, one example of autonomous solution is the expressive agents for symbiotic education and learning (EASEL) [17] project. The EASEL aims to deliver a new set of robotic-based tutoring solutions: a synthetic tutoring assistant. It considers the utilization of different subsystems as modules to automate the whole system. Speech recognition in noisy environments [18] and audiovisual scene analyzer [19], dialog and behavior planning [20] and dialog handler and behavior adapter [21], content and behavior generation using a Zeno [22] robot, and a FACE [23] robot compose the EASEL modules. The user-centered studies involve the children perceptions in relation to the robot [24], how children play with different robot shapes [25], how different games change the type of interaction [26], how children adopt different behaviors in symbiotic cooperation tasks [27], and how to model personalities in robots for symbiotic interaction in the educational context [28]. However, this proposal does not allow people from outside the project to design the interactions, and it does not provide accurate reports about the users' accomplishments after the interactions.

In this article, a new system named R-CASTLE is proposed that allows the students to do activities that improve their learning about the content approached as the teachers create new activities and generate reports about the students' performance during the activities.

III. Architecture Overview and Support Modules

The proposed architecture aims to offer a framework capable of being programmed for controlling the robot's resources, addressing previously known as well as new learning content, in interactive educational activities. There are some open-source programmed behaviors (such as object and face recognition, web research, and speech processing), and an interface is provided where the nonprogramming designer can use all robot features to create new activities. A NAO robot was used as an interactive device in our tests because it was one of our laboratory tools, but the architecture can be easily adapted to control other devices as well.

The R-CASTLE is constituted by several modules, according to their functionalities: vision, dialog, memory and content, and assessment, as illustrated in Fig. 1. All of these modules are detailed in the following sections.

A. Vision Module

The visual system goal is to recognize and classify the specific objects or images that were previously trained in the system along the interaction. In this case, the inputs were constituted by images obtained by the NAO's frontal camera

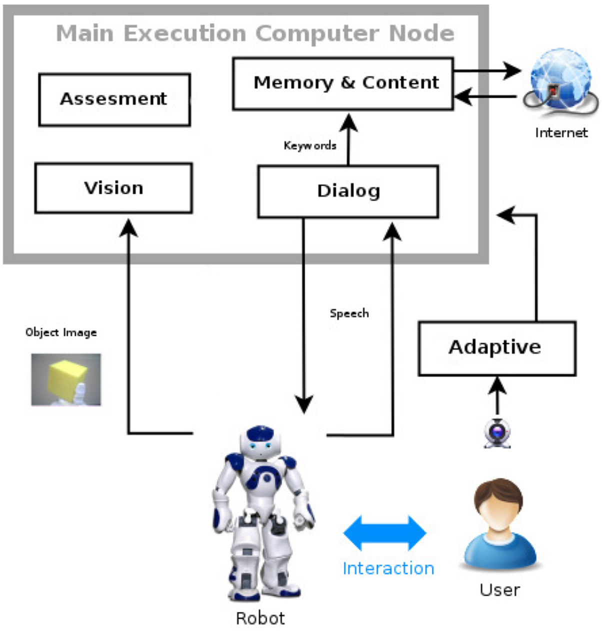

Fig. 1. Architecture overview summarizing the modules' functions.

and the images were collected and classified previously [29]. Several combinations of techniques of machine learning were tested to approach geometric figures. This content was chosen so that the children could handle solid objects [29], [30].

B. Dialog Module

The dialog module allows the interaction to be performed in a more human-like communication manner through verbal communication. Regarding the methods, the system is a composition of three core functions: 1) the speech recognition and text to speech functions; 2) a combination of natural language processing (NLP) techniques for basic sentence interpretation; and 3) a simple doubt/negation/affirmation analysis. The *Google's API* is used for speech recognition while the text to speech functions were provided by the software of the NAO robot. The NLP techniques came as part of the *TextBlob* package available for NAO robot users. In addition, more python packages for NLP are being employed in R-CASTLE. Finally, the affirmation/doubt/negation analysis is based on keywords defined by the designer that can indicate one of these statements. The checking order is doubt, negation, and affirmation to reduce the misinterpretations that some expression types may cause in the sentences. The keywords for each statement are held in specific files and can be changed before the interactions.

C. Memory Content Module

All the content regarding information aimed to to be addressed along the interaction is managed by the memory content module. The algorithms belonging to this module allow to manually program contents in the system database or search on the Internet for questions, along the interaction, that were not programmed by the designer. This procedure can be requested at any time by the dialog module. First, the system searches its local database, composed of answered questions. For new questions, it searches the Wikipedia database through an API and the new knowledge is added to the memory. All the answered questions are preprocessed by the dialog module, and only the resulting keywords that are mainly nouns are used for the web search.

TABLE I
OBJECTIVE MEASURES BY GROUP

Attention (α)	**Communication** (β)	**Learning** (γ)
Eye gaze (Eg)	Number of Words (nW)	Right/Wrong answers (RWa)
Posture (p)	Emotions (Em)	Time to answer (Tta)

D. Assessment

The assessment module has the functionality of collecting relevant information about the users along the interaction and displaying these reports to the designers any time after the interaction. Once the generated reports aim to work as interaction logs, all the measures extracted by the other modules are gathered and processed to show how they evolved along the time in the interaction. It should be done in a very intuitive manner. Graphic users interfaces (GUIs) are being developed to do so. These reports will help the designers plan the next activities based on the difficulties presented by the users and detected by the assessment module. Its objective is to be a supportive tool in modeling the users learning profile.

IV. Adaptive Module

An important part of the interaction design is to hold the attention span of the users and keep them interested as long as possible. Adaptation is the key to achieve this goal in all kinds of interactions and it may occur by changing the robot's behavior. The robot can change its behavior, assuming different roles, such as of a security guard or just guidance, error or correct performer, straight or interactive instructor, tutor or learner, and in all of them, the variation produces a different result.

The adaptive module aims to change the robot's behavior according to the observed users' indicators. For a better analysis, the indicators were divided into three main groups regarding the measures of attention, communication, and learning of the users. It is worthy to notice that these three measures are qualitative. Thus, a set of objective measures are required to conclude a consistent qualitative result. The objective measures of each group are summarized in Table I. The groups are shown in the first line (with the corresponding denotation function in parentheses) followed by the respective indicators in their columns. The indicators are: eye gaze (Eg) and posture for the attention; the number of words spoken by the user and emotions of users for communication; and the right/wrong answer and time to answer the proposed exercises for the learning group. All the objective measures are being modeled as a vector of each group. The functions are: α to attention, β to communication, and γ to learning.

These functions will process the corresponding vectors of objective measures and generate outputs according to particular rules from each function. In this way, the resulting

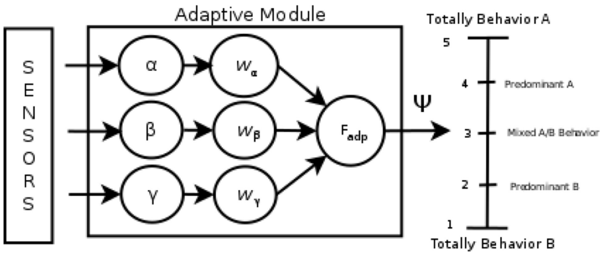

Fig. 2. Adaptive module representation.

functions with the corresponding vectors are represented as $\alpha = (Eg, P)$, $\beta = (nW, Em)$, and $\gamma = (RWa, Tta)$. The function outputs will have two utilizations: be saved and sent to the assessment module to produce the reports (to the designers about the interactions with the users) and to be used as input for the adaptive behavior function. The adapted robot's behavior, denoted as Ψ, is calculated by function F_{Adp} shown in the following equation:

$$\Psi(\alpha, \beta, \gamma) = F_{Adp}(w_\alpha * \alpha, w_\beta * \beta, w_\gamma * \gamma). \tag{1}$$

F_{Adp} is a function to adapt the resulting robot's behavior Ψ, trying to optimize the interaction engagement and learning rate. Parameters α, β, and γ are the group function outputs and w_α, w_β, and w_γ are the corresponding weights, calculated by some supervised machine learning algorithm.

Although the intention is to achieve many behavior degrees, the resulting robot's behavior Ψ so far is a discrete scale between two independent behaviors. This scale ranges from 1 to 5, in which 1 means exclusively one of the behaviors and 5 means exclusively the other one. More specifically, if the result is number 3, a mixed behavior with half of both scale behaviors is performed by the robot. This model is represented in Fig. 2, in which the adaptive module receives the entries read by the sensors, processes the F_{Adp} function according to the respective parameters, producing behavior Ψ, that will be placed on the behavior scale set by the designer.

For instance, consider two independent behaviors to be placed on the robot's behavior scale: "physical interaction" and "content learning." In the first one, the robot will take more physical interaction with the student, like inviting him/her to dance, perform some game or physical exercise, and so on. In the second, the robot will approach the programmed content in a very straightforward manner. In this case, as was mentioned previously, if F_{Adp} takes 1 as a result, the robot will assume the totally physical interaction behavior; if the result is 5, it will take the content learning behavior; and if the result is 3, there is a merged behavior between the conditions. In case the scale is 2, the robot will take a predominant physical interaction behavior, interchanging small content approaches along the game, whereas when the scale is 4, the robot will assume a predominant content learning behavior, interchanging small physical interactions along the content approach.

Adapting to a more continuous behavior scale, one solution being studied is to group the robot performances by behavior (e.g., all the listed interactions from physical interaction behavior are performances of this behavior). Thus, the behaviors on the scale are defined by its performances. Behavior Ψ will use variable λ to influence the probability of the system to choose a specific performance from the predominant behavior. λ is the probability of choosing the performance from scale behavior A and the performances from behavior B will be chosen with probability $(1 - \lambda)$. For example, taking the scale of Fig. 2, where $\Psi(5)$ = Totally behavior A, the system will choose performances from behavior A with probability 1 and group B with probability 0. It means that only performances from group A will be chosen. The same is valid for $\Psi(1)$, but with inverted values. For the mixed behavior, the system will choose performances from both groups with 0.5 of probability. For the scale values between the middle and the extremities, 2 and 4, the system will choose performances with probability 0.75 from one group and 0.25 from the other, depending on the scale value. Therefore, it is possible to extend the behavior scale to have more than five behavior values, calculated by the value λ. It is known that the probability variable λ will not assume totally continuous values, being biased on discrete numbers, but as already said, this solution is under investigation.

These performances may be proposed by the user, for example, to play soccer. The user teaches the robot to perform something and this performance will be saved in the memory module and can be used in future interactions, which also characterizes cognition. It is easy to notice that these solutions will require designer creativity to create as many performances as possible, in order to build a large and varied set of performances from each behavior.

Regarding the specific algorithms to collect the objective measures, those that depend on users' verbal communication (nW, RWa, and Tta) will be provided by the dialog system, whereas the others, will be analyzed with adaptive modules algorithms. The Eg is under development (see Section V) and can already detect face deviation. The posture and emotion analyzers were developed in previous works in the same research group. Cavalcante [31] developed a system based on Microsoft Kinect to detect key points in people to extract posture measurements, whereas Libralon and Romero [32] considered a regular web camera to classify emotions based on face features. Both systems need smaller adjustments that are already being installed. Furthermore, an emotion analyzer system using a convolution neural network (CNN) [33] is also being developed for this purpose.

Studies are being performed to create reliable data as the ground truth for training the weights w_α, w_β, and w_γ of F_{Adp}. Based on this, it will be possible to train supervised machine learning algorithms to optimize F_{adp} and also compare their efficiency.

V. Results and Discussion

The following sections show the results of the modules that provide the measures for the adaptive model.

A. Vision Module

The proposal is that a designer could change the programmed content and this change could be done anytime before the interactions. Thus, a low training time is desirable in order to change the database with few samples and training the system quickly. An ensemble classifier with methods offering acceptable accuracy with low training time has been implemented and it is presented as follows.

The ensemble classifier takes into consideration the prediction of three supervised machine learning methods: 1) KNN; 2) MLP; and 3) SVM [34]. These are classifiers that are well known. There are two most common approaches for images to process the input data in the three mentioned classifiers: 1) the image's raw pixel as an array (PXL) and 2) the image's histogram (HST). The ensemble takes into account six combinations of the approach classifier, corresponding to the possible combinations of using the three classifiers mentioned.

In configuration, the KNN with five near neighbors ($K = 5$) was set, the MLP with three layers (01 hidden) with 100 neurons each and limited max step in 500, and a linear kernel to SVM. Attributes not mentioned here were adopted with default values. The images content is a basic spatial geometrical figure, such as a cube, a pyramid, and a sphere. The classification results need to indicate one of these three figures or classes of objects. In this way, the database was trained with 90 images of each of these three classes, resulting in a total of 270 samples collected from the NAO's camera in an ideal scene, which means that the images have only the geometric figure with a monochrome background. The training times in this configuration were 0.49 s for KNN, 4.43 s for MLP, and 5.25 s for SVM, and combining the three models, the time spent was, on average, 10.18 s.

The test data set was divided into two groups of ideal and noise images. The ideal images were similar to those in the training set. On the other hand, the noise group were contained the images with a disturbed scenario in three levels: first only changing the scenario's background (light noise), then adding some other objects in the scene, such as bottles, pens, and backpacks (medium noise), and, finally, taking away the object from the robot's hand and showing it in different distances from the robot's camera (heavy noise). Each group was constituted by 90 samples separated in 30 samples for each class. Some examples of used images are shown in Fig. 3.

In the noise group, ten samples were collected for each class in each noise level. The classifier accuracy for the noise level is shown in Fig. 4. Comparing the results of the ideal scene with the noisy scenes, as is shown in Fig. 5, it is possible to notice an acceptable accuracy of the classifiers in the group with noises, in which the approach using ensemble with the voting of all classifiers stands out, and the MLP classifier with a histogram approach.

An unexpected contrast was found in the ensemble classifier compared with SVM using the pixel approach method, in which the SVM itself classified 100% of the samples versus 98.8% of the ensemble classifier. In this case, it is an undesirable situation, suggesting a processing waste. However, the ensemble classifier was better than this simple classifier for the noise images. Analyzing the results, for a noisy scene, the ensemble technique obtained greater scores than other classifiers. In this article, the best accuracy scored from a noisy scene is 93.25%, for MLP with the histogram approach. The ensemble classifier was the best, with majority voting, for all classifications and approaches.

It was noticed that most of the mistakes were made when the objects were positioned farther from the camera. A problem was noticed with the MLP classifier using the

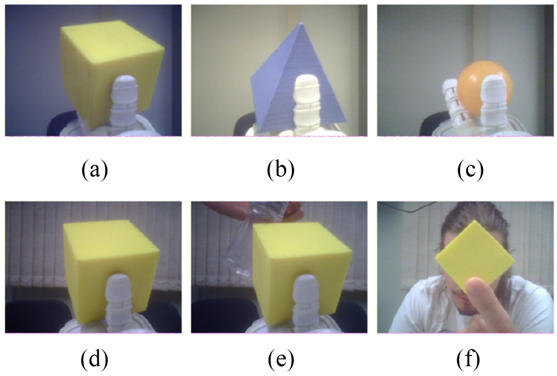

Fig. 3. Images used for tests in the two conditions. First row is the ideal scenes of (a) cube, (b) pyramid, and (c) sphere. Second row is a cube in three noise levels. (d) Light. (e) Medium. (f) heavy.

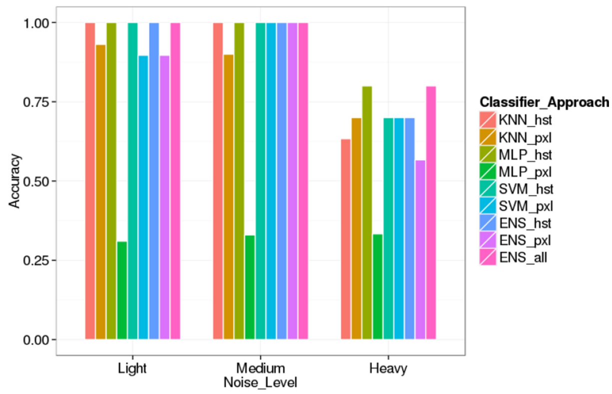

Fig. 4. Comparison between the classifiers separated by the noise levels from scene.

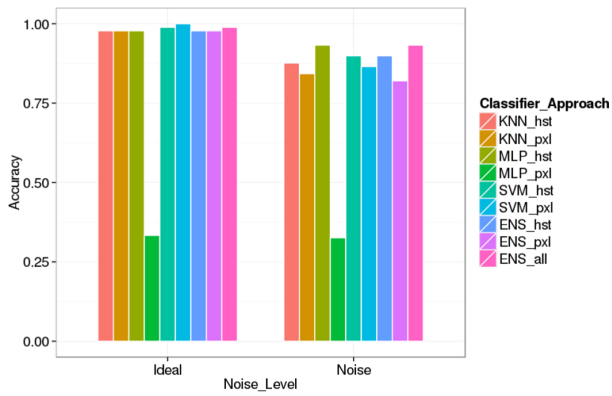

Fig. 5. Comparison between the ideal scene and the noise scene with three level noises.

TABLE II
RIGHT ANSWERS, WRONG ANSWERS, AND AVERAGE PROCESSING TIME OF THE THREE NOUNS FINDER OF THE DIALOG SYSTEM, GROUPED BY THE SENTENCE DIFFICULTY LEVEL

		Method 1			Method 2			Method 3		
Number of Nouns (Number of Sentences)	Total Expected Nouns	Right Awnsers	Wrong Awnsers	AVG Time (SD)	Right Awnsers	Wrong Awnsers	AVG Time (SD)	Right Awnsers	Wrong Awnsers	AVG Time (SD)
1 (3)	3	3	0	1e-04(4,7e-05)	3	3	5.32(1.00)	3	0	0.62(0.29)
2 (4)	8	8	0	2e-04(9,4e-05)	5	4	6.43(0.48)	6	0	1.45(0.24)
3 (3)	9	7	4	2e-04(9,4e-05)	5	3	8.85(1.18)	6	3	4.42(0.59)

TABLE III
NUMBER OF KEYWORDS FOUND PER QUESTION

	sun	**edges**	**vertex**	**sides**	**table**	**robot**	**triangle**	**planet earth**	**solar system**	**blue sky**	**world cup**	**where Atlantic Ocean**
Appears on Dialog	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No
Keywords	4	5	5	0	0	4	5	5	4	0	3	0

pixel approach (mlp_pxl), because classifications have become addicted. There is no information about why this happens, and, more research is needed.

B. Dialog Module Results

Only tests for the interpretation function have been performed so far. Three different methods have been evaluated to filter out the given phrase and return a topic, which was expected to be the core of the sentence. The topic was sent to the memory and content module.

Method #1 is a very simple effort to filter out *stopwords* from the input phrase, which are the words that have no meaning in the sentence, such as "not," "the," "what," etc. Method #2 was based on tagger analysis, translating each word of the phrase from Portuguese to English (the users are Brazilian Portuguese speakers and the methods used so far work only for the English language) and looking for nouns that could be interpreted as the main topic of the sentence. Method #3 is a combination of these two methods that first filters out the *stopwords* and, then, the tagger analysis is applied. It decreases processing time because of the fewer words needed to be translated.

The three methods were tested using ten sentences that were defined in three processing difficulty levels. The input set was divided into three easy sentences (with only one noun), four medium sentences (with two nouns), and three difficult sentences (with three or more nouns). It was expected that the more nouns the sentences have, the less accuracy will be shown in finding the right methods for the core words. The results are shown in Table II. The first column shows the number of nouns and the number of sentences in that condition, in parentheses. In the second column, the total expected number of nouns to be found in each group of sentence level difficulty is shown. In the rest of the columns there is efficiency of three methods, divided into the right answers (how many nouns were found correctly), the wrong answers (words not found that were not nouns), and the average time to process the sentence and return the topics.

The first method had an accuracy of 11 out 11 when tested with simple sentences composed of *stopwords* and one or two nouns. As the sentence's complexity level increases, the efficiency tends to decrease; this was an expected consequence, because this method works as a filter and does not deal with the different grammar classes.

The second method did not present satisfactory results due to the large distance between the meaning of isolated words and the meaning of the same word when used as part of a phrase. For instance, the word "cup" will have a different meaning when used as a standalone word than when it is combined with another word, for example, "world cup."

Furthermore, this inefficiency increases because in some instances, some isolated words in Portuguese have different meaning or taxonomy than in English.

As a combination of the two first methods, the third one obtained solid results as the number of individual words to translate was reduced. In this approach, it was expected that some cases of ambiguity were reduced, increasing the accuracy in more complex scenarios. According to the results, this method showed better performance when compared with methods #1 and #2.

In conclusion, the more simple the method, the less time was spent in the process. This suggests that an intelligent algorithm can be applied to predict how complex the conversation is going to be. Thus, this algorithm can switch between these methods to offer a simple question and less processing time or sophisticated processing, resulting in more time processing, according to the dialog necessity.

C. Memory Content Module Results

The evaluation for the memory content module was to first use the same sentences as used for the dialog module assess. For each sentence, five main words were prepared that were essential to have in a good answer, based on the Portuguese dictionary. We took the keywords given by the dialog module in the web search with the implemented method and evaluated the answers. An answer was considered very good when it had all the five main words and it was considered very bad when no main words were found.

As can be seen in Table III, the results are based on how well the dialog module perceives the question to find keywords to search. The average appearance of keywords per answer is 2.9, which is below the average.

To overcome this issue, another method that does not depend on the dialog module was used. Ten people were asked nonpersonal questions composed only by keywords in the search. For example, "what is a robot?" turns into "robot." It was requested that the questions be divided into easy, medium and hard and have one keyword, two keywords, and three keywords, respectively. Each subject asked five questions: two easy, two medium, and one hard. After receiving the answer, the subject rated it from 0 to 5, meaning 0 for "not at all" and 5 for "definitively," based on the usefulness and if it covered what he expected. A large increase was found. The average rates obtained were: for the easy questions, it was 4.5, for the medium ones it was 3.7, and for the hard ones, it was 3.1. The final average for all the questions together was 3.9, 1 point more than the previous assessment, which suggests that an improvement in the dialog processor would increase this module efficiency as well. However, the results suggest that the system has higher approval from the people in the easier questions.

D. Adaptive Module Results

In this article, only the results from the Eg algorithm are shown. Although its approach is quite simple, it takes a significant processing power. Thus, this module needs to run in a different computer node that the main process. When the main process is initialized, it sends a message to the assessment module, which triggers the deviation counter and the emotion classifier algorithms start to run. A socket implementation is used for nodes communication.

The method used to measure face deviation was the Haar Cascade [35], implemented in *opencv* library. This method is a machine-learning-based approach where a cascade function is trained from many positive and negative images and then used to detect any object in other images. An XML file was used to load the weights used in this algorithm, to classify faces. The cascade method works performing convolutions from masks of edge, line and center in a given number of neighbors, and this parameter is variable. A series of tests were conducted to analyze its impact.

After the deviation counter awakes, it runs capturing frames from a computer camera, counting the deviation number and time length until it receives a message from the main execution. For each frame, if a face is detected, the current time is assigned in two variables as initial and final time. Otherwise, it just updates the final time variable with the current time. When it detects a face again, it compares if the variables difference (final time–initial time) is greater than a chosen threshold. In the positive case, it counts the occurrence as a deviation, saving its time length. At the end of execution, the algorithm calculates the total number of deviations, the total time looking away, and the total time with the face gazing in the camera's direction.

The deviation counter algorithm allows some arguments to be changed. The arguments are the minimal time to count deviation, allowing more abrupt or soft deviations and the higher or lower values of minNeighbors (argument that is passed to detect the multiscale *opencv* function), which may let the attention deviation detection be more or less sensitive. The minNeighbors is the minimal number of neighbors that the cascade method needs in the rectangle convolution.

For this test, a video of approximately 65-s duration was recorded with a volunteer turning his face at some different angles and then with different speeds. The method was configured to allow a tolerance of 20° in four directions (left, right, up, and down) without counting a deviation. In total, 48 head turns were recorded but only eight of them were considered a real attention deviation for us. During the classification of this video, the scale factor (argument of the detect multiscale function) was initially changed, but this compromised the video duration, which disrupted the classification of time deviation.

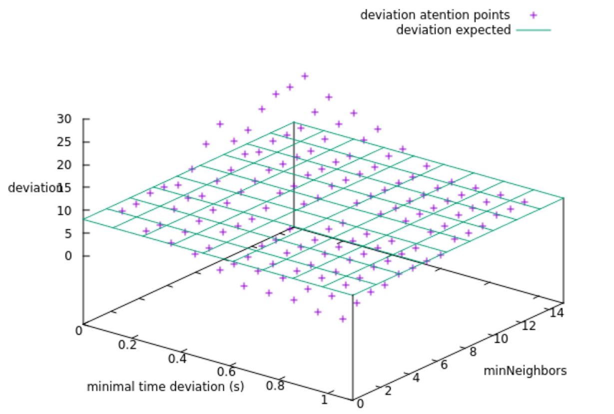

Fig. 6. Graph showing the points collected by varying two parameters (minNeighbors and minimal time deviation) and the plane with expected deviations.

TABLE IV
DETECTIONS AND NOT DETECTION DEVIATIONS OF THE ALGORITHM USING THE VALUES: POINT 1 = (0.6, 9, 8) AND POINT 2 = (0.7, 9, 7)

		Deviation	No Deviation
Detected	Point 1	7	1
	Point 2	7	1
No Detected	Point 1	1	39
	Point 2	0	40

After this, the minimum deviation time was varied from 0.1 to 1 s, being increased from 0.1, and the minNeighbors from 1 to 14, being increased from 1. The graph shown in Fig. 6 was generated by using the data collection

As can be seen in Fig. 6, the configuration with high minNeighbors and low minimal time deviation had a very large occurrence of type II error (because there are more pointers on top of the expected plane deviation attention). It means the errors that are false positives. On the opposite, it is possible to see also a very large occurrence of type I error (because there are more pointers below the expected plane deviation attention). Two points near the intersection of the planes were used, because this is the area of interest, where the errors of types I and II are minimized. These points are (time = 0.6, minNeighbors = 9) and (time = 0.7, minNeighbors = 9). In the first and second points, respectively, the algorithm counted 7 and 8 deviation, 18% and 17% of time lost in attention deviation, with accuracy of 97.92% and 95.83%. Complementary data can be observed in Table IV.

VI. CONCLUSION

This article presented the proposal, objectives, interactive studies, conclusion and ongoing developments of the R-CASTLE project: a robotic cognitive architecture that aims to deliver a new framework to assist in interactive educational activities design. Several contributions of this project can be highlighted.

For implementation matters, the proposed architecture will allow people without programming skills to easily design HRI tasks that will be done in an autonomous way. This type of intuitive framework is a good option for researchers coming from other areas, such as psychology and education, that are migrating to the HRI field, since this area has shown a significant growth in the last decades. Many teachers will benefit from its application since R-CASTLE can be used in different interactive devices, disposing of the need of expensive social robots.

Adaptive behaviors are important to achieve long-term goals and are viable observing user's body language, or physical indicators, along the interactions. Personal users' preferences saved by the system can be helpful to bring back their attention in long-term interaction. Memory and content and adaptive modules can provide this robot personalization, which can provide the feeling of rapport between the human and the robot, optimizing the engagement along the tasks. Combining communicative interaction and content approaches has shown to increase students' learning rates.

So far, the conceptual and technological developments have provided an architecture capable of capturing contextual information from audiovisual sensors, approaching and learning new concepts, creating individual personalization, and rapporting with students. As the quality of interaction can also vary according to different cultures, the presented studies performed in Brazil and those that are ongoing are helping to understand the perception of the users from developing countries that have less familiarization with high-tech robots.

REFERENCES

[1] R. Artstein *et al.*, "Listen to my body: Does making friends help influence people?" in *Proc. FLAIRS Conf.*, 2017, pp. 344–351.

[2] L. D. Riek, "Wizard of Oz studies in HRI: A systematic review and new reporting guidelines," *J. Human–Robot Interact.*, vol. 1, no. 1, pp. 119–136, 2012.

[3] G. Dorais, R. P. Bonasso, D. Kortenkamp, B. Pell, and D. Schreckenghost, "Adjustable autonomy for human-centered autonomous systems," in *Proc. 16th Int. Joint Conf. Artif. Intell. Workshop Adjust. Autonomy Syst.*, 1999, pp. 16–35.

[4] D. C. Tozadore, A. H. M. Pinto, and R. A. F. Romero, "Variation in a humanoid robot behavior to analyse interaction quality in pedagogical sessions with children," in *Proc. IEEE XIII Latin American Robot. Symp. IV Braz. Robot. Symp. (LARS/SBR)*, 2016, pp. 133–138.

[5] A. H. M. Pinto, D. C. Tozadore, and R. A. F. Romero, "A question game for children aiming the geometrical figures learning by using a humanoid robot," in *Proc. LARS/SBRp*, 2015, pp. 253–257.

[6] D. C. Tozadore, A. M. H. Pinto, G. Trovato, and R. A. Romero, "Wizard of Oz vs autonomous: Children's perception changes according to robot's operation condition," in *Proc. 26th IEEE Int. Workshop Robot Human Interact. Commun. (ROMAN)*, 2017, pp. 664–669.

[7] V. Montreuil, A. Clodic, M. Ransan, and R. Alami, "Planning human centered robot activities," in *Proc. IEEE Int. Conf. Syst. Man Cybern. (ISIC)*, 2007, pp. 2618–2623.

[8] M. Wald, "Using automatic speech recognition to enhance education for all students: Turning a vision into reality," in *Proc. 35th Annu. Conf. Front. Educ. (FIE)*, 2005, p. S3G.

[9] S. Lemaignan, F. Garcia, A. Jacq, and P. Dillenbourg, "From real-time attention assessment to with-me-ness in human–robot interaction," in *Proc. 11th ACM/IEEE Int. Conf. Human Robot Interact.*, 2016, pp. 157–164.

[10] P. Langley, J. E. Laird, and S. Rogers, "Cognitive architectures: Research issues and challenges," *Cogn. Syst. Res.*, vol. 10, no. 2, pp. 141–160, 2009.

[11] J. E. Laird, *The Soar Cognitive Architecture*. Cambridge, MA, USA: MIT Press, 2012.

[12] ACT-R. (2017). *Act-R Architecture Version 6*. [Online]. Available: http://act-r.psy.cmu.edu/

[13] P. Baxter *et al.*, "Long-term human–robot interaction with young users," in *Proc. IEEE/ACM Human–Robot Interact. Conf. (Robots Children Workshop)*, 2011, pp. 1–4.

[14] J. G. Trafton, L. M. Hiatt, A. M. Harrison, F. P. Tamborello, S. S. Khemlani, and A. C. Schultz, "ACT-R/E: An embodied cognitive architecture for human–robot interaction," *J. Human–Robot Interact.*, vol. 2, no. 1, pp. 30–55, 2013.

[15] P. Baxter, S. Lemaignan, and J. G. Trafton, "Cognitive architectures for social human–robot interaction," in *Proc. 11th ACM/IEEE Int. Conf. Human–Robot Interact. (HRI)*, 2016, pp. 579–580.

[16] T. Belpaeme *et al.*, "Child–robot interaction: Perspectives and challenges," in *Proc. Int. Conf. Soc. Robot.*, 2013, pp. 452–459.

[17] D. Reidsma *et al.*, "The EASEL project: Towards educational human–robot symbiotic interaction," in *Proc. Conf. Biomimetic Biohybrid Syst.*, 2016, pp. 297–306.

[18] S. Fernando *et al.*, "Automatic recognition of child speech for robotic applications in noisy environments," *arXiv preprint arXiv:1611.02695*, 2016.

[19] A. Zaraki, D. Mazzei, M. Giuliani, and D. De Rossi, "Designing and evaluating a social gaze-control system for a humanoid robot," *IEEE Trans. Human–Mach. Syst.*, vol. 44, no. 2, pp. 157–168, Apr. 2014.

[20] M. ter Maat and D. Heylen, "Flipper: An information state component for spoken dialogue systems," in *Proc. Int. Workshop Intell. Virtual Agents*, 2011, pp. 470–472.

[21] H. Van Welbergen, D. Reidsma, and S. Kopp, "An incremental multimodal realizer for behavior co-articulation and coordination," in *Intelligent Virtual Agents*. Heidelberg, Germany: Springer, 2012, pp. 175–188.

[22] D. Cameron *et al.*, "Congratulations, it's a boy! Bench-marking children's perceptions of the Robokind ZENO-R25," in *Proc. Conf. Auton. Robot. Syst.*, 2016, pp. 33–39.

[23] D. Mazzei, N. Lazzeri, D. Hanson, and D. De Rossi, "HEFES: An hybrid engine for facial expressions synthesis to control human-like Androids and Avatars," in *Proc. 4th IEEE RAS EMBS Int. Conf. Biomed. Robot. Biomechatron. (BioRob)*, 2012, pp. 195–200.

[24] D. Cameron *et al.*, "You made him be alive: Children's perceptions of animacy in a humanoid robot," in *Living Machines* (LNAI), vol. 10384, M. Mangan, M. Cutkosky, A. Mura, P. F. M. J. Verschure, T. Prescott, and N. Lepora, Eds. Cham, Switzerland: Springer, 2017. doi: 10.1007/978-3-319-63537-8_7.

[25] D. Cameron *et al.*, "Children's age influences their use of biological and mechanical questions towards a humanoid," in *Proc. Conf. 18th Towards Auton. Robot. Syst. (TAROS)*, 2017, pp. 290–299.

[26] D. Cameron *et al.*, "Children's age influences their perceptions of a humanoid robot as being like a person or machine" in *Living Machines*. Barcelona, Spain: Springer, 2015, pp. 348–353.

[27] V. Charisi *et al.*, "Towards a child-robot symbiotic co-development: A theoretical approach," in *Proc. AISB Convent.*, 2015.

[28] D. Cameron *et al.*, "Designing robot personalities for human–robot symbiotic interaction in an educational context," in *Proc. Conf. Biomimetic Biohybrid Syst.*, 2016, pp. 413–417.

[29] A. H. M. Pinto, A. X. Benicasa, L. O. Oliveira, R. C. G. Meneguetti, and R. A. F. Romero, "Attention based object recogniton applied to a humanoid robot," in *Proc. LARS/SBR*, 2014, pp. 136–141.

[30] A. H. M. Pinto, L. O. de Oliveira, A. X. Benicasa, R. C. G. Meneghetti, and R. A. F. Romero, "Inserção de um robô humanoide no ensino de objetos geométricos 2D sobrepostos," in *Proc. Anais do Simpósio Brasileiro de Informática na Educação*, vol. 25, 2014, pp. 632–641.

[31] F. Z. M. S. Cavalcante, "Reconhecimento de movimentos humanos para imitacao e controle de um robo humanoide," M.S. thesis, Dept. Comput. Sci., Univerisade de Sao Paulo, São Carlos, São Carlos, Brazil, 2012.

[32] G. L. Libralon and R. A. F. Romero, "Mapping of facial elements for emotion analysis," in *Proc. IEEE Brazil. Conf. Intell. Syst. (BRACIS)*, 2014, pp. 222–227.

[33] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2012, pp. 1097–1105.

[34] A. Smith and B. Jones, "On the complexity of computing," in *Advances in Computer Science*, A. B. Smith-Jones, Ed., Publ. Press, 1999, pp. 555–566.

[35] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 1, 2001, p. 1.

Daniel Tozadore received the B.Sc. degree and the M.S. degree in computer science from the University of São Paulo, São Paulo, Brazil, in 2014 and 2016, respectively, where he is currently working toward the Ph.D. degree in computation department with the ICMC.

His current research interests include pedagogical social robotics research, using artificial intelligence techniques in humanoid robots with the purpose of making human–robot interaction as natural as possible for the humans, its application focuses mainly on early childhood education, machine learning, and bio-inspired computing.

Adam H. M. Pinto received the B.Sc. degree from Goiás Federal University, Goiânia, Brazil, in 2011, and the M.S. degree in computer science from Institute of Mathematics and Computer Science, ICMC, University of São Paulo at São Carlos, São Carlos, Brazil, in 2015, where he is currently pursuing the Ph.D. degree.

His current research interests include educational social robotics and EEG signal analysis.

João Valentini is currently working toward the bachelor's degree with the Computation Department, ICMC, University of São Paulo, São Paulo, Brazil.

His current research interest includes natural language processing to allow human–robot verbal communication for educational purposes.

Marcos Camargo is currently working toward the bachelor's degree with the Computation Department, ICMC, University of São Paulo, São Paulo, Brazil.

His current research interest includes artificial intelligence techniques classify images in human–robot interaction for educational purposes.

Rodrigo Zavarizz is currently working toward the bachelor's degree with the Computation Department, ICMC, University of São Paulo, São Paulo, Brazil.

His current research interest includes cognitive techniques for robot learning applied to HRI designs.

Victor Rodrigues is currently working toward the bachelor's degree with the Computation Department, ICMC, University of São Paulo, São Paulo, Brazil.

His current research interest includes algorithms to detect user attention deviation applied in HRI context.

Fernando Vedrameto is currently working toward the bachelor's degree with the Computation Department, ICMC, University of São Paulo, São Paulo, Brazil.

His current research interest includes algorithms to analyze user emotions and adapt the robot's behavior.

Roseli Romero received the Ph.D. degree in electrical engineering from the University of Campinas, Campinas, Brazil, in 1993.

She has been an Associate Professor with the Department of Computer Science, ICMC, University of São Paulo, São Paulo, Brazil, since 1988. Her current research interests include artificial neural networks, machine learning, fuzzy logic, robotics, and computational vision.

CHAPTER

4

GRAPHICAL USER INTERFACE FOR EDUCATIONAL CONTENT PROGRAMMING WITH SOCIAL ROBOTS ACTIVITIES AND HOW TEACHERS MAY PERCEIVE IT

Contribution statement

D.C. Tozadore conceived the Graphical User Interface proposal and the proposed the scheme of interaction of Figure 1. D.C. Tozadore and R. A. F. Romero performed the literature review of Section 2. D.C. Tozadore and R. A. F. Romero proposed the pedagogical approach of Section 3. D.C. Tozadore proposed and implemented the windows of the interface, resulted in the Figures 2, 3, 4, 5, 6, 7, 8, 9 and 10. D.C. Tozadore run the experiments wit the GUI and the teachers, resulted in Figure 11. Both authors contributed on writing the manuscript. R. A. F. Romero,in addition, supervised all this research work.

Presentation

The paper in this Chapter presents specific details about the GUI development, its features and the teachers' perceptions about it. The GUI is part of the R-CASTLE that runs over the architecture presented in Chapter3, Thus, it was implemented after it. Herein, the experiments were about an R-CASTLE GUI presentation for 14 teachers that answered a questionnaire regarding their perception without interacted with the whole system. Findings were crucial to perform corrections and guide the next implementations of the GUI. A case study with 2 of these teachers that used the whole system will be shown in Chapter 5.

Revista Brasileira de Informática na Educação – RBIE
Brazilian Journal of Computers in Education
(ISSN online: 2317-6121; print: 1414-5685)
http://br-ie.org/pub/index.php/rbie

Submission: 14/02/2019; 1st round notif.: 28/05/2019; New version: 13/06/2019; 2nd round notif.: 02/07/2019
Camera ready: 29/08/2019; Edition review: 30/10/2019; Available online: 16/02/2020; Published: 16/02/2020

Graphical User Interface for educational content programming with social robots activities and how teachers may perceive it

Daniel Carnieto Tozadore
Institute of Mathematics and Computer Sciences
University of São Paulo
tozadore@usp.br

Roseli Aparecida Francelin Romero
Institute of Mathematics and Computer Sciences
University of São Paulo
rafrance@icmc.usp.br

Abstract
Interactive devices have been successfully applied in education in the last decades. The most used devices for such tasks are personal computers and tablets, due to its financial trade-off and popularization. Social robots are less used, mainly because of their cost and the complexity of being programmed. In this paper, a solution to work around the complexity of programming social robots is presented as a Graphical User Interface (GUI). The GUI system controls an interactive robot which plays with the students and adapts its behavior autonomously. During the activity execution, the adaptive algorithm detects student's body signals and verbal responses to adapt the addressed content to harder or easier questions. After creating and running an activity, all sessions' evaluation and information can be accessed for visual analysis, as well as students' preferences throughout the interaction. The proposal was presented to regular teachers from the elementary school that answered a questionnaire about their perception of this proposal. The answers were analyzed and, in general, they seemed to slightly notice the system potential in and how it can support them in after-classes exercises, despite it requires some time to fully get used with the interface.
Keywords: *Human-Robot Interaction; Graphic User Interface; Educational Tool, Educational Robotics*

Cite as: Tozadore, D. C., & Romero, R. A. F. (2020). Graphical User Interface for educational content programming with social robots activities and how teachers may perceive it. Brazilian Journal of Computers in Education (Revista Brasileira de Informática na Educação - RBIE), 28, 191-207. DOI: 10.5753/RBIE.2020.28.0.191

1 Introduction

Social robots are the robots that are capable of changing information between themselves or with humans (Goodrich & Schultz, 2007). They are widely used in several tasks from entertainment to medicine (Leite, Martinho, & Paiva, 2013). However, social robots are far from achieving popularization due to the high costs and lack of people's knowledge about designing and programming robots. Because of that, smartphones and tablets are the most common electronic devices used in learning activities, for being the opposite of the robots in this sense: they have low cost and a consolidate familiarization with the users. Mainly, in the educational field, the lack of training of teachers and their inclusion in the robot's programming are highlighted as one important concern in a worldwide scenario. Some authors pointed out this factor as an even more critical problem than the robot's higher costs (Johal, Castellano, Tanaka, & Okita, 2018).

Research that considers social robots to achieve success in tasks with humans are placed in the Human-Robot Interaction (HRI) field. These types of applications are known for increasing people's curiosity and motivation in social and intellectual activities because interactive robots are not common in our daily lives. Nonetheless, after the robot loses its novelty and the users get used them, a decrease in the users' motivation and attention span is commonly noticed. A robot programmed with memory about the users, personalized conversations and adaptation of content difficulty can hold students' interest in the pedagogical interactions for a longer period, compared to robots with simple and monotonous behavior. By searching in the literature, several types of research work can be found on building an efficient adaptive system (Belpaeme et al., 2018a). But a lack of studies to improve content programming effortlessness was noted. It means few social robotic systems allow non-programmers to design and execute HRI activities.

This paper is an extension of the published work entitled "Graphical User Interface for Adaptive Human-Robot Interaction Design in Educational Activities Creation", in the XXIX Simpósio Brasileiro de Informática na Educação (SBIE). The first version presented the Graphical User Interface (GUI) components and information that would be more helpful to the teachers, such as dialogue, content, and student and evaluation databases. This extended version presents also some system's implementation decisions and interfaces functionalities that would be more relevant to programmers, such as the vision subsystem and their methods, the adaptation algorithm and the interaction interface. It also deliveries an unprecedented study with regular teachers from the elementary school. They participate in a presentation about the proposed system and they were inquired about their opinion of how much this type of solution can be helpful to them in after-classes exercises.

The robotic architecture is a cognitive adaptive system (Tozadore et al., 2017) that encapsulates a module-based implementation to run over the robot's sensors and actuators. The interaction flow overview provided by this architecture is shown in Figure 1. The person who designs the interactions (the designer) can program the activities in the system's GUI. The designer is represented as the teacher in this illustration since it is expected that the teachers mainly perform this role. During the activities execution, the system controls the robot to autonomously interact and adapt to the student.

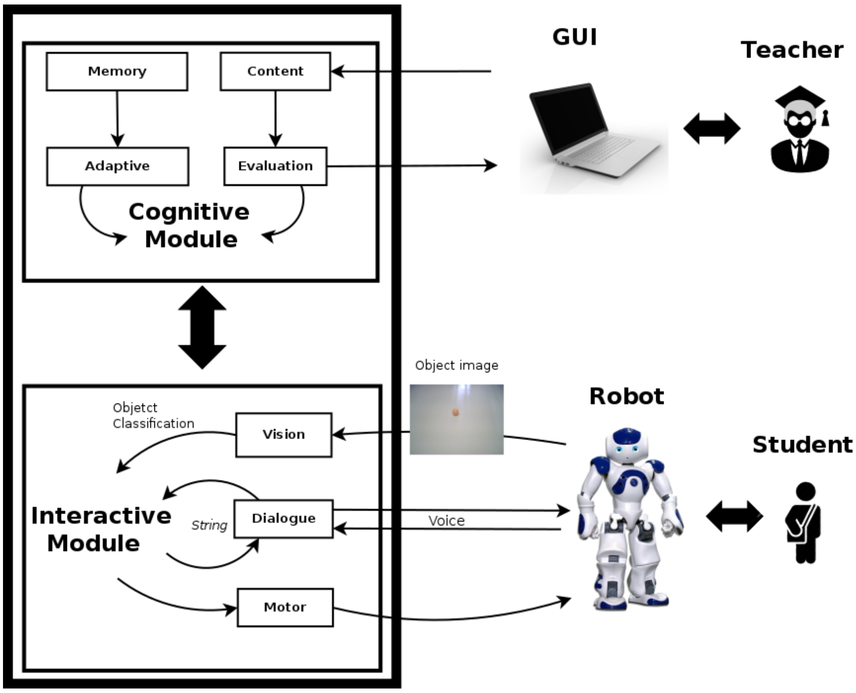

Figure 1: Architecture scheme.

The system interface with the student is a NAO robot, from Softbank robotics[1] due to being part of the available materials of the researcher institution. However, any interactive device with microphone, screen and speakers can be used with this system, after the right calibration. NAO is a 60 cm tall humanoid robot that has visual and sound resources (among others) designed to interact with humans in general purposes. Its application in educational tasks has been explored in several studies and shown well accepted.

2 Related Work

The predominance of mountable kits as Lego Mindstorms[2] and Pete[3] are often highlighted in the literature (Mubin, Stevens, Shahid, Al Mahmud, & Dong, 2013). They offer intuitive graphical interfaces to the users and they are most of the applications used as educational robotics (Benitti, 2012). However, their usage is limited to the STEM field[4] domain. After programming, the robots execute the student's code and there is no interaction between them. Hence, they are perceived by the students more than a learning tool than an agent that can play an active role in their cognitive process. Whereas social robots are not programmed by the students, but they are more supportive during the activities and capable to cover topics for more areas (Belpaeme et al., 2018b) due to the capability of verbal communication.

The development of systems that adapt to the students' difficulty is something commonly aimed in technological teaching designs. The Intelligent Tutoring Systems (ITS) are systems which adapt to the student necessities. They are often used in electronic learning and its implementation may vary from one application to another (Murray, 1999). Their usage increases the students' learning experience and provides better consequences in the content fixation. In general, the ITS also provide to the teachers an easy approach to plan the activity to be performed through GUI (Paiva, 2017). Nonetheless, embodied systems provide a more complete experience

[1] www.softbankrobotics.com
[2] www.lego.com/en-us/mindstorms
[3] www.pete.com.br
[4] Science, Technology, Engineer and Mathematics

than virtual learning environment and few of the existing ITS can be used along social robots (Platz, Krieger, Niehaus, & Winter, 2018). Those whose made use of ITSs with robots claimed they witness significant enhance in the young student's learning rate. These contributions go from pronunciation skills (Spaulding, Chen, Ali, Kulinski, & Breazeal, 2018) to Mathematics (Clabaugh et al., 2017). Moreover, GUI's allowed the settings of educational robotics that are not sociable to be more intuitive. Thus, it is expected to be helpful in this scenario as well (Rivas et al., 2015).

Among the advantages of social robots, one that can be highlighted is the advantage to aggregate human relation characteristics to the process such as shape and personalization of the robot. For instance, personalization in social interactions has shown to be an alternative in keeping the engagement of the users (Lucas et al., 2018). In the same way, by simulating the feeling of rapport building between robot and user, is possible to explore social techniques to enhance the results in the performed activity (Lucas et al., 2018).

Shiomi et al. (2006) investigated the influence of the free-play interaction and guidance of the robots. Their study highlights the importance of applying HRI personalization techniques to call the user's attention to scientific subjects. As a result, they found that the robot that performed personalized interactions by calling the visitors by name was better evaluated by the visitors themselves. Also, the robots that carried out a childlike free-play interaction and guided the visitors were the best in attracting attention to scientific explanations.

Research with interactive robotics architectures suggested significant improvement in multimodal interaction, achieved with a simple file management solution (Cortellessa et al., 2018). Multimodal emotional robots are playing an essential role when interacting with children. Results showed that the more human communication resources are demonstrated by the robot, the more the children's confidence in those systems increases (Kessous, Castellano, & Caridakis, 2010). In educational applications, studies about adaptive robots are recent and the authors report encouraging findings in their usage (Gao, Barendregt, & Castellano, 2017; Jones & Castellano, 2018).

However, little is known about how much these works with robots collaborate to place the teacher in a comfortable and active role in planning the activity. This is pointed as an issue to be enhanced in the area (Johal et al., 2018).

3 Pedagogical Model

The pedagogical model is based on constructivism, as the educational robotics in general (Kafai et al., 2017). The tutor (in this case the NAO robot, but can be any social robot) plays the main role in the interaction and measures by questions how much the student is rightly constructing its knowledge. The questions can consider objects to be handled by the robot and stories of daily problems to be addressed, characterizing the constructivism.

The ideal scenario is to use this system as practical exercise fixation after the regular classes about the topics' concepts. Each meeting between robot and student is called a session. During a session, the robot presents a concept to the student and evaluate if he/she has understood this explanation by asking questions. For that, every topic that aims to be addressed needs to be registered in the system. The topics have concepts - which is the topic explanation - and as many questions regarding this concept as the designer want to be approached. It is mandatory to divide the questions into five levels of difficulty and at least one question per level in order to guarantee content adaptation. Each session follows the same scheme divided in three phases: Welcome Dialogue, Content Approaching and Closure Dialogue.

In the first meeting, the Welcome Dialogue phase will recognize the student's face and insert it into the users database. If the student is already registered, all his/her information is recovered to be used in the following conversation. Content adaptation is mapped in the Content Approaching phase. The topics' concept is discussed followed by a random number of questions defined by the designer. The questions are chosen in the difficulty level set by the adaptive function in an instant *t* of the interaction. The instant *t* is considered the time to realize a task and it may vary from one task to another. Finally, in the Closure Dialogue, the robot makes a content summary to the student about what was approached in the session. As a feedback, some student's skills, such as average time to respond and correct answers rate, are reported and discussed by the robot. Additionally, some tips about how to improve or keep these skills for the next sessions are presented.

The computational mapping of this methodology is achieved by coding, which is a very unclear process for those who do not have programming knowledge. A graphical interface is presented in the next sections aiming to bring regular teachers closer to robotic solutions for education. The challenge in modeling the contents by following the presented guidelines is a secondary contribution of this research, since it also stimulates the teachers' creativity, pedagogical skills and motivation for new technologies to enclose their methodologies.

4 Graphical User Interface

The proposed GUI was implemented to operate over a cognitive adaptive system. Its code can be accessed in the project GitHub website[5]. The adaptive function's goal is to make the interactions as attractive as possible to the student, based on the indicators read along the session. The designer only needs to set up some variables in the GUI (detailed in Subsection 4.4).

The framework PyQt4[6] was used to the GUI's development. It facilitates the integration with the architecture that is also implemented in Python language. The software is preferably configured for 14 inches monitors and runs in the same window all the time. The system functionalities are handled in the bottom section of this window by changing the tabs, as detailed in the next subsections. For detailed technical development of each module, please check (Tozadore et al., 2017).

In Figure 2, the system interface is shown separated in two sections. The section (1) is related to the activity summary - highlighted in green - is fixed and the designer can use and the section (2) - highlighted in blue - to configure the activity by each functionality. The activity summary (1) is composed of the main menu buttons on the left of this figure, whereas the activity properties in the middle and an activity picture on the right.

Designers need to be registered in the system and sign in into the software for issues tracking. All their actions are registered and can be accessed whenever one wants. Along the software usage, all information is stored in files that can be reused in other activities or shared through storage devices and networks.

[5] https://github.com/LAR-Educational/Architecture_v2_0/tree/master/Arch_2_1/GUI
[6] https://www.riverbankcomputing.com/software/pyqt/download

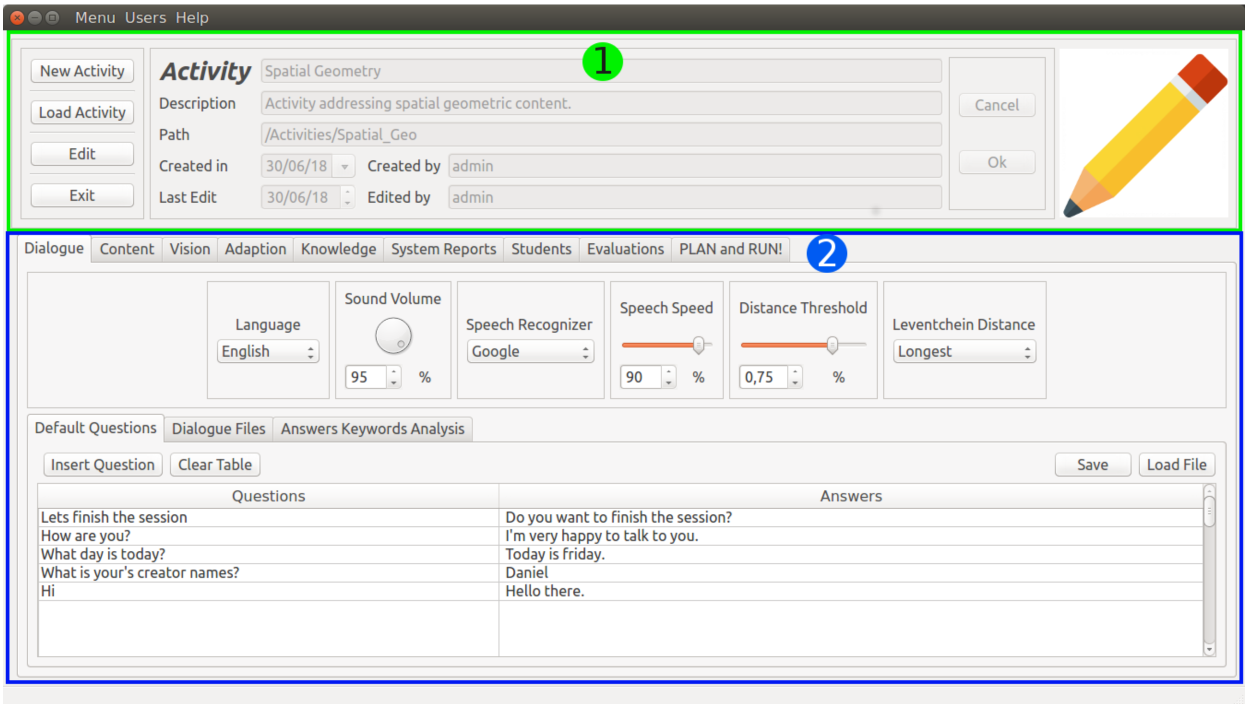

Figure 2: System interface divided in: 1 - the activity summary (fixed); and 2 - the functionalities tab.

The system is divided into the following functionalities, coded in the corresponding tabs.

4.1 Dialogue

In the Dialogue tab, it is possible to configure the system variables that will control the verbal interaction. In this project, all dialogues and expected answers must be registered in the system. This is mandatory because the system uses these sentences to compare their distance to the student answers and, based on a threshold set by the designer, judge if the answer is right or wrong. The chosen comparison algorithm is the Levenshtein distance, which is very used in Natural Language Processing techniques and DNA comparison.

In the Dialogue functionality tab (Figure 2) it is possible to set the components: Language (English or Portuguese), Volume (0 to 100 %), Speech Recognition Method (Google Recognition or NAO's Default), Robot's Speech Speed (0 to 100 %), Levenshtein Distance Threshold (0 to 1) and Levenshtein Distance Method (Longest or Shortest). The bottom section is a frame with tabs responsible for control the tables: "Default Questions" to register possible questions that can be made to the robot in any part of the interaction; "Conversation Set Up" to write Welcome and GoodBye dialogues; and "Answers Keyword Analysis" to set students vocabulary for affirmation, negation and doubt.

4.2 Content

The Content tab allows to create and manage the topics according to the adopted Pedagogical Model (Section 3). The activities can have as many topics as the designer wants and it is also possible to import topics from other activities. The content is defined by topics the designer aims to address during the content approaching phase. As can be seen in Figure 3, topics are easily inserted by clicking on the "New Subject" button. Already registered topics are handled in the corresponding combo box.

Topics' concepts are inserted or displayed in the concept field. The concept is the topic definition and it is exactly what the robot will explain about this topic to the student. Due to resource limitations, the robot exclusively counts on verbal explanation to address the contents. However, future works will include adding a visual display, such as tables or screens, to increase the explanation experience. In the bottom of the content tab, the designer registers the questions of every difficulty level. It is mandatory to register one question of each level and desirable to have as many as possible.

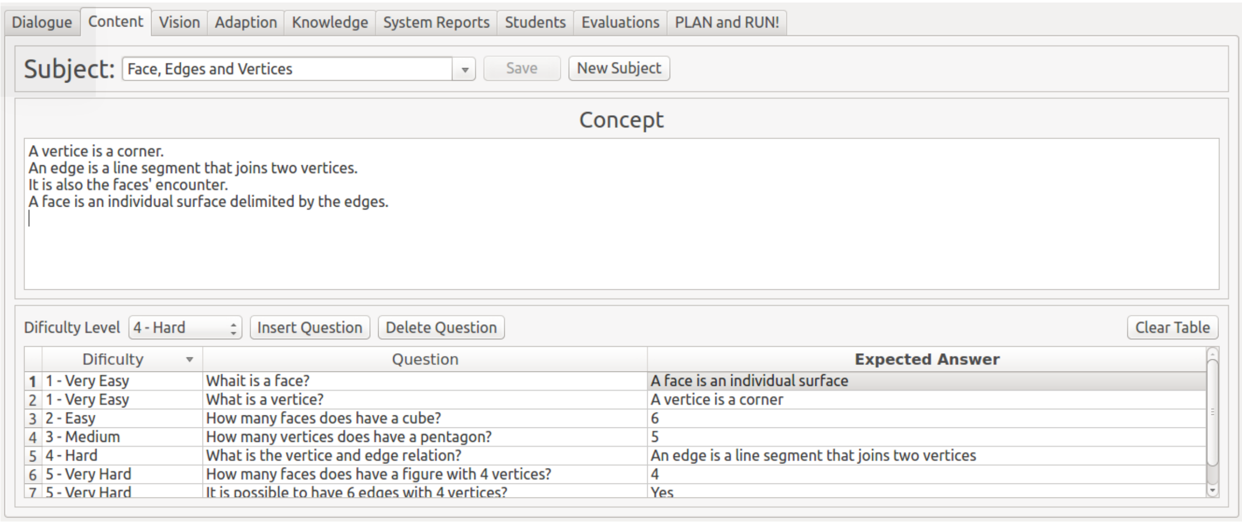

Figure 3: Content tab with 3D geometry example.

4.3 Vision

The Vision module is optional in the activities. It is responsible for recognizing and classifying the objects using Machine Learning methods. The implemented methods are Multilayer Perceptron (MLP) Networks, K-Nearest Neighbor (KNN), Support Vector Machines (SVM) and Convolutional Neural Networks (CNN). Objects database creation or reuse is required for every activity that uses the vision module. Methods train and validate the databases with specific speed and accuracy. For instance, the KNN and SVM have fast training time (time < 10 seconds) and perform it just before the sessions, whereas the MLP and CNN require more training time (time > 5 hours). More details about advantages and setbacks in using these methods can be found in (Tozadore & Romero, 2017). The vision tab has a section to manage the database, a section to manage the classification methods (middle) and a section for samples visualization (right), as can be seen in Figure 4.

4.4 Adaption

The Adaptive module aims to change the robot's behavior according to the observed student's indicators, expressed by body language and verbal answers. These indicators were divided into three main groups regarding the measures of Attention, Communication and Learning.

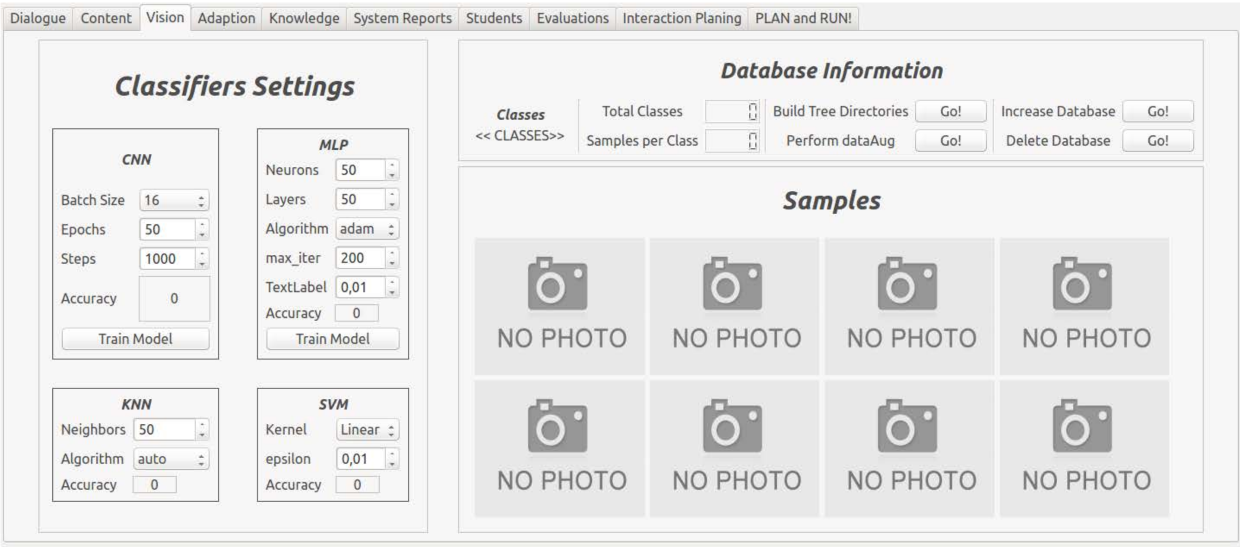

Figure 4: Vision tab view. The available methods and their parameter in the left section and the database information in the right section.

They are: Face gaze for the Attention; users' Emotions for Communication; and Right/Wrong answer and Time to answer the proposed exercises for the Learning group. The average of the objective measures of each group result in a final major value of the class, named as α to *Attention*, β to *Communication* and γ to *Learning*.

Table 1: Caption table 1.

Attention (α)	Communication (β)	Learning (γ)
Face gaze	Number of words	Right/wrong answer
	Emotions	Time to answer

Each major value is calculated by their respective measures' tolerance normalization between 0 and 1. The vectors correspond to the activation value for each weight in the F_{adp} calculation in Equation (1). The maximum limit for the major values is 1 and when applied to the F_{adp} equation, they will fully activate their corresponding weight from the class. 0 means that there was no detection of this class activity in the instant *t*. Thus, this value corresponding weight will not contribute to the F_{adp} in the instant $t+1$. In other words, the classes major values α, β and γ mean how much their respective class is being critical (from 0 for none to 1 for maximum) in the specific instant t, whereas the weights potentialize how much their respective class is contributing to the F_{adp} calculation and their values are the same all along the session. It is worth to notice that the sum of the weights should not be greater than one in order to fit the interval from 0 to 1 for the adaptive function F_{adp}.

The adapted robot's behavior, denoted as Ψ, is an iterative function calculated by its last value added by the function F_{adp}, as shown in Equation (1):

$$F_{adp}(t) = w_\alpha * \alpha(t) \ + w_\beta * \beta(t) \ + \ w_\gamma * \gamma(t) \tag{1}$$

where ***t*** is the instant in which the robot is approaching one question of a topic

The F_{adp} is a function to adapt the resulting robot's behavior Ψ, trying to optimize the interaction engagement and learning rate. The parameters α, β and γ are the group activation function outputs and the w_α, w_β and w_γ are the corresponding weights, set by the designer before the beginning of the session.

In Equation (2), the calculation of the robot's behavior state at the instant ***t*** is presented. The $\Psi(0) = 3$ state guarantees the system starts in neutral behavior.

$$\Psi(t) = \begin{cases} 3, & \text{if } t = 0 \\ \Psi(t-1)+1, & \text{if } F_{Adp}(t) \geq 0,66 \\ \Psi(t-1)+0, & \text{if } 0,66 < F_{Adp}(t) < 0,33 \\ \Psi(t-1)-1, & \text{if } F_{Adp}(t) \leq 0,33 \end{cases} \quad ,t \in \mathbb{N} \qquad (2)$$

Regarding the adaptive tab in the GUI, the values α, β and γ are set in the corresponding section as can be seen in Figure 5.

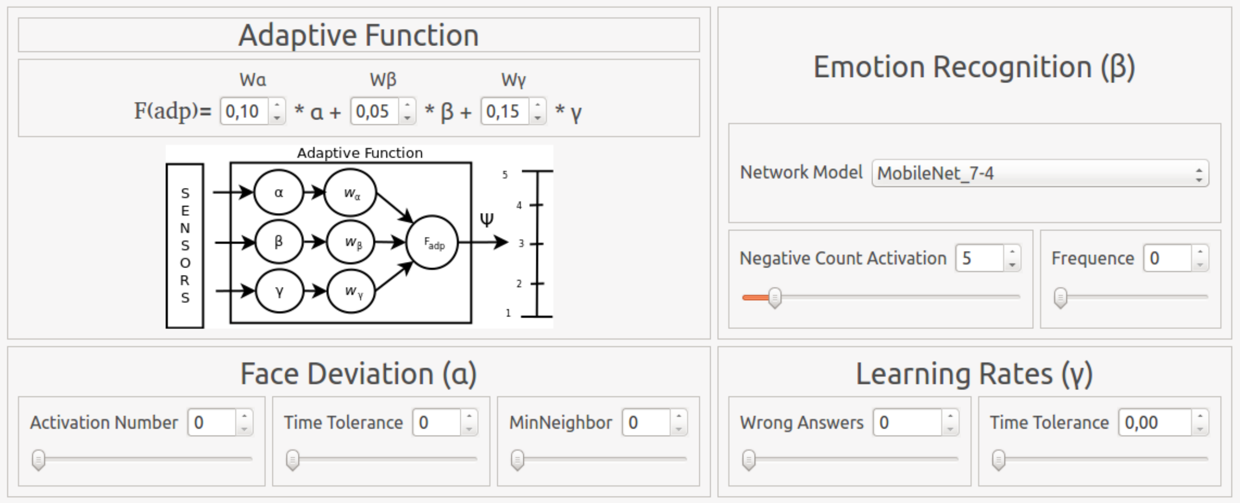

Figure 5: Adaptive tab view.

The face gaze is handled by a Haar Cascade algorithm in which the "MinNeighbor" and the "Time Tolerance" are considered to count a deviation (Viola & Jones, 2001). The "Activation Number" is the sum of the deviation counter in the end of instant ***t*** and is the variable that will be normalized to activate the weight w_α. The communication vector is the sum of the emotions recognized and number of words spoken by the student. The average of both measures is normalized and used to activate the weight w_β. Six different models of Convolutional Neural Networks were compared to classify the emotions, as it can be found in (Tozadore, Ranieri, Nardari, Guizillini, & Romero, 2018). "Counter Threshold" is the number of negative emotions detection to maximum activation. The number of words is counted by the Dialogue module. Leaning vector measures are provided by the Dialogue module. Its activation value is also the average of the "Wrong Answers" and "Time to Response" thresholds, which will influence the weight w_β.

4.5 Student Database

Students database stores personal information, such as, first name, surname, birthday date, school year, and 8 personal preferences as sport, dance, team, music, toy, hobby, game and food. The system searches in the Knowledge database about the choices of the students and uses them in talks as it is necessary, aiming to simulate long-term relations. These talks are performed according to the interaction flow (programmed in Section 4.8) or triggered by the adaptive module if it detects signs of students' low engagement. This information is collected by the robot in its first meeting with the student in short dialogues. Insertions, modifications or deletions can be made in the User tab, as shown in Figure 6.

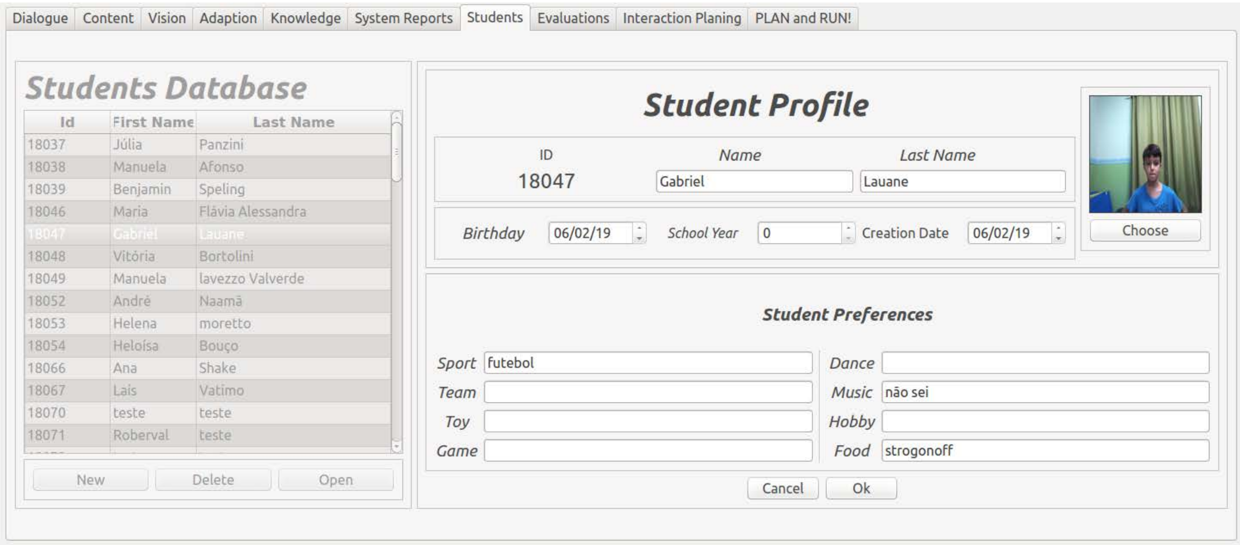

Figure 6: Students window. On the left side of the screen is the students database and in the right section the selected student's information is shown.

4.6 Knowledge

Similarly, in the knowledge database of the system all the nouns definitions are recorded through manual insertion or through automatic searched by the system on the internet. It is possible to insert content into the system through the Knowledge tab (Figure 7) or the system can search by them during the interaction. In case of new entries, the system searches in the Wikipedia website[7] through a python API. A small language processing technique is employed to extract the noun abstract. The personal database stores information created about the robot "personal life" is also stored, such as how old he/she is, how many brothers and sisters he/she has, what is his/her name and so on. Previous studies showed significantly higher rapport building when humanizing the robot (Pinto, Tozadore, & Romero, 2015).

4.7 Evaluation

Once the sessions are individual, it can be searched by the student that performed the activity in the evaluation database and this activity evaluation summary will be shown in the Evaluation Tab (Figure 8). The evaluation overview shows the Student Name, Execution Date, Supervisor (the designer that performed this activity), the time it was executed and a student picture at the section top. It also shows information that are relevant to the session, for instance, the total number of content questions asked by the robot, maximum number of attempts per question, student correctness rate and observations added by the supervisor after the sessions ended. In the right section, a multi-tab frame allows to navigate through the Topics Validation, a Timeline Session Evolution, the adaptive system metrics auto analysis and some pictures taken by the robot during the session.

[7] www.wikipedia.com

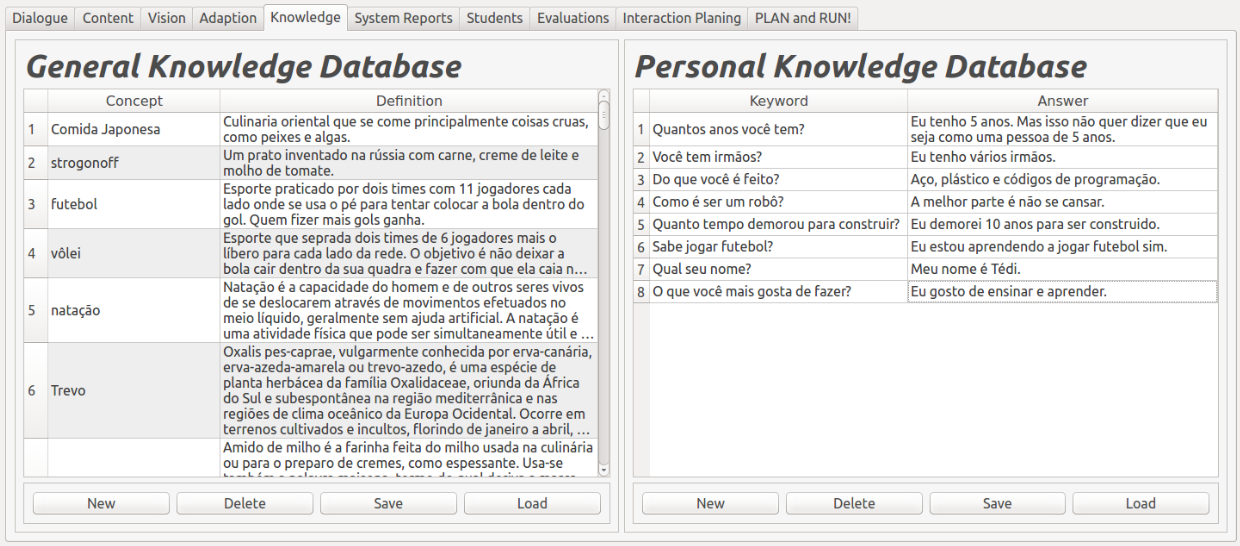

Figure 7: Knowledge database management. The left section is the general knowledge entries and in the right section the robot's "personal information".

The Topics Validation tab recovers all the content information exchanged between robot and the student along the corresponding session for validation. The supervisor needs to manually observe and validate if the system automatically classified the student answers right or wrong. Afterwards the questions validation, the system can generate an evaluation regarding its accuracy and the student performance.

This information about system and student performance in the interaction is shown as charts in the Information tab. It is considered one of the advantages of this system, because it is a helpful type of diagnostic about the student difficulty, in which the teachers can visualize the critical learning areas through charts and prepare the next activities focused in last sessions point of weakness.

4.8 Interaction Planning

The Interaction Planning tab is responsible for programming the interaction flow. It means the sequence of commands that will be executed by the robot during the session. There are three types of actions that the robot can perform, as shown in the left boxes of Figure 9: Content approach, Personalized talks and Extra actions.

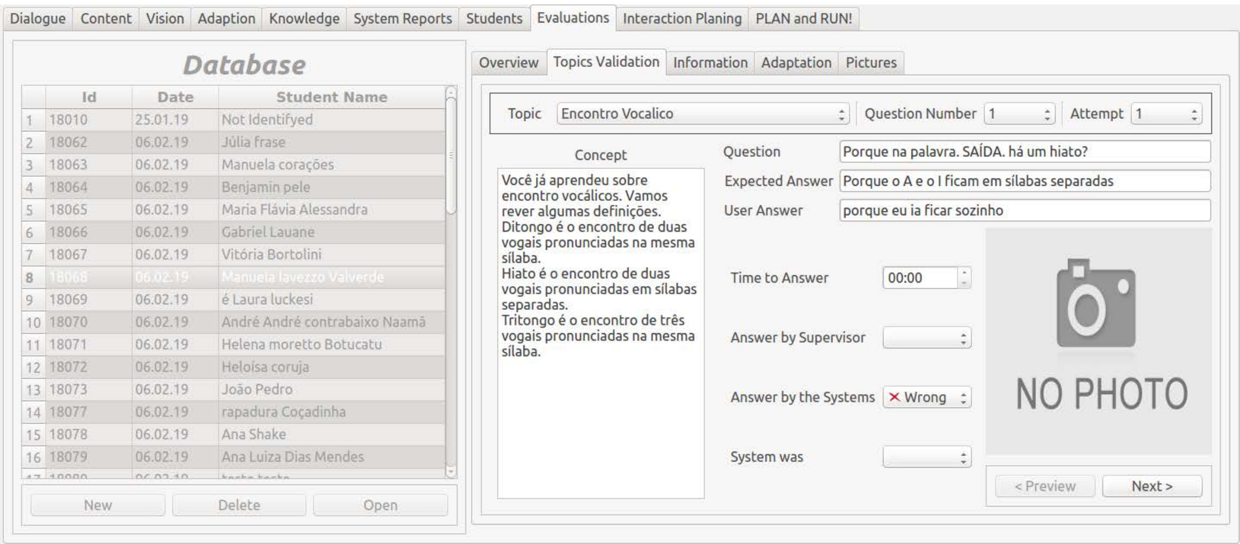

Figure 8: Evaluation tab. The left section is the evaluations database and the right section shows the selected evaluation's information.

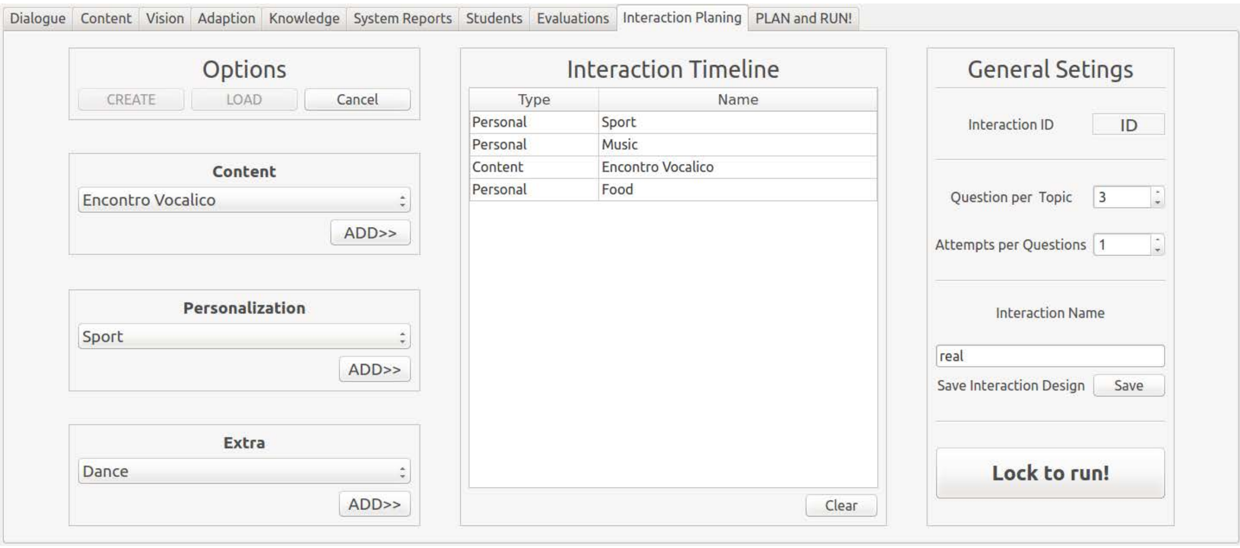

Figure 9: Interaction tab. The sequence of interaction steps is chosen by the designer.

In the Content approach actions, the robot will ask about the number of questions as set in the *number of questions* of the selected content. For each question, the student will have the maximum number of attempts set in the *attempts number*. All available contents to be approached are registered in the system through the content tab.

The Personalized talks are made of a small dialogue, in which the robot will ask the student what is his/her preference about the selected theme. The system searches for the student's answer in the local knowledge database and talks about it, if found. Otherwise, the system searches in a Wikipedia page, as explained in Section 4.6. If this conversation has already happened before, the system recovers the student's preferences and takes the same action as if he/she had answered that preference. These preferences can be manually set for each user, as shown in Figure 6.

The Extra actions include Dances, Games or any other performances that are programmed in the robot behavior set. The actions are added to the timeline table, in the center of the screen, and the session will flow in this exact order.

4.9 Run

After setting the described configurations, the Plan and Run tab (Figure 10) provides the session high-level scheme to the designer. In the left panels, the robot IP address and the robot communication port are required to initiate the session. It is possible to observe in execution time some system variables that change along the session, such as the adaptive function parameters, the robot behavior Ψ, previous values of robot behavior, user emotion read by the system, among others.

After these settings, the "Start" button is enabled. By starting the session, some variables that change over time are tracked in the bottom left section of the window in corresponding frames. The middle section of the screen displays the robot's camera image. Three fields of information tracking are placed below the robot view. They are the expected and understood answers field, that displays the corresponding information, and the user manual answer, in which is possible to insert the answers through the keyboard, if this option is enabled in the dialogue tab.

In the right panels, it is possible to force the system flow, jumping through the questions and topics - which is used is extreme cases - and take pictures and save videos from the robot's camera. In the bottom of these sections is the field corresponding to the robot's talks, that shows what the robot is speaking.

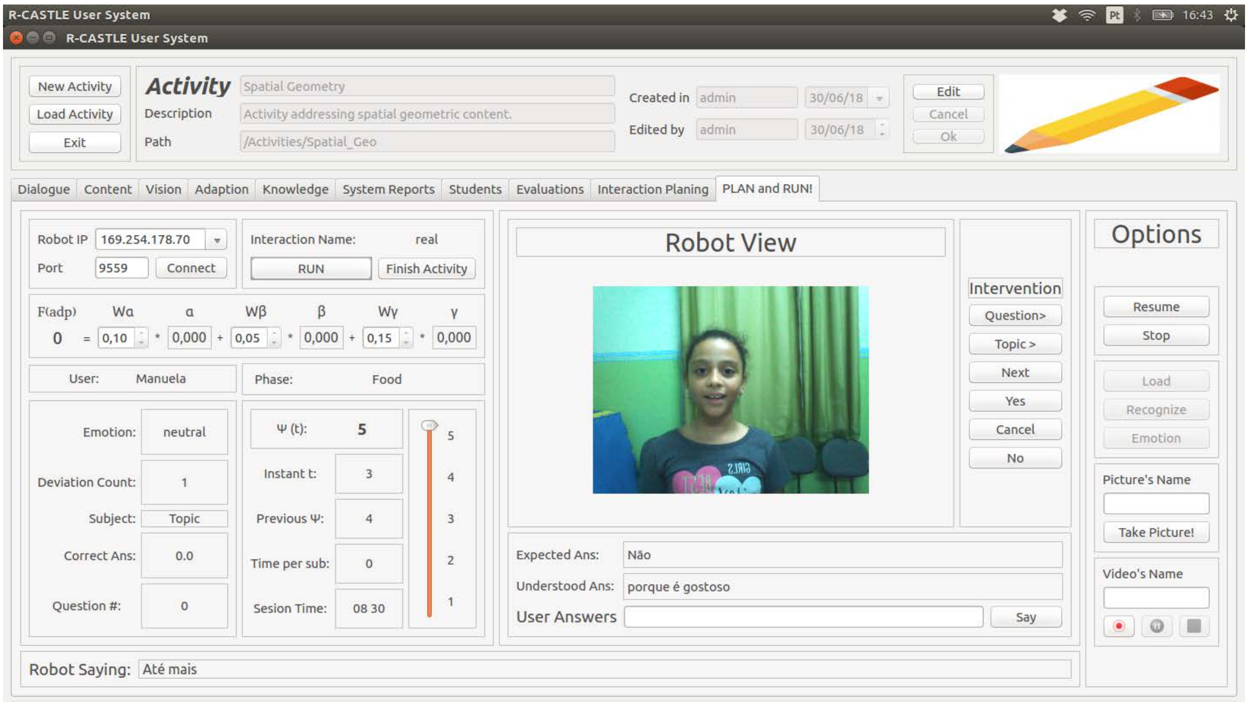

Figure 10: Plan and Run tab: The monitored variables in the left, the robot view in the middle and the extra actions panel in the right.

5 Teachers' perception

In order to analyze the teachers' first perception about the system's potential, 14 teachers from the elementary school "Oca dos Curumins", located in São Carlos city, participated in a 50 minutes demonstration about the GUI. The presentation approached the GUI components that are more relevant to the teacher's usage. As said previously, in general, it is expected that the teachers are more interested in some functionalities of the system such as the Content programming, the Student profiles and the Evaluations performed, than other GUI functionalities that may contain information more relevant to the developers, such as the vision, adaptation and system's log. During the presentation, explicitly mentioning the GUI advantages or disadvantages was avoided, to avoid inducing the participants (the teachers) to a bias in their perception about the whole system. In other words, only functionalities of the system and their use were explained, letting the perception of how the system can be useful to their own judgment. The teachers that wanted to interact with the interface could do so at the end of the explanation.

The aim was to evaluate the system strength of generalization, by giving an overall explanation, supporting the principle that the methodology can cover the areas of knowledge of the teachers. Finally, 9 out 14 of those teachers (2 males and 7 females) aged in average 43.2 (SD 12.7) years old, answered a 5-Likert Scale questionnaire about the GUI. The Likert Scale is very common in research about marketing that aim to know how much the customers would enjoy some products' characteristics. The answer's possibilities were a scale from 1 to 5, where 1 means *Absolutely Nothing* and 5 means *For Sure* for every item. Thus, 3 is the neutral score. The following questions were asked:

1. How much are you familiarized with technology?
2. How much do you think this application can support you in your class activities?
3. How much do you think it will be easy to create activities in this tool?
4. How much do you think this application has powerful to become a regular tool in teaching?
5. How much do you think this methodology is efficient for content approaching?

6. How much time do you think it is necessary to get used with this GUI?
7. How much do you think this methodology can enhance the students' learning?
8. How much do you think that technological methodologies are more efficient compared to the traditional ones?

The average score for each item can be seen in Figure 11. Beyond the 8 questions above, 2 open questions were optional. The points raised in these questions were computed and their number (n) of occurrence is shown below:

A. What are the system advantages from your point of view?

   Answers: Motivation (n=5), Adaptation (n=3), Technological Interaction (n=3), Teacher Support (n=2), Data Handling (n=1).

B. What are the system disadvantages from your point of view?

   Answers: Number of Variables (n=5), Robot Cost (n=2), Teachers' Adaptation (n=1).

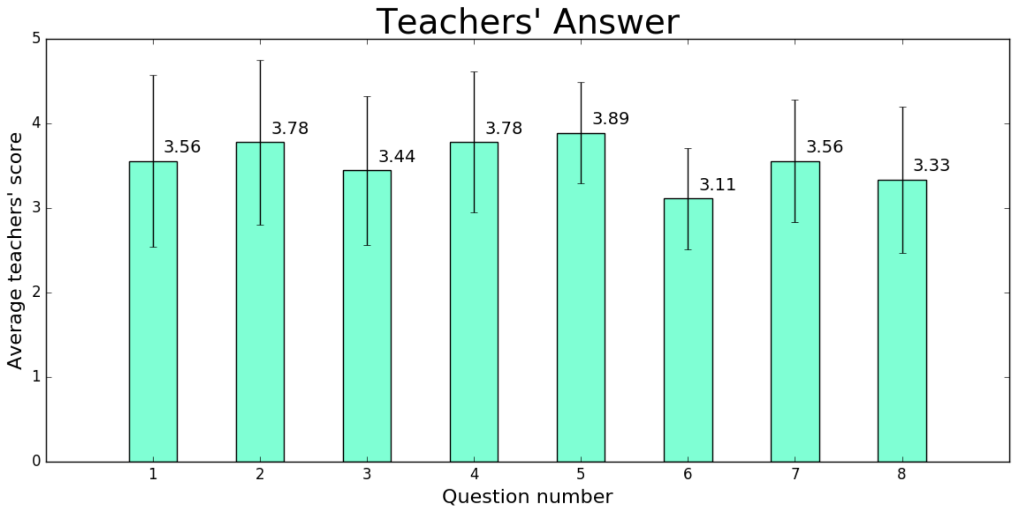

Figure 11: Teachers’ score average for Likert Items.

By analyzing the results, a curious fact noted was that they scored the facility in programming of the exercises in the GUI (Item 3) lower than they claimed they are familiarized with technology (Item 1). However, they show they understood the GUI's potential in supporting their classes (Item 2) with one of the highest averages of 3.78. The same score as they rated the system potential in becoming a regular tool in teaching (Item 4).

The highest score was achieved by the item that asked about the methodology efficiency in content approaching (Item 5). One fact that may explain this point is the similarity of the system pedagogical model with the regular methodology the school have. Thus, to adapt their regular methodology to the system would not be a difficulty for them.

They rated the Item 6, regarding the time to get used with the system, with a score of 3.11, which is very close to the neutral score. It suggests that they would take the same time to get used with this system as they would have to get used with another educational methodology.

The most surprising score was at Question 8, about this methodology efficiency compared to regular methods. It was slightly above the neutral score (3.33). Assuming the difficulty in changing consolidate methods, this score may raise a point in the sense of how teachers are noticing the necessity in starting to migrate from non-technological methods to innovating solutions.

Regarding the open questions, the most cited advantage of the system was the motivation that the robots provide, which is not novelty as shown by several researches in the area (Leite et al., 2013). The system's behavior adaptation was also noted as a strong point by some teachers, and only one participant claimed the facility in data handling as a notable advantage. Many of the participants pointed the amounts of variables that the system deals with as a potential problem to their usage. They said the adaptation system and its weight management is something they would take more time to understand. The robot cost was also cited as a disadvantage, because the robot used in the demonstration was the NAO robot, acquired from a FAPESP project by the host University. Nonetheless, the system can run along any interactive device that has cameras and microphones.

In summary, all the evaluated items had a score above the neutral score (3), but few of them got close to have an average in 4. This fact leads to believe that the system's potential was not fully perceived by all the participants that answered the questionnaire. Potentially, because they need more time for getting familiarity with the GUI and to know all the system's practices. However, the system advantages planned by the researchers were cited at least by one of the professors that answered the questionnaire. According to the teachers' statements themselves, perhaps more familiarization with the interface by using and testing it may result in perceptions more accurate about how this system can actually support their classes.

6 Conclusion

This paper showed the proposed interface for educational content management, providing autonomously robot behavior and questions difficulty adaptation. The system tries to simulate rapport building by keeping a database of students' preferences about daily subjects to interact with them along the session. Through face recognition, the system can support long-term interactions, retrieving the interactions optimization parameters for each user and approaching the programmed contents according to these variables. Following the adopted methodology, it is possible to set up the robotic system to interact in almost every area of elementary school subject. The proposed GUI is an alternative to work around the problem of lack of familiarization with social robots by non-programmers. After a short presentation of the interface to a group of regular teachers from elementary school, it was possible to identify their feedback as positive in the application of the proposed system as a pedagogical tool. All teachers of this study perceived some potential in using the presented system that can be enhanced by the time of usage and familiarization with the system variables. Future work includes performing several activities with the robot and a group of teachers and students to investigate the impact of the proposed system in their performance along their regular academic year.

Acknowledgements

This study was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001, the Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP) and the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq).

References

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018a). Social robots for education: A review. *Science Robotics*, *3*(21), DOI:10.1126/scirobotics.aat5954 [GS Search]

Belpaeme, T., Vogt, P., Van den Berghe, R., Bergmann, K., Göksun, T., De Haas, M. & Papadopoulos, F. (2018b). Guidelines for designing social robots as second language tutors. *International Journal of Social Robotics*, *10*(3), 325-341. DOI: 10.1007/s12369-018-0467-6 [GS Search]

Benitti, F. B. V. (2012). Exploring the educational potential of robotics in schools: A systematic review. *Computers & Education*, *58*(3), 978-988. DOI: 10.1016/j.compedu.2011.10.006 [GS Search]

Clabaugh, C., Tsiakas, K., & Mataric, M. (2017). Predicting preschool mathematics performance of children with a socially assistive robot tutor. In *Proceedings of the Synergies between Learning and Interaction Workshop@ IROS, Vancouver, BC, Canada* (pp. 24-28). [GS Search]

Cortellessa, G., Fracasso, F., Sorrentino, A., Orlandini, A., Bernardi, G., Coraci, L. & Cesta, A. (2018). ROBIN, a telepresence robot to support older users monitoring and social inclusion: development and evaluation. *Telemedicine and e-Health*, *24*(2), 145-154. DOI: 10.1089/tmj.2016.0258 [GS Search]

Gao, A. Y., Barendregt, W., & Castellano, G. (2017). Personalised human-robot co-adaptation in instructional settings using reinforcement learning. *In IVA Workshop on Persuasive Embodied Agents for Behavior Change*: PEACH 2017, August 27, Stockholm, Sweden [GS Search]

Goodrich, M. A., & Schultz, A. C. (2008). Human–robot interaction: a survey. *Foundations and Trends in Human–Computer Interaction*, *1*(3), 203-275. DOI: 10.1561/1100000005 [GS Search]

Johal, W., Castellano, G., Tanaka, F., & Okita, S. (2018). Robots for learning. *International Journal of Social Robotics*, 293–294. DOI: 10.1007/s12369-018-0481-8 [GS Search]

Jones, A., & Castellano, G. (2018). Adaptive robotic tutors that support self-regulated learning: A longer-term investigation with primary school children. *International Journal of Social Robotics*, *10*(3), 357-370. DOI: 10.1007/s12369-017-0458-z [GS Search]

Kafai, Y., Sawyer, C. I. R., Papert, S., Harel, S. C. I. I., Papert, S., & Duval, E. (2017). Technology and theories of learning. *Technology Enhanced Learning: Research Themes*, *17*(1), 169. [GS Search]

Kessous, L., Castellano, G., & Caridakis, G. (2010). Multimodal emotion recognition in speech-based interaction using facial expression, body gesture and acoustic analysis. *Journal on Multimodal User Interfaces*, *3*(1-2), 33-48. DOI: 10.1007/s12193-009-0025-5 [GS Search]

Leite, I., Martinho, C., & Paiva, A. (2013). Social robots for long-term interaction: a survey. *International Journal of Social Robotics*, *5*(2), 291-308. DOI 10.1007/s12369-013-0178-y [GS Search]

Lucas, G. M., Boberg, J., Traum, D., Artstein, R., Gratch, J., Gainer, A. & Nakano, M. (2018, February). Getting to know each other: The role of social dialogue in recovery from errors in social robots. In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction* (pp. 344-351). ACM. DOI: 10.1145/3171221.3171258 [GS Search]

Mubin, O., Stevens, C. J., Shahid, S., Al Mahmud, A., & Dong, J. J. (2013). A review of the applicability of robots in education. *Journal of Technology in Education and Learning*, *1*(209-0015), 13. [GS Search]

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 98-129. [GS Search]

Pinto, A. H., Tozadore, D. C., & Romero, R. A. (2015, October). A question game for children aiming the geometrical figures learning by using a humanoid robot. In *2015 12th Latin American Robotics Symposium and 2015 3rd Brazilian Symposium on Robotics (LARS-SBR)* (pp. 228-233). IEEE. DOI: 10.1109/LARS-SBR.2015.62 [GS Search]

Platz, M., Krieger, M., Niehaus, E., & Winter, K. (2018). Suggestion of an E-proof Environment in Mathematics Education. In *Classroom Assessment in Mathematics* (pp. 107-120). Springer, Cham. DOI: 10.1007/978-3-319-73748-5_8 [GS Search]

Rivas, D., Alvarez, M., Velasco, P., Mamarandi, J., Carrillo-Medina, J. L., Bautista, V. & Huerta, M. (2015, February). BRACON: Control system for a robotic arm with 6 degrees of freedom for education systems. In *2015 6th International Conference on Automation, Robotics and Applications (ICARA)* (pp. 358-363). IEEE. DOI: 10.1109/ICARA.2015.7081174 [GS Search]

Shiomi, M., Kanda, T., Ishiguro, H., & Hagita, N. (2006, March). Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction* (pp. 305-312). ACM. DOI: 10.1145/1121241.1121293 [GS Search]

Spaulding, S., Chen, H., Ali, S., Kulinski, M., & Breazeal, C. (2018, July). A Social Robot System for Modeling Children's Word Pronunciation: Socially Interactive Agents Track. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems* (pp. 1658-1666). International Foundation for Autonomous Agents and Multiagent Systems. [GS Search]

Tozadore, D., Ranieri, C., Nardari, G., Guizilini, V., & Romero, R. (2018, October). Effects of Emotion Grouping for Recognition in Human-Robot Interactions. In *2018 7th Brazilian Conference on Intelligent Systems (BRACIS)* (pp. 438-443). IEEE. DOI: 10.1109/BRACIS.2018.00082 [GS Search]

Tozadore, D. C., Valentini, J. P. H., de Souza Rodrigues, V. H., Vendrameto, F. M. L., Zavarizz, R. G., & Romero, R. A. F. (2018, November). Towards Adaptation and Personalization in Task Based on Human-Robot Interaction. In *2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE)* (pp. 383-389). IEEE. DOI: 10.1109/LARS/SBR/WRE.2018.00075 [GS Search]

Viola, P., & Jones, M. (2001). Rapid object detection using a boosted cascade of simple features. *CVPR (1)*, *1*, 511-518. [GS Search]

CHAPTER

5

WHEN SOCIAL ADAPTIVE ROBOTS MEET SCHOOL ENVIRONMENTS

Contribution statement

D.C. Tozadore conceived proposal scheme of interaction of Figure 1. D.C. Tozadore and R. A. F. Romero performed the literature review of the Section 2. D.C. Tozadore performed the experiments with the teachers, resulted in the images of Figure 2 and Table I. D.C. Tozadore, J. P. H. Valentini, V. H. Rodrigues, and J. Pazinni performed the tests with the students and analyzed its results presented in Table 2 Figure 3. All the authors contributed on writing the manuscript. R. A. F. Romero,in addition, supervised all this research work.

Presentation

The paper in this chapter presents studies with the robot in autonomous operation interacting with students in a real environment, approaching the content programmed by their teachers. An end-to-end experiment testing the experience of teachers with the GUI described in Chapter 4 and the architecture to autonomously run the robot interacting with the students, are presented in Chapter 3. The experiment analysis presents the performances of the system and the students and the perceptions of students and teachers regarding their experience with R-CASTLE.

When Social Adaptive Robots meet School Environments

Completed Research

Daniel C Tozadore
University of São Paulo
tozadore@usp.br

João P H Valentini
University of São Paulo
joao.valentini@usp.br

Victor H Rodrigues
University of São Paulo
victor.henrique.rodrigues@usp.br

Julia Pazzini
University of São Paulo
julia.pazzini@usp.br

Roseli A F Romero
University of São Paulo
rafrance@icmc.usp.br

Abstract

Few robotic adaptive architectures can be programmed by the teachers and used in school environments as a regular tool for teaching, compared to other information systems. This paper describes an experiment in a school environment. The experiment aimed to analyze the benefits of a cognitive adaptive system, which controls a social robot for educational activities, for teaching. The system performance was evaluated based on three criteria: (i) teachers perceptions in using the system, (ii) students perception with regard to the robot and (iii) the system accuracy. Teachers and students perceptions were obtained by means of questionnaires and the system's accuracy was evaluated by manual validation. Overall, the outcomes suggest that the teachers and students identified the benefits of the system for teaching. Furthermore, the system's classifications of verbal answers showed itself efficient when the answers were rightly understood by the speech recognizer.

Keywords

Human-Robot Interaction, Cognitive Robots, Social Educational Robots.

Introduction

Information Systems are commonly used in the educational field to support teachers and students and their advantages are sensible for both sides and since a long time ago (Murray 1999). To the teachers, they can provide tools for creating and approaching theoretical and practical exercises, provide information about the students' performance in these exercises. Whereas the students can take benefit from smart systems that count on difficulty adaptive function and also provide technological novelties.

Regarding social robotics to educational purpose, some disadvantages can be noted compared to other information systems, such as the small number of students that can participate in the interactions with the robot. Another disadvantage is that the students' regular teachers can rarely program the content and have access to the information. One of the main reasons is the lack of preparation of the teachers in using robotics devices in their teaching plan, due to the complexity in programming such systems (Johal et al. 2018). Even after being programmed, the information collected or generated by the student-robot interaction is usually used only by programming experts and the teachers may never have access to them.

Here, a cognitive adaptive system (Tozadore et al. 2018) which can provide adaptation and personalization through open source techniques was developed. It has autonomous analysis and recognition of speech. After creating and running an activity, all sessions' evaluation and information can be accessed for visual analysis, as well as students' preferences throughout the interaction. The system's goal is to intermediate the teachers and the agent, in this case, a NAO robot, to build complete interactions with no programming skill needs. It respects the idea of educational evaluation introduced by Luckesi (Luckesi 2014), which aims a social transformation, not its conservation, considering the learning process of each student.

In this paper, we analyzed how the combination of an Information System with a Cognitive Robotic System can impact both teachers and students in content approaching through pedagogical interactive activities. To validate the system, a set of Brazilian Portuguese grammar exercises was programmed by two teachers in the system Graphical User Interface (GUI) and performed with the students and an interactive robot. The exercises were about the content being addressing in classes by the teachers and approached by the robot in individual sessions.

The robot was controlled by the introduced system that also used rapport building, aiming to ease the children's shyness in interacting with the NAO and possible students demotivation caused by adaptation in the content difficulty. We also analyzed the students' acceptance via survey whether the Human-Robot Personality-Similarity (Robert 2018) is influenced by adapting the robot's behavior in the second meeting in which it said it likes the same things the student said he/she liked in the first meeting.

Related Work

In systems that autonomously interact for long periods with humans, it is essential that its behavior being adapted. Works that adopted this strategy constantly reported a enhance in the user experience. For instance, in a study (Leyzberg et al. 2014) in which the robot was assigned for tutoring a student during tasks that consisted in solve a logic challenge, the result showed that the exercises' resolution times were smaller in the group where there was an adaptive behavior of the robot when compared to the control group.

Adaptive robots can also leads to a more enjoyable experience by itself. Specially to drew attention of the users when tutoring (Vouloutsi et al. 2015). It was found a social adaptive robot can drew more attention from the users than an non-adaptive robot or even a human. Moreover, this method is still more effective for children than adults.

Rapport building and the feeling that the robot is building a relationship with the user is another strategy often used to enhance learning experiences. Some results showed that students tends to put more effort in educational tasks in order to do not disappoint the tutoring robot after create a friendship with it (Brown et al. 2013). This positive impact can also be seeing in group interactions in which teams performed better and were more viable when they were emotionally attached to their robot (You and Robert 2017). Robert (2018) pointed out that humans respond more favorably to extroverted robots, but this relationship is moderated, and humans respond favorably to robots with similar and/or different personalities from them.

However, there is a trade-off needed between content approaching and the robot's intervention in order to do not distract the student. Furthermore, studies have shown that children often do not make the best use of on-demand support and either rely too much on the help function or avoid using help altogether, both resulting in suboptimal learning. In general, the use of robots in educational tasks have shown positives results. Besides potentialize the students learning experience, robots can free up precious time for human teachers, allowing the teacher to focus on what people still do best: providing a comprehensive, empathic, and rewarding educational experience (Belpaeme et al. 2018).

Finally, we identified that the majority of the related studies did not take into account the teachers in their activity planning and evaluations. Thus, we also analyzed their perception of using such systems in this study.

System Scheme

The robotic architecture is a cognitive adaptive system that encapsulates a module based implementation to run over the robot's sensors and actuators. It is also designed for the robot cognition as well. In other words, it is expected the robot also learn from the interaction and use this learning in the next sessions. The agents' interaction flow overview provided by this architecture is shown in Figure 1. The system collects data from the students and uses it both in the current session and in the next ones. These data can be provided voluntarily by the students - such as the verbal answers - or read by the system - such as facial gaze recognition. The majority of them can be used by the adaptive function and other data, such as their personal preferences are used to simulate rapport building in next sessions. The data are stored and the system recovers and uses it after recognizing the user in the next sessions. Moreover, it can ask about some users' preferences and use them in the next conversations. These preferences are manually chosen in the GUI and they may be sport, dance, team, music, toy, hobby, game and food. Every new entry is searched on the internet for their meaning and stored it in the architecture's database.

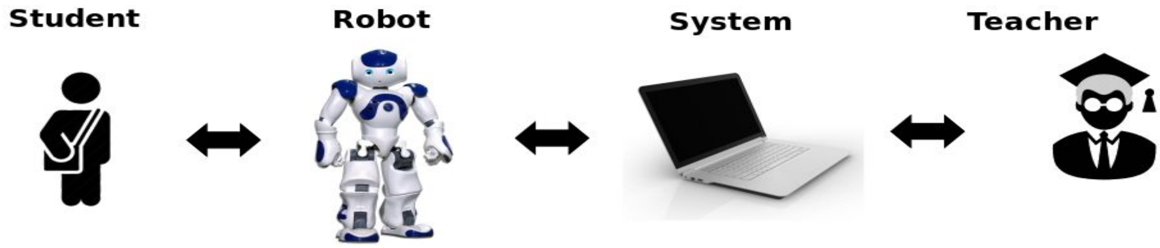

Figure 1: Agents information exchange flow.

The system's dialogue module is the system most important module. It is responsible for the information exchange between the system and the student during the interaction. Three functions compose this module core: read verbal information from the user, processing and understanding the read information and providing verbal information to the user. The user verbal information reading is made by the python SpeechRecognition library in Portuguese-BR idiom. It is one of the best Portuguese speech recognizers. However, it presents a delay to set up the microphone in every requisition and it is noisy sensible sometimes as well.

Sentences interpretation is achieved by a combination of Natural Language Processing (NLP) methods such as the Levenshtein Distance and an embeddings method. It basically compares two strings and returns a distance between them, where 0 means the sentences are the same and 1 means the sentences are totally different. A threshold is configured to accept a user answer as right or wrong compared to a registered expected answer.

The Levenshtein Distance approach is based on the classical implementation of this algorithm and uses the longest alignment between the two sentences, while the embeddings approach is based on the implementation suggested for the ASSIN competition (Hartmann 2016), in which this technique is combined with the TD-IF method. The result is an algorithm that has satisfactory performance in a sparse set of data, such as the user's sentences which the dialogue module is, usually, exposed. The most significant difference between the two approaches resides on the robustness of the methods. When we used the Edit distance the comparison between to sentences stays only into the syntactic field and this presents superficial results, as shown at (Omitted for blinding review), whereas the embeddings plus TD-IF reaches the similarity problem not only syntactically but semantically as well. A relevant weakness noted on this combination is not related directly to itself but to the verbal information caption. When the system does not discern the correct input, because of the pronunciation similarity, the classification tends to be useless.

The dialogue module can alternate among both algorithms by the virtue of some input sentences that can be composed solely by Stopwords. In this situation, the embeddings approach fails because stopwords are removed in the preprocessing stage and the edit distance has shown itself efficient on sentences of less complexity.

The pedagogical model is based on constructivism, as the educational robotic in general (Kafai 2017). The NAO robot plays the main role in the interaction and measure by questions how much the student is rightly constructing its knowledge. The goal is to use this system as practical exercise fixation after regular classes about the topics' concepts. Each meeting between robot and student is called a session. During a session, the robot presents a concept to the student and evaluate if he/she had understood this explanation by asking questions. Every topic aimed to be addressed needs to be registered in the system. The concepts are the topics explanation - and as many questions as the teacher wants to approach. It is mandatory to divide the questions into five levels of difficulty and, at least, one question per level.

The goal of the adaptive function is to make the interactions as attractive as possible to the student, based on the indicators read along the session. The system behavior adapting can take into account three measures read by observing body signals and verbal answers from the user which is interacting with the robot. It can use various reads to adapt the robot's behavior such as face gaze, posture, amount of spoken words per answer, time to answer and answer accuracy. In this case, only the answer correctness was considered to adapt the system and only the questions difficulty was altered. Beyond the importance of respect the following teaching strategies, the levels of difficulty are considered so the students don't have to answer questions that are not in their current level, which could make the student-robot interaction exhausting. In this case, the robot's behavior adapting is considered only the change in the difficulty of the questions the robot is asking.

Experiment

The system was initially tested in laboratory conditions, where the scenarios are considered ideals. It means the interactions were tested in an environment with no noise and performed by users that were already familiarized with the system limitations. So far, its impact with final users - teachers and students - was still unknown. In order to validate the system's data visualization, the programming content and the data visualization by the teachers, an experiment was designed and carried out. This experiment was the first trial with the robot outside the laboratory conditions. Exclusively visual and sound resources were considered in this experiment. The threshold to accept the users' answers as a right answer was set in 0.4

Thus, three independent variables were observed and analyzed in this study:

- Teachers acceptance in using the system's GUI.
- Students' perceptions in interacting with the robot.
- System accuracy in classifying the answers and adapt its behavior (difficulty level).

Further details with regard to the experiment settings are provided in the following sections.

Design

This experiment was divided in 7 steps:

1. *System GUI presentation:* one of the authors presented to 15 teachers of the school how the system should be used/works.
2. *First content programming:* 2 teachers, who participated of the presentation, programmed an activity in the system (Figure 2(a)).
3. *First robot-student interactions:* the robot performed the programmed activities with the students by the first time (Figure 2(a) and 2(b)).
4. *First results analysis and system adjustments:* results of the first interaction were analysed and some adjustments were made in the system.
5. *Second content programming:* evaluation of students performance was shown to the teachers and a second content was programmed by the teachers.

6. *Second robot-student interactions:* the robot performed the second programmed activities with the students by the second time.
7. *Final analysis:* results of the second interaction were analysed and shown to the teachers that answered a questionnaire.

(a) (b) (c)

Figure 2: (a) Teachers programming the content and (b) (c) students performing the activities.

Participants

The participants consisted of 2 female teachers (33 and 54 y.o.) and 32 students (M = 14, F = 18) from the 5th grade with an average age of 9.6 ($\pm$0.49) years old. The selection criteria were all the students that wanted to participate of the sessions and were allowed by their parents.

Interaction Scheme

For the experiment, the teachers programmed two contents in the system, i.e., *Vowel Encounter* and Digraph . The first topic (*Vowel Encounter)* was approached in the two sessions and the second content (Digraph) was only approached in the second session. In general, the questions were about to classify the Vowel Encounter category of a specific word. In the second session, if the words had Vowel or consonant digraph and its category.

The interactions followed a simple scheme. First, the students were called one by one in the classroom where they were developing their regular school activities. The students were head to the school library, where the robot and the researchers were waiting to start the activity. Some instructions were given by one of the researchers, such as how to get the perfect timing to answer the robot, speak louder if the robot did not understand the answer, and so on. Second, the student started the session, the researchers step back and took place behind the table, where the robot was. Each session followed the same scheme in which the robot initiated a dialogue to introduce itself. In the first meet, the robot recognized the student's face and inserted it into the users database. If the student was already registered, all his/her information was recovered to be used in the following conversations. The interaction approach was achieved by the robot asking 3 questions of each topic. The questions were chosen in the difficulty level set by the adaptive function. The level of difficulty of the questions was adjusted based students' answer of a previous question, i.e., from one question to another. In the end, the robot said goodbye to the student and the researches explained possibles system mistakes to the student or why some students' answers might be wrong.

Sessions

The first session was a pilot test for system calibration in the school environment and to the participant students get more used with the robot. In that, the robot asked the student his/her favorite sport, music and food. The system searched the topics on the internet and told them to the student. It also recorded these topics and respective explanations in its database for future use. Next, the robot asked 3 questions, which were randomly chosen according to the level of difficulty in which the student was. The robot was not programmed to repeat any question. It just gave feedback by saying "Congratulations! You are right." or "I'm sorry. You are wrong.". For this session, the robot only considered the content *Vowel Encounter*. The time average of this session was 8.02 ($\pm$3.51) minutes per student.

In the second session, the robot recognized the student by face recognition and recovered his/her information. It also said some things about the student's favorite sport and food to add a personal touch to the interaction. For instance, "I remember you like ice-cream. I tried it since the last time we met. Well... I prefer battery.". Later, the robot asked 3 question regarding *Vowel Encounter*, as in the first session, and 3 questions regarding Digraph, totalizing 6 questions. Differently from the previous session, the robot was able to repeat the questions and confirm the user's answers. Finally, the robot said goodbye to the student. The time average of this session was 9.92 ($\pm$2.83) minutes per student.

Assessment

To evaluate the users (teachers and student) opinion about their experience with the system , a 5-Likert Scale was used. In this scale, each question is an item where the participant gives a score from 1 to 5, in which 1 means absolutely nothing and 5 means totally and 3 is the neutral score.

For the students' part analysis, we applied the Wilcoxon paired test in the result in order to analyze if the users' perceptions in both conditions can be considered the same or if there is a statistical difference between them, it means, if the settings of the interaction may influence the participants. As the number of teachers was not significant - because we used only the teachers that were assigned to the students' classes in this school year - it was analyzed more as a case study than a statistical validation. For the system assessment, it was performed a human validation of the system classifications, taking into account the system's accuracy and tracking the causes of the mistakes.

Results and Discussion

The results of the experiment were grouped according to (i) teachers perceptions in using the system, (ii) students perception with regard to the robot and (iii) the system accuracy analysis.

Teachers perception in using the system

Understanding the teachers' perception of their experience with the system is an important step because the researchers can evaluate f the teachers can easily handle the information through GUI. Two regular teachers participated in the experiment. Both teachers are female and they have 33 and 55 years old. Between step 1 and step 2, they answered a questionnaire together regarding their perception about the system, after attempting the presentation of the system interface (Step 1). After step 7, they answered the same questionnaire again. The questionnaire has the following questions:

1. How much do you think this application can support you in your classes activities?
2. How much do you think it will be easy to create activities in this tool?
3. How much you think this application has potential to become a regular tool in teaching?
4. How much do you think this methodology is efficient for content approaching?
5. How much time do you think it is necessary to get used with this system?
6. How much do you think this methodology can enhance the students' learning?
7. How much do you think that technological methodologies are more efficient compared to the traditional ones?

The teachers answers before using the system (2nd row) and after using the system (3rd row) are presented in Table 1.

Question number	1	2	3	4	5	6	7
Before use the system	4	3	3	4	2	3	4
After use the system	4	4	4	4	3	4	4

Table 1: Teachers' scores before and after using the system.

By analyzing the first answers, it is possible to note that the teachers perceived the system potential, once their scores were above the average, except for question number 5 about the time to get used with the system. Although none of the questions got the maximum score, they kept a score above the neutral in the majority of the items (questions of the Likert Scale) after the teachers experienced the system. In open questions, they reported the students' motivation in interacting with the robot and system auto-reports as advantages and system misunderstandings as disadvantages.

Students perception with regard to the robot

We also analyzed the students experience with the system through the interactive robot NAO. A questionnaire was applied aiming to understand if the students were able to notice some robot skills provided by the system. For instance, we investigated if personal information would be a facilitator in rapport building and simulating the robot in becoming more close to the student (item 3). This questionnaire was applied twice to the same students, after the students' first interaction (Step 3) and after the second interaction (Step 6). The questions were:

1. How much do you think you enjoyed this activity?
2. How much do you think you learned with this activity?
3. How much do you think that you and the robot are personal friends?
4. How much do you think that the robot is intelligent?
5. How much do you think difficulty to perform this activity?

The students' average score and Standard Deviation (in parenthesis) after the first interaction (second row) and after the second interaction (third row) are presented in Table 2. We applied the Shapiro-Wilk test to verify if the data follow a normal distribution. Later, we applied the Wilcoxon test aiming to find a significant difference between the samples. In Table 2, the fourth row is the Wilcoxon test comparison between the scores of first and second interaction.

Question Number	1	2	3	4	5
1st Interaction	4.91 (0.29)	3.57 (1.25)	4.12 (0.89)	4.09 (0.94)	2 (1.25)
2nd Interaction	4.69 (0.58)	4.15 (0.93)	4.32 (0.83)	4.33 (0.92)	2.54 (1.11)
p-value ($\mu = 0.05$)	0.09210	0.05248	0.48979	0.19342	0.02808*

Table 2: Students' average scores after interacting with the robot by the first and the second time.

By observing the fourth row of Table 2 is possible to see that none of the items shown a significant difference but item 5, regarding the activities difficulty, in which the students rated the second activity as harder than the first one.

However, they also rated their learning perception higher in the second section as well. The fact that the robot used the users' stored information in the second meeting did not show a significant difference, despite this item average score was slightly higher in the second evaluation. Similarly, a slightly higher average happened in the second evaluation of item 4, in which they rated the robot as intelligent.

Conversely to other studies (Adam et al. 2016; Charisi et al. 2018; Robert 2018), no influence regarding how they felt about rapport building with the robot and their performance in the activity was found in this experiment, as they rated the second meeting less enjoyable than the first one (4.91 against 4.69) in item 1, although there was no significant difference. This can be associated with the novelty and their expectation of the unknown at the first interaction.

The fact that they rated the second interaction more difficult than the first one but also reported they thought they learned more in the second session leads to believe in the cognitive adaptation potential for learning support. However, the learning gain was not evaluated in this experiment, since it requires more assessment methods to present an accurate analysis.

System accuracy analysis

Only measures of the 2nd interaction (after the adjustments of the pilot test) were considered for a more accurate result. The second interaction, as said before, was composed of 6 questions of 3 questions of each topic. The x-axis points of the graphs in Figures 3(a) and 3(b) represent the questions from the Vowel Encounter (V.E.) and Digraph (D) topics, that happened in this chronological sequence. It means, it was performed 3 questions of Vowel Encounter at first and then 3 questions of Digraph topic. Worthy to remember that the first topic (VE) was already and exclusively approached in the first interaction.

The graph in Figure 3(a) shows the system's classification of the students' answers, in which the green markers are the right ones and the red markers are wrong ones, based on a human validation. A classification is considered right if both the system's classification and human validation about a student answer were the same (right or wrong).

The yellow markers are the answers that were not rightly understood by the speech recognition or that was clearly not a valid answer that some student may say, classified as "Listening problems". These classifications of problems in the system's listening were also validated by one of the researches.

In general, these problems were caused by a change in the outside noise or by a mistake by the student in say the answer while the speech recognition system was readjusting the noisy calibration. The same listening problems were considered for the graph of Figure 3(b) shows the answers read by speech recognizer and classified by a human supervisor. It is the ground truth of the answers evaluations since it is in that way the teachers do to correct the regular tests. The blue markers are the right answers and the red markers are the wrong answers.

The graph in Figure 3(c) shows the 5 levels of difficulty occurrences per question. according to his/her answer in the last question. The system started in medium difficulty level (3). Thus, it is possible to observe the students overall performance during the 6 questions. In other words, the higher this value is, the more the system is evaluating that the student is giving the right answer, according to the expected answer. A very close value to the distance threshold (0.4) - when comparing the student's answer to the expected answer - may not activate the adaptation function, what results in remaining asking questions of the same difficulty level in the next question.

Finally, the pie chart in Figure 3(c) shows the system's accuracy considering all the performed classification. The green area is the system's right classifications and the red area is the wrong ones. The yellow area corresponds to the same listening problems of Figures 3(a) and 3(b). By analyzing the graph in Figure 3(a) is possible to perceive the efficiency of the NPL classification when clearly understanding the users' answers. The same is noticed by analyzing the system overall accuracy graph in Figure 3(d). The system was considered classification had an accuracy of 92.8%, considering only the validated answers.

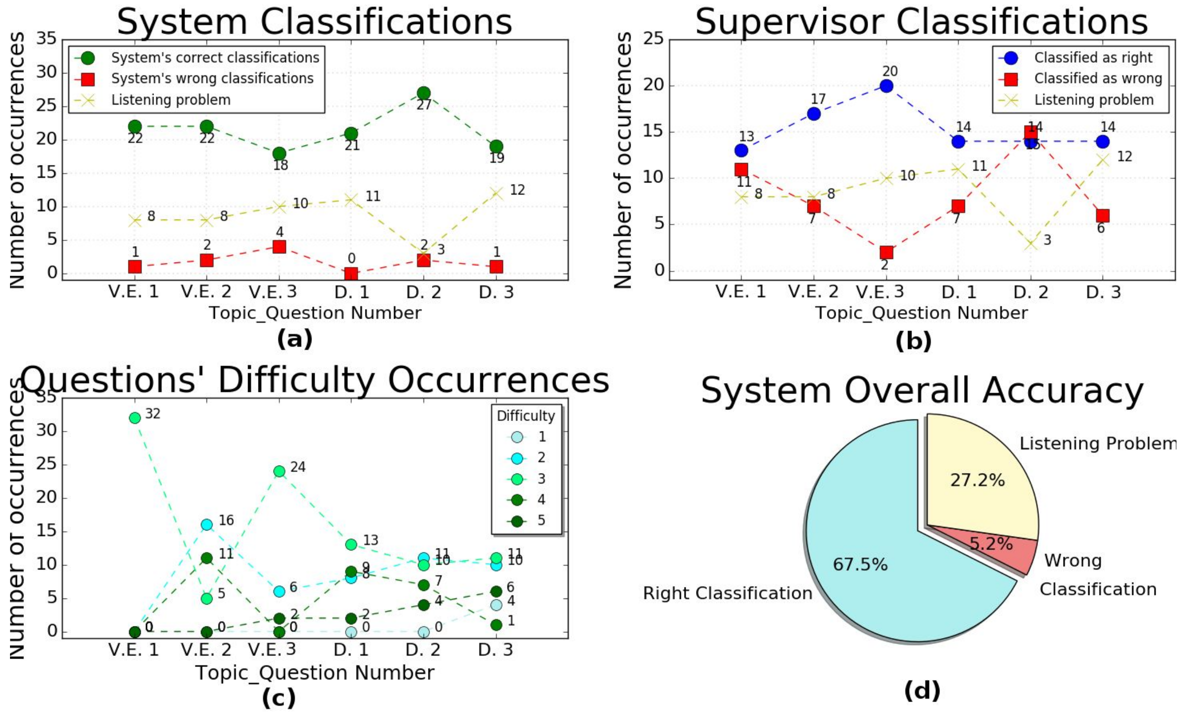

Figure 3: Graphs of (a) System's classifications , (b) Supervisor classifications, (c) Questions' difficulty occurrences and (d) System overall accuracy.

The most critical divergence between the system's classification and the human validation was in the third question of the first topic (V.E. 3) which was the question with the higher number of right answers by the supervisor validation and the lower number of right classifications occurrence by the system. This might has interfered in the adaptation function because it kept the difficulty level for the next topic, but it retook a good response in the second question of the second topic (D. 2). In general, the experiment shown that most demand improvement is the speech recognition treatment, once the system can not validate if the student was really wrong or misunderstanded by the speech recognition.

Conclusion

This paper described an experiment and its results to analyze a cognitive adaptive system by three points of view: by the teachers that used the system, by the students that interacted with the robot and by the researchers that analyzed the system performance. Scores above the neutral state in the teachers' questionnaire showed the teachers understood the potential of this proposal and this finding encourages to keep using this system as a supportive tool for handle the students' profile information along the activities. It seems that the students noticed the potential of the robot being controlled by the system as a tutor in practical exercises, despite it is impossible to analyze the students real learning gain with only correct answers in two interactions with the robot,

Although no statistical evidence was found, a curious point raised with this study was that an improvement in the students' learning perception was observed in the second meeting when the system presented more difficulty adaptation. However, they lower their scores about their interaction enjoyment. Moreover, according to the participants self-statement, their activity enjoyment was not enhanced by the robot's personalized treatment to them.

REFERENCES

Adam, C., Johal, W., Pellier, D., Fiorino, H., & Pesty, S. (2016). Social human-robot interaction: A new cognitive and affective interaction-oriented architecture. In *International conference on social robotics* (pp. 253-263). Springer, Cham.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. *Science Robotics*, *3*(21).

Brown, L., Kerwin, R., & Howard, A. M. (2013). Applying behavioral strategies for student engagement using a robotic educational agent. In *IEEE International Conference on Systems, Man, and Cybernetics*. 4360-4365. IEEE.

Charisi, V., Alcorn, A. M., Kennedy, J., Johal, W., Baxter, P., & Kynigos, C. (2018, June). The near future of children's robotics. In *Proceedings of the 17th ACM Conference on Interaction Design and Children* (pp. 720-727). ACM.

Hartmann, N. S. (2016). Solo queue at ASSIN: Combinando abordagens tradicionais e emergentes. *Linguamática*, *8*(2), 59-64.

Johal, W., Castellano, G., Tanaka, F. et al. Int J of Soc Robotics (2018) 10: 293.

Lee, M. K., Forlizzi, J., Kiesler, S., Rybski, P., Antanitis, J., & Savetsila, S. (2012, March). Personalization in HRI: A longitudinal field experiment. In *2012 7th ACM/IEEE International Conference on Human-Robot Interaction (HRI)* (pp. 319-326). IEEE.

Leyzberg, D., Spaulding, S., & Scassellati, B. (2014, March). Personalizing robot tutors to individuals' learning differences. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction* (pp. 423-430). ACM.

Luckesi, C. C. (2014). *Avaliação da aprendizagem escolar: estudos e proposições*. Cortez editora.

Murray, T. (1999). Authoring intelligent tutoring systems: An analysis of the state of the art. *International Journal of Artificial Intelligence in Education (IJAIED)*, *10*, 98-129.

Robert, L. P. (2018). Personality in the Human Robot Interaction Literature: A Review and Brief Critique, AMCIS 2018, Aug 16-18, New Orleans, LA. AIS eLibrary, 16-18.

Tozadore, D. C., Valentini, J. P. H., de Souza Rodrigues, V. H., Vendrameto, F. M. L., Zavarizz, R. G., & Romero, R. A. F. (2018). Towards Adaptation and Personalization in Task Based on Human-Robot Interaction. In *2018 Latin American Robotic Symposium* 383-389. IEEE.

Vouloutsi, V., Munoz, M. B., Grechuta, K., Lallee, S., Duff, A., Llobet, J. Y. P., & Verschure, P. F. M. J. (2015). A new biomimetic approach towards educational robotics: the distributed adaptive control of a synthetic tutor assistant. *New Frontiers in Human-Robot Interaction*, *22*.

You, S., & Robert, L. (2017). Emotional attachment, performance, and viability in teams collaborating with embodied physical action (EPA) robots. *You, S. and Robert, LP (2018). Emotional Attachment, Performance, and Viability in Teams Collaborating with Embodied Physical Action (EPA) Robots, Journal of the Association for Information Systems*, *19*(5), 377-407.

CHAPTER

6

A FUZZY DECISION-MAKING SYSTEM USING MULTIMODAL OBJECTIVE MEASURES FOR EDUCATIONAL ADAPTIVE SYSTEMS

Manuscript submitted to Elsevier Journal Expert Systems with Applications.

Contribution statement

D.C. Tozadore conceived Evaluation module and the Adaptive system proposal. D.C. Tozadore and R. A. F. Romero performed the literature review of Section 2. D.C. Tozadore implemented and tested all the methods presented in the manuscript. Both authors contributed on writing the manuscript. R. A. F. Romero, in addition, supervised all this research work.

Presentation

The Fuzzy Decision-Making Algorithm presented in this manuscript is the final proposal of R-CASTLE on this Ph.D. project. It was proposed taking into account the feedback of teachers about the GUI, presented in Chapters 4 and 5. They claimed the adaptive modeling was not so intuitive as expected. Then, a fuzzy model through linguistic variables and sets - which is at the level of the human idiom modeling - was proposed. Experiments herein were about the performance of the adaptation algorithms in a new dataset, made from recorded videos of a new activity. The new modeled proposal is still on hold to be evaluated with teachers. Its contributions, however, tells about evaluations of several algorithms for multimodal analysis for adaptation.

A Fuzzy Decision-Making System using Multimodal Objective Measures for Educational Adaptive Systems

PhD Daniel C. Tozadore[a,*] (Researcher), Professor Roseli A. F. Romero[a] (Researcher)

[a]ICMC - USP, Av. Trabalhador São Sancarlense, 400, São Carlos - SP, Brasil

ARTICLE INFO

Keywords:
Adaptive Systems
Fuzzy Logic
Human-Robot Interaction
Social Robots

ABSTRACT

Personalized Learning is hard to achieve through traditional methods, in which the teachers have to deal with a large number of students at the same time. Adaptive Systems have the potential to aid this scalability issue, plus presenting a more challenging procedure to the students. However, their decision-making mechanisms lack transparency of their calculations to the persons that will use them: the teachers. In this article, a fuzzy decision-making system for content difficulty adaptation is proposed. Opposite to conventional systems in education that only take into account the students' answers, this proposal takes into account five types of measures: face gaze, facial emotion expression, amount of spoken words and answering time, in addition to correctness of the given answer. These measures were taken from the behavior of the students during an educational activity with a social robot. The fuzzy decision-making system was compared to a simple rule-based system as baseline. A dataset was created for evaluation of the proposed system, in which the measures were autonomously extracted from videos regarding student-robot interactions, during practical exercises prepared by their teacher. Beyond providing more intuitive modeling by using linguistic terms, the proposed fuzzy system presented similar performance of classification regarding the baseline method and some supervised machine learning methods, having the advantage of does not depend on previous data.

1. Introduction

Artificial Intelligence (AI) is becoming more and more frequent in people's routines, present in their jobs and entertainments. AI can benefit the processes by automating repetitive tasks with decision making algorithms based on previous knowledge given from an expert or from data itself. Recommendation algorithms are the most visible example of AI supporting people's routine. The applications suggest to the users what videos to watch, what news can interest to them the most, what is the best time to leave their work due to the traffic, and all of this based on their previous choices crossed with other users' information. This customization feeling had shown itself a powerful ally to keep the people interested for a longer period of time while interacting with the devices equipped with such features. Thus, systems that adapt themselves to hold the users' interest and attention, the Adaptive Systems (AS), had an increased demand in the educational field. This is due to educational activities are daily performed for several times along the school year and this routine constantly results in boredom and decrease of students motivation (Wilson and Scott, 2017).

Although most of the people take advantage of AI, it is not easy to them to follow and know about the technological innovations happening constantly. The AI implementation has many costs, not only in terms of financial costs, but also in the changing of the mindset of stakeholders. If in one hand, AI has the potential of supporting teachers in their classes' preparation, evaluation and feedback to the students, on the other hand, AI is limited for the lack of teachers' preparation in dealing with such technologies (Johal, Sandygulova, De Wit, De Haas and Scassellati, 2019). Making new AI applications more understandable and accessible to everyone is no longer just a facilitator for users of the system. It is now also a guideline - and almost an appeal - from the European Union's scientific community (Hamon, Junklewitz and Sanchez, 2020).

The R-CASTLE project (Robotic-Cognitive Adaptive System for Teaching and Learning) (Tozadore, Pinto, Valentini, Zavarizz, Rodrigues, Vedrameto and Romero, 2019b) emerged from the aiming of connecting teachers and students, taking into count the advantages of AI. R-CASTLE is a framework for end-to-end educational activities, involving since the preparation of the exercises, passing to autonomously execution of the activities by using a social robot, until a final report of the evaluation of each student in these exercises. The main contributions of R-CASTLE is to offer its features in an intuitive way. The teachers can prepare their content through the windows of the Graphical User Interface (GUI), whereas the students are expected to take advantage of the interaction with a social robot through verbal and visual communication, in which they are more inclined to pay attention due to the novelty factor (Ahmad, Mubin and Orlando, 2016).

While the most of the AS for learning use only the students' answers to the exercises to calculate the adaptation, one of the biggest advantages of R-CASTLE is that the system also takes verbal and visual qualitative measures from the student and uses them to adapt itself. The aim is to provide a personalized experience of content difficulty and in-

*Corresponding author
tozadore@usp.br (D.C. Tozadore); rafrance@icmc.usp.br (R.A.F. Romero)
www.danieltozadore.com, dtozadore@gmail.com (D.C. Tozadore); www.icmc.usp.br/~rafrance/ (R.A.F. Romero)
ORCID(s): 0000-0003-0744-0132 (D.C. Tozadore); 0000-0001-9366-2780 (R.A.F. Romero)

teractive interventions according to the necessity of each student.

In this paper, a modeling of a Fuzzy Decision-Making Algorithm (FDMA) is proposed to be part of R-CASTLE. It allows the programmers or teachers that will run the activities to deal with linguistic variables in a more intuitive way compared to deal with raw numbers of Simple Rule-Based (SRB) algorithm or supervised ones. Furthermore, it is not necessary to provide previous information of the students and the teachers can quickly configure the system for autonomous classification in any activity. Thereby, we believe this fuzzy model can produce decision-making results similar to other classification methods, plus these cited advantages, being the analysis of these second advantages a subject for future works.

This article is organized as it follows. In Section 2, some works related to AS and decision-making systems are provided. In Section 4, the fuzzy decision-making system is proposed. In Section 5, the experimental setup is presented and the results are discussed. Finally, in Section 6, final considerations and future works are presented.

2. Background

In this section, the description of the main concepts and methodologies to better understand the proposed work is presented.

2.1. Adaptive Systems (AS)

In nowadays, systems that pursuit adaptation for each user are a tendency. In the educational context, a software for approaching the personalization learning problem is desirable due to address the needs of each student, working alongside the teachers in the exercises related to different contents. Educational activities are performed several times with the students along the year, which may lead to boredom.

Systems that adapt themselves to the students' needs can make them feel more challenged and consequently hold their attention and motivation for a longer period of time, compared to traditional ones. Adaptive Systems (AS) usually keep a model for each user in which they hold information about a specific user making the user enjoy the most the interaction. This information go from system settings to personal data about the users and all of them are intent to be used to improve the interaction experience in the learning scenario. For example, the system's volume the user likes the most or the frequency which the system may perform interventions and suggestions to the user along the activities.

There are three main types of User Model for AS: adaptive systems with no user model, systems based on static user model and systems based on dynamic user model. AS with no user model are characterized by a reactive behaviour with respect to the user's immediate feedback. They do not keep an explicit cache of information about the user. Systems based on static user model are systems that use predefined information about the user's relevant attributes, making use of this explicit knowledge for adaptation. Usually, they consider some groups of users' profile and try to place the current user in one of these groups, adapting its behavior according to the preferences of this group. Finally, systems based on dynamic user models, like the static ones, maintain an explicit model of the user, tailored to the task at hand. Additionally, they update information of the user as they operate. Therefore, dynamic user models seem to provide individual and exclusively adaptation. However, their implementation are more complex due to the various observed values that may change for each user in every interaction.

Regardless the types, AS are composed by common elements that communicate among themselves to the achieved adaptation. According to (Martins, Santos and Dias, 2018), the architectures of such systems are designed taking into account 4 components that serve as a framework for any of those systems. These components are: Operational Parameters, Decision Making module, User Model and User Interface. The *Operational Parameters* hold the variables that control the system e.g. the system volume, speed and used methods. *Decision making module* is the layer in which decision algorithms take as input the perceived information, and generate a future response action that will be externalized by the interface. The *User model*, as already discussed, can be seen as an explicit repository of knowledge on the users, which can be used to retrieve the information needed for adaptation of the system. *Interface* is the communication way with the users. It means the layer in which occurs the exchange of information between the computational process and the external world.

Systems composed by these components tend to present better results compared to conventional ones. Such systems often face the challenges of the internal communication between these modules and the external communication with users and environment. Once these systems deal with several data of different types and the efficiency of the final customization lays on the combination of these data collected autonomously, the propagated error is also an issue to be considered and overcome. For instance, the combination of an error in the User model module plus an error in the decision making module results in a higher mistake in the final customization.

2.2. Intelligent Tutoring Systems

Systems that adapt to the students' difficulty are addressed commonly in technological teaching designs. Intelligent tutoring systems (ITS) are those which adapt to the student necessities, mainly adapting the content difficulty based on previous experiences with that student. They are often used in activities using computer and tablets and its implementation may vary from one application to another Murray (1999). Their usage increases the learning experience of the students and provides better consequences in the content fixation. In general, ITS provides easy and intuitive ways to the teachers regarding their activity planning through graphical interfaces Paiva (2017).

These advantages could be combined to the advantages of embodied systems, which provide a more complete experience than virtual learning environment. Few of the existing ITS involves social robots Platz, Krieger, Niehaus and Winter (2018). Those whose made use of ITS with robots claimed they witness significant enhance in the young student's learning rate.

2.3. Rule-Based Decision Making

The most common method for decision-making is the Rule-Based method. Its popularity is due to its simplicity in both computational modeling and problems customization. In general, they use the values as input to the rules constituting the system in a specif state. These states are configurations of the system that will try to fit the users' expectation of enjoyment about their experience in the interaction. Rule-Based systems show results of satisfactory adaptation with a good trade-off in the observed mainly in those cases that their goals are very simple or very specific (Martins et al., 2018), such as adapting the robot's speech while talking to the users (Smith, Chao and Thomaz, 2015) or whether to talk or change the robot's personality (Shen, Dautenhahn, Saunders and Kose, 2015).

2.4. Fuzzy Systems

One of the biggest advantages of Fuzzy Logic is operating with linguistic variables. While the most common mathematical systems use numeric variables, the Fuzzy Logic aims to convert numeric values to linguistic variables based on pre-definitions.This will say how much the input belongs to each one of the possible fuzzy sets with a membership degree. Then, according to the pre-set rules, the defuzification process will place these (Pedrycz and Gomide, 2007). In theoretical language, this entire process is divided into three phases: fuzzification, applying Fuzzy rules and defuzzification.

Fuzzify a numerical input is the process of assigning it to a fuzzy set with some degree of membership. This degree of membership goes from 0 to 1, in which 0 means that the input does not belong to a fuzzy set and 1 means that the input belongs for sure to a fuzzy set. Intermediate values mean that the input may belong to a fuzzy set with a degree of uncertainty. A fuzzy set is a possible state that the input variable may assume in those linguistic variables, for example if the weather is cold, neutral or hot.

Fuzzy Rules are basically a set of *IF-THEN* rules that map the linguistic values of inputs into desirable output values of the system. For example, IF *weather is hot* THEN *fan speed is fast*. These rules are registered in the system and the result is computed by a chosen method that usually is the *modus ponens* or *modus tollens* (De Silva, 2018). The rules could be extended to as many variables as wanted with Fuzzy rules connectors. In general, they are the AND, OR and NOT operators that are mapped to fuzzy system in T-norms, T-conorms and Complement.

De-Fuzzify is the process of converting linguistic variables into a specific decision or real value (numeric values), given fuzzy sets and corresponding membership degrees. For that, it is needed a number of rules transforming a number of variables into a fuzzy result, that is, the result is described in terms of membership in fuzzy sets. For example, rules designed to decide how much changing the content difficulty to the students in "Decrease Difficulty (15%), Maintain Difficulty (34%), Increase Difficulty (72%)". Then, the *Center of Gravity* (CoG) is the most common and useful defuzzification technique. It defines the output as corresponding to the abscissa of the center of gravity of the surface of the membership function, characterizing the fuzzy set resulting from the aggregation of the implication results (Czabanski, Jezewski and Leski, 2017).

Studies with fuzzy rule-based systems were successful in adapting the robot speed to the users (Chiang, Chen and Lin, 2013) and also a powerful option for evaluating learning (Chao and Chen, 2009). In the most of the cases, they presented better results than other methods and more similar modeling and evaluations to human manners.

2.5. Evaluation of Adaptive Interactive Systems

Although the adaptation generated by user modeling techniques often tend to improve the user-system interaction, in the majority of the systems, these techniques turn the system more complex. Consequently, it should be evaluated whether the adaptation in fact improves the system and the user really prefers the adaptive version of it (Gena, 2005).A real and feasible evaluation of AS is a difficult task due to the complexity of such systems (Lee and Jiang, 2019).

In Intelligent Tutoring Systems community, the main practice of evaluation is to perform experiment with a particular set of users and checkout their common behaviors and perceptions (Mulwa, Lawless, Sharp and Wade, 2011). Nonetheless, there is no standard agreed measurement framework for assessing the effectiveness of the adaptation achieved by these systems. Thus, it is important not only evaluate but also to ensure that the evaluation uses the correct methods, since an incorrect method can lead to wrong conclusions (Nussbaumer, Steiner and Conlan, 2019).

According to the assessment framework of AS presented in Paramythis, Weibelzahl and Masthoff (2010), the success of adaptation is addressed at two distinct layers:

Layer 1: Evaluation of user modeling. At this layer only the user modeling process is being evaluated. Here the question can be stated as: "are the conclusions drawn by the system concerning the characteristics of the user-computer interaction valid?" or "are the user's characteristics being successfully detected by the system and stored in the user model?"

Layer 2: Evaluation of adaptation decision making. At this layer only the adaptation decision making is being evaluated. The question here can be stated as: "are the adaptation decisions valid and meaningful for the given state of the user model?"

The Layer 1 of R-CASTLE was evaluated in our previous studies (Tozadore, Valentini, de Souza Rodrigues, Vendrameto, Zavarizz and Romero, 2018) through the analysis of the user-measures reading algorithms. For evaluations matters, in this paper, we are analyzing the Layer 2 of R-CASTLE, the decision-making algorithms. Hence, we are assuming the er-

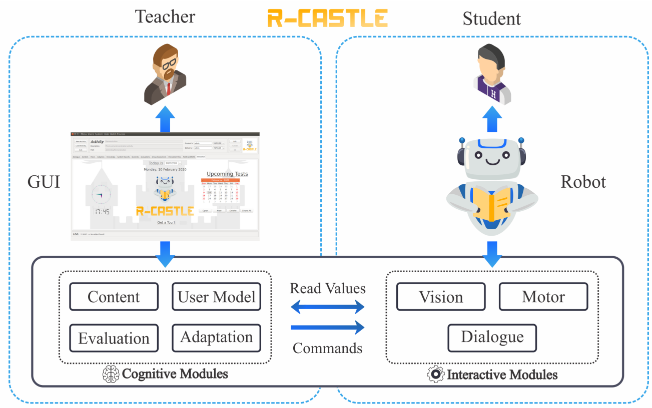

Figure 1: R-CASTLE's Scheme.

rors in the predictions based on the users reading are caused environmental noises, because they are currently facing the biggest challenges of adopting educational social robots in environment (Belpaeme, Kennedy, Ramachandran, Scassellati and Tanaka, 2018). It means, instead of cutting off the outliers resulting of the predictions, they will be considered in the final adaptive classification, which probably will imply in less accuracy.

Once the FDMA system presented in this article will be part of R-CASTLE, an educational system, proposed by us, in next section it will be described briefly.

3. The R-CASTLE System

The R-CASTLE (Robotic-Cognitive Adaptive System for Teaching and Learning) is a computational tool for supporting teachers and students through social robotics Tozadore, Hannauer Valentini, Rodrigues, Pazzini and Romero (2019a). It aims to allow the teachers planning, performing and evaluating fixation exercises regarding almost any content they want to approach with an interactive robot.

R-CASTLE works as a ITS alongside a social robot. After programming the content in the architecture through an intuitive GUI, the robot can autonomously perform the activities with the students and generate individual and group feedback. The feedback is regarding as well as the approached content and students' skills such as communication and concentration capabilities.

The architecture is divided in modules, as shown in Figure 1. The Vision, Dialogue and Motor modules form the group of the Interactive modules. They are responsible for autonomously exchanging information between robot and student, from which the users quantitative measures are taken.

On the other hand, the Cognitive group is formed by the modules of Content, User Model, Evaluation, and Adaptation. In summary, the functionalities of these modules are to process and/or store the information acquired by the Interactive Modules

The Content module allows teachers to insert the content desired to work with students. It is possible to add as many topics as they want in the system through the GUI. Each registered topic in the system consists of a explanation about this topic and as many questions as possible regarding it. The questions must be divided in level of difficulty from 1 to 5, in which 1 is very easy and 5 is very difficult, in order to ensure adaptation during the content approaching. This means that it is possible to correctly switch among them while performing the activities with the students. It is important to notice that the teachers - or the person inserting the questions - will be the people responsible to set up label the difficulty level of each questions. Then, the robot will approach this content in a quiz mode with the students in individual sessions. All the information is inserted in text mode in the Content window.

Students' information and their preferences are stored in the User Model module. As its name itself suggests, is one of the main part of an Adaptive System (as presented in Section 2.1). In R-CASTLE scope, the User Model is dynamic, i.e., it is often updated according to the new interactions with each student.

The Evaluation Module has the functionality of assessment to the users' performance and to obtain the accuracy of the algorithms implemented.It happens throughout the sessions, but the information is stored and available for assessment any time in the Evaluation window.Hence, this module can work as a laboratory to test many configurations of the assessment algorithms.

Finally, The Adaptation module uses the values read from the user to adapt the robot's behavior and the content difficulty, as detailed in the Section 4.

4. Decision-Making Method of R-CASTLE

Although in this paper is addressed only the difficulty content adaptation, the Adaptation module of R-CASTLE has also the functionality of adapting the robot behavior if desired to do so. Both of these adaptation kinds are performed based on the information it receives from the the modules responsible to exchange information with the users. The proposed scope is considering three student's skills to be analyzed by the Adaptation module. The measures are *Attention* (α), *Communication* (β) and *Learning* (γ).

These three measures reflect the processing in the users' body signals and verbal answers, as showed in Table 1. They are Face Gaze (for Attention skill), number of spoken words and facial emotions (for Communication skill) and right/wrong answer (Answer Correctness) for Learning skill.

Indicative values of the students are taken by using some Machine Learning (ML) algorithms for classification, named here the reading algorithms. The Face Gaze is detected by the Haar Cascade method, the emotion classification by a Convolutional Neural Network (CNN) , the number of words is counted by taking the user's verbal answer into string (with Python Speech Recognition [1]). The latest is also the input to the Answer classification algorithm , that matches the given answer by the student with the expected one.

[1] `https://pypi.org/project/SpeechRecognition/` Accessed in Feb 2020.

Table 1
Reading values grouped by observed users' skills.

Attention (α)	Communication (β)	Learning (γ)
Face gaze (Fg)	Number of Words (nW)	Right/Wrong answers (RWa)
	Emotions (Em)	Time to answer (Tta)

These measures are taken during a cycle of information exchanging called *Adaptive Window*, i.e, for each instant t of a cycle, an objective measure is taken. The *Adaptive Window* is predefined by the programmer and it is usually a set of questions or requisitions, to be presented by the robot, in order to evaluate the success rate of the user's response.

There are two ways to achieve adaptation based on the reading values of the users in the R-CASTLE. The first one is Simple Ruled-Based algorithm described in Subsection 4.1.The second is the Fuzzy Decision-Making Algorithm - or Fuzzy Rule-Based System - proposed here, is presented in the Subsection 4.2 In both of them, the designer or the teacher, responsible to run the activity or evaluation, needs to set some values of references for each one these variables, in order to guide the system into this process. In summary, these two ways aim to translate quantitative observations to qualitative analysis of the skills pursuing adaptation.

In the most of the cases, these measures are combined to achieve a unique adaptation in the system. However, the proposed division in these 3 skills has the advantage of analyzing them individually for different adaptation during the same interaction. For instance, if one wants to adapt both the robot behavior based on the skills of Attention and Communication and the content difficulty based on the user's Learning skill, in the same activity. This is possible thanks to this separation of the measures, instead of use all of them combined.

4.1. Simple Ruled-Based Algorithm (SRB)

The SRB decision-making algorithm is the one proposed in (Tozadore et al., 2019b). The skill measures are normalized in the interval [0, 1]. In equation 1, the β is calculated by the average of two numbers: Emotions counting and Spoken Words. These numbers are normalized in [0, 1], (considering the max tolerance for this value) and the same is valid to the γ parameter with its measures. The parameters correspond to the activation value for each weight in the calculation of the function F_{Adp} in Equation 1.

The resulting functions to calculate the skill measures are $\alpha(t) = (Fg(t))$, $\beta(t) = (nW(t), Em(t))$ and $\gamma(t) = (RWa(t), Tta(t))$, in which the functions of the read values are the average of their read values normalized (by values set by the programmer).

The function F_{Adp} (Equation 1) is the main rule to adapt the behavior resulting of the robot Ψ. It tries to optimize the interaction engagement and learning rate to each student, since they have different preferences and difficulties. The measures α, β and γ are the quantitative outputs in the instant t and the w_α, w_β and w_γ are the corresponding weights, set by the programmer before the session starts.

$$F_{Adp}(t) = (w_\alpha * \alpha(t) + w_\beta * \beta(t) + w_\gamma * \gamma(t)) \quad (1)$$

The $Act(t)$ is given by Equation 2 being equals to 1 (it means to increase the level of difficulty) if $F_{Adp}(t)$ is greater or equal to 0.66. This value will be equals to -1 (it means to decrease the level of difficulty), if F_{Adp} is smaller or equal to 0.33 and will be equals to zero (it means to remain the same value) if $F_{Adp}(t)$ is between these values.

$$Act(F_{Adp}(t)) = \begin{cases} 1, & \text{if } F_{Adp}(t) \geq 0.66 \\ 0, & \text{if } 0.33 < F_{Adp}(t) < 0.66 \\ -1, & \text{if } F_{Adp}(t) \leq 0.33 \end{cases} \quad (2)$$

The final result is given by Ψ variable that combines the previous rules as shown in Equation 3:

$$\Psi(t) = \begin{cases} 3, & \text{if } t = 0 \\ \Psi(t-1) + Act(F_{Adp}(t)), & otherwise \end{cases} \quad (3)$$

where t is the *Adaptive Window* in which the robot is approaching one question of a subject. The ψ values go from 1 to 5 in an incremental scale, where 1 is a behavior more focused in information, 5 is a behavior more focused in gestural interactions and 3 is either balanced or neutral behavior.

Next, it is proposed an alternative modeling of the Adaptation process using a Fuzzy Rule-Based System.

4.2. Fuzzy Decision-Making Algorithm (FDMA)

In this section, we describe what linguistic variables with corresponding fuzzy sets are considered to achieve customization in R-CASTLE. The fuzzy modeling of this proposal was implemented with the python library SkFuzzy 0.2 [2]. In this library, the programmer needs to set the mechanisms of fuzzification and defuzzification, the fuzzy sets for each linguistic variable and the corresponding fuzzy rules. For fuzzification, after testing the triangular, Gaussian and trapezoidal modeling and verify that they achieve very close accuracy , we choose the triangular one because it is simpler manner, as shown in the Figure 2.

In previous studies, mistakes in the readings of the users, due to the environmental conditions of luminosity and noise, were observed. Thus, the project decision was to design the parameters of tolerance about face deviation, emotion counting, word counting and time to answer, to be set up before every interaction, in order to decrease the number of outliers

[2] https://pythonhosted.org/scikit-fuzzy/

Table 2
Summary of the Fuzzy Rules for Attention.

IF **X1** is	THEN **A1** is
Frequent	Distracted
Neutral	Neutral
Rare	Concentrated

in each environment. For instance, the number of face deviations which results in the focusing lost, may vary according to the activity. The same can occur with other devices involved in it, such as, a tablet or support device for subtitles decreases the number of face gaze deviation because the student tends to read the subtitles and, consequently, not deviate its look too much. Conversely, a robot by itself that does not move is more susceptible to instigate face gaze distraction to the students or even make them turn their face to approximate the ear they think that listen better. In this way, these values are possible to change through the GUI as well as in running time or in later assessments (off-line evaluations). Fuzzy rules are also possible to be changed any time. Thus, the combination of different rules may enhance the adaptive system performance for each activity.

The Mandani method for inference and the Center of Gravity for defuzzification were chosen in this modeling (Czabanski et al., 2017). Influences of the reading values to the skill measures and, consequently, to the adaptation with the fuzzy decision making mechanism are presented next.

4.2.1. Attention

Attention skill measure will be inferred from the input values of the Face Gaze measure. The linguistic variables for the fuzzy set, representing the Face Gaze measure are: **X1** = { *Rare*, *Normal*, *Frequent* }.

The fuzzy set for the Attention measure is given by linguistic variables: **A1** ={ *Distracted*, *Medium*, *Concentrated*}. In this way, the degree of the Attention of the student will be inferred following the rule:

$$IF \quad \mathbf{X1} \implies \mathbf{A1}$$

and the corresponding rules are shown in Table 2.

4.2.2. Communication

Communication skill measure is calculated based on two observed values of the number of the emotions expressed through facial expression and the amount of words that users talked. Hence, the rules of this measure will count with two fuzzy sets: Emotions and Spoken Words, named here by **X2** and **Y2**, respectively.

The linguistic variables for the emotions are: **X2**= { *Frustrated*, *Sad*, *Neutral*, *Happy* and *Excited* } whereas those to the number of Spoken Words are: **Y2** = { *Contained*, *Neutral*, *Talker*}.

For the emotions variables, the more positive balance between good an bad emotions the user is displaying, the more the fuzzy set result will tend to the Excited output, and the more negative balance, the result will go to the Frustrated

Table 3
Summary of the Fuzzy Rules for Communication.

IF **X2** is	AND **Y2** is	THEN **C2**
Frustrated	NULL	Introverted
Sad	Talker	Neutral
Neutral	Contained	Neutral
Neutral	Talker	Extroverted
Happy	Talker	Extroverted
Excited	NULL	Extroverted

Table 4
Summary of the Fuzzy Rules for Learning.

IF **X3**	IF **Y3**	THEN **L3**
High	Fast	Efficient
Low	Slow	Inefficient

result.On the other hand, considering the number of words submitted to fuzzification process, the fewer words said by the user, the closer the result will get to the Introverted state and the greater the number of spoken words, the more the probability of the Talker result.Then, the inference rule is this case is:

$$IF \quad \mathbf{X2} \quad AND \quad \mathbf{Y2} \implies \mathbf{C2}$$

where **C2** represents the fuzzy set of Communication measure, i.e., **C2** = { Introverted, Extroverted }. and the fuzzy rules for communication are summarized in Table 3.

4.2.3. Learning

As the Communication skill measure, the Learning skill measure has two values to be processed: the time that student spent to answer and success rate of the given answer based in expected ones, represented here by **X3** and **Y3**, respectively. However, there are two fuzzy sets for each one of them.

The linguistic variables for **X3** = States of the success rate are: **X3** = { *Low*, *Medium*, *High* }. Once the success rate value is a number from 0 to 1, the closer this value is to 0, the bigger its membership to the state Low. On the other hand, the closer it is to 1, the higher its membership to the state High.In the time to answer value the states are: **Y3** = { *Slow*, *Average*, *Fast*}. So, the Learning measure is taken following the rule:

$$IF \quad \mathbf{X3} \quad AND \quad \mathbf{Y3} \implies \mathbf{L3}$$

where **L3** represents the fuzzy set of Learning measure, being: **L3** = { Efficient, Inefficient}, and its fuzzy rules are shown in Table 4.

4.2.4. Fuzzy Adaptive Function (FAF)

Once the three measures of Attention (A1), Communication (C2) and Learning(L3) are computed, they are used to calculate the adaptive level of the questions to be presented for the students. This final measure given by Fuzzy Adaptive Function (FAF) is the result of the composition of the inference rules for Attention, Communication and Learning, through the MaxMin operation (Czabanski et al., 2017).

We are considering 3 linguistic variables for **FAF**:

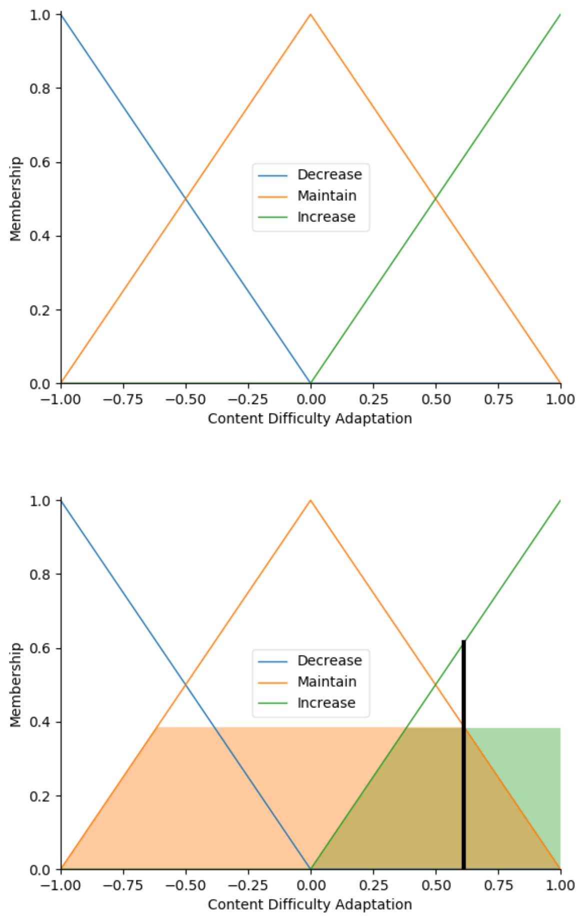

Figure 2: Defuzification of the FAF. The left graph shows the modeling and the right graph shows an example of defuzification with read values of $Fg = 3$, $Em = 100$, $nW = 3$, $Tta = 32$ and $RWa = 8.2$, which by the LoM defuzification method will result in the value 0.59.

Table 5
Rules of the final adaptation (FAF) based on the Attention (A1), Communication (C2) and Learning (L3) measures.

IF **A1** is	AND **C2** is	AND **L3** is	THEN **FAF** is
NULL	NULL	Inefficient	Decrease
Distracted	Introverted	Regular	Maintain
Medium	Neutral	Regular	Maintain
Distracted	Extroverted	Regular	Maintain
Concentrated	Extroverted	Regular	Maintain
NULL	NULL	Efficient	Increase

$$\text{FAF} = \{\ \textit{Decrease}, \textit{Maintain}, \textit{Increase}\ \}.$$

A graphic visualization of this fuzzification is shown in the left graph of Figure 2. The rules for adaptation are given by:

Obviously, the more the states presented by the skill measures are suggesting enjoyment and engagement, the higher the adaptation function will proceed to increase its level of content difficulty. Conversely, whether the undesirable states to attention and learning efficiency are more frequent, the adaptation will tend to lower the difficulty level.

The outcome of FAF function in the adaptive window t, $FAF(t)$, is a float number. However, the difficulty level of the questions approached can be an integer value from 1 to 5, as previously said. Thus, one problem is to discretize the FAF output in one to these values. For this, this modeling uses a similar threshold transformation as Equation (2) to determine the level of difficulty of the questions as it follows.

In our case, the threshold value was set up in 0.25 due to observations in our initial testes with fuzzy in the dataset.

$$F_{Adp}(t) = \begin{cases} 1, & \text{if } FAF(t) \geq 0.25 \\ 0, & \text{if } -0.25 < FAF(t) < 0.25 \\ -1, & \text{if } FAF(t) \leq -0.25 \end{cases} \tag{4}$$

where $F_{Adp}(t)$ is the changing result in the content difficulty, $FAF(t)$ is the crisp result of the FAF function in the Adaptive Window t. In other words, if FAF function value computed is lesser than -0.25, the difficulty of the next question will be decreased in one unit, if it is higher than 0.25, the difficulty will increase in one unit and, finally, if the result is in between these values, the difficulty is kept the same as the last one. The intervals of decrease and increase are bigger than the maintaining one because it is intended to provide to the users a notorious difference from a regular system. It means, the proposed fuzzy system tries to keep the difficulty changing during all the activity.

5. Experiment

In this section is presented the experimental setup and outcomes of performance tests to compare these two systems of adaptation, the SRB and the FDMA. For this experiment, a dataset was created, consisting of 117 samples of Adaptive Windows, in which each sample has the values that compose the users' skill measures. These values were autonomously extracted by the methods available in R-CASTLE system, from recorded videos of learning activities performed with students in their school.

The activity was performed in the primarily school "Oca dos Curumins", in São Carlos city, São Paulo state, in Brazil, specifically for this study. During the activity, participated 39 children from the 5th grade who individually interacted with a humanoid robot answering questions about "Environmental Health". All the students and their parents signed an agreement term for participating in this experiment. Their teacher programmed the content to be addressed through the GUI of R-CASTLE. This content was composed by 30 questions divided in the 5 levels of difficulty, as showed in Section 3.

The interaction sessions with the robot were run following the Wizard of Oz technique (Marge, Bonial, Byrne, Cassidy, Evans, Hill and Voss, 2017), in which a hidden person teleoperates the robot, in order to give the impression to the user that the robot has life by itself. However, is worthy to highlight that the system can work autonomously as well. The robot started making a question of difficulty level 3 and, after the current student answer the question, the person that was teleoperating the robot chose if the difficulty level of the next question would be increased, decreased or maintained, based on his/her answers and body signs. Thus, the true labels of the dataset were taken from the classifications of this person controlling the robot.

Each interaction session had 3 questions, generating the 117 samples of Adaptive Windows, as previously said. A sample (called by question in this paper) is formed by one tuple containing the five reading values described in Subsection (4), considering the time interval that current question started until the time it ended. Each tuple is considered an Adaptation Window, as also described in Section (4), and the labels are Decrease (-1), Maintain (0) and Increase (1) the content's difficulty level.

The dataset is unbalanced, as possible to check in the confusion matrix of Table 7. There are more samples of the Decrease class than the others. Nonetheless, we decided to consider the dataset as it is because an activity where the students will mistake more than hit the answer is expected, mainly the activities regarding new subjects. Thereby, and we want to understand the system behavior in this kind of situation as well.

Additionally, tests with the same dataset were performed using four different supervised methods of classification. The tests were used to also compare our proposal - that uses prediction models set by expertise of the programmer - to these algorithms that autonomously extract the best model of classification from previous data, after training the algorithm with it.

5.1. Results and Discussion

The search for the best configuration for the SRB algorithm consisted in trying every combination of the operational parameters (the values threshold tolerance) and skill measures weights w_α, w_β and w_γ. The weight values w_α, w_β were tested of a range from 0.0 to 0.9 of a maximum of 1, whereas the w_γ value started in 0.1. Operational parameters were tested varying them as it follows: Face gaze (Fg) from 5 to 25 deviations; Emotion balance (Em) from 300 to 500 emotions count; Number of spoken words (Nw) from 2 to 10 words; Time to answer (Tta) 10 to 90 seconds. Only the correctness of the answer (Ans) was kept in 1 for all the evaluations. The combination of all these values (Table 6) had achieved the measures presented in Table 7 in the part of the rule-based method.

The tests with the FDMA were performed with the following parameters (Table 6): the intervals of fuzzification were Fg in [0,30], Em in [0,500], Nw [0,3], Tta [10,90] and Ans [0,1]. The output intervals of all the defuzzification of variables **A1**, **C2** and **L3** were in [0,10]. The fuzzification values of these variables in the FAF also was in [0,10] and, finally, the defuzzification interval of the FAF was [-1,1].

In Table 7 is presented the confusion matrix as well as the measures of precision, recall and F-1 for both methods of rule-based and fuzzy-rule presented by each one of the classes: Decrease (-1), Maintain (0) or Increase (1) the difficulty level. The database is unbalanced, being 61, 13 and 42 samples of the classes Decrease, Maintain and Increase, respectively. Thus, evaluating the methods by their accuracy (correct predictions divided by the total classification attempts) is not a fair analysis, being left behind in this discussion. Hence, the values of precision, recall and F-measure (or F-1) were used.

5.2. Supervised Algorithms

To verify the performance of the FDMA face to ML algorithms, we have considered the methods: Multilayer Perceptron (MLP), Support Vector Machines (SVM), K-Nearest Neighbors (KNN) and Random Forest Classifier (RFC). We choose these methods because each one of them have a different approach to extracting the knowledge from the dataset as well as performing the search of the best model of prediction in the possible solution space. In this case, there was no division in the skill measures (α, β and γ) due to the fact that these supervised methods implementations are only being made for performances comparison and database analysis, whereas the division of the skills measures in the rule-based and fuzzy system allows to use and analyse the skills individually.

The tuples of readings were used as the inputs to the algorithms. All the algorithms were implemented by the Python Scikit-Learn library [3] After several tests, the MLP network topology was constituted by 3 intermediate layers of 1000 neurons each, max iterator of 2000 and momentum in 0.1. A linear kernel with $C = 3$ was assumed for SVM. Finally, the KNN used the 5 nearest neighbors to classify the samples. The best setting of the RFC was max depth of 5 and random state of 0. All the parameters not mentioned here were used as the library's default. The 10 fold cross validation was used in all the methods and their measures, and also the total measures of the SRB and the FDMA, are shown in Table 8.

5.3. Discussion

Analysing the videos from where the dataset was taken, it was possible to note frequent outlier readings from the Face gaze and emotion due to luminosity problems. Thus, the presented measures of classification are expected to be a little bit harmed. Overall, the results of the SRB algorithm showed a higher performances of this method with lower w_α, w_β weights.

The FDMA presented similar measures to the SRB. Results of their precision and recall showed that they have very close behaviour with regard to the false positives and true negatives (a small variation of 0.02 points), except for the Increase class that presented a precision 13% higher and a recall 11% smaller in the SRB compared to the FDMA. This means that the FDMA increased the difficulty more than it should whereas the SRB choose to decreased more than it should for this dataset.

[3] `https://scikit-learn.org/stable/`. Accessed Feb 2020.

Table 6
Best weights and operational parameters for SRB and FDMA methods.

Method	w_α	w_β	w_γ	Fg	Em	Nw	Tta	Ans
SRB	0.0	0.4	0.9	20	413	5	59	1
FDMA Rule	na	na	na	30	500	4	78	1

Table 7
Confusion matrix with the Precision, Recall and F-Measures of the SRB and FDMA.

Labels	SRB			FDMA		
True\Given	-1	0	1	-1	0	1
-1	**54**	7	0	**53**	3	6
0	10	**2**	1	11	**1**	1
1	10	9	**24**	11	2	**29**
Precision	0.73	0.11	0.96	0.70	0.17	0.83
Recall	0.89	0.15	0.56	0.87	0.08	0.67
F-1	0.80	0.13	0.71	0.77	0.11	0.74

Table 8
F-Measures for all classes of the supervised classifiers MLP, SVM, KNN and RFC as well as the proposed SRB and FDMA.

Classifier	MLP	SVM	KNN	RFC	SRB	FDMA
Precision	0.55	0.72	0.56	0.71	0.60	0.56
Recall	0.65	0.82	0.55	0.78	0.53	0.54
F-1	0.60	0.76	0.51	0.72	0.54	0.54

Usually, decrease the difficulty more than the real need may cause more boredom to the student, because it would feel less challenging, but tends to do not cause learning loses. Conversely, wrongly increase the difficulty may hit a challenger way, but may also cause learning loses. Thus, defining what of these methods had better measures only based on these findings, without analyzing their impact on the students' perception, may not define the real impact of such algorithms.

The Maintain class is very dissonant case. We noted that judging whether a student response is not clear to be accepted or denied is confusing for both human (true values) and autonomous classification, because the values of such samples were the most messy values. Thereby, we were not confident to go further with a deep analysis of this class in this study.

The RFC was one of the best tested models, alongside the SVM, presenting the higher measures. Checking the Feature Importance[4] of the RFC the results were $Fg = 0.08$, $Em = 0.12$, $Nw = 0.08$, $Tta = 0.16$, $Ans = 0.55$. It means that the rightness of the answer given by the users was also the most relevant feature in the classifications, just as observed with the SRB with the higher values to γ and int the FDMA in which more extreme values of **L3** leaded to a clear prediction.

Results from classic methods of ML corroborated with the outcomes of both approaches mentioned. These findings also supported the findings of other multimodal classification studies (Abbasi, Monaikul, Rysbek, Di Eugenio and Žefran, 2019), that report constant setbacks and difficulties aiming adaptive behavior in HRI.

Those predictions convergence can be seen as generalized knowledge about the problem itself, and not about the data, because, according to (Alpaydin, 2020), ML algorithms are generalized models used for understanding predictions in dataset independently from specialized knowledge about the data. Taking this statement into account,it is possible to concluded that the used methods have a certain degree of knowledge about the approached problem, regardless the used database.

Analysing the dataset and the videos, the biggest detected problem was the error propagation in the classifications. Users' readings are extracted autonomously through reading algorithms and they also have classification errors, that are propagated to the adaptive classification as outliers (Errors of the Layer 1 of Adaptive System Evaluation 2.5). Thus, the skill measures considered separately suggests a potential overcome in isolating possible errors from a reading algorithm that will propagate to the adaptive method, as it was shown for storytelling activities using the same proposed model of adaptation by multimodal skills (Tozadore et al., 2018).

Higher values in the F-1 of supervised methods (almost 20% more in the largest case) may influence one to believe that supervised algorithms are a better decision-making solution to these problems. However, it is also important to consider that they take time to collect previous data for training; their parameter configuration are not intuitive (mainly for non-programming persons); and they may present overfitting to this dataset. These facts are considered critical because they may compromise the viability in the whole system as a facilitator for teachers in practical exercises.

6. Conclusion

In this paper, a Fuzzy Decision-Making System for the content difficulty adaptation was proposed, jointly with a fuzzy adaptation function for modeling user-adaptive activities for inserting of social robotics in educational tasks. The adaptation was verified based not only in the users' answers but also based in indicative values of face gaze, facial emotion, number of spoken words and time to answer the questions. This decision-making was proposed to be part of R-CASTLE project.

The proposed fuzzy rule based system was compared to other methods to analyze their performance in adapting the difficult level of the questions presented to students. The

[4] A measure that goes from 0 to 1 of each feature, where 0 means not relevant at all and 1 means totally relevant to the classification.

methods used in the comparison were: a simple rule-based system, and some ML methods: MLP, SVM and KNN.

The performance of all methods were similar. However, the fuzzy rule-based system has the advantage of a modeling more closer to the human recognition of the qualitative measures - through the linguistic variables - providing a more familiar understanding of the process to non-expert persons, as the teachers that will use the R-CASTLE.

The next steps of this study consist in testing the R-CASTLE with the Fuzzy Adaptation Function with teachers and students to obtain a more conclusive result about the performance of the system.

Acknowledgment

This study was supported by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES), the Conselho Nacional de Desenvolvimento Científico e Tecnológico - Brasil (CNPq) and the São Paulo Research Foundation (FAPESP).

References

Abbasi, B., Monaikul, N., Rysbek, Z., Di Eugenio, B., Žefran, M., 2019. A multimodal human-robot interaction manager for assistive robots, in: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), IEEE. pp. 6756–6762.

Ahmad, M.I., Mubin, O., Orlando, J., 2016. Children views' on social robot's adaptations in education, in: Proceedings of the 28th Australian Conference on Computer-Human Interaction, pp. 145–149.

Alpaydin, E., 2020. Introduction to machine learning. MIT press.

Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., Tanaka, F., 2018. Social robots for education: A review. Science robotics 3, eaat5954.

Chao, R.J., Chen, Y.H., 2009. Evaluation of the criteria and effectiveness of distance e-learning with consistent fuzzy preference relations. Expert Systems with Applications 36, 10657–10662.

Chiang, H., Chen, Y., Lin, C., 2013. Human-robot interactive assistance of a robotic walking support system in a home environment, in: 2013 IEEE International Symposium on Consumer Electronics (ISCE), pp. 263–264. doi:10.1109/ISCE.2013.6570218.

Czabanski, R., Jezewski, M., Leski, J., 2017. Introduction to fuzzy systems, in: Theory and Applications of Ordered Fuzzy Numbers. Springer, Cham, pp. 23–43.

De Silva, C.W., 2018. Intelligent control: fuzzy logic applications. CRC press.

Gena, C., 2005. Methods and techniques for the evaluation of user-adaptive systems. The knowledge engineering review 20, 1–37.

Hamon, R., Junklewitz, H., Sanchez, I., 2020. Robustness and explainability of artificial intelligence .

Johal, W., Sandygulova, A., De Wit, J., De Haas, M., Scassellati, B., 2019. Robots for learning-r4l: adaptive learning, in: 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI), IEEE. pp. 693–694.

Lee, J.S., Jiang, J.W., 2019. Enhanced fuzzy-logic-based power-assisted control with user-adaptive systems for human-electric bikes. IET Intelligent Transport Systems 13, 1492–1498.

Marge, M., Bonial, C., Byrne, B., Cassidy, T., Evans, A.W., Hill, S.G., Voss, C., 2017. Applying the wizard-of-oz technique to multimodal human-robot dialogue. arXiv preprint arXiv:1703.03714 .

Martins, G.S., Santos, L., Dias, J., 2018. User-adaptive interaction in social robots: A survey focusing on non-physical interaction. International Journal of Social Robotics URL: https://doi.org/10.1007/s12369-018-0485-4, doi:10.1007/s12369-018-0485-4.

Mulwa, C., Lawless, S., Sharp, M., Wade, V., 2011. The evaluation of adaptive and personalised information retrieval systems: a review. International Journal of Knowledge and Web Intelligence 2, 138–156.

Murray, T., 1999. Authoring intelligent tutoring systems: An analysis of the state of the art. International Journal of Artificial Intelligence in Education (IJAIED) 10, 98–129.

Nussbaumer, A., Steiner, C.M., Conlan, O., 2019. Towards a multi-modal methodology for user-centred evaluation of adaptive systems, in: Adjunct Publication of the 27th Conference on User Modeling, Adaptation and Personalization, pp. 219–220.

Paiva, R.O.A., 2017. Autoria de decisões pedagógicas informadas por dados sob a perspectiva de um MOOC. Ph.D. thesis. Centro de Engenharia Elétrica e Informática, Universidade Federal de Campina Grande, Paraíba, Brasil.

Paramythis, A., Weibelzahl, S., Masthoff, J., 2010. Layered evaluation of interactive adaptive systems: framework and formative methods. User Modeling and User-Adapted Interaction 20, 383–453.

Pedrycz, W., Gomide, F., 2007. Fuzzy systems engineering: toward human-centric computing. John Wiley & Sons.

Platz, M., Krieger, M., Niehaus, E., Winter, K., 2018. Suggestion of an E-proof Environment in Mathematics Education. Springer International Publishing, Cham. URL: https://doi.org/10.1007/978-3-319-73748-5_8, doi:10.1007/978-3-319-73748-5_8.

Shen, Q., Dautenhahn, K., Saunders, J., Kose, H., 2015. Can real-time, adaptive human–robot motor coordination improve humans' overall perception of a robot? IEEE Transactions on Autonomous Mental Development 7, 52–64. doi:10.1109/TAMD.2015.2398451.

Smith, J.S., Chao, C., Thomaz, A.L., 2015. Real-time changes to social dynamics in human-robot turn-taking, in: 2015 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS), pp. 3024–3029. doi:10.1109/IROS.2015.7353794.

Tozadore, D., Hannauer Valentini, J.P., Rodrigues, V.H., Pazzini, J., Romero, R., 2019a. When social adaptive robots meet school environments doi:https://aisel.aisnet.org/amcis2019/cognitive_in_is/cognitive_in_is/4/.

Tozadore, D., Pinto, A., Valentini, M.C.J., Zavarizz, R., Rodrigues, V., Vedrameto, F., Romero, R., 2019b. Project r-castle: Robotic-cognitive adaptive system for teaching and learning. Accepted in IEEE Trans. on Cognitive and Developmental Systems .

Tozadore, D.C., Valentini, J.P.H., de Souza Rodrigues, V.H., Vendrameto, F.M.L., Zavarizz, R.G., Romero, R.A.F., 2018. Towards adaptation and personalization in task based on human-robot interaction, in: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), IEEE. pp. 383–389.

Wilson, C., Scott, B., 2017. Adaptive systems in education: a review and conceptual unification. The International Journal of Information and Learning Technology .

CHAPTER 7

CONCLUSION

This thesis presented the R-CASTLE project, which was the result of a Ph.D. research in HRI algorithms and user analyses that makes it possible to use Social Robotics and AI in the educational process for teachers and students.

Two crucial outcomes emerged from the evaluation of the specific objectives and hypotheses. They are in favor of the R-CASTLE's major hypothesis, that providing a smart and centralized platform equipped with new technologies for the educational agents would enhance the whole pedagogical process.

First, the teachers rated the R-CASTLE above average in every researched point, mainly in a comparison between this new proposal and traditional methods. Furthermore, after using R-CASTLE, they raised their scores in the majority of the requested items, suggesting that the time to get used to the system is very acceptable.

Second, the scores given by the students regarding Social Robotics increased as well as the observed educational items. It demonstrated the importance of placing together the presented methods of Social Robotics and ITS.

Although initially, the results pointed out interesting relations among the issues addressed in this thesis, that result in an enhancement in the educational process.

It's worthy to note that the R-CASTLE project was considered exceptionally unique, with ample potential and ambition in every conference and technological center where it was presented (e.g. the *Instituto Técnico Superior*, in Lisbon and *Istituto Italiano di Tecnologia* in Genoa).

According to Tony Belpaeme, one of the biggest researchers in HRI, "One of the goals of Human–Robot Interaction (HRI) is to research and develop autonomous social robots as tutors that are able to support children learning new skills effectively through repeated interactions. To achieve this, the interactions between child and robot should be pleasant, challenging, and pedagogically sound. Interactions need to be pleasant for children to enjoy, challenging so that children remain motivated to learn new skills, and pedagogically sound to ensure that children receive input that optimizes their learning gain."

Therefore, by presenting these findings and a framework where other researchers can easily keep studying in this research field, this thesis provides abundant reasons to believe in its contribution to HRI as a whole and also in people's quality of life through R-CASTLE and in the field of Social Robotics.

7.1 Discussion of results

Discussions regarding each one of the objectives and hypotheses are presented in this Section.

Unfortunately, in most of the cases, studies with the users did not achieve significant statistical difference due to multiple challenges faced by this thesis, such as the lack of specialized data to train the algorithm of automatic users' measure extraction; time constraints, due to the school's schedule and the number of teachers and students available to take part in the experiments. Furthermore the COVID-19 pandemic interrupted part of the final experiments. Nonetheless, the whole system implementation and its initial findings were crucial to support future works exploring HRI in both Education and other potential areas. Moreover, the adoption of technological solutions can be a viable alternative to level out the damage cause by social distancing and can support teachers' abilities of increasing educational out reach when human contact is not possible.

Detailed discussions of each objective and hypothesis are presented in the following subsections.

7.1.1 Technical approaches

The objective (Obj) and hypothesis (H) regarding the technical issues in question are:

Obj I *To evaluate algorithms for automatic users' measure extraction and multimodal adaptation that take into account a teachers-friendly setup and Social Robotics issues.*

H1 *Algorithms without previous data dependency for multimodal adaptation present a good trade-off between intuitive setup and performance compared to supervised algorithms.*

Autonomous algorithms of users' indicator extraction and adaptive computation were implemented, tested and compared. MLP, SVM, KNN and an ensemble of these algorithms were studied for image classification. The Haar Cascade method from the OpenCV library was used for face gaze detection. Hierarchical CNNs were compared to classify facial expressions and their corresponding emotions. Ageitgey face recognition was chosen for the users' recognition. Google Speech Recognition was considered for verbal understanding while Embeddings and Edit Distance were used for sentence matching. Algorithms showed high performances in ideal conditions, with accuracy above 90% most of the time (Chapter 3). In real conditions - school environments - they showed acceptable performances, taking into account the challenges in facing noisy environments, but obviously their accuracy was reduced in such conditions (Chapters 5 and 6).

Regarding the adaptation algorithms, a Rule-Based and a Fuzzy Modeling were proposed and compared (Chapter 6). Both of them had their own advantages but they presented a very similar overall performance in dataset that was tested. Their performances were also compared to supervised ML methods. Supervised methods showed better performances,with the worst case being a greater F-1 of 20%. Although they had insignificant training time, they depended on acquired data to set their prediction models. Conversely, no data dependency and modeling as close as possible to human interpretation through linguistic variables and sets were the strongest features of the proposed methods.

Thereby, **Objective I** was achieved since the evaluations of the algorithms were presented and their advantages were explored by the R-CASTLE implementation. On the other hand, even though the **Hypothesis H1** points out encouraging initial outcomes, it lacks studies and evidence to affirm a real worthy trade-off between the analyzed methods, mainly from the teachers' point of view.

It's important to note that results obtained here were from a specific database. More studies and data are required to accomplish relevant generalization and analysis on this point.

7.1.2 Users analysis

The objectives and hypotheses regarding the analyses of the user's perceptions are:

Obj II *To present an easy way for teachers to manage their educational activities through technology.*

Obj III *To deliver a natural user-experience with social robots that motivates the students.*

H2 *Students would present higher scores in the personal assessment items regarding their enjoyment and rapport building in questionnaires after performing activities with systems using personalization skills compared to their last meeting with the robot.*

Participant teachers gave scores above the average (above 3 out of 5) to every item with regards to their impression about the system. A controversial point was noted in open questions, where 5 out of the 14 teachers (35%) said that the high amount of R-CASTLE variables may be a setback. They also said this could be resolved by using the system and getting used to deal with it.

However, the most important discovery was that the two teachers which used the R-CASTLE to perform their activities raised their scores by one point on items regarding the system's advantages of intuitiveness, the amount of time it took to get used to it and learning enhancement. This result leads to believes that other teachers would also increase their scores if they had also used the full system's potential in some of their activities. Unfortunately, the time demanded to achieve a complete study in this sense played against this thesis.

In the experiments with students in Chapter 5, their motivation and natural interactions were observed. Students' motivation was intended to be autonomously measured by their eye gaze, however, lightning problems made it hard. By human observation, an excitement pointed out by the students' happy emotions and reactions was noted after personal interactions. One example of this is when the robot said it "tried the student's favorite food and it went pretty bad". This fact was also indicated by the high classification of happy emotion, in this case, measured by the emotion classification module.

Furthermore, from one meeting to another, only 4 out of 36 students (11%) did not perform the activity again with the robot due to a loss of interest. The rest of the class was labeled by the teachers as "joyful and enthusiastic interacting with the system".

Although a slight decrease in the students' enjoyment score in the second meet was noted compared to the first one, they still gave high scores to this item in the second interaction (a mean of 4.6 out of 5 in the worst case). On the other hand, the rapport building had a slight increase (even though it wasn't statistically significant). These outcomes suggests that the potential of the adopted techniques impacted the students' perceptions.

Although visual and verbal recognition methods presented flaws throughout the sessions, the flaws did not impacted the user's experiences, since they gave high scores to items regarding their enjoyment and robot's intelligence after the sessions. However, studies on this specific condition need to be conducted to confirm this hypothesis.

The emotional recognition system was precise (accuracy above 80%). Nonetheless, the Ekman emotions model (sadness, happiness, anger, dear, disgust, surprise and neutral) was not considered a good approach during the content approaching phase, due to it constant confusion of bad emotions. A model considering facial expressions for educational activities may present better support the adaptation algorithms.

Hence, no evidence to support **Hypothesis H2** was found in this study, since the students' perception of rapport feeling decreased. Specific studies comparing these points with a larger number of participants would benefit a better analysis regarding H2. Yet, they still gave high scores for these items in the next section and they were able to communicate with the system and presented high expectations in interacting with the robot again, supporting the goals of **Objective**

III. Teachers also showed high scores and the observed items regarding familiarization in dealing with R-CASTLE, suggesting the accomplishment of **Objective II**.

7.1.3 Impact on the educational process:

Obj IV *To increase the teaching experience by providing a scalable resource that can help teachers with personalized tutoring for their students.*

Obj V *To increase the learning experience through Personalized Learning and Social Robotic techniques.*

H3 *R-CASTLE would give more support and motivation to the teachers than traditional methods by assisting with planning, performing and evaluations.*

H4 *Students would present increasing scores in the items regarding learning statements in questionnaires after performing activities with the systems using adaptive behaviors compared to their last meeting with the robot.*

Significant findings were observed in the educational process.

The system provided an interesting alternative to the teaching process since the teachers did note its support and potential in helping them in their activities. These are exciting outcomes of **Objective IV**.

Conversely, the results studying the **Hypothesis H3** were not supportive. Mainly because it was hard to measure how much of the teachers' workload was reduced and how much of their motivation was increased when using the system with such few interactions with it. However, two points are worthy to highlight. First, and again, teachers who used the system claimed to have more ease in their work when using R-CASTLE. In the open questions, 5 out of 14 teachers (35%) claimed that an increase in motivation was an important strength of working with the R-CASTLE.

A significant difference was found on the content difficulty: the students rated the second session more difficult than the first. However, they also increased their scores on perception and the robot's intelligence in their personal statements. Although these factors do not necessarily mean a learning gain, they are often pointed out as factors strongly related to the improvement of the learning gain. The fact that teachers used the first activity's feedback to program the second one seemed to influence the students' experience in the second time as well. This is an inspiring observation since the R-CASTLE's goal was to help teachers achieve a greater influence on their students through its features. Therefore, by analyzing the presented points it is possible to conclude that **Objective V** was achieved and that **Hypothesis H4** was supported.

Finally, a summary of this thesis' objectives and hypotheses as well as their outcomes can be seen in Table 1.

Table 1 – Summary of Objectives and Hypotheses

Name	Analysis	Evaluation Method	Result	Reason
Obj I	Multimodal adaptation algorithms evaluation	Accuracy, training time, parameter configuration, execution time and application suitability	Achieved	Analysis and discussion of several algorithms were presented
Obj II	Teacher friendly system	Users' answers to the survey: Time to get used, intuitiveness, open answers.	Achieved	Scores above neutral state
Obj III	Student-Robot natural communication	Users' answers to the survey: Enjoyment, Rapport Building	Achieved	High scores achieved in the related items
Obj IV	Scalable resource	Users' answers to the survey: System's usefulness, System's potential	Achieved	High scores achieved in the related items
Obj V	Personalized Learning and Social Robotics	Users' answers to the survey: Learning perception, Robot's Intelligence, Content difficulty	Achieved	High scores achieved in the related items
H1	Better advantages of algorithms set by the user	F-Measure (or F-1), modeling intuitiveness and training time.	Not Supported	Measures 20% lower in worst case
H2	Robot personalization in students experience	Users' answers to the survey: Enjoyment, Rapport Building.	Not Supported	Measures have decreased
H3	Enhancement in the teaching process	Users' answers to the survey: System's usefulness, System's potential	Supported with no statistical significance	Measures have increased
H4	Enhancement in the learning process	Users' answers to the survey: Learning perception, Robot's Intelligence, Content difficulty	Supported	Measures have increased

7.2 Future works

The full potential of this thesis's implementations was limited due to time and human resource constraints of a Ph.D. project. Thus, the R-CASTLE proposal opens a wide field of future issues to be easily explored. Some of them already being performed and others are expected to guide future researches. Future studies can also contribute to different areas as follows.

7.2.1 Technical Improvements

Open studies in technical matters are:

- **Algorithms performances**: The evaluated algorithms presented satisfactory performance but it is possible to improve them. For instance, the acquisition of bigger and more specialized databases, and automatic fuzzy-rules extraction methods.
- **New emotion recognition model**: As previously said, considering emotion states specifics for educational activities, such as concentrated, frustrated, excited, bored and bothered

may increase the evaluation of the adaptation system.

- **Autonomous training feeding**: The automatic usage of incoming data to retrain the algorithms is also a goal for R-CASTLE.
- **Online Executions**: The system can operate in notebook mode only. Then, the next step of R-CASTLE is to allow its functionalities to be used online.

7.2.2 Studies with users

Once the presented studies were within-subject experiments, between-subject design experiments would provide better conclusions and understandings regarding R-CASTLE. Some of these identified opportunities are:

- **Personalized system vs Regular system**: Studies comparing the students' experience with a customization system compare to a regular system were interrupted due to COVID-19 pandemic, but it is intended to be tested as soon as possible.
- **Robot playing a tutor role vs robot working as a mate**: The robot's role in the activity plays a pivotal function in the interaction. Experiments considering these two conditions will be performed soon.
- **Studies with R-CASTLE in other areas**: Thanks to the natural communication with the users, R-CASTLE has the potential to perform HRI activities in several other areas, such as healthcare area and education for children with special needs.

7.2.3 Educational Impact

Research that is yet to be done with R-CASTLE in Education are:

- **Adaptive vs incremental difficulty**: Experiments comparing conditions where the content difficulty was customized or static were also interrupted due to COVID-19.
- **Learning gain studies**: Analysis in the content retention of the students was not addressed in this thesis, being left as a future step to be further investigated.
- **Teachers' time efficiency**: Studies comparing how much time teachers could save working with R-CASTLE compared to traditional ways are also required.

Note that other studies are also possible to be inspired throughout the discussions in Section 7.1.

7.3 Contributions

A summary of the technical and scientific R-CASTLE contributions are listed next.

7.3.1 Technical contributions

R-CASTLE's main technical contributions in this thesis went in two directions: the teachers and the students.

7.3.1.1 Teachers

Teachers can interact with the system through the graphical interface to plan and execute their activities and make use of Artificial Intelligence without knowing how to program it. Unlike many intelligent tutoring systems whose content is specific, teachers can input the content they want to exercise with students. This brings great flexibility to the system, since it can be widely used in several areas of teaching.

It's important to note that R-CASTLE automatically analyzes student behavior during student interaction. Thus, teachers can also receive individual reports of student performance in each activity. And the data obtained is stored so that the teacher can have a overall view of each student over a given period.

The data stored by the R-CASTLE system, unlike an intelligent tutor system, is not only related to the student's performance, but also the video corresponding to the student's period of interaction with the system. This feature brings a wealth of information to the teacher. They can check the students' behavior during the interaction; for example if the task caused the student to feel any discomfort or if it went smoothly.

Finally, R-CASTLE can be adapted to run considering the robot operating at simulation level, which makes it possible to be available online.

7.3.1.2 Students

Students will interact with the system through a social robot. It is believed that this fact can make the student exercise the content in a playful way, because, thanks to the vision and voice modules available in R-CASTLE, the student will be able to use voice to communicate with the system, that is, to answer the questions posed by the robot. It was observed that the students would be more willing to perform the exercises when the interaction is more authentic. However, new experiments are necessary to prove this and can be done in the future.

Another interesting point is that the system adapts as the student reaches a certain level of difficulty. Therefore the student will be up for the challenge of improving themselves every time instead of feeling unmotivated.

Students will be able to count on a system that aids in building their self confidence and learning skills through the cultivation of rapport with the robot and the robot's awareness of the student's preferences.

7.3.2 Scientific Contribution

R-CASTLE contributions to the scientific community can be found in the following papers.

7.3.2.1 Published

1. Tozadore, D., Pinto, A. H., Valentini, J., Camargo, M., Zavarizz, R., Rodrigues, V., ... and Romero, R. (2019). Project R-CASTLE: Robotic-cognitive adaptive system for teaching and learning. IEEE Transactions on Cognitive and Developmental Systems, 11(4), 581-589.

2. Tozadore, D. C., and Romero, R. A. F. (2020). Graphical User Interface for educational content programming with social robots activities and how teachers may perceive it. Revista Brasileira de Informática na Educação, 28, 191.

3. Tozadore, D., Ranieri, C., Nardari, G., Guizilini, V., and Romero, R. (2018, October). Effects of Emotion Grouping for Recognition in Human-Robot Interactions. In 2018 7th Brazilian Conference on Intelligent Systems (BRACIS) (pp. 438-443). IEEE.

4. Tozadore, D., Pinto, A., Romero, R., anad Trovato, G. (2017). Wizard of oz vs autonomous: children's perception changes according to robot's operation condition. In 2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN) (pp. 664-669). IEEE.

5. Tozadore, D., Hannauer Valentini, J.P., Rodrigues, V.H., Pazzini, J., Romero, R., 2019. "When social adaptive robots meet school environments" Doi:<https://aisel.aisnet.org/amcis2019/cognitive_in_is/cognitive_in_is/4/>.

6. Ranieri, C. M., Nardari,G. V., Pinto, A. H. M., Tozadore, D. C. and Romero, R. A. Francelin. 2018. "LARa: A Robotic Framework for Human-Robot Interaction on Indoor Environments," 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), Joao Pessoa, 2018, pp. 376-382.

7. Tozadore, D., Romero, R., 2017. Comparison of image recognition techniques for application in humanoid robots in interactive educational activ-ities. from portuguese: Comparação de técnicas de reconhecimento de imagens para aplicação em robô humanoides em

atividades interativaseducacionais, in: XXII Conferência Internacional sobre Informática na Educação, Fortaleza - CE.

8. Romero, R. A. F. ; Ranieri, C. M. ; Pinto, A. M. H. ; Tozadore, D. C. . Uma introdução à Interação Humano-Robô para cuidado com idosos. In: II Congresso Brasileiro de Geron-Tecnologia, 2017, Ribeirão Preto. Revista d a Faculdade de Medicina de Ribeirão Preto e do Hospital das Clínicas d a FMRP - USP, 2017. v. 50. p. 19-23.
9. D.C., Pinto, A.H.M., Ranieri, C.M., Batista, M.R., Romero, R.A.F., 2017. Tablets and humanoid robots as engaging platforms for teaching languages, in: 2017 Latin American Robotics Symposium (LARS) and 2017 Brazilian Symposium on Robotics (SBR), pp. 1–6. doi:10.1109/SBR-LARS-R.2017.8215290.
10. Tozadore, D.C., Valentini, J.P.H., de Souza Rodrigues, V.H., Vendrameto, F.M.L., Zavarizz, R.G., Romero, R.A.F., 2018b. Towards adaptation andpersonalization in task based on human-robot interaction, in: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics(SBR) and 2018 Workshop on Robotics in Education (WRE), IEEE. pp. 383–38
11. Pinto, A.H.M., Ranieri, C.M., Nardari, G., Tozadore, D.C., Romero, R.A.F., 2018. Users' perception variance in emotional embodied robots fordomestic tasks, in: 2018 Latin American Robotic Symposium, 2018 Brazilian Symposium on Robotics (SBR) and 2018 Workshop on Robotics in Education (WRE), pp. 476–482. doi:10.1109/LARS/SBR/WRE.2018.00090.

7.3.2.2 To be published

1. Tozadore, D. C., Valentini, J. P. H. and Romero, R. A. F., "Matching sentences in semantic and syntax level for human-robot dialogues".
2. Tozadore, D. and Romero, R. "Facing the challenges of Artificial Intelligence in Education through Adaptive Systems and Social Robotics"
3. Tozadore, D. and Romero, R. "A Fuzzy Decision-Making System using Multimodal Objective Measures for Educational Adaptive Systems"

BIBLIOGRAPHY

BELPAEME, T.; KENNEDY, J.; RAMACHANDRAN, A.; SCASSELLATI, B.; TANAKA, F. Social robots for education: A review. **Science robotics**, Science Robotics, v. 3, n. 21, p. eaat5954, 2018. Citation on page 20.

BREAZEAL, C.; DAUTENHAHN, K.; KANDA, T. Social robotics. In: **Springer handbook of robotics**. [S.l.]: Springer, 2016. p. 1935–1972. Citation on page 19.

BROEKENS, J.; HEERINK, M.; ROSENDAL, H. *et al.* Assistive social robots in elderly care: a review. **Gerontechnology**, Citeseer, v. 8, n. 2, p. 94–103, 2009. Citation on page 19.

BRUSILOVSKY, P. Adaptive Hypermedia: From Intelligent Tutoring Systems to Web-Based Education. In: GAUTHIER, G.; FRASSON, C.; VANLEHN, K. (Ed.). **Intelligent Tutoring Systems**. Berlin, Heidelberg: Springer Berlin Heidelberg, 2000. p. 1–7. ISBN 978-3-540-45108-2. Citation on page 18.

BRUSILOVSKY, P.; SCHWARZ, E.; WEBER, G. Elm-art: An intelligent tutoring system on world wide web. In: SPRINGER. **International conference on intelligent tutoring systems**. [S.l.], 1996. p. 261–269. Citation on page 18.

CLANCEY, W. J. Methodology for building an intelligent tutoring system. **Methods and tactics in cognitive science**, Lawrence Erlbaum Associates London, p. 51–84, 1984. Citation on page 18.

FUTTERSACK, M.; LABAT, J.-M. Quiz, a distributed intelligent tutoring system. In: SPRINGER. **International Conference on Computer Assisted Learning**. [S.l.], 1992. p. 225–237. Citation on page 18.

GHALI, M. J. A.; AYYAD, A. A.; ABU-NASER, S. S.; LABAN, M. A. An intelligent tutoring system for teaching english grammar. IJARW, 2018. Citation on page 18.

GLEASON, B.; GREENHOW, C. Hybrid education: The potential of teaching and learning with robot-mediated communication. **Online Learning Journal**, v. 21, n. 4, 2017. Citation on page 19.

GOODRICH, M. A.; SCHULTZ, A. C. *et al.* Human–robot interaction: a survey. **Foundations and Trends® in Human–Computer Interaction**, Now Publishers, Inc., v. 1, n. 3, p. 203–275, 2008. Citation on page 19.

GUNAWARDENA, C. N.; MCISAAC, M. S. Distance education. In: **Handbook of research on educational communications and technology**. [S.l.]: Routledge, 2013. p. 361–401. Citation on page 18.

HEERINK, M.; VANDERBORGHT, B.; BROEKENS, J.; ALBÓ-CANALS, J. New Friends: Social Robots in Therapy and Education. **International Journal of Social Robotics**, v. 8, n. 4, p. 443–444, 2016. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-016-0374-7>. Citation on page 19.

JAKE, B.; HEITZ, C.; SANGHVI, S.; WAGLE, D. **How artificial intelligence will impact K-12 teachers**. 2020. Accessed: 2020-01-30. Available: <https://www.mckinsey.com/industries/social-sector/our-insights/how-artificial-intelligence-will-impact-k-12-teachers>. Citations on pages 18 and 20.

JOHAL, W.; CASTELLANO, G.; TANAKA, F.; OKITA, S. Robots for learning. **International Journal of Social Robotics**, p. 293–294, Jun 2018. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-018-0481-8>. Citation on page 21.

JONES, A.; BULL, S.; CASTELLANO, G. "I Know That Now, I'm Going to Learn This Next" Promoting Self-regulated Learning with a Robotic Tutor. **International Journal of Social Robotics**, v. 10, n. 4, p. 439–454, 2018. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-017-0430-y>. Citation on page 20.

JUDSON, E. How teachers integrate technology and their beliefs about learning: Is there a connection? **Journal of technology and teacher education**, Society for Information Technology & Teacher Education, v. 14, n. 3, p. 581–597, 2006. Citation on page 18.

KNEZEK, G.; CHRISTENSEN, R.; TYLER-WOOD, T. Contrasts in teacher and student perceptions of stem content and careers. **Contemporary Issues in Technology and Teacher Education**, Society for Information Technology & Teacher Education, v. 11, n. 1, p. 92–117, 2011. Citation on page 19.

KOLIKANT, Y. B.-D.; MARTINOVIC, D.; MILNER-BOLOTIN, M. Introduction: Stem teachers and teaching in the era of change. In: **STEM Teachers and Teaching in the Digital Era**. [S.l.]: Springer, 2020. p. 1–16. Citation on page 20.

KOTSIANTIS, S. B.; ZAHARAKIS, I.; PINTELAS, P. Supervised machine learning: A review of classification techniques. **Emerging artificial intelligence applications in computer engineering**, Amsterdam, v. 160, p. 3–24, 2007. Citation on page 20.

KRAFT, K. **Robots Against Infectious Diseases: A Technologically Grounded, Human-Centered Exploration**. Master's Thesis (Master's Thesis) — Oregon State University, 2016. Master Thesis. Citation on page 19.

LEITE, I.; MARTINHO, C.; PAIVA, A. Social robots for long-term interaction: a survey. **International Journal of Social Robotics**, Springer, v. 5, n. 2, p. 291–308, 2013. Citation on page 19.

LI, D.; RAU, P. L. P.; LI, Y. A Cross-cultural Study: Effect of Robot Appearance and Task. **International Journal of Social Robotics**, v. 2, n. 2, p. 175–186, 2010. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-010-0056-9>. Citation on page 19.

MARTINS, G. S.; SANTOS, L.; DIAS, J. User-adaptive interaction in social robots: A survey focusing on non-physical interaction. **International Journal of Social Robotics**, Jun 2018. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-018-0485-4>. Citations on pages 18, 20, and 30.

MCTEAR, M. F. User modelling for adaptive computer systems: a survey of recent developments. **Artificial intelligence review**, Springer, v. 7, n. 3-4, p. 157–184, 1993. Citation on page 17.

MELIS, E.; SIEKMANN, J. Activemath: An intelligent tutoring system for mathematics. In: SPRINGER. **International Conference on Artificial Intelligence and Soft Computing**. [S.l.], 2004. p. 91–101. Citation on page 18.

METROPOLIS, N. **History of computing in the twentieth century**. [S.l.]: Elsevier, 2014. Citation on page 17.

NAKANO, T.; WATANABE, H.; HOMAE, F.; TAGA, G. Prefrontal cortical involvement in young infants' analysis of novelty. **Cerebral Cortex**, Oxford University Press, v. 19, n. 2, p. 455–463, 2009. Citation on page 18.

NIAJ, A. R. A. **The Manipulation of Digital Technology into The Classroom Settings: The Various Effects on The Teacher's role and Practices**. Phd Thesis (PhD Thesis) — The British University in Dubai (BUiD), 2019. Citation on page 18.

PIANFETTI, E. S. Focus on research: Teachers and technology: Digital literacy through professional development. **Language Arts**, JSTOR, v. 78, n. 3, p. 255–262, 2001. Citation on page 18.

RAMACHANDRAN, A.; SCASSELLATI, B. Adapting difficulty levels in personalized robot-child tutoring interactions. In: **Workshops at the Twenty-Eighth AAAI Conference on Artificial Intelligence**. [S.l.: s.n.], 2014. Citation on page 18.

ROSSI, P. G.; FEDELI, L. Empathy, Education and AI. **International Journal of Social Robotics**, v. 7, n. 1, p. 103–109, 2015. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-014-0272-9>. Citation on page 20.

SHARP, V. F. **Computer education for teachers: Integrating technology into classroom teaching**. [S.l.]: John Wiley & Sons, 2008. Citation on page 18.

STRUCK, E. L. **Digital transformation in the shipping industry: how Industry 4.0 is shaping the shipping industry?** Phd Thesis (PhD Thesis) — Universidade Católica Portuguesa, 2020. Citation on page 18.

TAPUS, A.; MATARIC, M. J. Socially assistive robots: The link between personality, empathy, physiological signals, and task performance. In: **AAAI spring symposium: emotion, personality, and social behavior**. [S.l.: s.n.], 2008. p. 133–140. Citation on page 19.

TOZADORE, D.; PINTO, A.; ROMERO, R.; TROVATO, G. Wizard of oz vs autonomous: children's perception changes according to robot's operation condition. In: IEEE. **2017 26th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)**. [S.l.], 2017. p. 664–669. Citation on page 42.

TOZADORE, D.; RANIERI, C.; NARDARI, G.; GUIZILINI, V.; ROMERO, R. Effects of emotion grouping for recognition in human-robot interactions. In: IEEE. **2018 7th Brazilian Conference on Intelligent Systems (BRACIS)**. [S.l.], 2018. p. 438–443. Citation on page 42.

TYLER-WOOD, T.; KNEZEK, G.; CHRISTENSEN, R. Instruments for assessing interest in stem content and careers. **Journal of Technology and Teacher Education**, Society for Information Technology & Teacher Education, v. 18, n. 2, p. 345–368, 2010. Citation on page 19.

WILSON, C.; SCOTT, B. Adaptive systems in education: a review and conceptual unification. **The International Journal of Information and Learning Technology**, Emerald Publishing Limited, 2017. Citation on page 18.

YUN, S.-S.; KIM, M.; CHOI, M.-T. Easy Interface and Control of Tele-education Robots. **International Journal of Social Robotics**, v. 5, n. 3, p. 335–343, 2013. ISSN 1875-4805. Available: <https://doi.org/10.1007/s12369-013-0192-0>. Citation on page 20.

APPENDIX

A

CONSENT FORMS

Attached are the terms of consent of the users involved in testing this system. In the case of students, they signed a term by themselves (**A.1**), which was also read to them if they had difficulty understanding the terms, and their parents or guardians also signed the term (**A.2**). In the case of teachers, they signed the term (**A.2**)., replacing "seu filho(a)" (your child) by "você" (you) when convenient.

A.1 Students (Children)

Termo de assentimento para criança e adolescente
(maiores de 6 anos e menores de 18 anos)

Você está sendo convidado para participar da pesquisa "Arquitetura cognitiva robótica para educação". Seus pais permitiram que você participe. Queremos saber o robô é capaz de ajudar as crianças a aprenderem mais. As crianças que irão participar desta pesquisa têm de 8 a 13 anos de idade. **Você não precisa participar da pesquisa se não quiser, é um direito seu e não terá nenhum problema se desistir**. A pesquisa será feita na Escola Oca dos Cumirins, onde as crianças irão interagir com um robô humanoide programado pelos professores. Para isso, será usado o robô humanoide NAO e materiais didáticos da escola. O uso do robô é considerado seguro, mas é possível ocorrer quedas do robô que podem te assustar e se você estiver muito perto, ele pode te acertar. Caso aconteça algo errado, você pode nos informar a qualquer momento, podendo até desistir das atividades sem problema algum. Você também pode nos procurar pelos telefones 3373-9661 da pesquisadora Professora Roseli Romero. Mas há coisas boas que podem acontecer como você aprender de forma divertida com o robô. Ninguém saberá que você está participando da pesquisa; não falaremos a outras pessoas, nem daremos a estranhos as informações que você nos der. Os resultados da pesquisa vão ser publicados, mas sem identificar as crianças que participaram. Quando terminarmos a pesquisa ficará mais claro de saber a melhor forma que os robôs podem contribuir para o aprendizado. Se você tiver alguma dúvida, você pode me perguntar. Eu escrevi os telefones na parte de cima deste texto.

CONSENTIMENTO PÓS INFORMADO

Eu ________________________________ aceito participar da pesquisa "Arquitetura cognitiva robótica para educação".

Entendi as coisas ruins e as coisas boas que podem acontecer.
Entendi que posso dizer "sim" e participar, mas que, a qualquer momento, posso dizer "não" e desistir que ninguém vai ficar bravo ou triste comigo.
Os pesquisadores tiraram minhas dúvidas e conversaram com os meus responsáveis.
Recebi uma cópia deste termo de assentimento e li e concordo em participar da pesquisa.

São Carlos/SP, ____de _________de __________.

____ ____
Assinatura do menor Assinatura do(a) pesquisador(a)

Denúncias e dúvidas podem ser sanadas em:

COMISSÃO NACIONAL DE ÉTICA EM PESQUISA - CONEP
SEPN 510 NORTE, BLOCO A, 3º Andar
Edifício Ex-INAN - Unidade II - Ministério da Saúde
CEP: 70750-521 - Brasília-DF
Contatos Conep:
Telefone: (61) 3315-5878
Telefax: (61) 3315-5879

A.2 Student's Parents and Teachers

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(para pais e/ou responsáveis de menores de 18 anos)

Título do estudo: Arquitetura cognitiva robótica para educação
Pesquisador(a) responsável: Roseli Aparecida Francelin Romero/ Daniel Carnieto Tozadore
Instituição / Departamento: Universidade de São Paulo – Instituto de Ciências Matemáticas e de Computação
Endereço do(a) pesquisador(a) responsável: Av. Trab. São-Carlense, 400, Centro, São Carlos, SP – CEP 13-566590
Telefone do(a) pesquisador(a) responsável para contato: 16 – 3373-9661
Local da coleta de dados: USP - São Carlos

Prezado(a) voluntário(a):

- Seu filho(a) está sendo convidado(a) a participar desta pesquisa de forma totalmente ***voluntária***.
- Antes de concordar com a participação dele/dela nesta pesquisa e responder este roteiro, é muito importante que você compreenda as informações e instruções contidas neste documento.
- Os pesquisadores deverão responder a todas as suas dúvidas antes que você se decidir autorizar a participação de seu filho(a).
- Ele/ela tem o direito de ***desistir*** de participar da pesquisa a qualquer momento, sem nenhuma penalidade e sem que isso acarrete qualquer ônus ou consequência para o mesmo(a);

Objetivo do estudo: A pesquisa **"Arquitetura cognitiva robótica para educação"** está sendo realizada pelo aluno de Doutorado em Computação Daniel Carnieto Tozadore, sob orientação da professora Drª. Roseli Ap. Francelin Romero do Curso de Computação do Instituto de Ciências Matemáticas e de Computação. Tal pesquisa tem como objetivo vivência das tecnologias no ambiente de ensino, e as melhorias que elas causam no aprendizado dos alunos.
Procedimentos: A participação de seu filho(a) nesta pesquisa consistirá em atividades de exercícios com um robô humanoide sobre matéria abordada pelo professor em sala de aula. Essas atividades serão realizadas na escola, em dia e horário de acordo com a disponibilidade dos alunos. Ele/ela tem o direito de se recusar a participar, caso ache as questões muito complicadas. A aula será gravada em vídeo e áudio para posterior transcrição e melhor aproveitamento dos dados, a partir da sua autorização.
Benefícios: Esta pesquisa trará maior conhecimento sobre o tema abordado, sem benefício direto para você, além de contribuir para uma melhor compreensão das questões envolvendo o uso de robótica nas escolas.
Riscos: Apesar de raríssimos, os riscos que seu/sua filho(a) corre são o de ser atingido pelo robô em movimento ou ser atingido por alguma peça que escape do robô por mau funcionamento, caso esteja muito perto do mesmo. A participação nesta pesquisa não envolve nenhum tipo de risco de ordem psicológica. Se o participante se sentir desconfortável em qualquer momento, o mesmo deverá comunicar imediatamente os responsáveis para tomem as devidas providências. A desistência é garantida em qualquer momento sem ônus nenhum.

A.3 Student's Parents and Teachers

Sigilo: O nome do estudante será mantido em sigilo, assim como outras informações pessoais. Os resultados deste trabalho científico estarão sob o cuidado dos pesquisadores, podendo ser divulgados em congressos e artigos, resguardando-se o sigilo quanto a qualquer informação pessoal a seu respeito.
Indenização e ressarcimento: Gastos realizados pelo participante decorrentes da participação na pesquisa serão ressarcidos em caso de acidentes.

Ciência e de acordo do participante (sujeito da pesquisa):
Ciente e de acordo com o que foi anteriormente exposto pelo(a) pesquisador(a), eu ________________________________, RG: ______________, estou de acordo que meu filho(a) participe desta pesquisa, assinando este consentimento **em duas vias**, ficando com a posse de uma delas.

São Carlos, _____/_____/_____

____
Assinatura do sujeito de pesquisa ou Representante legal

Ciência e de acordo do pesquisador responsável:
Asseguro ter cumprido as exigências da resolução 466/2012 CNS/MS e complementares na elaboração do protocolo e na obtenção deste Termo de Consentimento Livre e Esclarecido. Asseguro, também, ter explicado e fornecido uma cópia deste documento ao participante. Informo que o estudo foi aprovado pelo CEP perante o qual o projeto foi apresentado e pela CONEP, quando pertinente. Comprometo-me a utilizar o material e os dados obtidos nesta pesquisa exclusivamente para as finalidades previstas neste documento ou conforme o consentimento dado pelo participante.

Declaro que assinei 2 vias deste termo, ficando com 1 via em meu poder.

São Carlos, _____/_____/_____

____
Assinatura pesquisador responsável pelo projeto

ANNEX

A

FORMS LINKS

<https://forms.gle/BY92bK6gATpS3LjZ8> Questionnaire of experiments with students in Chapter 5;

<https://forms.gle/fXH6LGjXzkV1arm78> Questionnaire of experiments with teachers in Chapters 5 and 4;

<https://www.youtube.com/watch?v=GlNj98L1Mrc> R-CASTLE demo video;

ICMC USP
SÃO CARLOS