UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE TEXTO TEXTO TEXTO
PROGRAMA DE PÓS-GRADUAÇÃO EM CIÊNCIA DA COMPUTAÇÃO

BRUNO DE SOUZA JERONIMO

Human-Robot Interaction Engagement Strategies for Children's Education

Recife

2022

BRUNO DE SOUZA JERONIMO

Human-Robot Interaction Engagement Strategies for Children's Education

Trabalho apresentado ao Programa de Pós-graduação em Ciência da Computação do Centro de Informática da Universidade Federal de Pernambuco, como requisito parcial para obtenção do grau de Mestre em Ciência da Computação.

Área de Concentração: Mídia e Interação

Orientador (a): Prof. Judith Kelner

Coorientador (a): Dra Anna Priscilla de Albuquerque Wheler

Recife

2022

Catalogação na fonte
Bibliotecária Monick Raquel Silvestre da S. Portes, CRB4-1217

J56h Jeronimo, Bruno de Souza
Human-robot interaction engagement strategies for children's education / Bruno de Souza Jeronimo. – 2022.
80 f.: il., fig., tab.

Orientadora: Judith Kelner.
Dissertação (Mestrado) – Universidade Federal de Pernambuco. CIn, Ciência da Computação, Recife, 2022.
Inclui referências.

1. Mídia e interação. 2. Robôs sociais. I. Kelner, Judith (orientadora). II. Título.

006.7 CDD (23. ed.) UFPE - CCEN 2023-32

Bruno de Souza Jeronimo

“Human-Robot Interaction Engagement Strategies for Children’s Education”

Dissertação de Mestrado apresentada ao Programa de Pós-Graduação em Ciência da Computação da Universidade Federal de Pernambuco, como requisito parcial para a obtenção do título de Mestre em Ciência da Computação. Área de Concentração: Mídia e Interação.

Aprovado em: 11 de agosto de 2022.

BANCA EXAMINADORA

Prof. Dr. Jaelson Freire Brelaz de Castro
Centro de Informática / UFPE

Prof. Dr. Carmelo Jose Albanez Bastos Filho
Escola Politécnica de Pernambuco / UPE

Profa. Dra. Judith Kelner
Centro de Informática / UFPE
(**Orientadora**)

ACKNOWLEDGEMENTS

I want to thank my friends, family, and coworkers who contributed throughout this work, each in their own way, so I could reach this result. In particular, I would like to thank Professor Judith Kelner for welcoming me and betting on my capacity. Immensely glad that we have, in addition, built a solid and nurturing friendship. Equally, to Doctor Anna Priscilla de Albuquerque Wheler, a precious friend, my co-supervisor, and family nowadays. Who encouraged me to leave my comfort zone, made me face this challenge, accompanied me throughout this journey, and was fundamental in the most demanding moments of this work. To the two of you, my eternal thanks for making me the researcher I am today.

To my closest friends, who near followed my fears, and all the struggle. Sílvio Gatis, Fabiana de Andrade, Vinicius Jordão, Vinicius Fulgêncio and Felipe Brito. For all the love, affection, and refuge - yesterday, today, and always. My psychologist Rosália Albuquerque also deserves a special mention in this emotional support.

I thank my family, the foundation of everything, my safe haven, and whom I hope to bring pride with the fruits of this work. My father, Mário, and my mother, Betânia. My grandmother Maria, my brothers Mariozinho and Guilherme, and my sister-in-law Clara.

I am grateful to my research colleagues, whose efforts have also echoed in my work. In particular, Rodrigo Melo, José Paulo, and Rodrigo Monteiro. They were co-responsible for the core of the application used as the basis for this work. I also thank Priscila Lira, Val, and Fernanda, colleagues from GRVM, who walked with me on this journey.

My work colleagues from the Department of Electrical Engineering that also rooted for me. Gustavo Duarte, Marcos André and his wife Denise Ivone, Davidson Marques, Guilherme Soriano, Douglas Contente, Pedro Rosas, Luiz Henrique, Leonardo Limongi, Jeydson Lopes, Eduardo José, Eduardo Oliveira, Rafael Neto, Carolina Caldeira, Antonio Dutra, Aline Cavalcanti and Camila Bandeira. Thank you very much.

Finally, I would like to thank the examining board, Professors Carmelo Bastos and Jaelson Castro, for accepting our invitation and the provided suggestions. Their excellent considerations certainly improved the final quality of the manuscript. And I also thank CNPq, the PRONEX project 2014 - APQ-0880-1.03/14, and Professor Patrick Hung for providing the artifacts used in this research.

ABSTRACT

Assistive and Educational Robotics is one of the world's highest profile areas of Human-Robot Interaction. Regarding this aspect, recent literature review indicates that the use of social robots in the educational domain has attracted significant attention in recent years. Fueled in particular by the growing numbers of students per classroom, and the demand for greater adaptation of curricula for children with diverse needs, bringing efforts into the research of technology-based support that enhances the efforts of parents and teachers. It also elicits positive benefits for learners of various degrees and abilities, including musical education, driven in particular by the engagement generated within the physical presence of the robot. Computer-based technologies can support music education in developing an individual's acoustic performance and composition skills, including supporting distance learning and strengthening self-efficacy and independent skill learning. Technologies can reinforce existing learning strategies and encourage more people to learn music. Engagement, in this sense, is critical in avoiding giving up and inducing an individual's interest in developing further learning horizons. For instance, there is evidence that anthropomorphic embodiment features can engage users. As such, robot designers can improve successful HRI features, selecting from a window of validated robot embodiment features, such as robot motion, facial expressions, voice pitch, and voice speed, enhancing human perception while enriching social encounters. While the relationship between engagement, social robots, and learning is evident, hence the number of published works that intersect these topics, little is known about what it is in the robot's appearance that promotes learning. Even further, which physical attributes/features are ideal in such environments, representing an interesting HRI problem and a gap in the current state of the art. In order to better comprehend the role of the robotic embodiment in the context of learning and eliciting, which in these elements falls into the preferences and thus engages the target audience, the present work then focuses on the comparison of two social robot models running the same Human-Robot Interaction (HRI) applications targeting the context of music education for children, underlying the design choices favored by the target audience on the running tasks. The evaluation used an experimental remote protocol supporting collecting online feedback with users during the COVID-19 pandemic. Empirical results supported performing quantitative and qualitative evaluations of the HRI application and highlighting the perceived differences in robot embodiment features.

Keywords: social robots; children; user evaluation; robotic embodiment.

RESUMO

A Robótica Assistiva e Educacional é uma das áreas de amplo destaque no campo da Interação Humano-Robô. Em relação a esse aspecto, recente revisão de literatura indica que o uso de robôs sociais no domínio educacional tem despertado significativa atenção nos últimos anos. Alimentado, em particular, pelo crescente número de alunos por sala de aula, e pela demanda por uma maior adaptação dos currículos para crianças com necessidades diversas, o campo tem empreendido diversos esforços em pesquisas focadas em tecnologias que potencializem os esforços de pais e professores. A aplicação de Robôs Sociais na Educação traz benefícios positivos para alunos de vários graus e habilidades, incluindo educação musical, impulsionada em particular pelo engajamento gerado na presença física do robô. Recursos tecnológicos podem apoiar a educação musical no desenvolvimento de habilidades de um indivíduo, incluindo o apoio ao ensino à distância e o fortalecimento da autoeficácia e do aprendizado de habilidades independentes. As tecnologias podem reforçar as estratégias de aprendizagem existentes e incentivar mais pessoas a aprender música. O engajamento, nesse sentido, é fundamental para evitar a desistência e induzir o interesse do indivíduo em desenvolver novos horizontes de aprendizagem. Existem evidências que designs antropomórficos possuem a capacidade de engajar usuários. Dessa forma, designers de robôs e aplicações voltadas para esse contexto podem empreender melhorias nos recursos aplicados à Interação Humano-Robô, tais como movimentação, expressões faciais, tom de voz e velocidade da fala do robô, elevando a percepção humana e entregando encontros sociais mais ricos. Apesar da relação entre corporificação robótica, robôs sociais e aprendizado ser evidente, haja vista a quantidade de publicações que interseccionam esses campos, pouco é conhecido sobre o que é que há nos corpos destes robôs que promove o aprendizado e tão pouco quais os recursos nestes corpos que se apresentam como ideais nesse cenário. Essas questões representam um interessante problema no estado da arte da Interação Humano-Robô, O presente trabalho concentra-se então na comparação de dois modelos de robôs sociais que executam a mesma aplicação de Interação Humano-Robô (IHR) visando o contexto da educação musical para crianças, e a descoberta dos designs robóticos preferidos pelo público-alvo nas tarefas avaliadas. A avaliação usou um protocolo remoto experimental que apoia a coleta de feedback online com os usuários durante a pandemia do COVID-19. Os resultados empíricos apoiaram a realização de avaliações quantitativas e qualitativas de nossa aplicação, além de destacar as diferenças percebidas nos recursos presentes nos corpos destes robôs.

Palavras-chaves: robôs sociais; crianças; avaliação de usuário; corporificação robótica.

LIST OF FIGURES

LIST OF TABLES

CONTENTS

1 INTRODUCTION

Human-Robot Interaction is a multidisciplinary field dedicated to understanding, designing, and evaluating robotic systems for use by or with humans (GOODRICH; SCHULTZ, 2008). A social robot is a category of devices that supports Human-Robot Interaction (HRI) tasks through robot embodiment features (as in shape, size, motors, sensors, displays, etc.) and adapts its intelligence and behavior through the perception of specific social cues (e.g., voice commands, gestures, facial expressions, etc.) (BARTNECK; FORLIZZI, 2004). Social robots' human-like features include speech, gestures, movements, eye-gaze, and establishing a logical reasoning dialogue by processing personal data and users' social background, conquering a social presence to the robot (BARTNECK et al., 2020). Humans can perceive them as social actors since they represent a physical presence in the interaction environment. Social robots can assume different roles in HRI, such as supporting devices to manage HRI tasks, including displaying content or digital information, similar to a companion application running on a smartphone or tablet (WHELER et al., 2021). They can also support HRI by acting as active or passive social actors. Active social roles include social robots acting as co-participants of the HRI task (e.g., performing equal human tasks or sharing tasks steps). In contrast, passive roles include guiding the HRI task completion (e.g., the role of an educator or caregiver) or a companion that engages the user during HRI tasks (e.g., the role of a friend).

The state-of-the-art has many applications regarding the usage of social robots as tutors or peers in educational environments (BELPAEME TONY, 2018), with studies made using various robotic embodiments due to the availability, market distribution, among other factors. Belpaeme also highlights many challenges derived from the introduction of these devices in educational environments as in technical aspects, such as speech recognition, which is still insufficiently robust to allow the robot for to understand spoken utterances from young children. He questions about the role of the robot in the classroom, should it ever replace the tutor, or be best placed as peer and companion. And questions regarding the acceptance of these robots by the children.

In particular for music education, these applications leverage from the same advances that computer-based technologies have promoted in similar applications,enabling the development of an individual's ability to learning skills (WEBSTER, 2007). This leads to possibilities where children can practice or train new skills while using the actual musical instrument and be engaged by the electronic device. Also, these devices can reinforce the existing methodologies and encourage more people to learn music (SASTRE et al., 2013). Engagement, as in the process of encouraging people to be interested in a given subject or the fact of one being involved with a given activity, is a frequently used term in the vocabulary of educators and those involved in the scholarship of teaching. Bryson (BRYSON; HAND, 2007) emphasises the relationship between the perception and experience of the student, and how they make sense of those elements in the process of learning, as their engagement with education — a dynamic and constantly reconstructed relationship that is critical in avoiding giving up and inducing an individual's interest in developing further learning horizons.

While the evidences that robots physical presence can induce positive benefits for learners in various degrees and abilities, the question remains of what exactly it is about the robot's appearance that promotes learning. It is noticeable that direct comparisons between different robots are difficult with the available data, because of the absence of studies that used the same experimental design across multiple robots (BELPAEME TONY, 2018).

1.1 MOTIVATION

A social robot's embodiment features are set by its physical constraints, influencing how a robot perceives and behaves in the social world. Robot embodiment features such as human-likeness, robot emotion, verbal and non-verbal interaction, and spatial interaction can play significant roles in human perception, trust, and expectations towards social robots (GOODRICH; SCHULTZ, 2008; HANCOCK; BILLINGS; SCHAEFER, 2011). For instance, there is evidence that anthropomorphic design (as in the physical appearance and general hints of human-like demeanor) can engage users since such features will be more acceptable to humans. Such characteristics tend to align with the expectations of an interaction that provides a lifelike experience with an entity capable of developing social relationships and, therefore, introducing familiarity, as in the possibility of recognition and comparison with other regular social experiences (DUFFY, 2003).

Social robots interact with people in a natural and interpersonal manner, usually to achieve positive outcomes in diverse applications such as education, health, quality of life, entertainment, communication, and domestic chores, to name a few (BREAZEAL; DAUTENHAHN; KANDA, 2016). They are complex devices with a variety of form factors, embedded sensors, and capabilities. Positive outcomes go beyond task completion and are also related to the ability to create meaningful social interactions, as in interactions that resemble human encounters and are rich in socially expected cues. Therefore, it becomes crucial that social robots provide feedback respecting expected social behaviors (i.e., social norms, roles, and context) while providing an adequate response to human emotions and other user inputs (LI; JOHN-JOHN; TAN, 2011). Robot designers can increase successful HRI features, selecting from a window of validated robot embodiment features, such as robot motion, facial expressions, voice pitch, and voice speed, enhancing human perception while enriching social encounters.

Recent literature review indicates that the use of social robots in the educational domain has attracted significant attention in recent years, fueled in special by the growing numbers of students per classroom, the demand for greater adaptation of curricula for children with diverse needs, as children who require special education programming because of their behavioral, communicational, intellectual, learning, or physical characteristics or a combination of those characteristics, bringing efforts into research of technology-based support that enhances the efforts of parents and teachers. (BELPAEME TONY, 2018). Technological advances and the efforts of governments in adopting robotics as part of curricula, and efforts of research centers to create low-cost versions played a role in the general availability and popularity of these artifacts. It is possible in 2022 to acquire robots of this type for less than a thousand dollars, a price ten times lower than the ones practiced for the NAO in 2008 by its release. The literature also reveals positive benefits for learners in various degrees and abilities driven by the physical presence of the robot, although the question remains of what exactly is it about the robot's appearance that promotes learning, or even further, which physical attributes/features are ideal in such environments, representing an interesting HRI problem and a gap in the current state-of-the-art.

1.2 RESEARCH OBJECTIVES

Assistive and Educational Robotics is one of the highest profile areas of HRI in the world. (GOODRICH; SCHULTZ, 2008). While the literature reveals positive benefits for learners in various degrees and abilities driven by the physical presence of the robot, it is still unclear what is it about the robot's appearance that promotes learning, or even further, which physical attributes and features are ideal in such environments (BELPAEME TONY, 2018). We also found a gap in the evaluation and comparison of these robotic embodiment, with a lack of materials that explored applications across different embodiments.

In order to better comprehend the role of the robotic embodiment in the context of learning, and elicit which in these elements falls into the preferences and thus engages the target audience, the present research compares two social robot models running the same music education HRI application. The HRI application targets the context guitar learning for children aged 9-11, geographically located in Brazil, Canada and United States. The Guitar Tuner consists of two main functionalities: first, the robot helps the user to tune an actual guitar string by string, and similar to a metronome, a second module helps them to play a song and evaluate their performance (MELO et al., 2020). We implemented the same sequence of HRI tasks according to the available robot embodiment features of two robot models using the NAO (SOFTBANK, 2021) and Zenbo (ASUS, 2021) robots.

To achieve answers we evaluate both HRI applications with children and guardians online, via internet conferencing tools, using an adapted version of the System Usability Scale (SUS) (BANGOR; KORTUM; MILLER, 2008) for children – the SUS-Kids (PUTNAM et al., 2020). We also compare via qualitative interview about their preferences on the robot models looking to understand how perceived robot embodiment features impacted their evaluation. The user evaluation supported quantitative and qualitative evaluation of the HRI application and highlighting the perceived differences of robot embodiment features. The discussions center on improving a future version of the HRI application, plus children's considerations about their preferred robot embodiment features. Finally, we propose recommendations for robot embodiment design for children based on this case study and regard a successful online user evaluation protocol during the social distancing context in the SARS-Cov2 (COVID-19) pandemic (VINER et al., 2020).

A reminder that this work organizes as follows. The following chapter summarize existing knowledge on social robot embodiment features, robots supporting music education, and comparative studies with social robots. The methods and materials chapter details the HRI application according to each robot embodiment design and describes the user evaluation protocol. The results chapter shows quantitative analysis of the SUS-Kids scores and other statistical results, and a discussion section targets the qualitative comparative evaluation. Finally, we list design recommendations for using social robots in music education according to the children's perspective and expectations of perceived robot embodiment features. We also highlight the challenges and limitations during online user evaluation and recommend improvements in the evaluation protocol supporting future research.

2 THEORETICAL FOUNDATIONS

In this chapter, we present a body of definitions and explore the existing literature regarding the role of social robots in musical education, the relationship between the robotic embodiment and learning, and works that evaluate the robotic embodiment features across devices and scenarios, distinctively for educational purposes. We present these works here by organizing the information within the underlying theories, models, and definitions of key concepts we used as grounds for our research. We achieved this by using search engines such as google scholar and digital libraries such as ACM Digital Library, IEEE Explore, and Science Direct, with the keywords SOCIAL ROBOTS, EDUCATION, EMBODIMENT, and COMPARISON. We specifically selected the results from literature reviews since producing a new one was not in this work's scope. We used the h-index and g-index as measurements. We obtained the ranked indexes using the Publish or Perish tool provided by Harzing (HARZING, 2007). Publish or Perish is a software program that retrieves and analyzes academic citations. It uses various data sources to obtain the raw citations, then analyzes against a range of citation metrics, including the number of papers, total citations, and the h-index and g-index, as mentioned earlier. We selected those works that consisted of already well-established and cited works and then snowballed backwards the bibliography considering those themes. We also divided this chapter into two main sections, one with the frameworks and definitions and another with the related works that reflects fundamental aspects regarding the motivation and main objectives of this dissertation thesis.

2.1 FRAMEWORKS AND DEFINITIONS

2.1.1 Human-Robot Interaction

Human-Robot Interaction (HRI) is a multidisciplinary field dedicated to understanding, developing, and evaluating robotic systems for use (interaction) with or by humans (GOODRICH; SCHULTZ, 2008). Due to the nature of the observed subjects of study (humans and robots), HRI shares synergies with various other disciplines such as human-computer interaction, artificial intelligence, robotics, natural-language processing, design, and psychology, to name a few. Goodrich also defines the robot as an artifact capable of performing human work autonomously or semi-autonomously. Most of what we understand as robots within this paradigm comes from the post-World War II period, where several advances in autonomous systems and computational power have given rise to a series of capabilities, possibilities, and usage scenarios. Interaction with these artifacts can occur in different proximity settings. Namely, remote interaction and proximate interaction, these distance base typologies are well accepted in the current literature.

Remote interaction occurs when the human and the robot do not share the same space and temporal location (e.g., rovers used for underseas or even extra-planetary exploration). In contrast, proximate interactions are both co-located. While these classifications can introduce a range of expectations about requirements for mobility, physical manipulation, or social interaction, many other authors introduced different taxonomies considering various points of view. Scholtz (SCHOLTZ, 2003) proposed a taxonomy of roles that robots can assume in HRI scenarios (such as supervisor, operator, or peer). Yanco (YANCO; DRURY, 2004) on the other hand, focuses on six pillars for categorization: task, robot morphology, team composition, spatial setting, level of autonomy, and interaction role.

Task-wise robots have multiple known applications, from search and rescue, to assistive robotics (as in aid for people with special needs, therapy rehabilitation, or for the elderly), military and police, edutainment (as in companion robots, classroom assistants), space exploration and inhospitable environments, and home and industry (ranging from robotic appliances like robot vacuum cleaners, modern smart-toys, to construction and factory robots). Task classifications may encounter challenges in classifying multipurpose robotic designs since they can transit within these fields or even pertain to more than one category.

Design-wise robots can take many appearances, with shapes, sizes, and morphological resemblances varying according to the available feature set and capabilities and the general application objectives. Common embodiment design strategies encompass anthropomorphic characteristics (resembling human-like appearances), zoomorphic characteristics (portraying animal-like features), and functional/technical characteristics (more abstract and linked to the robots function, as in exploration rovers or industrial robots). Although these design strategies may sound divergent between themselves, they are not isolated nor mutually exclusive, with the possibility of mixing various degrees of these designs in the same robotic device (e.g. companion robots that have facial expressions but resembles animals, like iCat developed by Phillips (BREEMEN; YAN; MEERBEEK, 2005)).

Team composition relates to the possible combinations between humans and robots in the same interaction scenario (not necessarily in the same physical space). It has three main aspects regarding the ratio distribution between humans and robots, the composition of those teams (as in whether or not they are composed of more than one kind of robot), and the level of interaction shared between those teams.

The spatial setting is associated with the disposition of humans and robots in a given analyzed scenario. As previously discussed in the work of Goodrich, they can be co-located (with humans and robots sharing the same space) or remote. Yanco's classification refines these possibilities with a spectrum ranging from avoiding, passing, following, approaching, touching, and none.

The level of autonomy reveals the amount of human control and intervention necessary for the robot to perform the design task. Robots that can navigate spaces and respond automatically to environment variables dynamically are considered autonomous. Guide robots that navigate indoor spaces and perform delivery tasks are good examples of this category. Conversely, teleoperated robots, like industrial arms or search and rescue drones, require a great deal of human intervention and are classified as non-autonomous.

Lastly, as seen in Scholtz (SCHOLTZ, 2003), interaction roles describe the positions that robots are subject to in HRI scenarios, or in other words, the positions that humans occupy in these interactions, namely supervisor, operator, mechanic, peer, and bystander). Yanco (YANCO; DRURY, 2004) incorporates these roles in his work, describing these roles between humans and robots. Robots with supervisory necessities have their behavior monitored by humans but do not necessarily need direct manipulation. In the same fashion, an operator needs to receive more intervention from the human, being teleoperated when needed to change it's behavior. A teammate/peer works with the human to accomplish a task. It is very present in manufacturing, normally executing repetitive tasks to construct a given industrial part while the human counterpart works on other aspects. A mechanic or programmer needs to change the robot's hardware or software physically. A bystander is a human who does not control the robot but needs to understand what the robot is doing to be in the same space. Scenarios where movable autonomous robots transit and should be avoided by the bystander consist of examples of this classification.

2.1.2 Social Robots

A social robot is a device that supports Human-Robot Interaction (HRI) tasks through robot embodiment features (shape, size, motors, sensors, displays, etc.) and adapts its intelligence and behavior through the perception of specific social cues (e.g., voice commands, gestures, facial expressions, etc.) (BARTNECK; FORLIZZI, 2004). In this sense, we have requirements and implications within this definition. Firstly, the necessity of physical embodiment. The application cannot be merely a virtual agent in a screen. It needs to be present in a physical space. Also, the robot needs to have some degree of autonomy, since the portrayal of intelligence and behavior occur in a timely manner, with the communication being a key aspect. The more seamless and natural, the more it will be acceptable and attend to the social expectations of the users.

A key difference between other kinds of robots and socially interactive ones is that the way in which a human perceives the device establishes expectations that guide his interaction with it. This perception, especially of the robot's intelligence, autonomy, and capabilities is influenced by numerous factors, both intrinsic and extrinsic (FONG; NOURBAKHSH; DAUTENHAHN, 2003). In this sense, building a good robot with the goal to satisfy human interaction may prove to be quite a challenge. Designs that are excessively cute and childish can be deemed as a toy, thus narrowing the window of interest to children. In contrast, an over realistic design can introduce significant noise in the general perception and acceptance, making the robot seem uncanny and weird. This phenomenon presented first by Mori is called The Uncanny Valley (MORI, 1970), where users report discomfort while interacting with something that looks like a human but it's clearly not. Other examples of uncanny responses to artificial humans can be evidenced in computer-generated characters in films, humanoids in science-fiction and horror movies, or in photo-realistic sculptures.

Social robots' typically portray human-like features, including speech, gestures, movements, eye-gaze, and the capability to establish logical reasoning and dialogue by processing personal data and users' social background. The sum of these elements concedes the social presence of the robot (BARTNECK et al., 2020). Many social robots are humanoid or animal-like in form, although this does not have to be the case since people, in general, tend to anthropomorphize these artifacts (FINK, 2012) either by the observed embodiment features and/or the general perceived behavior. In such communications, information is often exchanged nonverbally and verbally, deeply embedded in affective or emotive factors. As a consequence, modeling emotions play a powerful role in communicating intent, fostering people's social connection to the robot, aiding people in learning how to use and achieve the presented tasks while also enhancing likeability, engagement, and the desire to collaborate (RIEK et al., 2009). Paiva et al. (PAIVA; LEITE; RIBEIRO, 2014) describe the current advances in emotion modeling for social robots. They begin by contextualizing the role of emotions in social robots and describe several nonverbal elements for synthesizing and expressing emotions through robotic embodiments, including facial expressions, limbic and corporal resources like movement and size, sound, voice and pitch, colors, and lights. Their work concludes that full facial expressions are more straightforward to model than limbic and corporal expressivity.

2.1.3 Human-centered Design

Human-centered design (HCD) is a strategy for problem-solving typically used in fields such as design, management, and engineering that develops solutions to problems by involving the human perspective in all steps of the problem-solving process (IDEO, 2015). HCD frameworks typically entangle observing the problem within its context, brainstorming, conceptualizing, developing, and implementing the solution.

As seen in ISO 9241-210:2019(E):

> Human-centered design is an approach to interactive systems development that aims to make systems usable and useful by focusing on the users, their needs and requirements, and by applying human factors/ergonomics, and usability knowledge and techniques. This approach enhances effectiveness and efficiency, improves human well-being, user satisfaction, accessibility and sustainability; and counteracts possible adverse effects of use on human health, safety and performance.

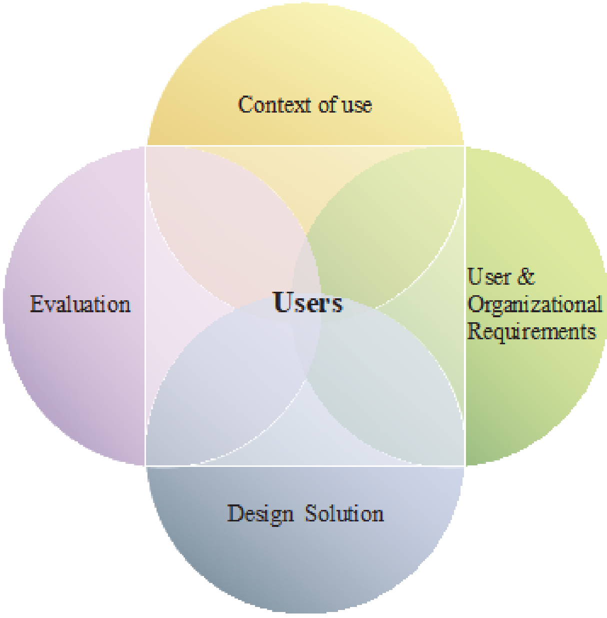
Figure 1 illustrates the processes' intersections and how the user permeates the activities in each scope of the HCD approach.

According to IDEO (IDEO, 2015) typical HCD cycles should encompass four activities during the design of any interactive system. These include understanding and specifying the context of use, the user requirements, producing design solutions, and evaluating the design.

The discovery phase consists of an analytic activity during which the design team learns who they will be designing the solution for and how they will use it. There is a wide range of activities that can facilitate this. Possibilities include surveying and bench-marking existing solutions, evaluating published research on the task domain, and interviews with potential users and stakeholders, among other resources.

Based on the analysis of the Discovery phase, the specification of the interaction design aspect of the project is set out. This becomes the blueprint that the interaction designer

Figure 1 – HCD Desing Process - as seen in The National Institute of Standards and Technology



Source: National Institute of Standards and Technology (NIST).

undertakes to satisfy the creation of the design.

As standardized by the HCD, the design phase should encompass the production of more concrete (development of prototypes) design solutions. By producing a testable version of the proposed solution, designers can plan how to assess and iterate via the users' feedback and fix the solution before committing to a final product.

Moreover, User-centred evaluation is a required activity of the HCD standard and should be carried out throughout the project.

2.2 RELATED WORKS

2.2.1 The Role of Social Robots in Music Education

Computer-based technologies can support music education in developing an individual's aural, performance, and composition skills, including supporting distance learning and strengthening self-efficacy and independent learning skills (WEBSTER, 2007). Technologies can reinforce existing learning strategies and encourage more people to learn music. Mobile applications are consolidating as pedagogical resources for music education since smartphones and tablets allow direct manipulation of objects and multi-tactile interactions, presenting excellent results in electronic musical instruments and supporting general music educational projects (SASTRE et al., 2013). In a study on technology use and self attitude toward music learning, authors surveyed 338 individuals using different devices supporting independent music learning skills and teaching music. Devices ranged from smartphones to tablets, laptops, computer desktops, smartwatches, television, audio and video recording, and playback devices (WADDELL; WILLIAMON, 2019). Survey results showed that individuals often use computer technology to run digital versions of classic music devices such as metronomes and tuners. Another finding is that they do not evaluate audio or video recordings in most situations, suggesting that automated performance feedback can become helpful. Interactive and Artificial Intelligence (AI) technologies can enhance those functionalities and improve the experience of using technology in music learning and teaching (GORBUNOVA; HINER, 2019). For instance, authors have used a game-based application supporting music learning in early childhood education to facilitate training sound perception skills and identifying sounds and notes in an octave of the musical scale (PAULE-RUIZ et al., 2017).

Several studies in music education have incorporated interactive technologies and artificial intelligence (AI), with a focus on children as the target audience. One study explored the use of virtual reality (VR) and augmented reality (AR) technologies in music learning, utilizing head-mounted displays and hand-held controllers to allow for the combination of VR training and musical instruments (SERAFIN et al., 2017). Another study combined VR and AI by creating a virtual social robot learning environment for music education, where two virtual versions of the NAO robot taught children how to play different notes and rhythms on a xylophone and music notes on a drum in a virtual setting (SHAHAB et al., 2021). This approach was evaluated with autistic children over 20 weeks, and it showed positive results in enhancing their music learning skills. Social robots can also contribute to music education in hands-on learning experiences, such as using modular robotic kits to develop STEAM projects involving dance, music, and culture, or even composing songs by turning the modular robots into physical instruments. These strategies can be especially helpful in environments where there are no physical instruments or robots available for learning.

Music education can be enhanced through the use of social robots in various ways. For example, a study on hands-on learning showed that children were able to develop STEAM projects involving dance, music, and culture by using the modular robotic kit KIBO. These projects involved assembling and programming modular robots to create dancing robots that were related to a specific culture and local music (SULLIVAN; BERS, 2018). Similarly, in another study utilizing modular robotics, children were able to compose songs without requiring prior musical knowledge by using the robots as physical instruments (NIELSEN; BÆRENDSEN; JESSEN, 2008). These approaches demonstrate the potential for social robots to expand the scope of music education and provide unique and engaging learning experiences for children.

Toy User Interfaces (ToyUI) combine hardware and software components to allow social and physical play experiences (WHELER et al., 2021). A ToyUI can combine toy components with companion devices like smartphones, tablets, and social robots. A social robot is a ToyUI component that can support active and passive social roles, acting as a co-player or guiding the play rules. Mainly, music education systems fall into the playful training ToyUI categorization (ALBUQUERQUE; KELNER, 2019). In this scenario, playful training examples usually mix interactive, mixed reality, and robotic technologies to enhance tangible interaction using physical musical instruments (LÖCHTEFELD et al., 2011; YAMABE; NAKAJIMA, 2013). The benefit is that the child can practice or train new skills while using the actual musical instrument. For instance, a study uses the NAO robot to teach children with autism to play xylophone (MALIK; YUSSOF; HANAPIAH, 2015). The robot is programmed to listen and assess students' music performance when playing a song using a musical instrument. Note detection occurs mixing audio processing techniques with image processing by using the xylophone keys as color descriptors. Several studies using social robots target autistic and neurodivergent children, which also occurs regarding music education (MALIK; YUSSOF; HANAPIAH, 2015; SHAHAB et al., 2021; TAHERI et al., 2021).

The present study aims to assess children's and guardians' expectations on social robots as companion devices supporting independent music learning skills using an acoustic guitar. The study also innovates by comparing their perception on different robot embodiment features and implementing an online evaluation protocol for HRI applications during the social distancing context in the COVID-19 pandemic.

2.2.2 Evaluating Social Robot's Embodiment

Many studies that compare different robot embodiment features tend to focus on user perception and task performance by comparing virtual agents and physical robots or teleoperation against co-located human-robot interaction experiences (WAINER et al., 2006; KENNEDY; BAXTER; BELPAEME, 2015; THELLMAN et al., 2016). For example, a study compares the game Tower of Hanoi supervised by a social robot in three different settings: virtual, teleoperated, and co-located (WAINER et al., 2006). Users experienced a virtual robot version running in the Gazebo (KOENIG; HOWARD,), teleoperation through video conference, and human and robot co-located in the same room. The user performed tasks of moving stacks while monitored by the system, and the robot assumed a role of an assistant, advising and reacting to the user's task decisions. The findings of this comparative evaluation, by focusing on task performance, suggest that users generally perform better in co-located settings as opposed to virtual or teleoperated settings.

Another comparative study evaluated children's preferences and performance while learning from different tutors: humans, tablets, and social robots (WESTLUND et al., 2015). Although there were no significant results in terms of learning retention, most children demonstrated substantial interest in learning activities with the robot as a tutor. While it is worth noting that these studies do not focus on long-term interactions nor validate if that given level of interest varies over time, we should also clarify that our work also faces this limitation.

Conversely, our present research aims to investigate children's perceived usability, likeability, and robot embodiment preferences comparing two robot models, NAO and Zenbo, running the same HRI application for music education. A similar feature-based approach compared 14 social robot models according to robot embodiment criteria (PAPAKOSTAS et al., 2018). Evaluation criteria included multimodality aspects (e.g., voice, movements, led blinking, etc.), flexibility towards the operational environment, cost, human-likeness, programmability, energetic autonomy, hardware performance (e.g., speed, readiness, and compute power), and built-in educational resources from the manufacturer. They used a grading system from 1 (negligible) to 10 (superior) for each category, indicating how much each social robot satisfies the given criterion. In the case of criteria directly connected with specific technical attributes, they used the original specifications' values instead of the artificial grades. While the authors mentioned evaluating the robot models with education specialists, it is unclear how they produced the artificial grades. They presented a performance evaluation running the TOPSIS optimization method, measuring the proximity (closeness) of each social robot alternative to the best and as well as to worst social robot cases.

Notice that both Zenbo and NAO robots were evaluated in this study, ranking second and seventh place, respectively. The ranking posed Zenbo as more affordable, with better energetic autonomy, and with an overall better educational package. NAO performed far better regarding its degrees of freedom, human resemblance, and programming capabilities. However, hardware-wise, it is not as powerful or energy-autonomous, ranking lower than the other selected devices.

In a review on research trends in social robots for learning, the authors noticed an increasing trend of user evaluation studies with children, and that reported outcomes focused on usability or feasibility studies or assessing affective or cognitive aspects, or a combination of both (JOHAL, 2020).

From 2015 to 2020, over 60% of user evaluation studies occurred in co-located settings compared to teleoperated systems. The majority of studies evaluated one-to-one experimental setups. Despite the many challenges of robots interacting with multiple users, some studies evaluated HRI applications in pairs, small groups (3-5 participants), and in the classroom (6 or more).

Learning systems often use social robots in combination with other tools and devices such as books, tangible interfaces, touchscreen displays, personal computers, and tablets. The authors of the study also identified different types of robot movements in human-robot interaction applications for learning. These interactions were categorized as either communicative gestures or manipulation. In the music education application discussed in the present study, social robots play a key role in providing task instructions, engaging users during and after tasks, and evaluating overall task performance by listening to the music played.

In Table 1 we have a comparison between the cited references in this section and our presented work. We separated in categories that highlight the kinds of agents present in these studies (if they were included virtual agents, the physical robot or both), the kind of interaction that the participants had with these agents (either simulated, co-located or remote), the general goals of these HRI experiments (being either the effectiveness of the task completion, evaluation of the learning retention, measurements of the engagement, or usability), the role of the agent (as an educational resource such as a book, a companion that stimulates the task execution or as an active tutor that replaces this human role), and the tested audience according to the age.

Summarizing, differently from the cited literature, where we notice a large focus on the comparison of the same depiction of a single robot embodiment in different interaction scenarios and how it affects their specific goals, our experiment contribution lies in the observations of different robotic embodiments, and the empirical data collected with the target audience in a remote circumstance.

The selected robot embodiment features range from robot motion as communicative gestures, smart speech, touchscreen interaction, image recognition, and audio signal processing. Due to the context of the COVID-19 pandemic, the evaluation occurred online and outside research facilities. Therefore, the evaluation used pre-recorded videos of the HRI tasks, and children engaged as observers.

Table 1 – Related work comparison

Reference number	Agent	Interaction	Goals	Role	Audience
(WAINER et al., 2006)	Virtual vs Physical	Simulation, remote, co-located	Task, engagement	Companion	Adults
(KENNEDY; BAXTER; BELPAEME, 2015)	Virtual vs Physical	Simulation, remote, co-located	Engagement, learning	Tutor	Children
(THELLMAN et al., 2016)	Virtual vs Physical	Simulation, remote, co-located	Engagement	Companion	Adults
(WESTLUND et al., 2015)	Virtual vs Physical	Simulation, co-located	Engagement, learning	Resource	Children
(JOHAL, 2020)	Physical robots	Co-located	Task, learning	Companion, tutor, and resource	Children
(PAPAKOSTAS et al., 2018)	Physical robots	Co-located, teleoperated	Engagement, usability	Companion, tutor, and resource	Children, adults, and elderly
The present work	Physical robots	Remote	Engagement, usability	Companion, tutor, and resource	Children

Source: Author.

2.2.3 General Considerations

This overview shows the theories and frameworks regarding Human-Robot Interaction and the design space for Social Robot applications, according to the surveyed literature in these fields. We highlighted the importance of creating both human-compatible and human-centered designs and how the robotic embodiment plays a crucial role not only in housing the general resources of the device but also in the mediation of social expectations of humans in interaction scenarios. These directly reflect the human perception against the projected role, behavior, and affectivity caused by or expressed by portrayed emotion or feeling. We also disclosed the relevant literature that relates to our research theme and objectives, positioning this work within the related literature. In the next chapter we present the Methods and Materials that are used in this work, with the description of the experiment and how it was conducted in order to meet the research objectives.

3 METHODS AND MATERIALS

The HRI application consists of a playful training application for music education with two learning modules: guitar tuning process and performance evaluation. The robot application pairs with an acoustic guitar to perform the HRI tasks. In order to better understand which feature set and robotic appearance produce the desired engagement and excitement for such educational environments, this research uses quantitative and qualitative methods, evaluating two social robot models running an HRI application with children and guardians.

The social robot guides the child in the tuning process by listening to them tuning a guitar string by string. The robot provides visual and speech feedback for each string by signaling to loosen or tighten the guitar's string. In the performance evaluation process, the robot listens to a song, provides information on music scores for the song, plays a metronome sound, and records the song to provide an AI evaluation performance. The robot reacts to music selection and music performance to engage the child in performing the task while improving their performance. Each version of the HRI application has particular design decisions according to available robot embodiment features (e.g., robot motion, touchscreen display, emotional expressivity). The following subsections explain the general purpose and the application's development context, compare its implementation in each robot model, and detail the user evaluation protocol during the COVID-19 pandemic.

3.1 THE GUITAR TUNER AND EVALUATION PERFORMANCE DESIGN

The first version of the playful training application supported implementation in the NAO 5 robot using the NAOqi SDK (MELO et al., 2020). It was a prototype developed in the context of the master's program, namely the discipline IN1169 - Advanced Topics in Media and Interaction, lectured by Professor Judith Kelner. This discipline used Human-Centered Design Tools to develop service robot applications. As defined in the curriculum, we used Human-Centered Design cycles to define the problem, a survey of state of the art as well as the available products in the market for alternatives, prototype a solution, and test it for feedback, all based on the attendees' collective interest. We achieved this by using brainstorming sessions and incremental refinements according to the application's target audience and the envisioned setup.

In our case, we have reached towards the Child-Computer Interaction Domain (CCI), we were interested in engaging/motivating children (ages 9 to 11 years old), into music learning, by facilitating tasks such as guitar tuning and practicing. This choice for ages from 9 to 11 years old was, at first, arbitrary because we thought that it would be ideal, aligned with the public's interest, in particular, given the robots' childish appearance, the application's complexity, and the learning window where children start to learn music. However, at that given time, it was not something that we had validated with the target audience. At the time, we had the NAO 5 robot available for implementation. We used its capabilities (e.g., sensing, motion) to interact with the users, collect and process musical signals into data, and give the children proper feedback. The NAO 5 robot embodies a humanoid appearance, showcasing articulated limbs, a torso with functional buttons, and a head with a static visage. Its general behavior resembles the one of a child, with a voice pitch and tone that is both boyish and infantile. This robot is a multipurpose device for a wide range of ages, perceivable as a colorful toy and a companion that responds to various environmental variables, including spacial awareness, world objects, and connected external devices.

The setup integration of one or more physical toy components with other hardware or software components constitutes a Toy User Interface (ToyUI) setup (ALBUQUERQUE; KELNER, 2019). Given the characteristics of the robot and our examined scenario, which includes children in a learning environment, we would benefit from the ToyUI setup framework and tools to undergo the design cycle towards the development of our application. The ToyUI framework categorizes applications and devices according to general objectives, genre, and in-toy capabilities (WHELER et al., 2021). The Serious Games and Applications (SGA) category, where our application fits, promotes content-driven play experiences that serve various operative purposes (e.g., learning and therapy). Beyond the exhibition of content, such devices should also be able to process relevant data to evaluate the users' performance and return actionable feedback. Our NAO application pertains to the Edutainment genre, introducing a ToyUI setup that supports theoretical and practical learning topics (Playful Training).

The ToyUI classification Tool proved to be very useful in narrowing down the scope by guiding the search for inspirational artifacts and the spectrum of available interactions for the scenario. We found a series of inspirational artifacts, relevant research, and applications in the SGA genre. We evaluated and benchmarked these against the range of opportunities presented in the alternatives we have generated and the constraints both in the application context and within the robot's feature set. We achieved this using the brainstorm ToyUI and the Robot Storyboard Ideas. The brainstorm ToyUI mixes traditional toys and play rule information to help creators generate ToyUI concepts. The Robot Storyboard Ideas Tool is a digital storyboard resource that offers specific robot embodiment information to creators. It consists of a digital slide template with embodiment features and scenario variables. It allowed us to quickly sketch ideas, produce group discussions, and plan the desired robotic behavior during the initial prototyping stages.

As seen in the related works, we have found some experiments using many form of digital signal processing techniques to acquire information about the users' performance while playing instruments or employing technologies to support small tasks related to music learning ((LÖCHTEFELD et al., 2011), (YAMABE; NAKAJIMA, 2013), (MALIK; YUSSOF; HANAPIAH, 2015)). Along with other inspirational artifacts such as video games like guitar hero and rock band, mobile platforms like PlayScore gave us a path to elaborate our solution. We chose the acoustic guitar as our instrument due to the group's familiarity with it, and a learning strategy that would encompass the training of two primary tasks: tuning and performing a music score. These two tasks are common to various other music learning frameworks, like the Suzuki Method (PEAK, 1996). Peak's work on this framework relates to practices analog to language learning and to our musical learning design, where the student learns by listening and experiencing and improves it with the introduction of techniques and practice by repetition of small tasks.

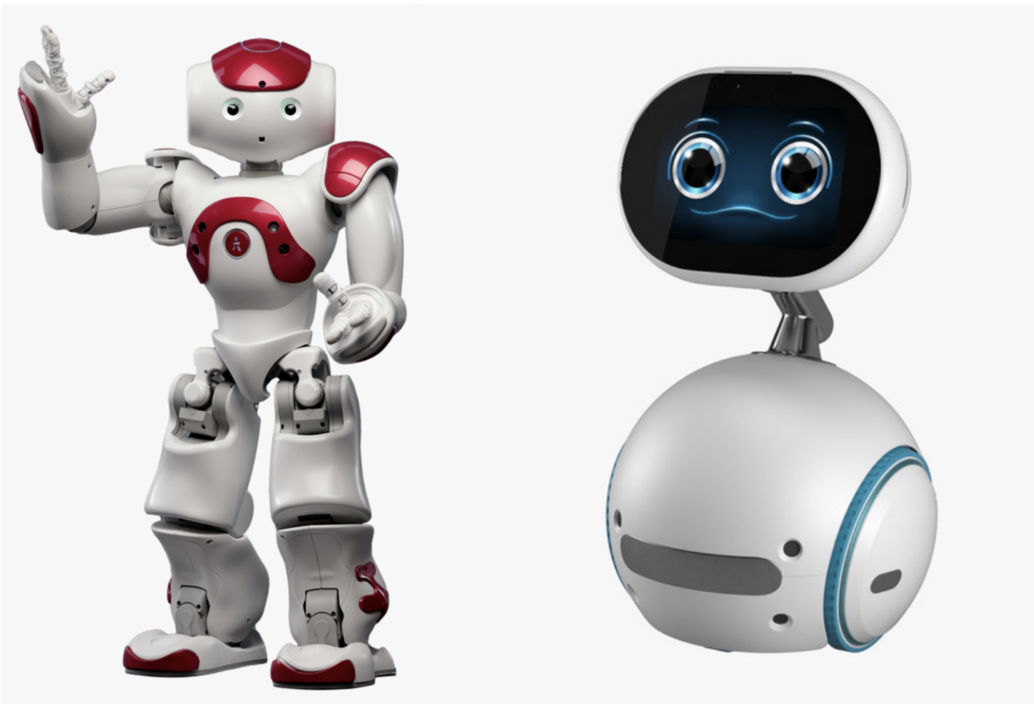
We leveraged the available microphones NAO robot. It has a frequency bandwidth from 150 Hz to 12 kHz. Some of the notes to be detected present a fundamental frequency below 150 Hz, e.g., the E note of the sixth string has a fundamental frequency equal to 82.42 Hz. The signal processing is performed through Goertzel's algorithm (GOERTZEL, 1958). It is a simplified version of Fourier analysis, where only the desired spectral component is evaluated. The decision about the presence of a note on the captured signal depends on its corresponding frequency's acoustic power. We use the total signal power as a floor reference to reduce the variability of how strong or weak the student plays the note. Then, the note is detected only when the amplitude given by the Goertzel's algorithm, relative to the signal power, is above a defined threshold. We use this detection procedure for both the tuning process and performance evaluation.

For instance, we could not find any approaches that used the same resources and design choices as seen in this project at the time of it's conception. More details about the technical aspects can be found in our publication Guitar Tuner and Song Performance Evaluation Using a NAO robot (MELO et al., 2020).

3.2 GUITAR TUNER AND EVALUATION PERFORMANCE DESIGN ADAPTATION

Here we adapted our existing NAO robot application to the Zenbo robot to compare different robot embodiment features with children and guardians. The goals are first to understand how robot embodiment features impact design decisions, then how these decisions impact children's perceived usability, likeability, and robot embodiment preferences. We chose the Zenbo robot as an alternative robot model due to the inherent robot embodiment differences compared to the NAO robot. Figure 2 depicts the actual embodiment of both devices. Table 2 compares NAO's and Zenbo's robot embodiment features. In overview, both robot models share humanoid design features, are movable, and support HRI through speech recognition and image processing. The NAO robot presents a traditional humanoid shape with articulated limbs, while the Zenbo robot does not have any limbs and uses wheels for navigation. The NAO robot has a static head limiting emotional expressivity. Differently, the Zenbo robot displays a set of facial expressions supporting greater emotional expressivity. Zenbo robot also offers more connectivity options, and its Operating System (OS) based on Android OS facilitates integration with mobile devices.

Figure 2 – Depictions of NAO (left) and Zenbo (right) robots



Source: Adapted from ASUS and Aldebaran Robotics.

Table 2 – Overview of robot embodiment features of each robot model.

Social Robot	Embodiment Features
NAO V5 (SoftBank, 2014—2018)	**Sensory:** Loudspeakers, microphones, video cameras, frs, imu, sonars, joint position sensors, contact and tactile sensors. **Connectivity:** Ethernet, Wi-Fi, and USB. **Emotion:** Static. **Movement:** Head, shoulder, elbow, wrist, hand (actuated hands and fingers), hip, knee, and ankle. Displays: RGB led on head, eyes, ears, and chest
Zenbo (ASUS, 2016)	**Sensory:** Digital microphone, 13M Camera, speaker, drop it sensor, Consumer ir CIR sensor, sonar sensor, line sensor, capacitive touch sensor. **Connectivity:** Wi-Fi and Bluetooth 4.0. **Emotion:** 24 cartoon facial expressions. **Movement:** Head, neck, and base. Displays: 12.6-inch touchscreen and wheels (RGB LEDs)

Source:https://www.softbankrobotics.com and https://zenbo.asus.com

We fully implemented the application in the NAO robot using the NAOqi Python SDK and implemented a rapid prototype in the Zenbo robot using the Zenbo App Builder. Note that the Zenbo robot does not incorporate the sound processing feature in this version, and we implemented this functionality using the Wizard of Oz (WoZ) technique (KELLEY, 2018).

The WoZ technique is a helpful resource to test interactive behaviors without fully implementing them. The goal is to implement the interaction and feedback, without hinting the user about the actual implementation status of the application while providing the expected functionality. We adapted the playful training application preserving the same sequence of steps and relevant HRI features to make it a fair comparison.

Tables 3 and 4 details the sequence of steps for both the tuning and performance evaluation processes, and how we implemented them in each robot. In general aspects, we kept the initialization sequence by touching the robot's head since both models offer similar touch sensors, and this helps to anchor the user to a common language between both devices. This same feature applies to starting the learning modules (e.g., starting the tuning process or playing a song).

Voice-based interaction remains challenging, and speech recognition services are still limited, often creating unexpected events and misbehavior. Initially, we decided to use NAO's head touch sensors to input and select tasks avoiding relying on voice inputs in general. In the Zenbo robot, we used its touchscreen and digital menus, making our design decision more explicit.

Table 3 – Comparison of robot design implementation: tuning process.

HRI task	NAO	Zenbo
Initialization: user activates the application.	The user touches the head sensor to activate the robot application.	The user touches the head sensor to activate the robot application.
Introduction: robot greets the user.	NAO stands up, and greets the user through speech and gestures.	Zenbo wakes up displaying a happy facial expression, and greets the user through speech and body movements.
Selection: robot offers options to available tasks.	NAO introduces the options through speech. The user selects between two different head sensors (A or B).	Zenbo introduces the options through speech, then shows a menu in the touchscreen display. The user selects options in the menu (1 or 2).
Select tuning process: robot provides instructions before starting up.	The user selects head sensor A. NAO provides instructions through speech about the tuning process, blinking the right and left eyes.	The user selects option 1 in the menu. Zenbo provides instructions through speech about the tuning process, blinking the right and left wheels.
Start tuning process: user is ready for the task.	The user touches the head sensor A.	The user touches the head sensor.
Tuning process: robot performs the tuning process with the user.	NAO blinks the right and left eyes to indicate whether the user should loosen or tighten the string, respectively. Robot indicates the process is complete by flashing both eyes at the same time, then proceeding to the next string.	Zenbo blinks the right and left wheels to indicate whether the user should loosen or tighten the string, respectively. Robot indicates the process is complete by flashing both wheels at the same time, then proceeding to the next string.
End tuning process: robot and user finishes the tuning process.	After the user is done tuning all desired strings, NAO informs that the process is complete through speech, and returns to an inactive position by sitting down.	After the user is done tuning all desired strings, Zenbo informs that the process is complete through speech, and returns to an inactive state by displaying a sleepy facial expression.

Source: Author.

The NAO robot offers several joints and articulations regarding robot motion, such as getting up from the floor, sitting down, and dancing. It also offers a mode of reproducing fine movements improving expressivity and lifelikeness.

However, implementing movements using the NAO robot demonstrated not to be a trivial task. We implemented facial expressions in the Zenbo robot when the NAO robot would significantly move towards emotional expressivity (e.g., dancing to celebrate or demonstrate sorrow). We developed a python script to support sound processing using the NAOqi SDK, which became a design priority instead of using robot motion at its best extent (MELO et al., 2020).

For that reason, the NAO robot would not move while running the sound processing script during the tuning and performance evaluation tasks, remaining static for most of that task. We implemented robot motion for initialization and feedback on each HRI task (e.g., greeting, dancing, standing up, and sitting down).

In turn, the Zenbo robot supports synchronizing the display of contents and animated facial expressions with body and neck movements to improve lifelikeness. The Zenbo robot can also move forward and adjust its head to look at the user, reinforcing emotional expressivity and attention. Despite not having any limbs, the Zenbo robot simulates dancing by making quick turns around its axis, displaying a singing facial expression, while flashing the available LED lights on its wheels.

Finally, the LED lights are another similar feature in both robots, which we preserved in the adaptation to maintain a recognizable pattern between the two versions. The NAO robot uses lights in the eyes as indicators to tighten or loosen the string in the tuning process, while Zenbo replicates this feature using lights in the wheels. Although the display could be used to perform this action, we have opted to maintain the similarity with the LEDs in order to facilitate comparison, since it conserves the same interaction parameters for the user.

The NAO robot uses a mobile companion application to support selection and display music scores concerning the performance evaluation process. The robot detects the selection by scanning NAOMarks on the mobile screen. The Zenbo robot's head is a touchscreen display, which facilitated us to implement all-in-one interaction, using the built-in screen and a selection menu for selecting the song and showing the music scores to the user. While this decision distanced the similarities in the interactions between both devices, it enabled a new set of comparisons with relevant results, as seen in our data analysis and conclusions presented in the further chapters.

A significant difference between the robots is that Zenbo offers a set of facial expressions, and we used them to compensate for the lack of lifelikeness regarding robot motion. We implemented facial expressions in the Zenbo robot when the NAO robot would significantly move towards emotional expressivity (e.g., dancing to celebrate or demonstrate sorrow).

Table 4 – Comparison of robot design implementation: performance evaluation.

HRI task	NAO	Zenbo
Select performance evaluation process: robot provides instructions before starting up.	The user selects head sensor B. NAO provides instructions through speech about the performance evaluation process. The user selects the song using a mobile app and shows a NAOmark to the robot tagged to music scores. Robot reacts to music selection through speech and gestures.	The user selects option 2 in the menu. Zenbo provides instructions through speech about the performance evaluation process. The user selects the song using another menu in the touchscreen display (a numbered list). Robot reacts to music selection through speech and facial expressions.
Start evaluation process: user is ready for the task.	The user touches the head sensor B.	The user touches the head sensor.
Performance evaluation process: robot starts the metronome, the user plays the song, the robot records it, and evaluates the user's performance.	NAO plays a metronome sound at 75bpm while recording and processing the user's audio. The user follows the music scores using the selection app.	Zenbo plays a metronome sound at 75bpm, displays the music scores on screen, while recording and processing the user's audio.
End performance evaluation process: robot finalizes the recording process, provides a score, and reacts to the user's performance.	NAO finalizes the recording process communicating through speech. The robot provides a score from 0 to 100, and reacts accordingly. A satisfactory score is above 70 points. The robot congratulates the user, dances while flashing rainbow lights, and plays a happy song. The robot reacts with sorrow by flashing blue lights, covering its face with its hands, and playing a sad song. After reaction, the robot returns to an inactive position.	Zenbo finalizes the recording process communicating through speech. The robot provides a score from 0 to 100, and reacts accordingly. A satisfactory score is above 70 points. The robot congratulates the user, dances while flashing rainbow lights, displays a happy facial expression, and plays a happy song. The robot reacts with sorrow by flashing blue lights, moving the head down, displaying a sad facial expression, and playing a sad song. After reaction, the robot returns to an inactive state.

Source: Author.

3.3 COMPARATIVE USER EVALUATION PROTOCOL

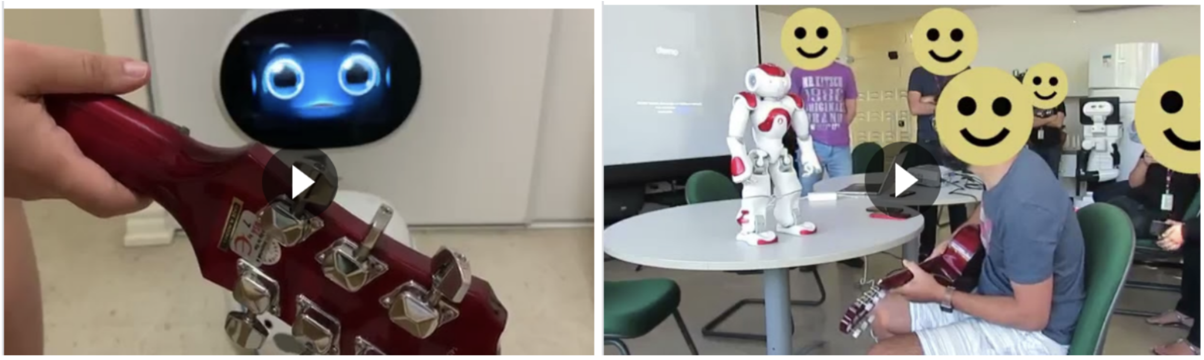
The online comparative user evaluation protocol consisted of the following steps and materials. First, recruitment occurred online using a call to action video disseminated in social media and instant messaging platforms (e.g., Instagram and Whatsapp). The recruitment targeted guardians with children who were English speakers (native or bilingual), residing in any country, and with any level of music education. We decided not to restrict age groups or gender aiming to assess the limitations of the application design. We scheduled interviews online after guardians fill out the informed consent forms via Google Forms. We included a short questionnaire to obtain sociodemographics on the guardians, including gender, age, location, occupation, and educational level. We also sent all research instruments beforehand, including an anonymous children's profile questionnaire and evaluation questionnaire.

Online interviews used either Zoom or Google Meet platforms, and recording was conditional to guardian approval. We stored automated recordings in the institutional cloud with restricted access to the researchers for further data analysis. We conducted the interviews in pairs to overcome casualties (e.g., weak or losing internet connection). Guardians could opt to participate or not in the interviews or supervise by distance, and we interviewed more than one child at the same time in family interviews settings. Children would fill their profile questionnaires before or at the beginning of the session, filling out independently or with assistance from the guardian or the researchers. The children's profile questionnaire was anonymous, covering gender, age, and experience with robots and music education.

The evaluation protocol followed a novel strategy to evaluate systems with children-guardians online during the COVID-19 pandemic (WHELER et al., 2021). We introduced the music education application using a storyboard template for each robot model first, followed by a recorded video of the actual robot in sequence. We leveraged the existing ToyUI storyboards from NAO to replicate and represent the same steps for Zenbo. The order of robot models presentation was randomized to avoid any preference bias. The storyboard and demonstration videos followed the same script and size, containing 15 scenes each and 5 minutes of duration, respectively. We edited the videos to introduce the same time frame, sequence of events, labels, and captions, but the audio and setup quality of the videos were substantially different. We recorded the NAO video in a public presentation on campus, showcasing the prototype to a live audience (MELO et al., 2020).

Differently, we recorded the Zenbo video at home without an audience or noise interference.

Figure 3 – Screenshot of video footage.



Source: Author.

Also, the Zenbo video displayed the robot on the floor, as referenced in Figure 3, and the NAO video showcased the robot on a table, and we anonymized all participant's faces. We could not record a second video using the NAO prototype due to limited access to the universities campi and research facilities. A reminder that we fully implemented the NAO prototype, but the Zenbo prototype used WoZ to demonstrate both tuning and performance evaluation processes, making it easier to script robot reactions and feedback to HRI tasks. Besides, the live audience would spontaneously react to HRI tasks along with the NAO robot prototype.

3.3.1 Robot Storyboard

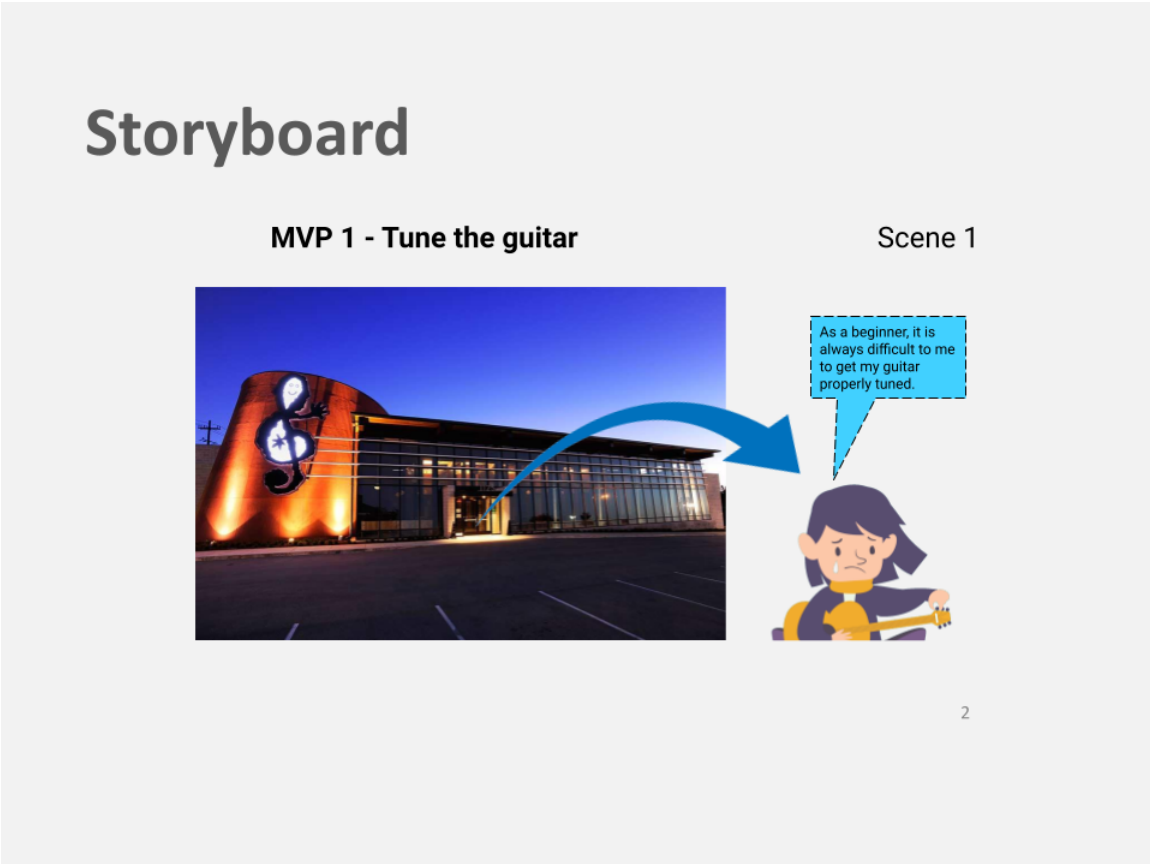
The Robot Storyboard Ideas Tool once used to sketch ideas, produce group discussions, and plan the desired robotic behavior during the initial prototyping stages was applied as part of this protocol to introduce the audience to the HRI application and enhance the comprehension of the context, the design space, the role of human and the robot in the interaction, the expected behaviors of these robots and put in evidence the utilized features.

3.3.1.1 Tuning Process

In figures 4 and 5 we introduce the the interviewees to designed scenario for our HRI application. The robot plays the role of an aide to the process of music learning for children. The users would seek help from the device on basic practicing activities such as tuning the instrument and the evaluation of performance. The interactions would take place in a one on one basis, in an indoor setting, like in a music school or in home.

Figure 6 demonstrates the first contact with the device. The activation of the loop relies

Figure 4 – Storyboard - Presenting the Interviewees to the Context



Source: Author.

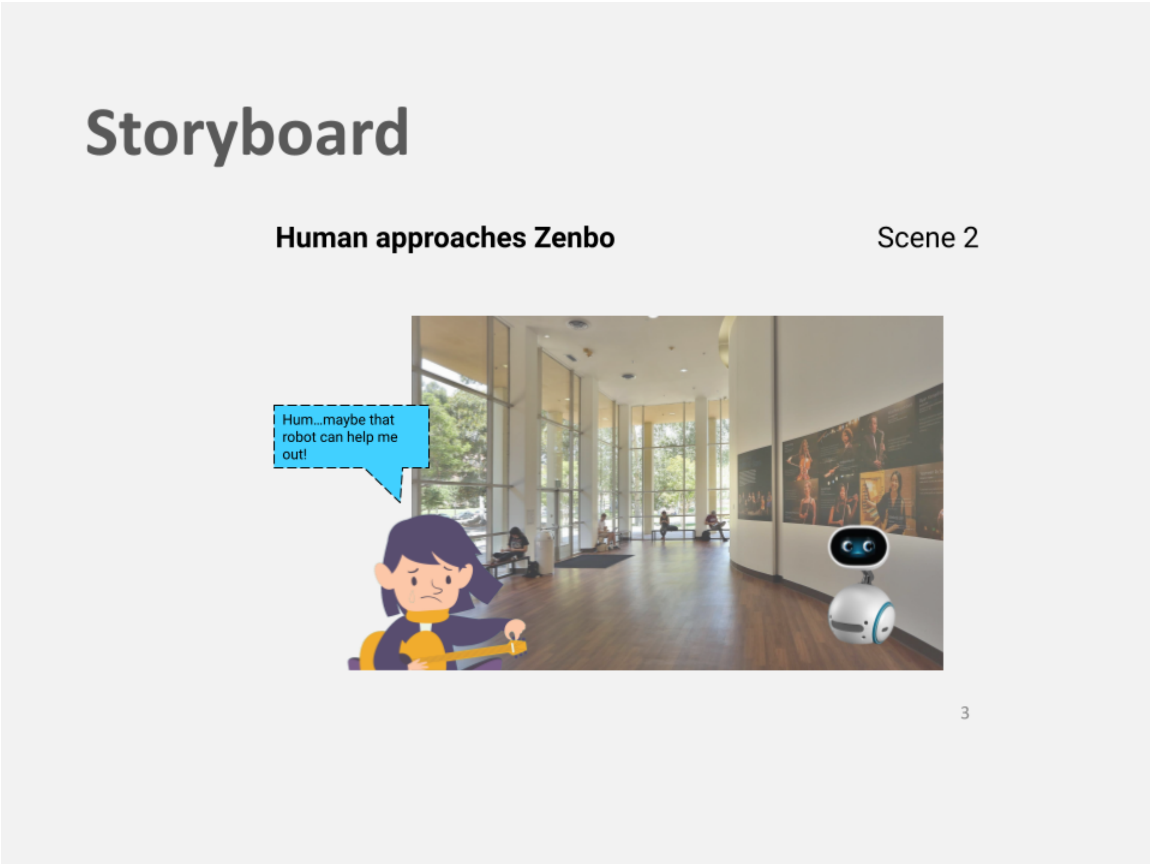
on the touch sensor existent in the head of the robot. This design decision was also made to the NAO robot as well. Although it would be interesting to have this activation in a more automatic manner, that would be costly to the robot and would increase the complexity of the prototype for something that isn't the core functionality of the application. We also avoided the usage of voice commands and activation. Both NAO and Zenbo perform poorly in that regard. Natural language processing is challenging and the native frameworks available for these devices proved to be insufficient for real-time fluid interactions.

Next steps illustrated by figures 7 and 8 the self introduction of the robot. Here Zenbo moves forward in order to present himself, and enable the user with two possible courses of action (tuning and performing respectively).

The user proceeds in figure 9 with the selection of the desired application within the robot's built in display. We firstly showcase the tuning process. As a reminder, NAO doesn't have a screen and his version of this step was firstly implemented with the help of the available touch sensors on top of the head of the device.

The robot confirms, as seen in figure 10 the selected activity and starts a brief tutorial for

Figure 5 – Storyboard - Presenting the Interviewees to the Scenario



Source: Author.

the interactions. The user should play each of the guitar's strings in the order dictated by the robot paying attention to it's signaling. When the left LED on Zenbo's wheel gets blue (figure 11) the user should loose the string. Otherwise, should the right LED on Zenbo's wheel get blue (figure 12), the user will need to tight the string. When both of those LEDs are blue, as in figure 13, the string is tuned and the user can proceed to next one 14. Similarly, NAO uses the LEDs present on his head (placed around the speakers) indicating the same actions.

Figures 15, 16 and 17 depicts the process explained in the tutorial phase, with the user loosening and tightening the string as indicated by the robot.

Once tuned, the user moves to the next strings (18), from the lowest to the highest, till there are no more strings to tune. Zenbo then returns to it's resting position (19), finishing the loop (20). NAO finishes his loop in the same fashion, returning to it's regular sitting position.

3.3.1.2 Performance Evaluation

In Figure 21 the users select the second application, regarding the performance evaluation process. Zenbo then invites the user to select one of the available tunes for practicing (Figure 22). NAO's version of this step was implemented using image recognition instead. We leveraged from the existing NAO mark structure to make these choices, since it would be more scalable then using the touch sensors, should we increase the number music scores at hand.

Once chosen the song (Figure 24), Zenbo explains how the process works and asks for confirmation of when should the evaluation start (Figures 25 and 26). Notice that we use the same confirmation pattern by pressing the top head sensor. This creates a common interaction language for both robots.

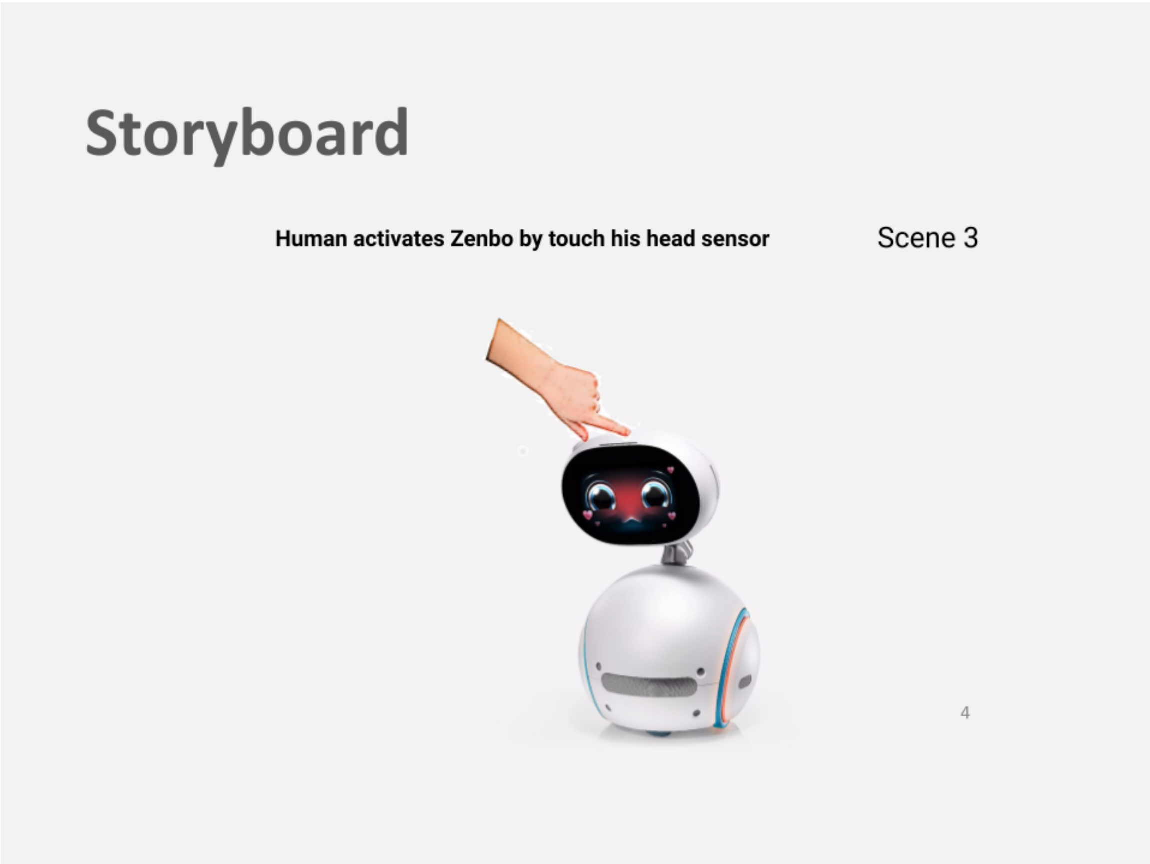
Figures 27, 28 and 29 depict the performance evaluation and metronome functionality respectively. While the user actively plays the song, the robot processes the signals from the guitar in order to produce the assessed feedback. Both NAO and Zenbo perform it in the same fashion, remaining static during the procedure while marking the pace by emitting metronome sounds. One difference worth of mentioning is that Zenbo displays the music score at the built-in display, while with NAO the user should follow his own printed music score or use the available companion app.

Once finished the song (Figure 30), Zenbo produces spoken feedback of the evaluated performance. Good performances, where the user plays at least 70 percent of the notes correctly, are awarded with positive feedback in the form of music, dancing and happy facial expressions. In insufficient performances, where the user plays at below 70 percent of the notes correctly, Zenbo warns that the user should practice more and try again, while manifesting sadness with position of the head, the portrayal of sad facial expressions and a sad song. As seen in table 4, NAO compensates the absence of full facial expression by the usage of LEDs in the eyes and the movements of his body. We achieved a representation of sadness by covering his face with the hands and slightly curving the body forward.

3.3.2 Interview and Data Collection

After presenting both storyboards and videos to the participants, the researcher would send or help the child to fill out the evaluation questionnaire. The evaluation questionnaire was adapted from the SUS-Kids (PUTNAM et al., 2020), consisting of 13 statements using

Figure 6 – Storyboard - The user activates the robot

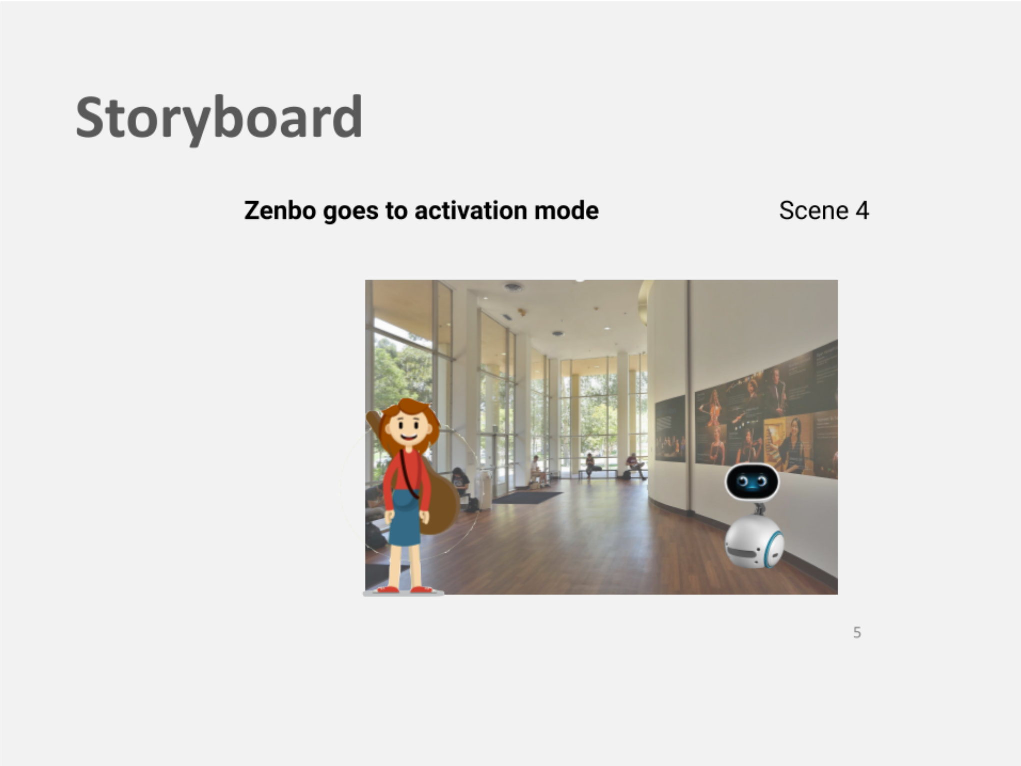


Source: Author.

a 5-point Likert scale ranging from (1) “I strongly disagree” to (5) “I strongly agree.” The authors adapted ten statements from the original scale to facilitate language for 9-11 years old, and added three additional statements on likeability and enjoyment based on related works (ZAMAN; ABEELE, 2010; READ, 2012). They also suggest using a visual Likert scale to facilitate assessment with children, so we used an Emoji-Likert scale for each statement.

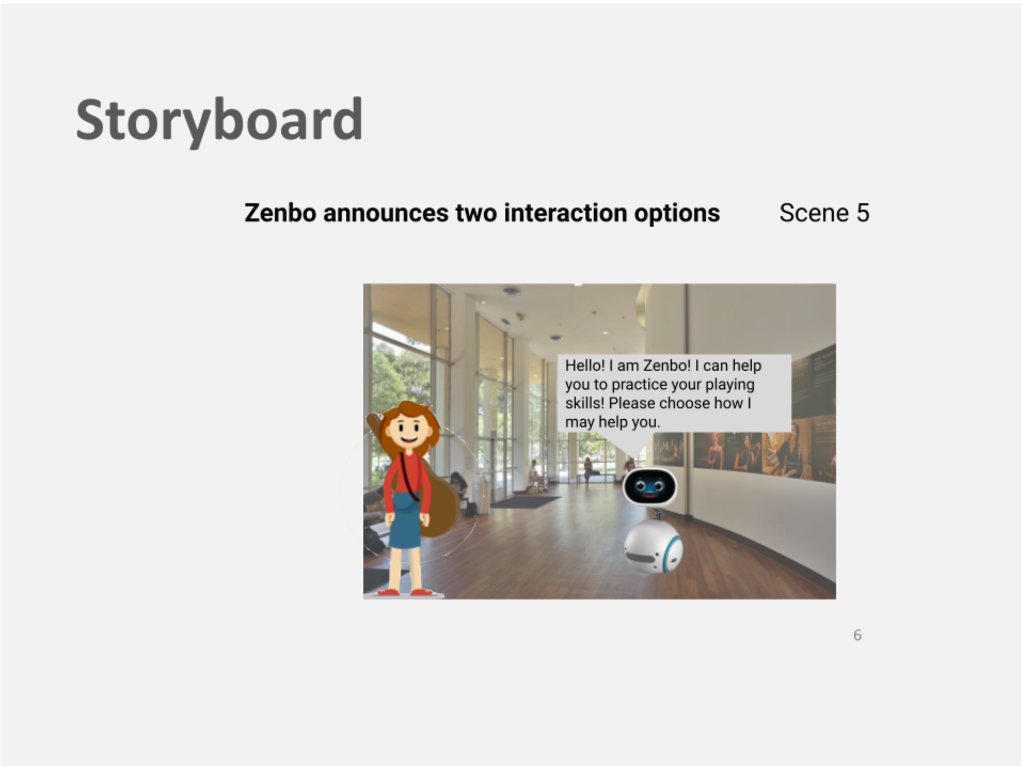
In Table 5, we adapted the SUS-Kids statements to the context of social robots, also considering that evaluation would use a video demonstration instead of an active usage scenario. Finally, we included three additional questions for the qualitative evaluation asking which robot they like the most (displaying name and picture of the robot), why, and if they had any suggestions. We randomized the order of the robots in the questionnaire to prevent misleading or bias.

Figure 7 – Storyboard - The robot gets closer to the user



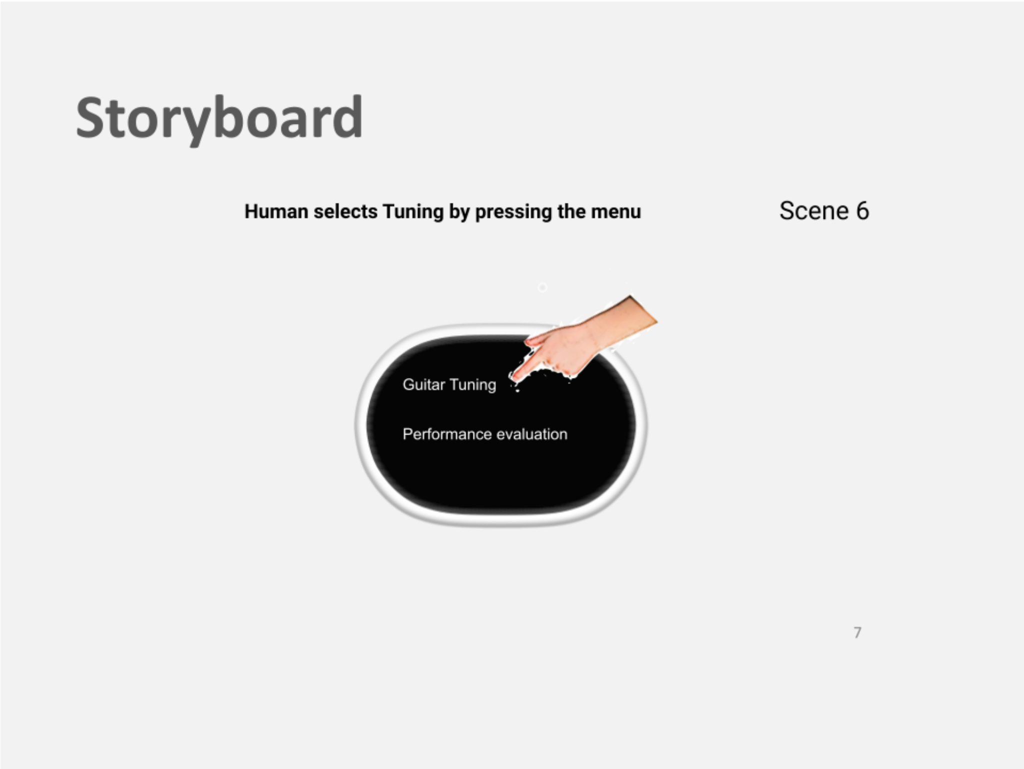
Source: Author.

Figure 8 – Storyboard - The robot introduces himself



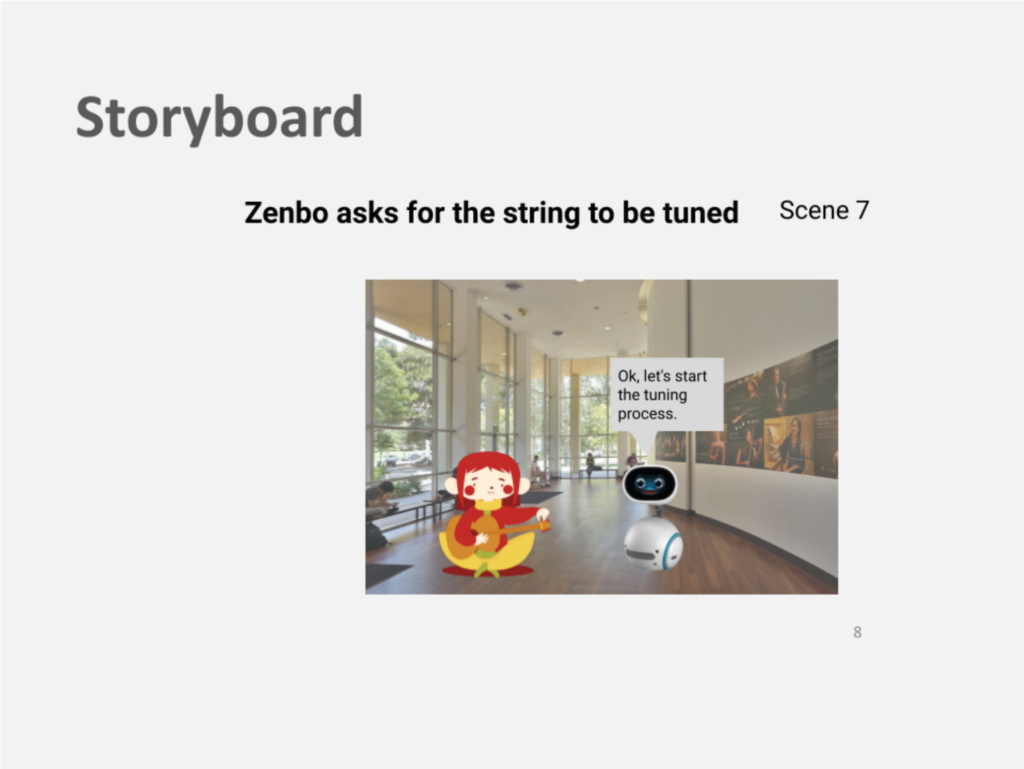
Source: Author.

Figure 9 – Storyboard - The user selects the first HRI application (tuning)



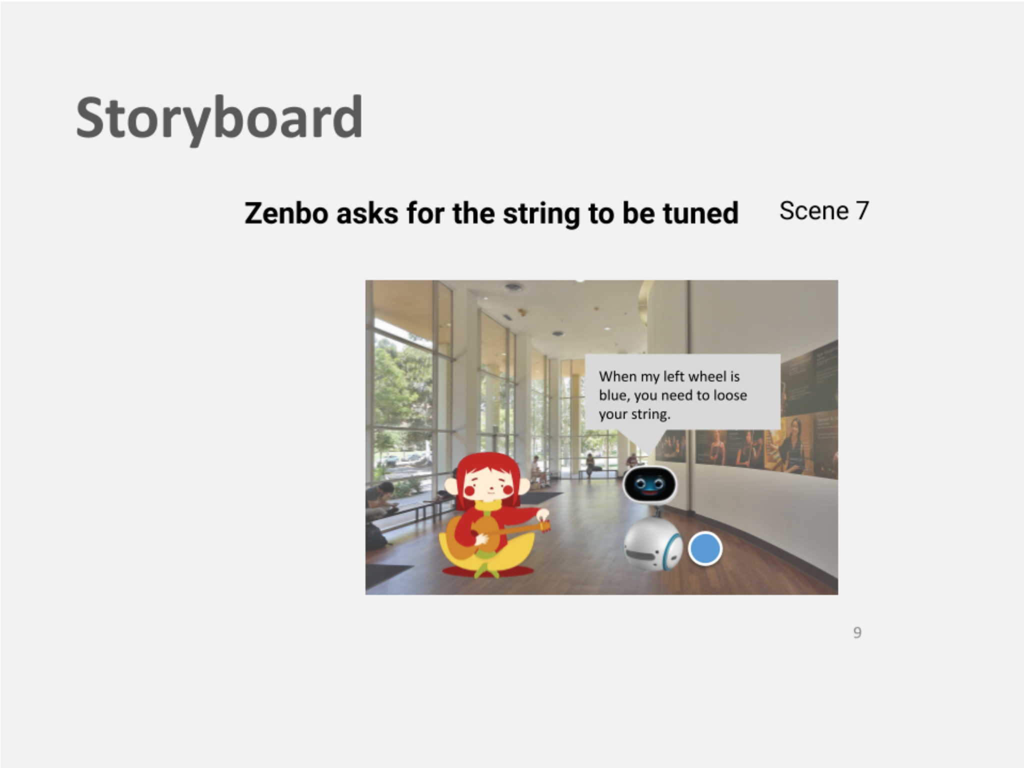
Source: Author.

Figure 10 – Storyboard - The robot confirms the selected application



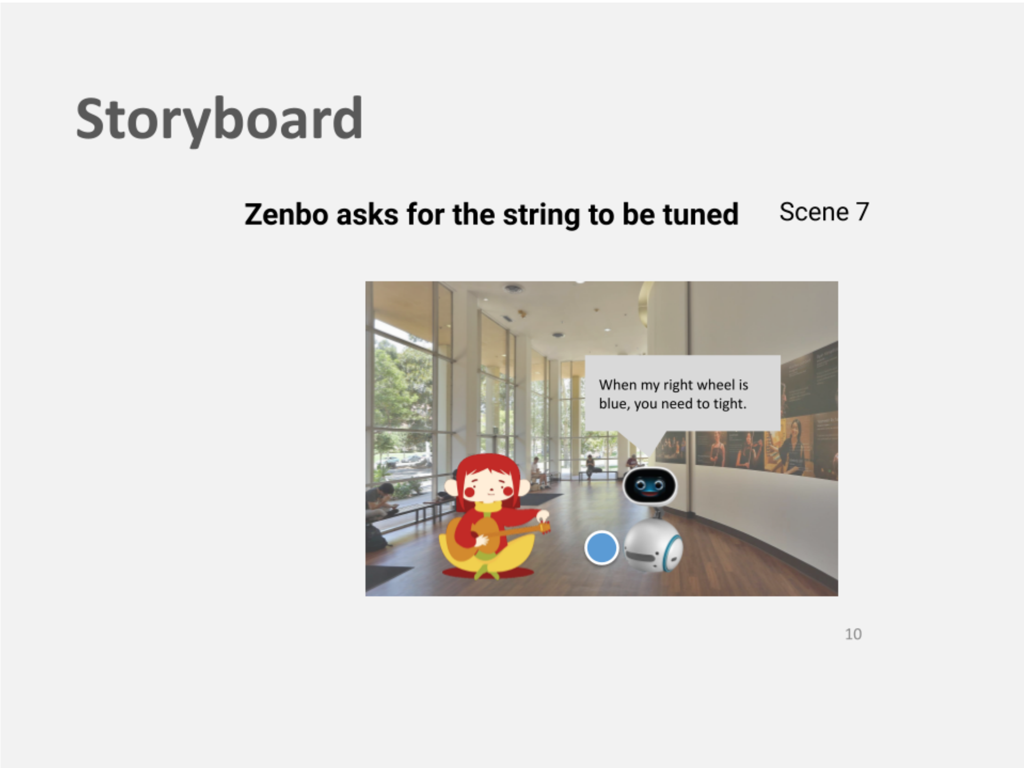
Source: Author.

Figure 11 – Storyboard - How to loosen the string



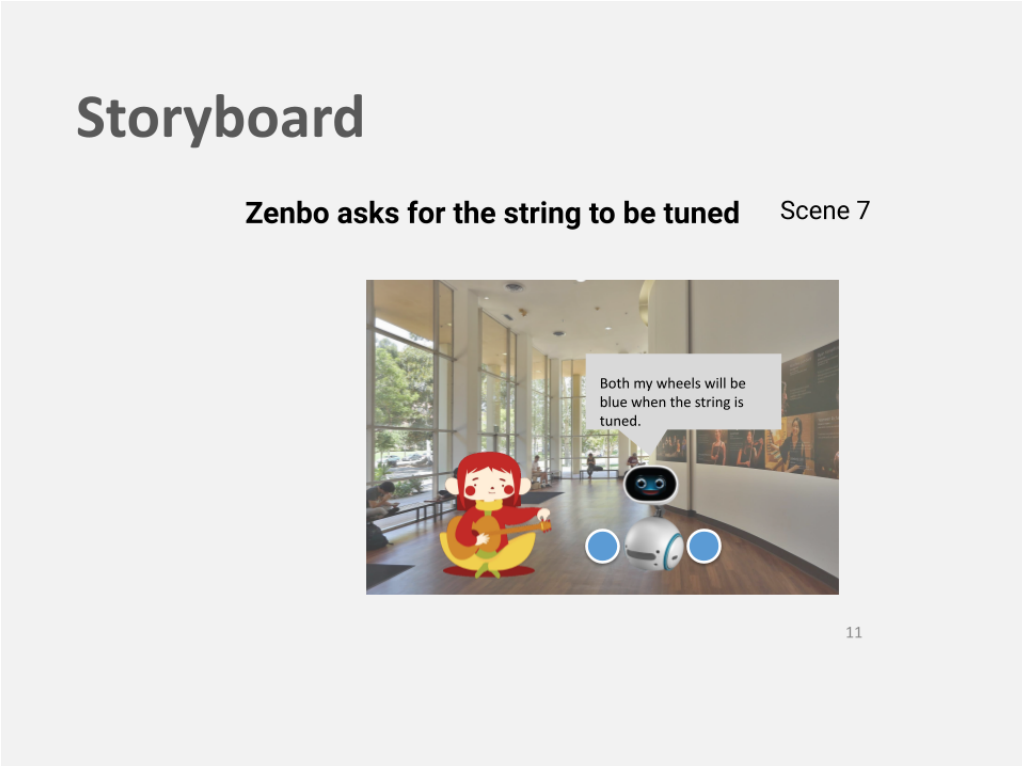
Source: Author.

Figure 12 – Storyboard - How to tighten the string



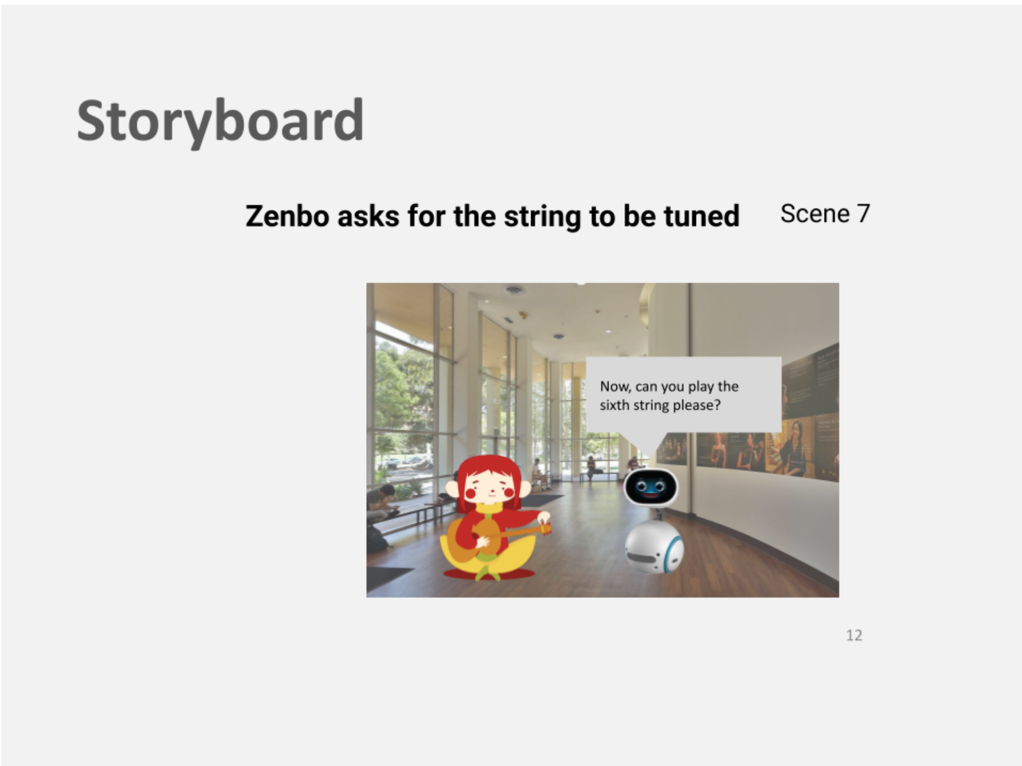
Source: Author.

Figure 13 – Storyboard - How the string is tuned



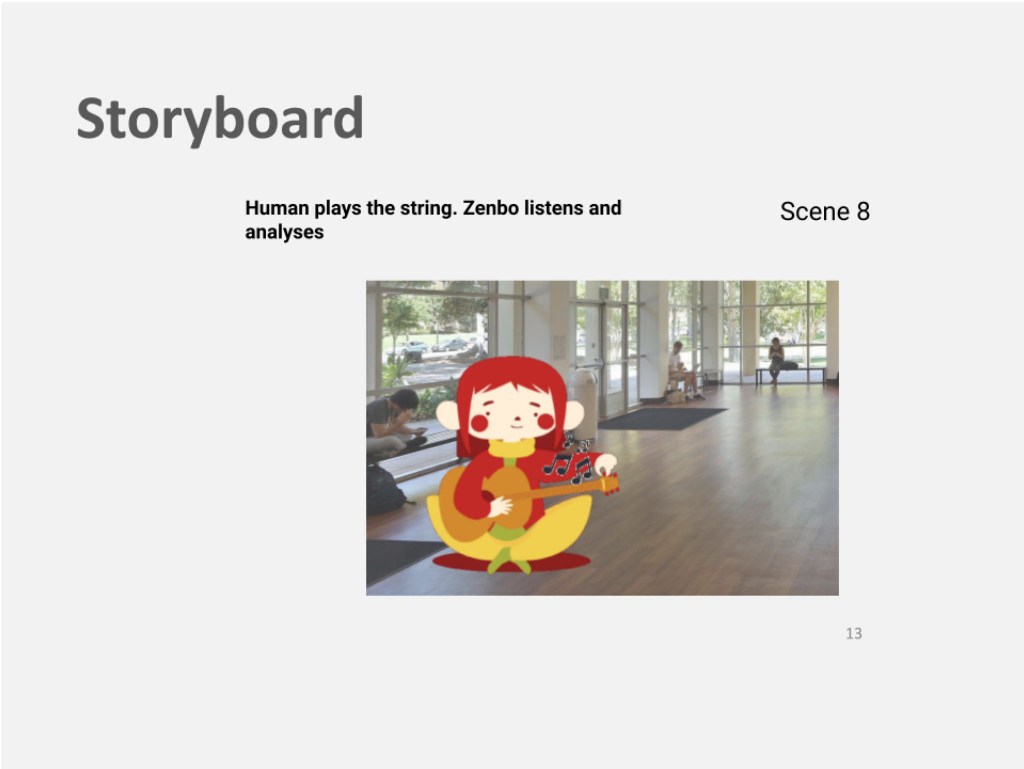
Source: Author.

Figure 14 – Storyboard - Zenbo asks for the next string



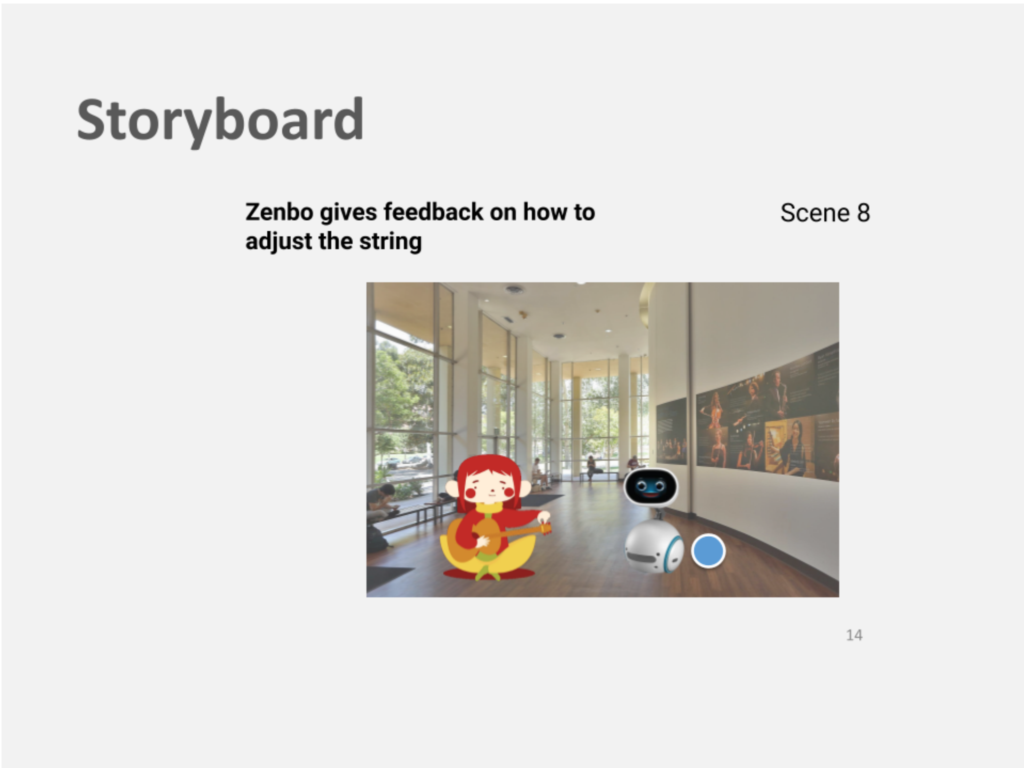
Source: Author.

Figure 15 – Storyboard - User plays the string while Zenbo listens and analyses



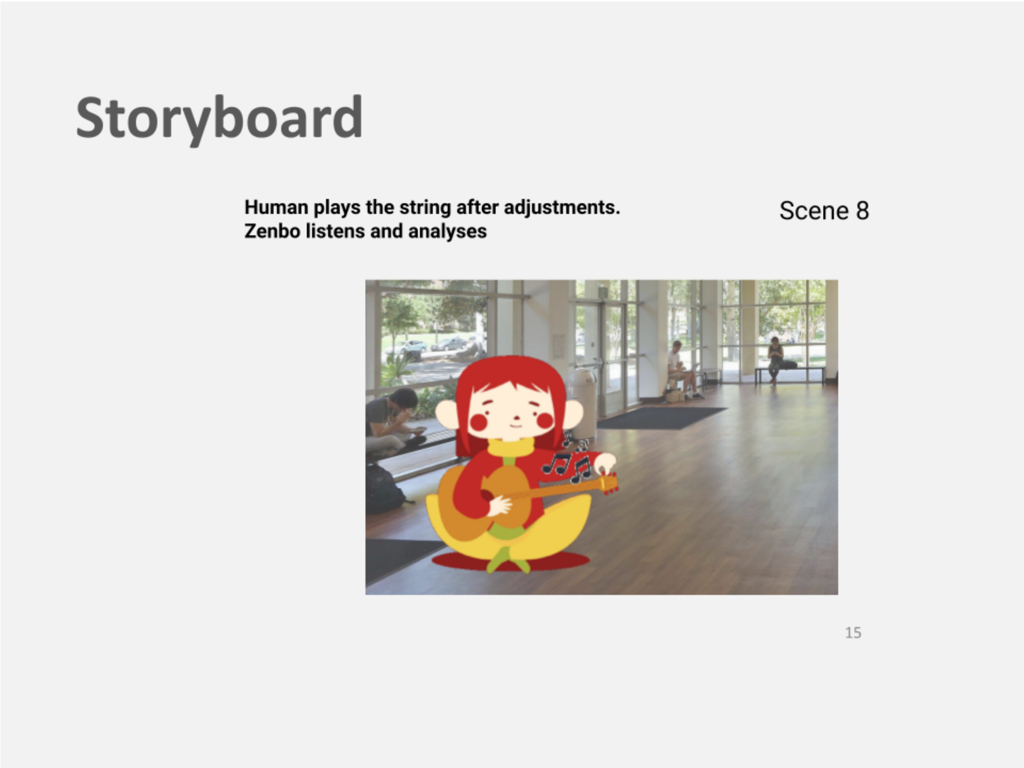
Source: Author.

Figure 16 – Storyboard - Zenbo gives feedback accordingly



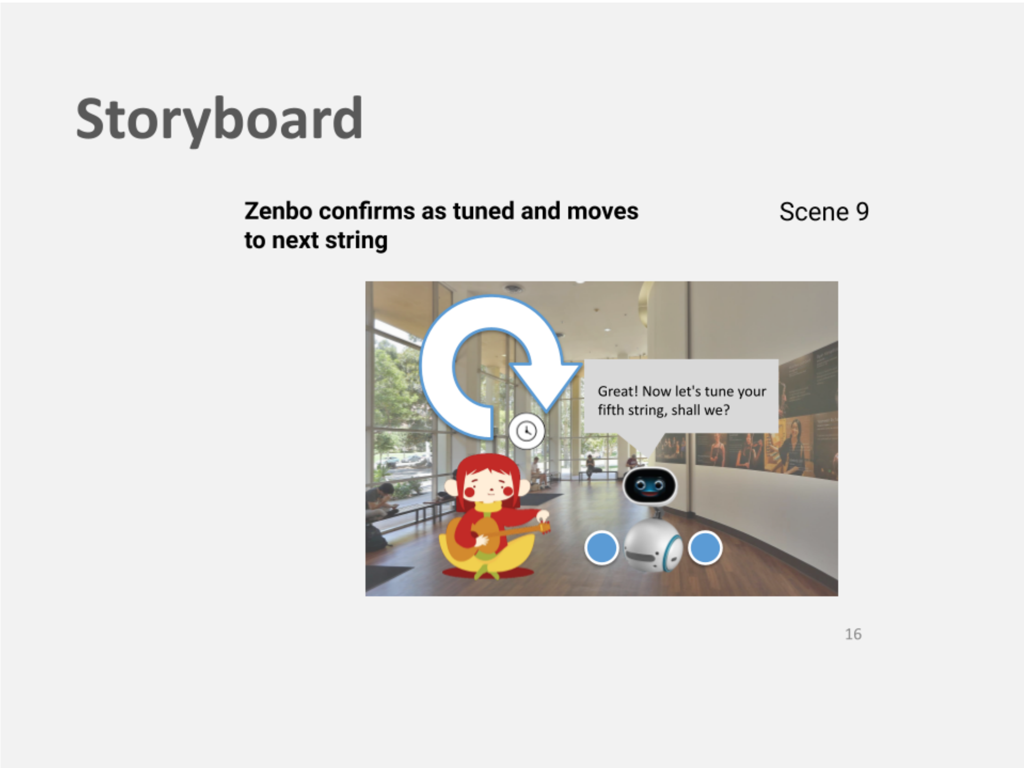
Source: Author.

Figure 17 – Storyboard - User makes the adjustments to the string



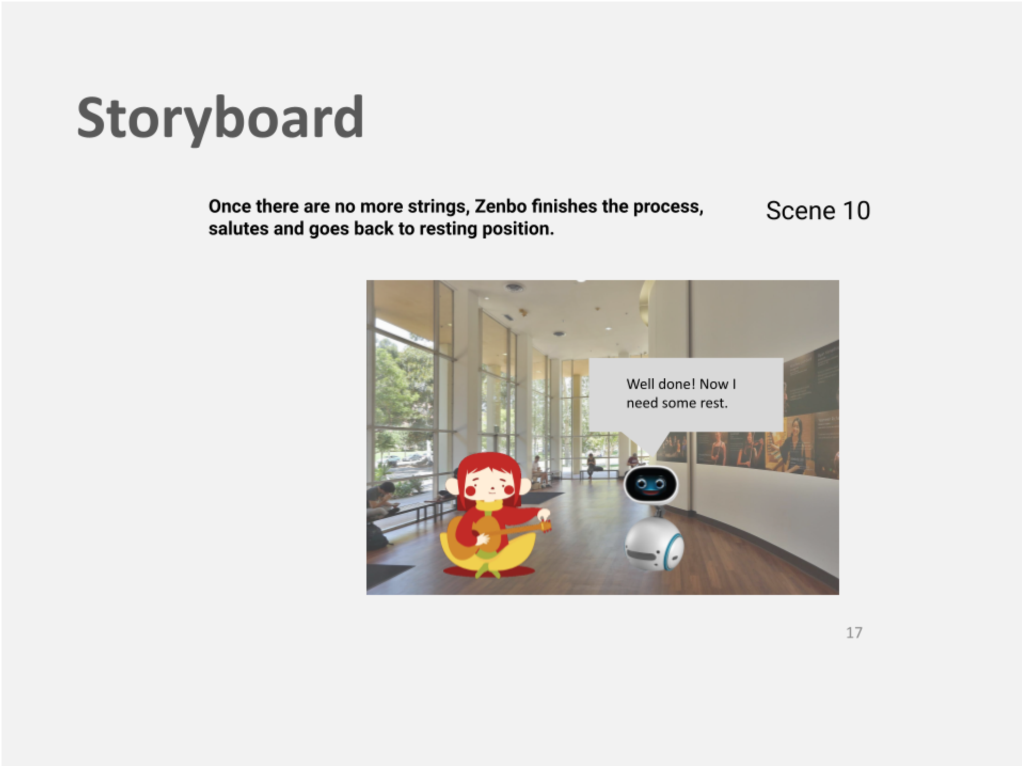
Source: Author.

Figure 18 – Storyboard - The string is tuned and the loop proceeds to next string



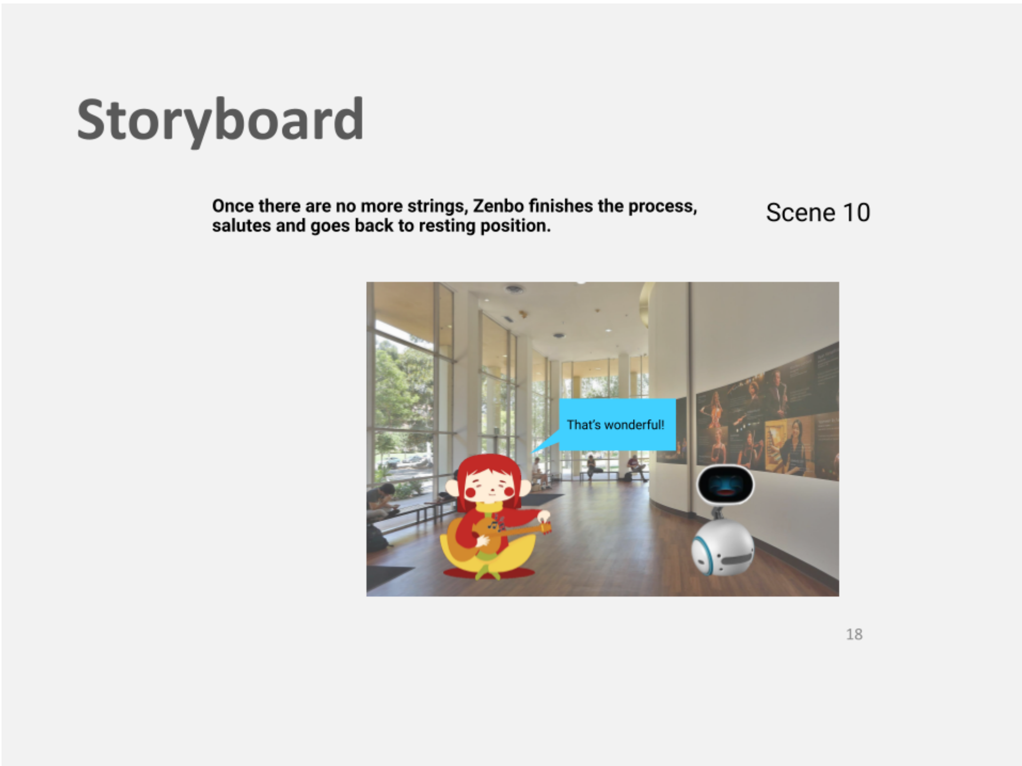
Source: Author.

Figure 19 – Storyboard - Once finished the robot salutes



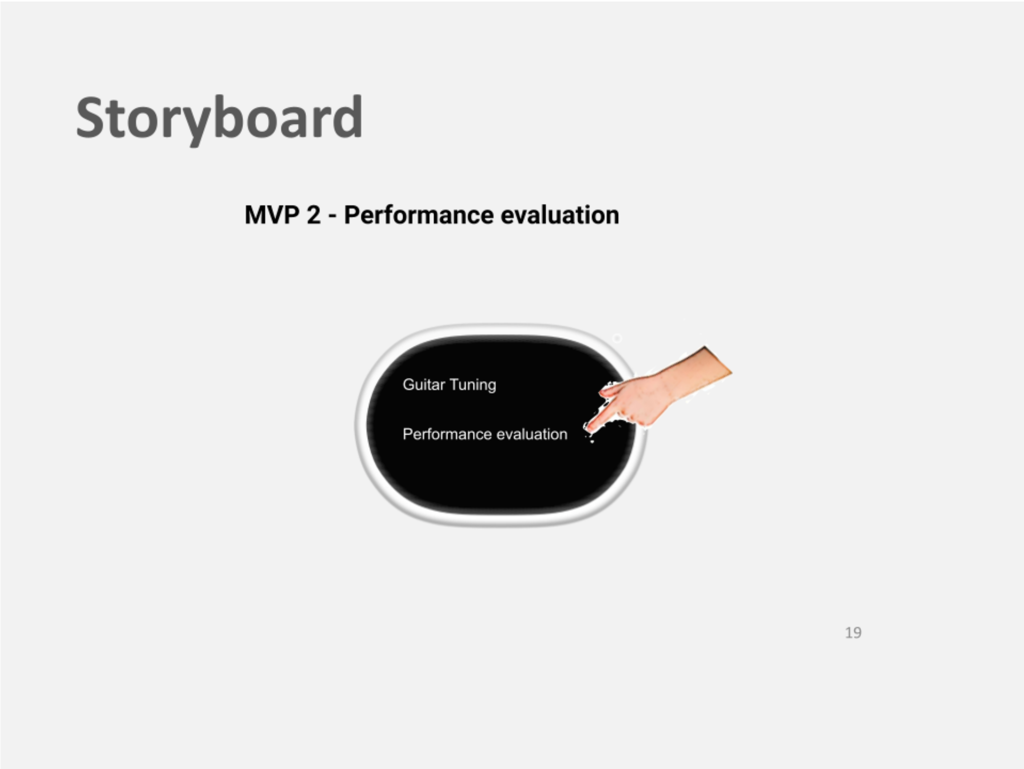
Source: Author.

Figure 20 – Storyboard - The guitar is tuned and the robot returns to it's resting position



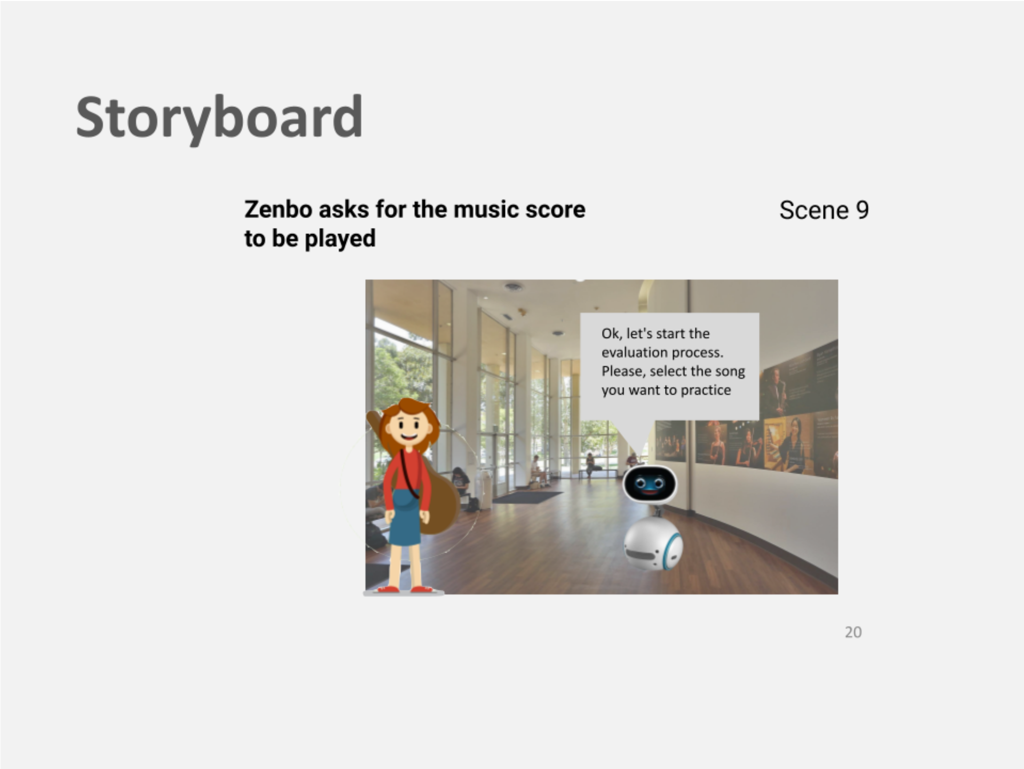
Source: Author.

Figure 21 – Storyboard - User selects performance evaluation



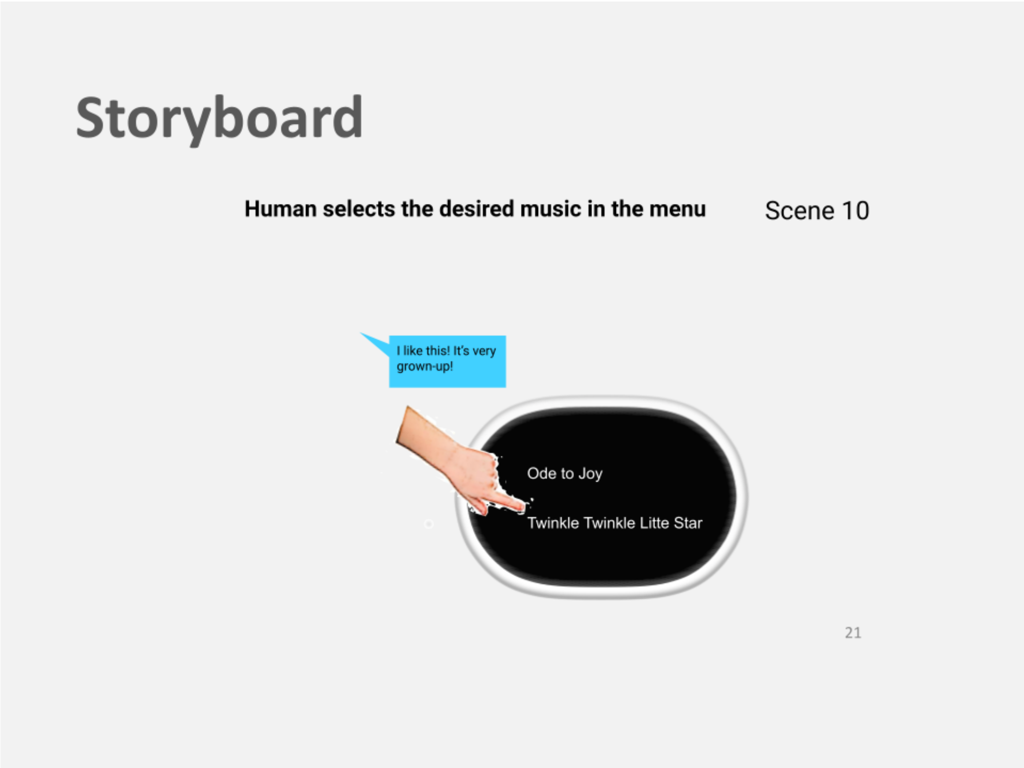
Source: Author.

Figure 22 – Storyboard - Zenbo asks for the music score to be played



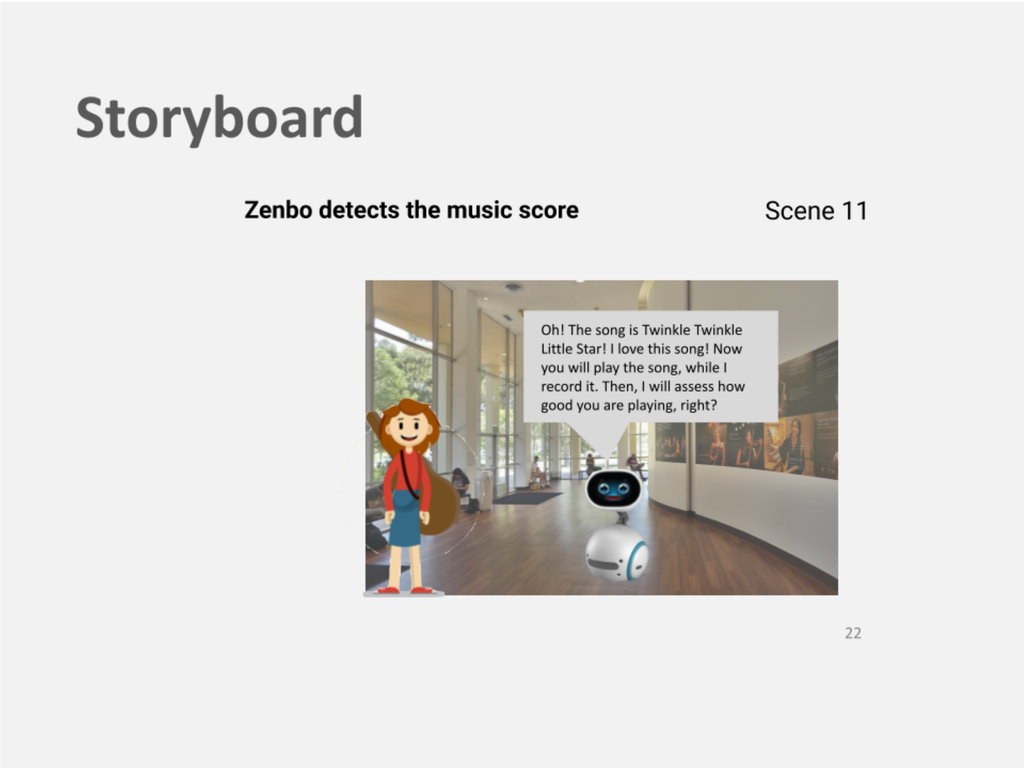
Source: Author.

Figure 23 – Storyboard - Human selects the desired music in the menu



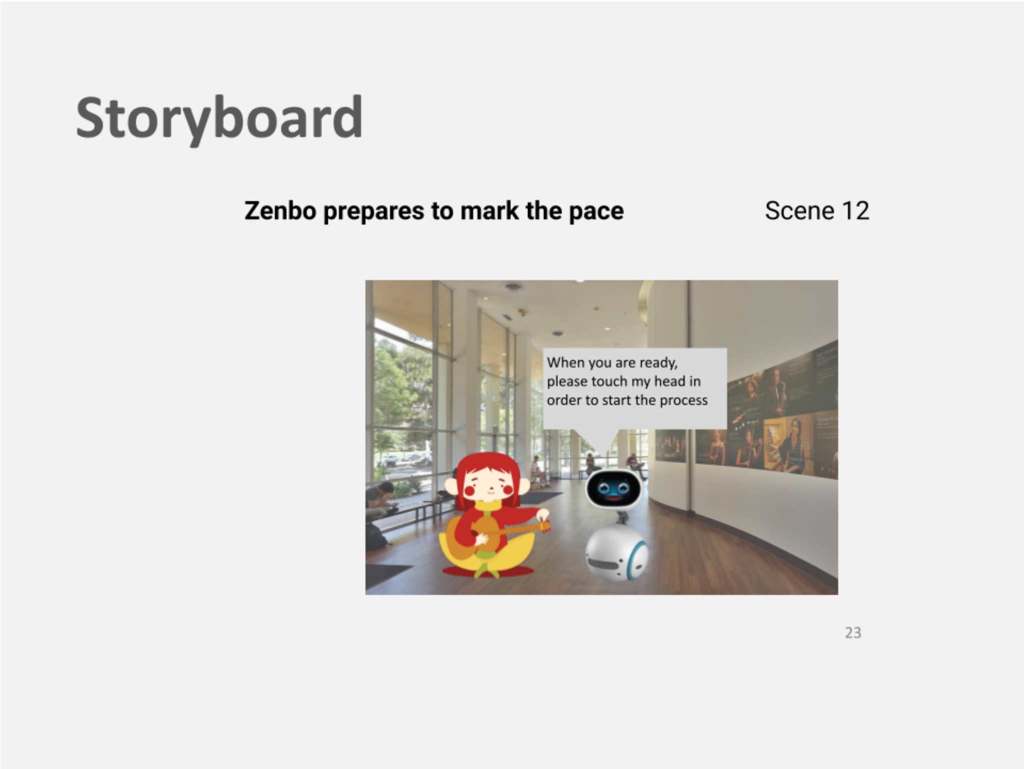
Source: Author.

Figure 24 – Storyboard - Zenbo detects the music score



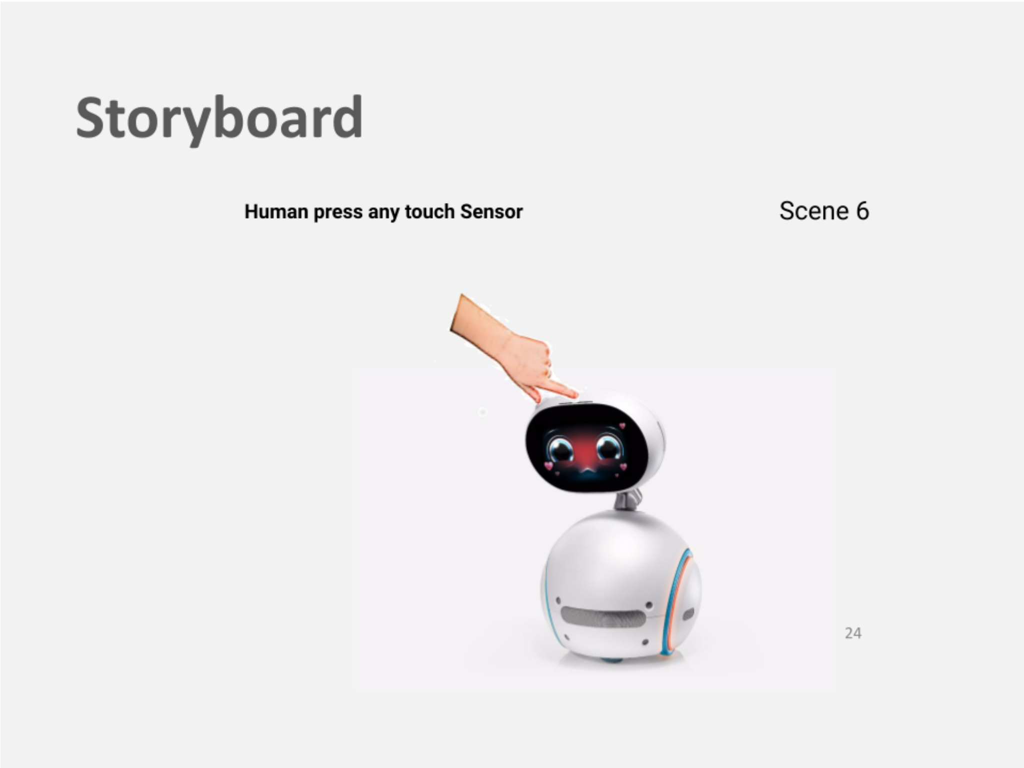
Source: Author.

Figure 25 – Storyboard - Zenbo then explains the evaluaion process



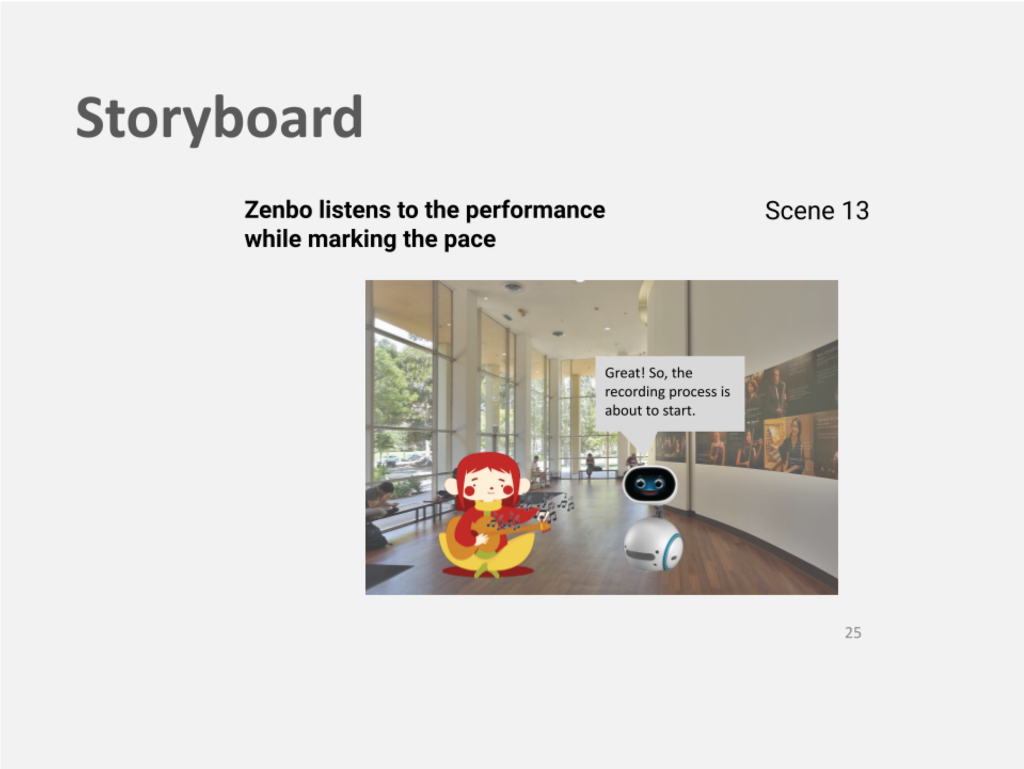
Source: Author.

Figure 26 – Storyboard - Human confirms the beginning of the evaluation process



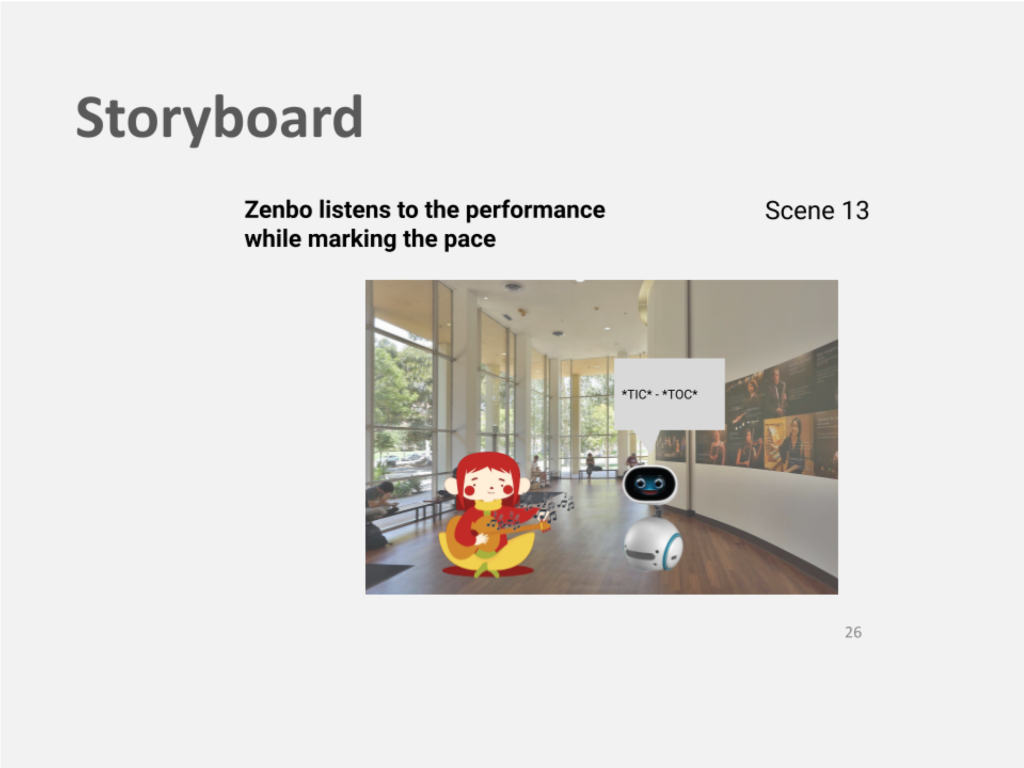
Source: Author.

Figure 27 – Storyboard - Zenbo listens and assesses the user performance



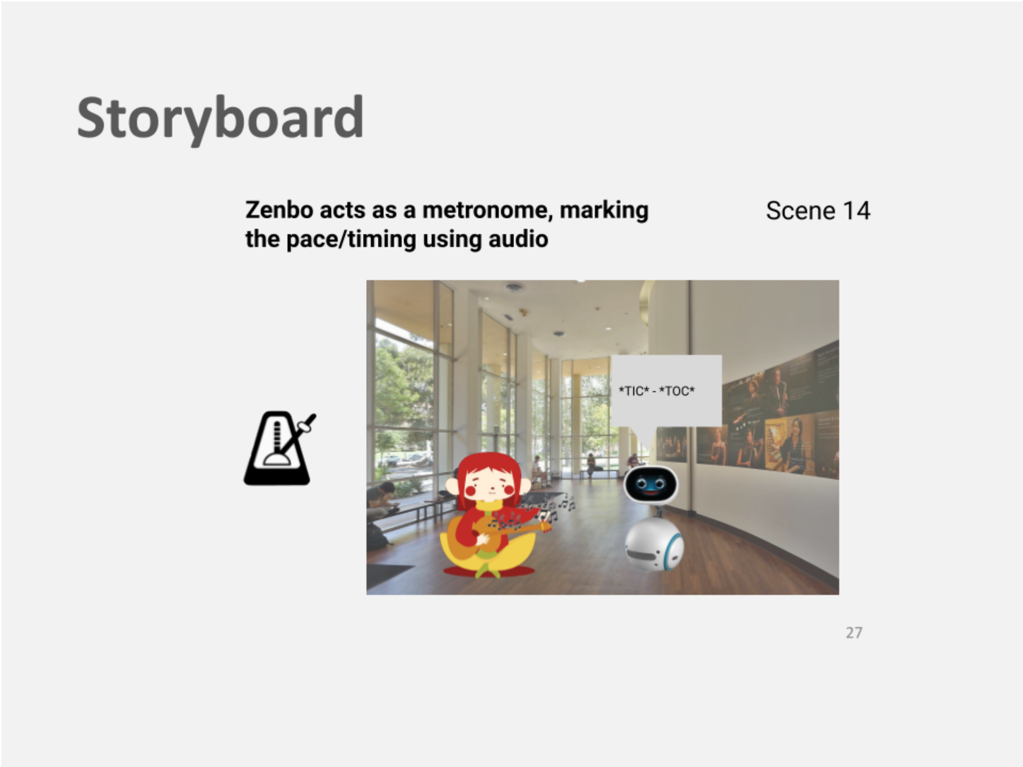
Source: Author.

Figure 28 – Storyboard - Zenbo marks the pace as a metronome



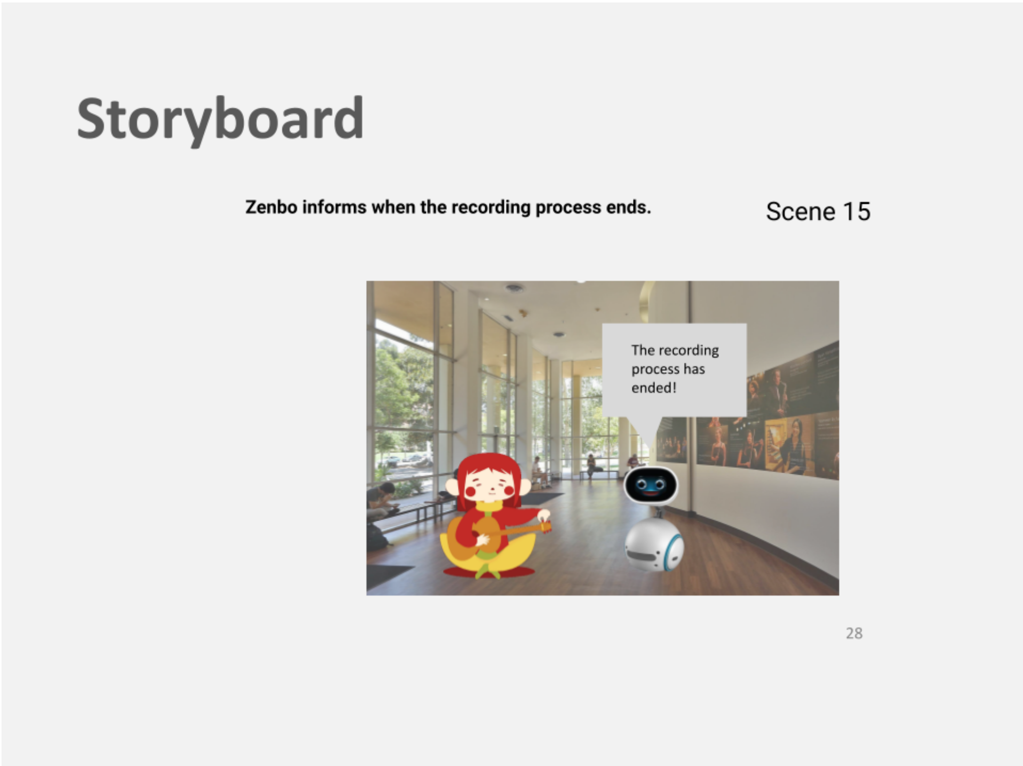
Source: Author.

Figure 29 – Storyboard - Zenbo marks the pace as a metronome



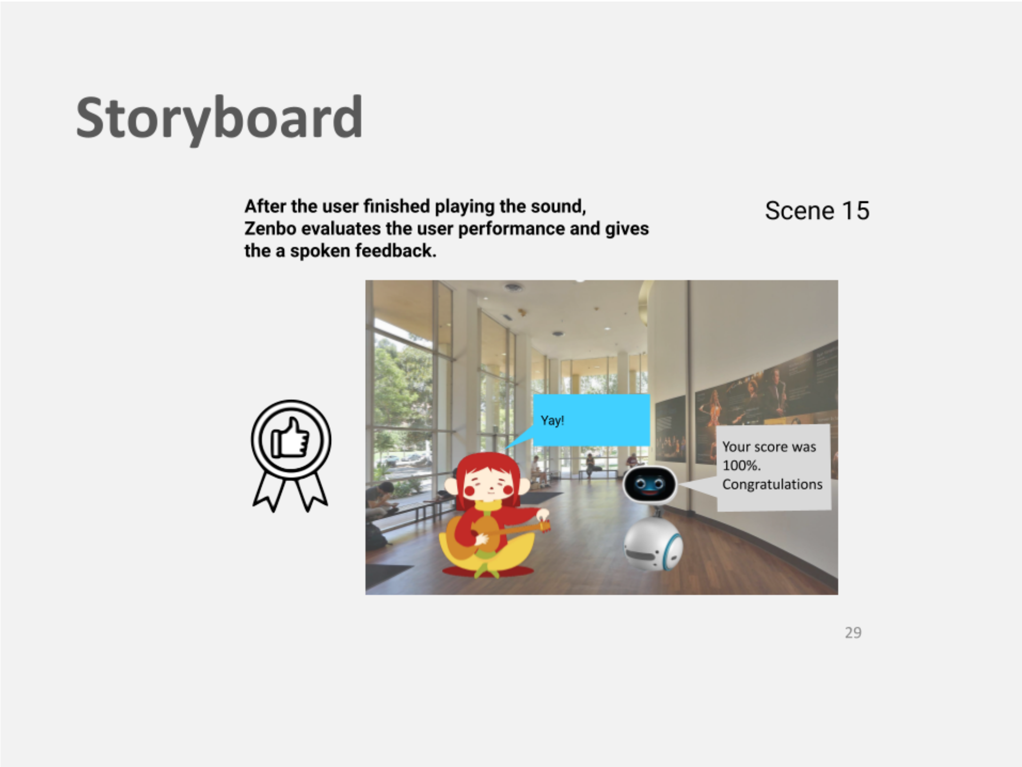
Source: Author.

Figure 30 – Storyboard - Zenbo informs the user when the evaluation ends



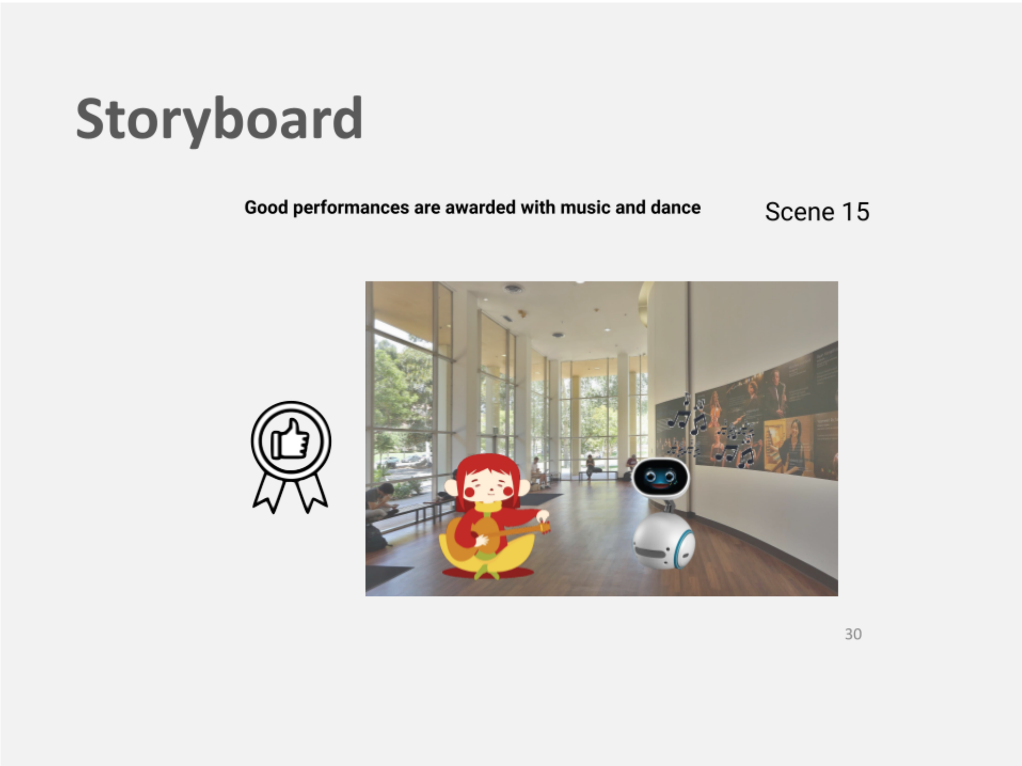
Source: Author.

Figure 31 – Storyboard - Zenbo feeds back the evaluated performance



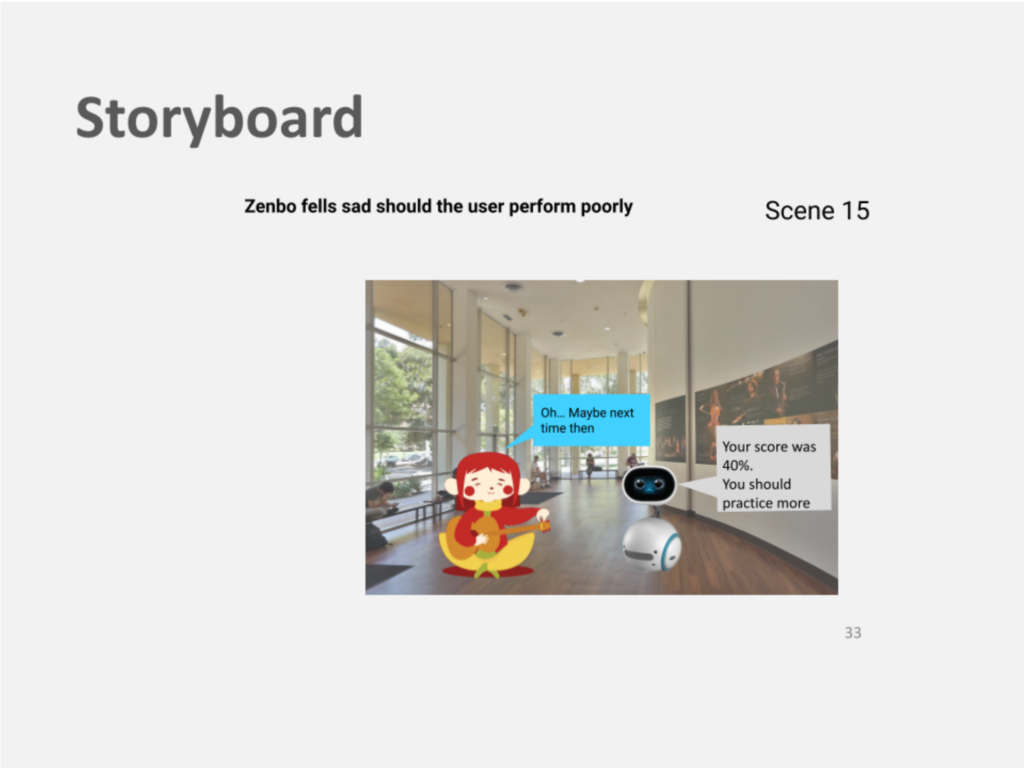
Source: Author.

Figure 32 – Storyboard - Good performances are awarded with music and dance



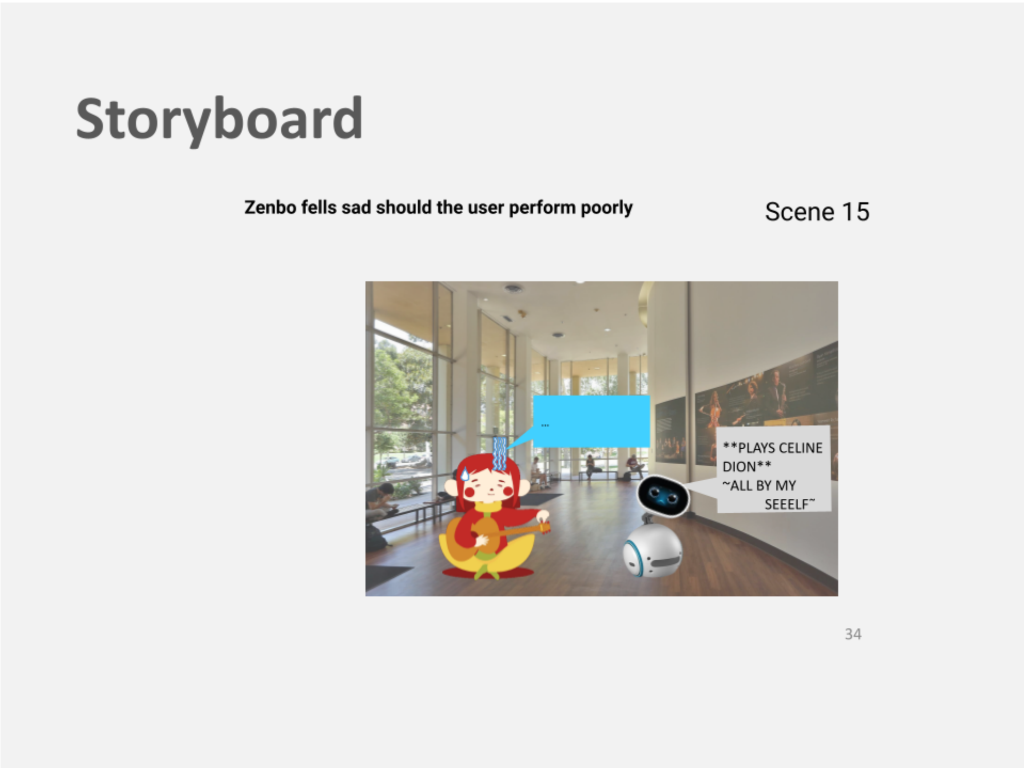
Source: Author.

Figure 33 – Storyboard - Zenbo feeds back the evaluated performance



Source: Author.

Figure 34 – Storyboard - Zenbo feels sad should the user perform poorly



Source: Author.

Table 5 – SUS-Kids adapted for social robot research and the online protocol.

SUS-Kids for Social Robots	**SUS-Kids (PUTNAM et al., 2020)**
If I had these robots, I think that I would like to play with them a lot.	If I had this [app] on my iPad, I think that I would like to play it a lot.
I was confused many times about how to play with the robots.	I was confused many times when I was playing [app].
I think these robots would be easy to use.	I thought [app] was easy to use.
I would need help from an adult to continue to play with the robots.	I would need help from an adult to continue to play.
I always felt like I would know what to do next when I watched those robots.	I always felt like I knew what to do next when I played.
Some of the things I would had to do when playing did not make sense.	Some of the things I had to do when playing [app] did not make sense.
I think most of my friends could learn to play with those robots very quickly.	I think most of my friends could learn to play [app] very quickly.
Some of the things I would had to do while playing sounded kind of weird.	Some of the things I had to do to play [app] were kind of weird.
I would feel confident when I was playing with the robots.	I was confident when I was playing [app].
I would have to learn a lot of things before playing well with the robots.	I had to learn a lot of things before playing [app] well.
I would really enjoy playing with the robots.	I really enjoyed playing [app].
If we had more time, I would keep playing.	If we had more time, I would keep playing [app]
I plan on telling my friends about these robots.	I plan on telling my friends about [app].

Source: Author.

3.4 GENERAL CONSIDERATIONS

In this chapter, we presented the Methods and Materials used in this work, describing the assessment and how it was conducted to meet the research objectives. We produced an HRI application that consisted of a playful training application for music education with two learning modules: guitar tuning process and performance evaluation. The robot application pairs with an acoustic guitar to perform the HRI tasks. We prepared a remote protocol adjusted for the COVID-19 pandemic scenario, leveraging online tools to produce the interviews, showcase the prototypes and collect data from the children. The ToyUI Robot Storyboard supported the planning and demonstration of the prototype with details about the adaptations made to achieve a fair comparison. Along with the SUS Kids, it also supported these events, generating sufficient information for analysis through the lenses of both qualitative and quantitative methods. In the next chapter, we describe the user evaluation, analyzing the obtained results in contrast to the surveyed literature.

4 USER EVALUATION

The recruiting sample gathered data from 22 children and 17 guardians. After excluding incomplete data, the final sample consisted of 20 children and 15 guardians. One guardian canceled their interview, a child opted to leave the study, and another failed to submit her evaluation form. Participants' locations varied from Brazil, the United States, Canada, and Europe. All guardians were the kids' parents, nine female and six male with post-secondary education, four parents with master's degrees, and four doctoral degrees. Most parents aged 36-45 (10 parents), three aged 26-35, and two over 46 years old. Occupations varied from university/college professors, school teachers, medical doctors, physiotherapists, entrepreneurs, lawyers, human resource professionals, Information Technology (IT) professionals, and one stay-at-home parent. The final sample of children consisted of 9 girls and 11 boys aged from 4 to 12 years old, but most children were aged 9-11 (14 children). Some of these children were siblings. All children were English speakers, either native or bilingual (English and Portuguese). Most children had limited knowledge of robotics (14 children), but they would recognize the fictional Star Wars BB8 robot. Five children recognized Zenbo, and 7 recognized NAO by either seeing them in person or resembling the robots' design. Most children had limited musical education levels (11 children had no experience and ten learned chord names), and only four children knew how to read tablatures or music scores.

Online interviews lasted between 30-50 minutes, and interviews with more than one child lasted the longest – we interviewed 1 to 3 kids simultaneously. A reminder that we conducted interviews online using Zoom and Google Meet platforms due to restrictions of the COVID-19 pandemic, which may have generated a perceptual noise and a series of limitations since the children did not interact with the robots live or teleoperated. We could not set up a teleoperation study due to not having access to both robots in the face of university closure and social distancing restrictions. Sessions started with a brief presentation of our research goals and tasks for the interviewees and data collection for the anonymous profile with the children. We alternated introducing NAO and Zenbo applications to prevent bias, consistently introducing the storyboard template before the video. In the end, each child evaluated both applications using a single evaluation form (per child), and then we discussed the open questions about their robot embodiment preferences. The form was adapted from the SUS-Kids survey (PUTNAM et al., 2020) with the regular ten questions concerning the usability and learnability of the prototypes and three added questions regarding engagement, enjoyment, and satisfaction of these users. For interviews with more than one child, we asked them to wait for each to fill out the evaluation form first. We performed the qualitative discussion together, considering parents' input, and three open questions about robot embodiment preferences at the end of the sessions. After all sessions, we analyzed the quantitative data using Google sheets associated with Google forms. We performed some statistical tests to verify some conclusions. Finally, we transcribed qualitative responses to text using the recorded videos from the interviews, permitting us to classify feedback into themes and tags (CHISM; DOUGLAS; JR, 2008).

4.1 QUANTITATIVE RESULTS

First, the quantitative results concern 20 responses to the adapted SUS-Kids survey, obtaining an average score of 75.4 for the music education application. We calculated the SUS-Kids scores following instructions provided in the original scale (BANGOR; KORTUM; MILLER, 2008). We also classified the individual 13 scores of SUS-Kids into four components following instructions provided in the related work: component 1 contains statements 1, 5, 9, 11, 12, and 13; (2) statements 2, 3, 6 and 7; (3) statement 8; and (4) statements 4 and 10 (PUTNAM et al., 2020). We noticed that the lower scores appeared in components 2 and 4, which are related to general usability aspects and requiring assistance or previous knowledge to use the system, respectively. Table 6 summarizes the SUS-Kids scores according to children's age, gender, musical level, and robot preference, and most kids preferred Zenbo (17 votes). Following, we show relevant graphs and make a statistical analysis of the results combining the SUS-Kids survey and children's anonymous profile. We linked survey results and child profile information based on the entry date and time to keep the data anonymous.

Table 6 – SUS-Kids scores related to children's profile information.

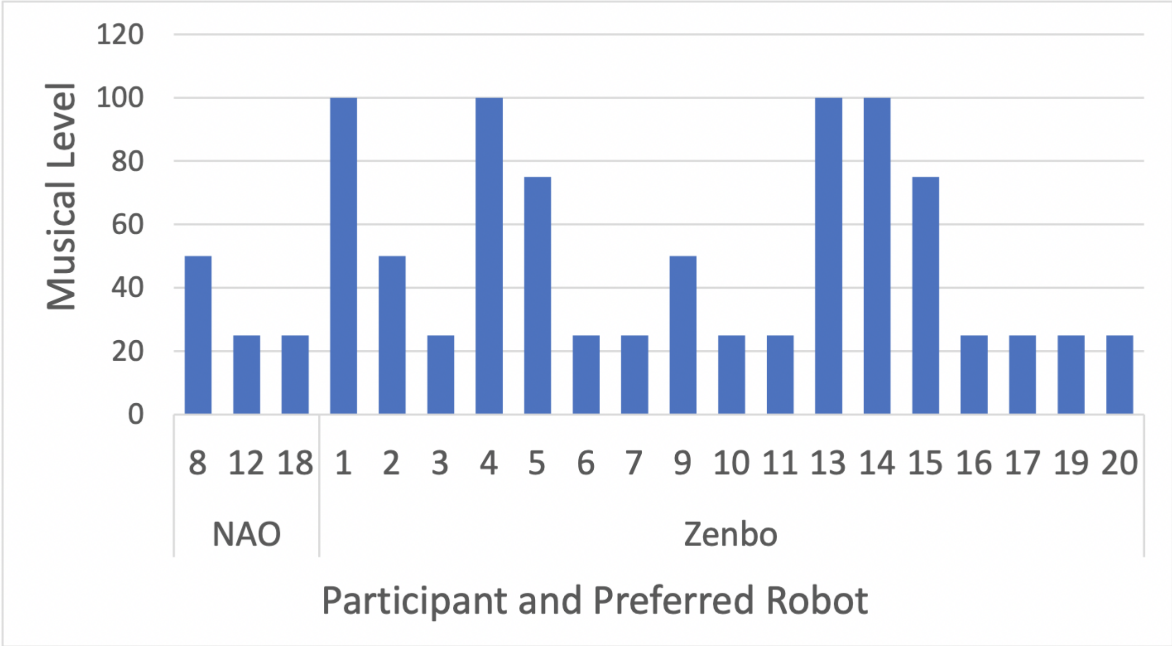
Participant	SUS Score	Musical level	Robot	Age	Gender
1	85	100	Zenbo	11	Boy
2	95	50	Zenbo	9	Boy
3	77.5	25	Zenbo	4	Girl
4	47.5	100	Zenbo	10	Boy
5	50	75	Zenbo	8	Boy
6	90	25	Zenbo	10	Boy
7	50	25	Zenbo	4	Girl
8	90	50	NAO	9	Boy
9	95	50	Zenbo	11	Girl
10	62.5	25	Zenbo	7	Boy
11	52.5	25	Zenbo	11	Boy
12	52.5	25	NAO	4	Boy
13	85	100	Zenbo	10	Girl
14	90	100	Zenbo	12	Girl
15	92.5	75	Zenbo	11	Girl
16	85	25	Zenbo	9	Boy
17	82.5	25	Zenbo	10	Girl
18	60	25	NAO	10	Girl
19	85	25	Zenbo	10	Boy
20	77.5	25	Zenbo	12	Girl

Source: Author.

4.2 MUSICAL LEVEL AND ROBOT PREFERENCE

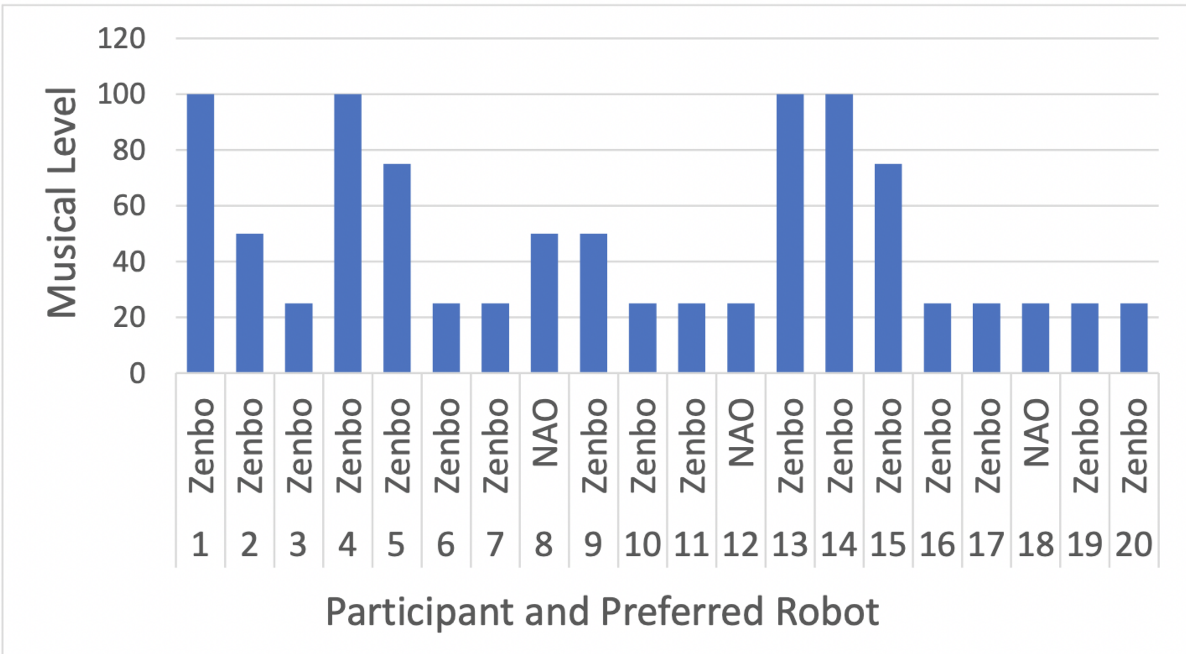
First, we performed some conversions on the raw data to enable the numerical processing of information. We quantified the musical level in four numerical values: none (25), chord names and symbols (50), chord names, symbols, and tablatures (75), and music scores (100). We also turned gender and robot preference into binary entries. Figure 35 illustrates the results relating to musical level and robot choice, and it is visible that there is no relationship between the level of musical knowledge and the preferred robot. Nonetheless, in Figure 36, it is noticeable that participants with higher music levels chose Zenbo, while participants who chose NAO have a lower musical level.

Figure 35 – Music level and preferred robot.



Source: Author.

Figure 36 – Higher music level and preferred robot.



Source: Author.

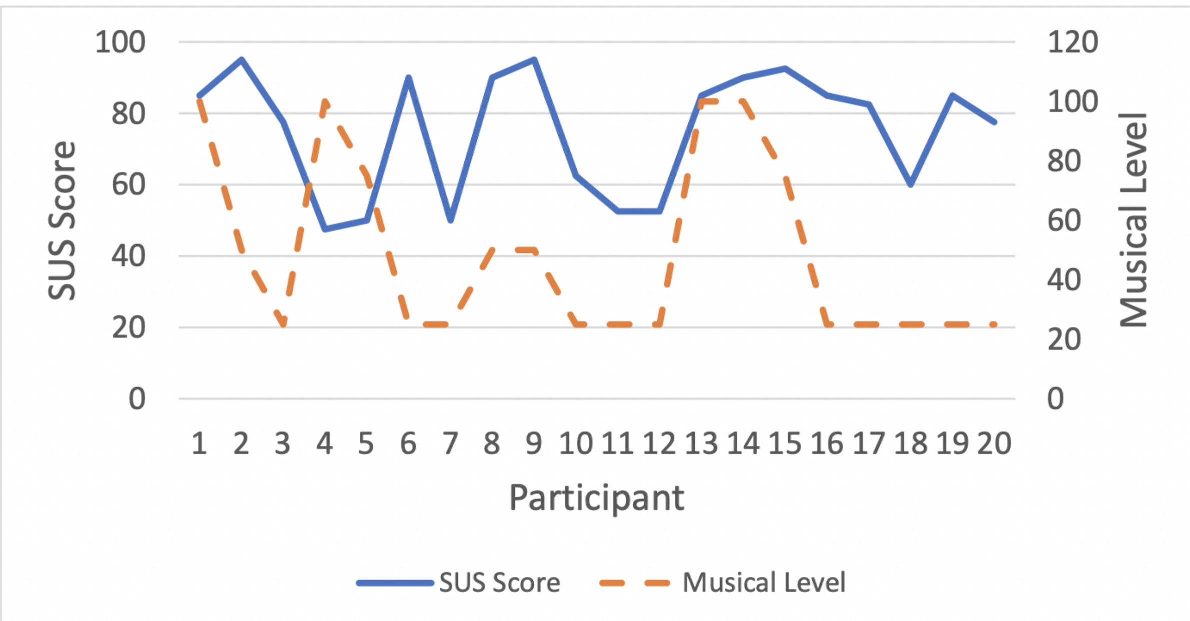
4.3 MUSICAL LEVEL AND SUS-KIDS SCORES

The calculated correlation is very low since the p-value is fairly above the significance level of 0.05, which indicates no rejection of the hypothesis that no correlation exists between the two samples. The graph in Figure 37 shows no relationship between the value of the SUS-Kids score and the musical level.

Figure 38 lists the SUS-Kids score and musical level parameters of the participants. Visually, there is no indication of dependence between these two variables. The correlation coefficient between each sample X and Y is calculated with function corr(X,Y) form Matlab:

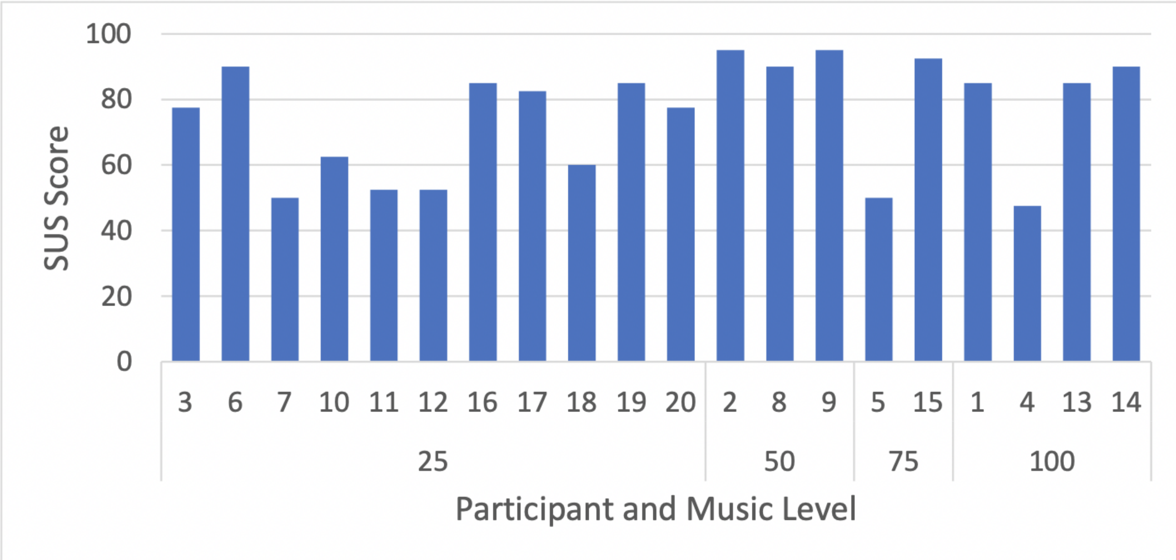
$$[\rho, p_{val}] = corr(SUS\_Score, Musical\_Level), \;\; \rho = 0.1428, \;\; p_{val} = 0.5481. \tag{4.1}$$

Figure 37 – SUS-Kids score and music level correlation



Source: Author.

Figure 38 – SUS-Kids score vs music level



Source: Author.

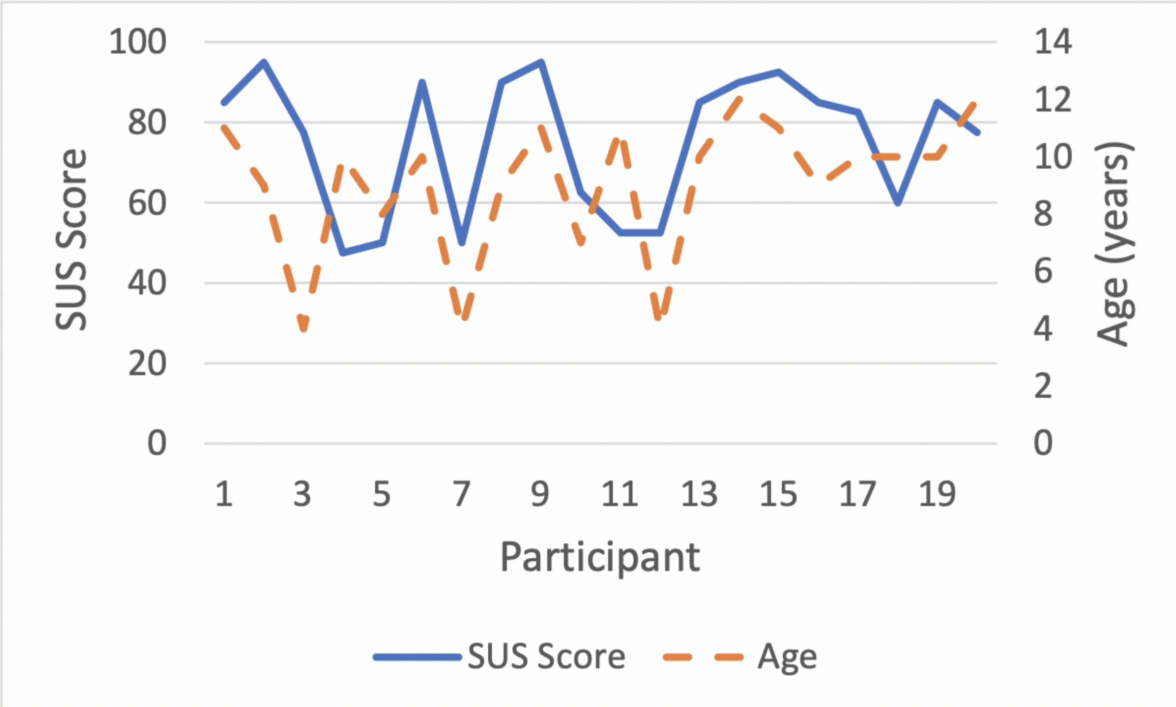
4.4 SUS-KIDS SCORES AND AGE

We use correlation tests in order to evaluate the association between the variables. For instance, we wanted to evaluate the presence of a relationship between the SUS Score and the ages of the participants, validating the initially proposed age range. From the graph in Figure 39, it is possible to detect a relationship between the values of the SUS-Kids score and participants' age, indicating the suitability of the proposed HRI application for older children. We performed the following correlation test:

$$[\rho, p_{val}] = corr(SUS\_Score, Age), \ \ \rho = 0.4583, \ \ p_{val} = 0.0421. \tag{4.2}$$

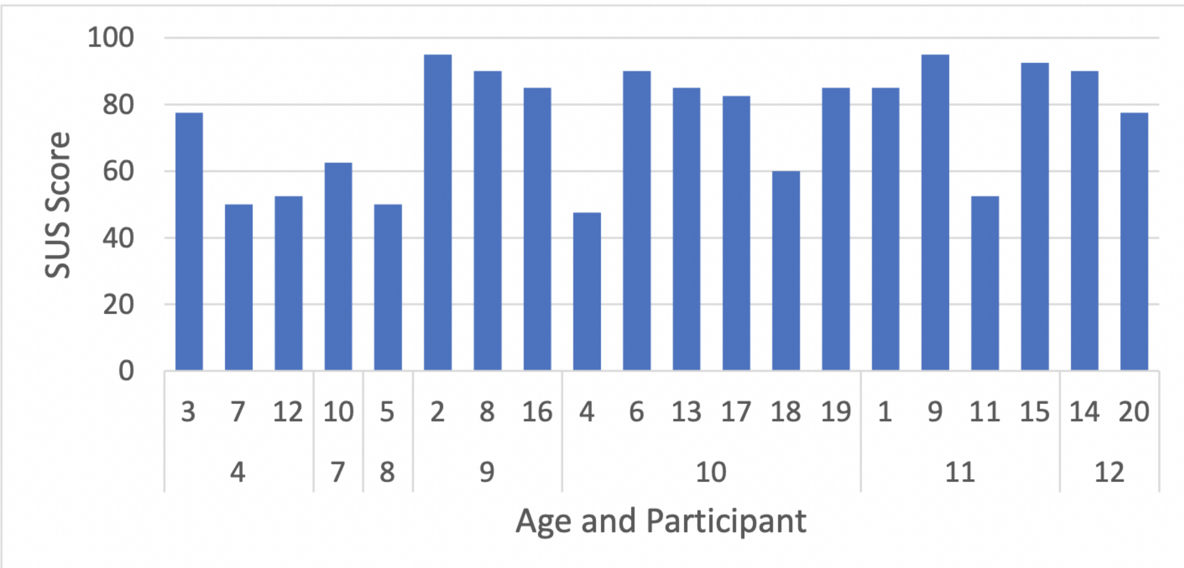
Although the p-value is less than the significance level of 0.05 – which indicates rejection of the hypothesis that no correlation exists between the two samples – the obtained correlation is low. Nevertheless, the graph in Figure 40 shows a concentration of higher values for the SUS-Kids score when the participant is older.

Figure 39 – SUS-Kids score and correlation with age group



Source: Author.

Figure 40 – SUS-Kids score by age group



Source: Author.

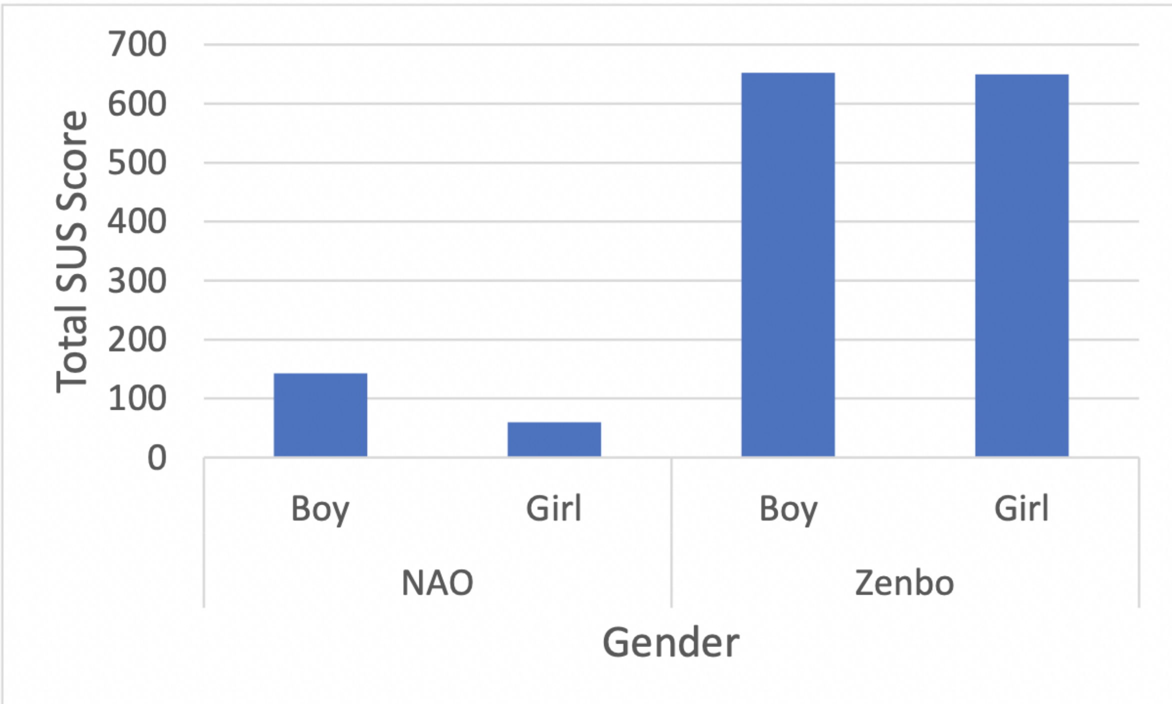
4.5 ROBOT PREFERENCE AND GENDER

Since we had a small sample size, not normally distributed, with a ranked SUS-Kids Score and ordinality present in the musical level and age data, we have opted for non-parametric testing using Kruskal-Wallys. The Kruskal-Wallis test (KRUSKAL; WALLIS, 1952) returns the p-value for the null hypothesis that the data in each column of [Robot, Gender] comes from the same distribution. The alternative hypothesis is that not all samples come from the same distribution. The distributions illustrated in Figures 41-42 indicate no differences between genders and SUS-Kids score values of the participants.

$$p_{val} = kruskalwallis([Robot, Gender]),\ p_{val} = 0.0088. \tag{4.3}$$

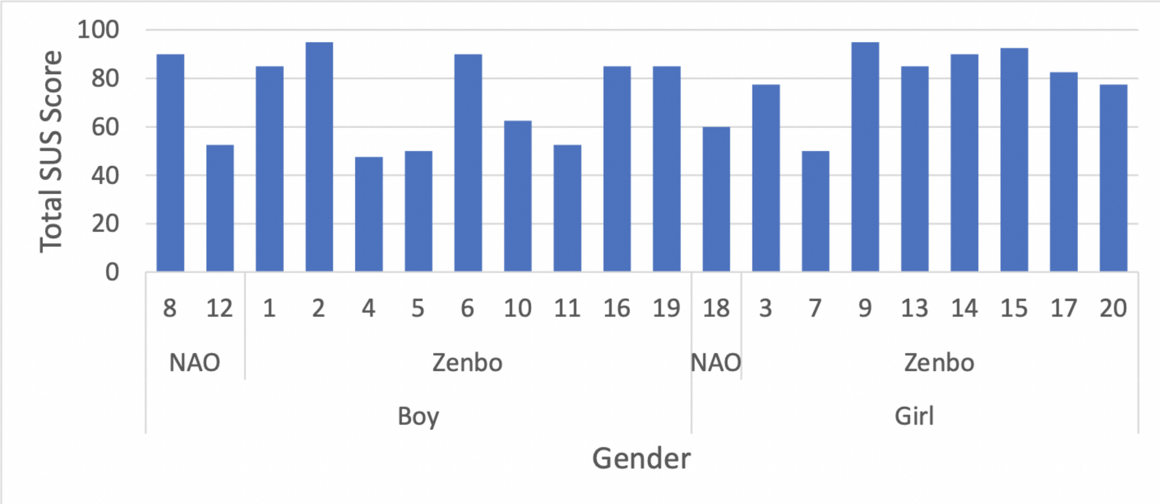
The returned value of p indicates that the Kruskal-Wallis test rejects the null hypothesis that all three data samples come from the same distribution at a 1% significance level.

Figure 41 – Robot preference and gender.



Source: Author.

Figure 42 – SUS-Kids score by gender and robot preference.



Source: Author.

4.6 QUALITATIVE RESULTS

The qualitative results concern the three open questions about robot embodiment preferences at the end of the SUS-Kids survey. Results compile text input in the Google Forms provided by the participants and additional oral transcripts from assessing the interview records. The qualitative analysis supported generating analytical categories underlining important information. We categorized queries into Robot Appearance and Usability, Robot Emotion and Behavior, and, lastly, Content, Additional Features, and Software. In Table 7, we remark children's positive and negative comments for each robot embodiment and the recurrent suggestions children made for the HRI application. In overview, 17 children preferred the Zenbo robot to use the music education application. Mainly, their comments concerned the robot's appearance and emotional expressivity using facial expressions. Participants also considered Zenbo easier to use since the built-in display makes it easier to select options without memorizing selection instructions or showing the NAOmarks to the robot. Another recurrent remark was that accessing the music scores on the robot's display was more convenient than relying on the NAO's companion application.

Table 7 – Analytical categories samples.

Category	NAO		Zenbo	
Robot Appearance and Usability	Positive	Negative	Positive	Negative
	"I like that it has hands, the way it moves."	"The eyes are too small. I couldn't see the lights correctly"	"I think it was really cute and easier because it has a screen, you can choose the music and you don't need to have a cell phone near"	"Zenbo's face is too cute and it made me feel uncomfortable?
Robot Emotion and Behavior	"NAO's reaction is funny when we play wrong"	"Zenbo has way more emotions than NAO"	"I liked Zenbo more because of its emotions"	"I felt sad too when Zenbo was sad"
Content, Additional Features and Software	**NAO & Zenbo**			
	"I would add more songs" "I would like to use them with other instruments"			

Source: Author.

4.7 ROBOT APPEARANCE AND USABILITY

Most comments highlight Zenbo as having a more pleasant, cute, or childish look. Some children expressed affective memory relating Zenbo to movie characters and animations, such as Disney's Wall-E or BB8. We believe that Zenbo's characterization, relative size, rounded shapes and edges, head movement and displacement around space, and the ability to convey more explicit facial expressions may have contributed to it. The screen established a significant element of distinction. Children's remarks highlighted its capacity to display facial expressions and the convenience of selecting options and displaying content in general. Some children claimed it was more practical to see the music scores on Zenbo's screen than on a companion device. Nevertheless, NAO also received positive comments associating robot motion and emotional expressivity. Children highlighted the robot's ability to stand up, dance, and conceal its face with its hands to express sadness. When we asked for suggestions for improvements, some children suggested that their ideal robot would have NAO's body (with limbs and articulations) and Zenbo's face (display). Another interesting topic was about using light feedback in the tuning process. Some children found it hard to notice the lights in the NAO robot (eyes), which was more noticeable in the Zenbo robot (wheels). A child suggested that it would be nice if NAO's eyes were bigger since it would make it easier to see. The video quality of the NAO robot might have compromised its visibility due to excessive brightness in the recording, noticeably making it challenging to discern lights and colors. Some children also criticized using the wheels in the Zenbo robot, although some found it adequate and visible. Children claimed it was too small, hard to remember which action it was representing, and less convenient than displaying the tuning instructions on the screen.

4.8 ROBOT EMOTION AND BEHAVIOR

Many children remarked on Zenbo's emotional expressivity through the facial expressions and built-in display. They considered its emotions more distinguishable and entertaining than the NAO robot. Although Zenbo's facial expressions are virtual animations, their opinion meets expectations from related literature. Full facial expressions are more straightforward to model than limbic and corporal expressivity, less ambiguous, and easier to identify (PAIVA; LEITE; RIBEIRO, 2014). In most cases, children perceived robot emotion as a desirable robot embodiment feature, but two particular cases took our attention. First, a six-year-old boy who claimed to enjoy the robot emotions revealed that he felt sorry for the robots when they expressed sorrow. In another case, a twelve-year-old girl felt uncomfortable with Zenbo's facial expressions. She affirmed that its eyes and expressions, in general, were exaggerated and overly cute, generating discomfort, and she enjoyed the dancing and corporal expressivity of the NAO robot more (she preferred the NAO robot in the survey).

4.9 CONTENT, ADDITIONAL FEATURES, AND SOFTWARE

Most participants suggested the application should have more music options and instruments, including singing. Some participants highlighted that the Zenbo display could guide the player note by note (as seen in rhythmic video games such as Activision's Guitar Hero) or teach them musical notes and scale. A singular observation came from a nine-year-old boy who thought about recording and training their music compositions using the robots. He also suggested that using robots in other learning contexts would be nice, such as replacing a tutor in homeschooling. Parents who participated in or supervised the interviews gave us spontaneous feedback during the evaluation. A remarking comment was about the robot's feedback on performance evaluation when the child performs poorly. Two participants said that negative feedback could cause discouragement, especially with children ages 5 to 7, since they typically do not cope well with this level of criticism. One of the participants was an early childhood educator. She highlighted the importance of keeping the feedback positive or neutral to inspire confidence in the child and motivate them to improve their performance. She also expressed concern that young children could perceive robots as living beings and empathize with their sadness, for example. Finally, some parents agreed on the potential of robot applications as helpful and entertaining resources, stating that robots are more stimulating for children than other resources such as private tutors or mobile applications.

4.10 GENERAL CONSIDERATIONS

In this chapter, we presented the obtained data resulting from the interviews and collection instrument. We conducted the analysis under the quantitative and qualitative lenses, verifying the data's consistency and relationships between age, gender, and skill level, while also understanding the scenario through the individual perspectives of the interviewees. Regarding the application itself, the preliminary data showed that the assumption of an age group from 9 to 11 years old was adequate for the prototype, given a concentration of higher scores among older individuals. The HRI application engaged the sample that expressed interest in other features and possibilities. The child's gender was not a relevant factor in the assessment. In terms of robotic embodiment, Zenbo came out as the favorite, with significant remarks about its appearance, level of emotional expressivity, and lifelikeness features. In the next chapter, we discuss the presented results, relate the evidence to the literature surveyed, and deal with the existing limitations of this work.

5 DISCUSSION AND STUDY LIMITATIONS

User evaluation results exposed both strengths and flaws in the HRI application's design decisions. First, the statistical analysis helped us confirm our target audience's adequacy (9-11 years old). Also, by evaluating the HRI application with younger children, we identified points for improvement that will help us make the application more accessible and suitable for a broader audience. Qualitative evaluation supported us in understanding our target audience's needs, which features are relevant to them, and the most suitable robot for the task. Emotional expressivity demonstrates to be a relevant factor favoring children's preference for the Zenbo robot. Although both robot models received positive comments regarding their appearance, most children preferred Zenbo's cute appearance, facial expressions, and ability to express joy and sadness. The NAO robot relies on voice pitch, body movements, and discreet lights in its eyes to express emotion, making it difficult for users to recognize emotions and for robot designers to model them. From a developer perspective, NAO emotional expressivity does not offer room for improvement, while Zenbo offers alternative skins for facial expressions and the possibility of displaying animation and other characters. The manufacturer (ASUS) also offers a customizing tool based on Unity 3D for making new faces and modeling expressions, making the system more flexible. Regarding the domains of emotional expressivity, we noticed that voice pitch and speed have not generated any significant comments by the interviewees. A single comment arose from an eleven-year-old girl who stated that Zenbo's speech was easier to understand due to the recording's audio, which can relate to the NAO robot's video quality rather than the text-to-speech services.

Regarding the playful training application itself, various feedbacks regarded content or feature additions, such as more songs, compatibility with other instruments, and other learning modules (e.g., teaching musical notes and scales, and even singing). Several comments mentioned the Zenbo robot's display and its ability to show relevant information. Other improvements regarding the display availability included showing the directives for tuning the guitar on screen. Another aspect is that the display facilitated the system's learning curve, reducing the load of information memorized by the child and enabling them to focus on the main HRI tasks and improving overall usability. Perhaps improving the NAO robot companion application would be worthy of achieving comparable results. However, this alternative still depends on a companion device, which was also a target of criticism. The companion device might disrupt the child's attention from the robot. Their comments indicate a desire for all-in-one interaction, especially considering they already have the musical instrument in the interaction environment.

Regarding the evaluation protocol, after conducting the interviews and data analysis, we identified points that need improvement. The first improvement is about the video conference rooms - we used Zoom and Google Meet. Initially, we planned the study to review children's video presentations to map attention and disruption behaviors during the robot's video presentation. Unfortunately, due to the nature of Google Meet, the presentation mode tends to hide other participants and favor the speaker's keynote, making this type of analysis unfeasible unless we pin the interviewee's video, which can become tricky during the presentation. Another limitation was the robot's video quality. We could not access the NAO robot to make a new video. We compared videos using different angles and perspectives, using different sound and lighting conditions and portraying incompatible social situations (live audience and homemade video). We do not know how video quality affected children's overall perception, including the SUS-Kids score and robot embodiment preferences. However, most of the children's comments on likeability aspects relate to the Zenbo's shape and facial expressions. We firmly agreed that their preferences would likely remain the same in different settings.

Another issue we experienced relates to children filling out the research instruments by themselves. At first, we encouraged the child to reply to the survey independently, attended by the guardian, and requested our help when needed. However, it led to losing data since kids would fill-up the form and forget to hit send at the end. Once we noticed the problem, we changed the protocol to prioritize assisted tasks. We would share the questionnaire screen and ask them to give us a verbal response. Unfortunately, we cannot measure whether this change had any impact on respondents' choices during feedback. Finally, a significant limitation concerns the fact that children did not experience the robot application live or teleoperated. Limited access to research facilities motivated us to proceed with this research using multimedia resources. We are satisfied with the quality of feedback we received and how suggestions will impact the future of our project.

6 CONCLUSION

This research compared two social robot models (NAO and Zenbo robots) with 20 children looking to assess their perceived usability, likeability, and robot embodiment preferences, establishing grounds for future comparisons of applications of similar gender. We evaluated the same HRI application using distinct robot embodiment features (e.g., robot shape, size, displays, robot motion, and emotional expressivity) in playful training for music education. The application aimed to support children in tuning an acoustic guitar's strings and providing automated feedback to playing skills through performance evaluation. We implemented an applicable online evaluation protocol in the COVID-19 pandemic using video conference platforms and online instruments.

Empirical results showed children's preferences using the Zenbo robot, consolidating this social robot model as the best fit for future versions of our playful training application, moving in the same direction of our surveyed literature. The Zenbo robot introduced a very enjoyable appearance, a satisfactory level of emotional expressivity, and lifelikeness features. Also, it is a flexible design resource for robot developers and HRI researchers, offering content creation freedom and character modeling, allowing for customization of expressions and face skins. Although the online evaluation introduced several limitations, we obtained valuable data on user's preferences and identified features needing improvements in both usability and entertainment aspects. For instance, regarding the age and knowledge requirements for the proposed application, additional functionalities can support expanding it to a broader audience (e.g., teaching younger children how to read music scores).

As final recommendations, our research suggests that HRI applications towards learning tasks should consider displaying and selecting content using a touchscreen display; preferred a built-in display demonstrated to be a better choice for robot embodiment features in this context. The embedded display removed the need to connect a companion device giving more freedom to introduce tangible and playful interfaces, potentially reducing learning requirements, providing content flexibility, precise inputs, and a more accessible environment for communicating robotic emotion. Another recommendation regards robot motion features since they presented of greater significance in children's perspectives. Regardless of motion level, robot motion helped the social robots to improve lifelikeness, reinforcing emotional portrayal abilities. In short, keep the robot alive.

Beyond the discussed results, the present work also delivered two academic papers. One indirect contribution regarding technical aspects of our NAO 5 implementation of the application - awarded as the best paper of the 11th Workshop of Robotics in Education (WRE 2020) (MELO et al., 2020) - and an article at the Journal of Intelligent & Robotic Systems (JERONIMO et al., 2022) as a direct contribution that summarizes our findings for the academic community.

As for future works, we must retake NAO's video footage to prevent perceptual noise in future data collections and re-run the tests both online and in person, as well as with the actual implemented version of Zenbo app with the new additions remarked by the audience in the iteration of the prototype, as it may generate relevant results. We are also interested in pursuing more data around other robot embodiment features such as voice pitch and speed, gender identity roles, anthropomorphism, and the role of color in robot emotion. Other study opportunities include comparing all-in-one solutions against multi-connected devices in different HRI learning scenarios, as well as the actual evaluation and comparison of the task results (e.g. learning performance between users of these robotic devices in the long term).

REFERENCES

ALBUQUERQUE, A. P. de; KELNER, J. Toy user interfaces: systematic and industrial mapping. *Journal of Systems Architecture*, Elsevier, v. 97, p. 77–106, 2019.

ASUS. *ASUS debuts New Robot -Zenbo Junior*. 2021. <https://zenbo.asus.com/whatsnew/events/zenbojunior/>.

BANGOR, A.; KORTUM, P. T.; MILLER, J. T. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction*, Taylor & Francis, v. 24, n. 6, p. 574–594, 2008.

BARTNECK, C.; BELPAEME, T.; EYSSEL, F.; KANDA, T.; KEIJSERS, M.; ŠABANOVIĆ, S. *Human-robot interaction: An introduction*. [S.l.]: Cambridge University Press, 2020.

BARTNECK, C.; FORLIZZI, J. A design-centred framework for social human-robot interaction. In: IEEE. *RO-MAN 2004. 13th IEEE international workshop on robot and human interactive communication (IEEE Catalog No. 04TH8759)*. [S.l.], 2004. p. 591–594.

BELPAEME TONY, e. a. Social robots for education: A review. science. *Science Robotics*, v. 3, n. 21, 2018.

BREAZEAL, C.; DAUTENHAHN, K.; KANDA, T. Social robotics. *Springer handbook of robotics*, Springer, p. 1935–1972, 2016.

BREEMEN, A. van; YAN, X.; MEERBEEK, B. Icat: An animated user-interface robot with personality. In: *Proceedings of the Fourth International Joint Conference on Autonomous Agents and Multiagent Systems*. New York, NY, USA: Association for Computing Machinery, 2005. (AAMAS '05), p. 143–144. ISBN 1595930930. Disponível em: <https://doi.org/10.1145/1082473.1082823>.

BRYSON, C.; HAND, L. The role of engagement in inspiring teaching and learning. *Innovations in Education and Teaching International*, Routledge, v. 44, n. 4, p. 349–362, 2007. Disponível em: <https://doi.org/10.1080/14703290701602748>.

CHISM, N. V. N.; DOUGLAS, E.; JR, W. J. H. Qualitative research basics: A guide for engineering educators. *Rigorous Research in Engineering Education NSF DUE*, v. 341127, p. 1–65, 2008.

DUFFY, B. R. Anthropomorphism and the social robot. *Robotics and autonomous systems*, Elsevier, v. 42, n. 3-4, p. 177–190, 2003.

FINK, J. Anthropomorphism and human likeness in the design of robots and human-robot interaction. In: GE, S. S.; KHATIB, O.; CABIBIHAN, J.-J.; SIMMONS, R.; WILLIAMS, M.-A. (Ed.). *Social Robotics*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012. p. 199–208. ISBN 978-3-642-34103-8.

FONG, T.; NOURBAKHSH, I.; DAUTENHAHN, K. A survey of socially interactive robots. *Robotics and Autonomous Systems*, v. 42, n. 3, p. 143–166, 2003. ISSN 0921-8890. Socially Interactive Robots. Disponível em: <https://www.sciencedirect.com/science/article/pii/S092188900200372X>.

GOERTZEL, G. An algorithm for the evaluation of finite trigonometric series. *The American Mathematical Monthly*, Mathematical Association of America, v. 65, n. 1, p. 34–35, 1958. ISSN 00029890, 19300972. Disponível em: <http://www.jstor.org/stable/2310304>.

GOODRICH, M. A.; SCHULTZ, A. C. *Human-robot interaction: a survey*. [S.l.]: Now Publishers Inc, 2008.

GORBUNOVA, I.; HINER, H. Music computer technologies and interactive systems of education in digital age school. In: *Proceedings of the International Conference Communicative Strategies of Information Society (CSIS 2018)*. [S.l.: s.n.], 2019. p. 124–128.

HANCOCK, P. A.; BILLINGS, D. R.; SCHAEFER, K. E. Can you trust your robot? *Ergonomics in Design*, SAGE Publications Sage CA: Los Angeles, CA, v. 19, n. 3, p. 24–29, 2011.

HARZING, A. *Publish or Perish*. 2007. Disponível em: <https://harzing.com/resources/publish-or-perish>.

IDEO (Ed.). *The field guide to human-centered design: design kit*. 1st. ed. ed. San Francisco, Calif: IDEO, 2015. ISBN 978-0-9914063-1-9. Disponível em: <http://www.designkit.org/resources/1>.

JERONIMO, B. de S.; WHELER, A. P. de A.; OLIVEIRA, J. P. G. de; MELO, R.; BASTOS-FILHO, C. J.; KELNER, J. Comparing social robot embodiment for child musical education. *Journal of Intelligent & Robotic Systems*, Springer, v. 105, n. 2, p. 1–16, 2022.

JOHAL, W. Research trends in social robots for learning. *Current Robotics Reports*, Springer, p. 1–9, 2020.

KELLEY, J. F. Wizard of oz (woz) a yellow brick journey. *Journal of Usability Studies*, Usability Professionals' Association Bloomingdale, IL, v. 13, n. 3, p. 119–124, 2018.

KENNEDY, J.; BAXTER, P.; BELPAEME, T. Comparing robot embodiments in a guided discovery learning interaction with children. *International Journal of Social Robotics*, Springer, v. 7, n. 2, p. 293–308, 2015.

KOENIG, N.; HOWARD, A. Design and use paradigms for gazebo, an open-source multi-robot simulator. In: IEEE. *2004 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)(IEEE Cat. No. 04CH37566)*. [S.l.]. v. 3, p. 2149–2154.

LI, H.; JOHN-JOHN, C.; TAN, Y. K. Towards an effective design of social robots. *International Journal of Social Robotics*, Springer Nature BV, v. 3, n. 4, p. 333–335, 2011.

LÖCHTEFELD, M.; GEHRING, S.; JUNG, R.; KRÜGER, A. guitar: supporting guitar learning through mobile projection. In: *CHI'11 Extended Abstracts on Human Factors in Computing Systems*. [S.l.: s.n.], 2011. p. 1447–1452.

MALIK, N. A.; YUSSOF, H.; HANAPIAH, F. A. Interactive behavior design in humanoid robot towards joint attention of children with cerebral palsy with human therapists. In: IEEE. *2015 IEEE International Conference on Rehabilitation Robotics (ICORR)*. [S.l.], 2015. p. 828–833.

MELO, R.; MONTEIRO, R. de P.; OLIVEIRA, J. P. G. de; JERONIMO, B.; BASTOS-FILHO, C. J.; ALBUQUERQUE, A. P. de; KELNER, J. Guitar tuner and song performance evaluation using a nao robot. In: IEEE. *2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE)*. [S.l.], 2020. p. 1–6.

MORI, M. *Bukimi no tani [the un-canny valley]*. [S.l.]: Elsevier, 1970. v. 97. 33–35 p.

NIELSEN, J.; BÆRENDSEN, N. K.; JESSEN, C. Robomusickids–music education with robotic building blocks. In: IEEE. *2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning*. [S.l.], 2008. p. 149–156.

PAIVA, A.; LEITE, I.; RIBEIRO, T. Emotion modeling for social robots. *The Oxford handbook of affective computing*, p. 296–308, 2014.

PAPAKOSTAS, G. A.; STROLIS, A. K.; PANAGIOTOPOULOS, F.; AITSIDIS, C. N. Social robot selection: a case study in education. In: IEEE. *2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM)*. [S.l.], 2018. p. 1–4.

PAULE-RUIZ, M.; ÁLVAREZ-GARCÍA, V.; PÉREZ-PÉREZ, J. R.; ÁLVAREZ-SIERRA, M.; TRESPALACIOS-MENÉNDEZ, F. Music learning in preschool with mobile devices. *Behaviour & Information Technology*, Taylor & Francis, v. 36, n. 1, p. 95–111, 2017.

PEAK, L. The suzuki method of music instruction. In: ______. *Teaching and Learning in Japan*. [S.l.]: Cambridge University Press, 1996. p. 345–368.

PUTNAM, C.; PUTHENMADOM, M.; CUERDO, M. A.; WANG, W.; PAUL, N. Adaptation of the system usability scale for user testing with children. In: *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*. [S.l.: s.n.], 2020. p. 1–7.

READ, J. C. Evaluating artefacts with children: age and technology effects in the reporting of expected and experienced fun. In: *Proceedings of the 14th ACM international conference on Multimodal interaction*. [S.l.: s.n.], 2012. p. 241–248.

RIEK, L. D.; RABINOWITCH, T.-C.; CHAKRABARTI, B.; ROBINSON, P. How anthropomorphism affects empathy toward robots. In: *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*. New York, NY, USA: Association for Computing Machinery, 2009. (HRI '09), p. 245–246. ISBN 9781605584041. Disponível em: <https://doi.org/10.1145/1514095.1514158>.

SASTRE, J.; CERDÀ, J.; GARCÍA, W.; HERNÁNDEZ, C.; LLORET, N.; MURILLO, A.; PICÓ, D.; SERRANO, J.; SCARANI, S.; DANNENBERG, R. B. New technologies for music education. In: IEEE. *2013 Second International Conference on E-Learning and E-Technologies in Education (ICEEE)*. [S.l.], 2013. p. 149–154.

SCHOLTZ, J. Theory and evaluation of human robot interactions. In: IEEE. *36th Annual Hawaii International Conference on System Sciences, 2003. Proceedings of the*. [S.l.], 2003. p. 10–pp.

SERAFIN, S.; ADJORLU, A.; NILSSON, N.; THOMSEN, L.; NORDAHL, R. Considerations on the use of virtual and augmented reality technologies in music education. In: IEEE. *2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning through Virtual & Augmented Reality (KELVAR)*. [S.l.], 2017. p. 1–4.

SHAHAB, M.; TAHERI, A.; MOKHTARI, M.; SHARIATI, A.; HEIDARI, R.; MEGHDARI, A.; ALEMI, M. Utilizing social virtual reality robot (v2r) for music education to children with high-functioning autism. *Education and Information Technologies*, Springer, p. 1–25, 2021.

SOFTBANK. *NAO*. 2021. <https://www.softbankrobotics.com/emea/en/nao>.

SULLIVAN, A.; BERS, M. U. Dancing robots: integrating art, music, and robotics in singapore's early childhood centers. *International Journal of Technology and Design Education*, Springer, v. 28, n. 2, p. 325–346, 2018.

TAHERI, A.; SHARIATI, A.; HEIDARI, R.; SHAHAB, M.; ALEMI, M.; MEGHDARI, A. Impacts of using a social robot to teach music to children with low-functioning autism. *Paladyn, Journal of Behavioral Robotics*, De Gruyter, v. 12, n. 1, p. 256–275, 2021.

THELLMAN, S.; SILVERVARG, A.; GULZ, A.; ZIEMKE, T. Physical vs. virtual agent embodiment and effects on social interaction. In: SPRINGER. *International conference on intelligent virtual agents*. [S.l.], 2016. p. 412–415.

VINER, R. M.; RUSSELL, S. J.; CROKER, H.; PACKER, J.; WARD, J.; STANSFIELD, C.; MYTTON, O.; BONELL, C.; BOOY, R. School closure and management practices during coronavirus outbreaks including covid-19: a rapid systematic review. *The Lancet Child & Adolescent Health*, Elsevier, 2020.

WADDELL, G.; WILLIAMON, A. Technology use and attitudes in music learning. *Frontiers in ICT*, Frontiers, v. 6, p. 11, 2019.

WAINER, J.; FEIL-SEIFER, D. J.; SHELL, D. A.; MATARIC, M. J. The role of physical embodiment in human-robot interaction. In: IEEE. *ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication*. [S.l.], 2006. p. 117–122.

WEBSTER, P. R. Computer-based technology and music teaching and learning: 2000–2005. In: *International handbook of research in arts education*. [S.l.]: Springer, 2007. p. 1311–1330.

WESTLUND, J. K.; DICKENS, L.; JEONG, S.; HARRIS, P.; DESTENO, D.; BREAZEAL, C. A comparison of children learning new words from robots, tablets, & people. In: *Proceedings of the 1st international conference on social robots in therapy and education*. [S.l.: s.n.], 2015.

WHELER, A. P. de A.; KELNER, J.; HUNG, P. C.; JERONIMO, B. de S.; JUNIOR, R. d. S. R.; ARAÚJO, A. F. R. Toy user interface design—tools for child–computer interaction. *International Journal of Child-Computer Interaction*, Elsevier, v. 30, p. 100307, 2021.

YAMABE, T.; NAKAJIMA, T. Playful training with augmented reality games: case studies towards reality-oriented system design. *Multimedia Tools and Applications*, Springer, v. 62, n. 1, p. 259–286, 2013.

YANCO, H.; DRURY, J. Classifying human-robot interaction: an updated taxonomy. In: *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No.04CH37583)*. [S.l.: s.n.], 2004. v. 3, p. 2841–2846 vol.3.

ZAMAN, B.; ABEELE, V. V. Laddering with young children in user experience evaluations: theoretical groundings and a practical case. In: *Proceedings of the 9th International Conference on Interaction Design and Children*. [S.l.: s.n.], 2010. p. 156–165.