Journal of Intelligent & Robotic Systems (2022) 105: 28
https://doi.org/10.1007/s10846-022-01604-5

REGULAR PAPER

Comparing Social Robot Embodiment for Child Musical Education

Bruno de Souza Jeronimo[1] · Anna Priscilla de Albuquerque Wheler[1] · José Paulo G. de Oliveira[2] · Rodrigo Melo[3] · Carmelo J. A. Bastos-Filho[2] · Judith Kelner[1]

Received: 30 July 2021 / Accepted: 24 February 2022 / Published online: 17 May 2022

Abstract

The present research focuses in the comparison of two social robot models running the same Human-Robot Interaction (HRI) applications targeting the context of music education for children aged 9-11, with the objective of underlying the design choices favored by the target audience on the running tasks. The Guitar Tuner consists of two main functionalities: tuning process and performance evaluation, which we implemented using the NAO and Zenbo robots. User evaluation included 20 children and assessed their perceived robot embodiment preferences (e.g., shape, robot motion, displays, and emotional expressivity) and perceived usability aspects. 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 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 during the observation sessions. Finally, we propose recommendations for robot embodiment design for children and learning based on this case study and discuss protocol limitations during the social distancing context, that we believe as a valid alternative to move forward with experimental designs, particularly in robotics, becoming a great contribution to other researchers facing similar hurdles.

Keywords Social robots · Musical education · Children · User evaluation

✉ Bruno de Souza Jeronimo
bsj@cin.ufpe.br

Anna Priscilla de Albuquerque Wheler
apa@cin.ufpe.br

José Paulo G. de Oliveira
jpgo@ecomp.poli.br

Rodrigo Melo
rodrigo.melo2@ufpe.br

Carmelo J. A. Bastos-Filho
carmelofilho@ecomp.poli.br

Judith Kelner
jk@cin.ufpe.br

[1] Centro de Informática, UFPE: Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil

[2] Engenharia da Computação, UPE: Universidade de Pernambuco, Recife, Pernambuco, Brazil

[3] Departamento de Eletrônica e Sistemas, UFPE: Universidade Federal de Pernambuco, Recife, Pernambuco, Brazil

1 Introduction

A social robot supports Human-Robot Interaction (HRI) tasks through robot embodiment features (e.g., 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.) [1]. HRI is an interdisciplinary approach to understand, design, and evaluate robots for use by or with humans. Examples of social robot models are ASUS' Zenbo and Softbank Robotics' NAO in Fig. 1. 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 [2]. 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 [3]. They can also support HRI by

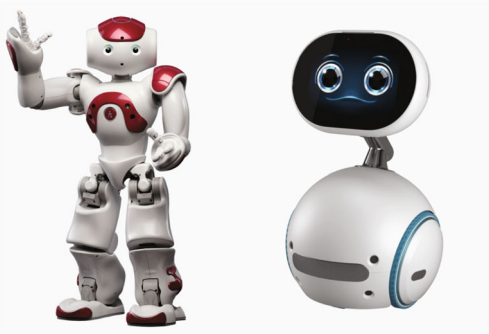

Fig. 1 NAO and Zenbo social robots

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 present research compares two social robot models running the same HRI application. The HRI application targets the context of music education for children aged 9-11. 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 [4]. We implemented the same sequence of HRI tasks according to the available robot embodiment features of two robot models using the NAO and Zenbo robots. Then, we evaluated both HRI applications with children and guardians online using an adapted version of the System Usability Scale (SUS) [5] for children – the SUS-Kids [6]. We also did a comparative 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 [7].

A reminder that this paper organizes as follows. The following subsections summarize existing knowledge on social robot embodiment features, robots supporting music education, and comparative studies with social robots. The methods and materials section details the HRI application according to each robot embodiment design and describes the user evaluation protocol. The results section 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.

1.1 Social Robot Embodiment Features

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 [8, 9]. For instance, there is evidence that anthropomorphic embodiment features can engage users [10]. 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 [11]. 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. 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 [12]. 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.

1.2 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 [13]. 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 [14]. 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 [15]. 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 [16]. 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 [17].

Interactive and AI technologies are present in several studies in music education, especially targeting children. In a study on virtual reality (VR) and augmented reality (AR) technologies in music learning, authors have integrated head-mounted displays and hand-held controllers, allowing combining VR training and musical instruments [18]. Another study combines VR and AI by developing a virtual social robot learning environment for music education [19]. In this proposed VR system, two virtual versions of the NAO robot teach children to play different notes and rhythms on a xylophone and music notes on a drum in the virtual environment. They evaluated the VR learning environment with autistic children over 20 weeks, and the results show a positive-sum in their music learning skills. This strategy can become helpful when there are no physical instruments or robots available in the learning environment. Social robots can benefit music education in different ways. For instance, in a hands-on learning study, children used the modular robotic kit KIBO to develop Science, Technology, Engineering, Arts, and Mathematics (STEAM) projects involving dance, music, and culture [20]. Children assembled and programmed modular robots to create dancing robots related to a particular culture and local music. In a similar study using modular robotics, children used them to compose songs regardless of prior musical knowledge by turning the robots into physical instruments [21].

Toy User Interfaces (ToyUI) combine hardware and software components to allow social and physical play experiences [3]. 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 [22]. In this scenario, playful training examples usually mix interactive, mixed reality, and robotic technologies to enhance tangible interaction using physical musical instruments [23, 24]. 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 [25]. 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 [19, 25, 26].

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.

1.3 Evaluating Social Robot's Embodiment

In most cases, studies comparing robot embodiment features usually focus on user perception and task performance by comparing virtual agents and physical robots or teleoperation against co-located HRI experiences [27–29]. For instance, a study compares the game Tower of Hanoi supervised by a social robot in three different settings: virtual, teleoperated, and co-located [27]. Users experienced a virtual robot version running in the Gazebo simulator, 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, providing hints and reacting to the user's task decisions. The comparative evaluation focused on task performance, and results suggest that users performed better in co-located settings than virtual and teleoperated. Another comparative study evaluated children's preferences and performance while learning from different tutors: humans, tablets, and social robots [30]. 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.

Differently, the 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 [31]. 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. The authors mentioned evaluating the robot models with education specialists but presented performance evaluation running an optimization algorithm. Notice that both Zenbo and NAO robots were evaluated in this study, ranking second and seventh places, respectively.

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 [32]. 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). Often learning systems combine social robots with other learning materials and devices, such as tangible interfaces, books, touchscreen displays, personal computers, and tablets. The authors also classified the types of robot motion in the HRI applications for learning, categorizing interactions in communicative gestures, and manipulation. In the present study, the roles of the social robots in the music education application are to provide instructions towards task completion (tuning a guitar), engage users during and after tasks (playing a song), and evaluate overall task performance (listening to the music).

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 a 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

Table 1 Related work comparison

Reference number	Agent	Interaction	Goals	Role	Audience
16	Virtual vs Physical	Simulation, remote, co-located	Task, engagement	Companion	Adults
33	Virtual vs Physical	Simulation, remote, co-located	Engagement, learning	Tutor	Children
36	Virtual vs Physical	Simulation, remote, co-located	Engagement	Companion	Adults
39	Virtual vs Physical	Simulation, co-located	Engagement, learning	Resource	Children
23	Physical robots	Co-located	Task, learning	Companion, tutor, and resource	Children
14	Physical robots	Co-located, teleoperated	Engagement, usability	Companion, tutor, and resource	Children, adults, and elderly
The present paper	Physical robots	Remote	Engagement, usability	Companion, tutor, and resource	Children

research facilities. Therefore, the evaluation used pre-recorded videos of the HRI tasks, and children engaged as observers.

2 Methods and Materials

The research method mixes quantitative and qualitative approaches, evaluating two social robot models running a HRI application with children and guardians. 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. 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 AI evaluation performance. The robot reacts to music selection and music performance to engage the child in performing the task and 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, etc.). The following subsections compare application implementation in each robot model and detail the user evaluation protocol during the COVID-19 pandemic.

2.1 Guitar Tuner and Evaluation Performance Design Adaptation

The first version of the playful training application supported implementation in the NAO 5 robot using the NAOqi SDK [4]. Here, we adapted the 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. Table 2 compares NAO's and Zenbo's robot embodiment features. In overview, both robot models are humanoid, 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.

We fully implemented the application in the NAO robot using the NAOqi 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 [33]. The WoZ technique is a helpful resource to test interactive behaviors without fully implementing them. The goal is to implement the interaction and feedback, not to notice that implementation is not fully functional. A full description of NAO implementation is available in previous work [4]. 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. 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 misbehaviors. 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.

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 [4]. 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

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/, https://zenbo.asus.com/.

emotional expressivity and attention. Despite not having any limbs, the Zenbo robot simulates dancing, making quick turns around its axis, displaying a singing facial expression, while flashing the 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. 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. 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

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.

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 NAO-mark 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.

towards emotional expressivity (e.g., dancing to celebrate or demonstrate sorrow).

2.2 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 a 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 [3]. 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. 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 [4]. Differently, we recorded the Zenbo video at home without an audience or noise interference. Also, the Zenbo video displayed the robot on the floor, as referenced in Fig. 2, 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.

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 [6], consisting of 13 statements using 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 [34, 35]. 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.

3 User Evaluation

The recruiting sample gathered data from 22 children and 17 guardians. After excluding incomplete data, the final sample remained on 20 children and 15 guardians. One guardian canceled their interview, a child opted to leave the study, and another one failed to submit her evaluation form. Participants' locations varied from Brazil, United States, Canada, and Europe. All guardians were the kids' parents, nine female and six male with post-secondary education, including four parents with master's degrees and four doctoral degrees. Most parents aged between 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 aged 9-11 (14 children). 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 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's names), 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 at the same time. 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. Each session started with a brief presentation of our research goals and tasks for the interviewees, and collection of data 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, and then we discussed the open questions about their robot embodiment preferences. For interviews with

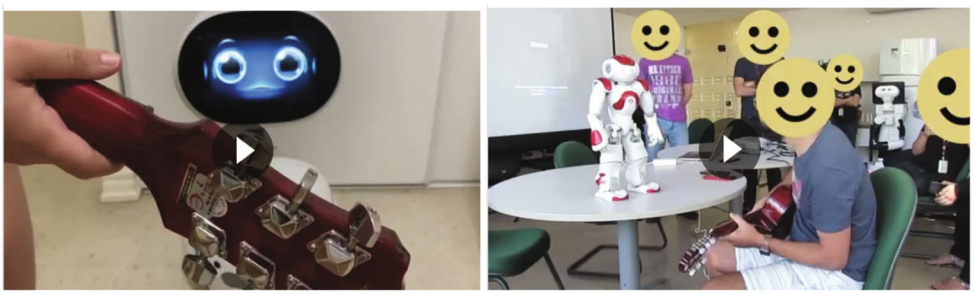

Fig. 2 Screenshot of video footage

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

SUS-Kids for Social Robots	SUS-Kids [6]
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].

more than one child, we asked them to wait for each one to fill out the evaluation form first. We performed the qualitative discussion together, also considering inputs from the parents. After all sessions, we analyzed the quantitative data using Google Sheets associated with Google forms, and we performed some statistical tests to verify some conclusions. Finally, we transcribed qualitative responses to text using the automated videos from the interviews, permitting us to classify feedback into themes and tags.

3.1 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 [5]. 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 [6]. 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.

3.1.1 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 3 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 Fig. 4, it is noticeable that participants with higher music levels chose Zenbo, while participants who chose NAO have a lower musical level.

3.1.2 Musical Level and SUS-Kids Scores

Figure 5 lists the SUS-Kids score and musical level parameters of the participants. Visually, there is no indication of dependence between these two variables.

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

The correlation coefficient between each sample X and Y is calculated with function corr(X,Y) form Matlab:

$$[\rho, p_{val}] = corr(SUS\_Score, Musical\_Level), \quad \rho = 0.1428, \; p_{val} = 0.5481. \tag{1}$$

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 Fig. 6 shows no relationship between the value of the SUS-Kids score and the musical level.

3.1.3 SUS-Kids Scores and Age

From the graph in Fig. 7, 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.

$$[\rho, p_{val}] = corr(SUS\_Score, Age), \; \rho = 0.4583, \; p_{val} = 0.0421. \tag{2}$$

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 Fig. 8 shows a

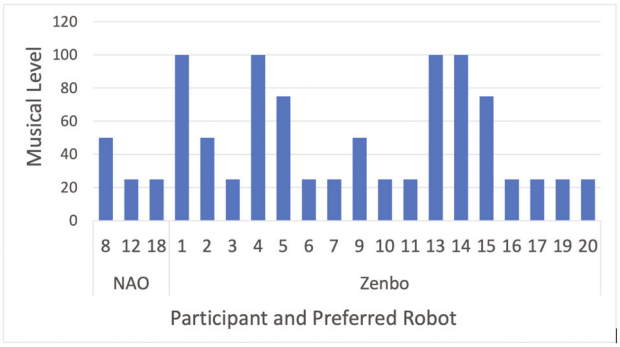

Fig. 3 Music level and preferred robot

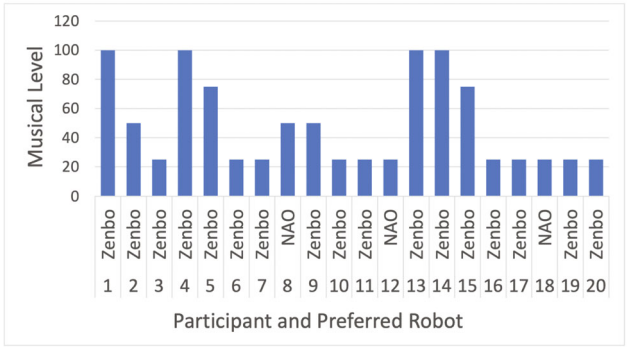

Fig. 4 Higher music level and preferred robot

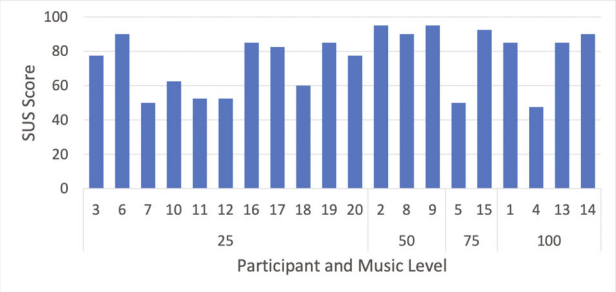

Fig. 5 SUS-Kids score vs music level

concentration of higher values for the SUS-Kids score when the participant is older.

3.1.4 Robot Preference and Gender

Finally, the distributions illustrated in Figs. 9–10 indicate no differences between genders and SUS-Kids score values of the participants. The Kruskal-Wallis test returns the p-value for the null hypothesis that the data in each column of [Robot, Gender] comes from the same distribution, using a Kruskal-Wallis test. The alternative hypothesis is that not all samples come from the same distribution.

$$p_{val} = kruskalwallis([Robot, Gender]), \quad p_{val} = 0.0088. \tag{3}$$

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.

3.2 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

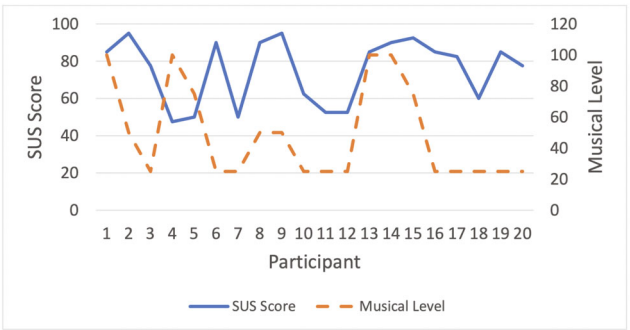

Fig. 6 SUS-Kids score and music level correlation

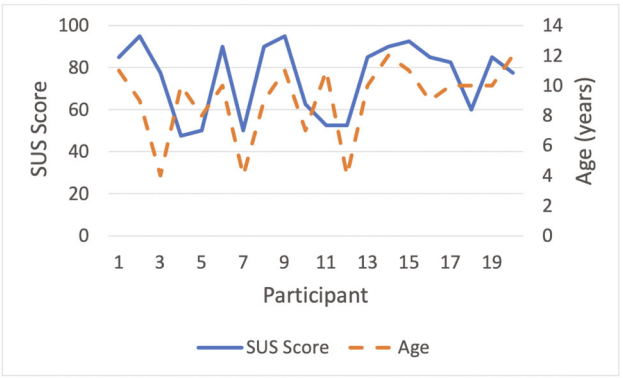

Fig. 7 SUS-Kids score and correlation with age group

Appearance and Usability; Robot Emotion and Behavior; and 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 NAO-marks to the robot. Another recurrent remark was that having access to the music scores on the robot's display was more convenient than relying on the NAO's companion application.

3.2.1 Robot Appearance and Usability

Most of the 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

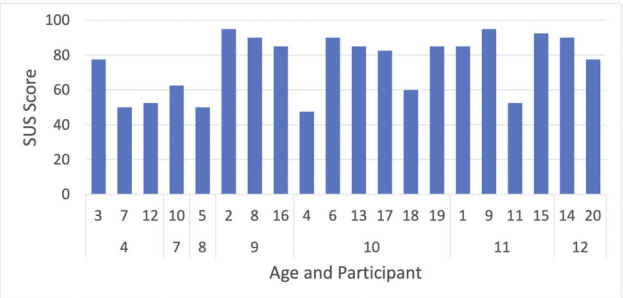

Fig. 8 SUS-Kids score by age group

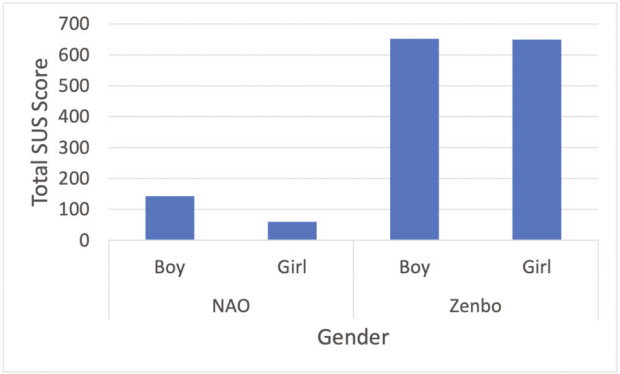

Fig. 9 Robot preference and gender

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 hide 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.

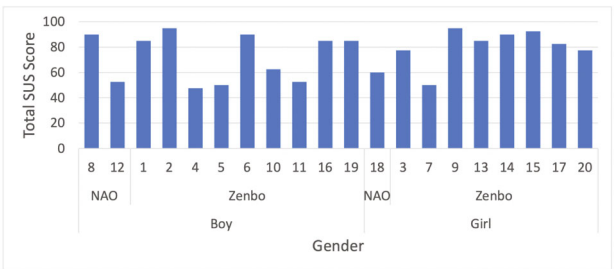

Fig. 10 SUS-Kids score by gender and robot preference

3.2.2 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 [36]. In most cases, robot emotion presented as a desirable robot embodiment feature for children, 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, making her feel discomfort, and she enjoyed the dancing and corporal expressivity of the NAO robot more (she preferred the NAO robot in the survey).

3.2.3 Content, Additional Features, and Software

Most participants suggested the application should have more music options, more 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 it would be nice to use robots in other learning contexts, such as replacing a tutor in homeschooling. Parents who participated or supervised the interviews also gave us spontaneous feedback during the evaluation. A remarking comment was about the robot's feedback to 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 therefore empathise 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.

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"			

4 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.

5 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. 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 a useful 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, also supported by the literature review. The Zenbo robot introduced a pleasant appearance, good 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 customizing 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 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. Ultimately, we must retake NAO's video footage to prevent perceptual noise in future data collections. 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.

Acknowledgements This study was financed in part by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES). The authors also thank Fundação de Amparo à Ciência e Tecnologia do Estado de Pernambuco (FACEPE) under grant IBPG-0844-3.04/17, as well as for the project PRONEX 2014 - APQ-0880-1.03/14.

Author Contributions Writing and reviewing: B. S. Jeronimo and A. P. A. Wheler; Data collection: B. S. Jeronimo, A. P. A. Wheler, and R. Melo; Research materials: B. S. Jeronimo, A. P. A. Wheler, and R. Melo; Supervising: C. J. A. Bastos-Filho and J. Kelner; Data analysis: B. S. Jeronimo and J. P. G. de Oliveira.

References

1. Bartneck, C., Forlizzi, J.: A design-centred framework for social human-robot interaction. In: RO-MAN 2004. 13th IEEE

International Workshop on Robot and Human Interactive Communication (IEEE Catalog No. 04TH8759), pp. 591– 594 (2004). IEEE
2. Bartneck, C., Belpaeme, T., Eyssel, F., Kanda, T., Keijsers, M., Šabanović, S.: Human-robot Interaction: An Introduction Cambridge University Press (2020)
3. de Albuquerque Wheler, A.P., Kelner, J., Hung, P.C., de Souza Jeronimo, B., Junior, R.D.S.R., Araújo, A.F.R.: Toy user interface design—tools for child–computer interaction. Int. J. Child-Comput. Interact. **30**, 100307 (2021)
4. Melo, R., de Paula Monteiro, R., de Oliveira, J.P.G., Jeronimo, B., Bastos-Filho, C.J., de Albuquerque, A.P., Kelner, J.: Guitar tuner and song performance evaluation using a nao robot. In: 2020 Latin American Robotics Symposium (LARS), 2020 Brazilian Symposium on Robotics (SBR) and 2020 Workshop on Robotics in Education (WRE), pp. 1–6 (2020). IEEE
5. Bangor, A., Kortum, P.T., Miller, J.T.: An empirical evaluation of the system usability scale. Intl. J. Hum.-Comput. Interact. **24**(6), 574–594 (2008)
6. 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, pp. 1–7 (2020)
7. 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 (2020)
8. Goodrich, M.A., Schultz, A.C.: Human-robot interaction: a Survey Now Publishers Inc (2008)
9. Hancock, P.A., Billings, D.R., Schaefer, K.E.: Can you trust your robot? Ergon. Des. **19**(3), 24–29 (2011)
10. Duffy, B.R.: Anthropomorphism and the social robot. Rob. Auton. Syst. **42**(3-4), 177–190 (2003)
11. Breazeal, C., Dautenhahn, K., Kanda, T.: Social robotics. Springer handbook of robotics, pp. 1935–1972 (2016)
12. Li, H., John-John, C., Tan, Y.K.: Towards an effective design of social robots. Int. J. Soc. Robot. **3**(4), 333–335 (2011)
13. Webster, P.R.: Computer-Based Technology and Music Teaching and Learning: 2000–2005. In: International Handbook of Research in Arts Education, pp. 1311– 1330. Springer (2007)
14. 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: 2013 Second International Conference on E-Learning and E-Technologies in Education (ICEEE), pp. 149– 154 (2013). IEEE
15. Waddell, G., Williamon, A.: Technology use and attitudes in music learning. Frontiers in ICT **6**, 11 (2019)
16. 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), pp. 124–128 (2019)
17. Paule-Ruiz, M., ÁLvarez-garcía, V., Pérez-Pérez, J.R., Álvarez-Sierra, M., Trespalacios-Menéndez, F.: Music learning in preschool with mobile devices. Behav. Inf. Technol. **36**(1), 95–111 (2017)
18. Serafin, S., Adjorlu, A., Nilsson, N., Thomsen, L., Nordahl, R., Considerations on the use of virtual and augmented reality technologies in music education. In: 2017 IEEE Virtual Reality Workshop on K-12 Embodied Learning Through Virtual & Augmented Reality (KELVAR), pp. 1– 4 (2017). IEEE
19. 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, 1–25 (2021)
20. Sullivan, A., Bers, M.U.: Dancing robots: integrating art, music, and robotics in singapore's early childhood centers. Int. J. Technol. Des. Educ. **28**(2), 325–346 (2018)
21. Nielsen, J., Bærendsen, N.K., Jessen, C.: Robomusickids–Music Education with Robotic Building Blocks. In: 2008 Second IEEE International Conference on Digital Game and Intelligent Toy Enhanced Learning, pp. 149–156 (2008)
22. de Albuquerque, A.P., Kelner, J.: Toy user interfaces: systematic and industrial mapping. J. Syst. Archit. **97**, 77–106 (2019)
23. 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, pp. 1447–1452 (2011)
24. Yamabe, T., Nakajima, T.: Playful training with augmented reality games: case studies towards reality-oriented system design. Multimed. Tools Appl. **62**(1), 259–286 (2013)
25. 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: 2015 IEEE International Conference on Rehabilitation Robotics (ICORR), pp. 828–833 (2015). IEEE
26. 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, J Behav. Robot. **12**(1), 256–275 (2021)
27. Wainer, J., Feil-Seifer, D.J., Shell, D.A., Mataric, M.J.: The role of physical embodiment in human-robot interaction. In: ROMAN 2006-The 15th IEEE International Symposium on Robot and Human Interactive Communication, pp. 117–122 (2006). IEEE
28. Kennedy, J., Baxter, P., Belpaeme, T.: Comparing robot embodiments in a guided discovery learning interaction with children. Int. J Soc. Robot. **7**(2), 293–308 (2015)
29. Thellman, S., Silvervarg, A., Gulz, A., Ziemke, T.: Physical vs. virtual agent embodiment and effects on social interaction. International Conference on Intelligent Virtual Agents, pp. 412–415 (2016). Springer
30. 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 (2015)
31. Papakostas, G.A., Strolis, A.K., Panagiotopoulos, F., Aitsidis, C.N.: Social robot selection: a case study in education. In: 2018 26th International Conference on Software, Telecommunications and Computer Networks (SoftCOM), pp. 1–4 (2018). IEEE
32. Johal, W.: Research trends in social robots for learning. Current Robotics Reports, 1–9 (2020)
33. Kelley, J.F.: Wizard of oz (woz) a yellow brick journey. J Usability Stud. **13**(3), 119–124 (2018)
34. 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, pp. 156– 165 (2010)
35. 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, pp. 241– 248 (2012)
36. Paiva, A., Leite, I., Ribeiro, T.: Emotion modeling for social robots. The Oxford handbook of affective computing 296–308 (2014)

Bruno de Souza Jeronimo is pursuing his master's degree in Computer Science at Universidade Federal de Pernambuco (CIn/UFPE). He has a BSc in Business and Management from the UFPE and a diploma in iOS Development from the Apple Developers Academy program. He is a researcher in the Virtual Reality and Multimedia Group - GRVM, with his areas of interest focused on Software Development, Mobile Devices, User Research and Experience, Human-Computer Interaction, Human-Robot Interaction, Social Robots, Social Robots for Education, and Aspects of Security and Privacy in Social Robots.

Dr. Anna Priscilla A Wheler is a Postdoctoral Fellow at Ontario Tech University in Canada, researching Inclusive Design and Social Robotics. She received her Ph.D. in Computer Science at Universidade Federal de Pernambuco (UFPE), Brazil, in 2021. She also holds a Computer Science master's degree and a Design bachelor's degree, both in UFPE. She has an equivalent BA degree in Computer Games Design (Story Development) at the University of East London in the UK. Her primary research domains are Child-Computer Interaction, Human-Robot Interaction, Human-Centered Design, and Rapid Prototyping.

J. P. G. de Oliveira is pursuing his Ph.D. at the Electrical Engineering Department at the Federal University of Pernambuco (UFPE) and is an Assistant Professor at the Computer Engineering Department at the University of Pernambuco (UPE), Brasil. His research interests span the areas of embedded systems, applied machine learning, industry 4.0, audio signal processing and optical communication through the atmosphere. He received the B.Sc. degree in electronics engineering and the M.Sc. degree in electrical engineering from the Federal University of Pernambuco (UFPE), in 2001 and 2003, respectively. He has more than 15 years of experience in embedded systems research and development.

Rodrigo Melo is pursuing his Ph.D. at the Electrical Engineering Department at the Federal University of Pernambuco (UFPE). His research interests involves robotics applications, computational intelligence and machine learning. He received his B.Sc. in Mechanical Engineering from the Federal University of Pernambuco (UFPE), with a one year scholarship at the University of Adelaide, and his M.Sc. degree in System Engineering from the University of Pernambuco (UPE), in 2015 and 2017, respectively.

Prof. Carmelo Bastos-Filho is a Senior Member of IEEE, was born in Recife, Brazil, in 1978. He received the B.Sc. degree in electronics engineering and the M.Sc. and Ph.D. degrees in electrical engineering from the Federal University of Pernambuco (UFPE), in 2000, 2003, and 2005, respectively. From 2016 to 2020, he was the Scientist-in-Chief of the Technological Park for Electronics and Industry 4.0 of Pernambuco. He is a Professor at University of Pernambuco. He is the Director for Innovation Parks and Graduate Studies of Pernambuco. He is also the Co-ordinator of the National Initiative for Promoting ICT in Pernambuco, jointly with the Ministry for Regional Development. He is a Research Fellow Level 1D of the National Research Council of Brazil (CNPq). He published roughly 300 full papers in journals and conferences and advised over 50 Ph.D. and M.Sc. candidates.

Prof. Judith Kelner is a full faculty member of the Center of Informatics in UFPE since 1979. She is the co-founder and research leader of the GRVM team since 1998. Prof. Kelner received her Ph.D. from the Computing Laboratory at the University of Kent at Canterbury, the UK in 1993. She currently teaches multimedia and interactive systems and applications, and coordinates numerous research projects in the areas of computer vision, robot interaction, cloud computing, data science visualization and smart communication devices. She has a partnership with Petrobras, Sapura, Chesf, Ericsson and several others companies. Lately she received a grant from the Bill & Melinda Gates Foundation in the call for Grand Challenges Explorations: Data Science Approaches to Improve Maternal and Child Health.