2024 IEEE International Conference on Robotics and Automation (ICRA 2024)
May 13-17, 2024. Yokohama, Japan

Design of Embodied Mediator Haru for Remote Cross Cultural Communication

Randy Gomez[1], Deborah Szapiro [2], Sara Cooper [3], Nabil Bougria[4], Guillermo Pérez[5], Eric Nichols[1], Javier Giménez-Figueroa[4], Jose M. Perez-Moleron[4], Matthew Peavy[4], Daniel Serrano[6] and Luis Merino[4]

***Abstract*— Social robots for children have focused mainly on conventional education domains such as teaching language, science, and math, while applications focusing on the enhancement of cultural competency are quite scarce. In this paper, we present a prototype of a robot-mediation framework for cross-cultural communication. This framework paves the way for a social robot to act as a mediator between groups of schoolchildren from different countries. First, we conducted a participatory design activity by an interdisciplinary team, resulting in the extraction of the design, robot's roles, and technical requirements. Based on these requirements, we built the robot-mediation system prototype. We conducted a pilot study using the system with groups of high school children in Japan and Australia and our results show the potential of the system to drive children's interest in communicating, sharing, and discussing cultural themes with their remote peers through the social robot.**

I. Introduction

The use of social robots for children has witnessed a remarkable surge in recent years, in tandem with the rapid advancements in artificial intelligence (AI) technologies. This surge has garnered considerable attention from influential organizations, including UNICEF [1], the Organization for Economic Co-operation and Development (OECD), and others, reflecting the growing recognition of the potential impact of social robots on childhood development.

Specific applications of social robots in human-child interaction domain includes the exploration of improving the learning experience of children. Researchers have deployed the general-purpose humanoid robot Pepper as a language tutor for Japanese children learning English as a foreign language [2]. Separate research also has explored the use of Pepper for STEM education [3]. Relevant research using QTRobot by LuxAI reported the potential use of a robot as a teaching assistant for traditional subjects making the learning experience more interactive and enjoyable [4]. Charisi et al. has carried out research on robot-children interaction for problem-solving with the robot Haru [5], [6].

[1]Randy Gomez is with Honda Research Institute, Wako, Japan r.gomez@jp.honda-ri.com
[2]Deborah Szapiro is with University Technology Sydney, Sydney, Australia deborah.szapiro@uts.edu.au
[3]Sara Cooper is with Honda Research Institute, Wako, Japan sara.cooper@jp.honda-ri.com
[4]N.B., J.G-F., J.P-M., M.P. and L.M are with Universidad Pablo de Olavide, Sevilla, Spain lmercab@upo.es
[5]Guillermo Pérez is with 4i Intelligent Insights, Sevilla, Spain g.perez@4i.ai
[6] Daniel Serrano is with Eurecat, Barcelona, Spain daniel.serrano@eurecat.org

Another facet of the use of social robots in child-robot interaction is the socio-emotional aspect of child development. Social robots have been actively developed to foster emotional intelligence and empathy. For example, Moxie by Embodied has been designed to engage through conversations and storytelling to build children's confidence and enhance both emotional and cognitive skills growth [7], [8]. Due to the promise of social robots in the context of social and emotional learning, there have been rich research works deploying robots for autism as well [9], [10].

While the surge in social robots for children has brought about a diverse landscape of innovative applications to enhance children's development both in the emotional and cognitive domains, it is crucial to acknowledge that the predominant focus remains on conventional education domains such as teaching language, mathematics, and sciences. A significant disparity exists in the attention subjects involving robotic applications that relate to enhancing cultural competency. This disparity raises pertinent questions about the holistic development of children, urging a shift toward more balanced and inclusive approaches that touch upon cultural competency on top of empathy, and interpersonal relationship-building from an early age. Therefore, it is imperative to recognize the role of social robots as tools for promoting cultural understanding and diversity among children to further drive the research in this field.

In this paper, we will show the methods for developing a prototype of a robotic framework that acts as a mediator for cross-cultural communication. In Sec II, we identify the problems impeding cross-cultural communication and then extract system requirements (i.e., design, robot roles, technical) as presented in Sec III. The system implementation based on the extracted requirements is discussed in Sec IV. Then, we discuss the interaction session for the experiments including the results in Sec V. Finally, we conclude the paper in Sec VI including future works.

II. Background

A. Haru the Robot

Haru is a new kind of robotic being designed to make people happy – a robot that makes people smile and fall in love with it, and through its rich, expressive capabilities, to create meaningful connections with people [11], [12]. One particularly important field of research in which Haru is engaged is intelligent communication and social interaction, where Haru acts as a mediator for humans, helping to bridge the gap of cross-cultural differences. In the long

979-8-3503-8457-4/24/$31.00 ©2024 IEEE

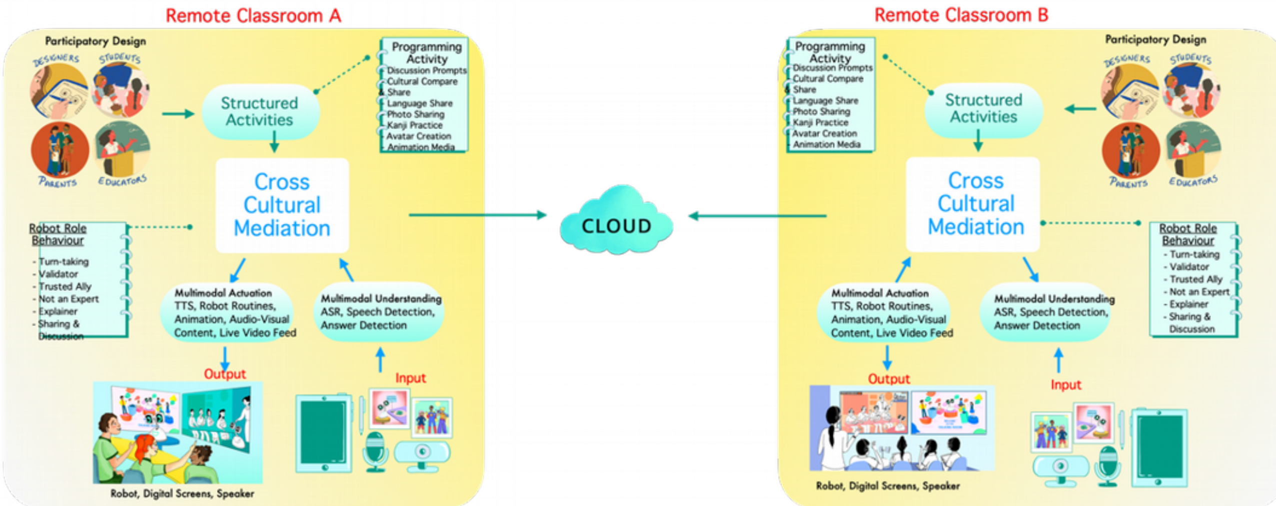


Fig. 1: Design Block Diagram of the Robot System for Cross-cultural Communication

term, Haru will develop and learn transformative perception and interaction capabilities that allow Haru to connect with people for the benefit of cultivating a better understanding of their differences and fostering good relationships. Haru's mediator role for cross-cultural communication was first declared in the Pilot study report for UNICEF [13]. From that time onward, we have been working with our consortium partners [14] to build the first robotic prototype for the cross-cultural mediation system reported in this manuscript.

B. Problems Impeding Cross-cultural Interaction

Cross-cultural mediation is a challenging task for humans, let alone for robots to facilitate. There are four main challenges that impede cross-cultural interaction:

(a) Geographical Location: Physical distance separates people from different regions, making face-to-face interactions challenging, and limits opportunities to connect with peers from diverse backgrounds
(b) Communication Barrier: Children's diverse readiness, confidence, and shyness levels may cause children to struggle to express themselves, leading them to withdraw from reaching out to their peers. Stimulation and recognition of children's multifaceted nature of interaction and respect for the individual's pace are important [28].
(c) Culture and Diversity Themes, not Common or Mainstream: Navigating subjects related to culture and diversity can present challenges for young minds not only due to their non-mainstream nature but also because of their inherent complexity. It is important to simplify and personalize the activities and contents by centering on their experience, making it more relatable. Sourcing content directly from the children themselves, educators, and designers would also help [29].
(d) Didactic Manner of Imparting Cultural Diversity: Teaching cultural diversity subjects can be intricate, and conventionally, these subjects are often delivered in a didactic manner, much like traditional subjects. It is important to explore different paths when using robots, steering clear of didactic methods. Instead of robots teaching cultural themes, robots only facilitate open dialogues and mutual sharing of personal experiences and let children drive the discussion themselves [30].

III. System Requirements Extraction

We gathered an interdisciplinary team [14] of child experts, educators, and young teens along with design and interaction experts, roboticists, and programmers to define the system requirements (i.e., Design, Robot's Role, and Technical).

A. Design Requirements

The first task of the inter-disciplinary team is identifying the design requirements that are fundamental in addressing the problems discussed in Sec II. The team identified eleven major requirements, 1. Meeting Face to Face; 2. Natural Interaction; 3. Structured Activity; 4. Turn-taking; 5. Multimodal Learning Strategies; 6. Technology Demystified; 7. Personalized and Relatable Activity; 8. Safe Space; 9. Power Balance; 10. Building Confidence; 11. Overcoming Shyness. These requirements are summarized in the first column of Table I and are consistent with the UNICEF SDGs [1].

B. Robot Roles Requirements

The system is built by centering on the role of the robot. Roles appropriate to the research purpose should be identified right from the very beginning as a response to the design requirements to ensure holistic robot functionality. The interdisciplinary team identified the robot roles as follows, 1. Social Connector; 2. Supporting Multimodal Interactivity; 3. Activity Navigator; 4. Equitable Turn-taking; 5. Fun and Engaging Experience; 6. Transparent and Equitable Exemplar; 7. Values Experience and Sharedness; 8. Trusted Ally; 9. A Peer, Not an Expert; 10. A Validator; 11. Ice Breaker. These roles, summarized in the second column of Table I are specific yet also have subtle overlaps to make them cohesive within the framework of the robot's personality and functionality.

TABLE I: Design and technical requirements for the project, and robot roles

Design Requirements	Robot Role Requirements	Technical Requirements
1. Meeting Face to Face The ability to see someone's face transcends beyond words, fostering genuine connection promoting active listening, open dialog and trust [15].	**Social Connector** Providing a social conduit to remotely connect and do meetups with children from various schools around the world, the robot is present in each location and is key to supporting two-way communication between the children	Remote Connection and video conferencing Synchronous and Asynchronous Communication
2. Natural Interaction Children need to interact naturally, a way that mirrors the real world. Multimodal interactivity reflects common communication practices in children's lives [16] .	**Supporting Multimodal interactivity** Provides the perception mechanism from input of raw data to contextual understanding using speech input, gestural input, hand-writing input tec.	Multimodal Perception system, including body and facial gestures, speech and language. Expanding support of modalities through tablet as input tool for text and speech.
3. Structured Activity Providing a structured and open-ended approach that involves complex interaction can simplify learning and bring positive impacts to the experience [17]	**Activity Navigator (More on System)** The robot guides the children through a set of curated activities for the entire duration of the interaction with the ability to navigate through, dialogue and interaction across modalities and seamlessly transition from one activity to the next.	Behavior control system that orchestrates all the interaction modalities (dialogue, non-verbal actuation) of the system to navigate the different activities
4. Turn-taking Interaction in groups is challenging. Group activities that facilitate turn-taking through explicit prompts for students to articulate the subject matter foster a fairer and more inclusive environment [18]	**Equitable Turn-taking** Ensures all children get the chance to talk and interact by providing a mechanism to detect interaction activities of children and decide which kid needs to talk next, give some ample time for the kid to be ready.	Speech activity detection is needed to avoid interrupting when children are talking, or to detect moments when the mediator should/can intervene.
5. Multimodal Learning strategies Increased engagement (fun), connection to peers, and knowledge acquisition when information is presented through a multimodal delivery system [19].	**Fun and Engaging Experience** The robot seamlessly facilitates the interactive, multimodal experience, creating a dynamic and fun environment through the various communication modes, animated content, question and subject prompts and the robot's personality.	Multimodal actuation set, including robot expressivity routines, audiovisual content. These elements need to be fully controllable from the orchestrator. Synchronization between the two classrooms is required
6. Kids need technology demystified Kids lack experience with robotic technology and AI. Being transparent it functionality and limitations supports children's confidence [1]	**A Transparent and Accountable Exemplar:** Haru is designed to UNICEF SDGs and introduces itself with clear explainability and transparency about privacy and functionality. This demystifies the technology and provides an exemplar for technological interactions. [13]	Explainability and transparency is built into the robot interaction with the children; Closed loop privacy system.
7. Personalized and Relatable Activity Relatable content connects students' interests and supports personalization [20] and builds a sense of community and belonging through sharedness [21]	**Values Experience and Sharedness** Haru facilitates communication about children's personal experiences through language and sharing of photos. This includes but not limited to the sharing of the different kinds of foods, school practices in their respective countries.	Participatory Design: Multi-stakeholder approach for the design of content and robot interaction. Contents are sourced from children, educators and designers.
8. Safe Space Respectful interactions creates a safe space for students to express themselves without fear of rejection [22]	**A Trusted Ally** The robot Haru creates a 'safe space' through non-judgmental conversation and feedback and projects a positive attitude in which children feel that there are no wrong answers.	Design of robot routines conveying positive personality using verbal and nonverbal cues. Remembering childrens faces, names and answers promotes a sense of connection and belonging and respect in social interaction [23]
9. Power Balance Social and emotional learning in adolescence is highly influenced by peer relationships [24] The power dynamic of adolescent peers is one of relative equals, which is different to the dynamic children have with parents and teachers.	**A Peer, Not an Expert** The robot is of a social character that engages children as a peer. It is not an expert or authority. Haru doesn't need to know everything, the robot lets the students know that it is learning from them in the same way that the children are learning things from each other.	Dialog system is designed for social sharing of topics and not for factual Q&A Haru's interaction is designed for that of a participant and not of a teacher
10. Building Confidence Validation is associated with improved emotional regulation in adolescents and greater satisfaction in relationships [25]	**A Validator** Haru encourages positive social conversation and validates the students' experiences and opinions through positive feedback, turn taking and the remembering of personal facts about each student.	Dialog detects answers and reinforces those For example I like pizza Sentiment analysis: When Haru detects sentiments, it reinforces them.
11. Overcoming Shyness Children can sometimes be shy about talking in a group. Conversational 'Ice Breakers' encourage students to bond, helps children overcome awkwardness and build bonds with their peers [26], [27].	**An Ice Breaker** Haru uses conversational interaction and 'silly' humour to break the ice. The content provides Haru with material for facilitating conversation and cracking jokes. Haru can proactively encourage the continuity of the conversation by sharing first when children hesitate to start.	The robot can recognize if there is a long pause or no reply that requires an 'Ice Breaking' prompt or joke.

C. *Technical Requirements*

Finally, the different robot's roles imply a set of technical requirements for the implementation which is discussed later in this paper. These requirements are described in the third column of Table I. As a summary, the technical requirements can be grouped into requirements for the system components (rows 1, 2), perception module (2, 4, 8, 11), behavior control (3), actuation modalities (5, 8), dialogue (9 to 11) and the implementation of children-robot interaction (6, 7, 9). The overall block diagram of the system requirements (design, robot roles, technical) is shown in Fig 1

IV. System Setup and Implementation

Following the technical requirements from Table I, this section describes the implementation of the main modules for an embodied mediator for cross-cultural exchange between students in different countries (see Fig. 1). Each section considers one or several of the requirements there.

A. *System components for a Social Connector*

The social connector consists of the following components:

- Two Haru robots, one in each classroom. Haru acts as the social mediator in each class and they coordinate to mediate between classes.
- Two screens for content display and video conference. The video conference supports meetings between schools around the world.
- Tablets (iPads) for each children, which are used for audio and textual inputs to the system.

All components of a classroom are integrated into a ROS system with its own roscore module. The two ROS systems are communicated by means of the FKIE multi-

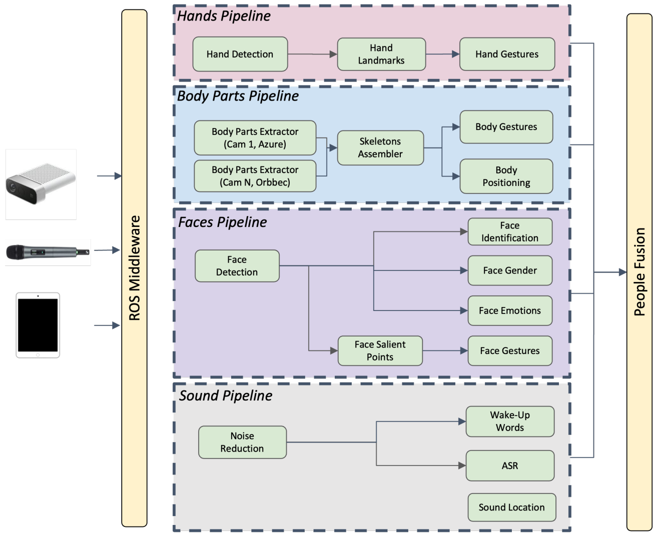


Fig. 2: Multi-modal perception system pipelines.

master module[1]. The two classrooms are connected between themselves using a VPN. This is used by the system to support synchronous and asynchronous communications between the classrooms.

B. Perception system to support multi-modal interactivity

A natural multi-modal interaction requires a perception mechanism for children's understanding. Here we summarize the perception system of Haru, presented in detail in [31].

The system uses as its main input an Azure Kinect camera, including its microphones, but it can incorporate information from additional cameras and microphones. It publishes through a ROS topic the `People` message, which contains information about all persons observed by the cameras. The system combines data from a set of feature extraction pipelines (see Fig. 2).

The body-parts pipeline extracts, for each person, a skeleton up to 32 body parts, and detects a range of body gestures (i.e. waving, looking at the robot). The hands and faces pipelines extract a set of features and gestures from the visible hands and faces respectively. The system uses the microphone of the iPads and Apple's Automatic Speech Recognition (ASR) to obtain the text from the children. All the features are integrated into a single data structure per person using data association [31].

C. Expressivity and Interaction Devices for a Fun and Engaging Experience

The robot also requires a multi-modal set of actuation modalities to participate in the learning experience. On one hand, Haru itself has a set of expressive routines that combine motion, eye, and sound to convey a plethora of expressions [32]. We have also developed a Unity application to display interactive content on a screen. This application is integrated into ROS, and the content can be fully controlled by calling the adequate services. Furthermore, we have integrated into the Unity application a commercial video conference system, Whereby[2]. The videoconference can thus be fully controlled from ROS, muting, and un-muting participants, initiating calls, etc.

D. Conversational capabilities

The dialogue system is a key component to meet many of the requirements in Table I, in particular, requirements 9 to 11. The conversational module is based on Google's Dialog Flow, and uses as inputs the transcribed text from each person provided by the ASR. On top of it, the module extracts Natural Language Understanding (NLU) features that are made available to the rest of the system.

On one hand, sentiment analysis is used to determine whether a given text is positive, neutral or negative. The analyzer operates on both input text (what is being said to the system) and output text (what the system is 'saying' back to the speaker). We analyze the text as a whole and on a per-sentence basis. The sentiment analyzer outputs a score from 0 to 1 on positive, neutral, and negative metrics. Also, a compound score is calculated, representing overall sentiment from -1 to 1, where the lower end of the scale represents negative sentiment and the higher end of the scale denotes positive sentiment.

The NLU module also performs emotion analysis on all transcribed text, as well as all text being generated for the speaker. The model classifies text into 6 basic emotions - anger, disgust, fear, joy, sadness and surprise, as well as a neutral class.

Entity detection allows for certain words to be extracted from a sentence if they belong to a certain word group. For instance, suppose it is important for the system to detect food terms in their speaker's replies, like ramen or sushi, etc. Entity detection allows for the extraction of words that belong to a certain entity type.

E. Behavior control for activity navigation, including turn-taking

The robot needs to control all the components described above to navigate the interaction with the children. All elements of the system are interfaced with ROS through topics, services and actions (we employ the ROS' `actionlib` package[3], which offers a simple interface to preemptable tasks). On top of those interfaces, we selected Behavior Trees [33] (BTs) as the method to implement this behavior control. BTs are becoming widespread in robotics, mainly due to their modularity and simplicity [33], [34]. The BT[4] is used to orchestrate all the elements (dialogue, non-verbal actuation, perception) to deliver the different activities that compose the interaction session.

An important aspect of the interaction is turn-taking. A very distinctive element of the system considered in this paper is that we have two robots mediating between two

[1] `https://wiki.ros.org/multimaster_fkie`

[2] `https://whereby.com`

[3] `http://wiki.ros.org/actionlib`

[4] We use the engine of `https://github.com/BehaviorTree/BehaviorTree.CPP/tree/v3.8`

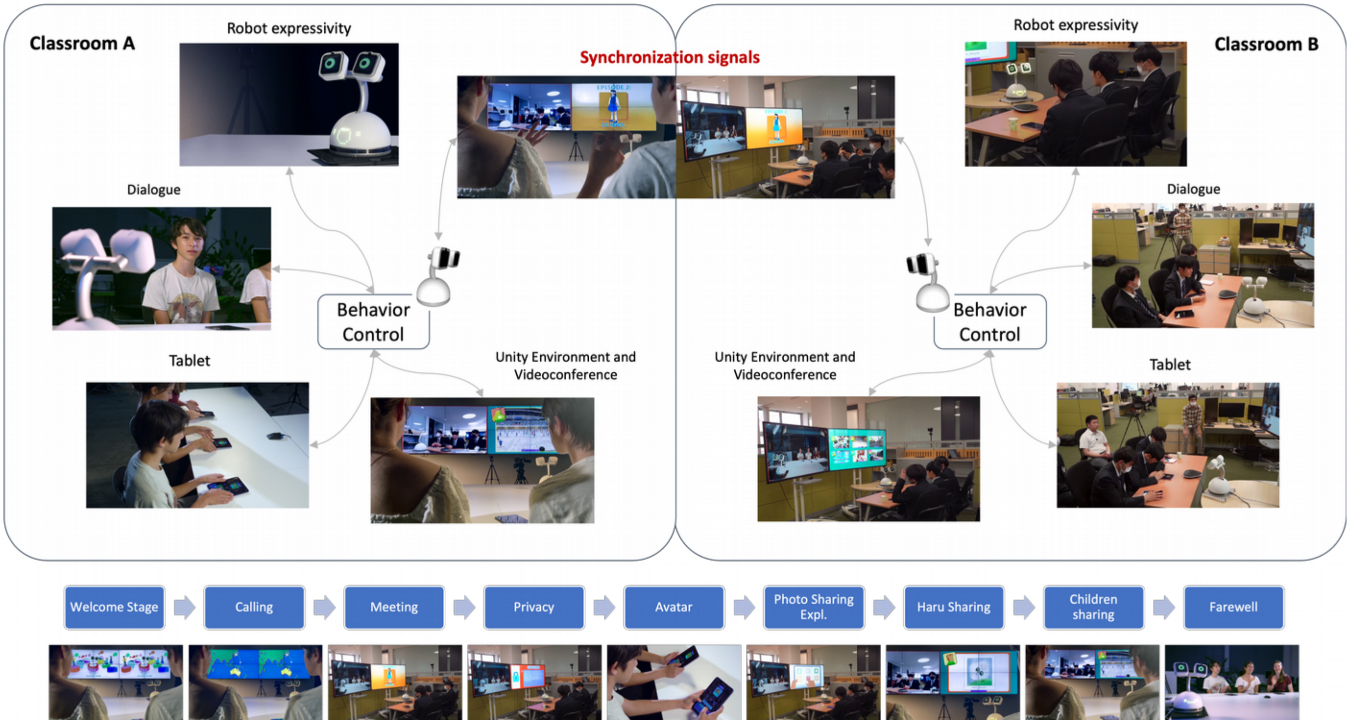


Fig. 3: Interaction session as deployed in our pilot experiment (Australia and Japan). Main elements: a robot capable of verbal/non-verbal communication; a screen for visual content and videoconferencing; a tablet for sound and text input. The robots need to synchronize during the interaction session. The whole session consists of 9 phases (bottom).

classrooms. Therefore, we consider two aspects in turn-taking: in-class and between-classes turn-taking.

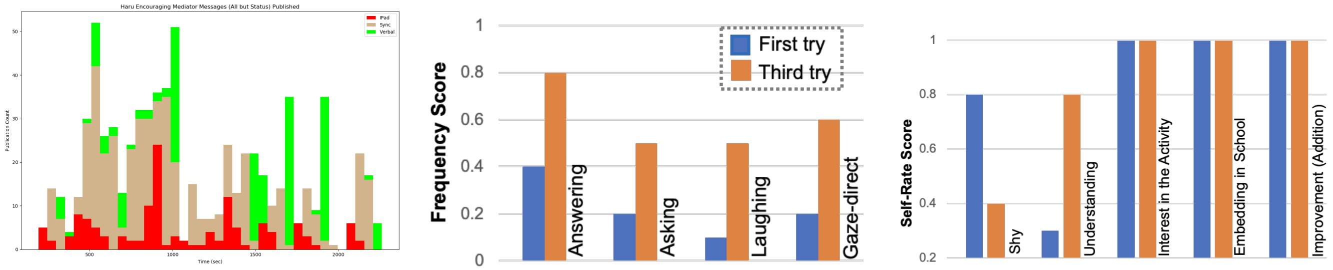
- In-class turn-taking is needed when interacting with the children in the same classroom as the robot, to facilitate exploration and discussion and to foster the sharing of experiences. This turn-taking is, thus, mainly based on speech detection, so that the robot does not interrupt the children when they are talking, and/or can intervene using ice-breakers to encourage equitable participation. The ASR system publishes real-time data on whether speech is detected in the audio stream or not. This functionality is achieved by continuously serving real-time events in ROS. The BT will receive continuous feedback on two event types - Speaker Starts Speaking and Speaker Stops Speaking - for each child.
- Between classes turn-taking and synchronization is required so that the two classrooms at the extremes of the conversation are aligned in the interaction session. We have added a synchronization mechanism to the Behavior Trees of each local robot. This synchronization mechanism is based on messages shared between the two locations and allows each robot to wait for its peer in different stages of the interaction.

V. Pilot Test Results

A. System deployment

The embodied mediator for remote cross-cultural communication has been tested in a pilot between students of the Jiyugaoka High School, in Tokyo, Japan, and the Tempe High School in Sydney, Australia (see Fig. 3). The tests involved, in each of the two locations, 3 children, one Haru robot, two screens for video-conference and audiovisual content, and one iPad for text and audio input for each child. The implemented interaction session follows the design in Table I. The accompanying video shows a summary of the pilot, which consists of the following phases:

1) **Welcome Stage**: The robot in each classroom initiates a welcome stage in which it introduces itself and plays an introduction video in the background.
2) **Calling Stage**: The robot proceeds to call the remote end using animation and video. The two robots synchronize.
3) **Meeting Stage**: The videoconference begins. The robots guide the turns for the children to introduce themselves to those at the remote location.
4) **Privacy Stage**: The robot addresses privacy concerns by informing all attendees about the information collected and how it is handled.
5) **Avatar Creation**: The robot provides instructions on creating personal avatars. Children are prompted to configure their avatars on iPads, and once completed, this information is included in the perception system.
6) **Photo Sharing Guide Explanation**: The robot explains the purpose of the meeting. Participants are informed that they will upload pictures to the system, associate these pictures with their digital avatars, and share them with remote participants. Both ends will view the pictures uploaded by each participant and engage in discussions about them, using hashtags to categorize the content. For example, if a participant uploads a picture



(a) (b) (c)

Fig. 4: Results from pilot tests. a) Interaction (tablet and verbal) and synchronization ROS messages for one session. b) Qualitative results from the study. Engagement evolution between the first and third trial. c) Shyness, understanding, and interest evolution between the first and third trials.

of themselves playing tennis, they tag it with #sports.

7) **Haru's Content Stage**: The robots from both ends demonstrate how to share pictures and initiate discussions by sharing their own pictures and encouraging participants to comment on them. Each end engages in dialogue with a random child from the other end, fostering discussions related to the pictures displayed.
8) **Shared Content Stage**: The robots guide an interactive session where all uploaded pictures are displayed one by one. Discussions are initiated with the individuals who uploaded each picture, delving into the content of the images and related topics. This process occurs in turns for each participant, first at one end and then at the other.
9) **Farewell Stage**: The robot concludes the session by showcasing all the pictures on the screen and reflecting on the overall experience. Following this, it guides a farewell to the remote end of the demo and disconnects the teleconference. Finally, it expresses gratitude to every participant for their involvement in the experience and formally concludes the session locally.

B. Quantitative aspects

The Behavior Tree that orchestrates all the system modules to implement the interaction described above contains $\approx$ 5K nodes and 150K transitions. These include more than 780 interaction commands through ROS for the different interactive elements (robot expressivity, tablet, dialogue, Unity app) local and remote (see Fig. 4a).

The conversational system includes 8 question events, related to the topics of the shared photos, with 15 follow-up questions. Each follow up question has 3 different branches depending on the answer of the children. Therefore, there are 53 unique reaction intents depending on how a speaker answers them.

The interaction contains also open-ended conversations between the children about the topics of the photos, in which the robot does not process the speech input, but employs speech detection for turn-taking. The interaction involves turn-taking for more than 70 x 2 verbal (dialogue) and tablet-mediated interactions. Furthermore, there are more than 150 synchronization points between the robots in the two classrooms.

C. Qualitative results

We conducted three trial interaction sessions for a span of two months consisting of 12 children in total (3 children per group). During the interaction, we measured their engagement by measuring the frequency (normalized score) of them answering and asking questions, laughing, and directing their gazes to other participants including the robot that initiated the interaction. The result as shown in Fig. 4 (b) shows a progression of improved engagement from the first trial to the third trial. We also asked the participants about their shyness level and understanding of the interaction session as shown in Fig. 4 (c) which shows improvements from the first trial. Moreover we also asked about their interest in the said activity and whether they see benefits for the embedding in their classrooms. Lastly the participants were unanimous of the need for the system to be further improved and provided their suggestions for future iterations.

VI. Conclusions

We have presented the first prototype of a robotic mediation framework for cross-cultural communication and interaction. To ensure a holistic design that reflects the pedagogy of cross-cultural learning, we conducted a participatory study that involved an interdisciplinary team of experts. We have conducted the preliminary pilot study with three trials among school children in Australia and Japan. This preliminary study shows promising results in terms of children's impression of the system. Currently, the study is very limited in which we selected school children who speak the English language. Moreover, the user study is also currently limited just focused on the viability and future potential of the proposed system. In the future, we are conducting a longitudinal study for multiple interactive sessions that span for months. In the planned longitudinal study we will focus on the development of actual content to embed the proposed system into school children's curriculum.

Acknowledgment

We are thankful to our multidisciplinary partners who helped us in this project: Vicky Charisi (European Commision-JRC), Steven Vosloo (UNICEF), Tomoko Imai (Jiyugaoka H.S. Tokyo) and the learning institutions involved in both the participatory and pilot studies [35].

References

[1] K. P. Virginia Dignum, Melanie Penagos and S. Vosloo, "Policy guidance on AI for children ," https://www.unicef.org/globalinsight/reports/policy-guidance-ai-children#:~:text=The%20policy%20guidance%20explores%20AI,inclusion%20of%20and%20for%20children, 2021, [Online; accessed 7-Sept-2023].

[2] H. Li, D. Yang, and Y. Shiota, "Exploring the possibility of using a humanoid robot as a tutor and oral test proctor in chinese as a foreign language," in *Expanding Global Horizons Through Technology Enhanced Language Learning*. Springer, 2021, pp. 113–129.

[3] T. Belpaeme, J. Kennedy, A. Ramachandran, B. Scassellati, and F. Tanaka, "Social robots for education: A review," *Science Robotics*, vol. 3, p. eaat5954, 08 2018.

[4] N. Azizi, S. Chandra, M. Gray, J. Fane, M. Sager, and K. Dautenhahn, "User evaluation of social robots as a tool in one-to-one instructional settings for students with learning disabilities," in *International Conference on Social Robotics*. Springer, 2022, pp. 146–159.

[5] V. Charisi, E. Gomez, G. Mier, L. Merino, and R. Gomez, "Child-robot collaborative problem-solving and the importance of child's voluntary interaction: a developmental perspective," *Frontiers in Robotics and AI*, vol. 7, p. 15, 2020.

[6] V. Charisi, L. Merino, M. Escobar, F. Caballero, R. Gomez, and E. Gómez, "The effects of robot cognitive reliability and social positioning on child-robot team dynamics," in *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2021, pp. 9439–9445.

[7] E. Ackerman, "Meet Moxie, a Social Robot That Helps Kids With Social-Emotional Learning," https://spectrum.ieee.org/moxie-a-social-robot-for-childhood-development, 2020, [Online; accessed 7-Sept-2023].

[8] N. Hurst, C. Clabaugh, R. Baynes, J. Cohn, D. Mitroff, and S. Scherer, "Social and emotional skills training with embodied moxie," *arXiv preprint arXiv:2004.12962*, 2020.

[9] B. Scassellati, H. Admoni, and M. Matarić, "Robots for use in autism research," *Annual review of biomedical engineering*, vol. 14, pp. 275–294, 2012.

[10] J. J. Diehl, L. M. Schmitt, M. Villano, and C. R. Crowell, "The clinical use of robots for individuals with autism spectrum disorders: A critical review," *Research in autism spectrum disorders*, vol. 6, no. 1, pp. 249–262, 2012.

[11] R. Gomez, D. Szapiro, K. Galindo, and K. Nakamura, "Haru: Hardware design of an experimental tabletop robot assistant," in *Proceedings of the 2018 ACM/IEEE international conference on human-robot interaction*, 2018, pp. 233–240.

[12] E. Ackerman, "Haru: An Experimental Social Robot From Honda Research," https://spectrum.ieee.org/haru-an-experimental-social-robot-from-honda-research, 2018, [Online; accessed 7-Sept-2023].

[13] UNICEF, "Honda Research Institute Japan European Commission, Joint Research Centre," https://www.unicef.org/globalinsight/media/2206/file, 2021, [Online; accessed 7-Sept-2023].

[14] "," https://mypersonalrobots.org/partners.

[15] L. A. Zebrowitz and J. M. Montepare, "Social psychological face perception: Why appearance matters," *Social and personality psychology compass*, vol. 2, no. 3, pp. 1497–1517, 2008.

[16] K. Crane-Deklerk, "Multimodality in early childhood education," *International Journal of Literacy, Culture, and Language Education*, vol. 1, pp. 73–87, 2020.

[17] P. Fritze, C. Johnston, and R. Kemm, "Structured open-ended learning activities using client-side interactive objects," *Flexibility: The next wave*, pp. 267–277, 1998.

[18] E. J. Theobald, S. L. Eddy, D. Z. Grunspan, B. L. Wiggins, and A. J. Crowe, "Student perception of group dynamics predicts individual performance: Comfort and equity matter," *PloS one*, vol. 12, no. 7, p. e0181336, 2017.

[19] P. Ginns, "Meta-analysis of the modality effect," *Learning and instruction*, vol. 15, no. 4, pp. 313–331, 2005.

[20] D. Wenmoth, M. Jones, and J. DiMartino, "Agency by design: Making learning engaging." *Aurora Institute*, 2021.

[21] B. Parekh, "The future of multi-ethnic britain: Reporting on a report," 2001.

[22] J. Roache and R. Lewis, "Teachers' views on the impact of classroom management on student responsibility," *Australian journal of education*, vol. 55, no. 2, pp. 132–146, 2011.

[23] M. B. Hargis, M. C. Whatley, and A. D. Castel, "Remembering proper names as a potential exception to the better-than-average effect in younger and older adults." *Psychology and Aging*, vol. 35, no. 4, p. 497, 2020.

[24] D. Pepler and K. Bierman, "With a little help from my friends: The importance of peer relationships for social-emotional development," *Edna Bennett Pierce Prevention Research Center: University Park, PA, USA*, 2018.

[25] C. E. Shenk and A. E. Fruzzetti, "Parental validating and invalidating responses and adolescent psychological functioning: An observational study," *The Family Journal*, vol. 22, no. 1, pp. 43–48, 2014.

[26] M. T. C. Team, "Ice Breakers-Communication Skills," https://prime.mindtools.com/pages/article/virtual-ice-breakers.htm, 2011, [Online; accessed 15-Sept-2023].

[27] J. F. Kilanowski, "Breaking the ice: a pre-intervention strategy to engage research participants," *Journal of Pediatric Health Care*, vol. 26, no. 3, pp. 209–212, 2012.

[28] J. Cummins, "Negotiating intercultural identities in the multilingual classroom," *Catesol Journal*, vol. 12, no. 1, pp. 163–178, 2000.

[29] C. Ward and A. Rana-Deuba, "Acculturation and adaptation revisited," *Journal of cross-cultural psychology*, vol. 30, no. 4, pp. 422–442, 1999.

[30] S. May, "Critical multiculturalism and education," *The Routledge International Companion to Multicultural Education*, pp. 33–48, 01 2009.

[31] R. Ragel, R. Rey, Á. Páez, J. Ponce, K. Nakamura, F. Caballero, L. Merino, and R. Gómez, "Multi-modal data fusion for people perception in the social robot haru," in *International Conference on Social Robotics*. Springer, 2022, pp. 174–187.

[32] R. Gomez, D. Szapiro, L. Merino, and K. Nakamura, "A holistic approach in designing tabletop robot's expressivity," in *2020 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2020, pp. 1970–1976.

[33] M. Colledanchise and P. Ogren, "How behavior trees modularize hybrid control systems and generalize sequential behavior compositions, the subsumption architecture, and decision trees," *IEEE Transactions on Robotics*, vol. 33, no. 2, p. 372–389, April 2017.

[34] C. Paxton, A. Hundt, F. Jonathan, K. Guerin, and G. D. Hager, "Costar: Instructing collaborative robots with behavior trees and vision." in *IEEE International Conference on Robotics and Automation (ICRA).*, 2017, p. 564–571.

[35] "," https://mypersonalrobots.org/pilot.