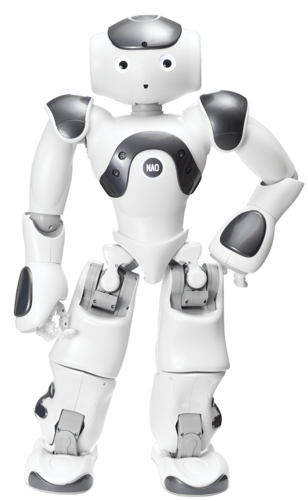
Final Chapters

Human-Robot Interaction

Radboud University

Motivating Movement in Children Through an Interactive Dancing Robot

Group 3



February 13, 2025

Contents

1 Prototype Development **2**
1.1 Initial Development Phase 2
1.2 Main Development Phase 3
1.2.1 Pose detection 3
1.2.2 Basic interaction 4
1.2.3 Main scenario 4
1.3 Refinement Phase 6

2 Evaluation **8**
2.1 Experiment Setup 8
2.2 Results 10
2.3 Discussion 11

3 Conclusion **15**

A Demo Video **19**

B Group **20**

C Ethical Application and Approval **21**
C.1 Application Form 21
C.2 Approval 22

D Participant Demographics **23**

E Questionnaire Results **25**

F Observation Phrases **26**

Chapter 1

Prototype Development

1.1 Initial Development Phase

The development process started with exploring Nao dance moves available online. After testing several dance codes, we selected three dances from a repository implementing a number of dancing commands[1]: the dab, air guitar and sprinkler. These dances were chosen for their simplicity and short duration, as well as relevance among the younger generations. The dances from the repository were developed and tested on a virtual robot. When tested on the physical Nao robot, we observed that the dances were performed faster than the robot could physically handle without tumbling over. To address this, we implemented a function to change the movement to a lower speed.

One large challenge we faced already early on was the fact that Nao runs on Python 2.7 – an outdated Python version. Since this project required packages available only for Python 3, the Nao bridge by Jan de Wit[2] was used. This bridge made it possible to write code in Python 3, and convert it to Python 2.7 during runtime. However, the bridge only allowed one-way communication: information could be sent from the machine running the code to the robot, but the robot could not send data back to the machine. Another issue was that only two group members were able to successfully install the bridge on their machines for unknown reasons. This however did not cause any issues, as we resolved the resulting unbalance in workload through clever task allocation.

Initially, we planned to use the built-in camera of Nao to take pictures. However, due to the limitations of the Python bridge, we chose to use an external camera placed in front of Nao to

[1] This is a repository containing movement commands for Nao in terms of joint movements over time. It can be found here.

[2] The repository containing the bridge can be found here.

capture the pictures instead. The same limitation applied to the microphone, requiring us to use an external one.

1.2 Main Development Phase

1.2.1 Pose detection

To give feedback on the dance performance of the user, we implemented a pose detection algorithm. We did this with the help of an existing Python library about human pose detection[3], which extracts the relative locations of different points of interest on a person's body. Pictures are taken at different times during the exercise to capture a range of poses and movements. By taking multiple pictures, we allow some leniency in the timing of the user's performance. We extract points of interest from these pictures (see Fig. 1.1) and compare them with the locations of the same points of interest of reference images captured beforehand for each exercise, showing someone correctly mimicking the robot.

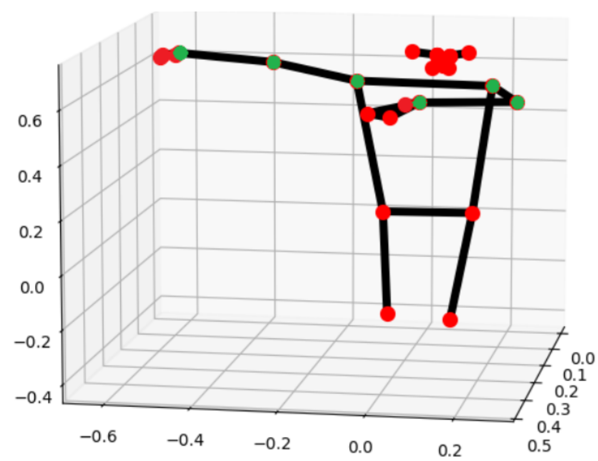


Figure 1.1: Example of a position skeleton extracted from a picture of someone dabbing. The green dots show which points we keep track of and use for the error calculation.

The comparison focuses on the angles of the upper and lower arms relative to the ground, the angle between the upper and lower arm, as well as the relative 2D coordinates of key points. To keep the complexity low, we only analyzed the positions of the arms and shoulders, as the other body parts were not very relevant for the chosen dance moves. Additionally, we limited our comparison to 2D coordinates instead of 3D to simplify the implementation. This analysis is performed for all captured images, both in their original orientation and as horizontally flipped versions, since the absolute direction of the exercise was not important. We then calculate the

[3]The repository of the human pose detection and extraction of points of interest can be found here.

mean differences of these points and angles as compared to the reference image and provide feedback based on the best-performing image. Feedback emphasizes the worst outlier, with Nao encouraging the user to pay extra attention to the location and pose of that specific body part (e.g., "left forearm").

1.2.2 Basic interaction

To make the robot robust and remove the need for the researcher's involvement in starting the interaction, we decided to implement basic motion detection. This way, Nao would be able to detect when someone walks into the frame and then start the scenario. This also enabled us to add another layer of interactiveness: whenever the user has left the frame during a session with the robot, Nao can detect that they are no longer in the frame and urge them to come back, encouraging them to continue dancing. The motion detection was realized by subtracting two subsequent frames from each other and checking whether the observed change was above a specified threshold. A key feature of the human-robot interaction is the robot's ability to understand the user's verbal prompts. For speech recognition, OpenAI's Whisper was used. After testing various speech recognition models, the tiny Whisper model was chosen due to its short processing time and ability to run locally. To ensure a smooth interaction, a phrase time limit of 3 seconds was set, meaning that the model begins processing the audio after recording for 3 seconds. To make the interaction intuitive and ensure the user only speaks whenever the robot is actively listening, a signaling beep sound is played before and after recording.

1.2.3 Main scenario

Once the aforementioned functionalities were implemented, they were integrated together into a main class of the overall robot behaviour. The general scenario structure was based on having two options for the user: learning a dance move and dancing together.

Firstly, whenever motion is detected, Nao is to give an introduction and ask for the user's name – which it will then use in a greeting. In order to handle a wide variety of answer structures (e.g., "I'm name", "My name is name", "name", etc.), the name was extracted from the user input utilizing a regular expression. Next, the robot is implemented to let the user choose whether the session will be about learning a dance move or just dancing together in free-style. The user's transcribed reply is passed through a set of if statements so that the corresponding and expected behaviour of Nao is triggered. This structure was put inside a while loop, so that the user could perform multiple dancing and/or learning cycles if desired. Starting from the second cycle, the option to stop the scenario is also provided by Nao. When the interpreted speech input matches none of the words in the if statements, the `misunderstand` variable is set to True, and Nao requests the user to repeat their prompt.

If the user requests to **learn a dance move**, Nao would first ask which of the available three dance moves they would like to see. This structure was similarly put within a while loop in case none of the words in the input matched the words in the if statements. Furthermore, in order to maximize natural interaction, the case wherein a user would request another dance move apart from the available three was handled separately. The main part of the `teach_move` method is contained within a while loop. This loop is terminated only when the user has performed the corresponding dance move correctly twice, in which case they are considered to have successfully learned it. If the performed dance was not correct based on the human pose error calculation, the corresponding feedback as described previously is provided and the cycle repeats until it was performed correctly twice. For increased interactivity, Nao would ask every 2 cycles whether the user wants to continue learning, or do something else instead (i.e., go back to the main scenario loop). Furthermore, at the end of each cycle, motion detection is utilized in order to check whether the participant is still there. If not, Nao would ask the user to come back and keep the motion detection algorithm running for another 10 seconds. If the user comes back into the frame, Nao would welcome them and continue. Otherwise, Nao would appropriately end the scenario with a sad tone.

If the user requested to **dance together**, Nao would perform the three dances sequentially, repeating each dance twice. During each dance, the robot also checks if the user has also performed that dance. This is achieved by continuously taking pictures in a while loop, and applying pose detection for evaluation. If Nao detects that a successful dance move has been performed by the user, it would compliment them. The error threshold is set to a lower value for the free-style dancing scenario than for the learning scenario.

During each dance performance by Nao or the user, music is played. The music fades in at the start and fades out at the end of the performance. While the robot and user are dancing together, the song plays continuously throughout the whole performance. The song used is an AI-generated track created with Suno[4], featuring a funky and danceable style.

A main challenge we faced in the implementation of the robot's behaviour, was that whenever Nao speaks or performs a dance move, the subsequent code would be executed immediately without waiting for Nao to finish. This led to numerous issues, such as the robot demonstrating a dance move and already, while doing that, taking pictures of the user to assess their performance. There is a built-in function that makes the code execution wait until Nao is done moving or talking, but it could not be utilized since the Python bridge did not allow for information retrieval (which was necessary for this method). Thus, a creative solution was developed. For the dancing, a `time.sleep(x)` was added after every location in the code where Nao executes a dance move, where `x` is the measured time required for Nao to fully execute the move. Speech necessitated a less trivial solution, as the time required for speech varied with every sentence. Thus, a method was created that would estimate the amount of time it takes for Nao to speak a given sentence based on its talking speed (words per minute). The words per minute variable was approximated through trial and error. A limitation of this approach is that it does not work exactly equally well for all possible sentences, as it depends on the length of words present in

[4]Suno is a generative AI music production program that can be found here.

the sentence. Nevertheless, these variations were in the order of deciseconds, implying that its impact on the general performance of the prototype is negligible.

Lastly, a slightly altered non-interactive version of the robot's behaviour was implemented for the video demonstration needed to answer our research question. Since no interaction would be possible, it was decided to simply execute all available options consecutively (i.e., learn each dance move, followed by dancing together). If-statements were placed at appropriate places in the behaviour class, such that the robot only asks for user input and performs speech recognition only when the global variable `INTERACTIVE` is set to True.

1.3 Refinement Phase

Once the overall integrated scenario structure was completed, it was time for refinement of several aspects. Firstly, for each sentence that Nao would speak (excluding very simplistic ones with few words), semantically identical alternative sentences were created with slight variation in phrasing and vocabulary use in order to make Nao seem more natural and less generic – an aspect most important and noticeable with continued use of the robot. These alternative sentences were obtained by prompting ChatGPT, and subsequently stored in lists. Note that ChatGPT was thus not utilized live during the experiments. Furthermore, the words per minute variable for estimating the speech time of Nao was adjusted repetitively through trial and error in order to find the ideal value for which the scenario seemed the most fluent and natural.

To enhance speech recognition capabilities, we explored using Rasa[5] to determine the intent behind users' responses. This approach aimed to move beyond simple keyword detection, enabling more accurate understanding of user intentions. By focusing on intent rather than specific words, we sought to avoid misinterpretations caused, for example by negations. As relying solely on detecting keywords like "dance together" could mistakenly interpret the response "I do not want to dance together" as a positive request to dance. However, it was discovered throughout prototype testing, that the utilized Whisper model for speech recognition had the tendency to hallucinate intermittently. This considerably decreased the fluency of the overall scenario, as Nao would repeatedly ask for the answer to be repeated until (one of the) desired word(s) were accurately interpreted. For this reason, in a few if-statements that check the user input, multiple phonetic neighbours of the desired word were added. For instance, if the user wants to learn the air guitar, the code checks for 'ergator' and 'eric' apart from 'air' and 'guitar', as the former words were found to be common in Whisper's interpretation when someone says 'air guitar.' Thus, these phonetic neighbours were acquired by observing Whisper's output throughout prototype testing. Despite these attempts to overcome the limitations of the model's performance, however, we were still not able to get Rasa to work well enough, so it was decided to stick to pattern matching words in the user's input using if-statements.

[5]Rasa is software utilising AI that enables semantic analysis of user prompts. It can be found here.

Lastly, an "emergency" stop button was implemented that we could press at any time throughout the experiments in case time was running out (due to tight scheduling). This button prompted Nao to accordingly explain time was up and ensured an appropriate ending to the scenario in such cases.

Chapter 2

Evaluation

2.1 Experiment Setup

Study design. We aim to assess the influence of the robot dance tutor's interactivity on the motivation for physical activity. Our main hypothesis is that the interaction between the child and the robot will be the biggest driving factor for a positive effect on the desire to move, and thus we devise two study conditions. In the first one, the participants get to interact with the robot in the same room, while in the second setting, the participants are only shown a video of the pre-recorded dance routine of the robot. The overall procedure is the same for both conditions: The participant reads the information form and signs the consent form[1]. Relevant participant information is collected, to be used in the analysis stage to attempt to identify possible correlations between the participant variables and results. Then the actual experiment with the robot is carried out, and finally the participants are asked to fill in a questionnaire to evaluate the experience. Throughout, one of the researchers is present in the room to make sure all is going smoothly and to take notes of the participants' body language and remarks in a non-invasive way.

The main experiment itself is what differs between the two participant groups. In both cases the robot follows a script, but in the interactive condition the direction of the interaction is determined by the participant: they can choose what type of dance to learn, whether to learn at all or just dance, when to stop, etc. The robot also evaluates the participants' dance moves and provides feedback, and the experience is made more personal by the robot asking and remembering the participant's name. The scenario used in the interactive experiment is as follows: Nao greets the participant and asks their name, offers them the choice between dancing freely or learning a dance, offers them a choice between different dances if they chose learning,

[1] The consent form can be found here and the assessment form, containing the initial information and questionnaire, can be found here .

shows them how to execute the dance, records their attempt and provides feedback on it. Occasionally, the robot asks the participant if they wish to stop or change activity. During the dance together mode, Nao plays music and recognizes and compliments the moves it taught the participant if they are performed. The scenario used in the non-interactive setting is as follows: the robot greets the participant, shows them the three dance moves one after the other, giving the participant time to repeat them, and finally plays music for them to dance together.

Participants. The target audience is individuals below the age of eighteen. However, due to ethical constraints, it was decided that instead of conducting our planned study with children, we will recruit adult participants. We attempt to involve healthy subjects with a wide breadth of activity levels and dancing proclivity. Because of restricted time and space availability, we aim for about N=20 participants in total, such that there are 10 per each studied group.

Evaluation. As mentioned, there are two moments at which we collect data from the participant. First, a pre-interaction questionnaire gathers basic demographic information, namely age, frequency of physical activity, and dancing habits. A summary of the participant demographics is included in Appendix D. Then, a post-interaction questionnaire assesses various dimensions of the interaction:

- **Interest and Enjoyment:** Explores enjoyment and interest in dancing with Nao through statements like "I enjoyed dancing with Nao."
- **Perceived Competence:** Measures participants' confidence and perceived improvement, e.g., "I felt I was able to follow Nao's dance instructions well."
- **Robot's Social Presence:** Evaluates the robot's interactivity, movements, and clarity of instructions with items such as "Nao was responsive to my actions."
- **Safety and Comfort:** Focuses on participants' comfort and safety during the interaction, with statements like "I felt safe while interacting with Nao."
- **Effort and Value:** Gauged participants' effort and perception of the activity's value through items like "I think dancing with Nao is a good way to be more active."

The questions are scored on a 7-point Likert-type scale (1 = strongly disagree, 7 = strongly agree), as it has been shown to provide optimal reliability and discriminating power [3] and is consistent with similar research in the field [4]. The full questionnaire can be found in the previously linked assessment form, and the self-reported scores from it are used to compare the interactive and non-interactive setting.

When it comes to the researcher notes on participant behaviour and actions, we use text mining techniques to extract the most commonly occurring n-grams and create a maximum likelihood estimate for their probability distribution. That can be easily visualised in a word cloud format, to highlight the most prevalent themes within the text. This is done separately for the two groups of participants. The participants' remarks during the experiment were analysed separately. We group them into themes that can be used as feedback for future iterations of development.

2.2 Results

An overview of the questionnaire results are presented in Fig. 2.1. The scores seem to mostly agree between the two study groups, with only three question scores appearing different. For Q9: "Nao was responsive to my actions," and Q10: "I felt Nao understood how well I was doing," the participants in the interactive scenario scored higher: on average 5.5 vs 3.0 for Q9 and 4.3 vs 3.0 for Q10. Conversely, for Q12: "I felt comfortable dancing with Nao," the participants in the interactive scenario scored lower: on average 4.4 vs 5.9. Significance testing confirms the difference in results for Q9 and Q12, resulting in p-values of 0.004 and 0.007 respectively.

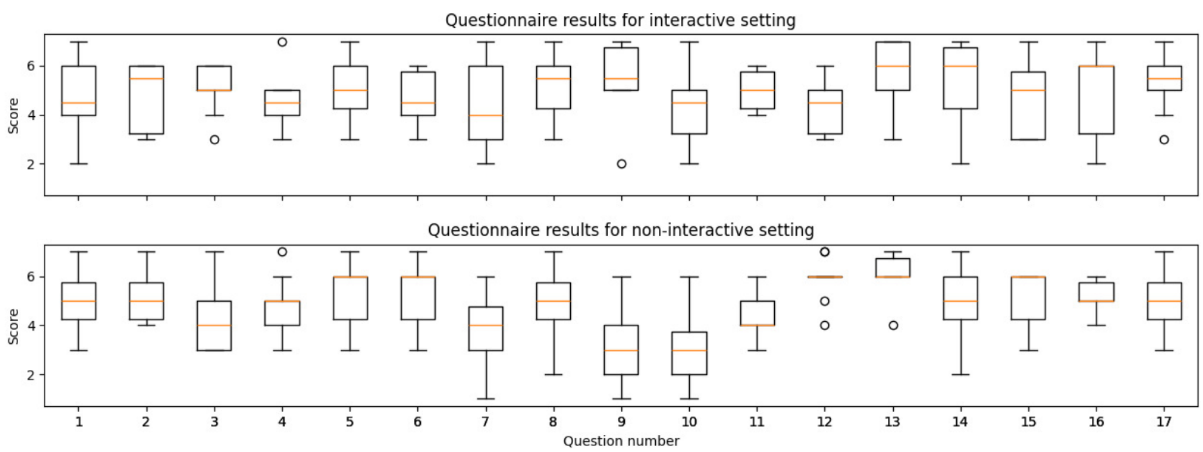


Figure 2.1: Summary of results. Note that 1 = strongly disagree and 7 = strongly agree. The question definitions and the average scores per question (overall and per setting) are included in Appendix E.

When it comes to the collected notes on the participants' behaviour and body language, distinct word clouds for the two settings were generated. They are included in Fig. 2.2. As can be observed, there are some phrases occurring in both conditions, but the most common ones are rather different: "laughing a lot," "doesn't understand," "unprompted reactions" and "actively dancing" for the interactive setting and "excitement dwindling," "forgets dance," "confused instructions," and "halfhearted attempts" for the non-interactive setting.

Figure 2.2: The most common 2-grams of the researcher notes for the interactive (left) and non-interactive (right) setting.

Finally, the grouped and themed verbal feedback of the participants is presented in Table 2.1.

Category		Occurrence
Robot movement	Too slow	6
Teaching dances	Doesn't demonstrate alongside participant	3
	Dance moves are easy to learn so in a next iteration, they would like to see other ones	1
Interaction	When moves are executed well, there is less feedback / interaction	1
Robot appearance	Unsettling because of tracking participant with its head	1
	Endearing because of its natural movements and voice	1

Table 2.1: Thematic analysis from the participants' verbal feedback during the experiment.

When it comes to the participant demographics, there is no significant difference in survey responses across groups categorized by how often individuals dance and their frequency of physical activity. Therefore, the obtained and reported results are not influenced by these participant variables and display truthfully the difference (or lack thereof) between the interactive and non-interactive study conditions.

2.3 Discussion

Study results. The questionnaire results evaluating the participants' enjoyment are mostly positive, at 4.7 and 4.95 out of 7. The gauged value and effort was evaluated at 4.85 and 5.1 out of 7, with the majority of participants willing to recommend dancing with Nao to others (5.2/7). The results of these questions are similar across the two conditions, indicating that no matter whether the robot is physically present in the room, the participants found the process pleasant and think this approach can be useful for tackling low activity levels.

Concerning the perceived competence and safety, results are once again mainly in agreement. The improvement in dancing is evaluated only at 4.05/7, but we believe this is not that important, given that our aim is to encourage movement and not specifically to hone dancing skills.

The participants could follow the instructions well (5.25/7), and felt confident (5/7) and safe (5.9/7) while dancing. This is encouraging given that our target audience is children, and it is desirable for them to find it easy and safe to use and interact with the dancing tutor. One of the significantly different results between the two study conditions is in perceived comfort with the robot, for which participants from the interactive group scored 1.5 points lower. Both results are satisfactory, but it is clear that the participants that interacted with the robot in person felt less comfortable. This could be due to the fact a camera was pointed at them throughout the experiment in that condition, which could introduce some unease[2]. Furthermore, the presence of all group members (although invisible for the participant) could have introduced additional unease. Moreover, one participant called the robot "creepy" due to the way it rotates its head to follow the person in front of it, possibly being another explanation of this result.

Finally, the questions regarding the robot's social presence indicate that participants agree the robot's movements are mostly easy to follow (5/7) and its instructions were mainly clear (4.65/7). There is disagreement between the two conditions about Nao's understanding of the participants' performance (4.3/7 in the interactive case and 3/7 in the non-interactive one) but not a significant one. There is, however, a significant difference in the perception of the robot's responsiveness, where participants in the interactive scenario scored 2.5 higher on average. This confirms the basic expectation that the designed interactive version of the study was indeed perceived as such compared to the non-interactive version.

Analysis of participant body language and behaviour. Based on the presented word clouds, it seems that the participants in the interactive setting had an overall positive experience and actively danced. On the other hand, for the video setting, a more negative experience is highlighted, with the participants' excitement diminishing over time and their effort into dancing seeming lower than the other participants'. However, these results are not supported by the self-reported scores. Questions Q1, Q2 and Q15, which reflect on the participants' enjoyment, level of fun and effort put into dancing, were not found to have significant differences between the two conditions. On the other hand, the phrase "unprompted reactions" only occurs in the word cloud for the interactive setting. It refers to e.g., the participants thanking the robot or waving goodbye – interactions that Nao has not initiated, but the participants do anyway, and this can be a direct result of the robot being seen as more responsive for that condition (Q9).

This analysis, however, is based on the notes taken by the executive researcher in the room, which can be seen as biased. In order to make the process as truthful to reality as possible and avoid the ambiguities and nuances that natural language inherently possesses, a pre-compiled list of behavior-describing phrases was created that the researcher could ascribe to the participant. The list is included in Appendix F. This drawback of the analysis could be why such a stark difference between the two conditions based on our notes is observed, but not based on participant questionnaire scores. In future work, a more sophisticated approach of analyzing the participants' body language can be deployed, such as video analysis [2].

[2]However, we did explicitly inform the participants that no recordings were made.

Thematic analysis of participants' feedback. The most common remark was about the robot's slow movements. This is partly unavoidable due to Nao's physical limitations, but there might be some room for improvement, especially since the robot would display the move twice with a pause in between. A large number of participants expressed a desire for Nao to perform the dance move alongside them. We believe this is valuable feedback that should be considered in future iterations of the prototype design, especially keeping in mind children's innate urge to copy what they are seeing. One participant thought that the robot feels less interactive when they perform dances well due to the briefer feedback; this is another point to be considered in the future. Furthermore, the robot's appearance was mentioned in contradicting opinions, which highlights the need to keep the user involved throughout the whole design process. Finally, the request for more dance moves in subsequent lessons is addressed in the next point of discussion.

Dances and novelty. We have developed a proof of concept prototype with a very limited skill set in terms of dancing. All dance moves were rather simple and only included movement with the arms. They were chosen with the intention to appeal to the younger audience, and we believe that was a good choice that we should continue following in later iterations with the addition of more dances. Adding more dances is a necessity, we believe, and one of the relatively effortless ways to keep the robot's abilities interesting and fresh.

We believe the novelty aspect of our prototype should be treated as a key feature for success. Some participants seemed excited because of Nao's ability to speak and move naturally and skill to dance, as it was something they had not seen before, and this amazement can be a driving factor in motivating interaction and movement. Unfortunately, due to time limitations only a single session with the robot was conducted, meaning we cannot gauge the effect on participants' desire to be active over time. It is possible that with more exposure to the robot and the three moves it can perform and teach, the excitement of the participants and their willingness to participate in dance tutorials might diminish. This issue is raised by previous literature on robot dance teachers, which reports declining participation over time due to decreased novelty and task repetitiveness [5]. As mentioned, expanding the dancing repertoire with longer and more elaborate dance moves, is the easiest way to keep the dancing tutor engaging. Furthermore, we initially considered enabling the robot to see a move made by the user and copy it – reversing the teacher-student roles. This was not implemented due to time constraints, but we believe it can add to the robot's innovativeness and open the possibility for an unlimited amount of unique interactions.

We are of the opinion that with a larger-scale study spanning multiple sessions, being able to interact with the robot and choose what happens next would result in more entertaining and unique experiences which would be impossible to replicate with a non-interactive video. We hypothesize that in that case, a bigger difference in results between the two study conditions will be visible, and believe this is something worth investigating further.

Participant age range. A limitation in the evaluation method is the discrepancy between the participants' age and the age of the target group. Due to privacy and ethical constraints, adults were used as a proxy to children. Although they were instructed to think like children, it is likely that the achieved results do not directly translate to children as well. Given the positive reception of the product, however, we believe it is worthwhile to conduct the study with the actual target group in future research.

Project goal. The results are inconclusive, with the questionnaire and the observational notes being in disagreement. Higher importance is placed on the self-reported scores, given the possible bias in the researcher notes. Based on the scores, the initial hypothesis is rejected: the participants in both conditions find the robot equally motivating. This could be due to any of the aforementioned study limitations. We believe this to be most likely due to the single-session setup, in which the robot even on a video is novel and entertaining and thus receives high satisfaction scores.

Chapter 3

Conclusion

This project set out to investigate how human-robot interaction could influence children's motivation for physical activity through an interactive dancing experience. With 81% of children being insufficiently active globally [1], we aimed to explore whether a robotic dance tutor could provide an engaging and accessible solution to promote physical activity. Specifically, we sought to understand if and how the interactive presence of a Nao robot would affect engagement compared to non-interactive video demonstrations of the same robot.

The prototype development and subsequent evaluation study yielded several interesting insights. While participants perceived the interactive robot as significantly more responsive than its video counterpart, there were no significant differences in self-reported enjoyment and effort between the two conditions. The observational notes indicated more positive engagement in the interactive condition, though these findings should be interpreted cautiously due to potential researcher bias. Since more importance is placed on the self-reported scores, the results suggest that the hypothesis does not hold.

Limitations and Learnings. Several limitations of our study provide important lessons for future work in this area:

- Participant Demographics: Due to ethical and practical constraints, the study was conducted with adult participants rather than the target audience of children. While this provided valuable initial insights, the results may not directly translate to younger users.
- Sample Size and Duration: The relatively small sample size and single-session format limited our ability to draw robust conclusions about long-term engagement and motivation differences between interactive and non-interactive approaches.
- Technical Constraints: Significant time was spent addressing voice recognition challenges, which could have been better allocated to improving the robot's movement fluidity and

conducting more user testing sessions.

- Data Collection: The absence of an open-ended feedback question in the questionnaire limited the ability to gather qualitative insights about participants' experiences.

Future Directions. Looking ahead, several opportunities exist to build upon this work. First and foremost, future research would benefit from an extended study design that addresses the current limitations. This should include a significantly larger participant pool to enable more robust statistical analysis and, critically, involve children as participants rather than adult proxies. The study duration should be expanded to include multiple sessions per participant, allowing researchers to assess long-term engagement patterns and the sustainability of motivation over time. Additionally, more attention should be paid to matching the interactive and non-interactive conditions more closely to isolate the specific effects of physical presence and interactivity. To capture richer insights about participant experiences, future studies should incorporate comprehensive qualitative feedback collection, including open-ended questions and structured interviews.

Additionally, an important consideration for future work is the need to ensure greater consistency between experimental conditions. In this study, the interactive robot was present in the room, whereas the non-interactive condition relied on video demonstrations of the robot. This difference could have influenced participant's engagement levels and perception of the experiment. To address this, future research could aim to create more comparable conditions, such as having a robot in the room for both situations but with different levels of interaction.

From a technical perspective, several improvements could enhance the robot's capabilities and the overall interaction experience. The dance repertoire could be expanded to include a wider variety of moves, helping maintain participant interest over multiple sessions. An exciting possibility would be implementing bi-directional learning, allowing the robot to learn and mirror moves from participants, which could create a more dynamic and engaging interaction. Future iterations should also focus on improving the fluidity of the robot's movements and creating more natural interactions. Given the limitations we encountered with the Nao platform, particularly regarding integrated sensors, exploring alternative robotic platforms that better support these features could prove beneficial.

The evaluation methodology could also be enhanced in future work. Rather than relying on potentially biased observational notes, more sophisticated approaches to analyzing participant engagement could be employed. For instance, automated video analysis of body language and movement patterns could provide objective measures of engagement and participation. This would not only reduce researcher bias but also enable more detailed analysis of how participants interact with the robot over time.

General Reflections. This project was challenging at times, especially due to the numerous technical issues faced. Nevertheless, it was a rewarding experience that taught us considerably regarding the importance of effectively designing and planning interventions with participants.

This included understanding their needs, anticipating potential barriers, and incorporating feedback iteratively to enhance outcomes... Furthermore, the team cohesion was of a high level, implying an efficient and seamless working experience throughout the project.

Final Reflections. This work has established a reusable framework while highlighting the complexity of using robotic systems to promote physical activity. The challenge extends beyond technical implementation to creating sustainable interventions that maintain user interest over time. Through this project, we have contributed to the field by developing and evaluating a novel paradigm for robot-led dance interaction, while also identifying key considerations for future research in this space. While more work is needed to fully understand the potential of robotic dance tutors for promoting children's physical activity, our findings provide a foundation for future investigations in this promising area.

Bibliography

[1] Guthold, R., Stevens, G. A., Riley, L. M., and Bull, F. C. Global trends in insufficient physical activity among adolescents: a pooled analysis of 298 population-based surveys with 1·6 million participants. *The Lancet. Child & Adolescent Health 4*, 1 (Jan. 2020), 23–35.

[2] Henkemans, O. a. B., Bierman, B. P., Janssen, J., Looije, R., Neerincx, M. A., Van Dooren, M. M., De Vries, J. L., Van Der Burg, G. J., and Huisman, S. D. Design and evaluation of a personal robot playing a self-management education game with children with diabetes type 1. *International Journal of Human-Computer Studies 106* (June 2017), 63–76.

[3] La Guardia, J. G., Ryan, R. M., Couchman, C. E., and Deci, E. L. Within-person variation in security of attachment: A self-determination theory perspective on attachment, need fulfillment, and well-being. *J. Pers. Soc. Psychol. 79*, 3 (2000), 367–384.

[4] Preston, C. C., and Colman, A. M. Optimal number of response categories in rating scales: reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica 104*, 1 (2000), 1–15.

[5] Ros, R., Baroni, I., and Demiris, Y. Adaptive human–robot interaction in sensorimotor task instruction: From human to robot dance tutors. *Robotics and Autonomous Systems 62*, 6 (2014), 707–720.

Appendix A

Demo Video

A demo video of the robot executing the basic scenario can be found here:
https://drive.google.com/file/d/1OgetFl49XwhsWuDw-A2sWDHho_sBBUiz/view?usp=sharing.

Appendix B

Group

Throughout the project, we worked closely as a team, meeting together once a week to collaborate and share updates. While we divided some tasks among us, Luuk and Thomas took the lead in implementing the different scenarios for the robot, as the rest of us encountered difficulties in establishing a connection between the bridge and Nao. Nadezhda focused primarily on creating the personas, working on data analysis and voice recognition, as well as collaborating with Maikel on pose detection. Maikel also worked on the thematic analysis of the feedback and integrating Rasa, working alongside Willem, though we were ultimately unable to utilize it, as explained in the chapter on prototype development. Willem also concentrated on developing the questionnaire questions. Despite our individual focuses, all design decisions were carefully discussed as a team to ensure alignment with our shared vision. The writing of the report chapters too was a joint effort, split equally between all of us. We are quite happy with our collaboration and teamwork!

Appendix C

Ethical Application and Approval

C.1 Application Form

Research Background and Questions
Our research aims to investigate whether human-robot interaction can effectively motivate children to be more physically active through dance. While the ultimate target group is children (as 81% of children aged 11-17 are insufficiently active), this initial study uses university students to evaluate our approach. We compare the effectiveness of a social robot dance teacher (NAO) against a traditional video-based dance instruction method. Our research question is: "How does human-robot interaction influence motivation for movement in children?" Participants will be asked to evaluate the system from a child's perspective.

Participants and Recruitment We will recruit 20 university students (18+) through campus advertising. Participants will be asked to imagine themselves as children while interacting with the system and provide feedback on its potential effectiveness for children. Participation is voluntary and participants can withdraw at any time.

Research Design and Methods The study uses a between-subjects design where participants are randomly assigned to one of two conditions:

1. Interactive session with NAO robot teaching dance moves (5 minutes)
2. Control session watching a dance instruction video (5 minutes)

Before participants start their session, we ask them some short questions about their personal data (see questionnaire in TODO).

After their session, participants complete a brief questionnaire measuring enjoyment, perceived competence, and comfort (7-point scale, 17 items, approximately 5 minutes). Sessions will be video recorded for later analysis of engagement and interaction patterns. Total participation time is approximately 10 minutes.

Impact on Participants
Physical Impact: The dance moves are simple and basic, presenting minimal physical risk. Participants can stop at any time if they feel tired or uncomfortable.
Psychological Impact: The study is designed to be enjoyable and non-stressful. The robot is programmed to be encouraging and non-judgmental. Participants are informed that they should evaluate the system from a child's perspective.
Time Commitment: One 10-minute session scheduled at the participant's convenience.

Privacy Protection
All data will be pseudonymized using participant codes. Video recordings will be stored on encrypted drives and deleted after analysis. Questionnaire data will be stored separately from consent forms Only researchers directly involved in the study will have access to the raw data. Publications will only use aggregated data. Participants can request deletion of their data within one month after participation.

The research involves minimal risk and aims to contribute to understanding how technology can promote physical activity in children through an initial evaluation with adult participants.

C.2 Approval

The experiment was submitted and approved through the "Light Track" by the Ethics Committee of Social Sciences (ECSW) and received supervisory approval from Pieter Wolfert. ECSW Reference ID: ECSW-LT-2024-12-5-53740

Appendix D

Participant Demographics

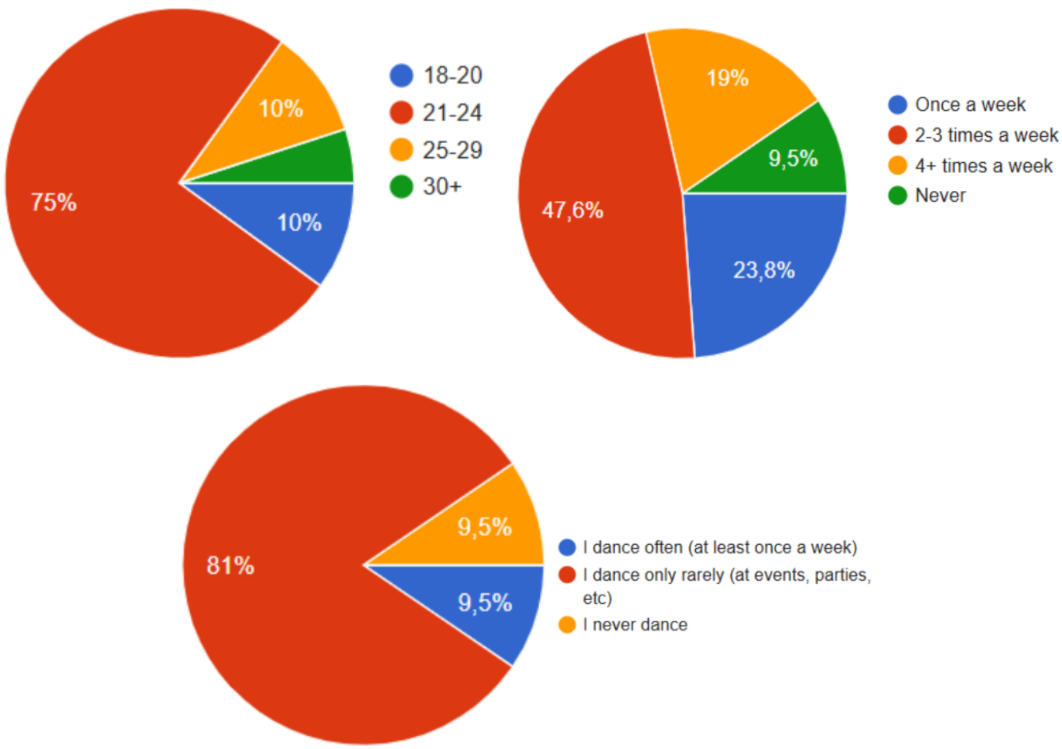


Figure D.1: Visualization of the spread of the participants' age (top left), physical activity level (top right) and dancing habits (bottom).

Appendix E

Questionnaire Results

Q#	Question description	Average score (overall)	Average score (interactive)	Average score (non-interactive)
1	I enjoyed dancing with Nao.	4.95	4.80	5.10
2	Learning dances with Nao was fun.	4.95	4.80	5.10
3	I would like to dance with Nao again.	4.70	5.10	4.30
4	I found the dance activities interesting.	4.60	4.50	4.70
5	I felt I was able to follow Nao's dance instructions well.	5.25	5.10	5.40
6	I felt confident while dancing with Nao.	5.00	4.60	5.40
7	I improved at dancing during the session.	4.05	4.40	3.70
8	Nao's movements were natural and easy to follow.	5.00	5.20	4.80
9	Nao was responsive to my actions.	4.25	5.50	3.00
10	I felt Nao understood how well I was doing.	3.65	4.20	3.00
11	Nao's instructions were clear and helpful.	4.65	5.00	4.30
12	I felt comfortable dancing with Nao.	5.15	4.40	5.90
13	I felt safe while interacting with Nao.	5.90	5.70	6.10
14	Nao's movements were predictable.	5.15	5.30	5.00
15	I put effort into learning the dances.	4.85	4.60	5.10
16	I think dancing with Nao is a good way to be more active.	5.10	5.00	5.20
17	I would recommend dancing with Nao to others.	5.20	5.40	5.00

Table E.1: Overall questionnaire average scores and per setting. Note that 1 = strongly disagree and 7 = strongly agree.

Appendix F

Observation Phrases

- doesn't understand speech
- understands speech
- seems confused about instructions
- finds everything clear
- doesn't understand the dance move
- forgets dance
- asks researcher for guidance
- seems engaged
- enjoying themselves
- laughs a lot
- some laughing
- serious and focused
- annoyed when robot doesn't understand a prompt
- excitement dwindling
- seems awkward
- mimicking robot and actively dancing in dance together mode
- enthusiastic about a dance
- half-hearted attempts
- unprompted reactions to the robot
- smooth interactions
- stunted interactions
- no interaction
- slow robot
- fast robot