Designing Parent-Child-Robot
Triadic Storybook Reading Interaction

by
Soo Jung Jang
B.S. Electrical Engineering and Computer Science,
Massachusetts Institute of Technology (2020)

Submitted to the Department of Electrical Engineering and Computer Science in partial fulfillment of the requirements for the degree of Master of Engineering in Electrical Engineering and Computer Science at the MASSACHUSETTS INSTITUTE OF TECHNOLOGY June 2021

© Massachusetts Institute of Technology 2021. All rights reserved.

Author ............................................................... Department of Electrical Engineering and Computer Science
May 27, 2021

Certified by .......................................................... Cynthia Breazeal
Professor of Media Arts and Sciences
Thesis Supervisor

Accepted by .......................................................... Katrina LaCurts
Chair, Master of Engineering Thesis Committee
Designing Parent-Child-Robot

Triadic Storybook Reading Interaction

by

Soo Jung Jang

Submitted to the Department of Electrical Engineering and Computer Science on May 27, 2021, in partial fulfillment of the requirements for the degree of Master of Engineering in Electrical Engineering and Computer Science

Abstract

With an increasing availability, social robots’ domains and applications have been expanding, yet the research on human-robot interaction (HRI) still mostly focuses on single person to single robot interactions. Contributing to the field of multi-party HRI in educational domain, this thesis presents a novel parent-child-robot interaction paradigm in the context of shared reading. Constructive parent-child shared reading is crucial in children’s early literacy learning, and as a result, we strive to aid children’s learning with productive and engaging parent-child-robot triadic reading interactions.

The thesis work designs and develops an interactive reading system consisting of a robot facilitator, a storybook tablet app, and a teleoperation controller. Using the implemented reading system, we conduct a pilot Wizard of Oz (WoZ) triadic interaction study with four families, observing and analyzing their triadic interactions in shared reading setting.

The pilot study investigates the effects of triadic reading on dyadic reading, and compares the effects of different robot interaction strategies. The study’s results suggest that the triadic reading experience generally have a positive influence on families’ reading behaviors and their perceptions on social robots, and that each robot strategy has a unique set of effects on the interaction. The thesis work’s results, along with its discussions, provide critical insights into parent-child-robot shared reading design considerations.

Thesis Supervisor: Cynthia Breazeal
Title: Professor of Media Arts and Sciences
Acknowledgments

It has been a great honor and privilege to be able to work with the Personal Robots Group at MIT Media Lab. The people at the Personal Robots Group are truly amazing and talented, and this thesis would not have been possible without the support and help from many wonderful people in the lab. First of all, I would like to express my sincerest gratitude to my supervisor, Cynthia Breazeal, whose vision and enthusiasm in robotics and in human-robot interaction greatly inspired me. It has been a great pleasure to be a part of her group, and to receive her guidance and insightful feedback.

I would also like to thank our research scientist, Hae Won Park, for sharing her vast knowledge and experience in interaction studies and robot deployments. Her weekly feedback helped me improve both technical and design aspects of the study, and I am grateful that I was able to receive advice from her throughout my thesis project.

Huili Chen - this thesis would have been impossible without her. She helped me in all areas of the project, from technical debugging to execution of the study. She actually introduced me to the world of social robots 4 years ago, back when I was an undergraduate researcher at the Personal Robots Group. I feel extremely lucky to be able to work with her for so long, and I cannot thank her enough for her help and support.

Sharifa Alghowinem - her help in data analysis and in recording system meant a lot to me, especially during Jibo deployment. Pedro Colon-Hernandez - it was great to receive his help and advice when designing Jibo questions. I am thankful that I got to work closely with them throughout the project. I would also like to thank Jon, Kejia, Sooyeon, Meng, and many other members of the Personal Robots Group, who have been supportive and friendly throughout my time at the lab. In addition, warm thanks go to all families who participated in the interaction study. I had a great time collaborating with them on studies and interviews; the parents and children were truly amazing and inspiring.

Last but not least, I would like to thank my family for always having my back and supporting me. I received so much love and support from my family throughout my years at MIT, and I would not have been where I am today without them. Mom, dad, Hajong, Dami, and Steve - I love you all very much.
Contents

1 Introduction 15
   1.1 Thesis Overview ........................................... 16

2 Related Work 17
   2.1 Multi-Party Human-Robot Interaction ......................... 18
   2.2 Child-Robot Interaction ..................................... 20
   2.3 Strategies for Parent-Child Shared Reading .................... 22
      2.3.1 Protocols for Dialogic Reading: CROWD and PEER .......... 22
      2.3.2 Strategies for Parent-Child-Coach Interaction ............. 22

3 Parent-Child Dyadic Interaction 25
   3.1 DAMI-P2C Survey Features .................................. 26
      3.1.1 Child’s Behavior Questionnaire (CBQ) .................... 27
      3.1.2 Parenting Relationship Questionnaire (PRQ) ............... 27
      3.1.3 Parenting Stress Index (PSI) .............................. 27
      3.1.4 Socio-Economic Status (SES) .............................. 28
      3.1.5 Home Literacy Environment (HLE) ........................ 28
   3.2 DAMI-P2C Reading Features .................................. 30
   3.3 Correlation Analysis on Survey Features ....................... 31
   3.4 Correlation Analysis on Survey-Reading Features .............. 32
   3.5 Results and Application to Triadic Interaction Study ........... 33
      3.5.1 Chat Frequency Thresholds ............................... 33
      3.5.2 Chat Interval Thresholds ................................. 34
4 Interactive Reading System

4.1 Robot Facilitator: Jibo

4.2 Storybook Tablet App

4.3 Teleoperation Controller
   4.3.1 Control Panel
   4.3.2 Recording Panel
   4.3.3 Observation Panel
   4.3.4 Annotation Panel
   4.3.5 Orientation Panel
   4.3.6 Execution Panel
   4.3.7 Question and Hint List Panel
   4.3.8 Keyword and Behavior List Panel

5 Parent-Child-Robot Triadic Interaction

5.1 Robot Conditions
   5.1.1 Baseline and Evaluation Conditions
   5.1.2 Model Condition
   5.1.3 Suggest Condition
   5.1.4 Mixed Condition

5.2 Participant Information

5.3 Study Procedure

5.4 Feature Extraction
   5.4.1 Survey Features
   5.4.2 Reading Features

6 Triadic Interaction Study Results

6.1 Pre-Study and Post-Study Interview Results
   6.1.1 Pre-Study Interview
   6.1.2 Post-Study Interview
   6.1.3 Pre-Study to Post-Study Improvements

6.2 Post-Session Survey Results
<table>
<thead>
<tr>
<th>6.2.1</th>
<th>Self-Reported Ratings</th>
<th>73</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.2.2</td>
<td>Liked and Disliked Jibo Features</td>
<td>75</td>
</tr>
<tr>
<td>6.3</td>
<td>Reading Feature Analysis Results</td>
<td>77</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Parent-Child Interaction Reading Features</td>
<td>77</td>
</tr>
<tr>
<td>6.3.2</td>
<td>Robot Behavior Reading Features</td>
<td>80</td>
</tr>
<tr>
<td>7</td>
<td>Discussions and Conclusion</td>
<td>83</td>
</tr>
<tr>
<td>7.1</td>
<td>Data Collection on Intervention Timings</td>
<td>84</td>
</tr>
<tr>
<td>7.2</td>
<td>Impact of Triadic Reading on Dyadic Reading</td>
<td>85</td>
</tr>
<tr>
<td>7.3</td>
<td>Comparison of Different Robot Interaction Strategies</td>
<td>86</td>
</tr>
<tr>
<td>7.4</td>
<td>Conclusion</td>
<td>87</td>
</tr>
<tr>
<td>A</td>
<td>Survey and Interview Protocols for Triadic Interaction Study</td>
<td>89</td>
</tr>
<tr>
<td>A.1</td>
<td>Pre-Study Interview Protocol</td>
<td>89</td>
</tr>
<tr>
<td>A.2</td>
<td>Post-Study Interview Protocol</td>
<td>91</td>
</tr>
<tr>
<td>A.3</td>
<td>Post-Session Survey Questionnaire Form</td>
<td>93</td>
</tr>
<tr>
<td>A.3.1</td>
<td>Post-Session Survey</td>
<td>93</td>
</tr>
<tr>
<td>A.3.2</td>
<td>Post-Evaluation-Session Survey</td>
<td>94</td>
</tr>
<tr>
<td>A.3.3</td>
<td>Translation of Jibo Features to Social Attributes</td>
<td>95</td>
</tr>
</tbody>
</table>
List of Figures

3-1  Regression of chat frequency \((Y)\) on child’s age \((X)\) .......................... 34
3-2  Regression of chat interval \((Y)\) on child’s age \((X)\) .......................... 35

4-1  Station for interactive reading system ............................................... 38
4-2  Overview of interactive reading system .............................................. 38
4-3  Jibo in various expressions ................................................................. 39
4-4  Overview of storybook app ................................................................. 40
4-5  Graphic user interface (GUI) for teleoperation controller ......................... 41
4-6  Annotated teleoperation controller GUI ............................................... 41
4-7  Recording panel ............................................................................... 43
4-8  Observation panel ............................................................................. 44
4-9  Annotation panel .............................................................................. 45
4-10 Orientation panel ............................................................................. 46
4-11 Execution panel ............................................................................... 47
4-12 Question and hint list panel ............................................................... 48
4-13 Keyword and behavior list panel ......................................................... 48
4-14 Hint list panel updated by behavior button ......................................... 48

5-1  Teleoperation controller GUI highlighted during model condition ............ 56
5-2  Teleoperation controller GUI highlighted during suggest condition .......... 58
5-3  Experimenter’s real-time view on dyads during reading sessions ............. 62

6-1  Self-reported scores on baseline, experimental, and evaluation conditions 74
6-2  Self-reported scores on experimental conditions ................................. 74
6-3 Self-reported scores over sessions ........................................ 75
6-4 Counts of liked and disliked Jibo features ............................... 76
List of Tables

2.1 Definitions and examples of CROWD protocol . . . . . . . . . . . . 21
2.2 Definitions and examples of PEER protocol . . . . . . . . . . . . 21
2.3 Descriptions of parent-child-coach triadic interaction strategies . . . . 23

3.1 Summary of DAMI-P2C survey features . . . . . . . . . . . . . . 26
3.2 Descriptions of DAMI-P2C reading features . . . . . . . . . . . . 29
3.3 Summary of DAMI-P2C reading features . . . . . . . . . . . . . . 29
3.4 Significantly correlated survey feature pairs (p-val < 0.01) . . . . . . . 31
3.5 Significantly correlated survey-reading feature pairs (p-val < 0.01) . . . . 32

4.1 Descriptions and examples of buttons in behavior list panel . . . . . . . 50

5.1 Descriptions of triadic interaction study participants . . . . . . . . . . 60
5.2 Descriptions of triadic interaction study procedures . . . . . . . . . . 61
5.3 Summary of triadic study survey features . . . . . . . . . . . . . . 64
5.4 Summary of triadic study baseline reading features . . . . . . . . . . 65

6.1 Parent-child interaction reading features of baseline and evaluation conditions 78
6.2 Parent-child interaction reading features of experimental conditions . . . 78
6.3 Robot behavior types in different robot conditions . . . . . . . . . . . 80
6.4 Robot behavior reading features in different robot conditions . . . . . . 81
6.5 Robot behavior reading features in different robot behavior types . . . . 81
Chapter 1

Introduction

As supported by much research, children’s ability to read is key to their later academic successes, and shared storybook reading practices with parents greatly influence children’s early literacy development. Good parent-child shared storybook reading is more than just reading a book together; it includes reinforcing associations between audio and visual word representations, stimulating interest with personalized curriculum, and facilitating dialogic reading or having rich parent-child conversations around the story context [14].

One of the major problems in children’s literacy learning, especially with low socioeconomic status (SES) families, is limited exposure to high quality dialogic reading. For instance, studies have shown that low-SES families tend to have less frequent and shorter conversations with more authoritative and limited lexical content [18, 36]. Such lack of high quality parent-child conversations in low-SES households imposes a society-wide disparity since active participation of parents or adults is crucial for children’s early literacy and educational success [26, 27]. Recognizing this problem, this work aims to reduce the disparity in parents’ participation by using a social robot as a dialogic reading facilitator. The thesis designs and develops a novel multi-party human-robot interaction paradigm involving one social robot and a parent-child dyad. The interaction paradigm encourages productive parent-child discussions with a robot agent which actively participates in parent-child-robot triadic interactions with constructive interventions and rich reactions.

In addition to the contribution to early childhood education, the novel interaction paradigm proposed in this work aims to provide design insights on multi-party human-robot interac-
tions, thereby advancing the research on human-robot interaction (HRI). Prior research on HRI have mostly been focused on single person to single robot interactions, causing HRI research in educational domain to be heavily weighted towards single person child-robot or adult-robot interactions [9, 12, 23, 39]. Although a growing number of studies have begun to explore the field of multi-party human-robot interaction (as explained in section 2.1), research on multi-party HRI, especially on parent-child-robot interaction, is still greatly unexplored. As a result, through the thesis, we strive to provide preliminary insights and findings on parent-child-robot triadic interaction during shared storybook reading, which is a subset of multi-party HRI in educational domain. To our knowledge, this is the first pilot study on robot-facilitated triadic shared reading.

Through exploring the concept of utilizing parent-child-robot triadic interactions to promote dialogic reading, the thesis provides three major contributions:

1. Analysis of parent-child dyadic reading styles and their correlations to the dyad’s developmental and relational profiles
2. Implementation of an interactive triadic reading system composed of a storybook tablet app, a robot facilitator, and a teleoperation controller
3. Execution of a parent-child-robot triadic interaction pilot study testing different robot interaction strategies and evaluating dyads’ perception and reactions to them

1.1 Thesis Overview

To summarize the overall flow of the thesis, related work (section 2) and dyadic interaction study (section 3) provide reasoning and background on our triadic interaction design choices. Interactive reading system (section 4) and triadic interaction study (section 5) explain the implementation and design of our novel triadic interaction paradigm, tackling various research questions. Finally, results (section 6) and conclusion (section 7) summarize the insights gained from the triadic study, and propose key design factors to consider when developing effective parent-child-robot shared reading.
Chapter 2

Related Work

This thesis is an inter-disciplinary work involving research in fields of human-robot interaction and early literacy education. As a result, we first review and summarize a handful of previous work in the following three areas: multi-party human-robot interaction, child-robot interaction, and parent-child shared storybook reading.

(1) The related research on multi-party human-robot interaction (section 2.1) explains the effects of different robot behavior modalities on human groups within multi-party HRI. The findings from the previous multi-party HRI studies are used to guide the design of robot modalities in our parent-child-robot triadic interaction.

(2) The previous studies on child-robot interaction (section 2.2) suggest many successful design principles on facilitating good robot interactions with children. We follow many design principles recommended by the past child-robot interaction studies, aiming to design our robot facilitator to be child-friendly.

(3) The educational protocols of parent-child shared storybook reading (section 2.3) provide common strategies used by human reading facilitators to improve parent-child shared reading experience. We design our robot facilitator’s interaction and reaction strategies based on the prior work on successful shared reading protocols.
2.1 Multi-Party Human-Robot Interaction

There is a significant difference between single person human-robot interaction (HRI) and multi-party HRI which involves more than one person. In addition to human-robot communication, multi-party interactions need to consider communications and dynamics within human groups, raising many interesting research questions and challenges. Since parent-child-robot interaction is one form of multi-party interaction, a number of papers on multi-party HRI are reviewed in this section as preliminary research.

There are various robot behavior modalities (speech, non-verbal sound, motion, gaze, facial expression etc.) and each modality has its own strengths and weaknesses when being used in a multi-party interaction. Gaze, facial expression, head movement, and non-verbal utterance are effective means to convey robots’ emotional intentions, while body movement, arm gesture, and speech allow robots to better communicate direct task-oriented intentions [2, 32, 33, 37]. Because each modality yields different effects in multi-party HRI, studies vary in their choices of robot behavior modalities when designing their interactions.

Many multi-party interaction studies have chosen robot speech as their key modality, and have altered the human-robot dialogue contents by applying various strategies and computational models [11, 21, 41, 45]. Many of these studies have discovered that the existence of robot speech, regardless of its content, may harm the quality of interaction due to its noticeable and disruptive nature.

For example, in Jung et al.’s and Short et al.’s multi-party HRI studies, task performance and people’s perceived satisfaction were majorly affected by frequency and length of robot speech rather than its content because the robot’s speech disturbed the communication between humans [21, 41]. In line with the observation, Unhelkar et al.’s paper concluded that taking "communication cost" into account, or penalizing high frequency of speech, yields more effective human-robot communication [45]. Semmens et al.’s study also suggested that technology agent’s speech assistance requires appropriate timing to be helpful and welcomed [38]. To summarize, speech is a powerful modality which is suitable to attention-grabbing robot behaviors and direct communication, yet without the right timing, it has a risk of disturbing interactions among humans in multi-party HRI settings.
In contrast to many studies that heavily rely on speech to facilitate interactions, some studies have only used non-verbal modalities such as body orientation and gaze to assist interactions among humans. For example, Rifinski et al.’s and Tennent et al.’s papers used a silent bystander-style robot to encourage better engagement and participation of members within a human group [35, 44]. In both studies, the silent robot was able to improve the quality of communication with only passive and indirect reaction behaviors.

A modality’s effects can be inferred since many studies have tested different behavior styles within the same modality, often comparing baseline case with random or no behavior against intentional cases. To summarize the insights gained from the studies’ comparisons, verbal speech is most commonly used and yields certain benefits from direct human-robot communication, yet has a risk of being disruptive. On the other hand, non-verbal modalities such as gaze and motion do not necessarily support straightforward communication, yet can indirectly influence people’s behaviors.

Because each modality has unique effects on multi-party HRI, it is important to contextualize a robot’s multimodal behavior design when optimizing interaction experience. For instance, intensity and timing of a robot’s multimodal behaviors should be purposefully designed and controlled. Therefore, when designing parent-child-robot triadic interaction, we utilize a various mix of different robot modalities, paying careful attention to balancing robot’s intervention intensity and timing.
2.2 Child-Robot Interaction

Many prior HRI studies have involved children in their work, exploring the field of child-robot interaction (CRI) [16, 24, 40, 42, 43]. One of the major differences between adult-robot interaction and child-robot interaction is that robot’s appearance and children’s engagement are significantly considered in the CRI study design.

Most CRI studies put much effort towards designing and using a friendly-looking robot that appeals to children. For example, both Gvirsman et al.’s and Short et al.’s studies found that a robot’s friendly appearance created by a fabric puppet over robot skeleton promotes children’s engagement and interest [16, 42].

In terms of robot behavior modalities, many CRI work have discovered that using a mixture of multiple modalities yields more positive experiences with children. Gvirsman et al. have observed that interacting with physical robots with body movement and gaze modalities improve children’s engagement and parent-child interaction, in comparison to interacting with tablets [16]. In line with the observation, Shen et al.’s study found that using speech coupled with non-verbal sound is more effective than using speech alone when catching children’s attention [40]. Furthermore, Short et al. observed that families with children prefer to interact with robot via multiple behavior modalities (speech, touch, and motion) [42].

Inspired by the prior CRI studies, the thesis work uses Jibo, a friendly-looking social robot with rich animations, body movements, and sound effects that can appeal to children. However, unlike most previous CRI studies that have focused on no-adult scenarios [16, 24, 40, 42, 43], the thesis investigates how a robot shapes the interactions between a parent and a child. Given a lack of research on robot-facilitated parent-child interactions, this thesis work uniquely contributes to the state-of-art CRI research.
Table 2.1: Definitions and examples of CROWD protocol

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
</table>
| C - Completion | Asking child to complete a sentence by filling in the blank (elliptical questions) | Adult: "The wolf was trying to blow down the house that belonged to ..."  
Child: "Three little pigs!" |
| R - Recall | Asking child to share details from book to demonstrate his understanding of the plot | Adult: "What happened to the house made of straw?" |
| O - Open-Ended | Asking child questions that require an explanation (more than just a few words) | Adult: "What do you think happened?" / "How do you feel about ..." |
| W - Wh | Asking child to expand on the plot with questions which start with "wh" (who, what, when, where, why, how) | Adult: "Where would they go after the first house blew down?" |
| D - Distancing | Asking child questions which make him relate aspects of the story to his own life or society | Adult: "Was there a time when you lost something like the piggy?" / "What would you do if ..." |

Table 2.2: Definitions and examples of PEER protocol

<table>
<thead>
<tr>
<th>Acronym</th>
<th>Definition</th>
<th>Example</th>
</tr>
</thead>
</table>
| P - Prompt | Prompt child to be engaged and talk about the story | Adult: "Look! What is that in the book?"  
Child: "That’s a dog." |
| E - Evaluate | Evaluate or praise child’s responses | Adult: "Correct, that’s a dog." |
| E - Expand | Expand child’s verbalization through repeating and adding information to it | Adult: "It’s a kind of a dog called a beagle." |
| R - Repeat | Encourage child to repeat expanded utterances | Adult: "Can you say - beagle?" |
2.3 Strategies for Parent-Child Shared Reading

In order to design a robot that can successfully serve as a reading facilitator, we investigated into common educational strategies and protocols used in improving parent-child shared storybook reading. In addition, we reviewed papers which have used human facilitator’s real-time coaching to improve shared reading in order to design the robot facilitator’s behaviors after the successful human facilitator’s strategies.

2.3.1 Protocols for Dialogic Reading: CROWD and PEER

Many studies have followed CROWD protocol to facilitate dialogic reading [13, 48, 47, 49]. CROWD stands for completion, recall, open-ended, wh, and distancing, as summarized in Table 2.1. According to prior work that have used the CROWD protocol, initiating conversations and asking questions based on CROWD yield successful dialogic reading, improving the quality of parent-child shared reading [13, 48, 47, 49].

In addition to the CROWD protocol, PEER protocol has also been commonly used by many studies to improve parent-child shared storybook reading [13, 48, 49]. PEER stands for prompt, evaluate, expand, and repeat, as described in Table 2.2. Interacting with and reacting to child following the PEER protocol enrich parent-child interactions, and therefore the protocol has been recommended by many studies for achieving productive and fun dialogic reading [13, 48, 49].

Given their effectiveness in facilitating good parent-child shared reading, both CROWD and PEER protocols are used in this thesis work as guidelines when designing robot facilitator’s question asking and reaction behaviors in triadic interaction.

2.3.2 Strategies for Parent-Child-Coach Interaction

A variety of coaching techniques have been used in intervention training programs that target parent-child dyads’ shared reading. Among them, demonstration, or modeling, method has been widely used [4, 19]. When using the demonstration method, a facilitator demonstrates a good dialogic reading example to parents by showing a model example in person or with a pre-recorded "read together, talk together" style video. The demonstration allows
Table 2.3: Descriptions of parent-child-coach triadic interaction strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Establish supportive context</td>
<td>Coach arranges environment to be supportive and fun</td>
</tr>
<tr>
<td>Acknowledge competence</td>
<td>Coach recognizes and praises good facilitative behaviors of parent and child</td>
</tr>
<tr>
<td>Focus attention</td>
<td>Coach comments on aspects of interaction to draw parent’s attention</td>
</tr>
<tr>
<td>Provide information</td>
<td>Coach provides information on child’s developmental agenda behind his behaviors</td>
</tr>
<tr>
<td>Model</td>
<td>Coach impersonates parent to demonstrate good dyadic interaction examples</td>
</tr>
<tr>
<td>Suggest</td>
<td>Coach gives specific suggestions as to what to do with child</td>
</tr>
</tbody>
</table>

Parents to be aware of what good reading behaviors are, and potentially leads them to emulate and apply the model behaviors to their own dyadic shared reading interactions. This method is highly scalable since the training can be done in a pre-recorded lecture-style way. However, the method lacks detailed tailoring to each parent-child dyad.

Some papers have used facilitator’s real-time coaching to improve parent-child shared reading [17, 25]. The real-time parent-child-coach interaction allows the coach to give direct and specific instructions to the parent, providing a tutor-style experience tailored to each dyad. McCollum et al. suggested specific strategies to follow when facilitating parent-child-coach interactions [25]. Table 2.3 provides the details of McCollum et al.’s strategies. Because our robot facilitator interacts with parent-child dyads in real time, many of McCollum et al.’s triadic interaction strategies are used to design the robot’s reading facilitation and coaching behaviors.
Chapter 3

Parent-Child Dyadic Interaction

In order to design a robot facilitator for parent-child dyadic interactions, it is crucial to first understand how parent-child dyads interact in shared reading setting without a robot, and how a dyad’s socio-demographic and relational profiles impact their reading behaviors. Hence, the thesis first analyzed parent-child dyadic reading interactions using a parent-child multimodal interaction dataset, DAMI-P2C [8, 10].

DAMI-P2C is an abbreviation for "dyadic affect in multimodal interaction – parent to child", and the dataset has been collected from 34 parent-child pairs with 3 to 7 years-old children. Each DAMI-P2C dyad pair has participated in two 45-minutes-long reading sessions in lab. During each session, the dyad has read digitized versions of various storybooks together, in a way that they would normally read at home.

The DAMI-P2C dataset provides a wide range of surveyed and observed information about the parent-child dyads' profiles and reading interactions. As a result, we used the dataset to extract "survey features" capturing dyads’ socio-demographic profiles and "reading features" capturing their shared reading interaction as described in sections 3.1 and 3.2. Then, using the extracted features, we ran correlation analysis to identify characteristics and relationships among the features, and used the insights gained from the analysis to design aspects of the triadic interaction study. Sections 3.3 and 3.4 describe the correlation analysis processes and results, and section 3.5 explains how the dyadic interaction data analysis results are applied to the triadic study.
3.1 DAMI-P2C Survey Features

The participants of DAMI-P2C have completed a number of questionnaire surveys capturing their socio-demographic and parent-child relational profiles. Families’ demographic information, home literacy environment, parenting style, parenting stress, and child’s temperament were collected with the following five questionnaires respectively: Socio-Economic Status (SES), Home Literacy Environment (HLE), Parenting Relationship Questionnaire (PRQ), Parenting Stress Index (PSI), and Child’s Behavior Questionnaire (CBQ).

The following sections 3.1.1 through 3.1.5 provide details on each questionnaire and explain how survey features were extracted from the families’ questionnaire responses. Furthermore, Table 3.1 provides a statistical summary of DAMI-P2C dyads’ survey features extracted from their questionnaire responses.
Overall, the DAMI-P2C families have generated a wide range of survey feature values, indicating that the dataset’s families vary in socio-demographic and relational profiles. Such diversity across participant families allows us to investigate the characteristics of different dyadic profiles.

3.1.1 Child’s Behavior Questionnaire (CBQ)

CBQ measures a child’s temperament [31]. It surveys 36 questions to output three sub-scales: Surgency, Negative Affect, and Effortful Control. Each response is recorded in a likert scale between 1 and 7, and the total CBQ score is an average of all likert responses, high score representing child’s negative temperament. Score to each sub-scale is an average of corresponding likert responses. Surgency averages activity level, high-intensity pleasure, impulsivity, and negative shyness. Negative affect averages anger, discomfort, fear, sadness, and negative soothability. Effortful Control averages attention focusing, inhibitory control, low-intensity pleasure, and perceptual sensitivity. As a result, CBQ results in four features: three sub-scale scores and a total score.

3.1.2 Parenting Relationship Questionnaire (PRQ)

PRQ measures a caregiver’s parenting style and its perception on parent-child relationship [22]. It is comprised of 60 questions, resulting in five sub-scales: Attachment, Discipline, Involvement, Confidence, and Frustration. Each response is recorded in a likert scale between 0 and 3, and the scores are calculated by summing the responses. PRQ total score is a modified sum of PRQ sub-scale scores with Frustration score inverted; high PRQ total means positive parenting style with high attachment, discipline, involvement, and confidence, and low frustration. PRQ results in six features: five sub-scale scores capturing different dimensions in parenting style, and one total score.

3.1.3 Parenting Stress Index (PSI)

PSI measures a parent’s level of parenting stress [1]. It is comprised of 36 questions, resulting in three sub-scales: Parental Distress (PD), Parent-Child Dysfunction (PCDI),
and Difficult Child (DC). Each response is in a likert scale between 1 and 5, and the scores are calculated by summing the responses. High PSI score means high parenting stress level. PSI results in four features: three sub-scale scores and one total score.

3.1.4 Socio-Economic Status (SES)

SES questionnaire can contain numerous parameters [3], yet for DAMI-P2C, only a subset of SES parameters were chosen based on families’ willingness to respond and on relevance to the context of shared storybook reading. As a result, families’ social, cultural, and developmental backgrounds were measured with following parameters of SES: child’s gender, child’s age, parent’s highest level of education, use of English at home, parent’s English proficiency, and parent’s years spent in USA.

Child’s gender and use of English at home are recorded as binary responses (1 for female and English used at home), and child’s age is recorded as integer years. Parent’s highest education, English proficiency, and years in USA are in different likert scales: between 0 and 4, 0 and 2, and 0 and 6 respectively. As a result, SES outputs six features, each capturing an unique information on dyad’s developmental and socio-economic status.

3.1.5 Home Literacy Environment (HLE)

HLE measures a family’s home literacy environment [30]. A single score is outputted from responses to six questions: number of children’s books at home, age of child when first read to, amount of time spent on reading with the child at home, amount of time spent by child reading alone, frequency of family members reading to child, and frequency of family members teaching child alphabet. Each response is in a likert scale between 1 and 5, and the total HLE score is calculated by converting the responses to percent of maximum possible score, and then averaging the responses. High HLE score represents good home literacy environment.
Table 3.2: Descriptions of DAMI-P2C reading features

<table>
<thead>
<tr>
<th>Behavior</th>
<th>Feature</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dialogic</td>
<td>Chat : Read</td>
<td>Ratio of time spent on chatting to time spent on reading</td>
</tr>
<tr>
<td>Reading</td>
<td>Chatting Duration</td>
<td>Average time spent on each chat session</td>
</tr>
<tr>
<td></td>
<td>Chatting Frequency</td>
<td>Average number of chat sessions per second</td>
</tr>
<tr>
<td>Child’s</td>
<td>Child : Parent in Reading</td>
<td>Ratio of child’s reading time to parent’s reading time</td>
</tr>
<tr>
<td>Participation</td>
<td>Child : Parent in Chatting</td>
<td>Ratio of child’s chatting time to parent’s chatting time</td>
</tr>
<tr>
<td></td>
<td>Child : Parent in Initiation</td>
<td>Ratio of child-initiated chats to parent-initiated chats</td>
</tr>
<tr>
<td>Active Turn</td>
<td>Turn Taking Rate</td>
<td>Average number of turns taken per second</td>
</tr>
<tr>
<td>Taking</td>
<td>Word Rate</td>
<td>Average uttered word count per second (rate of speech)</td>
</tr>
</tbody>
</table>

Table 3.3: Summary of DAMI-P2C reading features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat : Read of Parent</td>
<td>0.96</td>
<td>1.34</td>
<td>0.12</td>
<td>6.62</td>
<td>3.14</td>
</tr>
<tr>
<td>Chat : Read of Child</td>
<td>7.43</td>
<td>15.53</td>
<td>0.16</td>
<td>75.05</td>
<td>3.4</td>
</tr>
<tr>
<td>Chatting Duration</td>
<td>11.39</td>
<td>8.63</td>
<td>5.56</td>
<td>41.39</td>
<td>2.33</td>
</tr>
<tr>
<td>Chatting Frequency</td>
<td>0.04</td>
<td>0.02</td>
<td>0.01</td>
<td>0.08</td>
<td>0.8</td>
</tr>
<tr>
<td>Child : Parent in Reading</td>
<td>0.7</td>
<td>1.86</td>
<td>0</td>
<td>9.3</td>
<td>4.02</td>
</tr>
<tr>
<td>Child : Parent in Chatting</td>
<td>0.41</td>
<td>0.26</td>
<td>0.1</td>
<td>1.17</td>
<td>1.18</td>
</tr>
<tr>
<td>Child : Parent in Initiation</td>
<td>0.47</td>
<td>0.32</td>
<td>0.08</td>
<td>1.15</td>
<td>0.85</td>
</tr>
<tr>
<td>Turn Taking Rate in Reading</td>
<td>0.04</td>
<td>0.04</td>
<td>0</td>
<td>0.18</td>
<td>1.75</td>
</tr>
<tr>
<td>Turn Taking Rate in Chatting</td>
<td>0.12</td>
<td>0.05</td>
<td>0.05</td>
<td>0.2</td>
<td>0.5</td>
</tr>
<tr>
<td>Word Rate of Parent</td>
<td>2.07</td>
<td>0.61</td>
<td>0.54</td>
<td>2.95</td>
<td>-0.71</td>
</tr>
<tr>
<td>Word Rate of Child</td>
<td>1.38</td>
<td>0.47</td>
<td>0.62</td>
<td>2.39</td>
<td>0.52</td>
</tr>
</tbody>
</table>
3.2 DAMI-P2C Reading Features

DAMI-P2C reading features were extracted from the transcripts of parent and child during the shared reading activity in order to capture quality and characteristics of each dyad’s reading interaction. According to Whitehurst et al., high-quality reading behaviors are comprised of following three dimensions [47]:

1. Dialogic Reading: having discussions in parallel with story reading
2. High Child’s Participation: getting higher portions of child reading and speaking
3. Active Parent-Child Turn Taking: frequent turn switching between parent and child during reading and speaking

We contextualized the three high-quality reading dimensions by pursuing following steps. We first distinguished reading versus chatting (dialogic reading) in parent-child transcripts by comparing their conversations against the storybook script. If a block of conversation matched a script in the storybook, we classified it as reading. If not, we filtered out deliberate distractions (e.g. "I am bored") and classified the remaining parent-child conversations as constructive chatting. After dividing reading versus chatting, the portions read or spoken by parent or child were measured to extract reading features related to child participation and turn taking.

In total, eight different types of reading features were extracted as described in Table 3.2. By design, some feature types consist more than one feature; word rate and chat-to-read ratio have two values (for parent and child), and turn taking rate has two (for during reading and chatting). As a result, 11 reading features were outputted in total. The DAMI-P2C dyads’ resulting feature values are statistically summarized in Table 3.3.

High chat-to-read ratio, chat duration, and chat frequency values represent good dialogic reading behaviors, high child-to-parent ratio values represent high child’s participation, and high turn taking rate and word rate values represent active turn taking. Note that DAMI-P2C dyads have yielded large standard deviations for child-to-parent ratio in reading and chat-to-read ratio for child; this observation indicates that the children were widely different in their reading participation styles, potentially due to their varying literacy ability.
In order to understand the correlations within the survey features representing dyads’ socio-demographic and relational profiles, pairwise correlations were calculated between sub-scale scores of different questionnaires. All total scores except HLE total, and correlations within the same questionnaire were excluded to reduce redundancy. All values were min-max normalized to range within 0 and 1, and Pearson correlation coefficient was used to compute the pairwise correlations. As a result, a total of 122 pairwise correlations were calculated. Table 3.4 shows the ten most significantly correlated feature pairs by filtering pairs with p-values less than 0.01 and by sorting them by descending magnitude of correlation.

Looking at the pairwise correlations within the survey features provided by Table 3.4, many (six out of ten) most significantly correlated pairs are from PRQ and PSI. The significant correlations between many PRQ and PSI sub-scale scores indicate that a parent’s perspective on her parenting and parent-child relationship greatly influences her level of parenting stress, and vice versa. In addition, parents are less attached and less confident with older children, indicating that a child’s age significantly influences parent-child relationships and dynamics.

The correlation analysis result suggests that basic statuses such as child’s age and use of English can shape parental relationship and stress, and therefore customized thinking, even in the level of basic statuses, is needed when studying parent-child interactions.
In order to investigate the correlations between survey features and reading features, pairwise correlation values were calculated between 18 sub-scale survey features (excluding all total scores except HLE total) and all 11 reading features. All values were min-max normalized and Pearson correlation coefficient was used to generate a total of 198 pairwise survey-reading correlations. Table 3.5 gives a summary of the correlation analysis results by showing the five most significant correlations between DAMI-P2C survey features and reading features with the methods explained in section 3.3.

Looking at the summarized results in Table 3.5, there are some interesting correlations between survey and reading features. Parent’s rate of speech is highly positively correlated with her English proficiency and the use of English at home, reflecting that parents who are more comfortable with English are more likely to read and speak faster. Child’s rate of speech is negatively correlated to PRQ attachment; because PRQ attachment is negatively correlated to child’s age (Table 3.4), this correlation indicates that older children tend to read and speak faster.

It is noteworthy to observe that child participation (represented by child-to-parent ratio in initiation) and chat frequency are significantly influenced by child’s gender and age. Child’s gender and initiation ratio are positive correlated, implying that female children are more likely to initiate chats, and child’s age and chat frequency are negatively correlated, indicating that younger children yield more frequent chats. The correlations between survey and reading features imply that dyad’s socio-demographic and relational profiles do impact their shared reading behaviors.
3.5 Results and Application to Triadic Interaction Study

For both survey-survey and survey-reading correlation analysis, child’s age was included in the top five most significantly correlated feature pairs, indicating that the child’s age is an important factor in shaping the dyad’s relational dynamics and their reading behaviors.

Bracken et al.’s study with 233 preschool children support our observation since they also observed that age significantly influences child’s literacy skills and therefore correlates to parent-child reading interactions [5]. In addition to Bracken et al.’s study, there are handful of prior work that have found that children’s age and their reading performance are significantly correlated due to their developmental stages [28, 46].

As a result, when designing the parent-child-robot triadic reading interaction, we take child’s age into consideration and use it to project thresholds that can assess a dyad’s chat frequency and suggest good intervention timings. The thresholds are crucial during the triadic interaction study because the human experimenters launch robot facilitator’s behaviors based on the system’s intervention timing suggestions made by the thresholds. The following sections 3.5.1 and 3.5.2 explain how the thresholds are calculated and how they are applied to the triadic interaction study.

3.5.1 Chat Frequency Thresholds

As indicated in Table 3.5, there is a significant correlation between dyad’s chat frequency and child’s age. Building on this correlation, we applied linear regression on chat frequency against child’s age to get chat frequency thresholds as a function of child’s age. During a parent-child-robot interaction study, the outputted thresholds can serve as a guideline that indicates a dyad’s chat frequency level relative to the other dyads. For example, if a dyad’s chat frequency falls low compared to the other dyads with children in similar age group, it is an indication of infrequent chatting, and the experimenters can know that it may be a good timing for the robot facilitator to intervene and increase the dyad’s chat frequency to meet the threshold’s guideline.

In order to determine the chat frequency thresholds, linear regression is applied on chat frequency (number of chats per second) against child’s age (in decimal years). Note that
Figure 3-1: Regression of chat frequency ($Y$) on child’s age ($X$)

$$Y = -0.008X + 0.083 \text{ (Adjusted } R^2 = 0.245)$$

the child’s age used for the regression is different from the surveyed SES child’s age; the regression’s age uses birthday to compute the exact age in decimal years while the surveyed SES age uses conventional integer years. The fitted regression predicts a chat frequency given a dyad’s child’s age, and it is used as the middle chat frequency threshold. Then, the 60% prediction interval is computed around the regression in order to provide upper and lower chat frequency thresholds. The interval value of 60% is chosen because it splits the dataset into equal halves when dividing dyads by inside versus outside prediction interval. Figure 3-1 provides a plot of the lower, middle, and upper chat frequency thresholds calculated by the linear regression and the 60% prediction interval around it.

### 3.5.2 Chat Interval Thresholds

Chat frequency thresholds provide a good guideline of a dyad’s chat frequency levels specific to their child’s age. However, it is difficult to apply chat frequency thresholds to a real-time interaction study because frequency values are difficult to comprehend, and because annotating the exact occurrences of chats is challenging. As a result, we used chat interval ($1 / \text{chat frequency}$) instead of frequency when calculating thresholds for the parent-child-robot triadic interaction study.

Using chat interval thresholds yields two main advantages: easy comprehension and robustness to chat occurrence timings. First, chat interval (e.g. 20 seconds between chats)
Figure 3-2: Regression of chat interval \((Y)\) on child’s age \((X)\)

\[ Y = 5.285X + 1.360 \text{ (Adjusted } R^2 = 0.099) \]

is more intuitively comprehensible than chat frequency (e.g. 0.05 chats per second), and as a result, interval thresholds provide a more clear guideline to the human experimenters. In terms of robustness to chat occurrences, chat interval allows experimenters to apply thresholds independent from previous chats by resetting a timer whenever there is a chat occurrence and alerting infrequent chatting when the reset time passes the threshold (when \(t\) seconds has passed since the last chat occurrence where \(t\) is the threshold). The timer reset method with chat interval thresholds is an easy and effective way to assess and notify a dyad’s chat frequency levels.

In order to calculate the chat interval thresholds, we utilized the methods described in section 3.5.1. We first applied linear regression on chat interval (seconds between chats) against child’s age (in decimal years), then computed 60% prediction interval around the regression. Figure 3-2 provides a plot of the lower, middle, and upper chat interval thresholds calculated by the linear regression and its prediction interval.
Chapter 4

Interactive Reading System

The interactive reading system is majorly composed of following three components:

1. Robot facilitator: a social robot agent that interacts with parent and child
2. Storybook tablet app: an app that parent and child use to read books
3. Teleoperation controller: a controller that experimenters use to run interaction study

In terms of physical set up, the robot and the tablet are mounted onto a platform together forming a station shown in Figure 4-1. The station allows the components to maintain their configuration when being deployed to families’ homes. Following Jeong et al.’s work, the station integrates the devices needed for the study, including Jibo robot, Samsung Galaxy tablet, Intel NUC PC, wireless router, and Logitech USB camera [20]. The camera records the study, the router establishes wireless connection between the station components, and the PC creates an environment for the teleoperation controller to run.

The teleoperation controller allows the experimenters to view, access, and control the station. Because the controller in NUC can be accessed through remote desktop programs, the experimenters can control the station remotely in real time. For communication, we use Robot Operating System (ROS) which publishes and subscribes messages across various system components. Figure 4-2 provides an overview of the interactive reading system, and sections 4.1 through 4.3 provide descriptions for each system component.

---

1 The study was conducted during COVID-19, so the study was conducted at participants’ homes.
Figure 4-1: Station for interactive reading system

Figure 4-2: Overview of interactive reading system
4.1 Robot Facilitator: Jibo

As the robot reading facilitator, we use Jibo, an 11 inches tall table-top voice-user interface (VUI) agent with a touchscreen face and an expressive body (Figure 4-3). Jibo has been used in multiple short-term and long-term HRI studies in various settings such as lab, home, and school [6, 29], and therefore is appropriate for our triadic interaction study which requires deployment of robots to participants’ homes.

Jibo has a friendly and modern appearance with animation-style head and body. With its body movement, face display, LED light, sound, and speech, Jibo is capable of conveying various multimodal social cues. In addition, Jibo has a text-to-speech (TTS) feature which automatically executes expressions with various modalities given a text. For example, inputting "thank you" through TTS automatically makes Jibo say "thank you" with social cues of appreciation such as a nod. Because Jibo is highly expressive and likeable, we use Jibo as our robot facilitator when running the parent-child-robot interaction study.

During the study, Jibo sends ROS messages to the teleoperation controller, indicating real time statuses such as its current motion and sound, and receives ROS messages from the teleoperation controller for commands. When it is not executing commands, Jibo’s default behavior during the triadic reading interaction is to gaze the speaker, conveying silent attentive listening behaviors.
4.2 Storybook Tablet App

The storybook app on the tablet is a simple reading application that serves a role similar to an electric book. The app has been used in the prior dyadic interaction study when collecting the DAMI-P2C dataset [8]. As a result, the app ensures consistency in reading interface setting across the two studies.

On the app, the dyad can choose a book from the collection, and read it through a simple display interface. Figure 4-4 provides an overview of the storybook app. Similar to Jibo, the storybook app communicates with the teleoperation controller through ROS, notifying its current statuses such as book and page that the dyad is currently reading.

There are total 16 book options in the app: six short with less than 500 words, and ten long with more than 500 words. The specific children’s books are chosen because these books have abundant list of story questions created through the previous projects at the lab. The details of how questions are generated and used are provided in section 4.3.7.
Figure 4-5: Graphic user interface (GUI) for teleoperation controller without recording panel and partial observation panel

Figure 4-6: Annotated teleoperation controller GUI showing all eight panels
4.3 Teleoperation Controller

Because the parent-child-robot interaction study is run by human experimenters remotely, it is important to have a solid teleoperation system that is easy to operate, allows real-time observation, and provides various options for robot interaction behaviors.

For easy comprehension, we use a graphic user interface (GUI) implementation for the controller, which displays a wide range of features including recording, real-time status update, list of story questions and hints, and pre-coded interaction behaviors, using an intuitive design and layout.

In addition, in order to further reduce the experimenters’ cognitive load when running the study, the controller provides various suggestion and random selection features. Some factors such as robot’s interaction target and question type should be randomized throughout a session in order to assure diversity in interactions. The teleoperation controller automatically selects these factors randomly after each book page flip, yet it still allows the experimenter to override the selection. Such controller design ensures flexibility and autonomy of the human experimenters while offering easy operation.

The teleoperation controller can be broken down into eight categories, each corresponding to a panel grouping in the GUI: control, recording, observation, annotation, orientation, execution, question and hint list, and keyword and behavior list. Figure 4-5 provides a capture of the teleoperation GUI during an example interaction, and Figure 4-6 annotates the GUI to show its 8 panels. Sections 4.3.1 through 4.3.8 give detailed descriptions of each panel composing the teleoperation controller.

4.3.1 Control Panel

The control panel has two parts: Jibo scripts and systems. Jibo scripts section has three buttons that launch specific Jibo behaviors that guide the triadic reading interaction. "Jibo Intro" button is pressed at the beginning of each session, and it makes Jibo notify that the story time session has started, explain the task, and tell parent and child to choose a book from the storybook app. "Jibo Intro" also notifies the start of audio and video recording, and automatically starts the recording. "Jibo Book End" button is pressed in between books;
Jibo expresses a celebratory action for completing a book, then follows up by asking the dyad if they want to read another one. Finally, "Jibo Sess End" button is pressed at the end of the session, when the dyad have completed reading 2 or 3 books. At session end, Jibo does a celebratory dance, and notifies the dyad that the session is over. Jibo also notifies that its recording will stop and that it will go back to leisure mode, which is Jibo’s default mode when not doing story time.

In systems section, there are multiple buttons that control the logistics of the interaction. "Camera Remind" button is pressed when parent or child moves away from the camera view; the button makes Jibo ask the dyad to adjust their seating to be observable. The next four buttons ("Start Video Record", "Stop Video Record", "Start Rosbag Record", and "Stop Rosbag Record") are buttons that can start or stop the recordings. The video-related buttons control audio and video recordings, and the rosbag-related buttons control ROS message recordings. Section 4.3.2 provides more details about the recordings that happen during the interaction. Finally, "Unity App Quit" button allows the experimenters to force quit the storybook app on the dyad’s tablet. Whether the storybook app is currently on or off is displayed under "Unity App Quit"; green means the app is on, and red means the app is off. Overall, the control panel allows the experimenters to directly run and control certain parts of the reading interaction system.
4.3.2 Recording Panel

The recording panel (Figure 4-7) is composed of 4 windows: "GST-video" (black), "ROS-audio" (blue), "ROS-video" (purple), "rosbag-recording" (green). Each window is color-coded for easy visual comprehension. "GST-video" records audio and video and outputs an avi video. "ROS-audio" and "ROS-video" records audio and video ROS topics, and "rosbag-recording" records ROS commands sent to and received from Jibo and the storybook app. The three windows related to ROS ("ROS-audio", "ROS-video", and "rosbag-record") output one composite rosbag file which aligns all ROS messages by timestamps.

4.3.3 Observation Panel

The observation panel has two parts: observation on dyad’s tablet and on dyad. The panel on dyad’s tablet (Figure 4-8a) provides a real-time view of the storybook tablet app, allowing the remote experimenters to know what the dyad is currently reading. In addition to the view of book illustration and text, the tablet view provides book title and page progress; in Figure 4-8a example, the book title is "on_the_farm" and the page progress is "2 / 16", indicating that the dyad is on second page out of 16 total pages.

The observation panel on dyad provides a real-time camera view on the dyad via Zoom. The Zoom view is provided by the USB camera mounted in the interactive reading system station, and it allows the experimenters to listen and observe the dyad’s interactions in real time during the triadic interaction study.
4.3.4 Annotation Panel

The annotation panel (as shown in Figure 4-9) allows experimenters to annotate dyad’s conversation in real-time, provides information on current interaction statuses, and serves a crucial role in suggesting good robot intervention timings. First, throughout the triadic interaction session, experimenters annotate dyad’s conversation by pressing "Related Chat" and "Distraction" buttons. This annotation is used to calculate dyad’s chat frequency in real-time, and is also later used to extract reading features from the dyad’s conversation transcript. "Mistake" button allows experimenters to mark annotations or robot behaviors that are clicked or launched by human mistake.

In addition to the buttons, the annotation panel provides current interaction statuses though displaying the following: "# of Related" (number of related chats between dyad), "# of Book-Related" (number of related chats happened while reading current book), "# of Distracted" (number of distracted chats between dyad), "# of Jibo Intv" (number of Jibo interventions), "Session Elapsed" (total seconds elapsed during current interaction session), and "Book Elapsed" (total seconds elapsed while reading current book).

Finally, the annotation suggests good robot intervention timings by displaying the current chat interval relative to the thresholds. With the "Related Chat" button annotation, the dyad’s current chat interval can be calculated in real-time; whenever "Related Chat" is pressed, the interval resets to 0 and starts counting up again. As described in section 3.5.2, each dyad has high, middle, and low chat interval thresholds. Comparing the current chat interval against the thresholds, the bottom right display color-codes dyad’s relative chat frequency level: green (if lower than low threshold), light green (if between low and middle thresholds), orange (if between middle and high thresholds), and red (if higher than high
threshold). The orange and red colors imply relatively long chat intervals, and therefore they suggest good intervention timing for chat encouragement. The color display system provides real-time robot intervention timing suggestions to the human experimenters.

4.3.5 Orientation Panel

The orientation panel is composed of two parts: seating adjustment and orientation. The seating adjustment section, located on the left side of the orientation panel, allows the experimenters to input parent’s and child’s seating locations as 3 dimensional vectors (vectors (0,3,1) and (1,1,1) in Figure 4-10a example). The seating vectors are predetermined based on the dyad’s previous seating configuration, but if needed during interaction, the experimenters can manually change the configuration in the seating adjustment section.

The orientation section (Figure 4-10b) allows the experimenters to choose the target or the subject of a robot interaction. When the robot is launching an interaction behavior, it first orients itself towards the target; for example, if parent is selected in the orientation panel, Jibo first orients towards the parent’s seating and addresses its behavior towards the parent. Selecting "Parent" or "Child" sets Jibo’s target to parent or child, and selecting "Both" sets Jibo to orient towards the middle of the two.

During the triadic interaction study, the target is selected randomly between "Parent", "Child", and "Both" for each storybook page, automatically diversifying Jibo’s target choices. However, a randomly chosen target can be manually overridden by clicking buttons in the orientation section.
4.3.6 Execution Panel

The execution panel displays to-be-executed command in its text field, and provides launch and replay buttons to execute the command. The command can be loaded in the text field either by manually typing a command into the field, or by selecting a question or comment from the question and hint list panel. In Figure 4-11 example, the to-be-executed command is loaded from question list: command text "What does it mean when a rooster crows?" with Jibo TTS tag "<es cat='curious'>".

Pressing "Launch" button executes the command displayed in the text field, making Jibo speak the command text, and make expressions based on the command’s TTS tags. The TTS tags marked with "< ... >" enable various Jibo expressions ranging from Emoji display to music play. The replay button next to the "Launch" button loads the immediate previous command back into the text field display. The replay button allows the experimenters to quickly resend the command to Jibo when the dyad misses a behavior.
Figure 4-12: Question and hint list panel

Figure 4-13: Keyword and behavior list panel

Figure 4-14: Hint list panel updated by behavior button
4.3.7 Question and Hint List Panel

Each book used in the study has abundant auto-generated and manually-generated questions that are designed to serve as good dialogic reading questions. The manually-generated questions are created by educational experts, and they are rich with intellectually stimulating content and expressive Jibo TTS tags. The auto-generated questions are produced by passing each storybook sentence into paragraph-level commonsense inference model (PARA-COMET) [15], then transforming the outputted context inferences (subject, object, and relation) into usable questions via contextualization templates. After generation, all auto-generated questions are assessed by human evaluators in Amazon Mechanical Turk to ensure question quality. As a result, each auto-generated question has ratings indicating whether the question is child appropriate, educational, linguistically correct, and good overall. Using the average of the ratings, the auto-generated questions are sorted by descending quality. Assuming that the manually-generated questions have the best quality, the manually-generated questions are ranked top with highest ratings.

Whenever a dyad flips a page in the storybook app, the list of questions corresponding to the loaded page gets displayed in the question list panel (top part of Figure 4-12). Ordering of the question list is determined by sorted ratings, making the panel initially show five questions with the highest ratings. The experimenters can look at more questions with lower ratings by navigating with the arrow buttons below the list. The experimenter chooses an appropriate question based on various factors such as previous interaction context and child’s literacy level. When the experimenter selects a question from the list, the question gets loaded in the execution panel text field, ready to be executed onto Jibo with the "Launch" button. The question list panel allows Jibo to proactively ask good dialogic reading questions directly to parent and child.

Many questions in the list have hints and answer corresponding to the question. When a question is clicked from the list, the hints and answer corresponding to the selected question are loaded in the hint list panel (bottom part of Figure 4-12). When a follow up to Jibo’s question is needed, the experimenter can choose a hint or answer from the panel list. The hint list panel allows Jibo to efficiently react to the dyad’s responses.
Table 4.1: Descriptions and examples of buttons in behavior list panel

*keyword* is replaced with selected keyword (e.g. "rooster"),
*dr_type* is replaced with selected dialogic reading type (e.g. "recall"),
and *target* is replaced with selected target’s name or with "we" if "Both" is selected.

<table>
<thead>
<tr>
<th>Row</th>
<th>Button</th>
<th>Description</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Let’s Talk (kw)</td>
<td>Jibo suggests dyad to start a chat about <em>keyword</em></td>
<td>&quot;Let’s chat about <em>keyword!</em> <em>target</em>, maybe you can start the chat?&quot;</td>
</tr>
<tr>
<td></td>
<td>Ask us Q</td>
<td>Jibo suggests dyad to ask a <em>dr_type</em> question</td>
<td>&quot;<em>target</em>, I think now is a good time for you to ask us a <em>dr_type</em> question!&quot;</td>
</tr>
<tr>
<td></td>
<td>Ask us Q (kw)</td>
<td>Jibo suggests dyad to ask a <em>dr_type</em> question about <em>keyword</em></td>
<td>&quot;<em>target</em>, maybe you can ask us a <em>dr_type</em> question about <em>keyword</em>.&quot;</td>
</tr>
<tr>
<td></td>
<td>DR Q (kw)</td>
<td>Jibo asks dyad a <em>dr_type</em> question about <em>keyword</em></td>
<td>(if <em>dr_type</em> is &quot;R&quot; recall) &quot;Can you tell me what happened to <em>keyword</em>?&quot;</td>
</tr>
<tr>
<td></td>
<td>Vocab Q (kw)</td>
<td>Jibo asks dyad a vocabulary question about <em>keyword</em></td>
<td>&quot;Can you tell me what <em>keyword</em> means?&quot;</td>
</tr>
<tr>
<td>2</td>
<td>No</td>
<td>Jibo gives a negative reaction</td>
<td>&quot;I don’t think so.&quot;</td>
</tr>
<tr>
<td></td>
<td>Backchannel</td>
<td>Jibo gives a silent positive reaction</td>
<td>Jibo nods / Jibo smiles</td>
</tr>
<tr>
<td></td>
<td>Yes</td>
<td>Jibo gives a positive reaction (agreement)</td>
<td>&quot;I think so.&quot;</td>
</tr>
<tr>
<td></td>
<td>Correct</td>
<td>Jibo gives a positive reaction (affirmation)</td>
<td>&quot;I agree.&quot;</td>
</tr>
<tr>
<td></td>
<td>Praise</td>
<td>Jibo gives a positive reaction (praise)</td>
<td>&quot;Great!&quot;</td>
</tr>
<tr>
<td>3</td>
<td>Look Tablet</td>
<td>Jibo looks at tablet (as if Jibo is reading)</td>
<td>Jibo looks at tablet</td>
</tr>
<tr>
<td></td>
<td>Wait</td>
<td>Jibo tells dyad to wait (to be on current page)</td>
<td>&quot;Wait!&quot;</td>
</tr>
<tr>
<td></td>
<td>Prev Page</td>
<td>Jibo tells dyad to flip to previous page</td>
<td>&quot;Can we please go back to the previous page?&quot;</td>
</tr>
</tbody>
</table>
4.3.8 **Keyword and Behavior List Panel**

Each storybook page has a list of keywords; the keyword list includes characters, objects, and vocabularies in book text and illustration of its corresponding page. The keyword list is displayed as buttons in the keyword list panel (top part of Figure 4-13), and when a keyword is selected, it gets loaded in the text field labeled "Key Word". The keyword list panel also includes a section for dialogic reading type selection labeled "DR". The dialogic reading types are designed after the CROWD protocol described in section 2.3.1, but "C" (completion) type is excluded since completion prompts require keen timing control. Instead, "." (generic) type is added to allow general, any-type selection. The dialogic reading type is switched randomly for each storybook page, but the selection can be manually overridden with the buttons in a real-time interaction by an experimenter. The selected keyword and dialogic reading type are used when launching certain behaviors in the behavior list panel.

The behavior list panel (bottom part of Figure 4-13) is composed of buttons that provide various options for commonly used Jibo interaction behaviors. The "Call Parent" and "Call Child" buttons allow Jibo to specifically address its behavior towards parent or child. When "Call Parent" or "Call Child" is selected, Jibo calls parent’s or child’s name before speaking its launched command text.

The remaining 13 buttons in the behavior list panel all directly launch behaviors onto Jibo. Five buttons in the panel’s first row are used for robot intervention behaviors, and the remaining eight buttons are used for simple reactions and commands. Table 4.1 provides descriptions and examples of the 13 behavior buttons. Note that all behavior buttons have multiple command script options, and that the examples in Table 4.1 are only showing one script per button. In addition to launching script, clicking "Ask us Q" or "Ask us Q (kw)" updates the hint list panel to give examples and definition of the selected dialogic reading type. In Figure 4-14 example, "Ask us Q" button is clicked with "R" (recall) type, so the hint list panel provides three examples of recall questions and an explanation of recall.

Overall, the behavior buttons along with the keyword and dialogic reading type selections allow experimenters to run rich parent-child-robot triadic interaction.
Chapter 5

Parent-Child-Robot Triadic Interaction

In order to observe and investigate the impact of a robot facilitator in parent-child shared reading, a pilot study with four parent-child dyad pairs was conducted. Using the interactive reading system described in section 4, we conducted a remote triadic interaction study with the Wizard of Oz (WoZ) technique, in which a hidden human experimenter teleoperates Jibo to perform different robot strategies [34].

Jibo’s interaction behaviors are suggested by a rule-based system, but are ultimately chosen and executed by the human experimenter’s judgement. During an interaction, the teleoperation controller notifies potential intervention timings on its annotation panel (section 4.3.4) and highlights appropriate Jibo behaviors based on the interaction’s condition (section 5.1). Based on such suggestions made by the system, the human experimenter decides when and how to launch Jibo’s interaction behaviors.

There are two different robot interaction strategies that we focus on in the study: model and suggest. Regarding the model strategy, Jibo demonstrates model dialogic reading behaviors by directly asking good questions to the dyad. For the suggest strategy, Jibo coaches the dyad with suggestions, encouraging the dyad to initiate dialogic reading.

In order to investigate into the effects of triadic interaction in general and compare different robot strategies, the triadic interaction study consists of one baseline session, three experimental sessions with differing robot strategies, and one evaluation session. Section 5.1 explains the five interaction sessions and study conditions, and section 5.3 provides details of the overall study procedure.
The parent-child-robot triadic interaction study strives to achieve following goals:

(1) Data Collection: assuming that the human experimenter is launching interaction behaviors at optimal timings, we collect human input data on good robot intervention and interaction timings

(2) Study on Impact of Triadic Reading Experience on Dyadic Reading: by conducting two dyadic parent-child reading sessions before and after the dyad’s interactions with a robot facilitator, we study the impact of triadic interaction on dyads’ reading behaviors outside of sessions with robot

(3) Comparison of Different Robot Interaction Strategies: by comparing dyad’s reading behaviors and their qualitative measures in response to different robot interaction strategies, we study each strategy’s unique effects
5.1 Robot Conditions

The robot interaction conditions are designed after the "model" and "suggest" parent-child-coach interaction strategies described in section 2.3.2. The third condition "mixed" is added to allow a mix use of both model and suggest strategies. In order to study the effects of triadic reading in general, we add baseline and evaluation conditions with no robot proactively interacting as the study’s pre and post test measures. The five reading sessions always start with baseline condition and end with evaluation condition. However, the ordering of the middle three experimental sessions (with model, suggest, and mixed conditions) is randomized in order to avoid the ordering effect.

The major difference between the three conditions with proactive robot interactions is on how Jibo intervenes, or who it puts in charge of initiating conversations. In the model condition, Jibo is in charge of coming up with good questions and comments, and Jibo starts the chat. On the other hand, in the suggest condition, parent and child are in charge of initiating discussions; Jibo only suggests a topic or a prompt. Sections 5.1.1 through 5.1.4 explain each robot conditions in further detail.

Some interaction components such as Jibo’s intervention timings and its reaction behaviors remain the same across the three robot conditions. Intervention timings are decided based on dyad’s chat frequency and human judgement, and therefore are independent from the conditions. Likewise, Jibo’s overall interaction flow and its reactions remain consistent because they follow the child-robot interaction design suggestions (section 2.2) and the PEER protocol (section 2.3.1).

Inspired by Shen et al.’s design which adds non-verbal sound to more effectively catch children’s attention [40], Jibo plays a siren-like sound effect and displays a question mark Emoji on its face display before starting its intervention speech. We aim to better grab dyad’s attention with Jibo’s such multimodal intervention notifying behavior.

Following the PEER protocol, Jibo prompts ("P") dyad to be engaged with various interventions, evaluates ("E") and praises their comments and answers, and when appropriate, expands ("E") or repeats ("R") the comments with a follow up. Applying the PEER protocol to all three robot conditions allows all triadic interactions to be fun and rich, and
Figure 5-1: Teleoperation controller GUI highlighted during model condition

it highlights the differences in the contents of Jibo interaction strategies by keeping the interaction flow consistent.

5.1.1 Baseline and Evaluation Conditions

During baseline and evaluation conditions, Jibo does not proactively ask questions or make comments. Instead, it remains in its default listening behavior, only making speech when parent or child deliberately asks Jibo to participate. The baseline and evaluation conditions allow us to study the effects of triadic interaction on dyadic reading by observing how parent and child read in dyadic settings before and after the triadic reading sessions.

5.1.2 Model Condition

During model condition, Jibo demonstrates good dialogic reading behaviors by proactively asking questions and making comments. Jibo’s questions are designed after the CROWD protocol described in section 2.3.1, and they are launched by the prompts in the question list panel (section 4.3.7) and by "DR Q (kw)" and "Vocab Q (kw)" buttons in the behavior list panel (section 4.3.8). The GUI components specifically corresponding to Jibo’s model behaviors are highlighted during the model condition sessions as shown in Figure 5-1.
Following script is an example interaction during model condition:

- Jibo: "What does it mean when a rooster crows?"
  (asks a recall question from question list)

- Child: "Rooster makes a sound?" (answers)

- Jibo: "That’s Right." (affirms answer with "Correct" button)

- Jibo: "Have you ever heard a rooster crowing?" (expands with manual input)

The model condition is designed to yield more occurrences of higher quality dialogic reading since Jibo directly asks good reading prompts, yet it may yield less parent-child interactions since the condition focuses on parent-robot and child-robot interactions.

5.1.3 Suggest Condition

During the suggest condition, Jibo coaches parent or child to creatively come up with a question or a comment to initiate a discussion. Jibo’s suggesting behaviors are launched by "Let’s Talk (kw)", "Ask us Q", and "Ask us Q (kw)" buttons in the behavior list panel, and Jibo can specifically coach a dyad to ask a CROWD prompt by selecting a non-general (not "-") dialogic reading type button in the keyword panel (section 4.3.8). The buttons and panels corresponding to Jibo’s suggest behaviors are highlighted during the suggest condition sessions as shown in Figure 5-2.

Following script is an example interaction during suggest condition where Jibo makes a general suggestion:

- Jibo: “Can we start a discussion about rooster?”
  (suggests to start a chat with “Let’s Talk (kw)” button)

- Parent: “You have seen a rooster before, right?” (asks question to child)

- Child: “Yes, last week.” (answers)

- Jibo: “Nice. Do you like roosters?”
  (praises answer with “Praise” button and expands with manual input)
Following script is another example interaction during suggest condition where Jibo makes a specific suggestion to follow the CROWD protocol:

- Jibo: “Could you please ask us a recall question about rooster?”
  (suggests to ask a recall question with “Ask us Q (kw)” button)

- Parent: “I’m not sure what you mean.” (requires follow up)

- Jibo: “Here is an example. Can you tell me what happened to rooster?”
  (gives an example from hint list)

- Parent: "What happened to rooster?" (asks question to child)

- Child: “Rooster crowed.” (answers)

- Jibo: “Correct.” (affirms answer with “Correct” button)

Jibo’s suggestion behaviors are designed to prompt parent-child pairs to be in charge of initiating and carrying conversations, thereby leading to more frequent and diverse parent-child interactions. However, for dyads who are less familiar with dialogic reading prompts, we expect the suggest condition to be less effective.
5.1.4 Mixed Condition

In the mixed condition, Jibo may exhibit both model and suggest behaviors. When teleoperating the robot, the experimenter can use human judgement to freely choose an appropriate behavior from either model or suggest behaviors. None of the components in the teleoperation controller GUI is highlighted in order to allow the human experimenter to choose from all options. By using a mix of both model and suggest robot conditions, we expect the mixed condition to take the unique advantages of both model and suggest behaviors, and create a synergistic effect, facilitating rich and diverse interaction dynamics between parent, child, and robot.
Table 5.1: Descriptions of triadic interaction study participants

<table>
<thead>
<tr>
<th>ID</th>
<th>Child Gender</th>
<th>Child Age</th>
<th>Parent Gender</th>
<th>Parent Age</th>
<th>Recruitment</th>
<th>Order of Conditions</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>female</td>
<td>6</td>
<td>female</td>
<td>35</td>
<td>dyadic study</td>
<td>suggest, model, mixed</td>
</tr>
<tr>
<td>2</td>
<td>female</td>
<td>5</td>
<td>female</td>
<td>43</td>
<td>lab community</td>
<td>model, suggest, mixed</td>
</tr>
<tr>
<td>3</td>
<td>female</td>
<td>6</td>
<td>female</td>
<td>43</td>
<td>lab community</td>
<td>N/A</td>
</tr>
<tr>
<td>4</td>
<td>male</td>
<td>5</td>
<td>female</td>
<td>35</td>
<td>dyadic study</td>
<td>model, mixed, suggest</td>
</tr>
<tr>
<td>5</td>
<td>female</td>
<td>5</td>
<td>female</td>
<td>39</td>
<td>dyadic study</td>
<td>mixed, model, suggest</td>
</tr>
</tbody>
</table>

5.2 Participant Information

The dyads participating in our study were recruited from Media Lab community and from the past dyadic interaction study (section 3) mailing list. Our target recruits were families who reside in greater Boston area and have children between age 4 and 7. About 10 families volunteered to participate in our study, but at the end, five were chosen as final participants due to limitations on location and child’s age.

Because we randomize the ordering of the three robot interaction conditions (model, suggest, and mixed), each dyad has been assigned with an unique order of conditions for second through fourth reading sessions. Table 5.1 gives a breakdown of all five participating parent-child dyads, describing their demographic information, source of recruitment, and assigned ordering of robot interaction conditions.

Note that one family did not completed the entire study before the start of our analysis due to schedule conflict, and therefore was excluded from the thesis’s analysis. In total, four families completed all five reading sessions as well as interviews and surveys, and therefore only these four dyads were considered in our study analysis.

---

1 Dyad #3 did not complete all reading sessions.
Table 5.2: Descriptions of triadic interaction study procedures

<table>
<thead>
<tr>
<th>Procedure</th>
<th>Time</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Questionnaire Survey</td>
<td>Anytime during study (10-20 min)</td>
<td>Survey capturing dyad’s socio-demographic and relational profiles</td>
</tr>
<tr>
<td>Pre-Study Interview</td>
<td>Before reading sessions (20-30 min)</td>
<td>Phone call asking parent about her initial preferences and opinions on social robots</td>
</tr>
<tr>
<td>5 Reading Sessions</td>
<td>15-30 min each</td>
<td>Parent-child-robot story time with different robot conditions</td>
</tr>
<tr>
<td>5 Post-Session Surveys</td>
<td>After each reading session (1-5 min)</td>
<td>Short survey capturing parent’s sentiment and perception after each session</td>
</tr>
<tr>
<td>Post-Study Interview</td>
<td>After reading sessions (20-30 min)</td>
<td>Phone call asking parent about her overall triadic interaction experience and final opinions on social robots</td>
</tr>
</tbody>
</table>

5.3 Study Procedure

After recruitment, the interactive reading system stations (Figure 4-1) were deployed to the recruited families, allowing the dyads to participate the study from their homes. When not participating in reading sessions, Jibo remained in leisure mode, allowing families to freely and privately enjoy features and entertainments that Jibo commercially provides [29].

Each family completed a questionnaire survey, a pre-study interview, five reading sessions, five post-session surveys, and a post-study interview, which are described with an overview in Table 5.2. All families completed the entire study procedure within two or three weeks by participating in a session once every one to three days. The parents voluntarily scheduled the sessions based on their availability.

The questionnaire survey and the interviews are designed to provide detailed comprehension on the participants. The questionnaire survey captures dyads’ demographic and relational profiles, and the two pre-study and post-study interviews provide qualitative insight into dyads’ opinions and perceptions. The detailed descriptions, results, and discussions of the study’s questionnaire survey and qualitative interviews are provided in section 5.4.1 and section 6.1 respectively.

---

2 The study was conducted during COVID-19, so there was difficulty in hosting participants in lab.
Each reading session proceeds as follows. Before a session, parent and child sit in front of the station, staying within the station’s webcam view. When they are ready, they notify the experimenter via text. Then, the experimenter switches Jibo to story time mode and launches “Jibo Intro” behavior (section 4.3.1) to start the session.

Throughout the session, the experimenter observes the interaction through the observation panel (section 4.3.3). More examples of the experimenter’s view on dyads are shown in Figure 5-3. While observing the dyad and their tablet, the experimenter makes appropriate annotations on the dyad’s interactions using the annotation panel (section 4.3.4); the experimenter marks “Related Chat” whenever the dyad has on-task or story-related chats, and marks “Distraction” whenever the dyad has irrelevant chats. Section 5.4.2 describes how the annotations are used in analyzing dyads’ transcripts and extracting reading features.

In addition to making annotations, the experimenter executes Jibo interaction behaviors throughout the session following the protocols specific to each robot condition. As described in section 5.1, the interactive reading system highlights and suggests different protocols for each condition, allowing the experimenter to easily teleoperate Jibo to execute appropriate behaviors that are specific to the condition.

Each reading session ends after approximately 15 to 25 minutes, when the dyads finish reading two to three books of their choice. When the reading session ends, the experimenter notifies the end of the session with “Jibo Sess End” button (section 4.3.1), terminates and saves the recordings, and switches Jibo back to leisure mode. After each reading session, the parents are expected to complete a short post-session survey for self-reported evaluation. The results and details of the post-session surveys are provided in section 6.2.
5.4 Feature Extraction

Using similar methods from the dyadic interaction analysis (sections 3.1 and 3.2), we extracted survey features and reading features from the triadic interaction study’s questionnaire surveys and interaction transcripts. Sections 5.4.1 and 5.4.2 describe and summarize the features extracted from the triadic interaction study.

5.4.1 Survey Features

The triadic study’s questionnaire survey is a modified version of the survey used in the DAMI-P2C dyadic interaction study (section 3.1). The questionnaire survey outputs various survey features that capture families’ demographic information, exposure to English, home literacy environment, parenting stress, and child’s temperament with SES, ENG, HLE, PSI, and CBQ questionnaires respectively.

For CBQ, PSI, and HLE questionnaires, the same set of questions and scoring methods from the DAMI-P2C dyadic interaction study are used as described in sections 3.1.1, 3.1.3, and 3.1.5. PRQ questionnaire (section 3.1.2) is dropped because its scores are highly correlated to PSI scores as shown in section 3.3, and because the PRQ questionnaire is very long and time consuming.

Additional questions are added to SES (section 3.1.4) to capture more demographic information on the dyads, and a separate ENG category is created to focus on features related to dyad’s exposure to English. The new SES questionnaire additionally asks for parent’s gender and age, and moves the three features related to English ability (use of English at home, English proficiency, and years in USA) to ENG category. As a result, the triadic study’s SES outputs five survey features.

ENG questionnaire asks for parent’s English proficiency, use of English at home, parent’s years in USA, parent’s first language, and child’s first language. Use of English at home, parent’s first language, and child’s first language are recorded as binary responses (1 for English used at home and English is first language). Parent’s English proficiency and years in USA are recorded with likert scales between 0 and 2, and between 0 and 6 respectively. As a result, ENG questionnaire outputs five survey features.
### Table 5.3: Summary of triadic study survey features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>CBQ Total</strong></td>
<td>4.65</td>
<td>0.74</td>
<td>3.58</td>
<td>5.31</td>
</tr>
<tr>
<td>Surgency</td>
<td>4.19</td>
<td>0.43</td>
<td>3.75</td>
<td>4.75</td>
</tr>
<tr>
<td>Negative Affect</td>
<td>4.03</td>
<td>0.50</td>
<td>3.40</td>
<td>4.53</td>
</tr>
<tr>
<td>Effortful Control</td>
<td>4.96</td>
<td>0.99</td>
<td>3.92</td>
<td>6.00</td>
</tr>
<tr>
<td><strong>PSI Total</strong></td>
<td>86.00</td>
<td>7.07</td>
<td>76</td>
<td>91</td>
</tr>
<tr>
<td>Distress (PD)</td>
<td>27.75</td>
<td>8.30</td>
<td>21</td>
<td>39</td>
</tr>
<tr>
<td>Dysfunction (PCDI)</td>
<td>27.25</td>
<td>2.36</td>
<td>24</td>
<td>29</td>
</tr>
<tr>
<td>Difficult Child (DC)</td>
<td>31.00</td>
<td>6.98</td>
<td>23</td>
<td>40</td>
</tr>
<tr>
<td><strong>SES</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Child’s Gender</td>
<td>0.75</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Child’s Age</td>
<td>5.25</td>
<td>0.50</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Parent’s Gender</td>
<td>1.00</td>
<td>-</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Parent’s Age</td>
<td>38.00</td>
<td>3.83</td>
<td>35</td>
<td>43</td>
</tr>
<tr>
<td>Parent’s Education</td>
<td>2.50</td>
<td>0.58</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td><strong>ENG</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>English at Home</td>
<td>0.50</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Parent’s English Proficiency</td>
<td>1.25</td>
<td>0.96</td>
<td>0</td>
<td>2</td>
</tr>
<tr>
<td>Parent’s Years in USA</td>
<td>4.50</td>
<td>1.73</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Parent’s First Language</td>
<td>0.50</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>Child’s First Language</td>
<td>0.75</td>
<td>-</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td><strong>HLE</strong></td>
<td>0.76</td>
<td>0.17</td>
<td>0.52</td>
<td>0.89</td>
</tr>
</tbody>
</table>

As a result, the questionnaire survey yields 19 survey features in total for each dyad. The statistical summary of the survey features extracted from the four triadic study participants are provided in Table 5.3. Looking at the participating families’ questionnaire responses, our participants lack diversity in terms of gender; all parents are females, and only one child is male. However, the participants’ profiles vary in other areas, especially in their exposure to English represented by ENG.

The triadic study’s survey features are generally in line with the DAMI-P2C survey features except child’s gender. This indicates that the four pilot families are not heavily skewed away from the 30+ DAMI-P2C families, and therefore are relatively well sampled in terms of their socio-demographic and relational profiles.
Table 5.4: Summary of triadic study baseline reading features

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mean</th>
<th>SD</th>
<th>Min</th>
<th>Max</th>
<th>Skew</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat : Read of Parent</td>
<td>1.02</td>
<td>0.83</td>
<td>0.21</td>
<td>2.13</td>
<td>0.83</td>
</tr>
<tr>
<td>Chat : Read of Child</td>
<td>3.53</td>
<td>4.09</td>
<td>0.41</td>
<td>9.53</td>
<td>1.75</td>
</tr>
<tr>
<td>Chatting Duration</td>
<td>12.64</td>
<td>4.32</td>
<td>7.35</td>
<td>16.17</td>
<td>-0.54</td>
</tr>
<tr>
<td>Chatting Frequency</td>
<td>0.03</td>
<td>0.00</td>
<td>0.03</td>
<td>0.03</td>
<td>0.85</td>
</tr>
<tr>
<td>Child : Parent in Reading</td>
<td>0.86</td>
<td>1.31</td>
<td>0.01</td>
<td>2.78</td>
<td>1.75</td>
</tr>
<tr>
<td>Child : Parent in Chatting</td>
<td>0.55</td>
<td>0.29</td>
<td>0.21</td>
<td>0.92</td>
<td>0.39</td>
</tr>
<tr>
<td>Child : Parent in Initiation</td>
<td>0.52</td>
<td>0.12</td>
<td>0.38</td>
<td>0.67</td>
<td>0.11</td>
</tr>
<tr>
<td>Turn Taking Rate in Reading</td>
<td>0.10</td>
<td>0.12</td>
<td>0.02</td>
<td>0.28</td>
<td>1.59</td>
</tr>
<tr>
<td>Turn Taking Rate in Chatting</td>
<td>0.21</td>
<td>0.09</td>
<td>0.13</td>
<td>0.32</td>
<td>0.32</td>
</tr>
<tr>
<td>Word Rate of Parent</td>
<td>1.77</td>
<td>0.74</td>
<td>1.06</td>
<td>2.71</td>
<td>0.67</td>
</tr>
<tr>
<td>Word Rate of Child</td>
<td>1.28</td>
<td>0.45</td>
<td>0.79</td>
<td>1.75</td>
<td>-0.08</td>
</tr>
</tbody>
</table>

5.4.2 Reading Features

In order to capture the characteristics of dyadic parent-child interaction within the triadic interaction, especially during the baseline and evaluation conditions, the 11 reading features used in the dyadic interaction analysis (section 3.2) were extracted again. Note that the “Related Chat” and “Distraction” annotations collected throughout the reading sessions were matched to dyads’ interaction transcripts to more accurately filter out off-task distractions and label each interaction as reading or on-task chatting (dialogic reading).

Because the robot facilitator does not interact during the baseline condition, the reading features extracted from the baseline condition (Table 5.4) are reflective of our participants’ baseline dyadic reading behaviors. Comparing the triadic study participants’ baseline reading features against DAMI-P2C reading features, the triadic study participants yielded slightly better baseline reading behaviors with more child participation and turn takes in both reading and chatting. Our triadic study participants, however, had lower chatting ratio for children, implying that the pilot study children were generally more actively participating in reading than in chatting.

The 11 reading features on parent-child interactions were extracted for all five reading sessions, and the results of the extracted parent-child interaction reading features are explained and summarized in section 6.3.1.
In addition to the 11 reading features that characterize parent-child interactions, we extracted “robot behavior reading features” which capture the effects of each robot behavior on its facilitated triadic interaction.

The robot behavior feature extraction process is conducted as follows. We first disregard simple robot reactions such as “Yes” and “Correct”, and only focus on deliberate robot questions, comments, and follow ups that are designed to facilitate dialogic reading. Each deliberate robot behavior is then categorized as “manual”, “model”, or “suggest” based on its type or source; manual behaviors are generated by the human experimenters via the execution panel’s text field (section 4.3.6), model behaviors are launched by the controller components corresponding to the model condition (highlighted in Figure 5-1), and suggest behaviors are launched by the components corresponding to the suggest condition (highlighted in Figure 5-2).

For each deliberate robot behavior, the behavior’s reading features, or its effects on triadic reading, are captured with the time spent and the parent-child-robot turns taken during the chats that are facilitated by the specific robot behavior. A block of chat following a robot’s behavior is considered to be facilitated by the leading robot behavior, and as a result, a robot behavior’s reading features are defined by its following chat block’s duration and number of turn takes.

Below is an example of a parent-child-robot interaction sequence with annotations on chat blocks and robot behaviors. Note that the example is a simplified and shortened version of an actual interaction that occurred during the triadic interaction study:

Chat block 1 (facilitated by model robot behavior):

- Jibo: "Have you ever said no to a friend when they asked for your help?"  
  (model behavior)
- Child: "No."

Chat block 2 (follow up facilitated by manual robot behavior):

- Jibo: "Really? You must be a really nice friend."
  (manual behavior)
• Parent: "Mm-hmm. You helped me plant the garden last week."
• Child: "Yeah."
• Robot: "Nice!" (reaction)
• Child: "Now, you will help me carry my bag back because it’s heavy."
• Parent: "Yes."

In the above example, the model robot behavior yields one child-robot turn take, and its duration is the time spent on the first chat block. Likewise, the manual robot behavior which facilitates the follow up yields five turn takes in total, and its duration corresponds to that of the second chat block. Note that the simple robot reaction ("Nice!") in the second chat block is not a deliberate robot behavior, and therefore is considered as a part of the interaction facilitated by the manual behavior.

As in the provided example, manual robot behaviors are mostly launched to facilitate follow up conversations, elaborating on dyads’ short closed-ended responses. Occasionally, when the model and suggest behavior options provided by the reading system are not applicable, or when the experimenters want to launch context-specific intervention, manual robot behaviors are used to initiate dyad discussions.

The results collected from the triadic study’s robot behavior reading features are summarized and explained in section 6.3.2.
Chapter 6

Triadic Interaction Study Results

By analyzing the results collected from the various study procedure components, we provide both qualitative and quantitative insights into the impact of triadic reading experience on dyadic reading, and into the effects of different robot interaction strategies.

The baseline and evaluation conditions’ post-session survey results and reading features, along with the pre-study and post-study interview responses, indicate the influence of the overall triadic reading experience on the participants’ perceptions and reading behaviors. The results corresponding to the three experimental conditions allow us to compare the outcomes of different robot interaction strategies, studying each strategy’s unique effects on the participants.

Sections 6.1 through 6.3 summarize and explain the results from the pre-study and post-study interviews, the post-session surveys, and the reading features.
6.1 Pre-Study and Post-Study Interview Results

Pre-study and post-study interviews are designed to provide qualitative insights into parent-child-robot interactions. By asking parents their initial and final opinions on social robots and on triadic reading, we comprehend how the triadic reading sessions have influenced the families’ views and preferences. In addition, by asking parents to recap and evaluate the five reading sessions after the entire study, we collect verbal descriptions of families’ perceived triadic reading experience. The protocols used for pre-study and post-study interviews are listed in Appendix A.1 and A.2 respectively.

6.1.1 Pre-Study Interview

The pre-study interviews captured participating parents’ initial expectations on parent-child-robot reading and their pre-study dyadic reading habits. According to the pre-study interview responses, all four parents did not start with a clear imagination of how a social robot would join their parent-child story time. When challenged to picture ideal parent-child-robot interactions, they imagined a social robot answering and asking questions, and offering entertainment features such as music play and dance. Two families who use smart home devices such as Alexa and Google Home expected an ideal social robot to have smart-device-like features.

In terms of the participants’ typical parent-child dyadic reading experience, most parents read books often with their children, and they generally enjoyed their current dyadic reading experience. All participants valued “quality time” or “time together” the most in shared reading. However, the children’s reading participation styles varied a lot across families due to the wide range in their reading abilities; some children liked letting their parents read the entire story, while some children liked to try to read the stories on their own. Such interview response is in line with the dyads’ baseline reading features (section 5.4.2) where the children’s reading participation varied a lot with a large standard deviation value.

In general, the participating parents all responded positively about their pre-study dyadic reading experience, except one parent who recently has not been able to spend as much time on reading with her child.
6.1.2 Post-Study Interview

During the post-study interviews, parents were asked to recap and elaborate their overall triadic interaction experience. When asked if they felt any differences across the five reading sessions, all parents felt a big difference between the sessions with proactive Jibo (model, suggest, and mixed) and the sessions without proactive Jibo (baseline and evaluation). However, they did not feel much difference among the three experimental conditions with proactively interacting Jibo.

All parents enjoyed the triadic reading experience overall; they thought the reading sessions with Jibo were fun, engaging, facilitative, and helpful. One parent mentioned that “Jibo [helped them] to pause and think what just happened [in the story]”, and another parent mentioned that “even just the thought of [a] robot being more interactive or asking questions . . . made it more fun for [the child]”. In addition, many parents noted and liked the sound effect playing before each robot intervention because they could pause and expect Jibo’s intervention without having their speech disturbed.

When asked to mention dislikes, parents generally wanted more diversity in books and in Jibo behaviors so that they can customize the interaction to their child’s reading ability and interest. Some parents thought certain Jibo questions were too difficult or too obvious, yet they noted that the questions, regardless of their quality, still lead them to initiate good parent-child discussions.

One parent mentioned that Jibo’s interventions felt unnatural due to inappropriate timing. Although the intervention timings were carefully controlled by the human experimenters, Jibo’s interventions did sometimes overlap with the dyads’ speech when Jibo experienced one or two seconds of lag due to network issues. With the sound effect feature, most dyads did not note or mind the overlap, but the particular dyad noted the issue with Jibo’s intervention timing and perceived it as disturbing. This is partially because there were difficulties in finding smooth intervention timings when interacting with dyads who liked to focus on reading without many pauses. The particular dyad’s feedback on Jibo’s timing issue highlights the importance and need of adaptive and personalized robot intervention timings tailored towards each dyad’s reading style.
6.1.3 Pre-Study to Post-Study Improvements

In order to study the macro effects of triadic reading, we explicitly asked parents if the triadic interactions with Jibo effected their parent-child dyadic readings outside the study’s reading sessions. In general, the parents did not feel a drastic change in their dyadic readings, but some gave very positive feedback. One parent mentioned that the triadic reading sessions “made [her] want to make [reading] more fun for [the child] again; not just reading because it’s [a] routine, [but] reading because it’s fun and [connecting]”. Furthermore, another parent mentioned that the triadic sessions made her “[feel] a little bit more self-conscious about [having] more conversations [during reading]”.

In order to gauge pre-study to post-study improvements in families’ preferences, the same preference ratio question was asked in both pre-study and post-study interviews. During both interviews, the parents were asked to hypothetically divide ten reading sessions based on their preferences among the following three variations: parent-child, parent-child-robot, and child-robot. A split of $x$-$y$-$z$ would indicate that the parent prefers to do $x$ parent-child, $y$ triadic, and $z$ child-robot reading sessions proportionally.

Before the study, two parents gave even 3-3-3 splits, and one parent gave 3-2-5 split since she wanted her child to spend more time independently. Note that one parent’s answer was excluded since she did not understand the question clearly.

After the study, the two parents who originally gave even splits yielded opposite results; one ended up preferring parent-child reading more, yielding a 4-3-3 ratio, while the other preferred reading with robot more, yielding a 2-4-4 ratio. The parent who originally gave 3-2-5 ratio decided to give 4-4-2 ratio in post-study since she enjoyed the triadic reading and wanted more connection with her child through shared reading.

Overall, all parents liked the triadic reading experience with Jibo, and wanted to keep it in their final preference ratios. However, the parents also noted that there are irreplaceable qualities that are unique to parent-child reading such as the bond, and therefore wanted to give parent-child reading a fair share in their splits.
6.2 Post-Session Survey Results

After each reading session, parents completed a short post-session survey. For the baseline and experimental conditions, a post-session survey asking questions specific to the precedent reading session was used (Appendix A.3.1). On the other hand, for the evaluation condition, a modified post-session survey asking about the overall triadic reading experience was used (Appendix A.3.2). The evaluation condition used a modified survey because Jibo did not proactively interact during evaluation, making the scores specific to the evaluation sessions not reflective of Jibo’s performance.

With three self-reported rating questions, the post-session surveys captured parents’ satisfaction on overall experience, perceived Jibo’s intelligence, and perceived Jibo’s likeability. All self-reported ratings were in likert scales of 5, and the rating questions were designed after the robotic social attribute scales (RoSAS) presented in Carpinella et al.’s work [7]. Using RoSAS adjectives, Jibo’s perceived intelligence was recorded in incompetent to competent scale, and its likeability was recorded in unfriendly to friendly scale.

One additional rating question was added to the post-evaluation-session survey in order to explicitly ask how the triadic reading sessions have affected the participants’ dyadic reading experience. The question was answered on a scale of "very negative effect" (-2) to "very positive effect" (+2), where 0 is "no effect".

Furthermore, with two multiple-selection questions, the post-session surveys captured specific Jibo features that the parents’ liked and disliked about each reading session. The multiple-selection answers were translated into robotic social attributes to provide further information on Jibo’s perceived intelligence, likeability, and animacy (Appendix A.3.3).

6.2.1 Self-Reported Ratings

The parents’ self-reported ratings on overall experience, Jibo’s intelligence, and Jibo’s likeability let us understand parents’ perceptions and sentiments towards the triadic reading experience and on Jibo. All self-reported scores were generally high, remaining greater than or equal to 3 in all five reading sessions. The high scores indicate that the dyads’ did not have negative experience or sentiment throughout all sessions, even during the baseline
dyadic reading sessions. This observation supports the participants’ interview responses since most parents were already satisfied with their dyadic reading and enjoyed the triadic reading experience with Jibo.

Note that in Figure 6-1, the baseline score is reflective of parents’ initial ratings on not proactive Jibo, the experimental score averages the ratings on the three experimental conditions to reflect parents’ perceptions on proactive Jibo, and the final evaluation score reflects parents’ general perception on Jibo after completing all sessions.

The experimental and evaluation scores in Figure 6-1 are similar because both scores reflect parents’ general perceptions on proactive Jibo. However, there is a difference between baseline and experimental scores on perception towards Jibo, especially towards its intelligence. The difference indicates that the parents perceived the proactive Jibo in the experimental conditions more intelligent compared to the baseline, not proactive Jibo. Such observation is in line with the parents’ interview responses, in which they noted a big difference between proactive and not proactive Jibo interactions.

As shown in Figure 6-2, the parents gave relatively even intelligence and likeability ratings across all experimental conditions, implying that they developed similar perceptions when interacting with proactive Jibo, regardless of its interaction strategy. The observation in self-reported scores is in line with the parents’ interviews, as they generally did not note much difference across the three interactive robot conditions.
Figure 6-3: Self-reported scores over sessions where first and last sessions are baseline and evaluation, and middle three sessions are experimental conditions in random order

The lack of significant rating difference across the experimental conditions may have been caused by the ordering effect. As shown in Figure 6-3, which plots parents’ self-reported scores against progress in sessions, parents’ perceptions on Jibo, especially on its likeability, increased over sessions, as they spent more time with Jibo. Note that on third experimental session, all families gave perfect score (5) on Jibo’s likeability. This implies that the interactions with Jibo had a positive effect on families’ perceptions and sentiments.

Finally, based on the post-evaluation-session survey responses answering the effect of triadic reading on dyadic reading, parents reported that the triadic reading sessions with Jibo had some positive effects on their dyadic parent-child readings (+1 on average). This rating result, along with the positive feedback received during the post-study interviews, suggest that the triadic interaction has yielded successful pre-study to post-study improvements in dyads, positively influencing their reading behaviors outside the study.

6.2.2 Liked and Disliked Jibo Features

The multiple-selection questions asked parents to select specific Jibo features that they liked and disliked about each reading session. Note that the multiple-selection questions were not asked after the evaluation sessions since Jibo did not proactively participate in the evaluation sessions, not offering features to be assessed on.
Figure 6-4: Counts of liked and disliked Jibo features

(a) Total counts on liked versus disliked Jibo features for each condition

(b) Break down of liked Jibo features

(c) Break down of disliked Jibo features

Figure 6-4a provides total counts of liked and disliked Jibo features and compares them for each condition. Figures 6-4b and 6-4c break down the total counts by translating each Jibo feature into its corresponding robotic social attribute category.

With the information on liked and disliked Jibo features, we get a more detailed view into parents’ self-reported scores. In line with the observation made on the self-reported scores, parents liked more features with the proactive Jibo than with the not proactive baseline Jibo. The self-reported scores did not vary much across the three proactive experimental conditions, but with the multiple-selection responses, we observe a difference across the three conditions. Parents liked more Jibo features during the mixed condition than in the singular model or suggest condition, which is reflective of the fact that the mixed condition offers the most diverse range of Jibo features and behaviors.
6.3 Reading Feature Analysis Results

The reading features extracted from the interaction transcripts using the methods explained in section 5.4.2 were analyzed to capture the behaviors in parent-child reading and in robot-facilitated interactions. Section 6.3.1 analyzes parent-child interaction reading features, focusing on dyadic reading behaviors’ baseline to evaluation changes, and their differences across different experimental conditions. Section 6.3.2 analyzes the reading features related to robot behaviors, studying the effects of each robot condition and behavior type on its facilitated triadic interactions.

6.3.1 Parent-Child Interaction Reading Features

The extracted parent-child interaction reading features (section 5.4.2) were compared across different robot conditions. Note that the descriptions and the baseline results of the parent-child interaction reading features are provided in Table 3.2 and Table 5.4.2 respectively.

In order to study the effects of triadic reading experience on dyadic reading, we first measured baseline to evaluation improvements using the two conditions’ reading features. The average values of the reading features, along with the average percentage of change \((\text{evaluation} - \text{baseline}) / \text{baseline}\) are shown in Table 6.1.

According to Table 6.1, the children’s chat-to-read ratio ("Chat : Read of Child") greatly increased in the evaluation condition while their participation ratio in reading ("Child : Parent in Reading") decreased; this observation implies that the triadic reading sessions lead the children to be more actively participating in chatting than in reading. In addition, in the evaluation condition, the chatting duration increased while the frequency decreased slightly; this indicates that the dyads yielded longer, slightly less frequent chats after the triadic reading experience.

As a result, in terms of baseline to evaluation improvements in reading features, the triadic interaction had positive influence on children’s chatting behaviors, but not much on their reading and parents’ behaviors. The lack of influence on reading behaviors makes sense because the triadic reading interaction design focuses on promoting dialogic reading, or parent-child chats, and does not target the dyads’ book text reading behaviors.
Table 6.1: Parent-child interaction reading features of baseline and evaluation conditions

Δ% is average percentage of change from baseline to evaluation condition
green and red colors indicate over 50% and -50% change respectively

<table>
<thead>
<tr>
<th>Feature</th>
<th>Baseline</th>
<th>Evaluation</th>
<th>Δ %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat : Read of Parent</td>
<td>1.02</td>
<td>0.60</td>
<td>-20.9 %</td>
</tr>
<tr>
<td>Chat : Read of Child</td>
<td>3.53</td>
<td>6.32</td>
<td>292.4 %</td>
</tr>
<tr>
<td>Chatting Duration</td>
<td>12.64</td>
<td>15.74</td>
<td>30.4 %</td>
</tr>
<tr>
<td>Chatting Frequency</td>
<td>0.03</td>
<td>0.03</td>
<td>-16.7 %</td>
</tr>
<tr>
<td>Child : Parent in Reading</td>
<td>0.86</td>
<td>0.07</td>
<td>-57.9 %</td>
</tr>
<tr>
<td>Child : Parent in Chatting</td>
<td>0.55</td>
<td>0.54</td>
<td>1.7 %</td>
</tr>
<tr>
<td>Child : Parent in Initiation</td>
<td>0.52</td>
<td>0.51</td>
<td>10.3 %</td>
</tr>
<tr>
<td>Turn Taking Rate in Reading</td>
<td>0.10</td>
<td>0.03</td>
<td>-43.8 %</td>
</tr>
<tr>
<td>Turn Taking Rate in Chatting</td>
<td>0.21</td>
<td>0.18</td>
<td>-8.2 %</td>
</tr>
<tr>
<td>Word Rate of Parent</td>
<td>1.77</td>
<td>2.05</td>
<td>24.1 %</td>
</tr>
<tr>
<td>Word Rate of Child</td>
<td>1.28</td>
<td>1.49</td>
<td>36.9 %</td>
</tr>
</tbody>
</table>

Table 6.2: Parent-child interaction reading features of experimental conditions

Δ% is average percentage of change from baseline to experimental condition
green and red colors indicate over 50% and -50% change respectively

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mixed</th>
<th>Model</th>
<th>Suggest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chat : Read of Parent</td>
<td>1.51</td>
<td>1.81</td>
<td>2.59</td>
</tr>
<tr>
<td>Chat : Read of Child</td>
<td>8.99</td>
<td>7.80</td>
<td>13.87</td>
</tr>
<tr>
<td>Chatting Duration</td>
<td>21.24</td>
<td>22.84</td>
<td>27.99</td>
</tr>
<tr>
<td>Chatting Frequency</td>
<td>0.03</td>
<td>0.03</td>
<td>0.03</td>
</tr>
<tr>
<td>Child : Parent in Reading</td>
<td>0.15</td>
<td>0.32</td>
<td>0.22</td>
</tr>
<tr>
<td>Child : Parent in Chatting</td>
<td>0.64</td>
<td>0.60</td>
<td>0.70</td>
</tr>
<tr>
<td>Child : Parent in Initiation</td>
<td>0.69</td>
<td>0.49</td>
<td>0.75</td>
</tr>
<tr>
<td>Turn Taking in Reading</td>
<td>0.07</td>
<td>0.06</td>
<td>0.05</td>
</tr>
<tr>
<td>Turn Taking in Chatting</td>
<td>0.21</td>
<td>0.21</td>
<td>0.23</td>
</tr>
<tr>
<td>Word Rate of Parent</td>
<td>1.80</td>
<td>1.69</td>
<td>1.97</td>
</tr>
<tr>
<td>Word Rate of Child</td>
<td>1.36</td>
<td>1.22</td>
<td>1.41</td>
</tr>
</tbody>
</table>
In addition to studying the baseline and evaluation reading features, we also analyzed the three experimental conditions’ reading features. The parent-child reading features of mixed, model, and suggest conditions, and their percentage of change relative to baseline \((\text{condition} - \text{baseline})/\text{baseline}\) are provided in Table 6.2.

Looking at Table 6.2, all three experimental conditions with proactive robot behaviors significantly increased the dyads’ chatting ratio and duration. The children’s chat-to-read ratio was heavily influenced, being increased by over 500%, 180%, and 800% in mixed, model, and suggest conditions respectively. Such observation indicates that the dyads, especially children, spent more portions of their time on chatting relative to reading with proactive Jibo. Furthermore, the features related to child’s participation, such as child-to-parent ratio in reading, chatting, and initiation, increased during the experimental conditions, especially during the mixed and suggest conditions. Note that the suggest condition was most effective in increasing child’s participation and turn takes in chatting.

Overall, having proactive robot behaviors during shared reading positively influenced dyads’ dialogic reading behaviors and children’s participation. The suggest strategy, which prompts the dyad to creatively initiate conversations, was especially effective in improving parent-child dyadic reading features. Note that turn takes in reading and word rates were not heavily considered in the reading feature analysis because they were more correlated to dyads’ English abilities and literacy skills than to robot conditions.

The reading features’ results were generally in line with the qualitative observations made by the human experimenters when executing the study. The manual robot behaviors customized to the exact context yielded best interactions and reactions from the dyads. With the suggest behaviors, the experimenters observed more parent-child interactions and more occurrences of personalized discussions. However, some dyads lacked diversity in questions and topics when challenged to initiate chats; for example, some parents would resort to vocabulary questions when initiating chats. Furthermore, with the model behaviors, the experimenters observed more straightforward discussions that demand less cognitive load from the parents. As a result, based on the qualitative observations, the mixed condition with both model and suggest robot behaviors yielded best balance between having creative personalized contents and avoiding heavy dependence on parents.
Table 6.3: Robot behavior types in different robot conditions

<table>
<thead>
<tr>
<th>Feature</th>
<th>Mixed</th>
<th>Model</th>
<th>Suggest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Robot Behaviors</td>
<td>23.75</td>
<td>25.00</td>
<td>23.25</td>
</tr>
<tr>
<td>Divided By Total Time</td>
<td>0.018</td>
<td>0.015</td>
<td>0.019</td>
</tr>
<tr>
<td>Robot Behavior Types</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percentage of Manual Behavior</td>
<td>54.74%</td>
<td>44.00%</td>
<td>54.84%</td>
</tr>
<tr>
<td>Percentage of Model Behavior</td>
<td>29.47%</td>
<td>56.00%</td>
<td>-</td>
</tr>
<tr>
<td>Percentage of Suggest Behavior</td>
<td>15.79%</td>
<td>-</td>
<td>45.16%</td>
</tr>
</tbody>
</table>

6.3.2 Robot Behavior Reading Features

In order to study the effects of different robot strategies on reading, we analyzed and compared the robot behavior features extracted from the interaction transcripts. As explained in section 5.4.2, the robot behaviors can be categorized into three types (manual, model, and suggest) based on their source, and only the mixed condition uses all three behavior types since the model and suggest conditions only use two (manual and condition-specific).

Table 6.3 provides the average total number of robot behaviors launched during each condition, along with a breakdown of each condition’s robot behavior types. Note that the total number of robot behaviors divided by total time reports the average number of robot behaviors launched per second. Both the total number of robot behaviors and its values divided by time were consistent across the conditions, indicating that the experimenters launched relatively even number of robot interaction behaviors across all experimental sessions. Furthermore, looking at the behavior type breakdown, roughly half of the robot behaviors were manually generated, regardless of the condition. During the mixed condition, the model behaviors were used slightly more than the suggest behaviors.

Looking at the robot behavior reading features (chat duration and turn takes) averaged on each robot condition (Table 6.4), the model condition yielded longer chats while the suggest condition yielded more turn takes in chats, indicating that the different robot conditions had varying effects on triadic reading behaviors. However, because each condition used a mix of different behavior types, the results averaged on each condition may not be directly reflective of each robot behavior type’s effect.
Table 6.4: Robot behavior reading features in different robot conditions

<table>
<thead>
<tr>
<th>Feature</th>
<th>Robot Condition</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mixed</td>
</tr>
<tr>
<td>Duration of Robot-Facilitated Chat</td>
<td>18.40</td>
</tr>
<tr>
<td>Turn Takes in Robot-Facilitated Chat</td>
<td>4.99</td>
</tr>
<tr>
<td>Number of Parent-Child Turns</td>
<td>3.04</td>
</tr>
<tr>
<td>Number of Parent-Robot Turns</td>
<td>1.22</td>
</tr>
<tr>
<td>Number of Child-Robot Turns</td>
<td>0.73</td>
</tr>
</tbody>
</table>

Table 6.5: Robot behavior reading features in different robot behavior types

<table>
<thead>
<tr>
<th>Feature</th>
<th>Robot Behavior Type</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Manual</td>
</tr>
<tr>
<td>Duration of Robot-Facilitated Chat</td>
<td>22.46</td>
</tr>
<tr>
<td>Turn Takes in Robot-Facilitated Chat</td>
<td>5.45</td>
</tr>
<tr>
<td>Number of Parent-Child Turns</td>
<td>3.37</td>
</tr>
<tr>
<td>Number of Parent-Robot Turns</td>
<td>1.33</td>
</tr>
<tr>
<td>Number of Child-Robot Turns</td>
<td>0.75</td>
</tr>
</tbody>
</table>

As a result, in order to more deliberately show the effects of each robot behavior type, the robot behavior reading features across all experimental conditions were averaged on each robot behavior type. Based on the results summarized in Table 6.5, manually launched robot behaviors yielded the best triadic reading behaviors with longest duration and most turn takes. Such observation implies that customized robot questions and comments tailored to specific conversational context facilitate best parent-child-robot interactions.

Comparing the features in model and suggest conditions, the model condition resulted in more parent-robot and child-robot turn takes while the suggest condition resulted in more parent-child turn takes. The result is in line with our interaction strategy design since the model behavior makes the robot directly start interactions with the dyad, facilitating parent-robot and child-robot interactions, while the suggest behavior makes the robot coach the dyad to start conversations, encouraging parent-child chats.
Chapter 7

Discussions and Conclusion

As part of the thesis, we designed and implemented a novel parent-child-robot interaction paradigm to encourage constructive parent-child conversations during shared reading. Through the triadic interaction study, we applied and tested our interaction paradigm with four families, studying the effects of both the general triadic reading experience and the specific robot interaction strategies on the dyads’ reading behaviors and sentiments.

Various critical remarks have been gained from the rich set of results collected with the qualitative interviews, the self-reported survey responses, and the reading features. In addition to the analyzed results, there have been many insightful observations made by the human experimenters during their process of running the triadic reading sessions. The following sections 7.1 through 7.3 summarize and organize both analyzed and human-observed results based on their correspondence to each goal of the triadic interaction study.
7.1 Data Collection on Intervention Timings

Most robot interventions were launched in accordance with the reading system’s suggestions made based on each dyad’s chat interval thresholds. However, because the participating children did not vary much in age, the reading system ended up suggesting similar chat intervals for all participating dyads. No dyad noted the interventions to be too frequent or infrequent, implying that the system’s suggested interval was generally appropriate.

In most cases, an intervention was launched at the end of a storybook page, when parent and child paused to digest the page that they had just completed reading. Launching interventions at the end of a page worked well with most dyads, but was not optimal for some dyads who liked to focus on reading with a fast pace. As mentioned in the post-study interview results, having the multimodal notification feature with a sound effect and emoji display was very helpful in yielding smooth, nondisruptive interventions, yet there were still occurrences of the robot’s speech overlapping with the dyads’ speech.

In addition, throughout the study, there were many occurrences in which a robot intervention was very effective in retrieving dyad’s, especially child’s, engagement. Because children were generally excited about interacting with a social robot, the robot facilitator’s interventions were more effective in retrieving attention compared to parent’s attempts.

Such observations made on the triadic interaction study’s robot intervention timings highlight the strength of appropriate robot interventions, as well as the importance and need of adaptive and personalized robot intervention timings.
7.2 Impact of Triadic Reading on Dyadic Reading

The experience of having the rich triadic shared reading interactions generally had a positive effect on dyads’ reading behaviors and on their perceptions towards social robots. On average, after completing the triadic reading sessions, the children participated more in constructive chatting, and the parents perceived Jibo more competent and friendly.

Most importantly, the results from the qualitative interviews and the post-session surveys indicated that the triadic reading experience had a positive effect on the participants’ dyadic reading outside the study. Such result implies that our triadic interaction paradigm has been able to train the dyads, extending its effects onto their personal lives. The observed effects of the triadic reading experience on dyadic reading suggest that the triadic interactions have a potential in making long-term improvements on dyads’ shared reading.
7.3 Comparison of Different Robot Interaction Strategies

The thesis majorly focused on comparing the two robot interaction strategies: model and suggest. Although the model and suggest strategies were not perceived very differently based on the dyads’ interviews and the post-session surveys, they resulted in unique effects in terms of the dyads’ reading behaviors.

The different effects of the two robot interaction strategies were in line with our design intentions; the model strategy, which is designed to promote direct interactions between the dyad and the robot, yielded more parent-robot and child-robot turn takes, while the suggest strategy, which is designed to encourage parent-child chats through coaching, yielded more parent-child turn takes.

Furthermore, based on the qualitative observations made during the execution of the study, the suggest strategy worked very well for some dyads because they gave the dyads freedom to have conversations tailored to their personal interests. However, the suggest behaviors were sometimes challenging for the dyads, as they were not always able to come up with an appropriate question or topic. As a result, in some interaction scenarios, the model condition, which directly provides rich and diverse contents, was more optimal in facilitating productive parent-child interactions.

In addition, according to the results gathered by the robot behavior reading features, the manually executed robot interaction behaviors were most effective in facilitating constructive interactions because they could be keenly tailored to the story and interaction context. Such reading feature results along with the qualitative observations again highlight the need of adaptive and personalized interaction strategy and content.
7.4 Conclusion

Based on the qualitative and quantitative insights gathered from the various aspects of the triadic interaction study, the interaction contents, strategies, and timings all work best with detailed adaptation, indicating that customization must be heavily considered in all areas of parent-child-robot interaction designs.

Our parent-child-robot interaction paradigm has yielded promising results in the pilot triadic interaction study, and therefore we believe that there are many potentials in multi-party human-robot-interaction that are yet to be discovered.

As future work, additional triadic interaction studies can be conducted to support and verify the thesis’s preliminary results. In addition, taking dyads’ suggestions, more books, contents, and robot behaviors can be added to the interactive reading system to provide more diversity, offering a wider range of features for better customization.
Appendix A

Survey and Interview Protocols for Triadic Interaction Study

A.1 Pre-Study Interview Protocol

Dyadic Reading Experience (Pre-Study)

(1) (Explain study’s objective) We are designing some educational features on Jibo that potentially would help parent-child story time to be more fun and productive. So, we would like to first learn about your experience reading books with your child.

(2) Can you describe a typical story reading time between you and your child?
   (Follow up) How long does it last? Do you talk a lot? Do you take turns often?

(3) What do you value the most in parent-child reading?
   (Follow up) Having fun, having dialogues, learning vocabulary words, bonding etc.

(4) How satisfied are you with your current story time experience?
   (Follow up) Do you love reading books with your child? Or, is it a burden that you have to do but not necessarily enjoy?
Initial Opinions on Social Robot

(1) Have you gotten a chance to play with the robot, Jibo?
   (If yes) How much time have your family spent interacting with Jibo so far?

(2) (Introduce the idea of social robot) Jibo is a perfect example of a social robot. We
   would like to know your initial thoughts on Jibo or social robots in general and in the
   context of parent-child-robot interactions.

(3) Imagine an ideal robot which is capable of everything and is not limited by any
   technological barriers. What do you imagine the ideal "social robot" to be?
   (Follow up) What roles does it serve or what features does it have?
   (Follow up) What kind of activities do you see it doing with you and your child?

Initial Opinions on Parent-Child-Robot Story Time

(1) (Explain scenario) Now, imagine the ideal social robot coming in and participating
   in your parent-child story time.

(2) How would you feel having a robot joining the story time?
   (Follow up) How comfortable or uncomfortable would you feel?

(3) What would be the ideal reading interaction between the three of you?
   (Follow up) What roles do you want the robot to play?

Initial Preferences on Different Interactions

(1) Out of 10 reading sessions, how many would you like to read just with your child,
   how many would you like to read with your child and robot, and how many would
   you like your child to read with robot without you?
   (Follow up) How would you like to divide up the 10 reading sessions among the 3
   conditions: parent-child, parent-child-robot, and child-robot?
A.2  Post-Study Interview Protocol

Triadic Interaction Overview

(1) (Recap) We had five reading sessions in total. Do you remember how Jibo was acting during each session?
   (If yes) Could you please briefly describe Jibo’s actions for each session?
   (If no - give a brief summary)

(2) Did you feel any differences between the five sessions? What were the differences?

(3) Which session was your favorite and which was your least favorite? Why?

(4) How was the general story time experience with Jibo?
   (Follow up) How did you feel? What did you like or dislike about it?

Dyadic Reading Experience (Post-Study)

(1) Nowadays, what do you value the most in parent-child reading?

(2) Recently, how satisfied are you with your story time experience?

(3) Did the reading sessions with Jibo influence the story time outside of the sessions?
   (If yes) How did your personal story time change?

Final Opinions on Social Robot and Triadic Story Time

(1) How did your initial thoughts on social robots change after the sessions with Jibo?
   (Follow up) Were there behaviors that you did not expect or were different from your imagination?

(2) (Explain scenario) Imagine an ideal robot which is capable of everything and is not limited by any technological barriers. Now, imaging the ideal robot participating in your parent-child story time.

(3) How would you feel having a robot joining the story time?
(4) What would be the ideal reading interaction between the three of you?

(5) What are some things that we should consider when designing parent-child-robot reading interactions?

**Final Preferences on Different Interactions**

(1) Out of 10 reading sessions, how many would you like to read just with your child, how many would you like to read with your child and robot, and how many would you like your child to read with robot without you?
A.3 Post-Session Survey Questionnaire Form

Baseline and experimental conditions use the general post-session survey (Section A.3.1), and evaluation condition uses the modified post-evaluation-session survey (Section A.3.2). Section A.3.3 maps Jibo features in multiple-selections to robotic social attributes.

A.3.1 Post-Session Survey

(1) How was today’s story time experience overall?
   Answer: 1 (terrible) - 5 (great / perfect)

(2) How helpful or competent was Jibo during today’s story time?
   Answer: 1 (not helpful at all / incompetent) - 5 (very helpful / competent)

(3) How friendly was Jibo during today’s story time?
   Answer: 1 (very unfriendly) - 5 (friendly)

(4) What did you like about today’s story time with Jibo?
   Answer: (multiple selection)

   • Jibo’s questions or comments effectively facilitated reading
   • Jibo asked interesting questions
   • Jibo made good comments
   • Jibo made natural or organic speech
   • Jibo made natural or organic nonverbal behaviors
   • Jibo was responsive to us and understood us well
   • Jibo made our shared reading more engaging and fun
   • Jibo proactively engaged in shared reading with us
   • Jibo was a good story listener
   • Other (free response)
(5) What did you NOT like about today’s story time with Jibo?

Answer: (multiple selection)

• Jibo’s questions or comments disrupted reading
• Jibo asked irrelevant questions
• Jibo made irrelevant comments
• Jibo’s speech felt mechanical and distracting
• Jibo’s nonverbal behaviors felt mechanical and distracting
• Jibo was not responsive and did not understand us well
• Jibo made our shared reading less engaging and fun
• Jibo was too proactive
• Jibo was not a good story listener
• Other (free response)

A.3.2 Post-Evaluation-Session Survey

Please only consider the reading sessions when Jibo was proactively interacting.
Exclude the first and last sessions with not interactive Jibo.

(1) How was the overall story time experience with Jibo?

Answer: 1 (terrible) - 5 (great / perfect)

(2) How helpful or competent was Jibo during the story time?

Answer: 1 (not helpful at all / incompetent) - 5 (very helpful / competent)

(3) How friendly was Jibo during the story time?

Answer: 1 (very unfriendly) - 5 (friendly)

(4) How have the triadic reading sessions affect your parent-child reading outside the study without Jibo?

Answer: -2 (very negatively) - +2 (very positively)
where 0 is no effect
A.3.3 Translation of Jibo Features to Social Attributes

(1) Intelligence

• Jibo facilitated / disrupted reading
• Jibo asked good / bad questions
• Jibo made good / bad comments

(2) Likeability

• Jibo made reading more / less fun
• Jibo was proactive / too proactive

(3) Animacy

• Jibo’s speech was organic / mechanical
• Jibo’s nonverbal behaviors were organic / mechanical
• Jibo was responsive / not responsive
• Jibo was a good / bad listener
Bibliography


