



Spatiotemporal Aspects of Engagement during Dialogic Storytelling Child–Robot Interaction

Scott Heath^{*†}, Gautier Durantin[†], Marie Boden, Kristyn Hensby, Jonathon Taufatofua, Ola Olsson, Jason Weigel, Paul Pounds and Janet Wiles

School of Information Technology and Electrical Engineering, The University of Queensland, Brisbane, QLD, Australia

OPEN ACCESS

Edited by:

Bilge Mutlu,
University of Wisconsin-Madison,
United States

Reviewed by:

Fulvio Mastrogiovanni,
Università di Genova, Italy
Erol Sahin,
Middle East Technical University,
Turkey

*Correspondence:

Scott Heath
scott.heath@uq.edu.au

[†]These authors have contributed
equally to this work.

Specialty section:

This article was submitted
to *Humanoid Robotics*,
a section of the journal
Frontiers in Robotics and AI

Received: 13 January 2017

Accepted: 06 June 2017

Published: 22 June 2017

Citation:

Heath S, Durantin G, Boden M,
Hensby K, Taufatofua J, Olsson O,
Weigel J, Pounds P and Wiles J
(2017) Spatiotemporal Aspects of
Engagement during Dialogic
Storytelling Child–Robot Interaction.
Front. Robot. AI 4:27.
doi: 10.3389/frobt.2017.00027

The success of robotic agents in close proximity of humans depends on their capacity to engage in social interactions and maintain these interactions over periods of time that are suitable for learning. A critical requirement is the ability to modify the behavior of the robot contingently to the attentional and social cues signaled by the human. A benchmark challenge for an engaging social robot is that of storytelling. In this paper, we present an exploratory study to investigate dialogic storytelling—storytelling with contingent responses—using a child-friendly robot. The aim of the study was to develop an engaging storytelling robot and to develop metrics for evaluating engagement. Ten children listened to an illustrated story told by a social robot during a science fair. The responses of the robot were adapted during the interaction based on the children’s engagement and touches of the pictures displayed by the robot on a tablet embedded in its torso. During the interaction the robot responded contingently to the child, but only when the robot invited the child to interact. We describe the robot architecture used to implement dialogic storytelling and evaluate the quality of human–robot interaction based on temporal (patterns of touch, touch duration) and spatial (motions in the space surrounding the robot) metrics. We introduce a novel visualization that emphasizes the temporal dynamics of the interaction and analyze the motions of the children in the space surrounding the robot. The study demonstrates that the interaction through invited contingent responses succeeded in engaging children, although the robot missed some opportunities for contingent interaction and the children had to adapt to the task. We conclude that (i) the consideration of both temporal and spatial attributes is fundamental for establishing metrics to estimate levels of engagement in real-time, (ii) metrics for engagement are sensitive to both the group and individual, and (iii) a robot’s sequential mode of interaction can facilitate engagement, despite some social events being ignored by the robot.

Keywords: engagement, human–robot interaction, storytelling, social robotics, immediacy

1. INTRODUCTION

As robots become more prevalent in our lives, it is important to design robots that can share spaces with humans and to evaluate how robots can engage in social interactions. A robot’s social abilities will affect whether the robot is allowed into spaces occupied by humans, as well as the types of tasks that the robot will be trusted to perform. A key ability for social robots is that of establishing and maintaining human engagement.

Engagement during a social interaction is defined as a combination of attention and understanding of this interaction (Tomasello et al., 2005). Building social robots that can maintain user engagement has been recognized as one of the main challenges of human–robot interaction (HRI) (Sidner et al., 2005), in particular when interacting with children in a learning context (Walters et al., 2008 and Ioannou et al., 2015). A major component of HRI affecting engagement is the level of *immediacy* of the interaction (Kennedy et al., 2015). *Immediacy behaviors* are defined as “... those which increase the sensory stimulation between two interaction partners” (Mehrabian, 1968), where high immediacy is implemented as a greater number of socially contingent responses from the robot.

The level of immediacy of a social robot can be varied by using different forms of responsiveness to the user’s actions (Yanco and Drury, 2004). *Open-loop* modes of interaction correspond to the lowest level of immediacy, where the robot executes scripted actions without processing any of the user’s actions. In contrast, higher levels of immediacy are achieved when the robot implements *closed-loop* control by processing and adapting to user inputs. User inputs can either be invited by the robot at certain times (termed a *synchronous* mode of interaction) or provided whenever the user wants (an *asynchronous* mode of interaction).

Robot storytelling presents a benchmark challenge for creating engaging interactions, where different levels of immediacy can be used. As with other interactions, the level of immediacy during storytelling can be controlled through open-loop scripted responses, or closed-loop responses with synchronous or asynchronous modes of interaction. Closed-loop storytelling is also called “dialogic” storytelling (Whitehurst et al., 1988) and requires the storyteller to contingently respond and change how the story is delivered based on children’s reactions. Rather than considering the child as a passive listener, the aim of this approach is to give an active role to the child by initiating richer interactions with them. Dialogic storytelling by human storytellers has been shown to increase the level of engagement in children during storytelling (Whitehurst et al., 1988 and Mol et al., 2008). We hypothesize that implementing dialogic storytelling using robots has the potential to increase the level of engagement in a similar way.

In their study of the interaction between a robot and preschoolers Ioannou et al. (2015) used dancing, moving, and storytelling activities. They noted that the engagement of children remained high during dancing and moving activities that exhibited higher levels of interactivity and were enriched with emotions and gestures. In contrast, the storytelling activity was performed in an open-loop mode of interaction and used few gestures and emotions, which resulted in disengagement of the children. This study demonstrated the effect of different levels of immediacy on engagement in different types of human–robot interaction. Similarly, the majority of previous attempts at building robots capable of telling stories have used open-loop modes of interaction (e.g., Mutlu et al. (2006); Gelin et al. (2010); and Fridin (2014)), where the motions, utterances, and emotions displayed by the robot were not dependent on user inputs. Few studies have implemented closed-loop (dialogic) storytelling with synchronous interaction by allowing a human to program

or trigger robot actions through a control interface (Ryokai et al., 2009 and Kory, 2014) and by measuring the location of the user in the space surrounding the robot at specific points in time (Pitsch et al., 2009). Both methods resulted in good levels of engagement. Finally, to the authors’ knowledge, only one storytelling study can be considered both closed-loop and asynchronous. The approach used measurements of engagement estimated in real-time from brain signals to modify the behavior of the storytelling robot (Szafrir and Mutlu, 2012), which also resulted in high levels of immediacy and comprehension of the story.

In each implementation of closed-loop storytelling robots, enabling interactivity required preliminary programming of the robot by the user (Ryokai et al., 2009), control by an experimenter (Kory, 2014), or invasive measures of engagement (electroencephalography, see Szafrir and Mutlu (2012)) to manipulate the responses of the robot. None of these solutions are suitable for a robot capable of interacting with children in a public space with a high level of autonomy.

In the OPAL project, we aim to build a child-sized robot (Opie) that is capable of socially interacting in public spaces through a variety of activities including storytelling. Opie is inspired by the RUBI project at UCSD (Malmir et al., 2013) and is designed to be a safe social robot for children that encourages the use of haptic modalities such as touching, leaning on, and interacting in close proximity. In the current study, we implement and evaluate dialogic storytelling using Opie in a public setting (a science fair). The specific aims of our study are to explore reaction times (through the modality of touch), touch patterns, and location of the child in space around the robot during the course of a dialogic storytelling interaction. We aim to explore the following research questions:

1. *What level and duration of engagement can Opie facilitate?*
2. *How individual or stereotypical are the spatial and temporal reactions across different participants?*
3. *How do the patterns of spatial and temporal reactions relate to Opie’s synchronous behavior?*

These research questions will be explored by studying the patterns that are present within the different spatial and temporal reactions and what these patterns show. The location and task create a challenging context for gaining, maintaining, and estimating engagement. In order to create a responsive robot for a public location, we require methods to evaluate engagement that are non-invasive, provide high temporal resolution, and have the potential to be automated. As both immediacy and proxemics have been argued to play a role in user engagement (Mehrabian, 1972), we aim to develop methods for evaluating engagement that are based on temporal and spatial features.

In this paper, we present an implementation of dialogic storytelling using synchronous inputs of touch, which were designed to maintain the engagement of the children during the course of the interaction. The impact of dialogic storytelling on the engagement of children was evaluated while they took turns interacting with the robot during a science fair. The children’s behavior was monitored by a video camera and by recording their actions. Analysis focused on spatial and temporal features

(such as response times, patterns of touches and of motion in the space of the robot) extracted from these non-invasive recordings to provide estimates of children's engagement during the interaction.

2. MATERIALS AND METHODS

2.1. Robot

The robot platform used in this study is the child-friendly robot Opie, designed as a social robot for social interaction with children across different modalities. Opie is the result of a multidisciplinary, iterative design process (Wiles et al., 2016) and was previously used for interaction with children to investigate language (Heath et al., 2016) and elements of spatial proximity such as touch patterns (Hensby et al., 2016 and Rogers et al., 2016). Opie is intended to explore how robots can be used to facilitate social tasks, such as educating, conversing, or playing. Opie's torso, head, arms, and single neck actuator are intended to enable social functions.

Opie's torso is manufactured from soft materials to facilitate interaction through touch and increase safety. Opie's torso and child-sized stature were designed to make the robot appear friendly. The shape of Opie and the surrounding cushions and mat contribute to the creation of a safe area for children to occupy in front of the robot (see **Figure 1A**). Opie incorporates two tablets, one mounted on the head and one mounted on the torso. The 8-inch head tablet displays animated eyes that are capable of moving and expressing emotions. Opie's head is tilted slightly forward to help create an inviting space in front of the robot. The 12-inch torso tablet displays media and runs a speech synthesizer, while allowing children to interact through touch. The inclusion of two tablets allows the face tablet to be dedicated to social interactions. This version of Opie contains a single actuator in the neck, which allows Opie's head to yaw

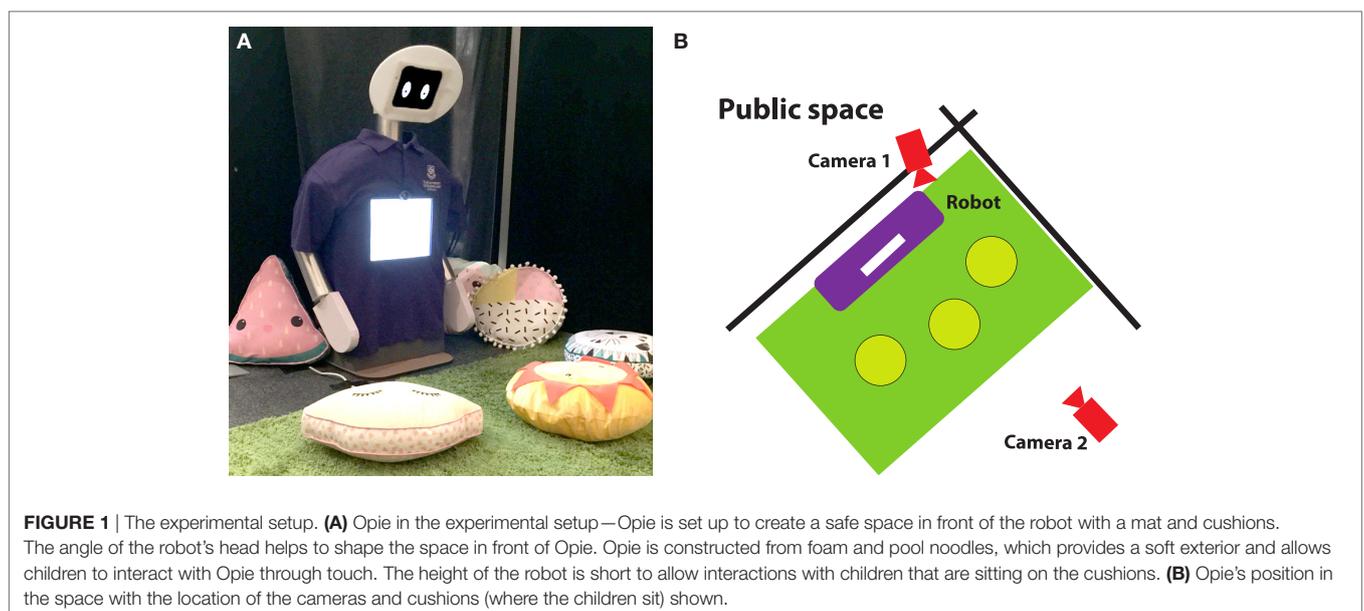
left and right around the neck. Opie also has arms, which rotate around the shoulder, but are not actuated. The robot's behavior can be controlled both by games that run on the torso tablet or by a Wizard of Oz (WoZ) through a phone interface. In this study, robot behaviors were autonomous during storytelling.

Opie's other electronics include a router that enables wireless information transfer between the robot parts and a Raspberry Pi running core server software. All the components are integrated using the robot operating system (ROS) middleware (Quigley et al., 2009), which allows the tablets, neck motors, and the WoZ phone to communicate with each other. The software running on the head and torso tablets is written using the Unity game engine and uses the Android native text-to-speech API to tell the story.

Opie was installed at a science fair in Brisbane (Australia) in a 1 m × 2 m space delimited by rugs. The robot had two separator panels of approximately 2 m height on the back and left sides and separating the robot's area from other activities at the science fair. The space around the robot was monitored by two video cameras: one facing the front of the robot at a distance of 1.5 m and the other attached to the separator panel above the robot's head and looking down at the mat from behind the robot (see **Figure 1B**).

2.2. Storytelling Game

A storytelling game was designed for Opie's torso tablet. The aim of the game developed for this study was to present an interactive narrative combined with a simple object finding task for children to perform. The storytelling game was built to facilitate (i) presentation of narrative content to children, (ii) the robot responding to touches on the tablet (allowing dialogic storytelling), (iii) a temporal measure of engagement based on touches, and (iv) expression of emotions from the robot that accompany narration. The storytelling game consisted of the presentation of a



scene and an accompanying narrative. Each of the three scenes within the game consisted of a background, a target animal, and several non-target elements (called distractors). Each of the scenes was designed to be as similar as possible in terms of difficulty. A target animal was presented on the torso tablet first, while Opie named and described the animal, and explained why the target animal was visiting that scene. The first level would then start.

2.2.1. Levels

Each scene within the game consisted of six levels of increasing difficulty. Each time a level started, Opie said a sentence about the target animal running away and asked the child to find that animal. When the child selected the target animal, the level ended and the next level would begin. The task (finding the animal) was then repeated with increased difficulty as the target became smaller or partially occluded in each successive level.

2.2.2. Overlay

Every time Opie started a level or a child pressed an object in a scene, Opie used an “overlay” to decrease the saliency of the scene or increase the saliency of an object. The *overlay* is a semi-transparent black rectangle that is used to increase the saliency of objects in the scene relative to the rest of the scene by darkening all other objects and the background. The event of adding or removing the overlay to the scene was significant, as in addition to changes in saliency, it designated when Opie started or stopped reacting to children’s touches. Opie did not respond to touches when the overlay was displayed.

2.2.3. Attentional Countermeasures

If 10 s elapsed since the last detected touch or utterance from Opie an “attentional countermeasure” was presented. An *attentional countermeasure* consisted of Opie telling the child that help was needed and reiterating to the child to find the target animal.

2.3. Ethics

Testing of the robot at the science fair was approved by a local ethics committee. Parents provided consent for their child’s participation in the study, and experimenters engaged parents prior to the children entering Opie’s space. The consent form was completed on an iPad and also included an optional media release consent. Parents were able to stay with their children during the study, either watching from behind the child or sitting with their child in front of the robot.

2.4. Procedure

The procedure of the study consisted of three phases—an introductory phase (which required the intervention of a human facilitator and WoZ), a storytelling phase (which was completely autonomous), and then a quiz phase (conducted by the human experimenters). During the entire procedure, the role of the human facilitator was to familiarize the children with the robot and supervise the interaction without taking part in it. As all robot behaviors were autonomous during the storytelling game, the only role of the WoZ was to trigger the start of the story.

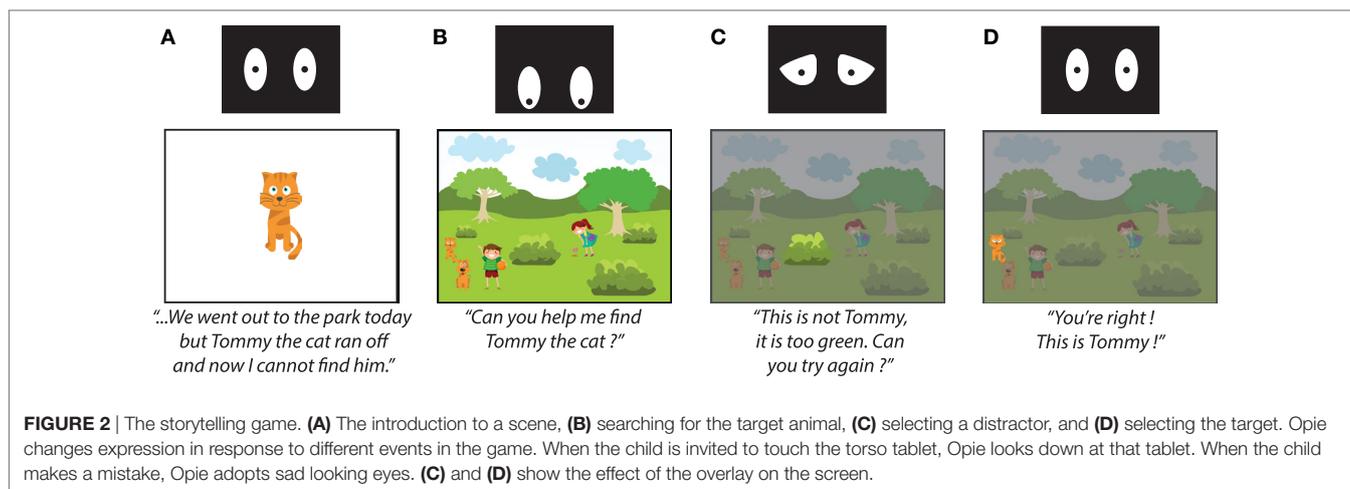
2.4.1. Introductory Phase

After obtaining consent, a human facilitator took up to three children and their parents over to Opie and introduced them to the robot. Any additional children had to wait until the end of the current interaction (out of sight). The children were encouraged to sit down on a cushion each. The facilitator started a pregame consisting of colored shapes displayed on Opie’s torso tablet. This pregame was designed to familiarize the children with the robot and prime them to touch Opie’s tablet during the storytelling game. The facilitator encouraged the children to touch the shapes. Upon touching a shape, that shape would become salient for a short period by darkening the rest of the screen, and then the shapes would return to their initial colors. After each child touched the screen once, the WoZ would press a button to begin Opie’s storytelling.

2.4.2. Storytelling Phase

During the storytelling phase, Opie would run the storytelling game. The facilitator remained next to the robot for the storytelling game and would select a child to take a turn if more than one child was present. During each story the facilitator did not interrupt the story or the interaction. The children remained sitting in front of Opie for the storytelling phase. The storytelling game proceeded as follows (see **Figure 2**):

- Opie began with a narrative while the children sat and listened. The torso tablet initially was blank.
 - Opie introduced itself and verbally greeted the child/children.
 - Opie presented an image of the target animal with a white background onto the torso tablet and introduced the story.
 - Opie then showed the first scene on the torso tablet while continuing to narrate.
- The children’s active participation (dialogic storytelling) began when Opie verbally asked them to find the target animal by name.
- The child would choose an object by touching the object on the torso tablet. The touched object would become salient by darkening the rest of the scene.
 - If the object was a distractor, then Opie would tell the child that the object they had selected was not the target, briefly describe the distractor, and then ask the child to try again. The darkening would then be removed.
 - If the object was the target animal, then the robot would congratulate the child and the game would move to the next level.
 - If the child did not choose an object during a given time limit (10 s), Opie would attempt to regain attention using an attentional countermeasure by telling the child that help was needed and asking them to find the target again.
- When changing level, the screen would be briefly darkened again, and Opie would tell the child that the animal had “... run away again.” The target animal was repositioned in the scene to increase the difficulty and the process was repeated from step 2. The process was repeated an additional five times (six levels per scene).



- For each child present (up to three children), Opie would change to the next scene and repeat the process from step 1c. The start of a new scene from the beginning was the only action controlled by the WoZ.

2.4.3. Quiz Phase

After the experiment, each child's comprehension of the story was estimated by asking a series of questions. One experimenter asked the set of questions to each child present. The questions were presented on additional tablets (one per child) and consisted of recognizing visual and audio content. Children were asked to visually identify the main character and the background scene of the story from sets of five pictures and identify the name of the main character from a set of five names read out by the experimenter. After the questions were answered, the study was complete. The order of the options for each question was shuffled across participants.

2.5. Data Analysis

The behavioral data (touches, performance, and spatial movements around the robot) measuring the interaction with the first story (cat story) of ten participants (five males and five females; mean age = 54.4 months; SD = 13.7 months) were collected and analyzed.

2.5.1. Touch Patterns

Screen touch data were collected using ROS logging functionality (rosviz) and processed using Matlab. The location of each touch on the screen was recorded as well as the touch duration. Touches were automatically classified into four types:

- **target touches**, when the child touched the target object (i.e., the cat);
- **distractor touches**, when the child touched another object (distractor object) in the scene;
- **background touches**, when the child touched the background of the picture (which did not trigger any reaction from the robot); and
- **overlay touches**, when the child touched the overlay.

Each of these different touch types was expected to give different information about the interaction. Target touches suggest that the child is understanding and completing the task given by the story. Distractor touches suggest that the child understands part of the task, but is not able to find the correct object. Background touches suggest that the child does not understand the task at all. Overlay touches suggest that the child does not understand the synchronous interaction mode of the robot and that it is not possible to interrupt the robot during this time. Touch patterns were analyzed by comparing the percentage of touches that were classified as each of these four types.

2.5.2. Spatial Movement of Children

Spatial position data were extracted from the camera looking down at the scene from behind Opie's head, in order to characterize the motions of the children in the space surrounding the robot. The relative position of the child with respect to the robot was extracted from the video every 2 s, using the center of the child's forehead. Due to the noise contained in the video data, this extraction was performed manually using Manual Video Analysis (MVA) software. The data were then used to create spatial heatmaps for each participant to look at the area the participant occupied during the study. The distance between the child and the robot over the course of the interaction was computed from the spatial position data (in pixels). We applied a linear regression model with robust fitting options in Matlab (*fitlm* function) to estimate the direction of evolution of child proximity during the interaction. The significance of the linear fit was estimated by applying an analysis of variance to the model (Matlab *anova* function).

2.5.3. Quiz Data

Quiz data were aggregated for each participant to give a score out of three. Data were also aggregated for each question to give the number of participants that answered correctly so that the questions could be compared against each other to better understand what elements of the story the children recalled best.

3. RESULTS

Out of the ten participants, nine approached the robot individually, while the remaining participant was part of a group of three (P1—see spatial results in Section 3.2 and **Figure 9** for differences in position). The storytelling interaction lasted on average 373 s (6 min 13 s; SD = 96 s), from the start of Opie narrating a scene (Opie's first utterance) until the end of the scene (Opie changing scenes) (see **Figure 4** for a typical interaction). During the interaction children frequently looked at the face of the robot, which indicated that they were attending to the social functions of the robot. The robot only used attentional countermeasures (when the children did not touch the screen for more than 10 s) with two participants (respectively, one and six countermeasures used). For the quiz at the end of the story children got 2.1 questions correct on average (SD = 0.99); with nine correctly identifying who the main character of the story was; four correctly identifying the name of the character; and eight correctly indicating the place where the story was located. A χ^2 test on these data showed that participants recalled the name of the character significantly less than its appearance ($\chi^2 = 5.5$; $p < 0.05$ after correction for comparisons across the three quiz questions).

All participant response times during the story (except one outlier) tended to converge to a stable value of level duration, after decreasing from a maximum value for the first level (see **Figure 3**).

3.1. Touch Patterns

Touches of the children mainly focused on the targets of the story. Touch data show the salience of zones containing a target at a

moment of the story compared to other areas of the picture (see **Figure 5**). On average, 91.1% of the time spent touching the screen was in areas containing targets. In addition, touches in areas containing targets lasted on average 456.2 ms (SD = 276.4 ms) and were significantly longer than touches outside of these areas (average = 43.3 ms; SD = 16.6 ms; Mann-Whitney U $p < 0.01$; adjusted $Z = 2.87$). The significant amount of extra time that child spent touching targets indicates that children engaged with the task of finding the target.

Despite the salience of the target zones, target touches represented only 63.1% of the touches observed during the experiment. All participants completed the story and, therefore, did exactly six target touches during the story. On average, participants also did 2.90 overlay touches (SD = 1.41). Among the 29 overlay touches, 24 were measured while Opie was speaking (with 23 of them being in the first 5 s of Opie's utterance). Only three participants did background touches (five touches in total), and one participant did one distractor touch.

A variety of touch behaviors were observed during the study (see **Figure 6**). There is a concentration of the overlay and background touches at the beginning of the story (scenes one and two) (82.3% of the total number of overlay and background touches, see **Figure 7**), representing interaction opportunities missed by the robot at these instants. There is also a contrast between the first two levels of the story (with a large concentration of touches that would not trigger a robot reaction) and the rest of the interaction (with less touches not triggering reactions and longer target touches).

The temporal dynamics of the interaction between the child and the robot shows changes across the levels: the interquartile interval of the level durations appears to increase in the last

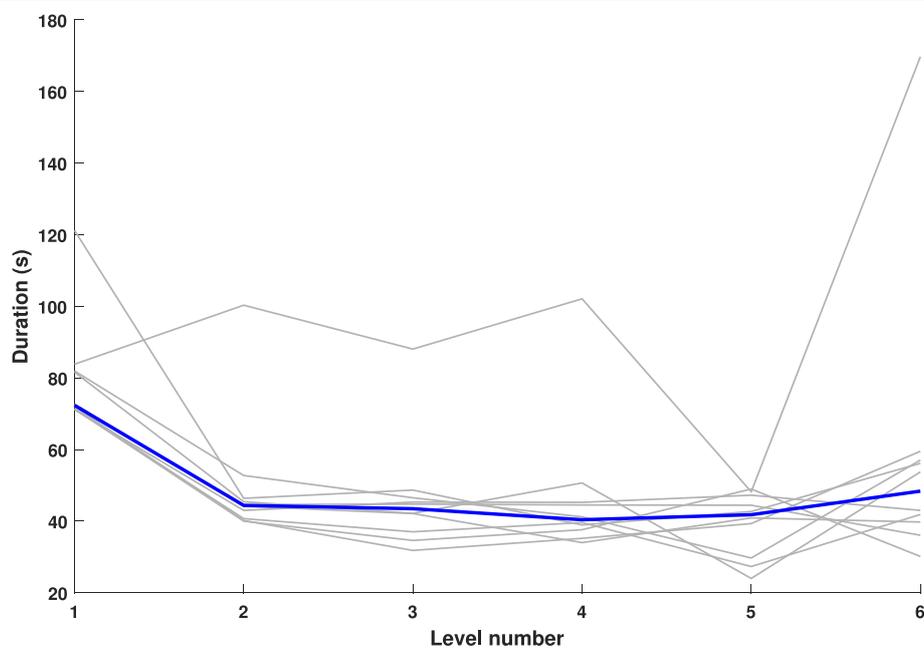


FIGURE 3 | Level duration (in seconds) per participant. The gray curves represent individual data. The blue curve shows the median across participants.

three levels, together with a stabilization of the median value of duration (see **Figure 7**). Finally, the large area covered by the interval between the third and fourth quartiles of response times for levels one, two, three, four, and six emphasizes the presence of an outlier (cf. **Figure 3**) for response times.

Among the small number of overlay touches when the robot was not talking, three were located in the fastest 25% of trials (first quartile) and four were below the median time taken (i.e., happening on shorter trials) (see **Figure 7**). Similarly, half of the background and distractor touches were located in the longest 25% of trials (fourth quartile) and four out of six were above median time taken (i.e., happening on longer trials).



FIGURE 4 | A typical interaction between Opie and a child. The child presses Opie's torso tablet within the interactive story.

3.2. Spatial Motion

Spatial position heatmaps (see Section 2.5.2) were extracted from the camera view to show preferred locations of children over time (see **Figures 8** and **9**). All the locations corresponded to a distance of less than 1 m away from the robot (the participants stayed on the rugs), and all children that approached the robot alone (from P2 to P10) stayed on the left-hand side of the robot. The participant that was part of a group of three—P1—had no outlying temporal behavior but was an outlier in spatial motion due to the constraints caused by the other children in the area. In particular, P1 was constrained to the right side of the robot, while all the other participants approached on the left.

For seven out of ten participants, there was a statistically significant decreasing linear trend for the distance, suggesting that the children got closer to the robot over time (see **Figure 10**). All participants' distances exhibited a large variance over time, due to the back and forth motions between active participation and listening to the story.

4. DISCUSSION

In this article, we describe an implementation of dialogic storytelling on a child-friendly robot (Opie), based on interaction with the robot through touch during a story. The interaction was implemented in a closed-loop synchronous way, as the robot invited the children to touch the screen at some moments and did not process touch inputs the rest of the time. We explored the interaction of children with the robot in a science fair environment and measured the time and space aspects of children's engagement with the robot, based on their response time, touch patterns, and motions in the peripersonal space of the robot.



FIGURE 5 | Average time spent touching the different areas of the picture, showing the focus of children's attention on the target of the scene (cat character). Note that the targets from all six levels are shown here, while a child only sees one target in each level.

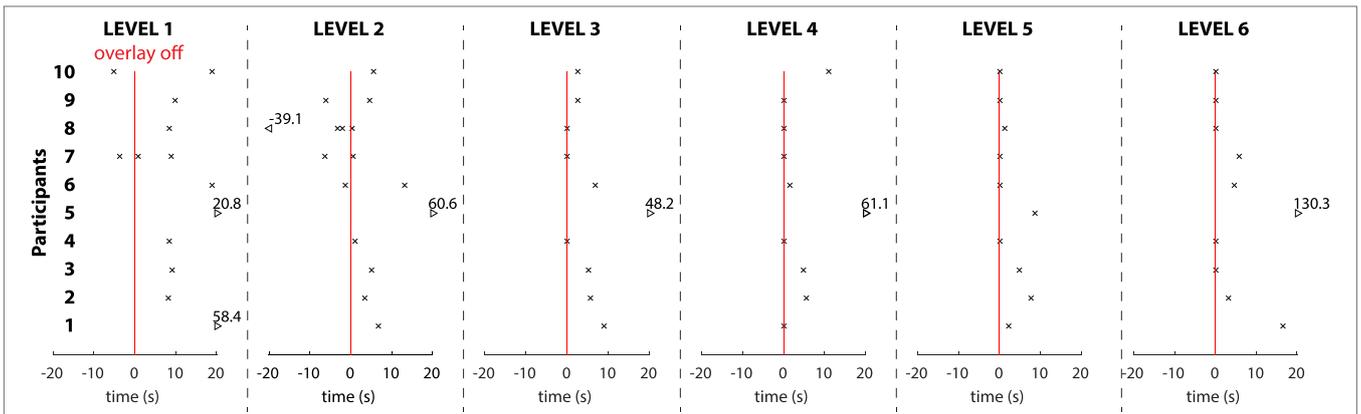


FIGURE 6 | Event-related touch raster showing touches (as x's) arranged around the removal of the overlay (the red line) on each level. The arrows indicate values that lie outside of the range of the plot. Each participant occupies one row on the Y-axis, and time is represented on the X-axis. The removal of the overlay is an important event within the interactive story as it indicates when the robot starts reacting to touches presented by the child. Children appear to adapt to the event of the overlay being removed—in earlier levels they touch the screen prior to removal and in later levels they do not.

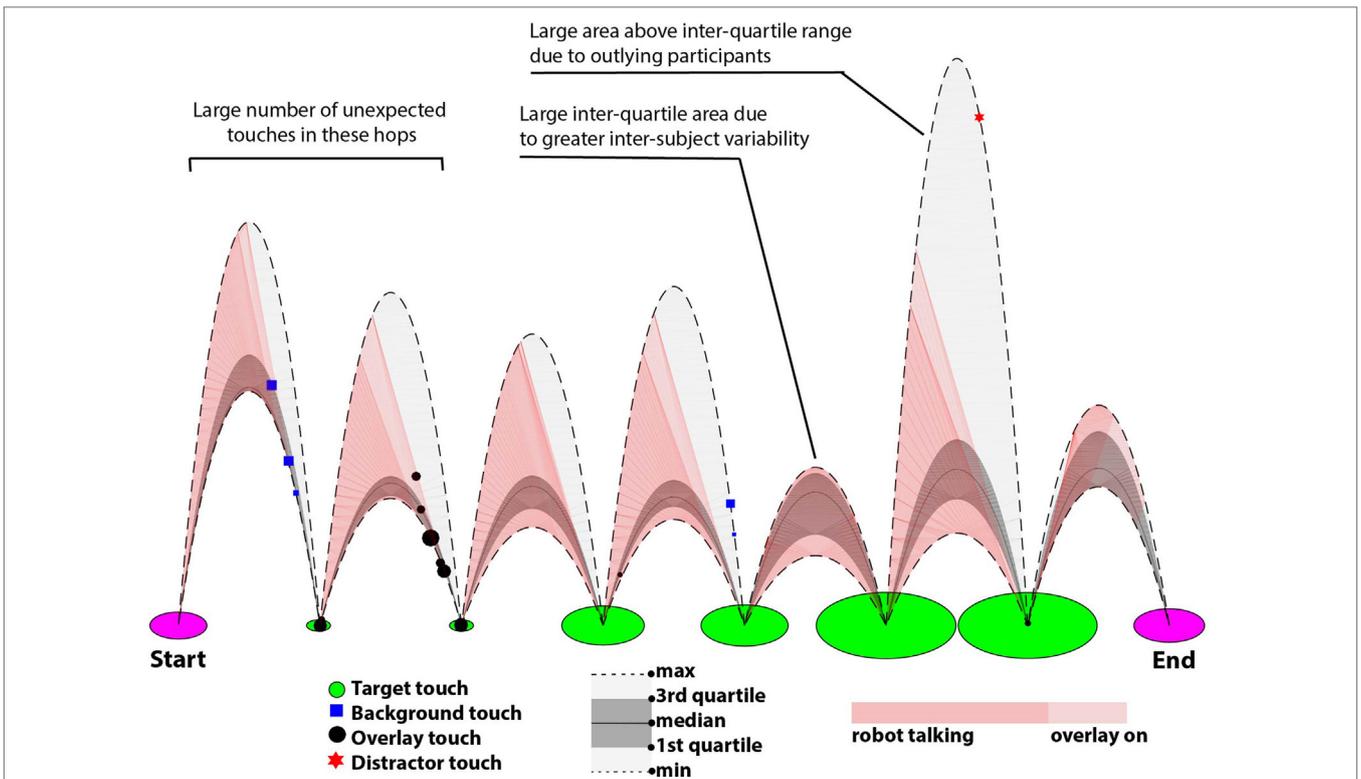


FIGURE 7 | Frog-hop plot showing the timeline of the interaction of all participants with the robot. The frog-hop plot shows the study as a series of “hops”—regions bounded by parabolas that represent a single level in the storytelling game. Each green circle in between hops represents a target touch (and, therefore, the end of a level; the size of the circle is proportional to the duration of the touch). Each hop shows the accumulated touches of all the participants, where the length of the curves representing a hop is proportional to the participant’s time spent on that level (e.g., the parabola at the top of the hop shows the time taken by the slowest participant and the parabola at the bottom of the hop shows the fastest participant). The gray shaded areas represent the interquartile range of the data. The pink shaded areas represent moments when the children could not interrupt the robot, i.e., when the robot was talking or when the overlay was on. The markers in each hop represent the unexpected events: respectively, background touches (blue square), overlay touches (black circle), and non-target touches (red star). The size of the markers is proportional to the duration of the touch. The plot exhibits the variability of the behaviors observed among the participants, with a high number of unexpected events (black circles and blue squares) that would not trigger a response from the robot, particularly in the first two levels.

In a survey of human participants, de Graaf et al. (2015) listed the top three important abilities for a robot to appear social as (i) participating in two-way interactions with users—both synchronous and asynchronous, (ii) displaying thoughts and feelings, and (iii) exhibiting social awareness. In our experiment, we combined those requirements in a dialogic storytelling context. The robot not only told the story by combining speech, head motions, and displayed emotions but also engaged in a social interaction with the children by responding to their touches on the screen and using attentional countermeasures to maintain their engagement.

4.1. What Level and Duration of Engagement Can Opie Facilitate?

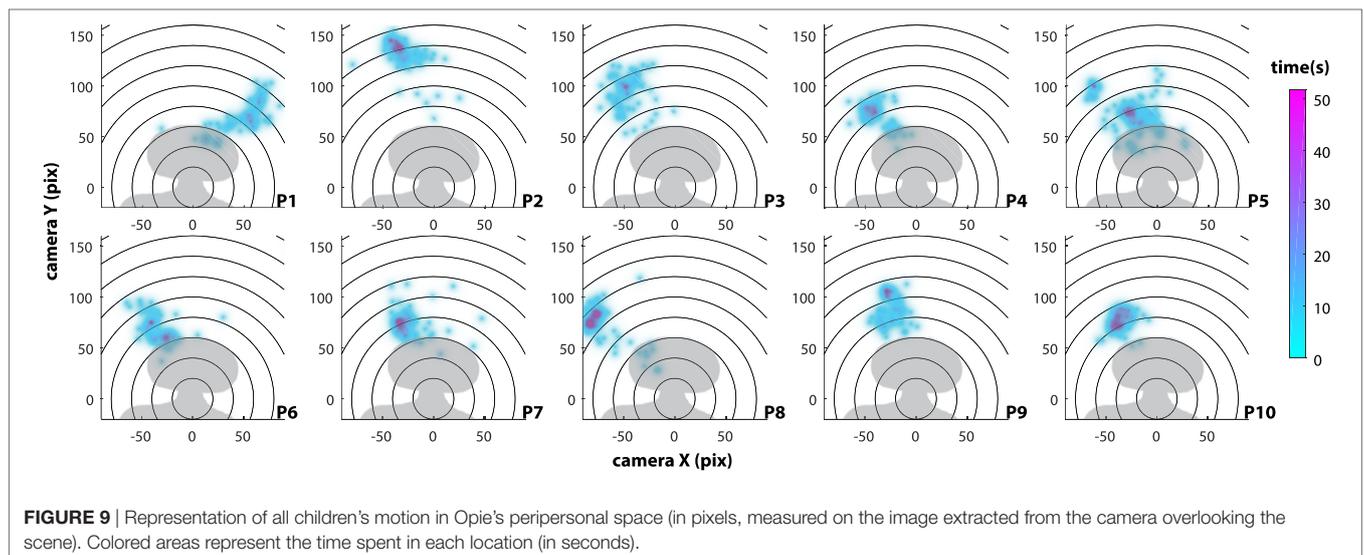
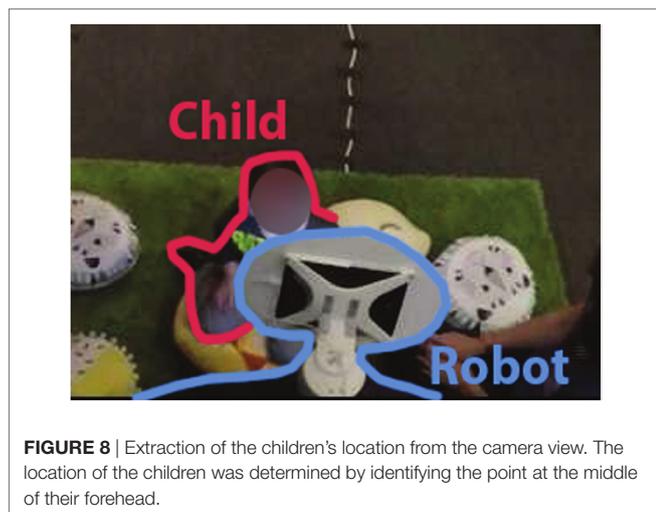
Within this study, we investigated performance at the interaction task (reaction times, completion of the task, and comprehension) and position and motion within close proximity of the robot. All participants completed the storytelling activity successfully,

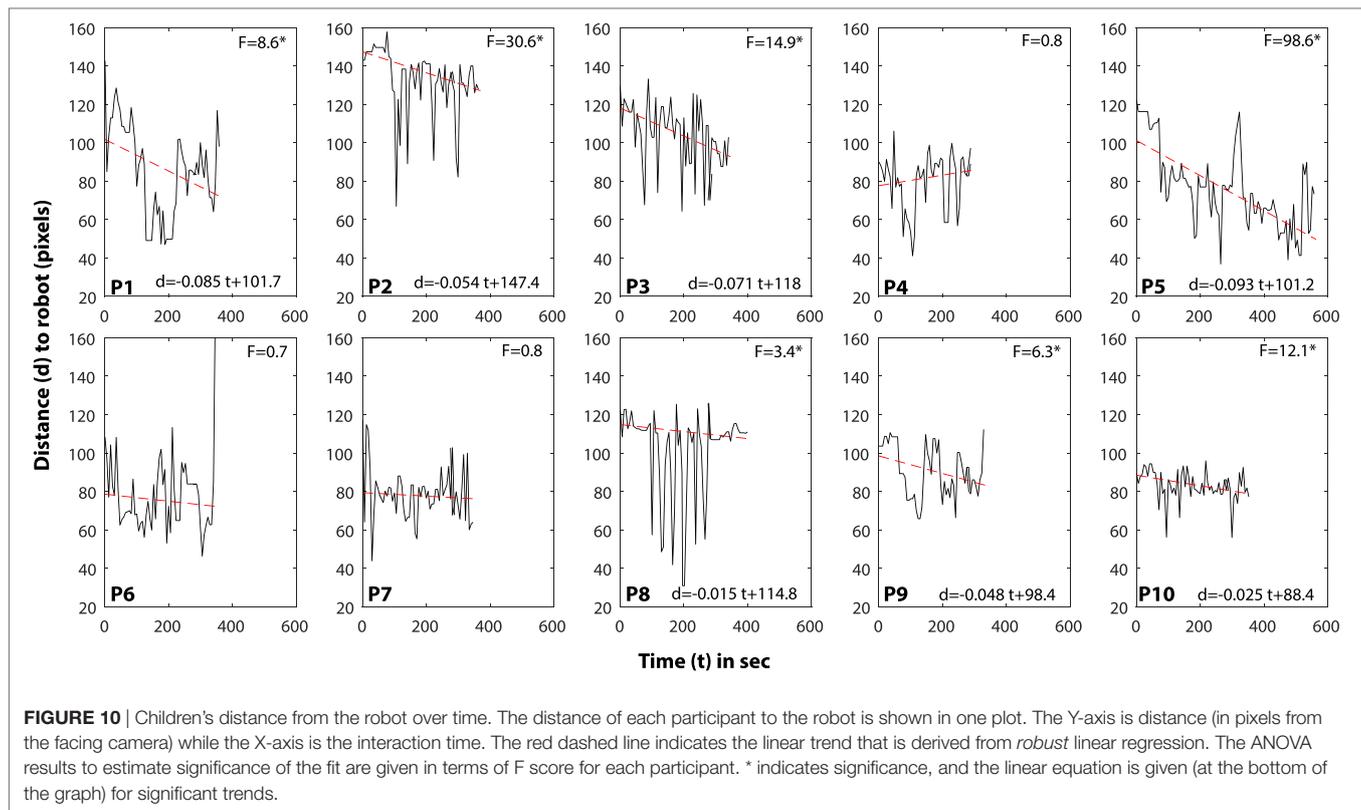
receiving directions from the social robot only. During the interaction, children remained in close proximity of the robot (<1 m) and touched the target (main character) preferentially and for a significantly longer time (cf. Figure 5). In addition, seven out of ten participants got closer to the robot during the interaction (the other three did not exhibit significant linear trends), suggesting greater engagement.

Temporal and spatial data together indicate that the robot succeeded in creating and maintaining engagement with children during the experiment. Spatial proximity data show the existence of preferred locations for interaction with the robot (see Figure 9), and that the children remained in the “personal” space (Hall, 1966) of the robot, which is an optimal distance for social interaction. Previous studies on human–robot interaction have supported the hypothesis that presence in the space less than 1 m away from the robot and greater closeness can be associated with engagement (Vázquez et al., 2014).

Touch data revealed a greater attentional focus directed toward the target, which shows that the robot succeeded at sharing the goal of the interaction with the children during the interaction. As shared intentionality has been argued to be a major correlate of engagement (Tomasello et al., 2005), the result suggests that children successfully engaged with the robot. In most cases, this maintenance of engagement did not require attentional countermeasures (only two participants out of ten received a countermeasure), which also supports the idea that the engagement was the result of a shared intentionality rather than forced by the use of countermeasures. Furthermore, the greater touch duration on the target compared to distractors or background areas also reinforces this conclusion, as duration of touch has been associated with greater engagement levels (Baek et al., 2014 and Silvera-Tawil et al., 2014).

Similarly, the performance in the quiz showed good recall of the elements of the story as a result of engagement. The poor performance at recalling the name of the main character could be due to the difficulty in recalling auditory compared to visual information (Jensen, 1971 and Cohen et al., 2009). The synthetic





speech used by the robot or difference to the other quiz questions (the experimenter read this question to the child) could also explain this result.

The robot maintained engagement with the children for an average of 6 min and 13 s, which is a long duration compared to other interactive settings in public spaces (between 3 and 4 min, see Hornecker and Stifter (2006)). However, this length of time is still short for educational purposes, where difficulties in engaging children can appear after a longer period of interaction (Ioannou et al., 2015). In particular, the implementation of dialogic storytelling proposed in this paper involves directional feedback and consideration of the turn taking rhythms, which are two of the three elements identified by Robins et al. (2005) to effectively maintain engagement. The current state of the robotic platform used in the study did not allow us to consider the third element: interaction kinesics (which would require moving limbs). Further investigation is required to study the impact interaction kinesics would have on sustained engagement.

4.2. How Individual or Stereotypical Are the Spatial and Temporal Reactions across Different Participants?

A possible explanation of the high level of engagement seen in this study is the adaptability of the robot's behavior during dialogic storytelling, as the robot was able to produce socially contingent responses to some of the children's actions by using verbal and emotional responses. Bartneck (2008) argued that one of the major bottlenecks of social robotics is that practical

implementation often requires producing a system that has generalizable features, but having an impact on society requires the capability to adapt to each user independently of group behaviors.

In our study, we introduced a novel visualization (the frog-hop plot, see Figure 7) which is intended to show an overview of different touch events and how they relate across participants and storytelling state. The frog-hop plot exhibits the unexpected individual behaviors of the children and reveals the temporal dynamics of the interaction across the different levels of a scene.

From a spatial perspective, the location in the space surrounding the robot also showed that patterns of engagement were different across children. In particular, although our data suggest that the children got closer to the robot over time, we also observed a large variance of the location of the children around their linear trends (see Figure 10). The variance observed was likely a result of the turn-taking dynamics of the interaction, which required the child to alternatively touch the robot or listen to it. This is similar to Michalowski et al. (2006)—despite the existence of optimal areas for social interaction (the personal space), individual patterns of motion in the space should be considered to fully understand the dynamics of engagement.

Implementing engaging social behaviors in social robots can benefit from an awareness of the spatial and temporal features at the individual level. This recommendation is akin to previous studies that exhibited physical, social, and cultural aspects of engagement with interactive technologies (Dalsgaard et al., 2011). We suggest that a multimodal approach will help account for all these aspects when designing for engagement.

4.3. How Do the Patterns of Spatial and Temporal Reactions Relate to Opie's Synchronous Behavior?

A large number of touches did not lead to a socially contingent response (background or overlay touches) during the beginning of the interaction (see **Figures 6** and **7**). Overlay touches were associated with smaller response times. They could be indicative of a high level of engagement: as the objective of the task remained the same during the story, some children could have had such a high level of understanding and performance that they responded before being prompted to. The concentration of background and overlay touches on levels one and two suggests that the children who touched these regions then changed their behavior as the story advanced and adapted to the limitations of the robot (as the robot would not respond to these touches). Each of these touches are missed opportunities of interaction for the robot, to which the children had to adapt. Interestingly, these missed opportunities suggest that higher levels of immediacy could be obtained by implementing storytelling in a closed-loop asynchronous manner and this modification would likely result in even higher levels of engagement. This issue is left for further investigation.

5. CONCLUSION

We proposed an implementation of dialogic storytelling using a closed-loop synchronous mode of interaction in a child-friendly robot. Based on spatial and temporal features of the interaction, we conclude that our robot succeeded in engaging children in a dialogic storytelling interaction. However, one outlying child disengaged during the story, and some touches of the children did not produce a response from the robot.

Consideration of spatial and temporal attributes of the interaction is important for evaluating the engagement of participants. Our study results show that touch timing data and spatial position data demonstrate different trends over the course of the study and provide insight into the child's engagement toward the task and robot. While the spatial, temporal, and quiz data collected generally suggest engagement, each of these measures is sensitive to both the group and individual level. Temporal touch responses reveal group trends such as a concentration of overlay and background touches during the first stages of the story, while also showing unique unexpected touches and response times for individuals. Spatial data have properties that reflect not only engagement at the group level (decreasing distance with the robot over time) but also show the existence of preferred areas for each child during the interaction.

REFERENCES

- Baek, C., Choi, J. J., and Kwak, S. S. (2014). "Can you touch me? The impact of physical contact on emotional engagement with a robot," in *Proceedings of the Second International Conference on Human-Agent Interaction* (Tsukuba: ACM), 149–152.
- Bartneck, C. (2008). "What is good? A comparison between the quality criteria used in design and science," in *CHI'08 Extended Abstracts on Human Factors in Computing Systems* (Florence: ACM), 2485–2492.

In addition to providing insight into engagement of children at the group and individual level, spatial and temporal measures also reflect the synchronous nature of the robot. This study demonstrates that closed-loop synchronous robots can facilitate engaging interactions; however, there is a distinct limitation created by the number of events that are not processed by the robot. It is not always feasible to have a reaction for every input that a robot receives, but each input could still be used to modify the robot's current state and future reactions. A future goal of the project is to implement asynchronous, closed-loop immediacy to enrich child-robot interactions with Opie, by processing all identified metrics automatically and online.

ETHICS STATEMENT

This study was carried out in accordance with the NHMRC statement on ethical conduct in human research (created in 2007—updated May 2015) and the protocol was approved by the UQ School of Psychology Ethics Committee (clearance no. 15-PSYCH-PHD-42-JH). Written informed consent was given by the parents of all subjects in accordance with the Declaration of Helsinki.

AUTHOR CONTRIBUTIONS

All the authors contributed to prototype design, results discussion, and paper redaction. The experiment was designed by SH, GD, MB, KH, JT, OO, and JW. Data collection was conducted by SH, GD, MB, KH, JT, OO, JWe, and JW. Analysis of the data and initial drafting of the article were performed by SH, GD, and JW.

ACKNOWLEDGMENTS

The authors would like to acknowledge Khoi Dhang (University of Queensland) for his help during the development of the storytelling game used in the study; Wilson Kong, Rollsy Ponmattam-Madassery, Sharon Koipuram, Jia Lim, Tamara Van Der Zant, Kelly Kirkland, Abigail Kelso, and Phoebe Mayne (University of Queensland) for their help during data collection.

FUNDING

The OPAL project was funded by the Australian Research Council (ARC) Centre of Excellence for the Dynamics of Language (CoEDL) (grant no. CE140100041).

- Cohen, M. A., Horowitz, T. S., and Wolfe, J. M. (2009). Auditory recognition memory is inferior to visual recognition memory. *Proc. Natl. Acad. Sci. U.S.A.* 106, 6008–6010. doi:10.1073/pnas.0811884106
- Dalgaard, P., Dindler, C., and Halskov, K. (2011). "Understanding the dynamics of engaging interaction in public spaces," in *IFIP Conference on Human-Computer Interaction* (Lisbon: Springer), 212–229.
- de Graaf, M., Allouch, S. B., and van Dijk, J. (2015). "What makes robots social? A users perspective on characteristics for social human-robot interaction," in *International Conference on Social Robotics* (Paris: Springer), 184–193.

- Fridin, M. (2014). Storytelling by a kindergarten social assistive robot: a tool for constructive learning in preschool education. *Comput. Educ.* 70, 5364. doi:10.1016/j.compedu.2013.07.043
- Gelin, R., d'Alessandro, C., Le, Q. A., Deroo, O., Doukhan, D., Martin, J.-C., et al. (2010). "Towards a storytelling humanoid robot," in *AAAI Fall Symposium: Dialog with Robots*, Arlington.
- Hall, E. (1966). *The Hidden Dimension. A Doubleday Anchor Book*. Anchor Books.
- Heath, S., Hensby, K., Boden, M., Taufatofua, J., Weigel, J., and Wiles, J. (2016). "Lingodroids: investigating grounded color relations using a social robot for children," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction* (Christchurch: IEEE Press), 435–436.
- Hensby, K., Wiles, J., Boden, M., Heath, S., Nielsen, M., Pounds, P., et al. (2016). "Hand in hand: tools and techniques for understanding children's touch with a social robot," in *The Eleventh ACM/IEEE International Conference on Human Robot Interaction, HRI '16* (Piscataway, NJ: IEEE Press), 437–438.
- Hornecker, E., and Stifter, M. (2006). "Learning from interactive museum installations about interaction design for public settings," in *Proceedings of the 18th Australia conference on Computer-Human Interaction: Design: Activities, Artefacts and Environments* (Sydney: ACM), 135–142.
- Ioannou, A., Andreou, E., and Christofi, M. (2015). Pre-schoolers interest and caring behaviour around a humanoid robot. *TechTrends* 59, 2326. doi:10.1007/s11528-015-0835-0
- Jensen, A. R. (1971). Individual differences in visual and auditory memory. *J. Educ. Psychol.* 62, 123. doi:10.1037/h0030655
- Kennedy, J., Baxter, P., Senft, E., and Belpaeme, T. (2015). "Higher nonverbal immediacy leads to greater learning gains in child-robot tutoring interactions," in *International Conference on Social Robotics* (Paris: Springer), 327–336.
- Kory, J. M. (2014). *Storytelling with Robots: Effects of Robot Language Level on Children's Language Learning*. Master's thesis, Massachusetts Institute of Technology.
- Malmir, M., Forster, D., Youngstrom, K., Morrison, L., and Movellan, J. (2013). "Home alone: social robots for digital ethnography of toddler behavior," in *Proceedings of the IEEE International Conference on Computer Vision Workshops* (Sydney: IEEE), 762–768.
- Mehrabian, A. (1968). Some referents and measures of nonverbal behavior. *Behav. Res. Methods Instrum.* 1, 203–207. doi:10.3758/BF03208096
- Mehrabian, A. (1972). *Nonverbal Communication*. Transaction Publishers.
- Michalowski, M. P., Sabanovic, S., and Simmons, R. (2006). "A spatial model of engagement for a social robot," in *9th IEEE International Workshop on Advanced Motion Control, 2006* (Istanbul: IEEE), 762–767.
- Mol, S. E., Bus, A. G., de Jong, M. T., and Smeets, D. J. (2008). Added value of dialogic parent-child book readings: a meta-analysis. *Early Educ. Dev.* 19, 7–26. doi:10.1080/10409280701838603
- Mutlu, B., Forlizzi, J., and Hodgins, J. (2006). "A storytelling robot: modeling and evaluation of human-like gaze behavior," in *6th IEEE-RAS International Conference on Humanoid Robots* (Genoa: Institute of Electrical and Electronics Engineers (IEEE)).
- Pitsch, K., Kuzuoka, H., Suzuki, Y., Sussenbach, L., Luff, P., and Heath, C. (2009). "The first five seconds: contingent stepwise entry into an interaction as a means to secure sustained engagement in HRI," in *RO-MAN 2009 - The 18th IEEE International Symposium on Robot and Human Interactive Communication*, Toyama.
- Quigley, M., Conley, K., Gerkey, B., Faust, J., Foote, T., Leibs, J., et al. (2009). "ROS: an open-source robot operating system," in *ICRA Workshop on Open Source Software*, Vol. 3 (Kobe, Japan), 5.
- Robins, B., Dautenhahn, K., Nehaniv, C. L., Mirza, N. A., François, D., and Olsson, L. (2005). "Sustaining interaction dynamics and engagement in dyadic child-robot interaction kinesics: lessons learnt from an exploratory study," in *ROMAN 2005. IEEE International Workshop on Robot and Human Interactive Communication, 2005* (Nashville: IEEE), 716–722.
- Rogers, K., Wiles, J., Heath, S., Hensby, K., and Taufatofua, J. (2016). "Discovering patterns of touch: a case study for visualization-driven analysis in human-robot interaction," in *ACM/IEEE Int Conf on Human-Robot Interaction*, Christchurch.
- Ryokai, K., Lee, M. J., and Breitbart, J. M. (2009). "Children's storytelling and programming with robotic characters," in *Proceeding of the Seventh ACM Conference on Creativity and Cognition - C&C 09* (Berkeley: Association for Computing Machinery (ACM)).
- Sidner, C. L., Lee, C., Kidd, C. D., Lesh, N., and Rich, C. (2005). Explorations in engagement for humans and robots. *Artif. Intell.* 166, 140–164. doi:10.1016/j.artint.2005.03.005
- Silvera-Tawil, D., Rye, D., and Velonaki, M. (2014). Interpretation of social touch on an artificial arm covered with an eit-based sensitive skin. *Int. J. Soc. Robot.* 6, 489–505. doi:10.1007/s12369-013-0223-x
- Szafir, D., and Mutlu, B. (2012). "Pay attention!," in *Proceedings of the 2012 ACM annual conference on Human Factors in Computing Systems - CHI 12*, Austin.
- Tomasello, M., Carpenter, M., Call, J., Behne, T., and Moll, H. (2005). Understanding and sharing intentions: the origins of cultural cognition. *Behav. Brain Sci.* 28, 675–691. doi:10.1017/S0140525X05000129
- Vázquez, M., Steinfeld, A., Hudson, S. E., and Forlizzi, J. (2014). "Spatial and other social engagement cues in a child-robot interaction: effects of a sidekick," in *Proceedings of the 2014 ACM/IEEE International Conference on Human-Robot Interaction* (Bielefeld: ACM), 391–398.
- Walters, M. L., Syrdal, D. S., Dautenhahn, K., Te Boekhorst, R., and Koay, K. L. (2008). Avoiding the uncanny valley: robot appearance, personality and consistency of behavior in an attention-seeking home scenario for a robot companion. *Auton. Robots* 24, 159–178. doi:10.1007/s10514-007-9058-3
- Whitehurst, G. J., Falco, F. L., Lonigan, C. J., Fischel, J. E., DeBaryshe, B. D., Valdez-Menchaca, M. C., et al. (1988). Accelerating language development through picture book reading. *Dev. Psychol.* 24, 552. doi:10.1037/0012-1649.24.4.552
- Wiles, J., Worthy, P., Hensby, K., Boden, M., Heath, S., Pounds, P., et al. (2016). "Social cardboard: prototyping a social ethnodroid in the wild," in *ACM/IEEE Int Conf on Human-Robot Interaction*, Christchurch.
- Yanco, H. A., and Drury, J. (2004). "Classifying human-robot interaction: an updated taxonomy," in *SMC* (The Hague: IEEE), 2841–2846.

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2017 Heath, Durantin, Boden, Hensby, Taufatofua, Olsson, Weigel, Pounds and Wiles. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.