Role of Parent Nonverbal Cues in Children's Engagement During Dialogic Reading

Jocelyn Shen
Cambridge, MA, USA
Massachusetts Institute of Technology
joceshen@mit.edu

Ying Li
Cambridge, MA, USA
Harvard University
yingli@mde.harvard.edu

Javaria Hassan
Cambridge, MA, USA
Harvard University
javariahassan@gse.harvard.edu

ABSTRACT

High-quality, social interactions between parents and their children are crucial for young children's development. In order to develop agent interventions that improve interactions between parents and children, it is important to both understand how the parent's nonverbal behavior influences the child's affective state and quantify and detect these effects computationally. In this paper, we explore the role of a parent's nonverbal cues on a child's engagement during an educational, dialogic reading interaction. We specifically focus on using the parent's body pose features to better inform a child's engagement during the task. Our analysis offers a high-level, holistic approach to the relationship between parent nonverbal cues and child's affect through both quantitative and qualitative methods. We find that the child's engagement is positively correlated with the joint engagement between the child and the parent, as well as the valence. However, the arousal of the child and the parent are mostly negatively correlated. Using a simple model, we find that training on the parent's body features and predicting the child's engagement score yields a model with a maximum accuracy of 64.8%. Furthermore, through feature analysis, we find that the most predictive features of a child's engagement are interpersonal features between the parent and child. To better understand the moments when children lose engagement in a task, we qualitatively assessed video clips and found that those moments are associated with either the parent or the child exhibiting actions unrelated to the reading stories, such as yawning or the parent checking their cellphone. Finally, we provide an overview of design implications for parent-child facilitator agents as well as future research directions in this area.

Author Keywords

Affective Computing; Children's Psychology; Nonverbal Behavior; Human-Robot Interaction

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04...$15.00

DOI: https://doi.org/10.1145/3313831.XXXXXXX

CCS Concepts

•**Human-centered computing** → **Interaction design;** •**Applied computing** → *Collaborative learning; Interactive learning environments;*

INTRODUCTION

Parent-child interactions are widely acknowledged as playing a central role in children's developmental outcomes. In educational tasks specifically, nonverbal behaviors are paramount in establishing communication between a parent and a child whose language skills are not fully developed. To improve parent-child interactions during educational tasks, understanding how a parent's nonverbal behavior relates to the child's engagement with the learning task is just as important as developing ways to detect the effects of the parent's behavior on the child.

We aim to study this relationship in the context of a dialogic reading task between a parent and a child. Dialogic reading provides an ideal setting to conduct our analysis, since it is a literacy learning task rich with both verbal and nonverbal behaviors. Dialogic reading interactions between a parent and child can promote literacy skills and social-emotional learning in young children [10]. Using an agent to improve dialogic reading interactions can provide better home literacy environments to families that have limited rich language exposure, which are disproportionately families from low socioeconomic statuses [27].

We study how a parent's nonverbal behavior affects their child's engagement during a dialogic reading task through two smaller explorations. The first is understanding how much synchrony or alignment there is between the parent's and the child's affective states. Parent-child synchrony can be a good metric to evaluate an interaction, as prior work demonstrates that high parent-child synchrony is associated with familiarity, a healthy parent, typical development, and more positive outcomes [19]. Quantifying parent-child synchrony during learning interactions is important since understanding human-human interpersonal dynamics can inform decisions regarding human-agent interactions.

The second exploration in our work is developing methods to predict a child's engagement using the parent's nonverbal cues. Few studies have started to model nonverbal cues of multiple interactants to recognize social-emotional states, despite the importance of context in affect detection and comprehension

[7], [20]. Incorporating a parent's nonverbal cues into models that predict a child's affective state can improve affect-related context awareness and ultimately inform better interventions from an automated agent.

We explore the two aforementioned areas through a quantitative and qualitative analysis of a rich parent-child dyad dataset, the DAMI-P2C (Dyadic Affect in Multimodal Interaction - Parent to Child) dataset [5]. This dataset includes pose data and annotated affect labels for parent-child dyads engaged in a story-reading activity.

With the dataset, a variety of simple computational methods are employed to analyze the interplay of the parent's body pose features and the child's engagement score. Pearson correlation coefficients are calculated for each session and we find parent's body pose features are not linearly correlated with child's engagement at all, showing coefficients of 0 in almost all sessions. We trained various models to predict the child's engagement score using the parents body pose features. Our model is able to achieve 64.8% accuracy with a simple linear model and standardization of the parent's body pose features. By performing feature analysis, we find that the nonverbal features most highly correlate with a child's engagement score are the interpersonal pose features between the parent and child. For our qualitative assessment, we visualize points where the dyad's valence are out of synchrony as well as points where the child's engagement score dropped significantly. From analyzing these videos, we identify that a child's disengaged moments are associated with either the parent or the child's actions unrelated to the reading task, for example, the child yawning, the child playing with the chair, or the parent checking their cellphone. We observe that it is hard to qualitatively classify the out-of-synchrony moments consistently across all sessions. Some out-of-synchrony moments are clearly a result of the disengagement, but this is not true in every interaction we observed.

Based on our results, we conclude with design implications for an agent facilitating parent-child interactions. These implications include adding active interactions into passive listening and recognizing the importance of physical interactions between the parent and child dyad. We recognize that our broad approach cannot cover the depth necessary for this topic, and therefore outline future work to explore nonverbal cues as context for understanding affect in human-human joint interactions.

BACKGROUND

People communicate with one another through two predominant means: verbal and nonverbal. While verbal communication tends to dominate human interactions, nonverbal communication extends spoken communication and can offer considerable information about both people and situations in its own right. Communication behaviors such as facial expressions, gestures, body postures, and vocal intonation or prosody are all very powerful in conveying social information.

A growing literature characterizes human communication in terms of reciprocal behavioral and physiological mimicry [4, 11, 15]. Garrod and Pickering argued that people's nonlinguistic aspects of communication, such as expression, become more aligned when they interact with each other during dialogic activities [15]. Similarly, Brennan and Clark showed how interlocutors use similar expressions when describing pictures of objects to each other [2]. Studies have also demonstrated how people tend to adopt the same postures as each other [28], laugh or yawn together [16], and imitate shaking the head or nodding [22].

Such behavior matching has been related to development of a common ground, improved rapport, and better collaborative performance [13, 23]. For example, addressees who align their gaze with speakers tend to align their interpretations with the speaker as well [26]. Fusaroli Tylén developed a quantitative approach enabling them to predict collective performance from interpersonal synergy (complementary pattern of behavior) [14]. Breazeal et al. found that during human-robot teamwork tasks, the human reads and interprets non-verbal cues from the robot, such as nods of the head, deictic gestures, and gaze, in order to coordinate their behavior in a way that improves teamwork efficiency and robustness [1]. Other work has shown that the higher the eye-synchrony between a speaker and listener, the greater the listener's score on a comprehension test [26].

Synchrony in parent-child interactions has been studied extensively, in addition to the aforementioned works; Leclère et al. reviewed studies on synchrony in mother-child interactions. Focusing on smile strength, tonal and temporal features in voice, mutual gaze, coordinated movements, etc., they showed that mother-child synchrony correlates to the child's cognitive processing, school adjustment, and learning of word-object relations [19].

Prior research has also explored the effects of nonverbal communication on learning. Sinha and Cassell studied speech in dyadic peer tutoring conversations and found that influence, convergence, and rapport in dialog are correlated with more significant learning gains [29]. Kory-Westlund et al. found that children were able to attend to non-verbal social cues to learn new words from both a robot tutor and a human partner equally as well [31]. The researchers also observed that children displayed signs of greater emotional engagement during a dialogic reading task when the robot learning companion was more expressive [17].

Synchrony and rapport have been measured in a number of different ways. For example, it has been demonstrated by the synchrony of heart rate and pupil-diameter during social interactions, the tendency to blush when an interaction partner blushes, and the contagiousness of crying or yawning [18, 25]. Research also suggests that during non-physical close interactions, mothers and infants synchronize their heart rhythms and breathing patterns [12, 25]. Tickle-Degnen and Rosenthal (1990) describe and measure the nature of rapport in terms three interrelating components: mutual attentiveness, positivity, and coordination [30].

Prior research has demonstrated the use of external affective cues to predict the affective state of a person of interest. Lee et al. show that the accurate inference of children's social-emotional state of attention depends on accounting for the nonverbal behaviors of their storytelling partner, namely their speaker cues [21]. Speaker cues have also been used in a multi-task learning approach to jointly learn the recognition of affective states [32]. Chen et al. designed end-to-end deep learning methods to recognize a person's affective expression in an audio stream with two speakers, automatically discovering features and time regions relevant to the target speaker's affect [8].

Nonverbal cues such as body posture, gestures, interpersonal distance, and touch have been widely investigated when assessing dyadic parent-child interactions. However, prior work has majorly focused on analyzing these cues on an individual scale, rather than how they are elicited though to each other [9, 24].

Also, most of the research in parent-child nonverbal interactions has been focused on infants, since at this age nonverbal cues are the main communication channel between the parent and the child. Colegrove et al. showed only a few studies have examined nonverbal behavior as an assessment tool for children older than 12 months in relation to child development [9].

To the best of our knowledge, the nonverbal behaviors that were assessed in the majority of these studies, besides paralinguistic features such as pitch and voice tone, were manually annotated by human observers, which can be burdensome. Given the importance of nonverbal communication during parent-child interactions, and considering the gap in the previous work, we chose to explore an automated approach to understanding parent-child nonverbal communication that can be correlated to learning gains. This can help to inform the design of automated agents as learning companions in dialogic reading activities.

DATASET

For our study, we used the DAMI-P2C dataset, which captures the affect (engagement, arousal, and valence) of parents and children engaged in natural story-reading dyads [6].

The DAMI-P2C dataset consists of audio-visual recordings of 34 families recruited from the Greater Boston area (Table 1), where a parent and child (3-7 years old) dyad each engage in two 45-minute in-lab sessions. For the first 20 minutes of each session, the parent and child read stories together, and for the remaining 25 minutes, the parent fills out surveys. Families that completed both sessions were given $75 as their compensation.

In contrast to existing public datasets for affect recognition, each instance for both speakers in the DAMI-P2C dataset has been annotated for the perceived affect by three labelers. This dataset is publicly available, and includes acoustic features of the dyads' raw audios, affect annotations, body pose, and a diverse set of developmental, social, and demographic profiles of each dyad.

Figure 1 illustrates the experiment setup during the in-lab parent-child co-reading sessions. Audio-visual recordings were captured using 1 boundary microphone and 7 cameras installed in the story reading station. The cameras were used to capture different angles of the dyadic interaction, i.e., frontal view, birds-eye view, parent-centered view, child-centered view, parent first-person view, and table first-person view.

	parent (age)	child (age)
Female	25 (38.78 ± 4.70)	13 (5.20 ± 1.96)
Male	9 (42.25 ± 6.90)	21 (5.65 ± 0.96)

Table 1. Gender identity and age range of the participant families in the dyadic dataset.

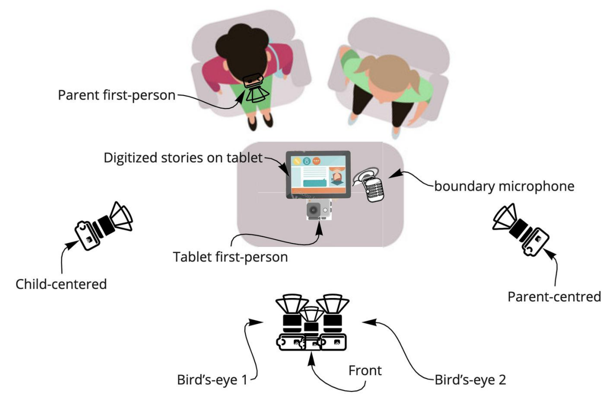

Figure 1. Experimental setup (taken from https://www.media.mit.edu/projects/dami-p2c/overview/)

Body Pose Features

The body pose features are extracted from the video data of the story-reading sessions. A pipeline using OpenPose is able to identify people of interest in a given frame and track their movement over time [3]. Once a person is identified, the pipeline finds, segments, and maps the body pose in 3D space.

These segments form the raw body pose features, from which touch, body and head orientation, gaze, movement, and interpersonal space can then be calculated. We identify body pose features such as the angle the child or parent is leaning forward, the orientation of the parent and child's head in relation to one another, as low level features in our dataset.

Affect labels

Three trained annotators with a psychology or education background annotated the audio-visual recordings of the families' co-reading interactions. The annotated dataset includes the child's arousal, valence, engagement, and join engagement, and the parent's arousal and valence. In total, 16,593 video fragments have been annotated with 488.03 ± 123.25 fragments from each family on average.

The annotators were provided with detailed coding schemes to score for the valence and arousal of the parent and child, as well as the child's engagement and joint engagement. These coding schemes were objective across all participants, that is, the score of a person's affect at a particular timestamp was given independent of the person's affect at previous timestamps.

Preliminary Data Analysis
We first preprocessed the datasets to transform the raw data into an aggregated format. To combine the affect features and body pose features, we added corresponding timestamps and clip IDs to both the datasets and used a database-style join based on the family, session and clip IDs to give us a merged pose and affect dataset.

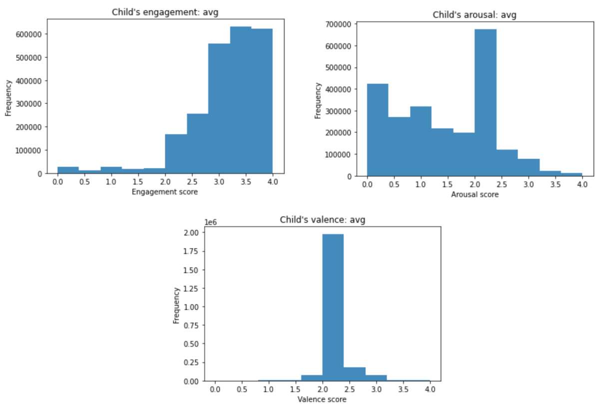

Figure 2. Histograms of the distribution for each affect attribute label after three individual ratings are averaged and standardized.

As a preliminary data analysis step, we plot the distribution of the child's engagement, as seen in Figure 2., and find that the distribution matches with results given in [8].

RESULTS AND ANALYSIS

Interplay of Parent's Nonverbal Behavior and Child's Affect

We first explore, at a high level, how closely tied a parent's nonverbal features are to a child's affective state.

Quantitative Approach
Our quantitative analysis begins with looking for correlations between the parent's body pose features (eg. parent body pitch, parent body yaw), the parent's affect, and the child's affect.

We calculated the Pearson Correlation Coefficient between the affective state of the parent and child within each session. Figure 3 plots the coefficients for all the video sessions. We found that the child's engagement is highly correlated with the joint engagement. Similarly, the valence between the parent and child is positively correlated, but less strongly than the engagement-joint engagement correlation. In contrast, the arousal between the parent and child does not have consistent correlation behaviours across sessions, and more than 60% of the sessions showed negative correlations.

We also looked at the correlation between the parent's valence and arousal, and the child's engagement. We found that half of the sessions showed positive correlations whereas the other half showed negative correlations. Our conclusion is that among the parent's affective states, the joint engagement is the primary factor influencing child's engagement.

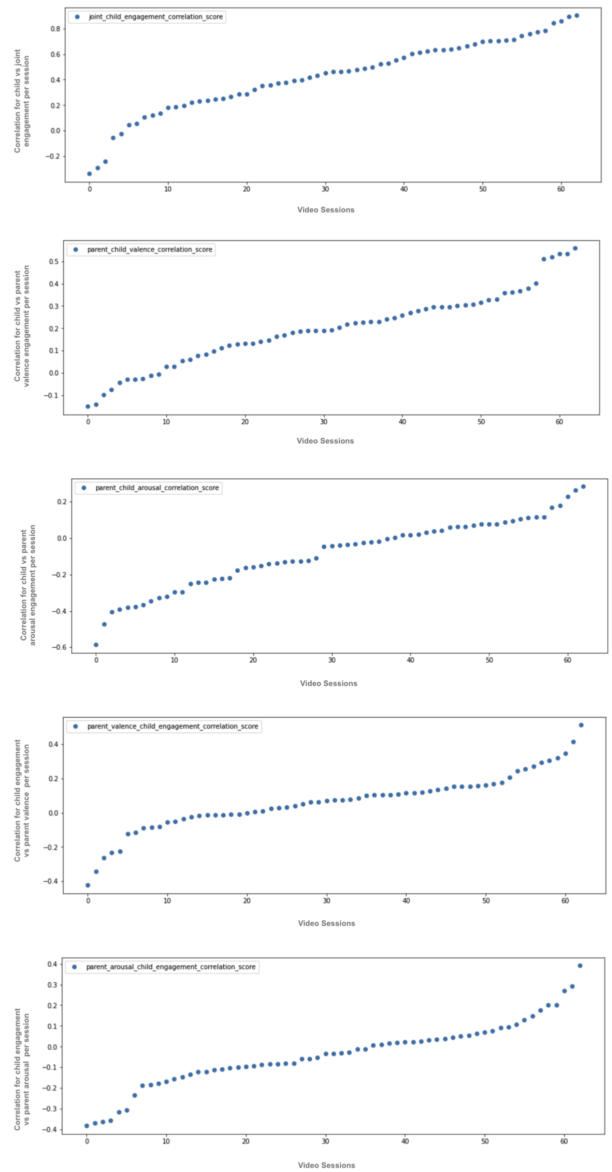

Figure 3. Correlation score between child and parent affect for all sessions

We also looked into the correlations between the low-level body pose features of the parent and child, and compared the parent's body pose features with the child's engagement in Figure 4. We found that the body pose features between the parent and child were not linearly correlated. This is unsurprising, as the features are too low-level to describe poses in a frame. We also found that the current body pose features barely correlate with the child's engagement score. These results indicate that there is a need to translate low-level body pose features into descriptive high-level features, which can be done in future work.

Qualitative Approach
We analyzed videos of the co-reading sessions to qualitatively assess what happens when the parent-child dyad are out of synchrony. Identifying moments with asynchronous engagement was easy, yet it was relatively harder to spot moments

when the valence was out of synchrony. This can be confirmed by comparing the parent and child valence in a session (Figure 7). In some sessions when the child is disengaged, the valence can also diverge. However, this correlation is by no means consistent spanning the full session, neither is it consistent across all family sessions.

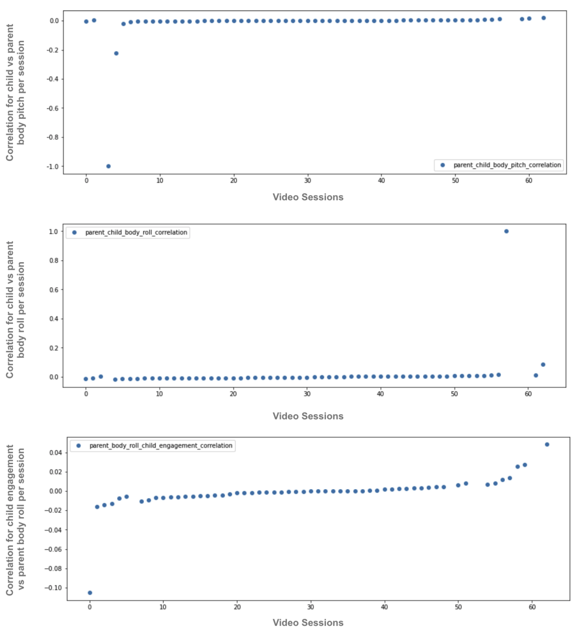

Figure 4. Correlation for body pose features

Predicting Child's Engagement Using Parent's Nonverbal Features

Quantitative Approach

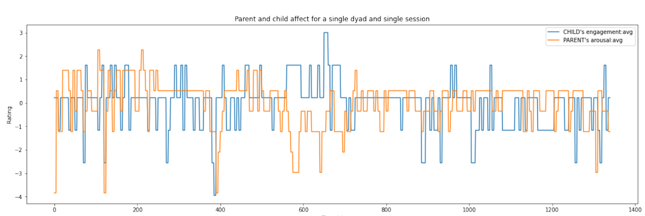

Figure 5. Childrens' behaviors during disengagement

To explore whether or not a parent's nonverbal behaviors are predictive of a child's engagement, we use a simple linear model, varying the features that the model is trained on.

First, we train a multi-class logistic regression model taking the normalized body pose features of both the parent and child as predictors of the child's engagement score and using an 80-20 train test split. We use two normalization methods, standard score and min-max feature scaling, and compare the model performance. Additionally, since the child's engagement scores are the average across annotators, we round the labels to discrete ratings from 0-4. The results for our preliminary model are shown in Table 3.

We attempted to further improve model accuracy by augmenting the subsets with family demographic data, including information on the parent's parenting style, parent stress, and child's behavior based on the survey results mentioned in the Dataset section. We additionally augmented the model with the parent's affect labels to see the effect on model performance. Since we are interested in both the task-related engagement and joint engagement of the parent-child dyad, we use these features to predict both engagement types. Results for these models are shown in figure 2.

We find that the model predicting a child's engagement using all body pose features, responses from the behavioral questionnaires, and parent affect labels has the highest accuracy (64.8%). Given that we are using logistic regression model, which has low model complexity, there are bounds to the prediction accuracy that we can obtain. However, it is interesting that even with no explicit feature selection or extraction of higher level features from the body pose data, the model is still predictive. One limitation to our approach is that parent affect labels are human labeled. This means that it is possible for the parent's body pose features to be highly correlated with the parent's affect labels, since annotators may have used these nonverbal cues implicitly during the labeling process. However, we see that even if we exclude the parent's affect labels, our model can still predict the child's engagement with an accuracy of 64.2%.

Additionally, since one of our goals is to uncover which body pose features are relevant in predicting a child's engagement, we perform feature importance analysis. We use the coefficients of the logistic regressions model as feature importance indicators. Figure 6 shows the feature importance for 5 engagement levels.

Our feature analysis results indicate that features of high coefficient magnitude are mainly interpersonal features between the child and parent. On the contrary, features of low coefficient magnitude are mainly parent's or child's body pose features. These results are interesting in that interpersonal features contain more information about the relationship between two people in space. It is perhaps unsurprising that a parent and child's interpersonal distance and head tilt angles are more informative than the parent or child's individual features. Focusing on these interpersonal features can help us quantitatively explore the physical and gestural synchrony between the parent-child dyad.

Qualitative Approach

It is important to not only understand the interplay between a parent's nonverbal behavior and their child's affective state, but also to understand the role of the parent's behaviors at specific times when the child disengages or re-engages with the learning task at hand. We first qualitatively assess patterns in the child's engagement during a single session. Figure 8 shows a sample plot of a child's engagement score plotted over time. Child engagement events (disengagement/re-engagement) are categorized as steps in the engagement ratings with a significant difference. For example, if the child's engagement rating drops from 4 to 2, we denote this as a disengagement event.

Engagement type	All body pose features	All body pose features + Behavioral	All body pose features + Behavioral + Parent affect labels
Joint engagement	57.9%	59.5%	59.6%
Engagement	63.4%	64.2%	64.8%

Table 2. Model performance for predicting the child's engagement and joint engagement across different feature combinations

Standard Normalization	Min-max Normalization
62.9%	63.1%

Table 3. Accuracy score for preliminary model across normalization methods

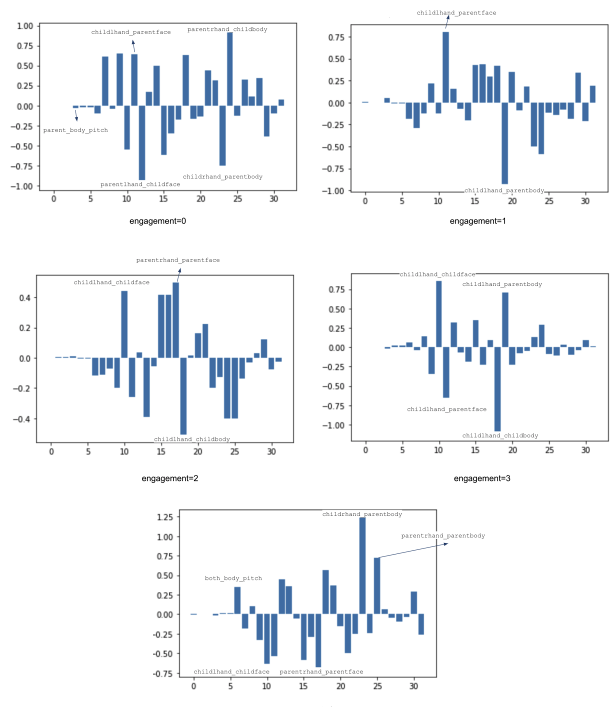

Figure 6. Feature Importance for 5 engagement levels

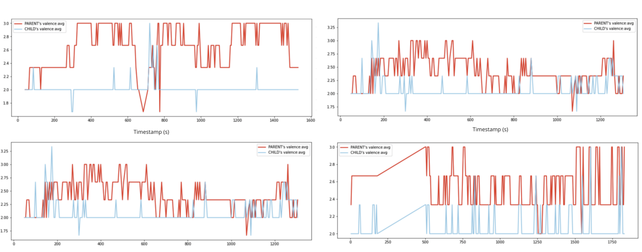

Figure 7. Valence of child and parent in four sessions

Alternatively, if the child's engagement rating increases from 2 to 4, we denote this as a re-engagement event.

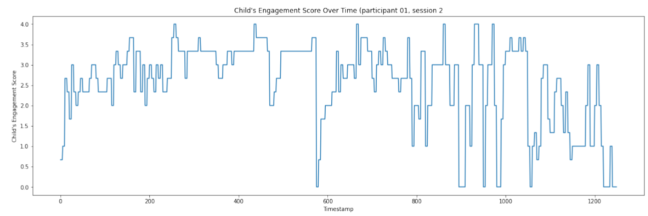

Figure 8. Sample plot of child's engagement over time during a single session

In reading activities, children's engagement with the material and their parent is important to their learning. We hypothesize that it is challenging for children to hold their attention to readings for 20-30 minutes. Parents usually intervened when they saw that the children were not engaged with the reading task, usually because the children lost interest in the activity or the activity continued for too long. Understanding how parents make the children re-engaged is critical to drawing insights for a facilitator agent interaction design.

We plotted the child's engagement scores in each session to identity timestamps where a child disengages or reengages. Then we watched the video clips for those periods to learn different parents' interventions used. Here we summarize our findings.

1. Some children can hold their attention to the readings through the session even though they would get tired of sitting and start moving around. Children would lean towards their parent and seek cuddles or stand up and move around, while still remaining engaged with the stories.

2. Some parents would ask their child whether or not they would like to continue listening to the stories when they noticed that their child was not engaged. This usually worked well to reengage the child. We hypothesize that children tend to be obedient to parents, or need to be reminded to draw their attention back to the reading task.

3. Asking the children questions related to the stories being read was a very efficient way to keep them engaged. It served as a break from passive listening and paved the way for active interactions with their parent. We spotted that many children tended to lean toward their parent or interact physically with them (e.g dragging mom's sleeves). We hypothesize that children like seeking their parent's attention, and thus, a parent asking questions provides the child with an opportunity to get their attention.

4. Some children would seek the parent's attention by starting conversations irrelevant to the readings or initiating physical

interactions like pulling on the parent's clothes. Some parents embraced such behaviors, and therefore both the parent and child disengaged from the reading task. However, in most cases, these pairs were able to get back to the reading shortly after, as long as the parent resumed readings. Some parents ignored such behaviors and the child appeared to be more frequently disengaged from readings after being ignored.

5. When children started to show disengagement behaviors like yawning or moving around in the chair, they tended to get physically closer to the parent by holding their arms out or seeking cuddling. Upon physical interaction, many children seemed to calm down and reengage with the readings.

Applying these findings to a human-agent interaction, an agent participating in parent-child reading activities can spot moments where children start to disengage and ask story related questions to create opportunities for children to interact with their parents. Validation of the hypothesis above requires a controlled experiment and thus can be applied in future work.

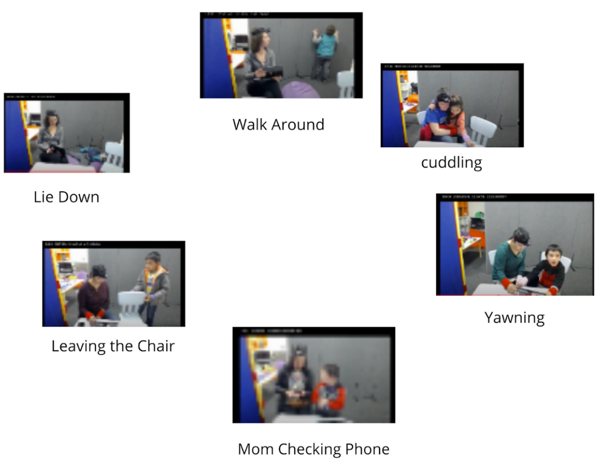

Figure 9. Childrens' behaviors during disengagement

DESIGN IMPLICATIONS

In this section, we discuss human-agent interaction design implications in the context of parent-child reading activities.

1. Add active interactions between passive listening to serve as a stimulant and increase the child's attention span. Most children have limited attention spans when acting as passive listeners. Active interactions can offer the child with an opportunity to control the reading progress, like navigation to next page or choosing which stories to read. Alternatively, the interactions can be initiated by the parent, for example, asking questions related to the stories that they read. The agent can prompt parents to talk to their child at appropriate times over the course of the reading session.

2. Our study shows that a child's engagement is highly correlated with joint engagement and a child's valence is highly correlated with parent valence, so designing the reading tasks to increase a parent's engagement could potentially increase child's engagement.

3. Physical interactions are important between the parent and child dyad. The most effective interaction we observed in the video data is when the parent cuddles the child in their arms during reading. We acknowledge that more research should be done on whether agents can incorporate physical interactions with children during learning tasks, and what the ethical implications of such a feature might be.

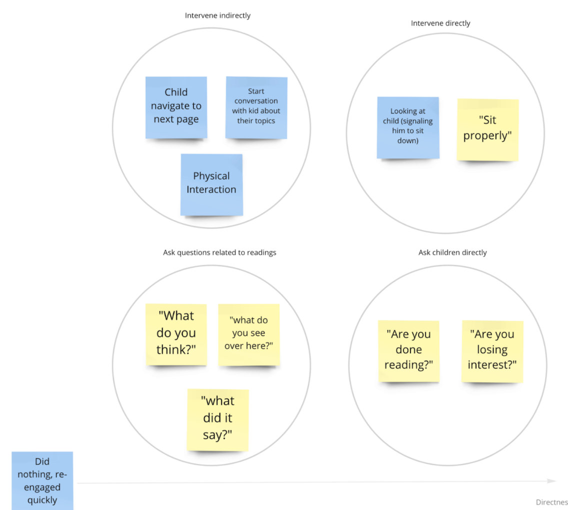

Figure 10. Childrens' behaviors during disengagement

CONCLUSION AND FUTURE WORK

Our work covers a broad, high-level overview of the role of a parent's nonverbal cues in their child's engagement to a dialogic reading task. Due to the scope of this analysis, there are certainly limitations to our work as well as more specific future explorations.

A limitation in our modeling approach is a lack of understanding of the most informative features in engagement prediction. While we conducted a preliminary feature analysis, further work is needed to expose the most predictive features. We noticed that a parent-child dyad's demographic features and behavioral questionnaires were related to the child's attention span. Future work may include augmenting the model with these demographic features. Additionally, many children sought attention from their parent during the task. With regards to using the parent's nonverbal pose features as our focus, we could improve model performance by extracting higher-level informative features. For example, "parent leaning forward," "parent child eye contact" might be more useful than raw numerical metrics. Additionally, incorporating time into our model would likely further improve performance.

Based on our results, there are a number of future research directions that can expand upon our work. In this paper, we focused solely on a child's engagement during a learning task. However, other concepts such as rapport, synchrony, reciprocity, attention-seeking behavior, and interpersonal connectedness would be interesting areas of focus. Since part of our motivation is to develop better agent interventions in facilitating learning tasks, studying the effects of nonverbal behavior across human-human and human-agent dyadic interactions is another important direction. We cannot naturally assume that people respond similarly to a human intervention as opposed to an agent (embodied or virtual), but we can understand what an agent cannot do and use human interventions to motivate further studies. Finally, we recognize that context of the interaction is crucial in understanding affect. A parent waving their hand to their child might not mean that the child is disengaged, unless we observe that the parent also asks the child to concentrate. This work could be expanded with increased modalities, such as linguistic characteristics of the utterances between parents and children.

ACKNOWLEDGMENTS

We thank Rosalind Picard and Neska Elhaouij for their continuous guidance on this project. We also thank Huili Chen, Hae won Park, and Sharifa Alghowinem helping us work with the DAMI-P2C dataset and providing additional guidance on the project.

REFERENCES

[1] Cynthia Breazeal, Cory D Kidd, Andrea Lockerd Thomaz, Guy Hoffman, and Matt Berlin. 2005. Effects of nonverbal communication on efficiency and robustness in human-robot teamwork. In *2005 IEEE/RSJ international conference on intelligent robots and systems*. IEEE, 708–713.

[2] Susan E Brennan and Herbert H Clark. 1996. Conceptual pacts and lexical choice in conversation. *Journal of experimental psychology: Learning, memory, and cognition* 22, 6 (1996), 1482.

[3] Zhe Cao, Gines Hidalgo, Tomas Simon, Shih-En Wei, and Yaser Sheikh. 2019. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. *IEEE transactions on pattern analysis and machine intelligence* 43, 1 (2019), 172–186.

[4] Tanya L Chartrand and John A Bargh. 1999. The chameleon effect: the perception–behavior link and social interaction. *Journal of personality and social psychology* 76, 6 (1999), 893.

[5] Huili Chen, Sharifa Alghowinem, Soo Jung Jang, Cynthia Breazeal, and Hae Won Park. Dyadic Affect in Parent-child Multi-modal Interaction: Introducing the DAMI-P2C Dataset and its Preliminary Analysis. (????), 16.

[6] Huili Chen, Sharifa Alghowinem, Soo Jung Jang, Cynthia Breazeal, and Hae Won Park. submitted. Dyadic Affect in Parent-Child Multi-Modal Interaction: Introducing the DAMI-P2C Dataset and Its Preliminary Analysis. *Transactions on Affective Computing* (submitted).

[7] Huili Chen, Hae Won Park, Xiajie Zhang, and Cynthia Breazeal. 2020a. Impact of Interaction Context on the Student Affect-Learning Relationship in Child-Robot Interaction. In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, Cambridge United Kingdom, 389–397. DOI: http://dx.doi.org/10.1145/3319502.3374822

[8] Huili Chen, Yue Zhang, Felix Weninger, Rosalind Picard, Cynthia Breazeal, and Hae Won Park. 2020b. Dyadic speech-based affect recognition using dami-p2c parent-child multimodal interaction dataset. In *Proceedings of the 2020 International Conference on Multimodal Interaction*. 97–106.

[9] Vivienne M Colegrove and Sophie S Havighurst. 2017. Review of nonverbal communication in parent–child relationships: Assessment and intervention. *Journal of Child and Family Studies* 26, 2 (2017), 574–590.

[10] Sheri E Davis. An Enhanced Dialogic Reading Approach to Facilitate Typically Developing Pre-School Children's Emergent Literacy Skills. (????), 81.

[11] Ap Dijksterhuis and John A Bargh. 2001. The perception–behavior expressway: Automatic effects of social perception on social behavior. (2001).

[12] Ruth Feldman, Romi Magori-Cohen, Giora Galili, Magi Singer, and Yoram Louzoun. 2011. Mother and infant coordinate heart rhythms through episodes of interaction synchrony. *Infant Behavior and Development* 34, 4 (2011), 569–577.

[13] Riccardo Fusaroli and Kristian Tylén. 2012. Carving language for social coordination: A dynamical approach. *Interaction studies* 13, 1 (2012), 103–124.

[14] Riccardo Fusaroli and Kristian Tylén. 2016. Investigating conversational dynamics: Interactive alignment, Interpersonal synergy, and collective task performance. *Cognitive science* 40, 1 (2016), 145–171.

[15] Simon Garrod and Martin J Pickering. 2009. Joint action, interactive alignment, and dialog. *Topics in Cognitive Science* 1, 2 (2009), 292–304.

[16] E Hatfield and J Cacioppo. 1994. Emotional contagion. New York: Cambridge University Press ISBN 0-521-44948-0. (1994).

[17] Jacqueline M Kory Westlund, Sooyeon Jeong, Hae W Park, Samuel Ronfard, Aradhana Adhikari, Paul L Harris, David DeSteno, and Cynthia L Breazeal. 2017. Flat vs. expressive storytelling: Young children's learning and retention of a social robot's narrative. *Frontiers in human neuroscience* 11 (2017), 295.

[18] Mariska E Kret. 2015. Emotional expressions beyond facial muscle actions. A call for studying autonomic signals and their impact on social perception. *Frontiers in psychology* 6 (2015), 711.

[19] Chloë Leclère, Sylvie Viaux, Marie Avril, Catherine Achard, Mohamed Chetouani, Sylvain Missonnier, and David Cohen. 2014. Why synchrony matters during mother-child interactions: a systematic review. *PloS one* 9, 12 (2014), e113571.

[20] Jin Joo Lee, Cynthia Breazeal, and David DeSteno. 2017a. Role of Speaker Cues in Attention Inference. *Frontiers in Robotics and AI* 4 (Oct. 2017), 47. DOI: http://dx.doi.org/10.3389/frobt.2017.00047

[21] Jin Joo Lee, Cynthia Breazeal, and David DeSteno. 2017b. Role of speaker cues in attention inference. *Frontiers in Robotics and AI* 4 (2017), 47.

[22] Max M Louwerse, Rick Dale, Ellen G Bard, and Patrick Jeuniaux. 2012. Behavior matching in multimodal communication is synchronized. *Cognitive science* 36, 8 (2012), 1404–1426.

[23] Kerry L Marsh, Michael J Richardson, and Richard C Schmidt. 2009. Social connection through joint action and interpersonal coordination. *Topics in Cognitive Science* 1, 2 (2009), 320–339.

[24] Fatemeh Noroozi, Dorota Kaminska, Ciprian Corneanu, Tomasz Sapinski, Sergio Escalera, and Gholamreza Anbarjafari. 2018. Survey on emotional body gesture recognition. *IEEE transactions on affective computing* (2018).

[25] Richard V Palumbo, Marisa E Marraccini, Lisa L Weyandt, Oliver Wilder-Smith, Heather A McGee, Siwei Liu, and Matthew S Goodwin. 2017. Interpersonal autonomic physiology: A systematic review of the literature. *Personality and Social Psychology Review* 21, 2 (2017), 99–141.

[26] Daniel C Richardson and Rick Dale. 2005. Looking to understand: The coupling between speakers' and listeners' eye movements and its relationship to discourse comprehension. *Cognitive science* 29, 6 (2005), 1045–1060.

[27] MEREDITH L. ROWE. 2008. Child-directed speech: relation to socioeconomic status, knowledge of child development and child vocabulary skill. *Journal of Child Language* 35, 1 (2008), 185–205. DOI: http://dx.doi.org/10.1017/S0305000907008343

[28] Kevin Shockley, Marie-Vee Santana, and Carol A Fowler. 2003. Mutual interpersonal postural constraints are involved in cooperative conversation. *Journal of Experimental Psychology: Human Perception and Performance* 29, 2 (2003), 326.

[29] Tanmay Sinha and Justine Cassell. 2015. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *Proceedings of the 1st workshop on modeling interpersonal synchrony and influence*. 13–20.

[30] Linda Tickle-Degnen and Robert Rosenthal. 1990. The nature of rapport and its nonverbal correlates. *Psychological inquiry* 1, 4 (1990), 285–293.

[31] Jacqueline M Kory Westlund, Leah Dickens, Sooyeon Jeong, Paul L Harris, David DeSteno, and Cynthia L Breazeal. 2017. Children use non-verbal cues to learn new words from robots as well as people. *International Journal of Child-Computer Interaction* 13 (2017), 1–9.

[32] Yue Zhang, Felix Weninger, Schuller Björn, and Rosalind Picard. 2019. Holistic affect recognition using PaNDA: paralinguistic non-metric dimensional analysis. *IEEE Transactions on Affective Computing* (2019).