SpecialTime: Automatically Detecting Dialogue Acts from Speech to Support Parent-Child Interaction Therapy

Bernd Huber
Harvard University
Cambridge, MA, USA
bhb@seas.harvard.edu

Richard F. Davis III
Auburn University
Auburn, AL, USA
rfd0005@tigermail.auburn.edu

Allison Cotter
Auburn University
Auburn, AL, USA
amc0093@tigermail.auburn.edu

Emily Junkin
Auburn University
Auburn, AL, USA
ekj0010@tigermail.auburn.edu

Mindy Yard
Carroll County Youth Service Bureau
Westminster, MD, USA
myard@ccysb.org

Stuart Shieber
Harvard University
Cambridge, MA, USA
shieber@harvard.edu

Elizabeth Brestan-Knight
Auburn University
Auburn, AL, USA
brestev@auburn.edu

Krzysztof Z. Gajos
Harvard University
Cambridge, MA, USA
kgajos@eecs.harvard.edu

ABSTRACT

Parent-child interaction therapy (PCIT) helps parents improve the quality of interaction with children who have behavior problems. The therapy trains parents to use effective dialogue acts when interacting with their children. Besides weekly coaching by therapists, the therapy relies on deliberate practice of skills by parents in their homes. We developed SpecialTime, a system that provides parents engaged in PCIT with automatic, real-time feedback on their dialogue act use. To do this, we first created a dataset of 6,022 parent dialogue acts, annotated by experts with dialogue act labels that therapists use to code parent speech. We then developed an algorithm that classifies the dialogue acts into 8 classes with an overall accuracy of 78%. To test the system in an actual clinical setting, we conducted a one month pilot study with four parents currently in therapy. The results suggest that automatic feedback on spoken dialogue acts is possible in PCIT, and that parents find the automatic feedback useful.

KEYWORDS

Feedback, Healthcare, Machine Learning, Parent-Child Interaction, Therapy

ACM Reference Format:

Bernd Huber, Richard F. Davis III, Allison Cotter, Emily Junkin, Mindy Yard, Stuart Shieber, Elizabeth Brestan-Knight, and Krzysztof Z. Gajos. 2019. SpecialTime: Automatically Detecting Dialogue Acts from Speech to Support Parent-Child Interaction Therapy. In *Proceedings of PervasiveHealth conference (PervasiveHealth'19)*. ACM, New York, NY, USA, Article 4, 10 pages. https://doi.org/10.475/123_4

PervasiveHealth'19, May 2019, Trento, Italy

ACM ISBN 123-4567-24-567/08/06.
https://doi.org/10.475/123_4

1 INTRODUCTION

Early behavior problems of children have been the focus of considerable theoretical and empirical work [6, 7, 13, 20, 35]. In addition to the high prevalence, ranging from 15 to 34% [26, 38, 49], early behavior problems have been shown to be predictive of more serious disorders such as attention-deficit/ hyperactivity disorder (ADHD) [8, 34, 36, 39, 41].

Parent-Child Interaction Therapy (PCIT) is designed to help parents of children with early behavior problems to improve their relationship with their child and to manage their child's behavior more effectively [5, 16]. A core skill taught and practiced in PCIT is the dos and don'ts of child-directed spoken interaction: parents are taught a set of dialogue acts that they should use frequently when interacting with their child (such as Labeled Praise, for example "Thank you for helping me") and another set that they should avoid (such as Negative Talk, for example "Don't do that"). Besides weekly coaching by therapists, the therapy relies on the deliberate at-home practice of skills by parents. During therapy sessions, parents benefit from feedback by therapists on their behavior. During their at-home practice, however, parents currently don't receive any feedback.

Informed by our formative study with four PCIT clinicians and two parents currently in therapy, we designed and implemented the SpecialTime system for providing parents with feedback during their at-home practice of PCIT skills. To do that, we first collected a set of 6,022 expert-generated and annotated utterances of child-directed utterances. The data were annotated using the Dyadic Parent-Child Interaction Coding System (DPICS), a dialogue act classification scheme used in PCIT [16]. Using this dataset, we developed a system that automatically classifies child-directed dialogue acts into the eight DPICS classes at an overall accuracy of 79%, a reasonable performance as compared to therapist agreement rates of 80% [5]. We then designed a real-time interactive system for parents to receive real-time feedback on the spoken dialogue acts they use when interacting with their child.

Child: It's yellow like her arms!
Parent: Good job noticing that connection (**Labeled Praise**)
Parent: I put the red one on mine (**Neutral Talk**)
Child: Now she needs a mouth. So does yours!
Parent: Both of our potato heads need mouths (**Reflection**)
Parent: You're looking through the bin (**Behavioral Description**)
Child: Your potato head can have the tooth.
Parent: Good choice (**Unlabeled Praise**)
Child: My potato head will have the red lips.
Parent: Are we missing anything else? (**Question**)

Table 1: Example DPICS codes for a natural parent-child interaction. The coding system assigns one of eight dialogue act labels to every dialogue act in the parent-child interactions. These dialogue acts are coded by therapists during clinic sessions.

To better understand how SpecialTime works in real therapy settings, we conducted a pilot study over the duration of one month with four parents who were enrolled in PCIT. In our study, we found that SpecialTime could detect the dialogue acts, and that dialogue act detections from SpecialTime aligned with therapist assessments of in-clinic parent behavior. Our study shows that parents generally found real-time feedback and progress tracking functionality provided by the SpecialTime app useful.

In summary, we make the following contributions:

- The design and development of the SpecialTime real-time feedback system to support parents in therapy.
- A dataset and detection algorithm for classifying dialogue acts (using the DPICS classification scheme) from speech in parent-child interaction.
- A pilot study with four parents currently in therapy showing the feasibility and value of SpecialTime as a way to amplify the effectiveness parent-child interaction therapy.

The results presented in this paper inform the next iteration of a redesign of SpecialTime and the design of a formal randomized controlled trial to assess the effectiveness of the approach on a larger sample.

2 BACKGROUND AND RELATED WORK

2.1 Parent-Child Interaction Therapy

Parent-Child Interaction Therapy (PCIT) is a behavior therapy designed to improve the quality of interaction between parents and their children, ultimately supporting parents to manage their child's behavior more effectively [5]. As part of the therapy, parents are taught a set of dialogue acts to include frequently in their interaction with their child, and another set of dialogue acts to avoid. During the weekly clinical visits, a parent's interaction with the child is observed by a therapist and the parent's child-directed speech is coded by a therapist using the Dyadic Parent-Child Interaction Coding System (DPICS, see Table 1 for a sample annotated dialogue) [16].

Table 2 describes the eight classes that are included in the DPICS scheme. The DPICS coding scheme is ordered; for cases in which one sentence matches multiple classes, and hence therapists could place it in multiple classes, therapists are instructed to code the utterance with the class highest up in the ordering. The default class—*Neutral Talk*—captures sentences that are not assumed to have a direct impact on the parent's interactions with the child.

There are three desirable dialogue acts that are of particular interest when assessing parents' therapy progress, which are Labeled Praise, Behavior Description and Reflection. The mastery criteria that are needed to progress through therapy are to achieve ten dialogue acts of each of these three classes during one 5-minute parent-child interaction session. In addition, parents need to have fewer than 3 dialogue acts in each of the three undesirable categories, which are Question, Command and Negative Talk. Neutral Talk and Unlabeled Praise are treated as inconsequential.

2.2 Health and Parenting Technology Support

There is previous research in human-computer interaction (HCI) on parenting that had the goal to develop novel technological intervention capabilities. Pina et al. [40], for example, studied just-in-time stress coping interventions for parents of children with ADHD and found that prompts delivered just before a full escalation of stress were more useful as parents were especially receptive to an intervention strategy at that time. MOBERO [46] is a smartphone-based system to assist families of children with ADHD by encouraging the children to become more independent. TalkLIME is a mobile system to provide intervention to the parents, with a focus on parent-child interaction for children with language delay [45]. WAKEY teaches parents communication strategies for conflict resolution in morning routines [9]. Slovak et al. [44] discuss opportunities for social-emotional learning in education and how HCI can help understanding how technology can support related strategies in child education. Finally, TalkBetter uses non-verbal cues to provide parents personalized, real-time feedback on communication strategies [21], showing the feasibility of socially-driven care. These findings suggest that feedback on language use is important in various parenting settings, and we extend these findings to the domain of PCIT, providing parents feedback on spoken dialogue acts.

2.3 Self-tracking and Feedback Systems

Personal informatics tools help people understand their habits and behaviors [29]. Existing devices and apps often focus on tracking physical fitness [1, 10, 30, 31], sleep [1, 24], diet [4, 12, 33], smoking [2], and stress [37]. The primary focus of many existing solutions is to support a high-level health goal, such as staying healthy or sleeping better. More communication-centered tracking devices have also been studied previously. For example, the SocioPhone tracks non-linguistic communication patterns over time, showing the feasibility of long-term tracking of such signals [28]. There are also child-speech focused feedback systems. For example, Hailpern et al show that visualizations could support communication of children with ASD [19].

Studies have shown that when using real-time feedback in combination with tracking capabilities, electronic support for health behavior change becomes more effective when compared to a system without feedback [3]. Lee and Dey showed that providing self-monitoring feedback on a tablet display was effective for adhering to a medical regime [27]. Our SpecialTime system builds on this

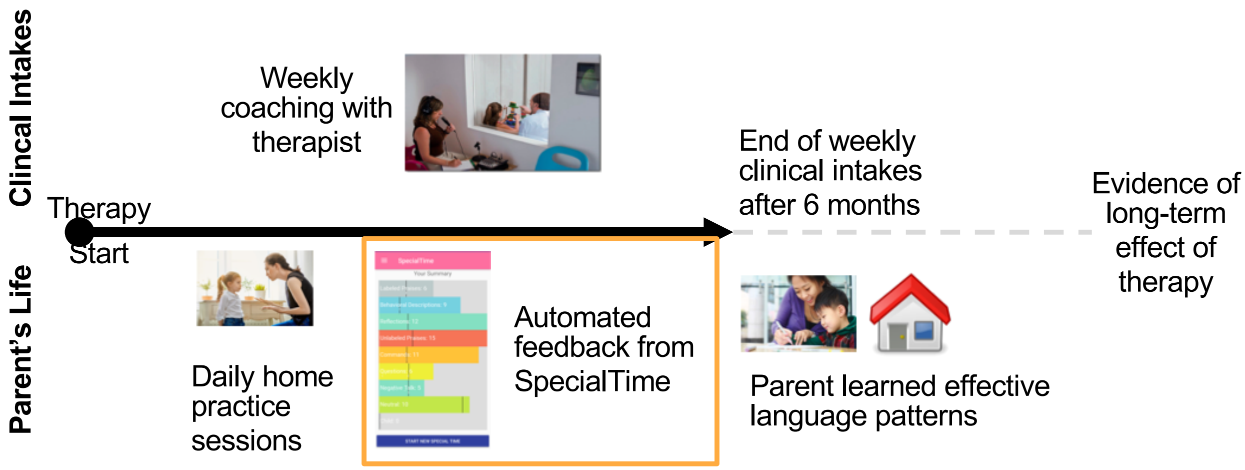

Figure 1: Outline of therapy flow and homework routine in parent-child interaction therapy. A typical therapy takes about six months from initial encounters and diagnosis until parents graduate from the therapy. An essential component of the therapy is skills practice at home and in-session coaching feedback from a therapist. The SpecialTime system supports the parent home practice through feedback, in which previously only paper-based tracking questionnaires were used.

work by providing parents feedback on their language behavior skills when interacting with their children.

2.4 Automatic Language-Based Assessment in Behavior Therapy

Language interactions are a major component of mental health assessment and treatment, and thus a useful lens for mental health analysis [11]. Few studies have investigated the labeling of dialogue acts in face-to-face interaction for therapeutic assessment. One study investigated automatic therapist and patient behavior coding [47]. As of recently, the social media industry provides tools for supporting mental wellness through automated monitoring systems, such as the mental wellness detection system on Facebook [42]. However, current technologies still struggle with correctly classifying such high-level language signals as dialogue acts, suggesting the challenging nature of detecting mental health states automatically. In our work, we develop algorithms to detect dialogue acts from spoken language for assessing parent-child interaction quality.

3 DESIGN OF SPECIALTIME

There is ample evidence that parents' practice with real-time feedback from therapists contributes to the effectiveness of PCIT [16, 17]. It has also been shown that active observation in therapy causes parents to more actively pay attention to the skills that they should practice, leading to more effective training [16]. For these reasons, we assumed that these two mechanisms (timely feedback and automatic monitoring of the frequency and quality of practice) would be important components of the solution.

3.1 Formative Interviews

To inform the design of the SpecialTime system, we conducted semi-structured interviews with four therapists (three female, one male, all were currently in training or in practice of parent-child interaction therapy), and with two parents who were currently in therapy. We asked therapists to describe their current workflow with patients and at-home practice, as well as the issues that they might observe with patients or that patients report. We asked parents whether they find homework paper sheets that are currently being used helpful, and what they found problematic about them.

The therapists consistently argued that at-home practice is an essential part of the therapy, however, they currently rely solely on self-reported data by parents. In current practice, parents fill out a tracking sheet with time of practice, things that went well, things that did not go well, and what the practice activities were. This process is burdensome to many parents, and parents reported that they did not see much value in filling out the paper tracking sheets and that they preferred practicing with the therapist because they would receive feedback on their skills. Furthermore, this burden is often mentioned as a reason for parents dropping out of the therapy. Parents reported that they often did not practice at home because of lack of time, because they did not feel it helped them very much with the progress in therapy, or because of a general lack of motivation. Aligned with what HCI research on self-tracking previously found [23], parents reported that they found the paper-based homework sheet complex and burdensome to complete during every practice.

These insights indicate that both therapists and parents might value an automated mechanism for logging the at-home practice: therapists because it offers additional value to self-reports, and parents because it frees them from having to perform a burdensome manual tracking task. The log should only capture when the practice occurred and, perhaps, how many dialogue acts of each class were used. It should not capture the content of the conversation. Furthermore, timely feedback on at-home practice may increase the perceived value of the practice for the parents given that feedback is an appreciated aspect of practice with the therapists.

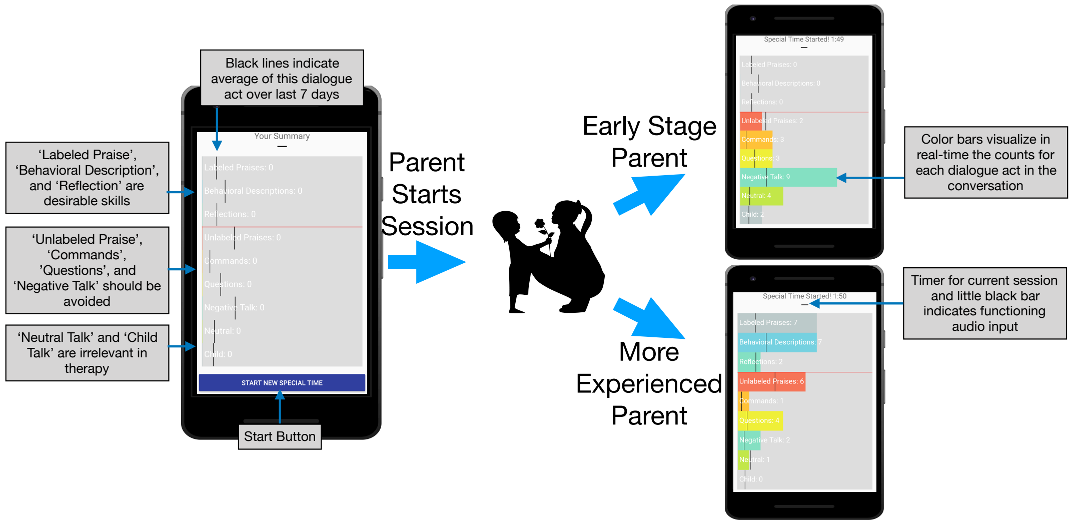

Figure 2: The *SpecialTime* user interface. The spoken conversation is transcribed and utterances classified into one of eight classes. Reference bars (in black) show the parent's average scores over the sessions in the last 7 days (same interval as clinical encounters). Parents start a new session by clicking on the 'Start New Special Time' button at the bottom of the screen. Dialogue acts are then classified in real-time by the system.

3.2 System Design

The design goals of SpecialTime are to provide parents with an easier way (compared to the paper sheets) to document their practice and to actively support their at-home practices. To mimic therapist coaching and because in-situ feedback has been shown pedagogically effective, the user interface of SpecialTime provides parents with feedback in real-time on how they are doing during in-home practice sessions. Feedback is shown visually and in aggregate form to make it glanceable and less distracting than feedback that would always require visual attention. Figure 2 shows details on the system design.

The system automatically transcribes parent speech and classifies dialogue acts while parents interact with their children. This design frees parents from having to fill out paper forms because practice is logged automatically.

SpecialTime presents a visualization to parents, providing an aggregate view of dialogue acts they used in a practice session. The counts are presented in a bar chart design, with the three topmost bars showing desirable dialogue acts. Once a parent decides to practice with her child and clicks the start session button, she will receive feedback in real-time about the DPICS label counts. For example, if a parent says *Thank you for playing with me*, the bar indicating counts for Labeled Praises will increase by one. The system provides counts for each of the eight DPICS labels, with counts being updated as they were detected by the system, throughout a practice session. This feedback does not require continuous monitoring and is designed to be glanceable throughout a practice session.

SpecialTime also provides reference bars with parents' average number of labels for each of the different DPICS classes over the last seven days, as indicated by black bars on the charts (see Figure 2). For example, if a parent had on average four Labeled Praises during the last seven days, and in the current session would have six Labeled Praises, the bar providing the counts for Labeled Praise of the current session would be higher than the black average bar, indicating good progress. At the end of a practice session, parents can review the visualization of the counts of different classes of DPICS dialogue acts.

4 AUTOMATICALLY DETECTING DIALOGUE ACTS IN PARENT-CHILD INTERACTION FROM SPEECH

Given an audio stream of a parent-child interaction session, SpecialTime outputs the DPICS parent-child interaction labels, automatically detected from speech. This section describes the processing steps involved in the speech classification, with Figure 3 illustrating these steps.

To classify spoken parent-child interaction, SpecialTime first uses the *Google Cloud Speech Recognition Service*[1] to transcribe the content of the conversation. We chose this transcription service as it reported high accuracies, in combination with word-level timestamps. While the service provided streaming-input for real-time transcription, we chose to chunk the audio stream into 30-second snippets due to technical difficulties with the streaming-input from Google. The audio stream, therefore, is processed in 30-second chunks, providing updates to the SpecialTime visualization feedback every 30 seconds.

SpecialTime then segments the transcript into sentences using a neural network–based sentence segmentation algorithm [48]. We chose this segmentation approach since the Google transcription service did not provide punctuation at the time SpecialTime was built. All of the further processing steps are done using custom

[1] https://cloud.google.com/speech/

Class	Definition & Examples	Do/Don't
1. Negative Talk (NT)	Negative Talk is a verbal expression of disapproval of the child or the child's attributes, activities, products, or choices. Negative Talk also includes sassy, sarcastic, rude or imprudent speech. (1) Parent: No, don't put that piece there. (2) Child: can I get some ice cream later? Parent: Yeah, that's gonna happen (*sarcastic tone*).	Don't
2. Command (CMD)	Parent commands are statements in which the parent directs the behavior of the child. Commands may be direct or indirect in form. Commands include statements directing the child to perform vocal or motor behaviors as well as mental or internal, unobservable actions (e.g., think, decide). (1) Come over here. (2) Hand me your jacket.	Don't
3. Labeled Praise (LP)	Labeled Praise provides a positive evaluation of a specific attribute, product, or behavior of the child. Thank you for playing Legos with me. (2) Thank you for cleaning. (3) You drew a beautiful picture.	Do
4. Unlabeled Praise (UP)	An Unlabeled Praise provides a positive evaluation of the child, an attribute of the child, or a nonspecific activity, behavior, or product of the child. (1) I love you. (2) Thank you. (3) Child: I made a tower. Parent: Good Job.	-
5. Question (QU)	Questions request an answer but do not suggest that a behavior is to be performed by the other person. There are two types of questions: Information Questions request a verbal response beyond a 'yes' or 'no,' whereas Descriptive Questions request a simple affirmative or negative response. (1) Where does that go? (2) What are you doing? (3) Do you want the raspberries too?	Don't
6. Reflection (RF)	A Reflection is a declarative phrase or statement that has the same meaning as the child's verbalization. The reflection may repeat, paraphrase, or elaborate upon the child's verbalization but may not change the meaning of the child's statement or interpret unstated ideas. (1) Child: I did it. Parent: You did it. (2) Child: Which car is the fastest? Parent: You want to know which car is the fastest.	Do
7. Behavior Description (BD)	Behavior Descriptions are non-evaluative, declarative sentences or phrases in which the subject is the other person and the verb describes that person's ongoing or immediately completed observable verbal or nonverbal behavior. (1) You are sitting in the chair. (2) You are driving the car.	Do
8. Neutral Talk (NTA)	Neutral talk is comprised of statements that introduce information about people, objects, events, or activities, or indicate attention to the child but do not clearly describe or evaluate the child's current or immediately completed behavior. (1) Yes (2) Zebras have stripes. (3) I don't know what to do next.	-

Table 2: Definitions and example sentences for each of the eight different DPICS classes as presented by Eyberg et al. [16], Eyberg and Robinson [17], in priority order of the coding scheme. The labels are assigned to every utterance in the parent-child interaction.

trained processing units to minimize exposure of the sensitive speech data that SpecialTime analyzes.

As a second step, the system separates parent from child speech. To do this, the speaker characteristics of every segment are analyzed using speech prosody features commonly used for such tasks as speaker recognition [14]. Note that this audio processing step requires timestamped transcripts in order to align text and audio. The prosody features were extracted using the openSMILE voice feature extraction toolkit [15]. To classify parent and child speech automatically, we trained a prosody classifier speech model on an existing parent-child speech corpus that provides a set of 50,000 parent and child utterances, labeled with child and parent labels [32]. A cross-validation analysis showed that this parent-child speech classification model can separate parent from child speech at over 99% accuracy.

PCIT therapists use tone of voice to code dialogue acts as question, even when the lexical features suggest otherwise. For example, a therapist might code the sentence *I am drawing* as Neutral Talk; however with a rising final boundary tone, the correct label is Question. Our system mimics this by analyzing the pitch contour of the final boundary tone [43]. Specifically, we use openSMILE [15] to extract the first derivative of the pitch contour of the last 0.5 seconds of the spoken segment. If the derivative is greater than 0, the segment is classified as rising final boundary tone independent of the spoken words in the segment. Note that this audio processing step also requires timestamped transcripts in order to align text and audio.

As a final step, the SpecialTime system classifies every segment that was labeled as parent speech and not already labeled as question, into one of eight DPICS classes. To automatically classify the dialogue acts, each transcript segment is represented as a feature vector using a text frequency-inverse document frequency (TFIDF) representation. Specifically, the TFIDF representation is learned on the training set and represents each sentence as a vector, giving each word in the sentence a score and a fixed position in the vector. SpecialTime also automatically tags the text with part-of-speech (POS) tags and uses both words and POS to train the TFIDF vectors. We train the TFIDF vectors with a combined uni- and bigram feature representation. Bigram models use two consecutive words as a feature, and therefore the word order is taken into account.

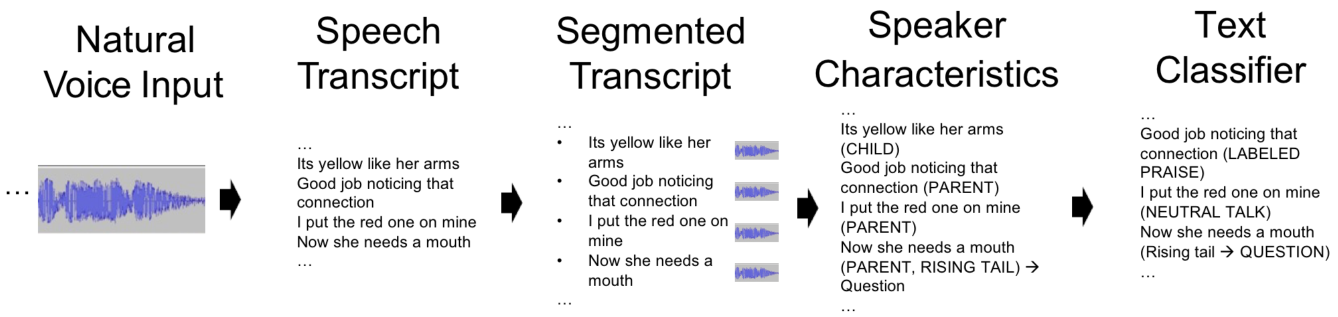

Figure 3: An overview of the information flow with the different computational units incorporated in the system architecture of SpecialTime. Speech first gets transcribed and segmented between pauses. Utterances are then automatically separated into child and parent utterances. The transcript is then further segmented into separate sentences using a neural network segmentation algorithm. Every utterance is then classified with our classifier. This gives the automatically tagged dialogue act label.

Due to simplicity and ability to handle relatively small datasets, we decided to train a linear support vector machine (SVM) as text classifier, which has previously proven effective for similar text classification tasks [22].

While this setup requires multiple processing steps, we decided to train a custom system and minimize data exposure due to privacy reasons. However, we provide the expert-curated dataset that we created to train the dialogue act classifier, and one could also use an external text classification service with this dataset, depending on the privacy requirements. The expert-curated dialogue act dataset that we used in this work is described in the next section.

To meet the privacy requirements that arise in actual clinical use, SpecialTime discards the captured audio and the transcripts as soon as the dialogue acts are classified. Only the dialogue act counts are retained.

	NT	CMD	LP	UP	QU	RF	BD	NTA
NT	0.81	0.04	0.02	0.04	0.03	0.02	0	0.06
CMD	0.05	0.79	0.01	0.03	0.04	0.01	0.01	0.06
LP	0.01	0	0.87	0.05	0	0	0.02	0.03
UP	0.02	0.02	0.02	0.89	0.01	0.01	0.01	0.02
QU	0.03	0.08	0.01	0.03	0.75	0.05	0.03	0.03
RF	0.02	0.05	0.02	0.05	0.04	0.41	0.18	0.23
BD	0.01	0.03	0.03	0.02	0.01	0.05	0.82	0.04
NTA	0.05	0.07	0.04	0.04	0	0.05	0.03	0.7

Figure 4: Confusion matrix of utterance classifier in the DPICS scheme with predictions in rows and actual labels in columns. The values are average fractions of ten-fold cross validations. The overall accuracy of the classifier trained on our dataset is 78.3%**, an acceptable performance as compared to reported therapist agreement rates of** 80% **in live-coding sessions.**

4.1 Parent-Child Interaction Dataset

To train the dialogue act classifier, we first created an expert-annotated dataset. Specifically, we collected 6,022 utterance samples provided voluntarily by 192 therapists working in the PCIT field. Therapists were recruited through a nation-wide PCIT mailing list. Therapists on this mailing list are clinical psychologists who are currently practicing PCIT in clinical settings. Therapists were asked to provide examples of parent utterances matching the different DPICS classes that they may observe in actual parent-child interaction settings. We could not use real transcripts and label them due to privacy concerns of therapists, so instead, we prompted therapists to come up with examples they could imagine occurring in actual therapy sessions. The dataset can be downloaded from https://doi.org/10.7910/DVN/C5Z3SC.

We used the expert-labeled corpus to learn the TFIDF text feature representations, as well as to train the dialogue act classifier. After pre-processing of the text, and automatically tagging of the text with part-of-speech (POS) tags, our model was trained on a total of 5,005 features.

4.2 Technical Evaluation

To evaluate the classifier, we used the TFIDF vectorization scheme with a linear support vector machine (SVM) classifier (C=0.1). We performed a ten-fold cross-validation analysis on the 6,022 utterances from our dataset. This led to a performance of 78.3% accuracy (79% precision, 77%recall) averaged over the eight DPICS classes (compare to 17% majority vote). This is an acceptable performance as compared to reported therapist agreement rates of 80% in live-coding sessions [5].

It is also important to note that not all DPICS codes are equally influential for therapy outcomes. For example, false positive *Neutral Talks* do not have much impact on the therapy assessment, whereas false positive *Reflections* may give a misleading positive signal to the parents. These different priorities are important when deploying such a system in real-world therapy settings.

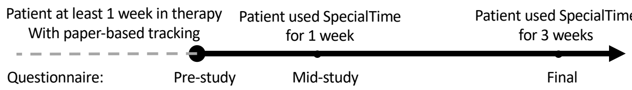

Figure 5: Distribution of the surveys and integration with the PCIT workflow. We provided parents a total number of three questionnaires, one time at the beginning of the therapy, after they used paper-based home practices for at least one week, then one week into the study, and then three weeks into the study.

5 PILOT STUDY: METHODS

We gathered preliminary participant feedback in an actual clinical setting in a primarily qualitative study. Such a qualitative study is the best practice for evaluating early-stage health technologies since it allows for understanding the *why* [25]. Our recruitment methods and study protocol were reviewed and approved by an Institutional Review Board.

Parents, who were enrolled in PCIT, received guidance from therapists on how to practice skills at their homes, and how to use the SpecialTime system for receiving feedback during their at-home practice sessions. In PCIT, parents are typically given a paper-based tracking sheet, and therapists review the records during therapy sessions. This process was replaced during our pilot study with handing out SpecialTime, and therapists reviewed the records from SpecialTime during therapy sessions. Our goal was to determine whether SpecialTime could successfully support parents in practicing their skills and whether it could successfully detect learning progress, and discover any challenges that parents encounter throughout.

For the purpose of the pilot study, parents were asked to use SpecialTime for three weeks. Parents were lent smartphones specifically for using SpecialTime. While SpecialTime works on any regular Android device with internet connection, we did not want to introduce any bias by only allowing participants that owned Android smartphones to participate in this study.

5.1 Recruitment

We recruited parents who either already started but were still at an early stage in the therapy, or who were about to start PCIT in a clinic. Parents were asked at the beginning of the therapy whether they would like to participate in our study.

5.2 Participants

We asked ten parents, out of whom four agreed to use SpecialTime. The four parents were on average 32.5 years old, three of them were female, and their children were on average 4.5 years old (range 3-6 years). Three parents were about to start therapy, one attended two sessions already.

5.3 Procedure

The study was divided into three parts: (1) a pre-study questionnaire that asked about parents' general habits with practicing skills in their homes, (2) completing the first week of practice with SpecialTime and then completing the mid-study questionnaire, and (3) completing two more weeks of practice with SpecialTime and completing a final questionnaire. Questionnaires were handed out and collected on paper during wait times at the clinic. Questionnaires included both Likert-scale and open-ended questions. No compensation besides the benefit of using SpecialTime was given to parents who participated. Figure 5 shows the timeline of our procedure.

Q1: I would recommend the app to other parents in therapy
Q2: I would continue using the app throughout therapy, if it were available
Q3: How often did you use the app when you practiced your skills at home during the last week?
Q4: The app was helpful for practicing my skills
Q5: What aspect of the at-home practice is most useful?
Q6: How did the at-home practice fit into your daily routine?
Q7: What did you learn by tracking your at-home practice?
Q8: Describe how you used the app for your practice at home:
Q9: The real-time feedback of the app was helpful? Why?
Q10: Tracking my homework with the app was helpful?
Q11: I think that I learned something from the app feedback.
Q12: What aspect of the app was most useful?
Q13: What aspect of the app was most problematic?

Table 3: Questionnaires handed out to participants during therapy encounters. These questions probe the overall helpfulness of the system, parents adherence to their homework sessions and the parents perceived learning of skills .

5.4 Data Analysis

Quantitative analysis consisted of calculating participant adherence (self-reported frequency of use of SpecialTime and data logs from the system), analyzing usability ratings, and analyzing the change in PCIT skills during the study using both automatic assessments made by SpecialTime during at-home practice and the assessment made by therapists during in-clinic visits.

For qualitative data analysis, we collected all responses from all participants from the questionnaires. We then clustered the responses using affinity diagramming [18].

6 PILOT STUDY: RESULTS

6.1 Data Overview

In total, we collected the three questionnaires per parent. Furthermore, we collected a total of 45 completed, 5-minute at-home practice sessions with SpecialTime, with parents' median usage of three sessions per week in which they used the app.

6.2 Helpfulness

We asked parents whether the app was overall helpful for practice, which aspects they found most useful or problematic, how useful the real-time feedback was, and whether the tracking functionality was helpful (see *Q4,5,9,10,12,13*, Table 3).

Parents had positive experiences with PCIT practice and SpecialTime. We probed the usefulness in the mid and final questionnaires. Parents agreed that the app was overall useful for their home practices, and all parents found the real-time feedback and tracking

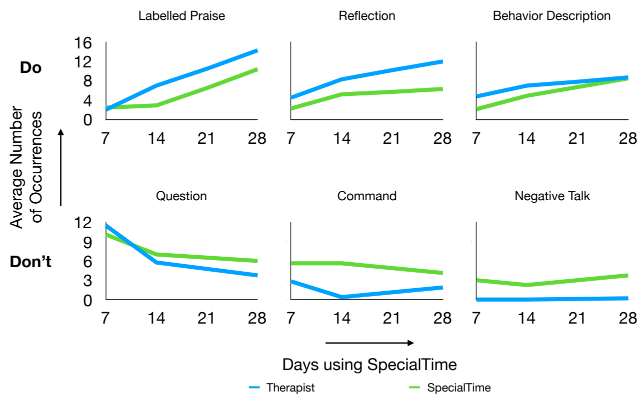

Figure 6: Number of occurrences of the dialogue acts per 5 minute session separated by *do* and *don't*, over the days when SpecialTime was used. Values are averaged over participants, per week.

functionalities helpful. Compared to their prior attempts to track their homework, participants appreciated the feedback and tracking functionalities: *"It helps me to remember which skills need to be practiced, and how I did overall."* (P1). Parents also liked the fact that the app tracked just the fact that they did their practice: *"It helps me remembering that I did the practice in the first place"* (P3).

6.3 Therapy Progress

To evaluate if SpecialTime can accurately assess parents' performance, we compared per-session counts of the use of the DPICS dialogue acts reported by SpecialTime to the counts produced by the therapists. Note that SpecialTime was used during at-home practice while therapists observed in-clinic sessions. Thus, we expected the two sets of counts to differ, but we also expected that week-to-week trends observed in the clinic would be also visible during at-home practice.

The results, illustrated in Figure 6, show that the counts reported by SpecialTime and the therapists were aligned, particularly for the three dialogue acts most relevant for assessing parents' progress in treatment, which are Labeled Praise, Behavior Description, and Reflection.

When asked how SpecialTime supported their learning (see *Q7,11* in Table 3), all parents agreed that the feedback through SpecialTime made them reflect on the interactions with their children and helped them practice their skills: *"[It helps me to see] which skills need more work, which ones are better"* (P3). One parent mentioned that the app kept classifying some negative speech while they thought that it was wrong, which then affected their trust in the app: *"Sometimes I wasn't sure if it detects all my skills correctly"* (P4). In the open-ended responses, we also found that some parents seemed to doubt themselves, while others doubted the accuracy of the app *"It showed me where I should improve, although I am not sure it was correct about all my skills"* (P3).

6.4 Skill Practice

Finally, we wanted to see how SpecialTime affected parents' skill practice routines. To assess this, we asked parents how SpecialTime fit into the daily skill practice, how parents used the app in their homes, and whether they could imagine themselves using SpecialTime after the end of our study (see *Q1,2,3,6,8*, Table 3). Three of the four participants reported having used SpecialTime most of the time when they did a practice session. Parents reported that SpecialTime *"fit into their daily routine the same way the paper-based homework sheet"* (P3). Reasons for not using the app varied. One parent reported that the app did *"not always recognize my voice"* (P4) and that the *"technology was distracting to the child"* (P4). Three out of four parents reported that they would recommend the app to other parents and would continue to use it if it were available.

7 DISCUSSION

We evaluated SpecialTime by deploying it for one month with four parents who were participating in PCIT. Our results suggest that parents found the SpecialTime system is useful in two ways: First, it provided feedback on their performance. Even though parents were not always certain if the exact counts reported by SpecialTime were accurate, they still reported that the system gave them a useful idea of how they were doing. Second, parents appreciated the fact that SpecialTime freed them from having to fill out paper reports on their at-home practice.

Our results show that the counts of dialogue acts captured by SpecialTime during at-home practice aligned over time with the counts coded by therapists during in-clinic sessions. While we do not have further insights into the effect of such different sources of error as background noise on performance, this relationship suggests the feasibility of SpecialTime for capturing the parents' skill gain.

Our findings suggest that SpecialTime was effective through its automatic dialogue act detection, which could not have been achieved with a similarly designed manual tracking app. This implies that it is possible to amplify the effectiveness of therapy, and potentially reduce drop out rates, with speech detection. In our current design, the dialogue act labels are shown without any coaching module, and without any recommendations of what to do next, as therapists do during therapy. One possible future direction is to develop and evaluate such a coaching capability.

Our data also revealed a median of three practice sessions per week by parents in their homes. Since PCIT therapists commonly rely on self-reported data on how many times parents practice, SpecialTime offers potential value to therapists by providing a more objective way to verify how much practice actually takes place.

Some technical challenges remain. For example, the dialogue act class Reflection shows lower recall rates in our technical evaluation, as compared to the other DPICS classes. We suspect that this lower accuracy comes from the fact that the label requires a contextual understanding of the parent's and child's intentions, as expressed by a preceding child utterance. While our current system does not include analysis of context, we suspect that future work including contextual information will improve the performance.

Some parents showed skepticism about whether the app was classifying all of their dialogue acts correctly. This is a valid concern, as our current system has multiple sources of errors, from speech transcription to classifier accuracy. We designed the user interface of SpecialTime such that it provided the feedback in real-time of spoken dialogue acts. This made erroneous classifications potentially more impactful. In the next iteration of the system, we will explore delaying feedback to the end of the session. We will also present the feedback in a manner more mindful of the inherent uncertainty about the exact counts returned by SpecialTime. We can accomplish this, for example, by telling parents which of the desired dialogue acts they did particularly well, and which they should continue working on. It is likely that, given how demanding caring for a child with behavioral problems is, it is more important for SpecialTime to motivate parents by offering encouragement and credit for their effort than to provide them with exact numeric feedback.

Our data analysis revealed various directions of potential improvements for the design of the SpecialTime user interface. First, as mentioned in the previous paragraph, parents may benefit equally from SpecialTime when being shown the assessment only after the session, as compared to our current design which shows feedback in real-time. One further step towards building trust with the SpecialTime system may include just showing the three desirable skills (the Do's: Labeled Praise, Reflection, and Behavior Description), and not the full list of skills. While parents are trained on all of the skills, a more encouraging design could highlight the three positive skills, ultimately leading to potentially better therapy.

8 LIMITATIONS

One of the limitations of our study is the small sample size available, limiting the range of analyses we were able to do on user experience. However, given the relatively high sensitivity and expense of collecting data in the domain we studied, our results still show valuable insights. Secondly, there are still open questions regarding how to most effectively design an interface for therapy feedback systems like SpecialTime, and we hope that our work initiates further HCI research in this direction. Finally, our recruitment approach may have oversampled parents who are more receptive to technology and from higher socio-economic groups. Future studies will need a broader variety of user background.

9 CONCLUSION

Parent-child interaction therapy (PCIT) trains parents of children with behavioral problems to use language that supports the development of children. Through weekly encounters with therapists and daily at-home practice, the therapy trains parents to use a set of dialogue acts (Labeled Praise, Reflection and Behavior Description) and avoid another set of dialogue acts (Negative Talk, Command and Question). Until now, parents did not have a way to receive feedback on their skills without the presence of therapists, limiting the effectiveness of at-home practice and therefore the therapy overall.

We designed and developed SpecialTime, a system that analyzes parents' speech and provides real-time feedback on PCIT skills. In a pilot study with four parents currently in therapy, we found that SpecialTime can provide such feedback without the presence of therapists, detecting skills aligned with what therapists coded during weekly practice sessions.

Our research motivates further development of designing language feedback systems, as well as studies on their effect on behavior change. Future work should explore how such feedback could improve the efficiency of therapy workflows.

10 ONLINE APPENDIX

The expert-annotated dataset that we collected for this work can be found at https://doi.org/10.7910/DVN/C5Z3SC.

11 ACKNOWLEDGEMENTS

The work was funded in part by NIH grant R01CA204585 as part of the NSF/NIH Smart and Connected Health program, and a Harvard Mind Brain Behavior Graduate Student Award. We thank Mary Czerwinski for feedback on early versions of this work, and David Ramsay for work on initial prototypes. A special thank you goes to the experts who participated in our data collection efforts.

REFERENCES

[1] 2018. Fitbit. https://www.fitbit.com/.

[2] Amin Ahsan Ali, Syed Monowar Hossain, Karen Hovsepian, Md Mahbubur Rahman, Kurt Plarre, and Santosh Kumar. 2012. mPuff: automated detection of cigarette smoking puffs from respiration measurements. In *Proceedings of the 11th international conference on Information Processing in Sensor Networks*. ACM, 269–280.

[3] Jakob E Bardram, Mads Frost, Károly Szántó, Maria Faurholt-Jepsen, Maj Vinberg, and Lars Vedel Kessing. 2013. Designing mobile health technology for bipolar disorder: a field trial of the monarca system. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 2627–2636.

[4] Eric PS Baumer, Sherri Jean Katz, Jill E Freeman, Phil Adams, Amy L Gonzales, John Pollak, Daniela Retelny, Jeff Niederdeppe, Christine M Olson, and Geri K Gay. 2012. Prescriptive persuasion and open-ended social awareness: expanding the design space of mobile health. In *Proceedings of the ACM 2012 conference on Computer Supported Cooperative Work*. ACM, 475–484.

[5] Elizabeth V Brestan and Sheila M Eyberg. 1998. Effective psychosocial treatments of conduct-disordered children and adolescents: 29 years, 82 studies, and 5,272 kids. *Journal of clinical child psychology* 27, 2 (1998), 180–189.

[6] Lisa M Broidy, Daniel S Nagin, Richard E Tremblay, John E Bates, Bobby Brame, Kenneth A Dodge, David Fergusson, John L Horwood, Rolf Loeber, Robert Laird, et al. 2003. Developmental trajectories of childhood disruptive behaviors and adolescent delinquency: a six-site, cross-national study. *Developmental psychology* 39, 2 (2003), 222.

[7] Susan B Campbell. 2006. *Behavior problems in preschool children: Clinical and developmental issues*. Guilford Press.

[8] Susan B Campbell and Linda J Ewing. 1990. Follow-up of hard-to-manage preschoolers: adjustment at age 9 and predictors of continuing symptoms. *Journal of child psychology and psychiatry* 31, 6 (1990), 871–889.

[9] Meng-Ying Chan, Yi-Hsuan Lin, Long-Fei Lin, Ting-Wei Lin, Wei-Che Hsu, Chia-yu Chang, Rui Liu, Ko-Yu Chang, Min-hua Lin, and Jane Yung-jen Hsu. 2017. WAKEY: Assisting Parent-child Communication for Better Morning Routines. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*. ACM, 2287–2299.

[10] Sunny Consolvo, David W McDonald, Tammy Toscos, Mike Y Chen, Jon Froehlich, Beverly Harrison, Predrag Klasnja, Anthony LaMarca, Louis LeGrand, Ryan Libby, et al. 2008. Activity sensing in the wild: a field trial of ubifit garden. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 1797–1806.

[11] Glen Coppersmith, Mark Dredze, Craig Harman, Kristy Hollingshead, and Margaret Mitchell. 2015. CLPsych 2015 Shared Task: Depression and PTSD on Twitter.. In *CLPsych@ HLT-NAACL*. 31–39.

[12] Felicia Cordeiro, Elizabeth Bales, Erin Cherry, and James Fogarty. 2015. Rethinking the mobile food journal: Exploring opportunities for lightweight photo-based capture. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*. ACM, 3207–3216.

[13] Kenneth A Dodge and Gregory S Pettit. 2003. A biopsychosocial model of the development of chronic conduct problems in adolescence. *Developmental psychology* 39, 2 (2003), 349.

[14] Florian Eyben, Klaus R Scherer, Björn W Schuller, Johan Sundberg, Elisabeth André, Carlos Busso, Laurence Y Devillers, Julien Epps, Petri Laukka, Shrikanth S Narayanan, et al. 2016. The Geneva minimalistic acoustic parameter set (GeMAPS) for voice research and affective computing. *IEEE Transactions on Affective Computing* 7, 2 (2016), 190–202.

[15] Florian Eyben, Martin Wöllmer, and Björn Schuller. 2010. Opensmile: the munich versatile and fast open-source audio feature extractor. In *Proceedings of the 18th ACM international conference on Multimedia*. ACM, 1459–1462.

[16] SM Eyberg, MM Nelson, NC Ginn, N Bhuiyan, and SR Boggs. 2013. Dyadic Parent–Child Interaction Coding System (DPICS) comprehensive manual for research and training. 4th PCIT International. *Gainesville, FL* (2013).

[17] Sheila M Eyberg and Elizabeth A Robinson. 1981. *Dyadic parent-child interaction coding system: A manual*. American Psychological Association.

[18] Barney G Glaser. 1992. *Basics of grounded theory analysis: Emergence vs forcing*. Sociology press.

[19] Joshua Hailpern, Karrie Karahalios, and James Halle. 2009. Creating a spoken impact: encouraging vocalization through audio visual feedback in children with ASD. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 453–462.

[20] Stephen P Hinshaw. 2002. Preadolescent girls with attention-deficit/hyperactivity disorder: I. Background characteristics, comorbidity, cognitive and social functioning, and parenting practices. *Journal of consulting and clinical psychology* 70, 5 (2002), 1086.

[21] Inseok Hwang, Chungkuk Yoo, Chanyou Hwang, Dongsun Yim, Youngki Lee, Chulhong Min, John Kim, and Junehwa Song. 2014. TalkBetter: family-driven mobile intervention care for children with language delay. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing*. ACM, 1283–1296.

[22] Armand Joulin, Edouard Grave, Piotr Bojanowski, and Tomas Mikolov. 2016. Bag of tricks for efficient text classification. *arXiv preprint arXiv:1607.01759* (2016).

[23] Ravi Karkar, Jessica Schroeder, Daniel A Epstein, Laura R Pina, Jeffrey Scofield, James Fogarty, Julie A Kientz, Sean A Munson, Roger Vilardaga, and Jasmine Zia. 2017. Tummytrials: a feasibility study of using self-experimentation to detect individualized food triggers. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. ACM, 6850–6863.

[24] Matthew Kay, Eun Kyoung Choe, Jesse Shepherd, Benjamin Greenstein, Nathaniel Watson, Sunny Consolvo, and Julie A Kientz. 2012. Lullaby: a capture & access system for understanding the sleep environment. In *Proceedings of the 2012 ACM conference on ubiquitous computing*. ACM, 226–234.

[25] Predrag Klasnja, Sunny Consolvo, and Wanda Pratt. 2011. How to evaluate technologies for health behavior change in HCI research. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 3063–3072.

[26] Janis B Kupersmidt, Donna Bryant, and Michael T Willoughby. 2000. Prevalence of aggressive behaviors among preschoolers in Head Start and community child care programs. *Behavioral Disorders* 26, 1 (2000), 42–52.

[27] Matthew L Lee and Anind K Dey. 2014. Real-time feedback for improving medication taking. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2259–2268.

[28] Youngki Lee, Chulhong Min, Chanyou Hwang, Jaeung Lee, Inseok Hwang, Younghyun Ju, Chungkuk Yoo, Miri Moon, Uichin Lee, and Junehwa Song. 2013. Sociophone: Everyday face-to-face interaction monitoring platform using multi-phone sensor fusion. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*. ACM, 375–388.

[29] Ian Li, Anind Dey, and Jodi Forlizzi. 2010. A stage-based model of personal informatics systems. In *Proceedings of the SIGCHI conference on human factors in computing systems*. ACM, 557–566.

[30] James J Lin, Lena Mamykina, Silvia Lindtner, Gregory Delajoux, and Henry B Strub. 2006. Fish'n'Steps: Encouraging physical activity with an interactive computer game. In *International conference on ubiquitous computing*. Springer, 261–278.

[31] Rebecca M Lovell and Alexander C Ford. 2012. Effect of gender on prevalence of irritable bowel syndrome in the community: systematic review and meta-analysis. *The American journal of gastroenterology* 107, 7 (2012), 991.

[32] Brian MacWhinney. 2014. *The CHILDES project: Tools for analyzing talk, Volume II: The database*. Psychology Press.

[33] Lena Mamykina, Elizabeth Mynatt, Patricia Davidson, and Daniel Greenblatt. 2008. MAHI: investigation of social scaffolding for reflective thinking in diabetes management. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, 477–486.

[34] Judi Mesman, Ilja L Bongers, and Hans M Koot. 2001. Preschool developmental pathways to preadolescent internalizing and externalizing problems. *The Journal of Child Psychology and Psychiatry and Allied Disciplines* 42, 5 (2001), 679–689.

[35] Terrie E Moffitt. 1993. The neuropsychology of conduct disorder. *Development and psychopathology* 5, 1-2 (1993), 135–151.

[36] Terrie E Moffitt, Avshalom Caspi, Nigel Dickson, Phil Silva, and Warren Stanton. 1996. Childhood-onset versus adolescent-onset antisocial conduct problems in males: Natural history from ages 3 to 18 years. *Development and psychopathology* 8, 2 (1996), 399–424.

[37] Margaret Morris and Farzin Guilak. 2009. Mobile heart health: project highlight. *IEEE Pervasive Computing* 2 (2009), 57–61.

[38] Edith E Nolan, Kenneth D Gadow, and Joyce Sprafkin. 2001. Teacher reports of DSM-IV ADHD, ODD, and CD symptoms in schoolchildren. *Journal of the American Academy of Child & Adolescent Psychiatry* 40, 2 (2001), 241–249.

[39] Sheryl L Olson, John E Bates, James M Sandy, and Elizabeth M Schilling. 2002. Early developmental precursors of impulsive and inattentive behavior: From infancy to middle childhood. *Journal of Child Psychology and Psychiatry* 43, 4 (2002), 435–447.

[40] Laura Pina, Kael Rowan, Asta Roseway, Paul Johns, Gillian R Hayes, and Mary Czerwinski. 2014. In situ cues for ADHD parenting strategies using mobile technology. In *Proceedings of the 8th International Conference on Pervasive Computing Technologies for Healthcare*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), 17–24.

[41] Sandra Cluett Redden, Steven R Forness, Sharon L Ramey, Craig T Ramey, and Carl M Brezausek. 2003. Mental health and special education outcomes of Head Start children followed into elementary school. *NHSA Dialog: A Research-to-Practice Journal for the Early Intervention Field* 6, 1 (2003), 87–110.

[42] Guy Rosen. 2017. Getting Our Community Help in Real Time. *https://newsroom.fb.com/news/2017/11/getting-our-community-help-in-real-time/* (2017).

[43] Elizabeth Shriberg, Andreas Stolcke, Daniel Jurafsky, Noah Coccaro, Marie Meteer, Rebecca Bates, Paul Taylor, Klaus Ries, Rachel Martin, and Carol Van Ess-Dykema. 1998. Can prosody aid the automatic classification of dialog acts in conversational speech? *Language and speech* 41, 3-4 (1998), 443–492.

[44] Petr Slovák and Geraldine Fitzpatrick. 2015. Teaching and developing social and emotional skills with technology. *ACM Transactions on Computer-Human Interaction (TOCHI)* 22, 4 (2015), 19.

[45] Seokwoo Song, Seungho Kim, John Kim, Wonjeong Park, and Dongsun Yim. 2016. TalkLIME: mobile system intervention to improve parent-child interaction for children with language delay. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, 304–315.

[46] Tobias Sonne, Jörg Müller, Paul Marshall, Carsten Obel, and Kaj Grønbæk. 2016. Changing family practices with assistive technology: MOBERO improves morning and bedtime routines for children with ADHD. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, 152–164.

[47] Michael Tanana, Kevin Hallgren, Zac E Imel, David C Atkins, Padhraic Smyth, and Vivek Srikumar. 2015. Recursive Neural Networks for Coding Therapist and Patient Behavior in Motivational Interviewing.. In *CLPsych@ HLT-NAACL*. 71–79.

[48] Ottokar Tilk and Tanel Alumäe. 2016. Bidirectional Recurrent Neural Network with Attention Mechanism for Punctuation Restoration. In *Interspeech 2016*.

[49] Carole Upshur, Melodie Wenz-Gross, and George Reed. 2009. A pilot study of early childhood mental health consultation for children with behavioral problems in preschool. *Early Childhood Research Quarterly* 24, 1 (2009), 29–45.