HHS Public Access
Author manuscript
Proc ACM Interact Mob Wearable Ubiquitous Technol. Author manuscript; available in PMC 2019 July 25.

Published in final edited form as:
Proc ACM Interact Mob Wearable Ubiquitous Technol. 2019 June ; 3(2): . doi:10.1145/3328935.

Automated Detection of Infant Holding Using Wearable Sensing: Implications for Developmental Science And Intervention

XUEWEN YAO,
The University of Texas at Austin

THOMAS PLÖTZ,
Georgia Institute of Technology

MCKENSEY JOHNSON, **KAYA DE BARBARO**
The University of Texas at Austin

Abstract

Physical contact is critical for children's physical and emotional growth and well-being. Previous studies of physical contact are limited to relatively short periods of direct observation and self-report methods. These methods limit researchers' understanding of the natural variation in physical contact across families, and its specific impacts on child development. In this study we develop a mobile sensing platform that can provide objective, unobtrusive, and continuous measurements of physical contact in naturalistic home interactions. Using commercially available motion detectors, our model reaches an accuracy of 0.870 (std: 0.059) for a second-by-second binary classification of holding. In addition, we detail five assessment scenarios applicable to the development of activity recognition models for social science research, where required accuracy may vary as a function of the intended use. Finally, we propose a grand vision for leveraging mobile sensors to access high-density markers of multiple determinants of early parent-child interactions, with implications for basic science and intervention.

Keywords

Mother-infant Interaction; Attachment; Infant Holding; Wearable Sensor; Accelerometer; Assessment Scenarios

---

Authors' addresses: X. Yao, Department of Electrical and Computer Engineering, The University of Texas at Austin; T. Plötz, School of Interactive Computing, Georgia Tech; M. Johnson and K. de Barbaro, Department of Psychology, The University of Texas at Austin.

ACM Reference Format:
Xuewen Yao, Thomas Plötz, McKensey Johnson, and Kaya de Barbaro. 2019. Automated Detection of Infant Holding Using Wearable Sensing: Implications for Developmental Science And Intervention. Proc. ACM Interact. Mob. Wearable Ubiquitous Technol. 3, 2, Article 64 (June 2019), 17 pages. https://doi.org/10.1145/3328935

1 INTRODUCTION

Research on the functions and importance of physical contact lags vastly behind that of other sensory modalities. For example, there are nearly 13 times more publications on vision than touch [22]. While the importance of physical contact is evident across the lifespan, it appears to play a particularly important role during infancy, with benefits spanning a number of domains.

Touch enhances infant physiological functioning, with meaningful implications for physical health. A meta-analysis of 21 randomized control design studies involving over 3042 infants [9] indicated that increasing skin-to-skin physical contact between parents and their preterm infants reduces rates of infant mortality (risk ratio (RR): 0.67) and infection (RR: 0.5), and can shorten the duration of hospital stays (typical mean difference: 0.9 day). Other forms of touch, such as infant massage, also appear to have similar functions. For example, a review of nine studies of massage therapy indicated that preterm newborns who received 5–10 days of massage therapy showed a 21–48% gain in weight and 3–6 days shorter hospital stay relative to preterm neonates who received standard care [17]. Physical contact is thought to improve health outcomes via a variety of mechanisms. For example, touch can promote the development of the parasympathetic nervous system, including enhancing infant respiratory and thermoregulatory functioning [6].

Touch also appears to play an important role in establishing caregiver-infant bonding and attachment. In a randomized control study, mother-infant pairs receiving two hours of skin-to-skin contact immediately following their infants' birth showed enhanced patterns of reciprocity, maternal sensitivity, and infant self-regulation one year later relative to those separated for 2 hours after birth in a "standard care" condition [8].

While the benefits of massage and skin-to-skin contact in the earliest weeks of life are relatively well-established, clothed physical contact, experienced when a caregiver simply holds their child, also appears to have important developmental consequences. In particular, both skin-to-skin and clothed holding behaviors are thought to play a key role in the regulation of infants' arousal and emotional state [16, 21]. Parents' participation in such cycles of arousal and regulation during infants' first year is critical for the development of infants' independent ability to regulate their emotions, termed self-regulation, as well as the attachment relationship between children and their caregivers. Both self-regulation and attachment relationships emerge within the first year of life and together are a foundation for social-emotional functioning across the lifespan, with implications for the development of mental health disorders as well as numerous related behaviors across the lifespan[4, 41]. For example, attachment style classified at 12 months of age has been shown to predict high-quality play and exploration behaviors in the first year, better problem-solving, sociability, and independence in toddlers, and more curiosity, flexible management of behavior, and ego control in the preschool years [13, 40].

Parent holding behaviors within the first year contribute directly and indirectly to the regulation of infants' distress. Physical touch is known to function as a pain analgesic for infants [18]. Indirectly, the physical act of holding also allows caregivers to be more attentive to their infants' needs. For example, while being held, infants can indicate to their caregivers that they are hungry by rooting or clawing at them, thereby avoiding a full-blown distress episode [31]. A small number of experimental studies have shown that holding can promote infant regulation and increase attachment security. For example, a supplemental carrying intervention was found to reduce infant crying in the first 3 months of life by 43% [24]. In another intervention study [2], caregivers of young infants were randomly assigned to either an experimental holding condition or a control condition. Caregivers in the experimental condition received soft carriers which promoted physical contact while those in the control condition received infant seats. Mothers in the experimental condition were more responsive to their infants' vocalizations at 3.5 months and their infants were more securely attached at 13 months, two key indicators of adaptive social-emotional development.

While suggestive, research on the impacts of clothed physical contact is relatively limited, in large part due to the lack of a consistent way to assess the duration and timing of holding behaviors as they occur in natural settings. Indeed, research on the role of touch in infancy is typically conducted in hospital settings where research participants can be directly monitored [8, 28]. Alternatively, it relies on self-report diaries filled out by participating caregivers[24]. This means that we have little understanding of the natural variations in the amount of physical contact behaviors between families, and the broader impacts of those variations on child development. More broadly, similar issues plague research on many aspects of early interaction between caregivers and their children. While many aspects of daily activity are hypothesized to have consequences for children's development, most studies can only capture approximations of these behaviors in the laboratory.

The widespread adoption and presence of mobile and wearable sensors, paired with the coming of age of powerful algorithms to automatically extract meaningful activities from raw sensor data, could provide unprecedented access to the daily contexts in which development happens. Sensor platforms could be used to simultaneously capture the activity of the child and the caregiver, in addition to a vast array of possible determinants of their actions: from their perceived environments and internal states, to ecological factors such as household chaos [11]. This is critical as it is widely believed that developmental outcomes are not determined by a single factor, such as genes or caregiving, but are rather the outcome of a complex dynamic ecosystem [41]. Captured repeatedly over months or years, high-density data on these various determinants of development could provide a radical new opportunity to understand how and why differences between individuals develop -from school success to physical and mental health.

The ultimate goal of our research is to develop a sensors-to-analytics platform to capture high-density markers of parent-child activity "in the wild" i.e. as they go about their typical day-to-day activities. Such a platform will provide unprecedented access to objective, unobtrusive, high-density measurements of developmental determinants, pushing the limits of basic developmental psychology research. In particular, such novel datasets have the potential to lead to a better understanding of children's individual trajectories of risk and resilience, with the potential for earlier diagnosis of potential issues and abnormalities as well as opportunities for timely intervention and improved care. Wearable and ubiquitous computing methods have the potential to play a key role in this endeavor, serving as enabling technology that facilitates systematic and automated assessments of natural behavior on a large scale.

In this paper, we lay the foundations for our research agenda. Specifically, we pursue one aspect of the envisioned framework, namely, the automated assessment of holding behavior - a highly critical yet understudied component of early mother-infant interactions. The contributions of this paper are as follows:

- We present the first attempt, to our knowledge, to build a comprehensive system to automatically detect multiple dimensions of caregiver-infant interactions. Our system will provide unique access to early caregiving behaviors for developmental scientists while also advancing ubiquitous computing research via a novel multi-person use case.
- We developed a model to detect mother-infant holding behaviors using body-worn accelerometers. We trained and evaluated our model with 26 mother-infant pairs, wearing our sensor for on average 45 minutes (std =11) during naturalistic interactions. Our model successfully distinguishes holding from non-holding activities with an accuracy of 0.870 (std: 0.059) at second-by-second resolution.
- We present and discuss the accuracy of our model within five specific assessment scenarios, including event-based accuracy and comparisons of absolute and relative activity summaries across participants. These additional assessment scenarios are of particular relevance to the developmental science community, and more generally to social-science approaches interested in understanding individual differences.
- In a sub-sample of participants, we assessed the accuracy of our infant-holding detection model with sensors placed on the wrist rather than chest. Our results indicate that chest-worn sensors are more accurate in all assessment scenarios. However, wrist-worn sensors perform adequately for some assessments and could be used for longer-term monitoring as they are more comfortable for participants. We discuss these usability trade-offs in the context of survey results detailing participant's comfort in our discussion.

2 MOTIVATION AND BACKGROUND

Much of the research on the developmental impacts of touch and physical contact is limited to direct observations of families during hospital stays (common when infants are born premature and must remain under supervision). This means we know relatively little about the role of touch in healthy infants (i.e. those not born premature) as well as older infants (who are no longer in the hospital). Another common approach within holding research is to compare developmental outcomes of children of families randomly assigned to a “holding intervention” vs. a “usual care” control intervention. Families in the holding intervention are provided a baby carrier in order to encourage holding behaviors, whereas families in the control condition receive an infant seat [2]. While end-of-study survey measures have been used to validate that families assigned carriers report the use of carriers more than the control families, this method provides no actual indication of the amount of time infants in either condition are carried. This limits both the interpretability of such research as well as its accuracy, as actual time spent holding is unknown.

Self reports are the gold-standard method for assessing holding in home environments. However, self-report diaries are burdensome on families, meaning they are conducted only for short periods of time and they are prone to errors and biases. Additionally, they provide highly imprecise data on the timing of physical contact. This is critical as a key hypothesis of attachment theory is that soothing contact temporally contingent to infant distress serves to regulate (i.e. reduce) infant arousal [41], providing key input for developing stress and arousal systems [10] as well as infants' expected patterns of distress and regulation.

Wearable sensors have the potential to greatly enhance the status quo for assessment of physical contact. We hypothesize that it will be possible to leverage body-worn inertial sensors to automatically and objectively capture parent-child holding behaviors with great precision. Activity recognition can be considered one of the pillars of wearable and ubiquitous computing, and a large number of systems and methods have been developed that successfully demonstrated the feasibility of activity recognition with wearable motion sensors specifically in health applications (e.g., [3, 19, 23, 29, 37, 38]).

Despite the great popularity of wearable sensors and their wide usage in many applications, within the ubiquitous computing community, research involving infants is not common. Exceptions are mainly focused on medical applications. For example, Hayes et al. [20], developed motion sensing systems to support premature babies in the transition from hospital to home. Additionally, Fan et al. [14] developed a Markov model based technique for recognizing gestures of Cramped Synchronized General Movements which were highly correlated with an eventual diagnosis of Cerebral Palsy. [35] described a child activity recognition approach using accelerometer and barometer to prevent child accidents such as unintentional injuries.

Additionally, there is a consumer market for wearable devices that can detect infant activity and report it to parents (as critically reviewed by [43]). These devices typically assess infant physical motion and vitals, including blood oxygenation, temperature and breathing rate. However, these applications only begin to scratch the surface of the potential and useful tools that could be developed. Thus, more research into reliable methods to objectively access parent-child interactions is warranted.

2.1 Research Design and Sensor selection

Our overall aim is to develop a platform to collect objective high-density markers of mother-infant interaction. In particular, this platform should have the following functionalities. First, we wish to be able to detect markers related to caregiving activities (such as physical contact, feeding, sleeping); the social-emotional quality of caregiver-child interactions (such as presence of warm vs. harsh tones, or high vs. low synchrony between caregiver and infant); and parent and child affect, including distress signals (such as fussing and crying) and parental stress. Second, this platform should be able to collect such data for caregiver-child activity for at least a full week or longer in their natural environment. While such extended recordings are still relatively rare, one week of data could likely provide sufficient variability in natural activity without being too burdensome for our families.

The current paper is focused on the development of a model to automatically detect holding behaviors via mother and infant motion signals. However, the additional requirements of our system affect the design considerations and sensor selection of the current research and, as such, we provide additional context on our overall platform.

Our current sensor platform includes high-quality heart rate and motion sensors worn by mother and child, as well as a continuous audio recording device, worn by infants [30]. These devices were selected to provide access to various aspects of parent and child activity. In particular, body-worn motion sensors can ostensibly be used to detect infant behavior and and caregiving activities (detailed in later paragraphs). Additionally, we envision using audio recordings to automatically detect qualitative and affective aspects of the parent and child interactions [27, 42]. High-quality heart rate data provides an opportunity to examine physiological measures related to social-emotional well-being and parent-child regulation, including vagal tone and parent-child heart rate synchrony [15]. Additionally, heart rate data can improve the quality of collected sleep data, another key behavior in our study. Given these varied goals, there were a number of competing considerations in choosing the form factor for the motion sensor used in the current study. While a wrist-worn sensor would provide a more comfortable experience for long-term wear, the need for high-quality heart rate data, as stated above, necessitates a chest-worn sensor. Ultimately, we selected a single chest-worn sensor, collecting both motion and high-quality heart rate signals, as our primary data collection device. However, we also recruited a small sample of participants to wear wrist-worn motion sensors to assess the feasibility of this alternative placement.

In our ongoing data collection, families with infants aged six weeks to nine months participate in two sessions, a 90-minute “Introductory” session (henceforth referred to as the Intro session) and a 72-hour “Home Recording” session, in which they use our sensor platform to collect 72 hours of data over the course of a week. In the current paper, we focus exclusively on data collected in the Intro session, which is detailed in Section 3.1 below. Intro sessions are videotaped, thus providing us a highly reliable source of ground truth data to develop and assess our holding detection models. We describe the detailed protocol for the Intro Session in Section 3, as well as the accuracy of our models and their implications for future studies in Section 4 and 5.

2.2 Assessment Scenarios

The literature suggests that both the duration and the timing of holding behaviors can impact infant outcomes [1, 2]. In light of this body of knowledge we defined five relevant assessment scenarios (listed in Table 1). In the simplest case, knowing the accurate onset and offset of individual holding events at second-by-second level (Assessment Scenario 1) can be valuable for we can infer exact interaction patterns. Some hypotheses are more specific with reference to timing: attachment theory suggests that knowing whether the parent is contingently holding following a stimulus, such as crying, can determine its outcome. Thus, knowing the timing of holding events would be valuable (Assessment Scenario 2). More broadly, we also want to know whether or not the parent picks up the infant contingent to the onset of distress episode (Assessment Scenario 3). In addition, knowing how much a parent is holding relative to other families (Assessment Scenario 4 & 5) would be valuable for predicting outcomes such as attachment security or maternal mental health systems. As reported above, more daily carrying can effectively reduce infant crying and fussing [24]. Each of these scenarios suggest a different way to conduct automated activity recognition. Developmental scientists may use an automated assessment method to study any of these scenarios (summarized in Table 1). Furthermore, differences among the evaluation scenarios can be considered through an automated assessment system. For example, while an algorithm may be able to perform well on average for Assessment Scenario 1, it may not have individual differences for Assessment Scenario 4 & 5. So it is important to evaluate and iterate on the models to get to adequate performance for each one of these scenarios. We will explore this in Section 4.1. These different scenarios are of value to developmental scientists as well as more broadly to social scientists interested in leveraging activity recognition within studies of daily activity.

3 METHOD

3.1 Data Collection

3.1.1 Protocol.—The current study uses data collected within an Intro session lasting approximately 90 minutes. Following an introduction to the study goals and the collection of informed consent, parents and their infants participated in developmental assessments and naturalistic activities while wearing our sensor suite (detailed in Section 3.1.3), and each session was recorded on video. At the start of every video recording, research personnel completed a sequence of movements with distinct motion and visual pattern to synchronize the video and sensor data [36].

The Intro session tasks included two commonly-used developmental assessments: a 5-minute standardized “free-play” task, in which caregivers were asked to play with their infants as they normally would, while seated on the floor, and a 5-to-7-minute “reactivity assessment” [26] in which infants are presented with a number of novel or “intense” stimuli (including diluted lemon juice and noisy toys) in a controlled manner to assess individual differences in their reactions. Additionally, in a final activity named “around-the-house” task, caregivers were asked to mimic typical care-giving routines in various locations around their own homes (10–20 minutes). For example, we asked mothers to show us how they typically played with, changed, and fed their infants, in the locations they would commonly do so. This task was developed specifically to provide a variety of natural physical contact and proximity behaviors, including numerous unprompted opportunities for parents to pick up, carry and hold their infants as they moved them around in their homes. Video recorded data from the entire Intro session was included as training data in our holding recognition model.

3.1.2 Participants.—33 healthy mother-infant pairs were recruited to participate in our study. The average age of the infants was 5.33 months (std: 2.54, range: 1.23–10.8), and the sample included 13 females and 20 males. Maternal age was collected from all but two participants in our sample, and the average age of the mothers was 30.71 (n:31, std: 3.94, range: 23–39). 19 mothers identified as White, 5 as Hispanic, 2 as African-American, 1 as Asian, and 6 as Multiracial.

3.1.3 Sensors.—As detailed in Section 2, chest-worn sensors were used for our main data collection efforts, due to requirements for high-quality heart rate data in a larger ongoing study (N=26). Additionally, to assess the feasibility and trade-offs with a likely more comfortable option for long-term wear, a smaller sample of participants (N =12) wore wrist-worn sensors during the Intro session. Note that 5 of the 33 participants collected both chest and wrist-worn data during their session.

We used the Movisens EcgMove 3 (Movisens) [34] to collect chest-worn sensor data. This sensor collects electrocardiogram (ECG; up to 1024Hz), 3-axis acceleration (64 Hz), and barometric pressure (1 Hz) data. It can collect continuous data for 3 days without charging and store up to 2 weeks of data. Our holding recognition model relies on the acceleration data alone. The sensor was attached to both caregivers and infants using either a chest belt or adhesive electrodes. Figure 1a shows an infant wearing a Movisens on the chest.

We used the MetaMotionR (MMR) [32] to collect wrist-worn sensor data. The MMR is a 9-axis IMU and environmental sensor, but for purposes of comparison, only 3-axis accelerometer data was used in the current study. It collects movement data at 25 Hz. Caregivers wore the MMR sensor on their dominant wrist; infants wore the MMR attached to their ankle, as shown in Figure 1b.

3.1.4 Annotation.—Trained coders annotated various states of physical contact (e.g. bouncing, carrying) occurring throughout the entire duration of the Intro session. Proximity states were collapsed into a binary category of holding and not holding (Mean: 45 minutes, SD: 11 minutes, kappa = 0.844). Holding activity includes both holding and carrying where a mother is physically supporting her infant while not holding mainly relates to those activities where a mother is away from her infant. The detailed definition for holding and not holding can be found in Table 2. In this way, we obtained fine-grained distinctions of ground truth at second-by-second resolution. We estimated that during a session, mothers held their infants on average 41.4% of the time (std: 0.130, range: 0.163 – 0.744), which was about 18.6 minutes for an average 45-minute session.

3.2 Development of Holding Detection Algorithm

Our multi-stage algorithm has been designed for automatic holding detection and contains four main steps: *i)* preprocessing and windowing (to extract small, consecutive portions of sensor data for processing); *ii)* feature extraction; *iii)* machine learning based classification; and *iv)* post-processing of prediction results through smoothing. We applied the algorithm to 3-axis acceleration data streams recorded using the sensing system mentioned in Section 3.1.3 (separately for chest-worn and wrist-worn setting). Model validation and evaluation is based on a leave-one-participant-out protocol.

Preprocessing calculates the magnitude of raw acceleration data for both mother and infant, and then smoothes the resulting sensor data stream using a standard Savitzky-Golay filter. Sliding window based analysis employs a 7-second analysis window that is shifted in increments of 1 second. For each window we compute five features: *i)* correlation of acceleration between mother and infant; *ii)* squared correlation; *iii)* variance of mother's acceleration; *iv)* variance of baby's acceleration; and *v)* the difference between variances. The correlation features were chosen to capture synchrony between caregiver and infant when holding or carrying, because in those cases their data streams are expected to be similar, which is in contrast to when both are moving independently. We expected to see high-correlation patterns during holding and little to none correlation during not holding. Variance features were chosen to characterize the individual variations in movements. Small differences suggest similar movement distributions, whereas larger differences suggest very different movement patterns. The machine learning based classification, which is derived from supervised training, then assigns activity labels to each window, and final predictions are determined through majority vote on predictions for overlapping windows.

We explored four machine learning models: *i)* AdaBoost; *ii)* Logistic Regression; *iii)* Random Forest; and *iv)* Support Vector Machines (SVM). Random Forest provided the highest and the most stable accuracy across all sessions and therefore we limit the presentation of the results in Section 4. to that classifier.

A final post-processing step involved smoothing the stream of predictions to remove implausible outliers in the predictions that would suggest unrealistically short holding (or not-holding) episodes. In our model, non-holding episodes that were shorter than or equal to 30 seconds, and holding episodes that were shorter than 10 seconds were eliminated by the smoothing procedure through reassignment to the particular other class.

4 RESULTS

As motivated in Section 2, we introduced 5 assessment scenarios (Table 1) which are necessary to consider in social science studies. We will show the results derived from our model and chest-worn Movisens sensor one-by-one here and discuss its implications in Section 5. In addition, we present the results for all 5 scenarios using data collected by wrist-worn sensors and compare it to chest-worn sensors.

4.1 Assessment Scenario 1: High Temporal Precisions

Assessment Scenario 1 represents conventional accuracy metrics used in activity recognition. It classified holding vs. not holding activities at a second-by-second level. This level of granularity (knowing what the participant is doing every second) is important to infer exact interaction patterns.

Model performance for classifying holding and not holding second reached 87% of accuracy and 83.1% of F1 score as in Table 3. Precision and recall were both higher than 80%, meaning our model is of practical relevance for correctly distinguishing holding from not holding episodes.

4.2 Assessment Scenario 2: Event-based accuracy

Our second scenario assesses model performance for identifying events, that is, periods of continuous holding between periods of not-holding. To assess event-related accuracy, we first assessed event recall, that is, how often the model accurately detected any frames of holding during a holding event identified in the ground truth data. Additionally, we set different thresholds to specify varying proportions of overlap between detected and ground truth holding events, which could theoretically range from 0% (no detected holding for a given ground truth holding event) to 100% (complete overlap between a detected holding event and the corresponding ground truth event). We present mean recall and mean precision (summarized in Figure 2) across overlap thresholds from 0%−80%. Our results indicate that we can obtain recall and precision larger than 75%, with a threshold of 60% of overlap matching between ground truth and predicted events. In particular, at the 60% threshold, the mean recall was 0.828 and mean precision 0.774. However, this threshold also results in a number of false positives. Where on average the model detected 9.308 (Std: 3.739) events, 2.077 (Std: 1.639) were false positives.

Visual inspections suggested that these false positives were more likely to occur when ground truth events were very short. Considering only those events that were 15 seconds or greater, the number of false positives was reduced. In particular, an average number of 1.615 (Std: 1.003) false positives was generated for an average session with 6.962 (Std: 1.951) holding events.

These results indicate that while our model does not accurately capture the entire duration of every holding event, it can be used to calculate the frequency of holding events, especially those longer than 15 seconds. Measures of holding frequency have previously been used to assess variation in physical contact in the developmental literature [39].

4.3 Assessment Scenario 3: Contingency Analysis

The timing of caregiving behaviors is of special interest to developmental psychologists. In particular, caregiving behaviors which are contingent upon children’s activities have been shown to promote learning and development across a number of domains, from language learning to attachment and social-emotional outcomes [5, 7, 25]. Parental holding contingent upon infant distress is one input considered to be important for development of attachment behaviors [24]. Thus in our third scenario we assessed the accuracy of our model to detect the presence or absence of any holding activity within specific windows of time. In future research efforts these windows could be determined via the onset of periods of infant distress.

We considered windows of 5 seconds (one second stride) and 2 minutes (one minute stride) in length. If the model predicted any holding events within the window, the whole window was labeled as “holding”, otherwise it was labeled as “not holding”. We chose two different window lengths in consideration of two cases common in home scenarios. First, mother can be feet or rooms away from the baby or she may be busy with another child, which makes window length of 2 minutes an appropriate metric, whereas window length of 5 seconds is for cases when a caregiver is next to the infant, typical during interactions such as home play.

Our results are reported in the form of confusion matrices in Table 4, Considering 5-second windows (13,312 windows total) summed across all 26 sessions, the precision for holding activities reached 0.850 while the recall was 0.873 (Accuracy: 0.881, F1-score: 0.861). Considering 2-minute windows (1,097 windows total) summed across all 26 sessions, the precision for holding activities reached 0.843 while the recall was 0.943 (Accuracy: 0.847, F1-score: 0.891).

Results considering each participant separately (using a window size of 2 minutes) are reported in Table 5. The high values in all 4 measures across all sessions mean that our model can identify contingencies with high confidence.

4.4 Assessment Scenario 4 & 5: Absolute and relative activity summaries

Our final assessment scenarios characterize relative differences in summed activity between individuals (or in our case, families). A key goal within developmental science is to understand how differences in the quantity or quality of behaviors between individuals or families predict children's future developmental outcomes. This requires confidence in the model's predictions for each *pair of participants* rather than average model performance across the entire sample. We note that it is common for model performance to vary across participants: in our assessment of second-by-second accuracies, the standard deviation of our model's recall was larger than 10% (see Section 4.1), which is not uncommon in published activity recognition models. For this reason we developed two additional scenarios to directly test the model performance to preserve absolute and relative differences in activity summaries between participants.

For the fourth assessment scenario, we compared ground truth and predicted holding duration during each session using Pearson's correlation. Due to slightly varying session lengths we calculated relative holding durations after normalizing session length to 45 minutes. The R2 of the correlation was strong (correlation =0.915, R2 = 0.815; Figure 3 Left Plot), indicating that our model is accurate in predicting the amount of holding across our participants.

For the fifth assessment scenario, we assessed whether our model could preserve the rank of the amount of holding, which in some cases may be preferable to absolute differences between individuals. Holding time was again normalized to adjust for differences in recording durations between participants and compared between ground truth and predictions. The right plot of Figure 3 shows the rank of holding time in both ground truth and predictions and all data points are close to the line of expected output. Thus, our model preserved most of the order and can be used to assess the individual difference in the amount of holding. The Spearman's rank correlation reaches 0.876 and R2 is 0.771.

4.5 Comparison between chest-worn and wrist worn sensor

In our final analyses we compare the accuracy of models developed using data from wrist-worn sensors with our original model developed using data from chest-worn sensors. As stated above, this comparison has implications for participant usability as wrist-worn sensors are known to be more comfortable for long-term use. To develop the model from the wrist-worn data we used the same methods described in Section 3.2 using 12 total sessions.

Overall model accuracy for the 12 wrist-worn sessions was 0.738 (std: 0.076, F1: 0.690, Precision: 0.744, Recall:0.679) for Scenario 1, with similar accuracies for the other 4 scenarios. Additionally, five participants wore wrist and chest-worn sensors systems simultaneously, allowing us to directly compare their results for each of the five assessment scenarios. These are summarized in Table 6.

The data in Table 6 indicates that wrist-worn sensors show worse classification of holding than chest-worn sensors. Interestingly, these discrepancies in classification accuracy are not consistent throughout all five scenarios. While results in scenarios 2 and 5 appear inadequate, wrist-worn sensing is appears adequate for use cases according to scenarios 3 and 4.

5 DISCUSSION

5.1 Implications

Our results indicate that high precision and recall were achieved for all assessment scenarios, meaning that our model can automatically detect holding activities accurately and consistently within and across all participants in our study.

The strong performance of our model means that it can be used to precisely and objectively quantify holding behaviors in naturalistic settings, useful for both basic science of child development as well as development of interventions. For example, data on absolute or relative amounts of holding across a day could be used to predict relevant child and caregiver outcomes ranging from stress system neurobiology and attachment security to maternal mental health symptoms. The temporal precision of the model paired with its accuracy across participants means it could be used to collect data speaking to specific hypotheses regarding the role of the timing of physically soothing responses to infant distress. Research indicates that holding can be an important mechanism to reduce infant crying and promote attachment behaviors fundamental to lifelong outcomes in social and emotional domains.

Our model could be incorporated into interventions designed to increase caregiver holding behaviors. Providing families with objective feedback about their holding behaviors could function to increase holding frequencies and durations, with benefits for both caregivers and infants. The LENA system [30] is a similar intervention model that has had much success in the developmental science community. LENA is a wearable audio recorder which automatically detects patterns of contingent speech between parents and their children. This system has been used by thousands of parents to provide objective feedback used to increase speaking to children and ultimately vocabulary and school success.

Finally, our assessment scenarios could be usefully applied to many different activity recognition problems. In particular, we highlight the need to directly measure the consistency of model performance across different scenarios. In particular, models with acceptably high average accuracy may obscure or mislead researchers interested in using activity recognition to identify individual differences across participants [11]. Assessment and reporting of accuracy across participants or sessions can lead to more consistent models or at the very least more informed use of these models.

5.2 Assessment of Sensor Usability

One potential concern with our sensing platform is that it may be physically uncomfortable for our participants. Attaching a sensor to the chest of an infant may limit their natural movements or restrict interactions with the caregiver. Additionally it may simply be uncomfortable to wear for extended periods of time. To assess these potential concerns, we conducted an exit survey among a subset of our participants (N = 18) to evaluate their experiences with our current sensing platform. The survey was conducted following the home recording session, in which participants were instructed to wear our complete sensor platform for 72 hours over the course of a single week.

On a scale of 1 to 5, we asked caregivers to indicate how comfortable they were with the sensor attached to the chest, where 1 represents "very uncomfortable" and 5 "very comfortable". We achieved an average score of 3.67, which is reassuring for the developed approach. In particular, 10 caregivers said that the chest-worn sensor was *comfortable/very comfortable* while only one participant said the sensor was uncomfortable. In addition, using the same scale, 17 (out of 18) mothers provided ratings of the perceived comfort of their infants. Overall, mothers perceived the infant to be comfortable with the chest-worn sensors, with an average rating of 4.12. Only one mother reported that she perceived her infant was uncomfortable with the sensors while 13 mothers reported that their infants were either comfortable or very comfortable.

Overall, our survey results indicate that a chest-worn sensor was sufficiently comfortable for our participants and thus does not represent a reason for concern at this time.

5.3 Limitations and Future Work

Relative to wrist-worn sensors, chest-worn sensors provided higher accuracy across all assessment scenarios. While our participants of our exit survey indicated that chest-worn sensors were sufficiently comfortable, (Section 5.2) 14 of 18 respondents indicated that they would prefer to use wrist-worn sensors over chest-worn sensors if provided the option. Thus, there is an apparent trade-off between sensor positioning (and therefore accuracy of holding detection) and participant comfort. As a consequence, in future work we will continue to optimize our sensing platform. We note that prototypes of miniaturized inertial measurement units have recently become available. These prototypes resemble a band-aid like form factor and are thus substantially smaller than the devices used for the study presented in this paper [33]. This prospective form factor could provide a more comfortable user experience while maintaining the higher accuracy of chest-worn sensing, effectively eliminating any remaining concerns.

While our model showed reliable performance in detecting the presence or absence of holding behavior, we recognize that different types of physical contact may have different implications for children's immediate and long term outcomes. For example, in studies of infant massage, babies who received moderate pressure had better gastric motility and vagal tone than those who received light pressure [12]. It is likely that other types of touch, such as bouncing or carrying versus a more procedural contact (e.g., fastening or adjusting clothing) also have differing implications for attachment and stress system development across the first year. We are not aware of any research differentiating these in the developmental science literature, likely because it has been impossible to accurately characterize or report upon such variations in touch in natural daily activity. While beyond the scope of this paper, more precise characterization of holding activities may be possible with future research efforts. In particular, in our current dataset we have annotated a total of eight distinct types of physical contact, including: holding, carrying, bouncing, picking up, putting down, hovering, touching and nearby. We will work to incorporate these into our framework in order to provide more fine-grained analysis of physical contact behaviors in the future.

5.4 Conclusions

In the current paper, we described our use of wearable sensors to build an overarching system to capture multiple aspects of caregiver-infant interaction. We presented a model that combines chest-worn accelerometer data from caregiver and infant to detect holding behavior. Our model allows objective assessments of the amount and duration of holding in naturalistic home settings, with implications for basic science and intervention. Our results show that our model achieves high average precision and recall at second-by-second resolution. In addition, we proposed and discussed accuracy metrics for four other potential assessment scenarios, namely, event-based accuracy, contingency analysis, absolute and relative activity summaries. A subset of our 26 total participants wore both wrist- and chest-worn sensors, allowing us to compare accuracy between these two form factors. Our results indicate that chest-worn sensors are more accurate while still reasonably comfortable, even in deployments lasting up to a week. However, wrist-worn sensors may be preferable for longer-term applications when only summary measures of holding time are desired.

REFERENCES

[1]. Mary D Salter Ainsworth. 1979 Attachment as Related to Mother-Infant Interaction. Advances in the Study of Behavior, Vol. 9 Academic Press, 1–51. 10.1016/S0065-3454(08)60032-7

[2]. Anisfeld Elizabeth, Casper Virginia, Nozyce Molly, and Cunningham Nicholas. 1990. Does infant carrying promote attachment? An experimental study of the effects of increased physical contact on the development of attachment. Child development 61, 5 (10 1990), 1617–1627. 10.2307/1130769 [PubMed: 2245751]

[3]. Avci Akin, Bosch Stephen, Mihai Marin-Perianu Raluca Marin-Perianu, and Havinga Paul. 2010. Activity Recognition Using Inertial Sensing for Healthcare, Wellbeing and Sports Applications: A Survey. In 23th International Conference on Architecture of Computing Systems 2010 1–10.

[4]. Beauchaine Theodore. 2001. Vagal tone, development, and Grays motivational theory: Toward an integrated model of autonomic nervous system functioning in psychopathology. Development and Psychopathology 13, 2 (2 2001), 183–214. 10.1017/s0954579401002012 [PubMed: 11393643]

[5]. Beebe Beatrice, Jaffe Joseph, Markese Sara, Buck Karen, Chen Henian, Cohen Patricia, Bahrick Lorraine, Andrews Howard, and Feldstein Stanley. 2010. The origins of 12-month attachment: A microanalysis of 4-month mother-infant interaction. Attachment & Human Development 12, 1–2 (2010), 3–141. 10.1080/14616730903338985 [PubMed: 20390524]

[6]. Bohnhorst Bettina, Heyne Tim, Peter Corinna S., and Poets Christian F.. 2001. Skin-to-skin (Kangaroo) care, respiratory control, and thermoregulation. The Journal of pediatrics 138, 2 (2 2001), 193–197. 10.1067/mpd.2001.110978 [PubMed: 11174616]

[7]. Bornstein Marc H. and Tamis-Lemonda Catherine S. 1997. Maternal responsiveness and infant mental abilities: Specific predictive relations. Infant Behavior and Development 20, 3 (1997), 283–296. 10.1016/S0163-6383(97)90001-1

[8]. Bystrova Ksenia, Ivanova Valentina, Edhborg Maigun, Matthiesen Ann-Sofi, Anna-Berit Ransjö-Arvidson Rifkat Mukhamedrakhimov, Uvnäs-Moberg Kerstin, and Widström. Ann-Marie 2009. Early Contact versus Separation: Effects on Mother-Infant Interaction One Year Later. Birth 36, 2 (6 2009), 97–109. 10.1111/j.1523-536X.2009.00307.x [PubMed: 19489802]

[9]. Conde-Agudelo Agustin, Belizán José M., and Diaz-Rossello Jose Lius. 2016. Kangaroo mother care to reduce morbidity and mortality in low birthweight infants. Cochrane Database of Systematic Reviews 23, 8(8 2016). 10.1002/14651858.CD002771.pub4

[10]. Cozolino Louis J.. 2014 The neuroscience of human relationships: attachment and the developing social brain. W. W. Norton & Company, New York, NY.

[11]. Barbaro Kaya de. 2019 Automated sensing of daily activity: A new lens into development. Developmental Psychobiology (3 2019), 1–21. 10.1002/dev.21831

[12]. Diego Miguel A., Field Tiffany, and Hernandez-Reif Maria. 2005. Vagal Activity, Gastric Motility, and Weight Gain in Massaged Preterm Neonates. The Journal of pediatrics 147, 1 (7 2005), 50–55. 10.1016/j.jpeds.2005.02.023 [PubMed: 16027695]

[13]. Egeland Byron and Farber Ellen A.. 1984 Infant-Mother Attachment: Factors Related to Its Development and Changes over Time. Child Development 55, 3 (6 1984), 753–771. 10.2307/1130127 [PubMed: 6734316]

[14]. Fan Mingming, Gravem Dana, Cooper Dan M., and Patterson Donald J.. 2012. Augmenting Gesture Recognition with Erlang-cox Models to Identify Neurological Disorders in Premature Babies In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp ‘12). ACM, New York, NY, USA, 411–420. 10.1145/2370216.2370278

[15]. Feldman Ruth, Romi Magori-Cohen Giora Galili, Singer Magi, and Louzoun Yoram. 2011. Mother and infant coordinate heart rhythms through episodes of interaction synchrony. Infant Behavior and Development 34, 4 (12 2011), 569–577. 10.1016/j.infbeh.2011.06.008 [PubMed: 21767879]

[16]. Field Tiffany. 1995 Touch in Early Development. Lawrence Erlbaum Associates, Mahwah, NJ.

[17]. Field Tiffany, Diego Miguel, and Hernandez-Reif Maria. 2010 Preterm Infant Massage Therapy Research: A Review. Infant behavior and development 33, 2 (4 2010), 115–124. 10.1016/j.infbeh.2009.12.004 [PubMed: 20137814]

[18]. Gray Larry, Watt Lisa B., and Blass Elliott. 2000. Skin-to-Skin Contact Is Analgesic in Healthy Newborns. Pediatrics 105, 1 (1 2000), e14 10.1542/peds.105.1.e14 [PubMed: 10617751]

[19]. Hammerla Nils Y., Fisher James M., Andras Peter, Rochester Lynn, Walker Richard, and Plotz Thomas. 2015 PD Disease State Assessment in Naturalistic Environments Using Deep Learning. In Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (AAAI’15). AAAI Press, 1742–1748.

[20]. Hayes Gillian R., Cheng Karen G., Hirano Sen H., Tang Karen P., Nagel Marni S., and Baker Dianne E.. 2014 Estrellita: A Mobile Capture and Access Tool for the Support of Preterm Infants and Their Caregivers. ACM Trans. Comput.-Hum. Interact 21, 3, Article 19 (6 2014), 28pages. 10.1145/2617574

[21]. Hertenstein Matthew J.. 2002. Touch: Its Communicative Functions in Infancy. Human Development 45, 2 (2002), 70–94. 10.1159/000048154

[22]. Hertenstein Matthew J., Verkamp Julie M., Kerestes Alyssa M., and Holmes Rachel M.. 2006. The Communicative Functions of Touch in Humans, Nonhuman Primates, and Rats: A Review and Synthesis of the Empirical Research. Genetic, social, and general psychology monographs 132, 1 (2 2006), 5–94. 10.3200/MONO.132.1.5-94

[23]. Hoey Jesse, Thomas Plötz Dan Jackson, Monk Andrew, Pham Cuong, and Olivier Patrick. 2011. Rapid Specification and Automated Generation of Prompting Systems to Assist People with Dementia. Pervasive Mob. Comput 7, 3 (6 2011), 299–318. 10.1016/j.pmcj.2010.11.007

[24]. Hunziker UA and Barr Ronald. 1986. Increased carrying reduces infant crying: A randomized controlled trial. Pediatrics 77, 5 (6 1986), 641–648. [PubMed: 3517799]

[25]. Jaffe Joseph, Beebe Beatrice, Feldstein Stanley, Crown Cynthia L., Jasnow Michael D., Rochat Philippe, and Stern Daniel N.. 2011. Rhythms of dialogue in infancy: Coordinated timing in development. Monographs of the Society for Research in Child Development 66, 2 (2011), i–viii +1–149.

[26]. Jerome Kagan J Steven Reznick, and Nancy Snidman. 1987. The Physiology and Psychology of Behavioral Inhibition in Children. Child development 58, 6 (12 1987), 1459–1473. 10.2307/1130685 [PubMed: 3691195]

[27]. Kim Jonathan C. and Clements Mark A.. 2015 Formant-based feature extraction for emotion classification from speech. In 2015 38th International Conference on Telecommunications and Signal Processing (TSP). 477–481. 10.1109/TSP.2015.7296308

[28]. Klaus Marshall H., Jerauld Richard, Kreger Nancy C., Willie McAlpine Meredith Steffa, and Kennell John H.. 1972. Maternal Attachment. New England Journal of Medicine 286, 9 (2 1972), 460–463. 10.1056/NEJM197203022860904 [PubMed: 5009748]

[29]. Kranz Matthias, Andreas Möller Nils Hammerla, Diewald Stefan, Thomas Plötz Patrick Olivier, and Roalter Luis. 2013. The mobile fitness coach: Towards individualized skill assessment using personalized mobile devices. Pervasive and Mobile Computing 9, 2 (2013), 203–215. 10.1016/j.pmcj.2012.06.002SpecialSection:MobileInteractionswiththeRealWorld.

[30]. LENA. Retrieved. LENA. [n. d.].April 2, 2019 fromhttps://www.lena.org/

[31]. Emily Elysia Little Cristine H. Legare, and Carver Leslie J.. 2018. Mother-Infant Physical Contact Predicts Responsive Feeding among U.S. Breastfeeding Mothers. Nutrients 10, 9, Article 1251 (2018). 10.3390/nu10091251

[32]. MMR. Retrieved. MbientLab. [n. d.].April 2, 2019 fromhttps://mbientlab.com/metamotionr/

[33]. Miller Ross. 2015 The craziest fitness tracker ever is basically a smart band-aid. https://www.theverge.com/2015/1/4/7492199/ces-2015-craziest-fitness-tracker-bandaid-ampstrip

[34]. movisens. [n. d.]. ECG and Activity Sensor - EcgMove 3. Retrieved April 2, 2019 from https://www.movisens.com/en/products/ecg-and-activity-sensor-ecgmove-3/

[35]. Nam Yunyoung and Park Jung Wook 2013. Child Activity Recognition Based on Cooperative Fusion Model of a Triaxial Accelerometer and a Barometric Pressure Sensor. IEEE Journal of Biomedical and Health Informatics 17, 2 (3 2013), 420–426. 10.1109/JBHI.2012.2235075 [PubMed: 24235114]

[36]. Thomas Plötz Chen Chen, Hammerla Nils Y., and Abowd Gregory D.. 2012. Automatic Synchronization of Wearable Sensors and Video-Cameras for Ground Truth Annotation – A Practical Approach In Proceedings of the 2012 16th Annual International Symposium on Wearable Computers (ISWC) (ISWC ‘12). IEEE Computer Society, Washington, DC, USA, 100–103. 10.1109/ISWC.2012.15

[37]. Thomas Plötz Nils Y. Hammerla, Rozga Agata, Reavis Andrea, Call Nathan, and Abowd Gregory D.. 2012. Automatic Assessment of Problem Behavior in Individuals with Developmental Disabilities In Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp ‘12) ACM, New York, NY, USA, 391–400. 10.1145/2370216.2370276

[38]. Thomas Plötz Paula Moynihan, Pham Cuong, and Olivier Patrick. 2011 Activity Recognition and Healthier Food Preparation. Atlantis Press, Paris, 313–329. 10.2991/978-94-91216-05-3_14

[39]. Roopnarine Jaipaul L., Talukder Enayet, Jain Deepa, Joshi Priti, and Srivastav Parul. 1990. Characteristics of holding, patterns of play, and social behaviors between parents and infants in New Delhi, India. Developmental Psychology 26, 4 (1990), 667–673. 10.1037/0012-1649.26.4.667

[40]. Alan Sroufe L. 1979. The coherence of individual development: Early care, attachment, and subsequent developmental issues. American Psychologist 34 (10 1979), 834–841. 10.1037/0003-066X.34.10.834

[41]. Alan Sroufe L, Egeland Byron, Carlson Elizabeth A., and Collins W. Andrew. 2005 The development of the person: The Minnesota study of risk and adaptation from birth to adulthood (1st. ed.). The Guilford Press, New York, NY.

[42]. Tobin Erin, Kane Heidi, Saleh Daniel J., Wildman Derek, Elizabeth Crabb Breen Elizabeth Secord, and Slatcher Richard. 2015. Asthma-Related Immune Responses in Youth With Asthma: Associations With Maternal Responsiveness and Expressions of Positive and Negative Affect in

Daily Life. Psychosomatic medicine 77, 8 (10 2015), 892–902. 10.1097/psy.0000000000000236 [PubMed: 26407226]

[43]. Wang Junqing, Aisling Ann O’Kane Nikki Newhouse, Sethu-Jones Geraint Rhys, and Barbaro Kaya de. 2017. Quantified Baby: Parenting and the Use of a Baby Wearable in the Wild. Proc. ACM Hum.-Comput. Interact 1, CSCW, Article 108 (12 2017), 19pages. 10.1145/3134743

Author Manuscript

Author Manuscript

Author Manuscript

CCS Concepts:

Human-centered computing → Ubiquitous and mobile computing design and evaluation methods; Empirical studies in ubiquitous and mobile computing.

(a) (b)

Fig. 1.

Images of sensors as worn by participants in our study. Chest-worn sensors were attached with a chest belt or adhesive electrodes as in Figure 1a. Wrist-worn sensors were attached to mothers' dominant wrist or infants' ankle with tight-fitting fabric cuffs (shown in pink), as indicated in 1b.

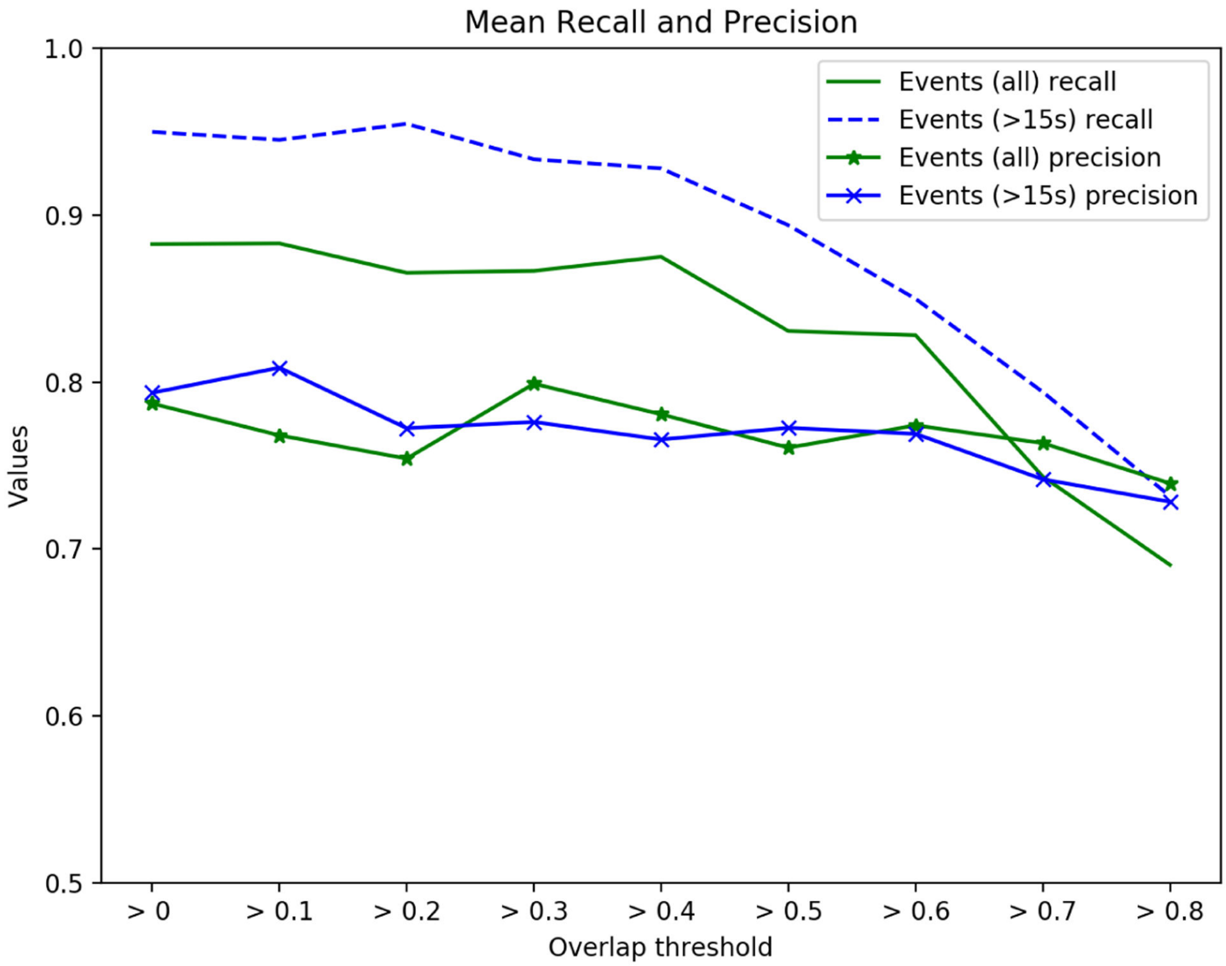
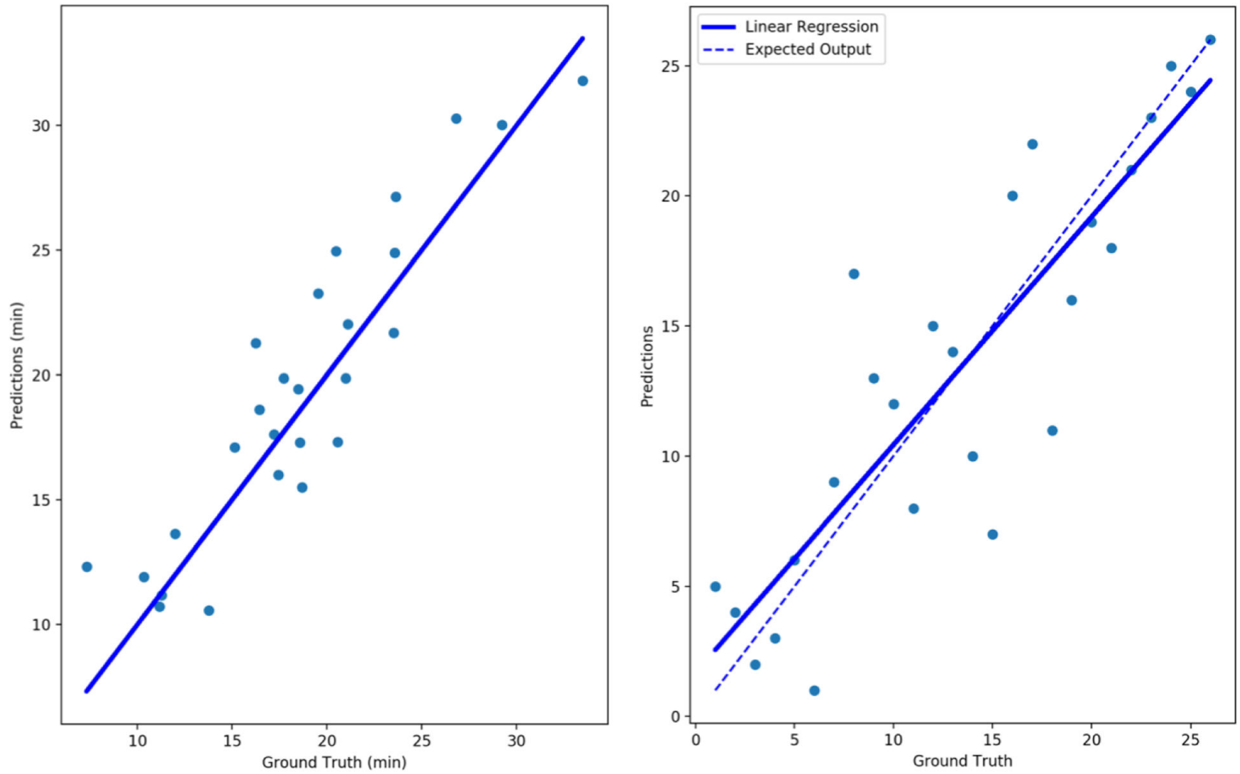


Fig. 2.

Mean recall and precision values of all 26 sessions for assessment scenario 2. When all events are considered, the model can capture at most 88% of the holding events and the number was increased to 95% as short holding events without a strong pattern are missed. The mean precision didn't change much for different overlap thresholds overall.



(a) Duration of Holding Time (R-squared = 0.815)

(b) Rank of Holding Time (R-squared = 0.767)

Fig. 3.

These plots shows the correlations between ground truth and predictions. In particular, the left plot corresponds to Scenario 4 and right plot Scenario 5. Each individual point represents one session. For the left plot, x-axis represents the holding time normalized to a 45-minute session in ground truth for each individual session and y-axis represents the predicted holding time. For the right plot, x-axis represents the rank for the amount of holding time in ground truth and y-axis represents the predicted rank. In the ideal situation, we should obtain a diagonal line where the predicted rank is the same as the ground truth rank.

Author Manuscript

Author Manuscript

Author Manuscript

Table 1.

Descriptions And Examples of Assessment Scenarios for Activity Recognition

Assessment Scenario	Description	Example
1. High Temporal Precision	Accurate onset and offset of individual holding events	What is the timing of holding behaviors over the course of the day? Is holding or holding duration contingent upon infant distress or is it unrelated to infant activity?
2. Event-based Accuracy (with Thresholds)	Successful identification of a specific activity obtaining a certain amount of overlap between ground truth and prediction	How many times is the infant held over the course of a day? Is rate of holding decreased in cases of maternal depression?
3. Contingency Analysis	Determine whether an activity happens with a certain window of time	Is infant held at all within two minutes of crying onset?
4. Absolute Activity Summaries	Total sum of the amount of an activity	How much time is a child held over the course of a day?
5. Relative Activity Summaries	The amount of an activity comparing to others and different days.	Does the model accurately identify the rank order of holding time across children? Are children who are held more often than others accurately identified by the model as such?

Table 2.

Definitions of holding activities between mothers and infants

Activities	**Definition**
Holding	Caregiver is physically supporting the infant while either standing or sitting. Includes carrying, picking up, putting down and bouncing so long as child's weight is completely supported by caregiver.
Not holding	Caregiver is not supporting the weight of the baby. Includes all periods without physical contact as well as touching or leaning behaviors in which physical contact does not completely support the child's weight.

Table 3.

Second-by-second Accuracy

	Accuracy	Fl Average	Holding precision	Holding recall
Mean	0.870	0.831	0.818	0.854
Std	0.059	0.091	0.105	0.102

Table 4.

Confusion matrices of Contingency-Analysis Assessment (Scenario 3) for Windows Summed across All 26 Sessions

Ground Truth	**Predictions**	
	Not Holding	**Holding**
Not Holding	6801	870
Holding	717	4924
(a) 5 sec confusion matrix		

Ground Truth	Predictions	
	Not Holding	Holding
Not Holding	245	127
Holding	41	684
(b) 2 min confusion matrix		

Table 5.

Performance of Contingency-Analysis Assessment (Scenario 3) for Each Participant (considering 2 min windows)

Session	Accuracy	Fl average	Hold Precision	Hold Recall
Mean	0.855	0.894	0.855	0.944
Std	0.081	0.067	0.088	0.069

Table 6.

Comparison of accuracy between wrist-worn sensors and chest-worn sensors in all five scenarios (recorded simultaneously in the same five sessions). Note that for Scenario 2, precision and recall were calculated for all events longer than 15 seconds with over 60% of overlap. In addition, scenario 3 is the mean measurements of 2 minute contingency windows.

	Wrist-worn		**Chest-worn**	
Scenario 1	Precision	Recall	Precision	Recall
	0.784	0.716	0.826	0.868
Scenario 2	Precision	Recall	Precision	Recall
	0.693	0.709	0.859	0.949
Scenario 3	Precision	Recall	Precision	Recall
	0.84	0.919	0.862	0.955
Scenario 4	Correlation	R2	Correlation	R2
	0.878	0.747	0.952	0.878
Scenario 5	Correlation	R2	Correlation	R2
	0.600	0.360	0.800	0.640