Authors' accepted manuscript to appear in Behavior Research Methods

Quantifying Infants' Everyday Restrained Experiences in the Home Using Wearable Inertial Sensors

Hanzhi Wang, Hailey N. Rousey, and John M. Franchak

Department of Psychology, University of California, Riverside

Author Note

Hanzhi Wang https://orcid.org/0009-0000-0520-5935

Hailey N. Rousey https://orcid.org/0009-0000-9828-0693

John M. Franchak https://orcid.org/0000-0002-0751-2864

Correspondence concerning this article should be addressed to John M. Franchak, Department of Psychology, University of California, Riverside, Email: franchak@ucr.edu

Abstract

Physical restraint—including being held, carried, and restrained in devices—is a common feature of infants' everyday lives. However, previous survey-based and video-based methods cannot simultaneously provide continuous, full-day accounts of infant restrained experiences. This study developed and validated a machine learning model to quantify infants' restrained time moment-to-moment across the full day in the home environment using wearable sensor data. We used a dataset that includes 146 home-visit sessions from 66 infants, with 30 younger infants aged 4-7 months, and 36 older infants aged 11-14 months. We annotated infants' restrained states in the first 1.5-hour video recordings of each session as ground truth labels. The supervised machine-learning model achieved high accuracy (89%) and substantial kappa agreement ($\kappa = .73$) compared with human-coded ground truth. The model presented a slight bias towards overestimating unrestrained periods compared to restrained periods, but the bias was mitigated when we used a longer data-aggregating window length. The model also showed convergent validity by corroborating prior studies that showed an age-related decrease in infants' overall restrained time throughout the day. In short, the current study demonstrated the utility of using wearable sensors to quantify infants' real-world restrained experiences, offering a new tool for studying how daily restraint influences early development.

Keywords: infant, restraint, everyday experiences, wearable sensors, machine learning

Quantifying Infants' Everyday Restrained Experiences in the Home Using Wearable Inertial Sensors

In day-to-day life, infants are often held or carried by caregivers (i.e., caregiver restraint), or routinely placed in devices such as highchairs, walkers, or strollers (i.e., device restraint) to ensure safety, facilitate transport, and provide comfort (Birken et al., 2015). Defined as the restriction of an infant's ability to move or change their posture freely, restraint occurs across a wide range of durations, from brief episodes where a caregiver picks up an infant to reposition them to prolonged periods, such as 30 minutes in a highchair during mealtime. Ingrained in everyday caregiving routines, restraint fundamentally shapes infants' experiences by limiting their physical location and altering what they can do, see, hear, or interact with. For example, restraint directly shapes infants' motor experiences by limiting their muscle movement (e.g., Jiang et al., 2016). Restraint also serves as a learning context that provides different opportunities for object interaction (Franchak, Kadooka, et al., 2024), face looking (Kretch, 2025) and speech input (Malachowski et al., 2023; Rousey et al., 2026). Therefore, characterizing infants' restrained experiences helps profile infants' daily learning contexts and situate their motor, perceptual, and cognitive development within everyday experiences.

Previous studies have used a variety of methods to document infants' restrained time in the home. Some used parent surveys, including questionnaires, diaries, or ecological momentary assessment (EMA) (e.g., Abbott & Bartlett, 2001; Franchak, 2019; Karasik et al., 2018). Some used video recording and human annotation (e.g., Lee et al., 2026; Springfield & Lee, 2025; Wang et al., 2025). Although surveys provide daily overviews and video recordings provide continuous time series data, neither method can simultaneously provide full-day, continuous data to account for infants' restrained experiences. In the current study, we propose that combining wearable sensors and machine learning algorithms can be a solution to record and process long-form, real-time data to characterize infants' restrained experiences.

Previous Measurements of Restraint

Despite different approaches, prior work converges to show that restraint is common in infant daily life, and that the time infants spend restrained changes over development. In terms of caregiver restraint, 2- to 6-month-old infants spent around 40% of their awake time being held or carried (Malachowski et al., 2023; Siddicky et al., 2020). This percentage reduced to around 19% for older 10- to 13-month-olds (Franchak, Kadooka, et al., 2024). In terms of device restraint, younger 4- to 6-month-olds spent around 23% of awake time in seating devices (Malachowski et al., 2023). Meanwhile, the total amount of device restraint time ranged widely from 0 to 16 hours per day for infants younger than 5 months (Callahan & Sisler, 1997), but decreased with age from 2 to 6 months (Carson et al., 2022). Older 10- to 13-month-olds spent around 27% of awake time in restraining furniture and devices (Franchak, Kadooka, et al., 2024).

Many other studies have documented some portion of infant restrained time, but they differed in restraint definitions, making it difficult to compare results across studies. Some definitions only focused on caregiver holding and carrying (e.g., Airaksinen et al., 2024; Franchak et al., 2021; Franchak, Tang, et al., 2024; Yao et al., 2019), some only focused on seating devices (e.g., Callahan & Sisler, 1997), some looked at different surfaces of placement (e.g., lap, cradle, and seat) (e.g., Graciosa et al., 2024), some focused broadly on containment equipment (e.g., Bartlett & Kneale Fanning, 2003; Carson et al., 2022; Hesketh et al., 2015; Karasik et al., 2022), and others covered a wider range of restraint like furniture placement, seating devices and holding (Franchak, Kadooka, et al., 2024; Malachowski et al., 2023; Siddicky et al., 2021; Springfield & Lee, 2025).

In the current study, we take a broader scope and consider any practices that limit infants' freedom of movement as restraint. We do not pre-define a list of restraint practices but focus on infants' real-time opportunities for movement. By including both caregiver and device restraint, we can understand how restraint as a whole constrains and facilitates infants' real-time opportunities for action, perception, and interaction.

Restraint Can Constrain and Facilitate Experiences

As an external constraint, restraint can be detrimental. In the domain of motor development, various studies have shown that higher levels of both device and caregiver restraint experienced by infants as early as 8 months are correlated with decreased proficiency and temporary delays in the acquisition of motor skills (e.g., sitting, crawling, and walking) in both full-term and low-risk preterm infants (Abbott & Bartlett, 2001; Bartlett & Kneale Fanning, 2003; Carson et al., 2022; Karasik et al., 2023; Pin et al., 2007). Kinesiology measurements suggest that reduced body movement might be the underlying mechanism. For example, being tucked in a car seat significantly reduces infants' spinal muscle activity (Siddicky et al., 2021), hip position and muscle activity (Siddicky et al., 2020), and leg movement quantity (Jiang et al., 2016) compared to being unrestrained. Since motor skill acquisition has cascading effects on infants' object learning (Soska et al., 2010), spatial cognition (Clearfield, 2004; Frick & Möhring, 2013), and language development (e.g., Oudgenoeg-Paz et al., 2015; Walle & Campos, 2014), excessive restraint time could have downstream consequences if restraint delays motor development. Restraint is also linked to limitations in real-time experiences in other developmental domains. For example, infants have fewer opportunities for object interactions when restrained compared to unrestrained (Franchak, Kadooka, et al., 2024). Infants are also exposed to a reduced quantity and consistency of adult language when restrained in devices (Malachowski et al., 2023).

However, some aspects of restraint can be positive. For infants with limited motor control, restraint can facilitate learning by stabilizing their body posture. For example, pre-sitting infants often rely on their hands to maintain posture, which can limit opportunities for visual-manual coordinated exploration. Placing infants in supported sitting positions, such as on a caregiver's lap or in a highchair, can free their hands and support higher-quality object exploration compared to unsupported sitting (Kretch, 2025; Woods & Wilcox, 2013) or compared to non-sitting positions (Soska & Adolph, 2014). Having a free hand to point at objects also elicits more verbal responses from caregivers (Wu & Gros-Louis, 2015). Moreover, restraint can increase certain learning inputs. Being held and carried by caregivers is associated with more speech input (Rousey et al., 2026). Some restraint devices (e.g., high chairs) and caregiver holding can place infants high off the ground, providing them with richer visual input of faces, distant toys, and elevated portions of their environment compared to when prone on the ground (Franchak et al., 2018; Kretch et al., 2014; Luo & Franchak, 2020).

Therefore, restraint cannot be categorized as either harmful or beneficial. Rather, these examples underscore that real-time observation of restraint is essential for understanding how restraint shapes infants' multimodal learning opportunities from moment to moment.

Limitations of Previous Research Methods

Researchers have used survey methods, including traditional retrospective reports and ecological momentary assessment (EMA), to quantify infants' restrained time, but survey methods have limitations. Retrospective parent report asks parents to estimate their infant's average daily restrained time over the past few weeks or months (Abbott & Bartlett, 2001;

Bartlett & Kneale Fanning, 2003; Callahan & Sisler, 1997; Carson et al., 2022; Hesketh et al., 2015; Karasik et al., 2018; Siddicky et al., 2020). Despite its widespread use, retrospective reports lack precision by relying on parents' memory—it is difficult to recall and sum every bout of restraint and then report an overall time estimation (Bradburn et al., 1987). There are exceptions where caregivers were asked to complete structured daily diaries reflecting on the previous day's restraint practices in short intervals (e.g., every five minutes) (Karasik et al., 2018; Majnemer & Barr, 2005). However, this method might still be sensitive to memory loss and thus not as accurate in providing data. EMA survey techniques reduce caregivers' memory burden by sending momentary phone-survey prompts to caregivers multiple times (e.g., 5, 10, or 12 times) randomly throughout a day (Franchak, 2019; Franchak, Kadooka, et al., 2024; Malachowski et al., 2023). Despite the benefits of gathering immediate observations, EMA cannot provide continuous time-series data. Thus, a method is needed that provides time-sequence information without relying on parent report.

Video recording and human annotation overcome the above drawbacks of survey-based methods, but video is subject to other limitations. Video recording by handheld cameras can be obtrusive and potentially impact caregivers' and infants' behaviors. Stationary cameras (e.g., fixed on a tripod) may lose sight of the infant due to restricted fields of view or obstructions from furniture and people in the room. Furthermore, the video annotation process can be labor-intensive and time-consuming, which limits its use in large-scale or long-duration studies. Indeed, prior work that videotaped and manually annotated infants' restraint had short durations of recording (e.g., half an hour in Lee et al. (2026) and Springfield and Lee (2025), and one hour in Wang et al. (2025)). Additionally, within the short duration of recording, researchers tend to record the period when infants are awake and active. In contrast, periods when the infants are restrained, like seated in front of the TV or seated in highchairs, would likely be skipped. Thus, video recording is subject to sampling bias and might not be representative of infants' whole-day experiences.

Promises and Challenges of Wearable Sensors

A system that uses wearable sensors (i.e., inertial movement units) to record real-time movement data and machine learning models to classify movement categories can overcome the limitations listed above. Wearable sensors capture behavior-agnostic raw movement data, including linear acceleration from accelerometers and angular velocity from gyroscopes. Researchers define target behaviors and provide ground-truth labels based on video footage. A supervised machine learning model is then trained to map the movement data onto predefined behavioral categories.

This system addresses the aforementioned limitations of survey and video methods in the following ways. First, by collecting infants' movement data through sensors, the system can objectively capture infants' motion without relying on a respondent or a video annotator's interpretation. Second, the sensors can provide moment-to-moment observation that is continuous and rich in time-sequence information. Third, sensors embedded in baby garments are mobile and unobtrusive—an experimenter does not need to follow the infant with a camera to record movements. Fourth, the system can record and process full-day and even multiple-day data due to the sensors' sufficient battery life and automated classification of machine learning models.

The past decade has witnessed an uptick in the use of wearable sensors to record and classify movement categories to enjoy these advantages. The application of wearable sensors progressed from adults (Arif & Kattan, 2015; Preece et al., 2009), to children (Nam & Park, 2013; Ren et al., 2016; Stewart et al., 2018), and then to infants (Airaksinen et al., 2020, 2024; Franchak et al., 2021; Franchak, Tang, et al., 2024; Yao et al., 2019). For infants, there are established models that classify infants' body postures—categories include supine, prone, sitting, and being held (Airaksinen et al., 2020; Franchak et al., 2021; Franchak, Tang, et al., 2024; Kretch et al., 2026). There are also models that detect only holding and carrying periods (Airaksinen et al., 2024; Yao et al., 2019).

However, there is no established model for detecting infants' general restrained periods (both device restrained and caregiver restrained) from wearable sensor data. Detecting restrained periods might be more challenging than classifying body postures. First, infants may show similar body movements whether restrained or not, making restrained and unrestrained classes more difficult to differentiate. For example, infants might kick their legs when they are lying unrestrained and also when they are strapped in a highchair. Possible distinctions might be manifested in a reduced range of motion or repetitive motion due to constraints, which require high-precision sensors and sensitive models to detect. Second, there are lots of possible restraint forms infants can present—a caregiver can hold an infant in an upright position or lay the infant in arms while breastfeeding; seating restraint can happen in countless infant seats, which vary in affordances of moving (Alghamdi et al., 2024). Furthermore, naturalistic home environment often impairs model performance (Airaksinen et al., 2024; Franchak, Tang, et al., 2024) compared to controlled laboratory settings (Airaksinen et al., 2020; Franchak et al., 2021), likely because infants' movements at home are more variable. Accordingly, the wearable sensor system should also be robust enough to generalize across heterogeneous restraint scenarios.

One way to meet the dual requirement of sensitivity and robustness is to adjust the temporal scale of predictions. Wearable sensor studies usually use a windowing process

(Airaksinen et al., 2020, 2024; Franchak et al., 2021; Franchak, Tang, et al., 2024; Yao et al., 2019), where they apply a sliding window of a certain length onto the raw movement data, and summarize motion features, such as mean, standard deviation, and correlation, within each window. The machine-learning model then uses the motion features to predict the behavioral category at the window center. Theoretically, shorter windows have higher temporal resolution and are sensitive to abrupt changes, but fewer samples in a shorter window might mean the model lacks context to make an accurate classification. In contrast, longer windows can incorporate more data and provide more robust prediction, but may lack temporal sensitivity or may introduce irrelevant data from heterogeneous movement within an event. Studies with different prediction purposes used different lengths of windows. For example, window lengths of 2.3–4 seconds were validated for classifying infants' body positions (Airaksinen et al., 2020; Franchak et al., 2021; Franchak, Tang, et al., 2024), and window lengths of 1.15–10 seconds were validated for detecting holding and carrying episodes (Airaksinen et al., 2024; Yao et al., 2019). Given the lack of evidence on optimal window lengths for detecting general restrained episodes, the current study evaluates how the window lengths affect the model performance on restraint classification.

Current Study

We propose that wearable sensors offer a more objective, unobtrusive, information-rich, and efficient approach to capturing infants' full-day restrained experiences in natural settings compared to prior methods. To test the approach, we used an existing dataset with infants aged from 4 to 14 months (Franchak, 2023).The dataset consists of 146 sessions of full-day wearable sensor recording and 1.5-hour video recording of infants' movement in the home. We coded infants' restrained status from the videos to provide ground truth labels. We then trained a supervised machine learning model based on the human-annotated labels.

We validated the model in two ways. First, we compared how accurately model predictions matched human annotation for the 1.5 hours of each video-recorded session. We also examined how varying window length influences model performance. Second, we applied the model to full-day data to estimate the proportion of time each infant spent restrained. We examined the convergent validity of the model against past research by testing whether restrained time decreased with age (Carson et al., 2022; Franchak, Kadooka, et al., 2024).

Method

Dataset

The dataset is shared on a web-based repository (Franchak, 2023; https://databrary.org/volume/1637) hosted by Databrary. The dataset consists of two cohorts of infants: 30 younger infants aged 4 to 7 months, and 36 older infants aged 11 to 14 months. Among the 66 infants, 39% (*n* = 26) participated only once, while 61% (*n* = 40) participated 2–4 times, resulting in a total of 146 sessions. Thirty-five of the infants are female and thirty-one are male. Caregivers reported the infants' race as White (*n* = 23), Hispanic or Latinx (*n* = 8), Asian (*n* = 3), African American (*n* = 2), more than one (*n* = 15), or not fitting into the above races or chose not to answer (*n* = 15). More detailed descriptions of the sample are in other publications derived from the dataset (Franchak et al., 2026; Rousey et al., 2026).

Apparatus

In the dataset, a special infant garment was customized to secure four lightweight wearable sensors (MC10 Biostamp) at four specific locations (i.e., right thigh, left thigh, right ankle, and left ankle). Each sensor collects raw data as accelerometer signals (in x, y, and z orientations) and gyroscope signals (in roll, pitch, and yaw orientations). The sampling rate of 62.5 Hz allows for high-resolution motion recording, and the battery life of more than one day and sufficient built-in storage allow for daylong recording. An action camera fixed on a mini tripod (GoPro HERO9 Black) was used to record synchronized third-person video of the first 1.5 hours into the session as ground truth.

Procedure

On the day of the home visit, an experimenter turned on the equipment by the doorstep of the participant's house and synchronized the video camera and wearable sensors by dropping the sensors in front of the camera. This action left a distinct time stamp in both the video footage and the sensors' data. The experimenter then left the house and instructed the caregiver over the phone to set up the camera in the room where the infant spent the most time in, and to move it as needed to ensure the infant remained in view. The caregiver was instructed to put the garment on the infant and complete a brief set of structured movement tasks for model-training purposes. Caregivers were asked to position the infant supine, prone, sitting, and upright; to hold the infant while walking; and to place the infant in a commonly used restraint device, such as a highchair. Each position lasted approximately 1–4 min, depending on the activity and the infant's tolerance. After this, the caregiver was asked to play with the infant on the floor for 10 minutes before they continued their daily routines. The caregiver was also provided with a paper log form to manually record the exact clock time (both start and end times) when the infant was napping or when the garment was removed for reasons such as diaper changes, baths, or errands. These times were later excluded from estimates of daily restrained time. For convenience, we refer to the remaining, included recording time as infants' "awake time."

Human Annotation of Restraint

Using the 1.5-hour videos, our team annotated the periods when the infant was restrained or unrestrained using Datavyu software (https://datavyu.org/). Restrained and unrestrained were defined as a binary category, meaning that all times were coded as one of the two categories. A period was considered missing if the infant went out of view; missing periods were excluded from model training and evaluation. Restraint was defined as when infants could not initiate or control their own movements, which included being held or carried by caregivers, or being restrained in devices that constrain them from changing body position. Unrestrained periods, in contrast, were when infants could freely change their body position or locomote. For example, an infant was considered restrained when strapped into a highchair, but not when they were placed in a big crib and free to change position. Each video file was coded by two coders: a primary coder annotated the whole 1.5-hour video, and an independent reliability coder annotated the first 30 minutes of each video. Inter-rater reliability was calculated as the proportion of video frames where the two coders coded the same label. The human annotation reached a high agreement with an overall accuracy of 98.79%, ranging from 89.28% to 99.99%, and an average kappa of 0.98, ranging from 0.80 to 1.00.

Model Classification of Restraint

The supervised machine-learning classification process followed procedures similar to prior work (Franchak et al., 2021; Franchak, Tang, et al., 2024; Nam & Park, 2013). First, we aligned the time series of human annotations and sensor data by matching the synchronization time stamp in the video and sensors. This resulted in a dataset with human-coded labels and motion data from four sensors—each capturing acceleration (x, y, z) and angular velocity (roll, pitch, yaw)—available at every sampling point. We then aggregated raw motion data within the sliding window of *n* seconds into motion features, which included statistical features—mean, median, minimum, 25th percentile, 75th percentile, maximum, skewness, kurtosis, standard deviation, and sum—for all six signals across the four sensors, resulting in 240 features. We added 196 additional features based on cross-sensor and cross-orientation metrics such as correlations and pairwise differences, resulting in 436 features in total within each window. We varied the window length *n* from 4 s to 16 s and 30 s.

We started the windowing process from the synchronization point and slid the window every *n*/2 seconds. Consequently, each time point spaced every *n*/2 seconds contains 436 motion features, aggregated from the *n*/2 seconds before and after that point. For model training and testing, we included only windows where the infant remained in a single status for over 90% of the time and labeled the window with that status. For the 4-s window length, 1.78% of windows were excluded, leaving 304542 windows. For the 16-s window length, 5.75% of windows were excluded, leaving 73002 windows. For the 30-s window length, 8.99% of windows were excluded, leaving 37563 windows.

The third step was to train and validate supervised machine learning models that classify the motion features into restrained status. We applied the random forest algorithm (Breiman, 2001) in modeling and used a leave-one-out cross-validation approach, where we held out one session at each iteration as a testing set and trained the model using the other sessions. To evaluate how well the model generalized to unseen data compared to the human annotation, we calculated performance metrics at each iteration, including accuracy, kappa, correlation between actual and predicted prevalence, sensitivity, positive predictive value (PPV), and binary F1. Accuracy shows the proportion of windows for which the model prediction agreed with the human annotation. Kappa adjusts the accuracy for chance-level agreement. Prevalence correlation examines whether the proportion of windows classified as restrained (i.e., prevalence of restraint) predicted by the model correlates with the one annotated by humans. Sensitivity reflects the proportion of true restrained observations (marked by humans) correctly detected by the model, whereas PPV reflects the proportion of predicted restrained observations that were truly restrained. Binary F1 summarizes the balance between sensitivity and PPV. For all metrics, higher values indicate better model performance. Based on prior work and benchmarks in the field (Airaksinen et al., 2020, 2024; Franchak et al., 2021; Franchak, Tang, et al., 2024; Landis & Koch, 1997; Yao et al., 2019), we interpreted accuracy above 85% and kappa values above 0.70 as high agreement. We categorized a prevalence correlation coefficient above 0.80 as a strong correlation, and values above 0.80 for sensitivity, PPV, and F1 as high performance.

Whereas accuracy, kappa, and prevalence correlation reflect overall model performance, sensitivity, PPV, and F1 are category-specific metrics. Especially for sensitivity and PPV, the differences in the values between the restrained and unrestrained categories can reveal potential prediction bias. If sensitivity is higher for the restrained than the unrestrained category, the model is better at detecting true restrained periods than unrestrained periods. If the PPV is higher for the restrained than the unrestrained category, the model-predicted restrained periods are more likely to be credible than the predicted unrestrained periods. If a model shows higher sensitivity and lower PPV for the restrained category, the model is biased toward predicting restraint by capturing the true restrained periods but also producing false alarms. We therefore compared sensitivity and PPV across categories as indicators of prediction bias.

The leave-one-out cross-validation could inform us which window length brought the optimal result. The overall accuracy and kappa value were the primary criteria, with minimizing prediction bias serving as a secondary criterion in the case that different window lengths had similar overall metrics. After the window length was picked, we trained a final model using all the human annotations to process the whole-day sensor data of all sessions. We examined the model's convergent validity by replicating a previously-reported effect that restraint time decreases with age (Carson et al., 2022; Franchak, Kadooka, et al., 2024).

Reproducibility and Data Sharing

Machine-learning process was conducted in Julia version 1.11.1 using the DecisionTree package (Sadeghi et al., 2022). This reproducible manuscript, including the model evaluation, figures, and statistical analyses, was made with Quarto version 1.5.57 (Posit, PBC, 2024) using R version 4.5.0 (R Core Team, 2025) with the R packages tidyverse (Wickham, 2023), flextable (Gohel & Skintzos, 2025), and papaja (Aust & Barth, 2024), as well as the APA Quarto template (Schneider, n.d.). Human annotation results, processing scripts, and analyses codes are publicly shared on OSF (https://osf.io/a698g/overview?view_only=221d1f2bc8d34b5d9838354224d6da8f).

Results

The first part of the validation was based on the first 1.5 hours of each of the 144 sessions for which human annotations were available. Two sessions were excluded because no video was available to provide ground truth. As shown in Figure 1, infants were annotated as restrained for an average of 35.29% (*SD* = 21.16%) and unrestrained for an average of 47.52% (*SD* = 21.92%) time among the 1.5-hour recording period (the rest of the time was coded as missing), with substantial individual variability. We report the model performance using accuracy, kappa, sensitivity, PPV, binary F1, and prevalence of restraint as compared to human annotation. We also report the model's prediction bias by comparing the sensitivity and PPV between the two categories.

The second part of the validation was based on the full-day sensor recordings. We report whether the predictions made by the model with optimal window length converges with past results (Carson et al., 2022; Franchak, Kadooka, et al., 2024) to show that older infants had less restrained time than younger infants.

Model Prediction Compared to Human Annotation

Figure 2 shows an exemplar timeline of restraint from one session, comparing human ground-truth annotation with model predictions across the three window lengths. Overall, the model showed strong agreement with human annotation. To quantify this agreement, we calculated the model performance metrics, which were then averaged across sessions.

Overall Model Performance

Table 1 shows that all window lengths reached high accuracy (0.87 for 4 s, 0.88 for 16 s, and 0.89 for 30 s), and high kappa values (0.70 for 4 s, 0.72 for 16 s, and 0.73 for 30 s). For both categories, all window lengths presented high sensitivity (0.82-0.86 for restrained, and 0.91-0.92 for unrestrained), high PPV (0.88-0.88 for restrained, and 0.83-0.86 for unrestrained), and high F1 scores (0.83-0.86 for restrained, and 0.86-0.87 for unrestrained). Moreover, the three window lengths did not significantly differ for any of the metrics (ps $\geq$ .057).

At the session level, all three window lengths also provided reliable estimates of the overall restrained time as compared to human annotation (Figure 3 A), as the prevalence predicted by the model was highly correlated with the prevalence annotated by human coders (r = 0.79 for 4 s, r = 0.81 for 16 s, and r = 0.81 for 30 s). Thus, model predictions are suitable for capturing individual differences in overall restrained time across infants.

Model Bias Toward Predicting Unrestrained Periods

Despite the overall strong accuracy, the model was biased towards predicting more unrestrained periods than restrained periods, regardless of the window length. Figure 3 B shows that even though the model predicted prevalence highly correlated with the ground truth, the model tended to underestimate the occurrences of restrained periods. To quantify this bias, we calculated the prevalence difference by subtracting human-annotated restraint prevalence from model-predicted restraint prevalence. Thus, negative values indicate underestimation by the model. The prevalence difference was -4.70% for the 4 s window, 95% CI [-6.44%, -2.97%]; -2.40% for the 16 s window, 95% CI [-4.07%, -0.65%]; and -1.70% for the 30 s window, 95% CI [-3.46%, 0.11%]. Although the prevalence difference was not significantly different between the three window lengths ($F(2, 423) = 1.82$, $p = .16$), the 30 s window presented a slightly smaller prevalence difference.

Category-specific metrics further demonstrated this bias. Figure 4 shows the differences in sensitivity and PPV between the restrained and unrestrained categories. The model was less sensitive in detecting true restrained periods than in detecting true unrestrained periods, as shown by the negative value when subtracting the sensitivity of unrestrained prediction from restrained prediction (Figure 4 A). The sensitivity difference was -0.11 for the 4 s window, 95% CI [-0.13, -0.08]; -0.06 for the 16 s window, 95% CI [-0.09, -0.04]; and -0.05 for the 30 s window, 95% CI [-0.08, -0.02]. The sensitivity difference was not significantly different between the three window lengths ($F(2, 418) = 2.80$, $p = .06$). Meanwhile, the model predicted restrained periods were more credible than the predicted unrestrained periods, as shown by the positive value when subtracting the PPV of unrestrained prediction from restrained prediction (Figure 4 B). The PPV difference was 0.05 for the 4 s window, 95% CI [0.02, 0.09]; 0.03 for the 16 s window, 95% CI

[-0.01, 0.06]; and 0.02 for the 30 s window, 95% CI [-0.01, 0.06]. The PPV difference was not significantly different between the three window lengths ($F(2, 418) = 0.26$, $p = .77$).

However, although differences among window lengths were not statistically significant, the 30 s window showed the most favorable descriptive pattern, with the smallest prevalence difference of restraint compared with human annotation and the smallest category asymmetry in sensitivity and PPV. We therefore selected the 30-s window as the optimal window length for full-day prediction.

Model Estimates Replicate Previous Studies

We applied the model with the 30 s window length to the full-day recording for all sessions. We excluded eleven sessions whose recording time was less than 3 hours, which were useful for training and validating the models but insufficient for measuring day-long experiences. We also excluded another two sessions where the family took the garment off for a substantial portion of time but did not log the time stamp of removal. In the remaining 133 sessions, infants had an average awake time of 6.18 hours ($SD = 1.88$).

Model-estimated restrained time converged with previous studies (Carson et al., 2022; Franchak, Kadooka, et al., 2024). Younger infants aged from 4 to 7 months spent $M = 55.38\%$ ($SD = 14.03\%$) of awake time restrained, and older infants aged from 11 to 14 months spent $M = 33.42\%$ ($SD = 12.49\%$) of awake time restrained. Moreover, Figure 5 shows a significant group difference in restrained time between older infants and younger infants ($b = -0.21$, $SE = 0.03$, $p < .001$), supporting the convergent validity of the current system.

Discussion

In the current study, we developed and validated a machine learning algorithm to detect infants' restrained experiences from wearable sensors in the home environment. The model presented high accuracy and substantial kappa agreement compared to human-annotated ground truth, which are comparable to other wearable sensor algorithms classifying other infant movement categories in the home (Airaksinen et al., 2020, 2024; Franchak, Tang, et al., 2024; Yao et al., 2019). The model presented a slight bias towards unrestrained—the model tended to overestimate unrestrained periods, but was not as credible when labeling a period as unrestrained. The bias might be due to the inherent imbalance in the training set where unrestrained periods were more prevalent than restrained periods, indicated by human annotated prevalence (Figure 1). Using a longer window length (i.e., 30 s, rather than 16 s or 4 s) mitigated but did not eliminate the prediction bias. Ultimately, the model with 30 s window length presented convergent validity in the overall estimate of restrained time and age-related decrease aligned with previous findings (Carson et al., 2022; Franchak, Kadooka, et al., 2024).

Why could a longer window length (i.e., 30 s) provide more balanced detection? One explanation is that a longer window length can better capture the contextual information of restraint. Consider the contextual nature of restraint: infants may exhibit similar body movement whether restrained or not for most of the time, but one subtle occurrence of distinctive movement can signal restraint. For example, an infant can be sitting on the floor unrestrained, on a caregiver's lap restrained, or strapped in a seat restrained. It is hard to distinguish if the infant is restrained while sitting still, but when the infant moves their limbs, restraint may present differences, such as a stronger damping in the motion decay because the support (i.e., caregiver's laps, or seats) absorbs the motion. These differences only present themselves in the episodes of movement, not throughout the whole period. Thus, a longer window length has a better chance of capturing the episodes. Another explanation is that a longer window length can improve generalization across a variety of restraint forms. Infants can be restrained in various ways—such as in stationary seats, jumpers, swings, or caregivers' arms, either still or while being carried around. Different restraint forms are presented by different motion features. By encompassing a broader range of movement patterns, a longer window can provide more informative and generalizable features for the model to account for variability within the restrained class. Given a restrained bout usually lasts minutes rather than seconds, a 30 s window length presents a good balance.

Our analysis of window length informs future modeling studies to consider temporal resolution as a behavior-informed decision. In infants' everyday activities, behaviors unfold across various timescales—from second-by-second movement (e.g., locomotion) to longer periods of states (e.g., on the ground vs. lifted). Our optimal window length for restraint detection turns out to be longer than the window lengths in other models for body position classification (Airaksinen et al., 2020; Franchak et al., 2021; Franchak, Tang, et al., 2024; Yao et al., 2019), corresponding to the fact that restrained bouts last longer than position bouts. We recommend selecting window lengths that align with the typical duration and transition of the targeted behavior for machine learning modeling.

We also inspected the top-ranked features in the model, which suggests that restraint classification relied on the signals from all four sensors, from both accelerometers and gyroscopes, and from various kinds of motion feature calculation (e.g., mean, median, SD, 75th percentile, correlation between sensors, magnitude). The top 10 features for each window length are reported in Appendix A. However, the feature ranking should be interpreted cautiously. It does not mean other features were not important, since a large number of total features (436) is likely to contain strong correlations between similar features. In addition, it should not be interpreted as evidence that the top-ranked features alone would be sufficient for model training.

What the feature ranking does suggest is to use multiple sensors, use both sensor modalities, and incorporate different features, corroborating with previous work (Airaksinen et al., 2025; Franchak, Tang, et al., 2024).

Advances in Methodology

The key contribution of this work is introducing a novel approach for measuring infant restraint in the home environment using wearable sensors. Wearable sensors offer several advantages over previous methods when quantifying day-long movement in naturalistic settings. Compared with survey methods, wearable sensors provide objective, continuous measurements with fine-grained temporal resolution. Compared with video recording and human coding methods, wearable sensors are less obtrusive and enable substantially longer recordings. With the machine learning model established, researchers can extend their investigation timescales to a full day or even multiple days at low cost. Although the current machine learning model achieves slightly lower accuracy (89%) than human annotation (99%), the trade-off between accuracy and human labor favors the sensor approach—especially when the target behavior fluctuates over longer timescales beyond the practical limits of human coding. Restraint is such a behavior that does not recur repeatedly every hour but follows the rhythm of daily routines and activities. Therefore, a wearable sensor system capable of generating restraint profiles across full days offers greater scalability for measuring infants' everyday experiences.

de Barbaro et al. (2026) pointed out that machine learning models are constrained by their training data. Thus, the strong performance on the current dataset does not guarantee a comparable performance on new datasets. Applying the model to new populations or new settings (e.g., different inertial sensor brands) will require coding ground truth data to check the model's accuracy. In addition, the model may need to be re-tuned when applied to new restrained forms it has not encountered before. For example, some cultures involve distinctive forms of restraint, such as the Tajik "gahvora" (Karasik et al., 2018), Tseltal slings that may either fully wrap the infant or leave the head and torso free (Casey et al., 2026), and the use of sandbags or heavy blankets in other contexts (Adolph et al., 2010; Super & Harkness, 1986). These forms of restraint may produce hip and ankle movement patterns that were not represented in the current dataset, and whether the current model can detect them remains uncertain.

Moving beyond restraint, researchers should consider using the sensor system to capture a broader repertoire of infant behaviors—such as locomotion, steps, and falls. If proven feasible, the new models will substantially reduce the human labor required to code those behaviors. Furthermore, other than expanding the range of behaviors, future works should also broaden the scenarios in which wearable sensors are applied. To date, most applications of wearable sensors have been limited to laboratory or in-home environments. Infants' outdoor experiences remain largely underexplored, whereas outdoor environments provide infants with rich and unique opportunities for learning and exploration (Fjørtoft & Sageie, 2000; Herrington & Studtmann, 1998; Schneider, 2025). How much time do infants spend restrained in strollers, car seats, or in carriers worn by caregivers? Which positions are they frequently in when restrained out of the house? Validating sensor-based models in outdoor scenarios can open new avenues for investigation.

Advances in Understanding Full-Day Real-Time Restraint

The advances in methodology can facilitate a deeper understanding of infants' full-day real-time restrained experiences. When and how to restrain infants is decided and imposed by caregivers, which in turn, shapes infants' real-time experiences—such as when and how often infants have the opportunities for movement, which portions of the environment they can perceive, and what objects and people they can interact with. Thus, full-day restraint data can be used to answer two complementary questions: what factors shape caregivers' moment-to-moment restraint decisions, and how do those decisions structure infants' real-time sensorimotor experiences?

Prior works provided valuable insights into broad parental beliefs and attitudes of restraint. For example, parental concerns over floor play, beliefs about physical activity, and practices of co-play are found to be associated with restraint frequency and infants' physical activity level (Hänisch et al., 2025; Hesketh et al., 2015; Hnatiuk et al., 2013; Karasik, 2025; Prioreschi et al., 2017). However, few studies have examined how each restraint decision is formed in real time–when caregivers choose to restrain or release an infant, what immediately precedes those decisions, and how these decisions vary across infants, caregivers, and contexts. For example, restraint decisions may vary with infant characteristics (e.g., motor skill, activity level, attachment style, or emotion expression), caregiver characteristics (e.g., beliefs, stress, habits), family daily routines, and broader socio-cultural factors, such as cultural norms (Karasik et al., 2015, 2018). By locating each individual bout of restraint, the current system enables researchers to examine restraint not only as a static summary metric, but also as a dynamic caregiving practice embedded in everyday life.

On the other hand, a full-day continuous profile of restraint can also help understand how infant real-time sensorimotor experiences are embedded in different restrained and unrestrained periods. An overall estimate of restrained time is not enough to examine the mechanisms between restraint practices and developmental outcomes. Take motor development as an example. Karasik et al. (2023) found that Tajik infants achieved motor milestones later than US infants, even though both groups share comparable amounts of overall restrained time outside of naps (Karasik et al., 2022). However, researchers further found that Tajik infants had more prolonged individual restrained bouts (Karasik et al., 2022), suggesting that how the overall restrained time is distributed across the day is important to investigate. Moreover, restraint timing (i.e., when infants are restrained) can be linked to other real-time experiences, such as speech input and social interaction. Questions can be answered, including what learning opportunities are offered or lost during restrained periods, how the opportunities are distributed across a day, and what distribution is optimal for infant learning. Additionally, restraint timing can also be linked to infants' self-initiated explorations. More questions can be answered, such as how infants' own vocalizations, attention, locomotion, or object interactions differ across restrained and unrestrained periods, and whether the opportunities lost in the restrained periods can be compensated in later unrestrained periods. In short, full-day quantification of infant restraint offers a foundation for linking caregiver practices to infants' moment-to-moment sensorimotor experiences, thereby shaping infants' developmental outcomes.

Limitations and Future Directions

One limitation of the current study was its reliance on parent reports to exclude infants' naps and times when the garment was removed. Caregivers may forget to log these periods or may record imprecise times, and we did not have an independent way to verify the accuracy of parent logs. Future work should incorporate objective approaches to improve the identification of these periods. For example, Bang et al. (2023) developed an algorithm to detect sleep periods from LENA (https://www.lena.org/) audio recordings. Although automated detection, like any machine learning predictions, is not perfect, using multiple sources of information—such as caregiver logs, audio-based sleep detection, and sensor-based movement detection—could allow for cross validation and more reliable estimates of infants' awake wearing time.

Another limitation lies in the lack of contextual information that prevents us from drawing definite developmental conclusions. Restraint takes on various forms, and even the same form of restraint can bring distinct experiences for infants. Detecting *when* restraint happens does not grant knowing *what* experiences occur. For example, an infant placed in a highchair while their caregiver provides face-to-face feeding interaction would receive rich linguistic and social input. In contrast, an infant placed in the same device while the caregiver is occupied with household chores would experience a less stimulating, interactive environment. To map restraint onto broader developmental outcomes, future research should integrate complementary data streams that capture the concurrent situational context. Incorporating multimodal tracking—such as audio recordings and caregiver-infant proximity detection (Salo et al., 2022)—will provide the vital environmental backdrop needed to fully interpret restrained experiences.

Conclusions

To conclude, the current study provided a powerful tool for quantifying infants' everyday restrained experiences in the home using wearable sensors and machine learning modeling. The model achieved high accuracy and strong agreement with human annotations, as well as convergent validity by showing an age-related decrease in the overall restrained time. We explored the window length in the modeling procedure and demonstrated that a longer window enhanced the model's balanced performance. Together, the current study advanced the method for measuring infant full-day, real-time restrained experiences, and opened avenues for research on how restraint imposed by caregivers can shape early development.

Declarations

Funding

This work was supported by National Science Foundation Grants BCS #1941449 and BCS #2521429.

Conflicts of Interest

The authors report no conflicts of interest related to this article.

Ethics Approval

All study procedures were reviewed and approved by the Institutional Review Board at the University of California, Riverside (Protocol HS-15-050).

Consent to Participate

Written informed consent was obtained from all caregivers before the study began.

Consent for Publication

Additional written consent was obtained from caregivers to allow the sharing of audio and video data.

Availability of Data and Materials

Sensor data and video recordings are publicly shared on Databrary (Franchak, 2023). Compiled human annotation results are shared on OSF (https://osf.io/a698g/overview?view_only=221d1f2bc8d34b5d9838354224d6da8f).

Code Availability

All processing scripts, including model training and validation analyses, are publicly shared on OSF as well (https://osf.io/a698g/overview?view_only=221d1f2bc8d34b5d9838354224d6da8f).

References

Abbott, A. L., & Bartlett, D. J. (2001). Infant motor development and equipment use in the home. *Child Care Health Dev.*, *27*(3), 295–306.

Adolph, K. E., Karasik, L. B., & Tamis-LeMonda, C. S. (2010). Moving between cultures: Cross-cultural research on motor development. In M. Bornstein (Ed.), *Handbook of cross-cultural developmental science* (pp. 61–88). New York: Taylor & Francis.

Airaksinen, M., Räsänen, O., Ilén, E., Häyrinen, T., Kivi, A., Marchi, V., Gallen, A., Blom, S., Varhe, A., Kaartinen, N., Haataja, L., & Vanhatalo, S. (2020). Automatic posture and movement tracking of infants with wearable movement sensors. *Sci. Rep.*, *10*(1), 169.

Airaksinen, M., Räsänen, O., & Vanhatalo, S. (2025). Trade-offs between simplifying inertial measurement unit-based movement recordings and the attainability of different levels of analyses: Systematic assessment of method variations. *JMIR MHealth UHealth*, *13*, e58078.

Airaksinen, M., Vaaras, E., Haataja, L., Räsänen, O., & Vanhatalo, S. (2024). Automatic assessment of infant carrying and holding using at-home wearable recordings. *Sci. Rep.*, *14*(1), 4852.

Alghamdi, Z. S., Orlando, J. M., & Lobo, M. A. (2024). Evaluation of the movement and play opportunities and constraints associated with containers for infants. *Pediatr. Phys. Ther.*, *36*(4), 458–466.

Arif, M., & Kattan, A. (2015). Physical activities monitoring using wearable acceleration sensors attached to the body. *PLoS One*, *10*(7), e0130851.

Aust, F., & Barth, M. (2024). *papaja: Prepare American Psychological Association journal articles with R Markdown* (Version 0.1.3) [Computer software]. https://github.com/crsh/papaja

Bang, J. Y., Kachergis, G., Weisleder, A., & Marchman, V. (2023). An automated classifier for periods of sleep and target-child-directed speech from LENA recordings. *Language Development Research*, *3*(1), 211–248.

Barbaro, K. de, Madden-Rusnak, A., & Timmons, A. (2026). Thinking critically about algorithms for automated detection of behavior: 11 guidelines for social and behavioral scientists. *Dev. Sci.*, *29*(3), e70144.

Bartlett, D. J., & Kneale Fanning, J. E. (2003). Relationships of equipment use and play positions to motor development at eight months corrected age of infants born preterm. *Pediatr. Phys. Ther.*, *15*(1), 8–15.

Birken, C. S., Lichtblau, B., Lenton-Brym, T., Tucker, P., Maguire, J. L., Parkin, P. C., Mahant, S., & TARGet Kids! Collaboration. (2015). Parents' perception of stroller use in young children: A qualitative study. *BMC Public Health*, *15*(1), 808.

Bradburn, N. M., Rips, L. J., & Shevell, S. K. (1987). Answering autobiographical questions: The impact of memory and inference on surveys. *Science*, *236*(4798), 157–161.

Breiman, L. (2001). Random forests. *Mach. Learn.*, *45*(1), 5–32.

Callahan, C. W., & Sisler, C. (1997). Use of seating devices in infants too young to sit. *Arch. Pediatr. Adolesc. Med.*, *151*(3), 233–235.

Carson, V., Zhang, Z., Predy, M., Pritchard, L., & Hesketh, K. D. (2022). Longitudinal associations between infant movement behaviours and development. *Int. J. Behav. Nutr. Phys. Act.*, *19*(1), 10.

Casey, K., Elliott, M., Mickiewicz, E., Bergelson, E., & Casillas, M. (2026). Daylong patterns of object-centric interaction in two subsistence societies. *Infant Behav. Dev.*, *83*(102197), 102197.

Clearfield, M. W. (2004). The role of crawling and walking experience in infant spatial memory. *J. Exp. Child Psychol.*, *89*(3), 214–241.

Fjørtoft, I., & Sageie, J. (2000). The natural environment as a playground for children: Landscape description and analyses of a natural playscape. *Landscape and Urban Planning*, *48*, 83–97.

Franchak, J. M. (2019). Changing opportunities for learning in everyday life: Infant body position over the first year. *Infancy*, *24*(2), 187–209.

Franchak, J. M. (2023). Inertial Sensing and Language Recording of Infants across a Day. *Databrary*. Retrieved November 20, 2025 from https://databrary.org/volume/1637.

Franchak, J. M., Kadooka, K., & Fausey, C. M. (2024). Longitudinal relations between independent walking, body position, and object experiences in home life. *Dev. Psychol.*, *60*(2), 228–242.

Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant-caregiver social looking during locomotor free play. *Dev. Sci.*, *21*(4), e12626.

Franchak, J. M., Rousey, H. N., & Wang, H. (2026). Natural statistics of infants' everyday motor experiences relate to sitting and walking development. *Dev. Psychol.*

Franchak, J. M., Scott, V., & Luo, C. (2021). A contactless method for measuring full-day, naturalistic motor behavior using wearable inertial sensors. *Front. Psychol.*, *12*, 701343.

Franchak, J. M., Tang, M., Rousey, H., & Luo, C. (2024). Long-form recording of infant body position in the home using wearable inertial sensors. *Behav. Res. Methods*, *56*(5), 4982–5001.

Frick, A., & Möhring, W. (2013). Mental object rotation and motor development in 8- and 10-month-old infants. *J. Exp. Child Psychol.*, *115*(4), 708–720.

Gohel, D., & Skintzos, P. (2025). *Flextable: Functions for tabular reporting*. https://ardata-fr.github.io/flextable-book/

Graciosa, M. D., Ferronato, P. A. M., Drezner, R., & Jesus Manoel, E. de. (2024). Emergence of locomotor behaviors: Associations with infant characteristics, developmental status, parental beliefs, and practices in typically developing brazilian infants aged 5 to 15 months. *Infant Behav. Dev.*, *76*(101965), 101965.

Hänisch, R., Carl, J., Hesketh, K. D., & Barnett, L. M. (2025). Parental influence on children's motor competence for active play: A longitudinal analysis. *J. Sports Sci.*, *00*(00), 1–8.

Herrington, S., & Studtmann, K. (1998). Landscape interventions: New directions for the design of children's outdoor play environments. *Landscape and Urban Planning*, *42*, 191–205.

Hesketh, K. D., Crawford, D. A., Abbott, G., Campbell, K. J., & Salmon, J. (2015). Prevalence and stability of active play, restricted movement and television viewing in infants. *Early Child Dev. Care*, *185*(6), 883–894.

Hnatiuk, J., Salmon, J., Campbell, K. J., Ridgers, N. D., & Hesketh, K. D. (2013). Early childhood predictors of toddlers' physical activity: Longitudinal findings from the melbourne InFANT program. *Int. J. Behav. Nutr. Phys. Act.*, *10*(1), 123.

Jiang, C., Armendi, J. T. de, & Smith, B. A. (2016). Immediate effect of positioning devices on infant leg movement characteristics. *Pediatr. Phys. Ther.*, *28*(3), 304–310.

Karasik, L. B. (2025). Cultural cascades and infant resilience: Insights from tajik *gahvora* cradling practices. *Curr. Dir. Psychol. Sci.*, *34*(2), 131–139.

Karasik, L. B., Adolph, K. E., Fernandes, S. N., Robinson, S. R., & Tamis-LeMonda, C. S. (2023). Gahvora cradling in tajikistan: Cultural practices and associations with motor development. *Child Dev.*, *94*(4), 1049–1067.

Karasik, L. B., Kuchirko, Y. A., Dodojonova, R., & Elison, J. T. (2022). Comparison of U.S. And tajik infants' time in containment devices. *Infant Child Dev.*, *31*(4).

Karasik, L. B., Tamis-LeMonda, C. S., Adolph, K. E., & Bornstein, M. H. (2015). Places and postures: A cross-cultural comparison of sitting in 5-month-olds: A cross-cultural comparison of sitting in 5-month-olds. *J. Cross. Cult. Psychol.*, *46*(8), 1023–1038.

Karasik, L. B., Tamis-LeMonda, C. S., Ossmy, O., & Adolph, K. E. (2018). The ties that bind: Cradling in tajikistan. *PLoS One*, *13*(10), e0204428.

Kretch, K. S. (2025). Effects of sitting support and positioning on infant-parent coordinated attention. *Dev. Psychol.*

Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Dev.*, *85*(4), 1503–1518.

Kretch, K., Enriques, F., Franchak, J. M., Abney, D., Bell, C., Jerry, C., Lindig, K., & Steffen, G. (2026). Body position classification using wearable sensors in infants with cerebral palsy. *PsyArXiv*.

Landis, J. R., & Koch, G. G. (1997). The measurement of observer agreement for categorical data. *Biometrics*, *33*(1), 159–174.

Lee, D. K., Springfield, A., & Patel, P. (2026). Postural practices in infancy: How skill status and environment shape early motor development. *Infant Behav. Dev.*, *84*(102214), 102214.

Luo, C., & Franchak, J. M. (2020). Head and body structure infants' visual experiences during mobile, naturalistic play. *PLoS One*, *15*(11), e0242009.

Majnemer, A., & Barr, R. G. (2005). Influence of supine sleep positioning on early motor milestone acquisition. *Dev. Med. Child Neurol.*, *47*(6), 370–6; discussion 364.

Malachowski, L., Salo, V. C., Needham, A. W., & Humphreys, K. L. (2023). Infant placement and language exposure in daily life. *Infant Child Dev.*, *32*(3).

Nam, Y., & Park, J. W. (2013). Child activity recognition based on cooperative fusion model of a triaxial accelerometer and a barometric pressure sensor. *IEEE J. Biomed. Health Inform.*, *17*(2), 420–426.

Oudgenoeg-Paz, O., Leseman, P. P. M., & Volman, M. J. M. (2015). Exploration as a mediator of the relation between the attainment of motor milestones and the development of spatial cognition and spatial language. *Dev. Psychol.*, *51*(9), 1241–1253.

Pin, T., Eldridge, B., & Galea, M. P. (2007). A review of the effects of sleep position, play position, and equipment use on motor development in infants. *Dev. Med. Child Neurol.*, *49*(11), 858–867.

Posit, PBC. (2024). *Quarto* (Version 1.5.57) [Computer software]. https://quarto.org/

Preece, S. J., Goulermas, J. Y., Kenney, L. P. J., & Howard, D. (2009). A comparison of feature extraction methods for the classification of dynamic activities from accelerometer data. *IEEE Trans. Biomed. Eng.*, *56*(3), 871–879.

Prioreschi, A., Brage, S., Hesketh, K. D., Hnatiuk, J., Westgate, K., & Micklesfield, L. K. (2017). Describing objectively measured physical activity levels, patterns, and correlates in a cross sectional sample of infants and toddlers from south africa. *Int. J. Behav. Nutr. Phys. Act.*, *14*(1), 176.

R Core Team. (2025). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. https://www.R-project.org/

Ren, X., Ding, W., Crouter, S. E., Mu, Y., & Xie, R. (2016). Activity recognition and intensity estimation in youth from accelerometer data aided by machine learning. *Appl. Intell.*, *45*(2), 512–529.

Rousey, H. N., Tang, M., Garcia, S., & Franchak, J. M. (2026). Within-day variations in infant body position predict caregiver speech input. *Dev. Sci.*, *29*(2), e70120.

Sadeghi, B., Chiarawongse, P., Squire, K., Jones, D. C., Noack, A., St-Jean, C., Huijzer, R., Schätzle, R., Butterworth, I., Peng, Y.-F., & Blaom, A. (2022). *DecisionTree.jl: A Julia implementation of the CART decision tree and random forest algorithms* (Version 0.12.4) [Computer software]. Zenodo. https://doi.org/10.5281/zenodo.7359268

Salo, V. C., Pannuto, P., Hedgecock, W., Biri, A., Russo, D. A., Piersiak, H. A., & Humphreys, K. L. (2022). Measuring naturalistic proximity as a window into caregiver-child interaction patterns. *Behav. Res. Methods*, *54*(4), 1580–1594.

Schneider, J. L. (2025). Environments “develop”: Infant motor development can inform the study of physical space. *New Ideas Psychol.*, *76*(101125), 101125.

Schneider, W. J. (n.d.). *apaquarto* [Computer software]. https://github.com/wjschne/apaquarto

Siddicky, S. F., Bumpass, D. B., Krishnan, A., Tackett, S. A., McCarthy, R. E., & Mannen, E. M. (2020). Positioning and baby devices impact infant spinal muscle activity. *J. Biomech.*, *104*(109741), 109741.

Siddicky, S. F., Wang, J., Rabenhorst, B., Buchele, L., & Mannen, E. M. (2021). Exploring infant hip position and muscle activity in common baby gear and orthopedic devices. *J. Orthop. Res.*, *39*(5), 941–949.

Soska, K. C., & Adolph, K. E. (2014). Postural position constrains multimodal object exploration in infants. *Infancy*, *19*(2), 138–161.

Soska, K. C., Adolph, K. E., & Johnson, S. P. (2010). Systems in development: Motor skill acquisition facilitates three-dimensional object completion. *Dev. Psychol.*, *46*(1), 129–138.

Springfield, A. L., & Lee, D. K. (2025). Investigating the effects of age and autonomous locomotion on infant location use during the first year. *Infant Behav. Dev.*, *81*(102155), 102155.

Stewart, T., Narayanan, A., Hedayatrad, L., Neville, J., Mackay, L., & Duncan, S. (2018). A dual-accelerometer system for classifying physical activity in children and adults. *Med. Sci. Sports Exerc.*, *50*(12), 2595–2602.

Super, C. M., & Harkness, S. (1986). The developmental niche: A conceptualization at the interface of child and culture. *Int. J. Behav. Dev.*, *9*(4), 545–569.

Walle, E. A., & Campos, J. J. (2014). Infant language development is related to the acquisition of walking. *Dev. Psychol.*, *50*(2), 336–348.

Wang, Y., Karasik, L., Hewlett, B., & MacGillivray, T. (2025). Cultural diversity in infant motor development: A comparison of early locomotor experience. *J. Cross. Cult. Psychol.*, *56*(6), 711–725.

Wickham, H. (2023). *Tidyverse: Easily install and load the tidyverse*. https://tidyverse.tidyverse.org

Woods, R. J., & Wilcox, T. (2013). Posture support improves object individuation in infants. *Dev. Psychol.*, *49*(8), 1413–1424.

Wu, Z., & Gros-Louis, J. (2015). Caregivers provide more labeling responses to infants' pointing than to infants' object-directed vocalizations. *J. Child Lang.*, *42*(3), 538–561.

Yao, X., Plötz, T., Johnson, M., & Barbaro, K. D. E. (2019). Automated detection of infant holding using wearable sensing: Implications for developmental science and intervention: Implications for developmental science and intervention. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, *3*(2), 1–17.

Table 1

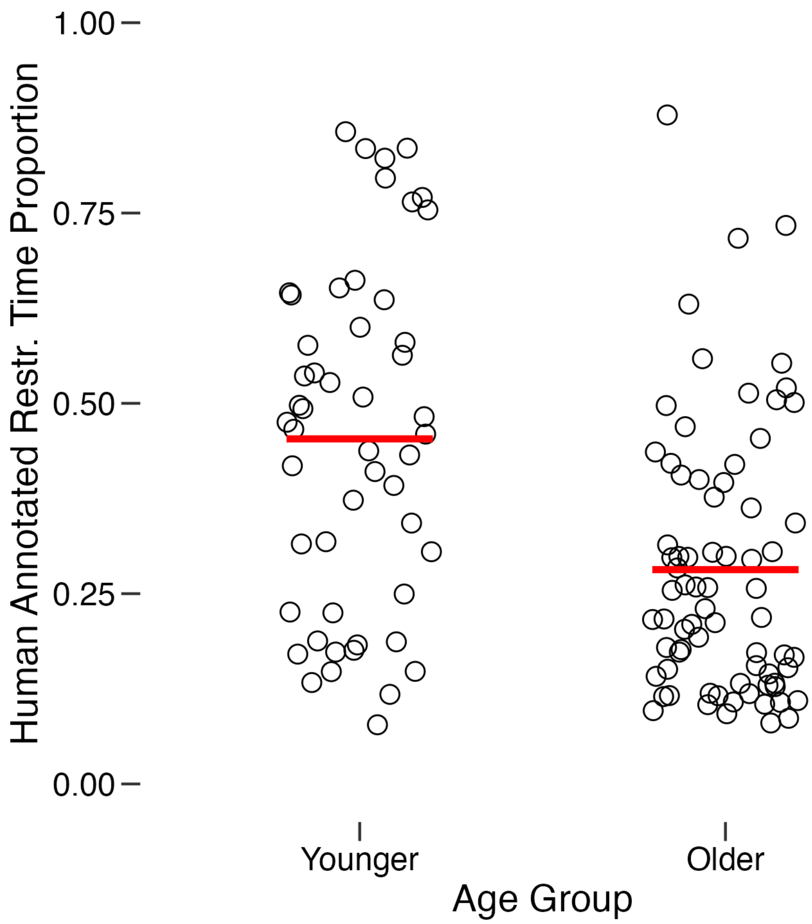
Performance Metrics According to Window Length (4 s, 16 s, and 30 s)

Metric	Category	4 s	16 s	30 s	F	df	p
Accuracy		0.87	0.88	0.89	0.58	(2, 423)	.561
Kappa		0.70	0.72	0.73	0.44	(2, 423)	.644
Sensitivity	Restrained	0.81	0.85	0.86	2.88	(2, 421)	.057
	Unrestrained	0.92	0.91	0.91	0.32	(2, 420)	.727
PPV	Restrained	0.88	0.88	0.88	0.02	(2, 423)	.982
	Unrestrained	0.83	0.85	0.86	0.68	(2, 423)	.505
F1	Restrained	0.83	0.85	0.86	1.14	(2, 421)	.322
	Unrestrained	0.86	0.87	0.87	0.12	(2, 420)	.89

Note. F, df, and p value indicate one-way ANOVA on each metric comparing across the three window lengths.

Figure 1

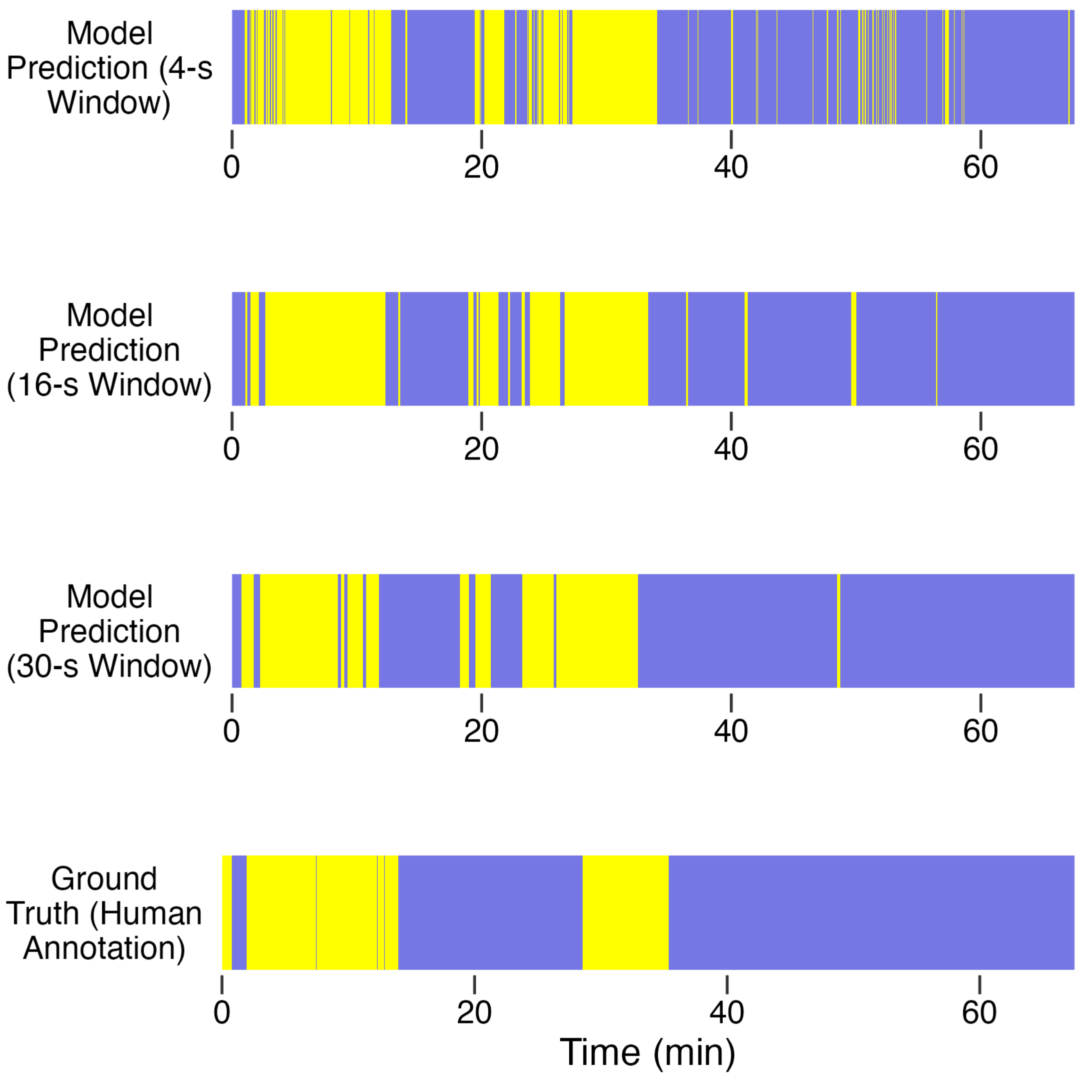
Human Annotated Restrained Proportion From the 1.5-hour Video Recording



Note. Each circle denotes one session. The horizontal line denotes the mean of restrained time proportion among the 1.5 hours for each session.

Figure 2

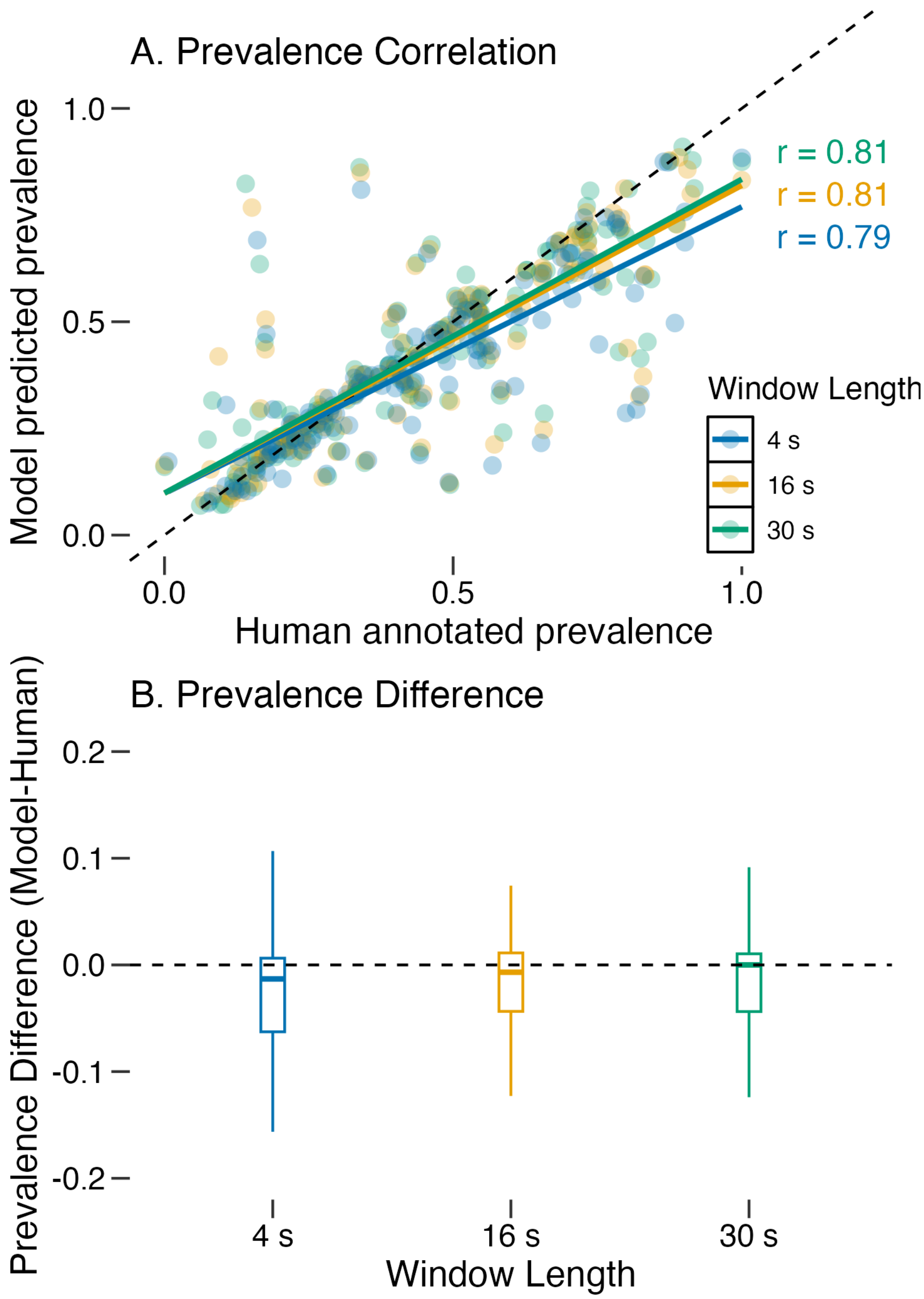
Exemplar Timeline of Model Prediction and Human Annotation



Note. This session was chosen as an example because it represents the average model accuracy. Purple bars in each timeline represent model predictions/human annotation of "restrained" and yellow bars represent "unrestrained".

Figure 3

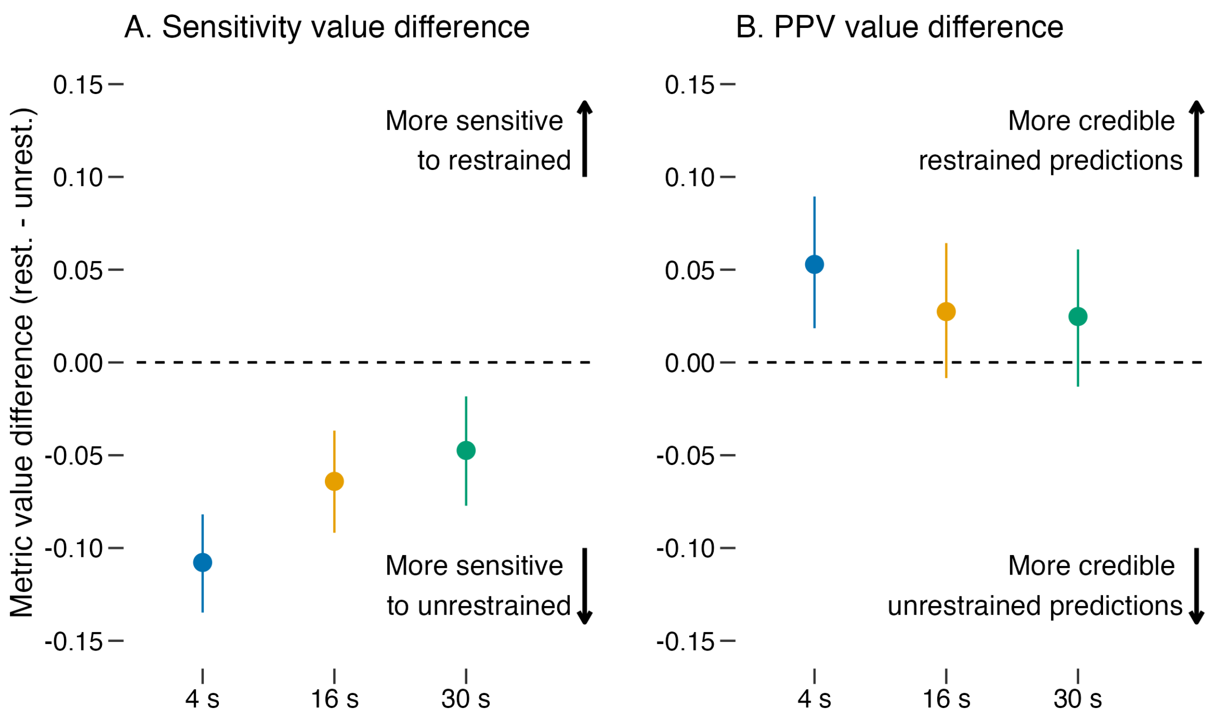
Agreement Between Model-Predicted and Human Annotated Prevalence of Restraint



Note. Prevalence is defined as the proportion of windows predicted (or annotated) as restrained among all windows. (A) shows the correlation of prevalence between model prediction and human annotation. The black dashed identity line indicates perfect agreement. Each circular symbol represents one session for a specific window length (blue = 4 s, yellow = 16 s, and green = 30 s). (B) shows the difference of prevalence between human annotation and model prediction. A negative prevalence difference indicates the model estimated less amount of time infants were restrained compared to human annotation. The horizontal line in the box denotes the median, the borderlines of the box denotes the first (Q1) quartile and the third (Q3) quartile, and the lines extended denotes the min and max data points within the range of 1.5 IQR (i.e., Q3-Q1).

Figure 4

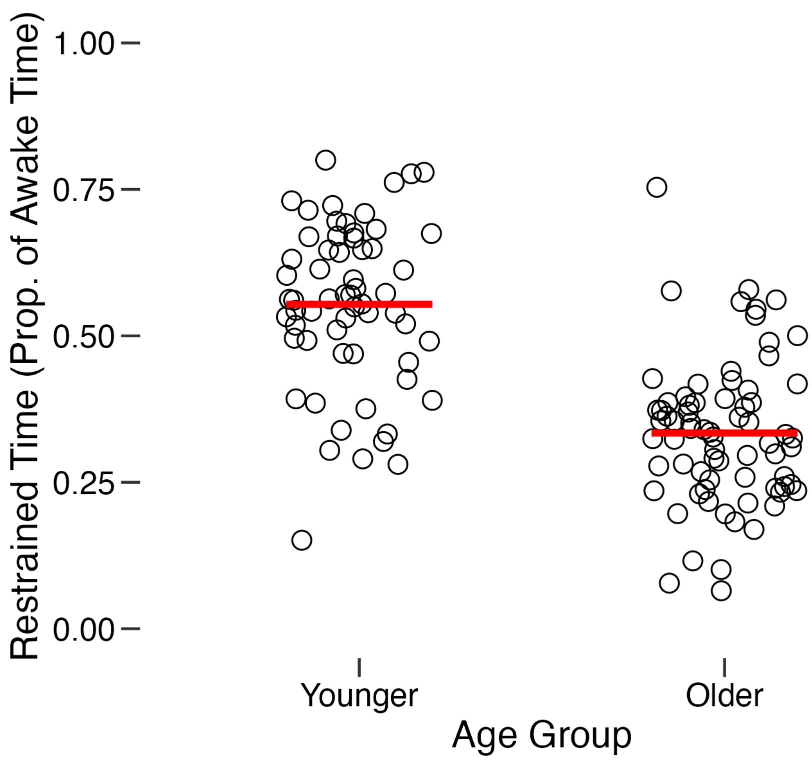
Differences in Sensitivity and PPV Between Restrained and Unrestrained Prediction According to Window Length



Note. Each vertical line indicates the 95% CI of the metric's difference between restrained prediction and unrestrained prediction. The dot on each line denotes the mean of the difference.

Figure 5

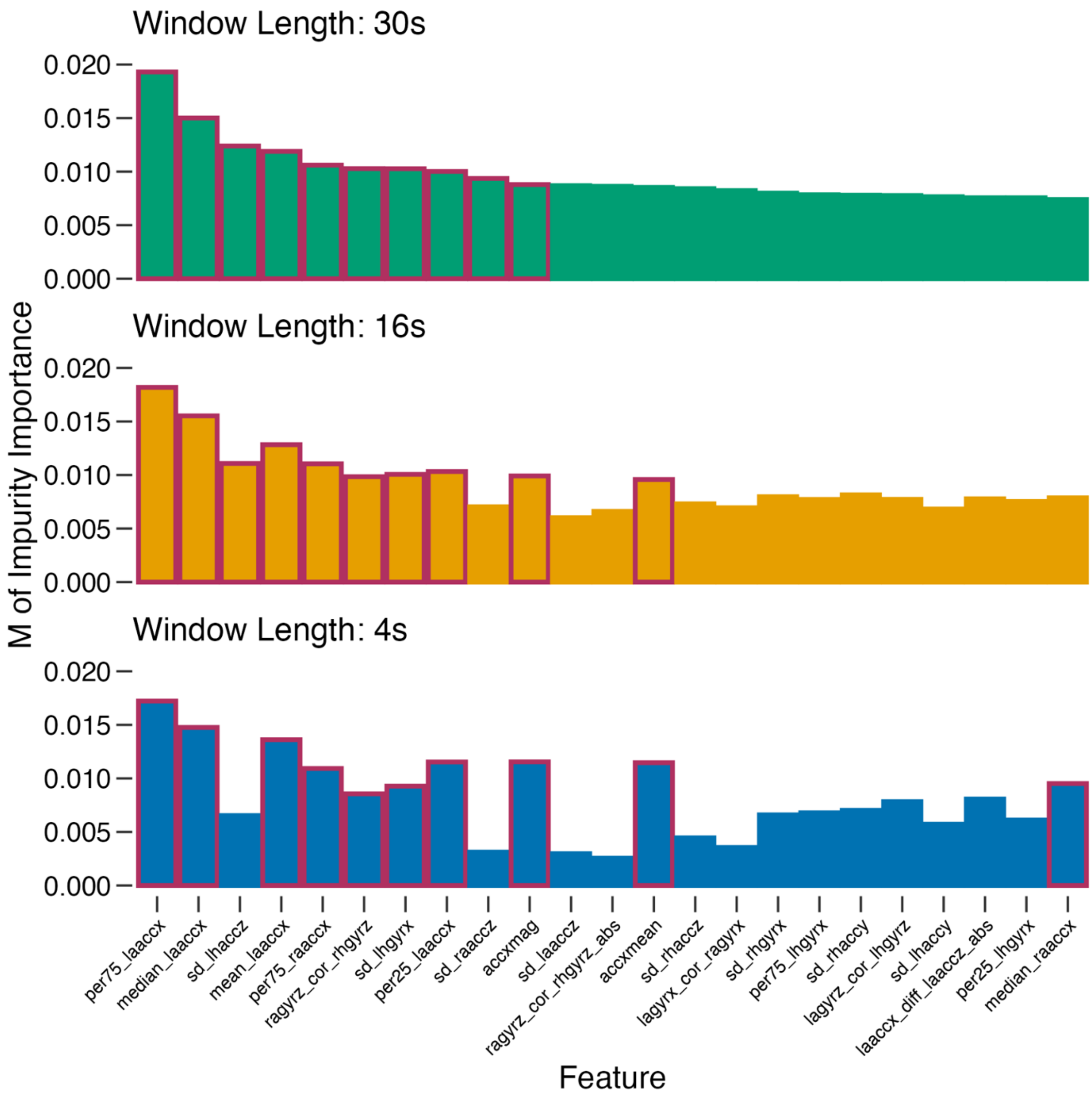
Younger Versus Older Infants' Full-Day Restrained Time



Note. Each circle denotes one session. The horizontal red line denotes the group mean of restrained time proportion (out of awake time).

Appendix A

Feature Importance by Window Length



Note. Feature plotted include the top 10 features for each of the three window lengths out of the total 436 features. Each bar represents one feature. Feature names use the following abbreviation to indicate the location, sensor, and axis: “la” stands for left ankle, “ra” stands for right ankle, “lh” stands for left hip, and “rh” stands for right hip. “acc” stands for “accelerometer” and “gyr” stands for “gyroscope”. “x” “y” and “z” are axes in accelerometer or gyroscope. The features are ordered along the x-axis based on their importance in the 30 s window model. The bars with red outline are the top 10 features for each window length.