

Full-day, in home validation of infant body position measurements from inertial sensors

John M. Franchak¹, Maximilian Tang¹, Hailey Rousey¹, & Chuan Luo¹

Author Note

Add complete departmental affiliations for each author here. Each new line herein must be indented, like this line.

Enter author note here.

Correspondence concerning this article should be addressed to John M. Franchak, UC Riverside Department of Psychology, 900 University Avenue, Riverside, CA 92521. E-mail: franchak@ucr.edu

Abstract

Abstract

Keywords: body position, motor development, everyday experiences, sitting, machine learning

Word count: X

Full-day, in home validation of infant body position measurements from inertial sensors

From moment to moment, infants’ movements facilitate and constrain how they can interact with their surroundings. Changes in *body position*—whether infants are supine on their backs, prone on their bellies, sitting, upright, or held by a caregiver—have immediate consequences for vision, object exploration, and social interaction. When sitting and upright, infants have a better view of faces and distant objects compared to their view while in a prone position (Franchak et al., 2018; Kretch et al., 2014; Luo & Franchak, 2020). Infants struggle to manipulate objects while supine and prone, but sitting affords object exploration (Soska & Adolph, 2014). Upright walking changes how infants and caregivers interact compared with crawling in a prone position; while walking infants move farther away, share toys in different ways, and hear different language from caregivers (Chen et al., 2022; Karasik et al., 2011, 2014; West & Iverson, 2021). As infants grow older and acquire new abilities, such as independent sitting and walking, they spend more time sitting and upright and less time held, supine, and prone (Adolph & Tamis-LeMonda, 2014; Franchak et al., 2018; Franchak, 2019; Thurman & Corbetta, 2017). Thus, characterizing individual differences in the day-to-day accumulation of body position experiences informs developmental theory by revealing differential opportunities for learning (Franchak, 2020).

In this paper, we present an inertial sensing method to classify infant body position from moment-to-moment across an entire day, and validate its accuracy using over 100 hours of video recorded across 35 in-home data collection sessions. Our method takes inspiration from a more mature technology: Long-form audio recordings of infants’ language experiences using wearable audio recorders. We begin by describing the impact of long-form audio recordings on research in developmental psychology, and identify the key features that should be replicated in long-form recordings of motor behavior. Next, we review the current state-of-the-art in measuring infant motor behavior—video and survey data—and their limitations in capturing real-time, full-day behavior. Finally, we discuss

the advantages of using inertial sensing to classify motor behavior. Despite promising past results in brief, supervised sessions (Airaksinen et al., 2022, 2020; Franchak et al., 2021), the current investigation takes a needed step forward by testing accuracy over long, unsupervised recordings. Although in the current investigation we focus on classification of body position, this approach can be extended to categorize other aspects of movement, such as locomotion (Airaksinen et al., 2020) and infant-caregiver contact (Yao et al., 2019). Accompanying our paper, we have created a database [CITE] containing synchronized video and IMU data, allowing researchers to annotate and train models for behaviors beyond body position.

Inspiration from Long-Form Audio Methods

The LENA® recorder is a commercial device that is worn by infants in a custom shirt pocket that has sufficient battery life and storage to record for an entire day. Closed-source LENA® algorithms analyze the audio recordings to provide automatic counts of useful metrics, such as the number of words spoken by adults in the vicinity of the participant. Other long-form audio methods rely on custom-built recorders (Wass et al., 2022), apply alternative classification algorithms to LENA® data (Micheletti et al., 2022; Räsänen et al., 2020), or manually transcribe audio recorded by LENA® devices to improve accuracy or identify behaviors beyond the built-in categories (Bergelson, Casillas, et al., 2019; Mendoza & Fausey, 2021).

Long-form audio recordings have had a transformational impact on language development research by allowing researchers to characterize opportunities for learning in daily life. Measuring the amount of speech heard by infants in the home (Weisleder & Fernald, 2013) or in a daycare setting (Perry et al., 2018) revealed individual differences in input that predict later vocabulary. Full-day language recording synchronized with other data sources allows researchers to identify how auditory input and vocal production interact with other processes. Beyond individual differences in aggregated data, long-form

recordings can be used to determine the temporal schedule of experiences. For example, infants’ daily experiences hearing music are clustered in time, with “bursty” episodes of hearing music separated by relatively long periods during which music is absent (Mendoza & Fausey, 2022). Synchronizing audio recordings with other data sources extends researchers’ ability to characterize daily experiences. Linking LENA® speech measurements to repeated, time-stamped text-message surveys about infant device placement revealed that infants heard less caregiver speech during moments that they were restrained in devices such as swings, exersaucers, and car seats (Malachowski et al., 2023). A custom-built wearable ECG and audio recorder allowed Wass et al. (2022) to discover that infant arousal increases the likelihood of infant vocalization across the day.

We identified five key features of long-form audio methods that should be replicated in analogous studies of motor behavior. First, wearable audio recorders are *mobile*. Measurement is not limited to a particular room because the recording device travels with the participant. Data are recorded to onboard device memory, so participants do not need to be in range of a receiver. Second, wearable audio recording is *unobtrusive*. Participants’ reactivity to observation, such as from a video camera, may influence behavior. For example, caregivers spoke more frequently to infants during a video-recorded portion of a home recording compared with audio-only segments captured by a LENA® device (Bergelson, Amatuni, et al., 2019). Third and fourth, recordings capture *real-time data* over a *full day*. The ability to record real-time data is vital for making inferences about processes that happen on the timescale of minutes or even seconds within an individual as opposed to comparisons of aggregated data between infants. Synchronizing real-time data to other data streams helps to reveal sources of variability within an individual (e.g., Wass et al., 2022). Full-day recordings are essential for capturing experiences across the heterogeneity of daily routines that moderate behavior (e.g., play, feeding, errands) (Kadooka et al., 2021, April). “Burstiness” of behavior means that long recordings are needed to capture clusters of events amid long periods in which they may be absent

(Barbaro & Fausey, 2022; Warlaumont et al., 2021). Fifth, *automatic classification* means that the approach can scale to analyze large numbers of participants over long recordings without the bottleneck of manual annotation/transcription. Automatic classification can only replace human annotation if it is sufficiently accurate and unbiased. An independent assessment of the LENA® algorithms found mixed results about the accuracy of different outcomes. For example, correlations between human transcribed counts of adult words and child vocalizations against LENA®’s automatic counts were strong, $r = .698$ and $r = .649$, respectively (Cristia et al., 2020). However, poor agreement was found for other metrics, such as the number of “conversational turns” between the child and communicative partners and the identification of male speakers.

Thus, for some use cases (and for some metrics), long form audio recordings provide a mobile, unobtrusive way to automatically score real-time data over a full day. In the remainder of the paper, we turn to the question of how to replicate these qualities in long form recordings of motor behavior.

Limitations of Video and Survey Methods

Video and survey methods are the current state-of-the-art in assessing motor behavior. Although, each method has advantages and disadvantages for characterizing infants’ everyday motor experiences that complements the other, neither method on its own can provide comparable data to the long form audio recordings reviewed in the previous section.

Video observation is the most common way of measuring infant motor behavior in home recordings. Most often, an experimenter with a handheld camera follows infants from room to room to ensure that their movements are visible throughout the recording session (Chen et al., 2022; Herzberg et al., 2021; Karasik et al., 2011). The primary advantage of video recording is that it captures real-time behavior. Infants’ body position, locomotion,

and reaching comprise events that occur on the timescale of seconds, so standard video recording is adequate to score gross motor behavior. However, requiring an experimenter to operate a camera is obtrusive, whereas relying on a stationary camera means that infants will be absent from view as they move from place to place. Moreover, video observation cannot easily scale to long durations or large numbers of participants. Logistically, an experimenter cannot follow behind infants to record their behavior from morning to night (and were they to do so, they would likely alter infants' and caregivers' behavior). Typical video recording sessions last 45-120 minutes (Chen et al., 2022; Herzberg et al., 2021; Karasik et al., 2011), short of capturing the variety of activities across the full daily routine. Even if full day videos were available, the lack of suitable automatic classification tools means that the human cost of annotation makes it difficult for video methods to scale to testing large numbers of participants. Our annotation of body position takes approximately 2-5 hours to complete for every hour of video (depending on how often infants switch positions), meaning that a full "waking day" of approximately 11 hours for a 12-month-old (Galland et al., 2012) could take 22-55 hours of labor to full annotate.

In contrast, survey methods such as daily diaries/inventories or ecological momentary assessment (EMA) are mobile, unobtrusive, can be applied across an entire day, and do not need laborious annotation. Diary studies provide caregivers with logs or structured interviews to estimate from memory how much time infants spend in particular activities (Karasik et al., 2022; Majnemer & Barr, 2005). Ecological momentary assessment uses text-message or app-based notifications to prompt caregivers to make repeated estimates of infants' behavior over the course of a day (Franchak, 2019; Kadooka et al., 2021, April). And although the responses are valuable in aggregate, survey methods lack the real-time temporal resolution to describe moment-to-moment changes in behavior. For diaries and interviews, limits on caregivers' memory mean that they will report what was most frequent, but cannot remember events that happen on the scale of seconds and minutes. At best, EMA surveys prompt caregivers to make hourly observations; increasing the number

of surveys per day would be too burdensome for the respondent. Thus, despite being a useful tool for estimating broad developmental changes and individual differences in infants' motor experiences, survey methods are not suited for capturing within-participant temporal dynamics.

Promise of Inertial Sensing Methods

Measuring infant movement with inertial movement units (IMUs) is a promising avenue for long form recordings of motor behavior in the home (Barbaro, 2019; Bruijns et al., 2020; Cliff et al., 2009; Lobo et al., 2019). Lightweight sensors (10-30 g) can be embedded in garments to make recordings fully *mobile*, and they are *unobtrusive* because they do not require a researcher's presence nor do they record sensitive audio/video that might influence participants' behavior. Many commercially-available and inexpensive IMUs have > 12 hour battery life with onboard storage to record *real-time*, *full-day* motion data at a sampling rate that is higher than typical video (e.g., 50-100 Hz). Moreover, past work has successfully recorded the rate of leg kicks (e.g., Deng et al., 2019) and activity intensity (Schneller et al., 2017) in infants and children across multiple days.

The open question is whether *automatic classification* is sufficiently accurate to measure movement categories that are relevant to developmental and clinical research, and whether measurement validity is acceptable over long recording periods. IMUs typically contain accelerometers that measure linear acceleration paired with gyroscopes that measure angular acceleration. Unlike motion tracking systems that might be used in a lab, IMUs do not provide data about the position of the body in space. Thus, data processing algorithms are needed to classify the raw sensor data (i.e., linear and angular acceleration timeseries) into meaningful categories (e.g., supine, prone, sitting, upright). The difficulty of the classification task depends on the categories of interest. More basic aspects of movement, such as overall activity intensity, can be identified by taking the magnitude of acceleration (irrespective of direction) and applying thresholds or cut-points to define when

a movement has occurred at a particular intensity (Armstrong et al., 2019; e.g., Hager et al., 2017).

Categorizing body position—supine on the back, prone on the belly, sitting, upright, or held off the ground by a caregiver—is too complex for cut-point definitions to be accurate. First, the magnitude of movement can vary greatly *within* a body position. An upright infant can be standing still or can be walking briskly across the room. A prone infant can be stationary in “tummy time”, or they can crawl in myriad ways (Adolph et al., 1998). Moreover, the configuration of the arms, legs, and torso within a body position can vary greatly in everyday contexts. Infants can sit on the floor in a tripod position with support from an arm, in a “V” position with legs fully extended, in a “W” position with knees bent. Sitting on a caregivers’ lap without the need to maintain balance means that the legs can dangle and the torso can lean in different directions. Sitting in a high chair or car seat reduces the magnitude of postural sway within sitting, and creates even more possibilities for how the arms and legs may move relative to the torso. Indeed, creating an all-encompassing set of rules for how to annotate sitting from video is no trivial task because of the various ways that sitting can occur in daily life. Finally, caregivers frequently pick up and transport infants, creating motion signals that need to be differentiated from independent activity (Kwon et al., 2019; Patel et al., 2019). Thus, modern approaches to human activity recognition have used machine learning to classify activity categories based on features derived from IMU data in adults (Arif & Kattan, 2015; Preece et al., 2009), children (Nam & Park, 2013; Ren et al., 2016; Stewart et al., 2018), and infants (Airaksinen et al., 2020; Franchak et al., 2021; Yao et al., 2019). Synchronized video with ground-truth human annotations creates training data for a machine learning algorithm, such as a random forest model, that can be later used to predict categories from IMU data that was not annotated. Crucially, this allows automatic classification to scale to full day recording by relying on a relatively smaller set of video annotation.

Three prior investigations have used machine learning to categorize infant body position from IMUs towards the goal of collecting full-day data. Airaksinen et al. (2020) tested 4- to 8-month-olds in a laboratory visit, and found 95% accuracy in distinguishing between body position categories that crawling infants could perform on the floor (excluding times that infants were held by caregivers). Using a wider age range of 6-18 months, Franchak et al. (2021) found 98% accuracy ($kappa = 95\%$) in a laboratory validation study in categorizing body position that included infants who could both crawl and walk and also included a category for caregiver holding. Most recently, Airaksinen et al. (2022) conducted a validation study of body position classification in either a home or clinic testing in 4- to 19-month-olds, refining their previous method to detect moments that infants were carried by caregivers. Classification accuracy did not vary between lab and home settings, and was generally high (95%, $kappa = .93$). Although all three studies yielded promising classification accuracy, accuracy was assessed in brief (15-60 minute) sessions supervised by a researcher, leaving the open question of how well body position classification will scale to testing across an entire day of natural home life.

Goals of the Current Study

Accordingly, the overarching goal of the current study is to test the feasibility and validity of long-form body position recording in the home during unsupervised, everyday behavior. Supervised recordings from past work (Airaksinen et al., 2022, 2020; Franchak et al., 2021), whether in the home or in the lab, let researchers set up the situation to encourage or restrict certain behaviors. Usually, caregivers are asked to play with the infant. However, across a real day, non-play activities (e.g., eating lunch in a high chair) create challenging situations for applying automated classification of body position. Will models trained on video-recorded observations at the beginning of the day generalize to predict behavior at a later time? Assessing the validity of temporally *distal* periods is a crucial step to establish whether automatic classification can be used to measure body

position across a day.

In the current study, we report the feasibility and validity of body position classification over the full day in the home based on 35 testing sessions from 22 infants aged 4-14 months. Participants received a custom pair of infant leggings embedded with 4 IMUs (one on each ankle and one on each hip) and a video camera to collect ground truth data about infant body position. A *proximal comparison* period began when participants received the equipment and completed a guided phone call during which caregivers were asked to elicit different body positions based on prompts from the experimenter. Although not directly supervised, this period was most similar to previous recordings because it occurred during a convenient time for the infant and caregiver to play while they received instructions from the experimenter. The **first goal** of the current study was to determine the accuracy of body position classification during the proximal comparison period using this novel, semi-supervised procedure in participants' homes.

A second, *distal comparison* period followed the proximal comparison period and captured approximately 90 minutes of home behavior that was completely unsupervised. Caregivers and infants could (and did) do whatever they wished. Because this recording happened a considerable amount of time after the initial setup and instructions from the experimenter, accuracy could decline if caregivers or infants moved the garment or sensors. Moreover, increasing variation in everyday activities during the distal comparison creates a greater challenge for classification, testing whether machine learning models can generalize to novel test cases. Thus, the **second goal** of our study was to assess accuracy during the distal comparison.

After this second video recording period, we asked caregivers to keep infants wearing IMUs for the rest of the day until their regular bedtime, creating the **first real-time, full-day dataset of infant body position**. Interpreting such data required caregivers to log when infants napped, when they removed the sensor garment for diaper changes or

other reasons, and when infants went to bed at the end of the day. The **third goal** of the study was to examine the quality of the full-day data. Could body position be successfully recorded over the desired period? Finally, the novel full-day dataset affords us a unique chance to ask whether the estimated time infants spent in different body positions align with past results using video and survey methods. Thus, the **fourth goal** of the current study was to determine whether full-day body position measurements conformed to expectations about age differences in body position. Based on past results (Franchak, 2019), infants should spend increasingly more time sitting and upright but less time supine over the age range tested (4 to 14 months).

Methods

Participants

Apparatus

Procedure

Body position annotation

Body position classification

Results

Goal 1: Assess the proximal accuracy of body position classification models

Goal 2: Assess the distal accuracy of body position classification models

What can we learn about the two outliers? What do they tell us about how this will work in the future.

One outlier (106-4) had significant confusion between sitting and supine. Both their proximal and distal performance was bad, but only for sitting and supine. Watching the videos showed that this participant spent a long period of time in a seating device that was

reclined between supine and sitting. Also spent long time in mother's arms in a similar supine/sitting position that was hard to classify.

The other outlier (107-3) had good proximal accuracy but poor distal accuracy. Confusion between upright and suspended. Infant was in a baby walker for most of part 2. Moved more like a baby held by a CG than a baby walking upright.

For both, other class performance was good, suggesting that it's about the particular events, not about the quality of the algorithm in applying to the participant overall.

Goal 3: Examine the data quality of full-day home recordings

Goal 4: Assess the suitability of full-day predictions for capturing age differences in body position

Discussion

References

- Adolph, K. E., & Tamis-LeMonda, C. S. (2014). The costs and benefits of development: The transition from crawling to walking. *Child Development Perspectives*, 8, 187–192. <https://doi.org/10.1111/cdep.12085>
- Adolph, K. E., Vereijken, B., & Denny, M. A. (1998). Learning to crawl. *Child Development*, 69, 1299–1312. <https://doi.org/10.1111/j.1467-8624.1998.tb06213.x>
- Airaksinen, M., Gallen, A., Kivi, A., Vijayakrishnan, P., Häyrynen, T., Ilén, E., Räsänen, O., Haataja, L. M., & Vanhatalo, S. (2022). Intelligent wearable allows out-of-the-lab tracking of developing motor abilities in infants. *Communications Medicine*, 2(1). <https://doi.org/10.1038/s43856-022-00131-6>
- Airaksinen, M., Räsänen, O., Ilén, E., Häyrynen, T., Kivi, A., Marchi, V., Gallen, A., Blom, S., Varhe, A., Kaartinen, N., et al. (2020). Automatic posture and movement tracking of infants with wearable movement sensors. *Scientific Reports*, 10(1), 1–13.
- Arif, M., & Kattan, A. (2015). Physical activities monitoring using wearable acceleration sensors attached to the body. *PLoS ONE*, 10, e0130851.
- Armstrong, B., Covington, L. B., Hager, E. R., & Black, M. M. (2019). Objective sleep and physical activity using 24-hour ankle-worn accelerometry among toddlers from low-income families. *Sleep Health*, 5(5), 459–465.
- Barbaro, K. de. (2019). Automated sensing of daily activity: A new lens into development. *Developmental Psychobiology*, 61(3), 444–464.
- Barbaro, K. de, & Fausey, C. M. (2022). Ten lessons about infants’ everyday experiences. *Current Directions in Psychological Science*, 31(1), 28–33. <https://doi.org/10.1177/09637214211059536>
- Bergelson, E., Amatuni, A., Dailey, S., Koorathota, S., & Tor, S. (2019). Day by day, hour by hour: Naturalistic language input to infants. *Developmental Science*, 22, e12715.
- Bergelson, E., Casillas, M., Soderstrom, M., Seidl, A., Warlaumont, A. S., & Amatuni, A. (2019). What do North American babies hear? A large-scale cross-corpus analysis.

Developmental Science, 22, e12724.

- Bruijns, B. A., Truelove, S., Johnson, A. M., Gilliland, J., & Tucker, P. (2020). Infants' and toddlers' physical activity and sedentary time as measured by accelerometry: A systematic review and meta-analysis. *International Journal of Behavioral Nutrition and Physical Activity*, 17(1), 14.
- Chen, Q., Schneider, J. L., West, K. L., & Iverson, J. M. (2022). Infant locomotion shapes proximity to adults during everyday play in the u.s. *Infancy*.
<https://doi.org/10.1111/inf.12503>
- Cliff, D. P., Reilly, J. J., & Okely, A. D. (2009). Methodological considerations in using accelerometers to assess habitual physical activity in children aged 0–5 years. *Journal of Science and Medicine in Sport*, 12(5), 557–567.
- Cristia, A., Lavechin, M., Scaff, C., Soderstrom, M., Rowland, C., Räsänen, O., Bunce, J., & Bergelson, E. (2020). A thorough evaluation of the language environment analysis (LENA) system. *Behavior Research Methods*, 53(2), 467–486.
<https://doi.org/10.3758/s13428-020-01393-5>
- Deng, W., Trujillo-Priego, I. A., & Smith, B. A. (2019). How many days are necessary to represent an infant's typical daily leg movement behavior using wearable sensors? *Physical Therapy*, 99(6), 730–738. <https://doi.org/10.1093/ptj/pzz036>
- Franchak, J. M. (2019). Changing opportunities for learning in everyday life: Infant body position over the first year. *Infancy*, 24, 187–209.
- Franchak, J. M. (2020). The ecology of infants' perceptual-motor exploration. *Current Opinion in Psychology*, 32, 110–114.
- Franchak, J. M., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant-caregiver social looking during locomotor free play. *Developmental Science*, 21, e12626.
- Franchak, J. M., Scott, V., & Luo, C. (2021). A contactless method for measuring full-day, naturalistic motor behavior using wearable inertial sensors. *Frontiers in Psychology*, 12.
<https://doi.org/10.3389/fpsyg.2021.701343>

- Galland, B. C., Taylor, B. J., Elder, D. E., & Herbison, P. (2012). Normal sleep patterns in infants and children: A systematic review of observational studies. *Sleep Medicine Reviews, 16*(3), 213–222. <https://doi.org/10.1016/j.smr.2011.06.001>
- Hager, E., Tilton, N., Wang, Y., Kapur, N., Arbaiza, R., Merry, B., & Black, M. (2017). The home environment and toddler physical activity: An ecological momentary assessment study. *Pediatric Obesity, 12*(1), 1–9.
- Herzberg, O., Fletcher, K. K., Schatz, J. L., Adolph, K. E., & Tamis-LeMonda, C. S. (2021). Infant exuberant object play at home: Immense amounts of time-distributed, variable practice. *Child Development, 93*(1), 150–164.
- Kadooka, K., Caufield, M., Fausey, C. M., & Franchak, J. M. (2021, April). Visuomotor learning opportunities are nested within everyday activities. *Paper Presented at the Biennial Meeting of the Society for Research in Child Development.*
- Karasik, L. B., Kuchirko, Y., Dodojonova, R. M., & Elison, J. T. (2022). Comparison of U.S. And Tajik infants' time in containment devices. *Infant and Child Development, 31*(4). <https://doi.org/10.1002/icd.2340>
- Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2011). Transition from crawling to walking and infants' actions with objects and people. *Child Development, 82*, 1199–1209. <https://doi.org/10.1111/j.1467-8624.2011.01595.x>
- Karasik, L. B., Tamis-LeMonda, C. S., & Adolph, K. E. (2014). Crawling and walking infants elicit different verbal responses from mothers. *Developmental Science, 17*, 388–395. <https://doi.org/10.1111/desc.12129>
- Kretch, K. S., Franchak, J. M., & Adolph, K. E. (2014). Crawling and walking infants see the world differently. *Child Development, 85*, 1503–1518. <https://doi.org/10.1111/cdev.12206>
- Kwon, S., Zavos, P., Nickle, K., Sugianto, A., & Albert, M. V. (2019). Hip and wrist-worn accelerometer data analysis for toddler activities. *International Journal of Environmental Research and Public Health, 16*(14), 2598.

- Lobo, M. A., Hall, M. L., Greenspan, B., Rohloff, P., Prosser, L. A., & Smith, B. A. (2019). Wearables for pediatric rehabilitation: How to optimally design and use products to meet the needs of users. *Physical Therapy*, 99(6), 647–657.
- Luo, C., & Franchak, J. M. (2020). Head and body structure infants’ visual experiences during mobile, naturalistic play. *PLoS ONE*, 15, e0242009.
- Majnemer, A., & Barr, R. G. (2005). Influence of supine sleep positioning on early motor milestone acquisition. *Developmental Medicine and Child Neurology*, 47, 370–376.
- Malachowski, L. G., Salo, V. C., Needham, A. W., & Humphreys, K. L. (2023). Infant placement and language exposure in daily life. *Infant and Child Development*.
<https://doi.org/10.1002/icd.2405>
- Mendoza, J. K., & Fausey, C. M. (2021). Everyday music in infancy. *Developmental Science*, 24(6). <https://doi.org/10.1111/desc.13122>
- Mendoza, J. K., & Fausey, C. M. (2022). Everyday parameters for episode-to-episode dynamics in the daily music of infancy. *Cognitive Science*, 46(8).
<https://doi.org/10.1111/cogs.13178>
- Micheletti, M., Yao, X., Johnson, M., & Barbaro, K. de. (2022). Validating a model to detect infant crying from naturalistic audio. *Behavior Research Methods*.
<https://doi.org/10.3758/s13428-022-01961-x>
- Nam, Y., & Park, J. W. (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, 420–426.
- Patel, P., Shi, Y., Hajiaghajani, F., Biswas, S., & Lee, M.-H. (2019). A novel two-body sensor system to study spontaneous movements in infants during caregiver physical contact. *Infant Behavior and Development*, 57, 101383.
<https://doi.org/10.1016/j.infbeh.2019.101383>
- Perry, L. K., Prince, E. B., Valtierra, A. M., Rivero-Fernandez, C., Ullery, M. A., Katz, L. F., Laursen, B., & Messinger, D. S. (2018). A year in words: The dynamics and

- consequences of language experiences in an intervention classroom. *PLOS ONE*, 13(7), e0199893. <https://doi.org/10.1371/journal.pone.0199893>
- 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 Transactions on Biomedical Engineering*, 56, 871–879.
- Räsänen, O., Seshadri, S., Lavechin, M., Cristia, A., & Casillas, M. (2020). ALICE: An open-source tool for automatic measurement of phoneme, syllable, and word counts from child-centered daylong recordings. *Behavior Research Methods*, 53(2), 818–835. <https://doi.org/10.3758/s13428-020-01460-x>
- 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. *Applied Intelligence*, 45(2), 512–529.
- Schneller, M. B., Bentsen, P., Nielsen, G., Brønd, J. C., Ried-Larsen, M., Mygind, E., & Schipperijn, J. (2017). Measuring children’s physical activity. *Medicine & Science in Sports & Exercise*, 49(6), 1261–1269. <https://doi.org/10.1249/mss.0000000000001222>
- Soska, K. C., & Adolph, K. E. (2014). Postural position constrains multimodal object exploration in infants. *Infancy*, 19, 138–161. <https://doi.org/10.1111/inf.12039>
- 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. *Medicine and Science in Sports and Exercise*, 50(12), 2595–2602.
- Thurman, S. L., & Corbetta, D. (2017). Spatial exploration and changes in infant-mother dyads around transitions in infant locomotion. *Developmental Psychology*, 53, 1207–1221.
- Warlaumont, A. S., Sobowale, K., & Fausey, C. M. (2021). Daylong mobile audio recordings reveal multitime-scale dynamics in infants’ vocal productions and auditory experiences. *Current Directions in Psychological Science*, 31(1), 12–19. <https://doi.org/10.1177/09637214211058166>

- Wass, S., Phillips, E., Smith, C., Fatimehin, E. O., & Goupil, L. (2022). Vocal communication is tied to interpersonal arousal coupling in caregiver-infant dyads. *eLife*, 11. <https://doi.org/10.7554/elife.77399>
- Weisleder, A., & Fernald, A. (2013). Talking to children matters: Early language experience strengthens processing and builds vocabulary. *Psychological Science*, 24, 2143–2152.
- West, K. L., & Iverson, J. M. (2021). Communication changes when infants begin to walk. *Developmental Science*, 24(5). <https://doi.org/10.1111/desc.13102>
- Yao, X., Plötz, T., Johnson, M., & Barbaro, K. de. (2019). Automated detection of infant holding using wearable sensing: Implications for developmental science and intervention. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 3(2), 1–17.

Table 1

Summary statistics for model performance metrics shown separately for group and individual models.

Metric	Group			Individual		
	Median	Mean	SD	Median	Mean	SD
Overall Accuracy	0.913	0.852	0.145	0.926	0.917	0.074
Balanced Accuracy	0.918	0.899	0.074	0.918	0.910	0.074
F1	0.838	0.830	0.102	0.879	0.870	0.095
Sensitivity	0.859	0.840	0.122	0.856	0.846	0.125
Pos Pred Value	0.841	0.826	0.115	0.925	0.892	0.118
Kappa	0.783	0.764	0.156	0.862	0.826	0.148

Table 2

Model performance metrics for each body position category, shown separately for group and individual models.

Metric	Position	Group			Individual		
		Median	Mean	SD	Median	Mean	SD
Balanced Accuracy	Supine	0.974	0.923	0.109	0.998	0.963	0.082
	Prone	0.997	0.953	0.105	0.965	0.924	0.124
	Sitting	0.918	0.864	0.155	0.963	0.944	0.069
	Upright	0.929	0.861	0.154	0.935	0.872	0.134
	Held	0.915	0.884	0.107	0.861	0.851	0.147
F1	Supine	0.949	0.863	0.185	0.994	0.945	0.120
	Prone	0.971	0.921	0.139	0.943	0.898	0.160
	Sitting	0.912	0.810	0.250	0.955	0.926	0.098
	Upright	0.782	0.758	0.214	0.850	0.756	0.259
	Held	0.832	0.782	0.196	0.845	0.782	0.227
Sensitivity	Supine	1.000	0.921	0.166	1.000	0.959	0.123
	Prone	1.000	0.915	0.205	0.985	0.864	0.249
	Sitting	0.911	0.795	0.288	0.985	0.916	0.136
	Upright	0.870	0.755	0.293	0.907	0.783	0.274
	Held	0.875	0.793	0.218	0.793	0.715	0.297
Pos Pred Value	Supine	0.993	0.814	0.300	1.000	0.941	0.127
	Prone	1.000	0.892	0.217	0.986	0.903	0.191
	Sitting	0.926	0.841	0.242	0.974	0.947	0.071
	Upright	0.842	0.738	0.292	0.888	0.803	0.235
	Held	0.935	0.839	0.195	0.941	0.868	0.218
Kappa	Supine	0.934	0.762	0.311	0.989	0.921	0.158
	Prone	0.963	0.885	0.216	0.929	0.860	0.224
	Sitting	0.838	0.720	0.313	0.930	0.890	0.129
	Upright	0.735	0.688	0.284	0.820	0.730	0.264
	Held	0.808	0.757	0.205	0.827	0.740	0.268

Table 3

Correlations between human-coded and model-predicted body position durations across the entire long delay period. Correlations are provided within each posture and overall, and computed separately using group and individual models with and without outlier participants.

Position	With Outliers		Without Outliers	
	Group	Individual	Group	Individual
Held	0.02	0.04	0.73	0.60
Prone	0.97	0.86	0.97	0.84
Sitting	0.79	0.97	0.91	0.95
Supine	0.88	0.98	0.94	0.97
Upright	0.63	0.83	0.99	0.95
Overall	0.80	0.91	0.95	0.96

Table 4

Correlations between human-coded and model-predicted body position durations using 10-minute bins during the distal comparison. Correlations are provided within each posture and overall, and computed separately using group and individual models with and without outlier participants.

Position	With Outliers		Without Outliers	
	Group	Individual	Group	Individual
Held	0.51	0.46	0.67	0.63
Prone	0.96	0.90	0.96	0.89
Sitting	0.72	0.93	0.89	0.92
Supine	0.76	0.96	0.88	0.93
Upright	0.91	0.93	0.98	0.96
Overall	0.80	0.94	0.92	0.94

Table 5

Summary of age differences in full-day body position for younger (4- to 7-month) and older (11- to 14-month) infants. Values shown are the mean percent of time for each body position averaged across infants in each group. Standard deviations are shown in parentheses. Descriptive statistics are shown separately for group and individual models.

Position	Group		Individual	
	Younger	Older	Younger	Older
Upright	7.7% (9.3)	18.6% (7.4)	10.4% (13.6)	18.7% (8.4)
Sitting	24.9% (11.5)	44.4% (10.1)	18.8% (16.3)	46.9% (13.3)
Prone	14.4% (13.8)	14.4% (6.0)	12.5% (10.2)	16.9% (10.6)
Supine	38.5% (24.0)	14.0% (8.4)	40.2% (31.6)	10.0% (9.5)
Held	12.7% (7.1)	8.5% (5.4)	17.1% (20.8)	7.4% (7.6)

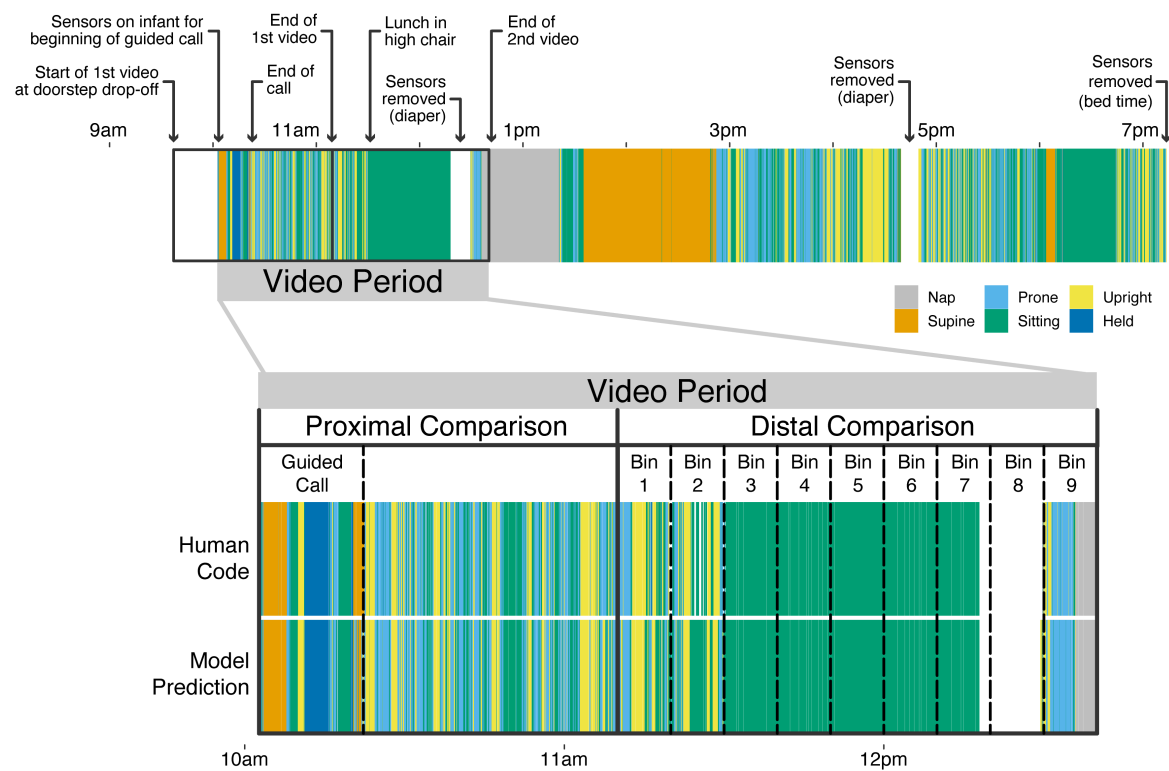


Figure 1. Example Timeline Caption.

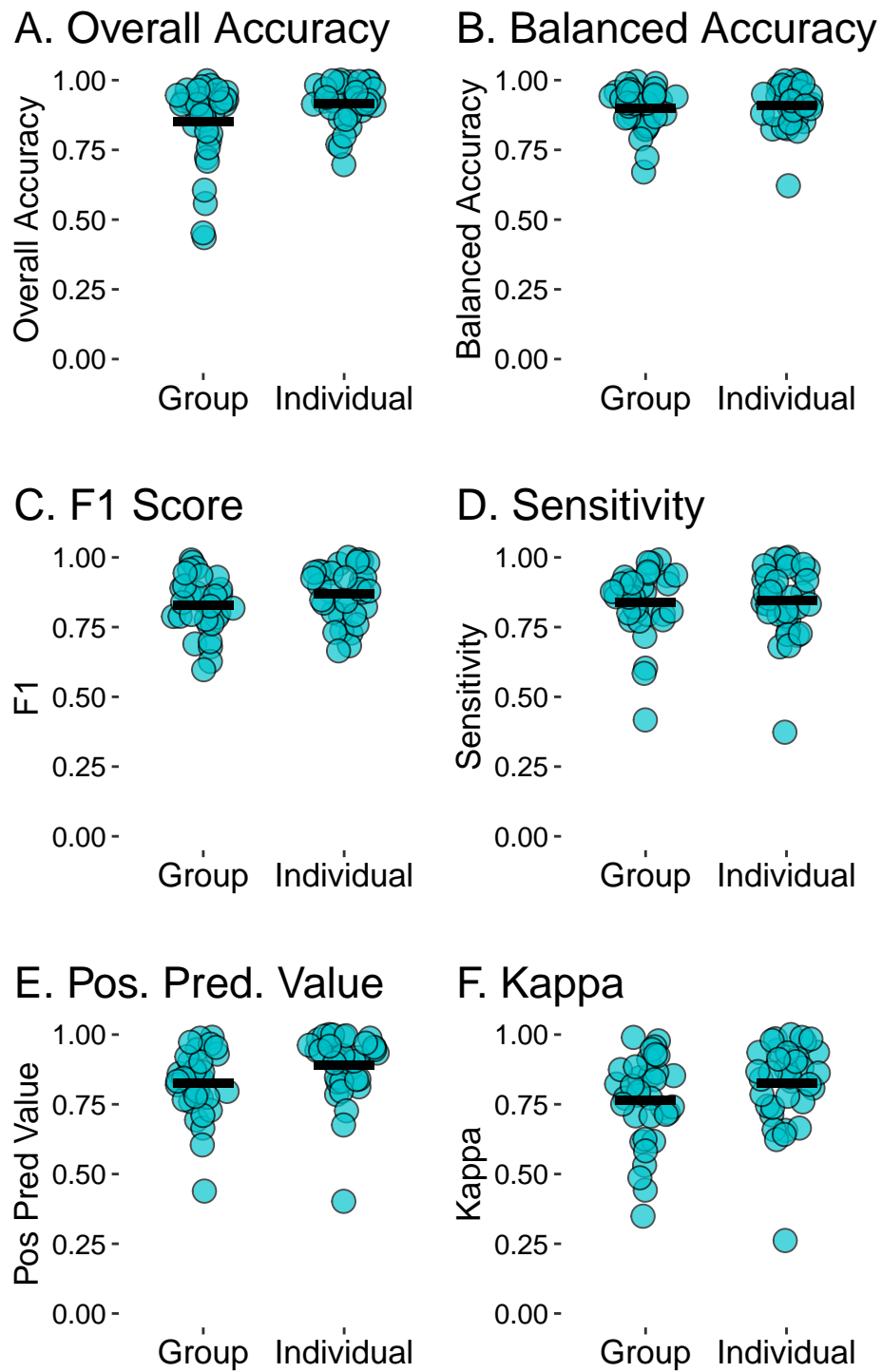


Figure 2. Metrics

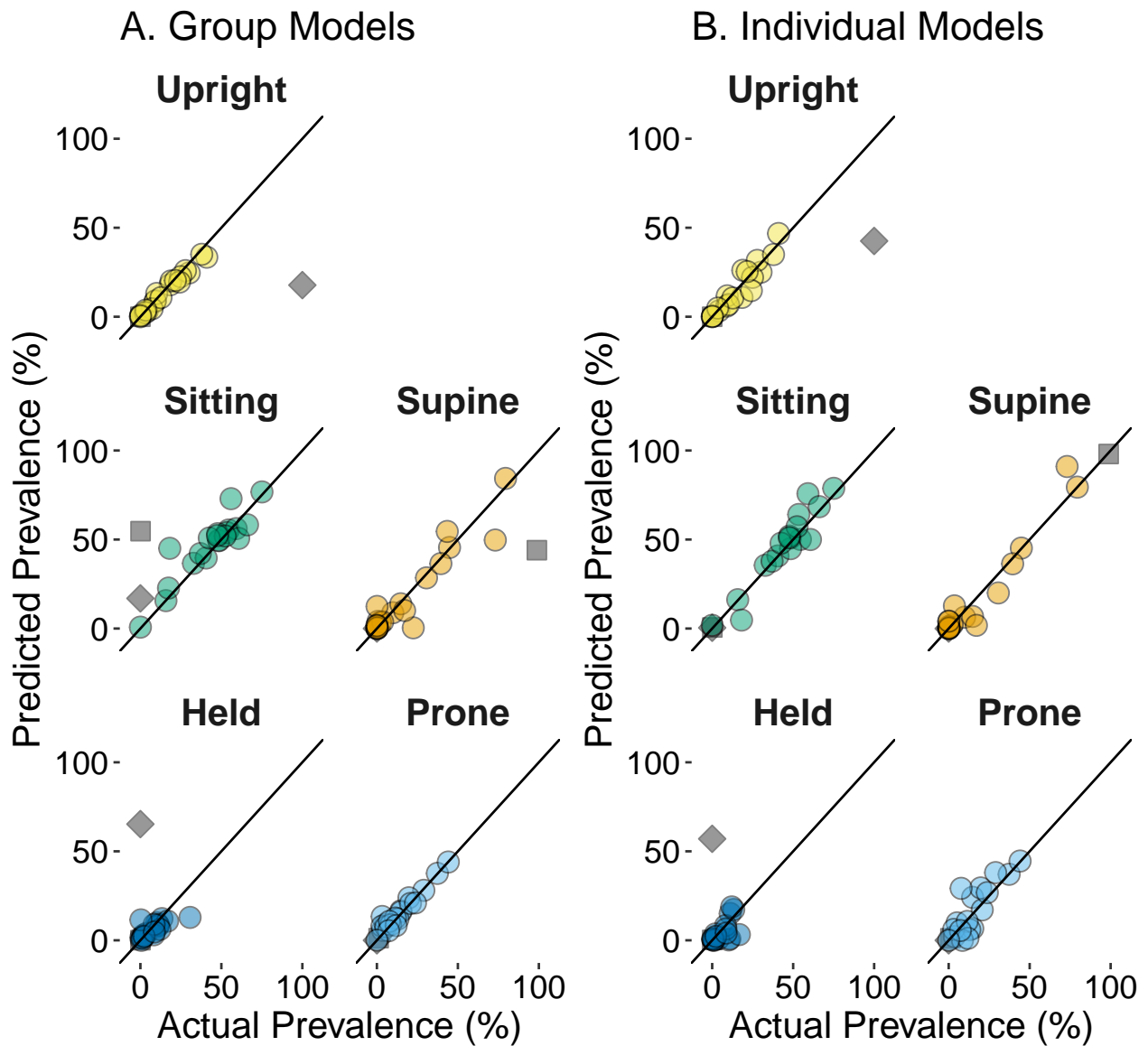


Figure 3. Overall agreement between human-coded body position and model-predicted body position in the distal comparison. Agreement for group models is shown in (A) and agreement for individual models is shown in (B). Plots are shown separately for each body position with a reference line that indicates perfect agreement; each point in a plot represent data for a single participant. The three outlier participants are plotted in dark gray, with a different shape marking each individual.

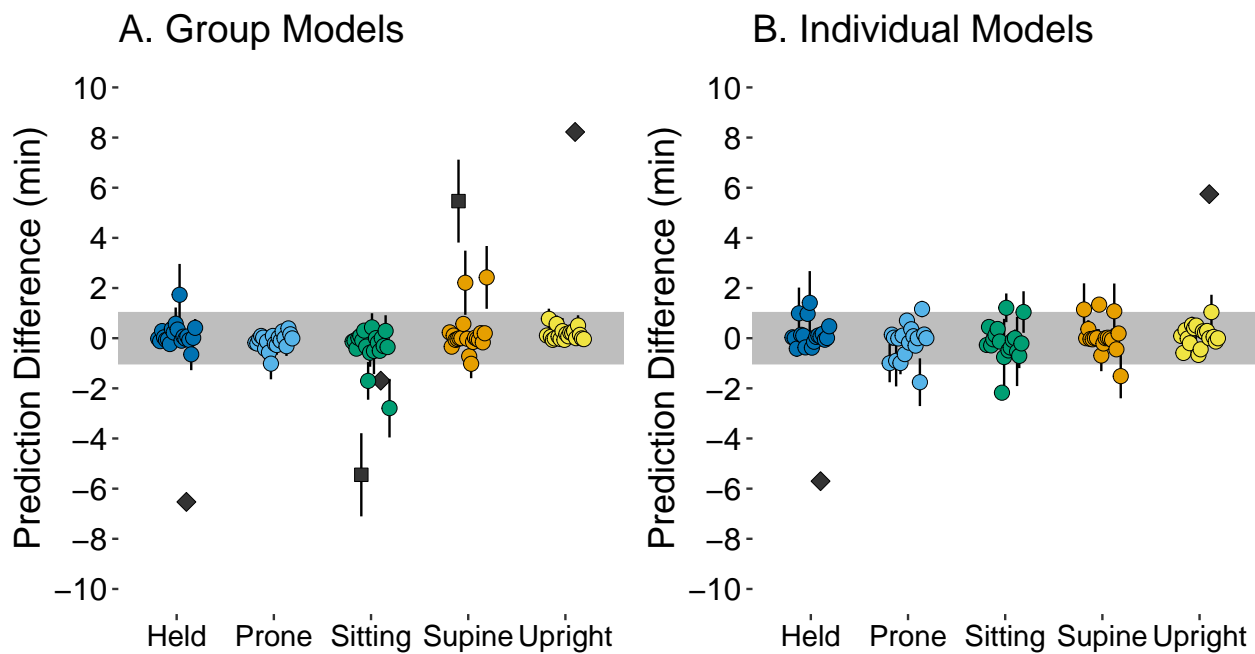
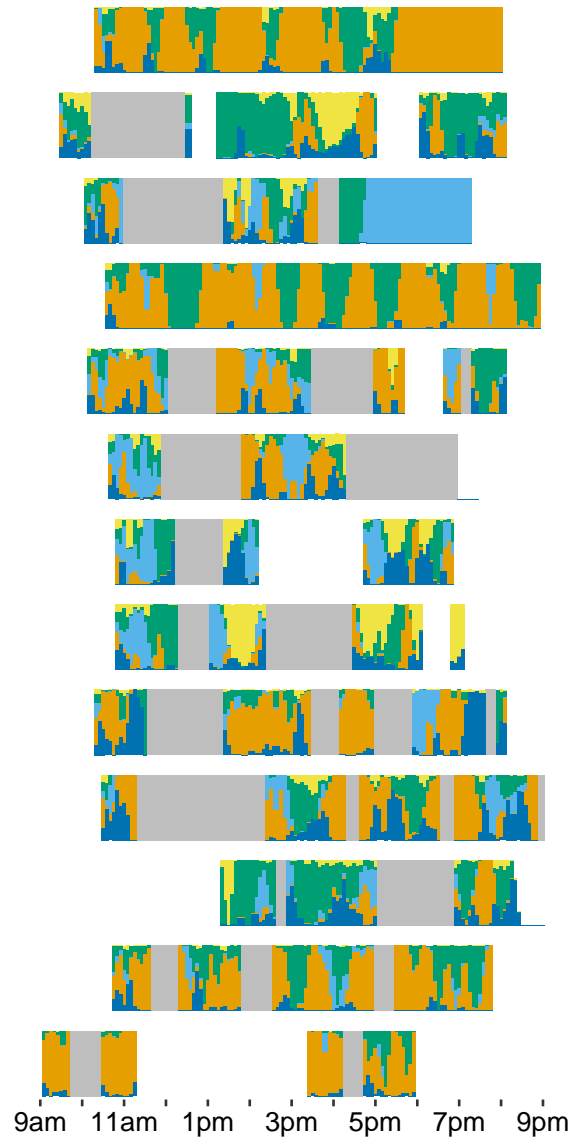


Figure 4. Prediction performance (difference in minutes between human-coded and model-predicted body position) for 10-minute bins in the distal comparison. Each point shows the mean and SE for a single participant for each body position, summarizing the prediction difference for each of their 10-minute bins. Points falling within the gray shaded region indicate that average prediction errors were less than 1 minute. Performance is plotted separately for (A) group models and (B) individual models. The three outlier participants are plotted in dark gray, with a different shape marking each individual.

A. 4–7 Months



B. 11–14 Months

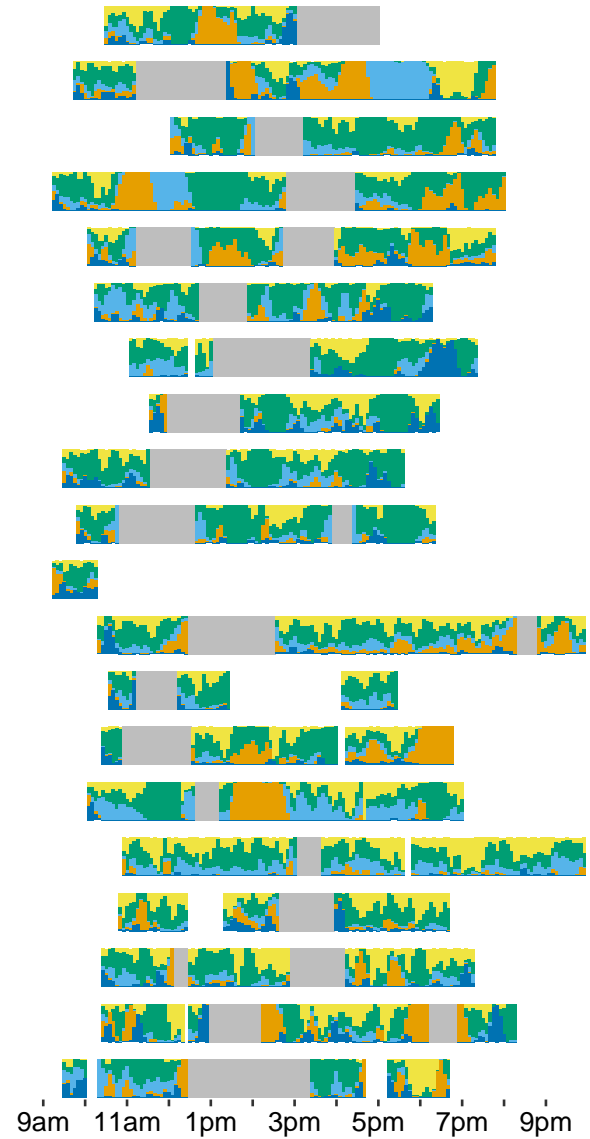


Figure 5. Timelines

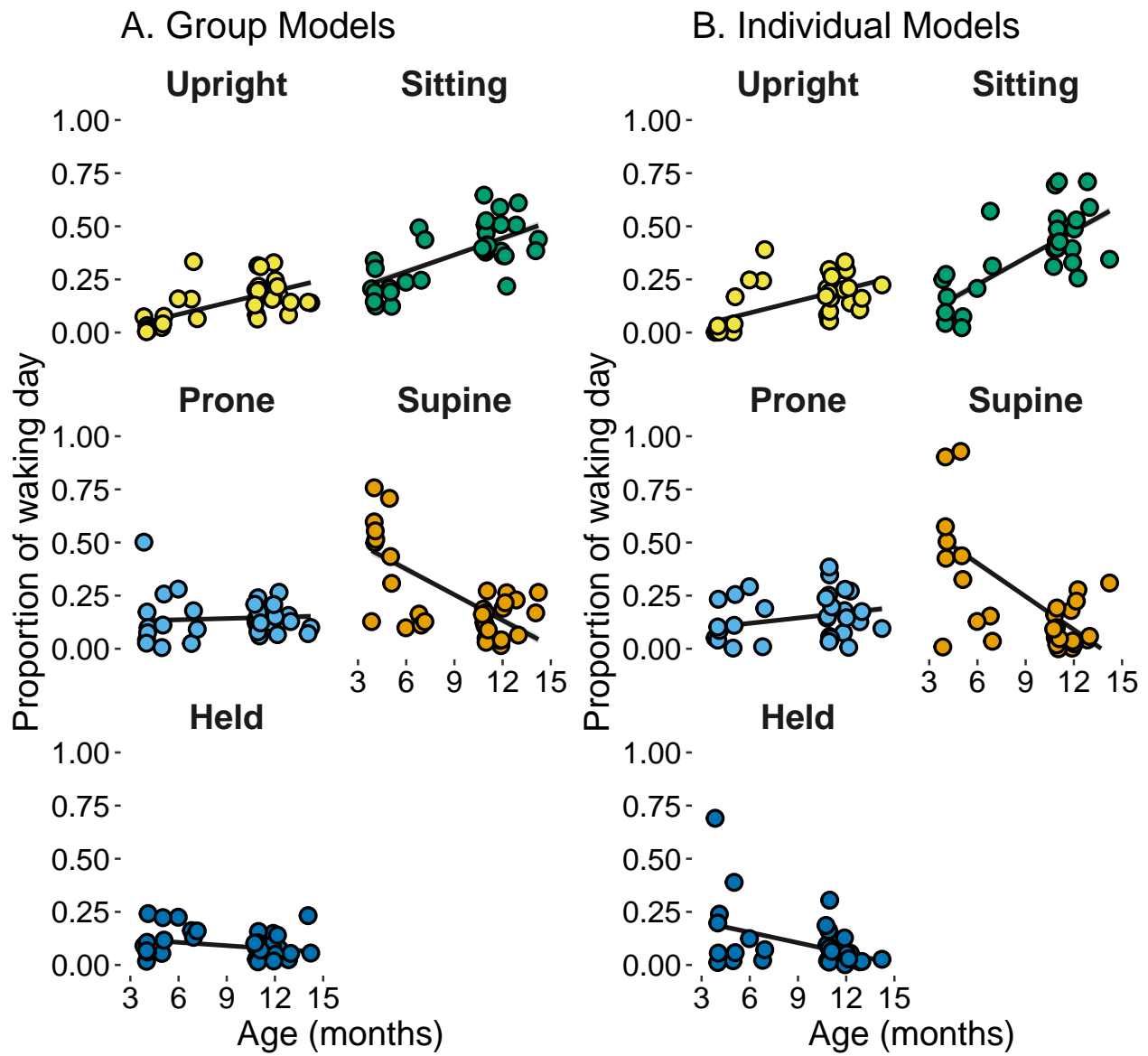


Figure 6. Age trends