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

John M. Franchak[1], Maximilian Tang[1], Hailey Rousey[1], & Chuan Luo[1]

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

Infants' movements facilitate and constrain how they interact with their surroundings from moment to moment. 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, the 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.

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 XX hours of video recorded across XX 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 [CITES]. 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 sessions [CITES], the current investigation takes a needed step forward by testing accuracy over long recordings.

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 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 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 input that predict later vocabulary. Full-day language recording synchronized with other data sources allows researchers to identify how vocal input and production interacts 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 to replicate in analagous 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 stored on device, 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 in 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. Full-day recordings are essential for capturing experiences across the heterogeniety of daily routines that moderate behavior (e.g., play, feeding, errands) [CITE]. 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).

- automatic classification with good-enough accuracy (scaling)

Limitations of Video and Survey Methods

Video. Only advantage is real time

Survey. Everything but real time

Promise of Inertial Sensing Methods

Wear time is not an issue. Lots of Smith and other studies with long recordings.

Classification accuracy from ML is good. older papers; Airaksinen; Frontiers

What's missing

Goals of the Current Study

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

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

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.

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/infa.12503

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., Kretch, K. S., & Adolph, K. E. (2018). See and be seen: Infant-caregiver social looking during locomotor free play. *Developmental Science*, *21*, e12626.

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

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

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

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

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

Soska, K. C., & Adolph, K. E. (2014). Postural position constrains multimodal object exploration in infants. *Infancy*, *19*, 138–161. https://doi.org/10.1111/infa.12039

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 multitimescale 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

Table 1

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

	Group			Individual		
Metric	Median	Mean	SD	Median	Mean	SD
Overall Accuracy	0.908	0.848	0.143	0.936	0.919	0.071
Balanced Accuracy	0.909	0.897	0.074	0.916	0.907	0.072
F1	0.813	0.827	0.101	0.873	0.870	0.089
Sensitivity	0.840	0.836	0.123	0.848	0.840	0.121
Pos Pred Value	0.822	0.815	0.125	0.915	0.899	0.108
Kappa	0.769	0.757	0.160	0.846	0.820	0.141

Table 2

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

		Group			Individual		
Metric	Position	Median	Mean	SD	Median	Mean	SD
Balanced Accuracy	Supine	0.977	0.935	0.099	0.995	0.957	0.090
	Prone	0.997	0.950	0.110	0.983	0.921	0.132
	Sitting	0.932	0.861	0.151	0.973	0.943	0.071
	Upright	0.930	0.867	0.146	0.937	0.877	0.123
	Held	0.907	0.872	0.111	0.892	0.854	0.153
F1	Supine	0.952	0.873	0.173	0.987	0.937	0.130
	Prone	0.965	0.923	0.124	0.958	0.889	0.186
	Sitting	0.897	0.805	0.247	0.963	0.927	0.101
	Upright	0.781	0.748	0.235	0.868	0.769	0.227
	Held	0.813	0.783	0.163	0.862	0.786	0.240
Sensitivity	Supine	1.000	0.945	0.125	1.000	0.948	0.141
	Prone	1.000	0.907	0.216	0.988	0.856	0.266
	Sitting	0.901	0.787	0.291	0.977	0.917	0.136
	Upright	0.878	0.767	0.275	0.893	0.790	0.255
	Held	0.867	0.774	0.229	0.817	0.722	0.309
Pos Pred Value	Supine	0.973	0.802	0.300	1.000	0.939	0.131
	Prone	0.991	0.896	0.196	0.993	0.901	0.197
	Sitting	0.910	0.840	0.238	0.967	0.947	0.078
	Upright	0.848	0.743	0.285	0.875	0.817	0.205
	Held	0.917	0.803	0.247	0.949	0.872	0.232
Kappa	Supine	0.932	0.770	0.307	0.984	0.913	0.169
	Prone	0.951	0.887	0.208	0.948	0.851	0.242
	Sitting	0.836	0.712	0.308	0.946	0.889	0.134
	Upright	0.729	0.700	0.277	0.829	0.744	0.232
	Held	0.786	0.733	0.219	0.845	0.744	0.280

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.

	With Outliers		Without Outliers	
Position	Group	Individual	Group	Individual
Held	-0.03	0.12	0.55	0.58
Prone	0.97	0.85	0.97	0.84
Sitting	0.73	0.93	0.91	0.97
Supine	0.84	0.94	0.94	0.97
Upright	0.84	0.93	0.99	0.94
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.

	With Outliers		Without Outliers	
Position	Group	Individual	Group	Individual
Held	0.45	0.43	0.57	0.55
Prone	0.96	0.89	0.96	0.88
Sitting	0.72	0.92	0.91	0.93
Supine	0.75	0.95	0.90	0.94
Upright	0.93	0.95	0.97	0.95
Overall	0.80	0.94	0.93	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.

	Group		Individual	
Position	Younger	Older	Younger	Older
Upright	7.6% (8.9)	18.6% (7.4)	9.9% (13.1)	18.7% (8.4)
Sitting	26.3% (12.1)	44.4% (10.1)	20.2% (16.3)	46.9% (13.3)
Prone	13.8% (13.5)	14.4% (6.0)	11.9% (9.9)	16.9% (10.6)
Supine	37.9% (23.2)	14.0% (8.4)	39.4% (30.3)	10.0% (9.5)
Held	12.7% (6.9)	8.5% (5.4)	17.6% (19.9)	7.4% (7.6)

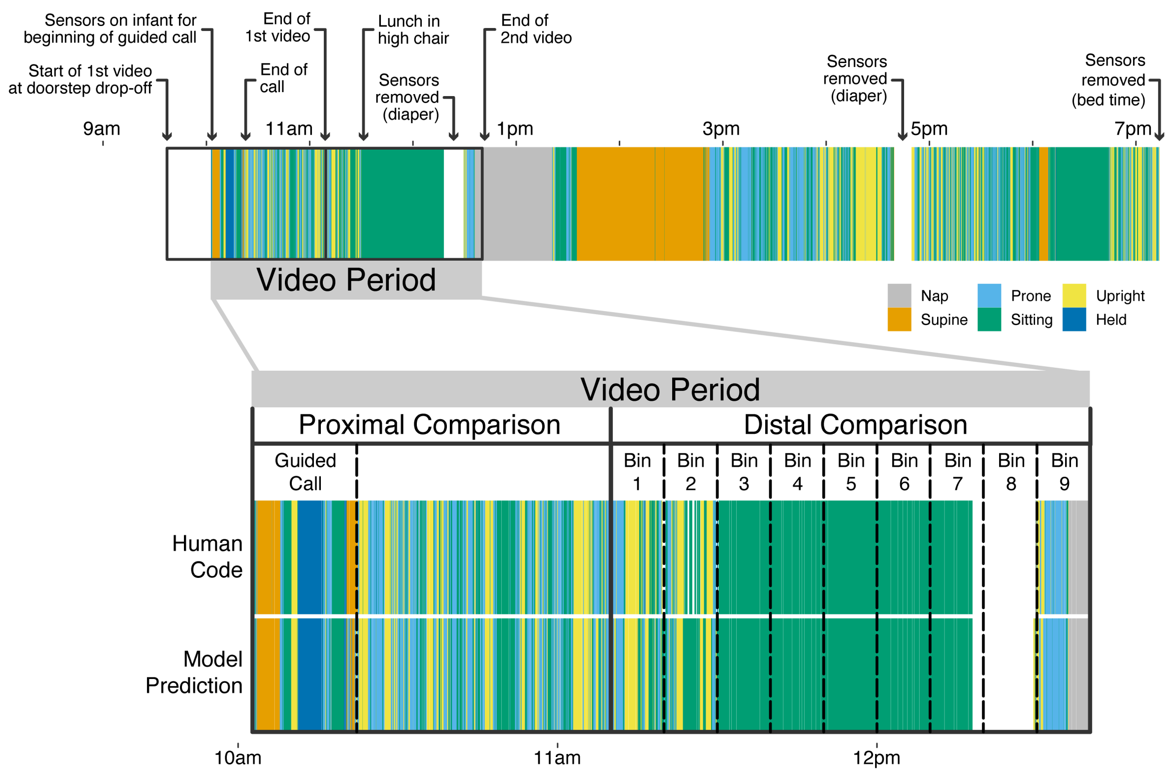

Figure 1. Example Timeline Caption.

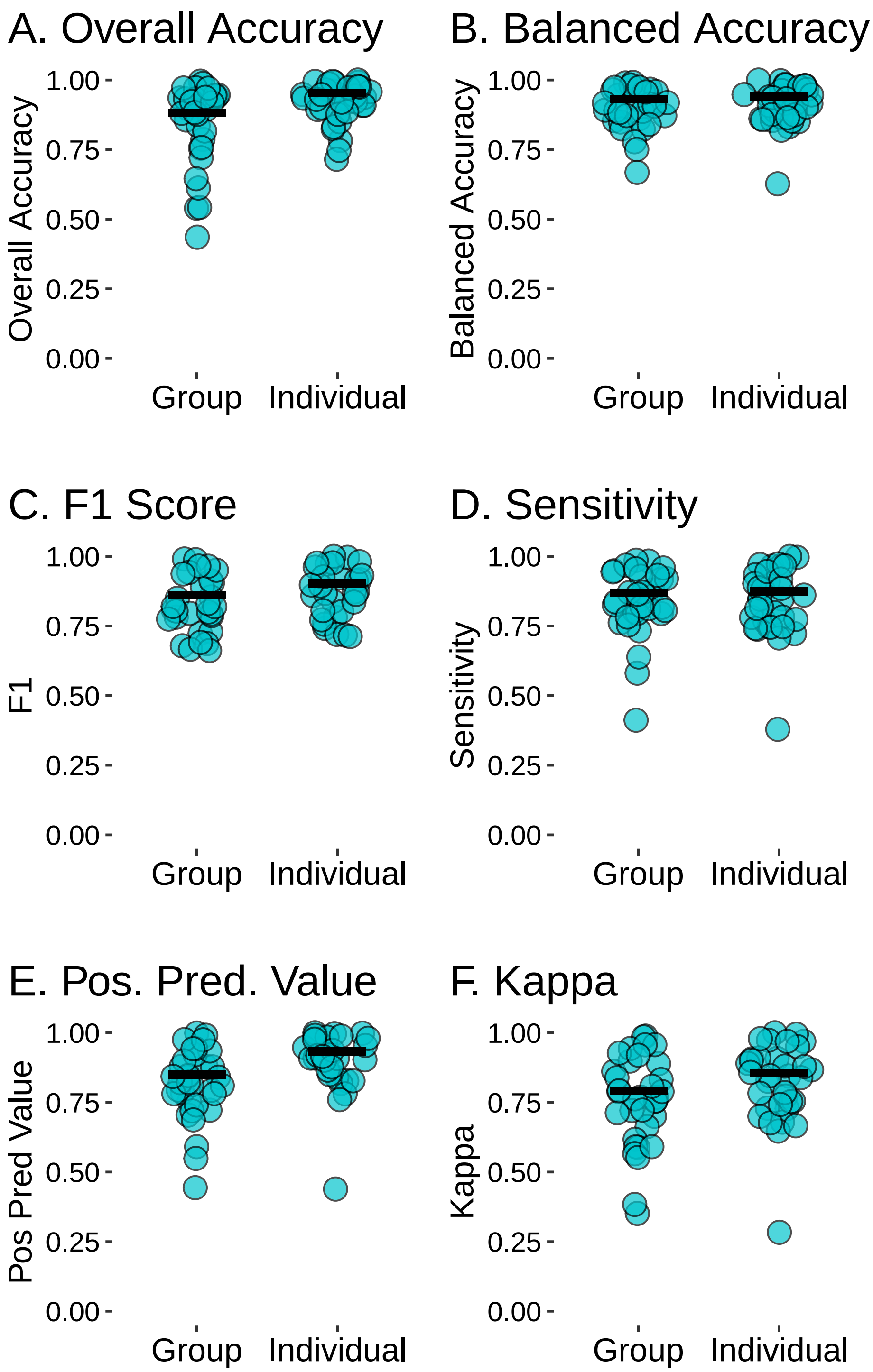

Figure 2. Metrics

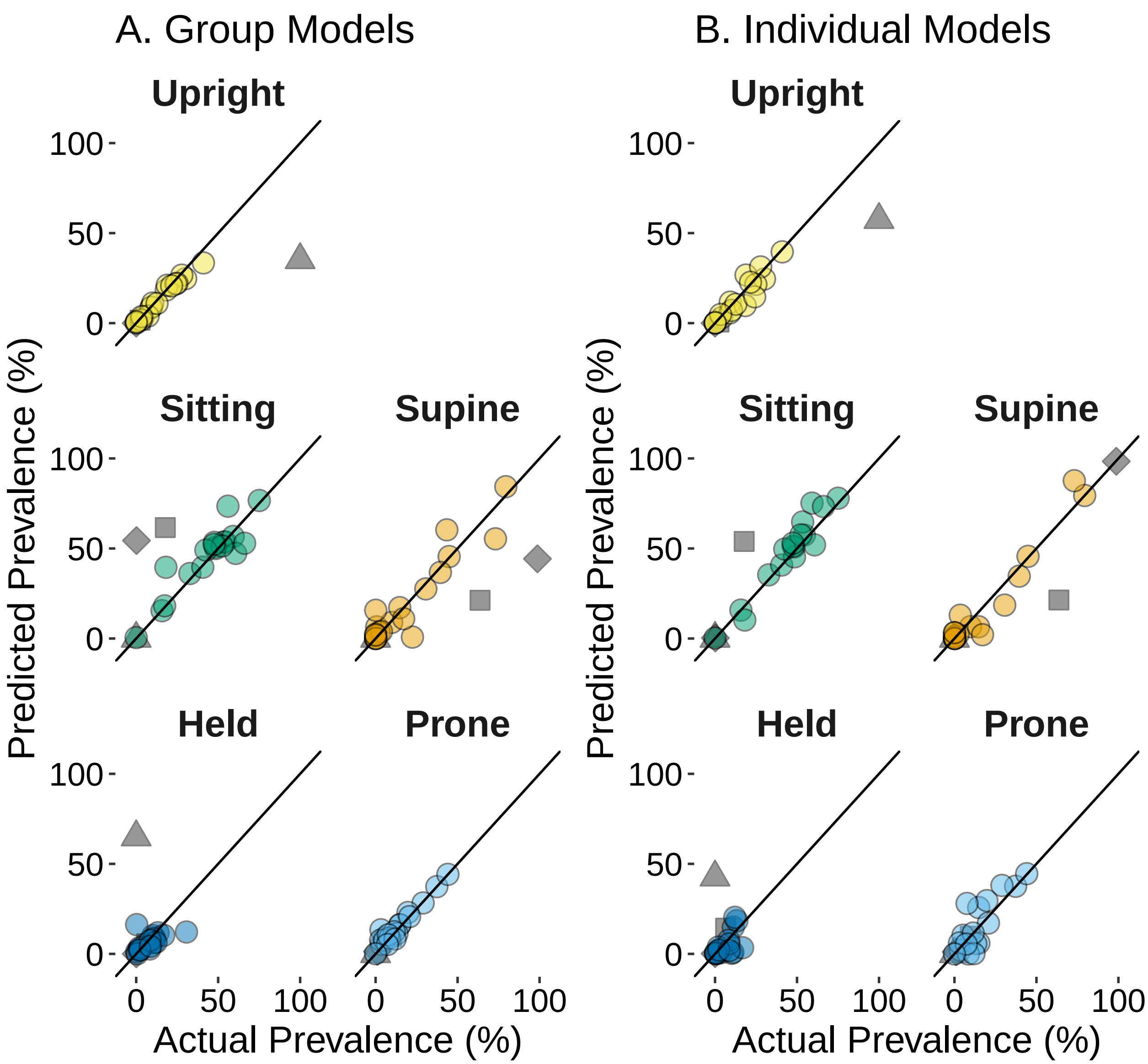

Figure 3. Overall agreement between human-coded body position and model-predicted body position in the distal comparison. Agreement for group models is shown in (A) and agreement for individual models is shown in (B). Plots are shown separately for each body position with a reference line that indicates 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.

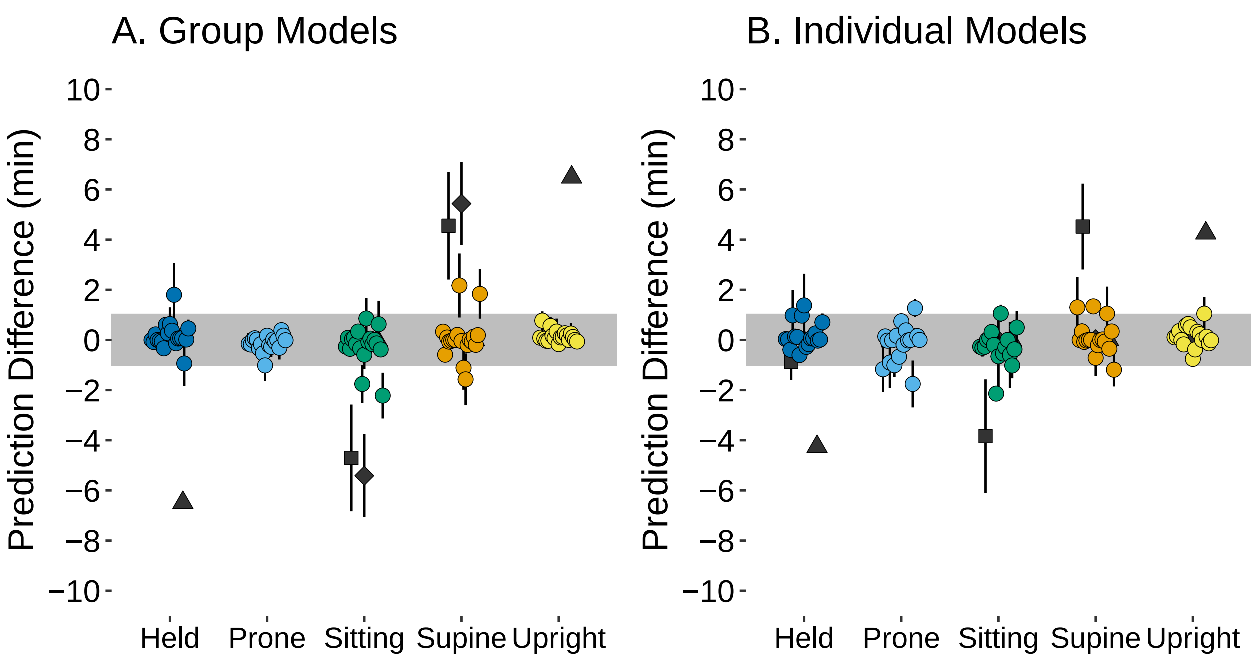

Figure 4. Prediction performance (difference in minutes between human-coded and model-predicted body position) for 10-minute bins in the distal comparison. Each point shows the mean and SE for a single participant for each body position, summarizing the prediction difference for each of their 10-minute 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.

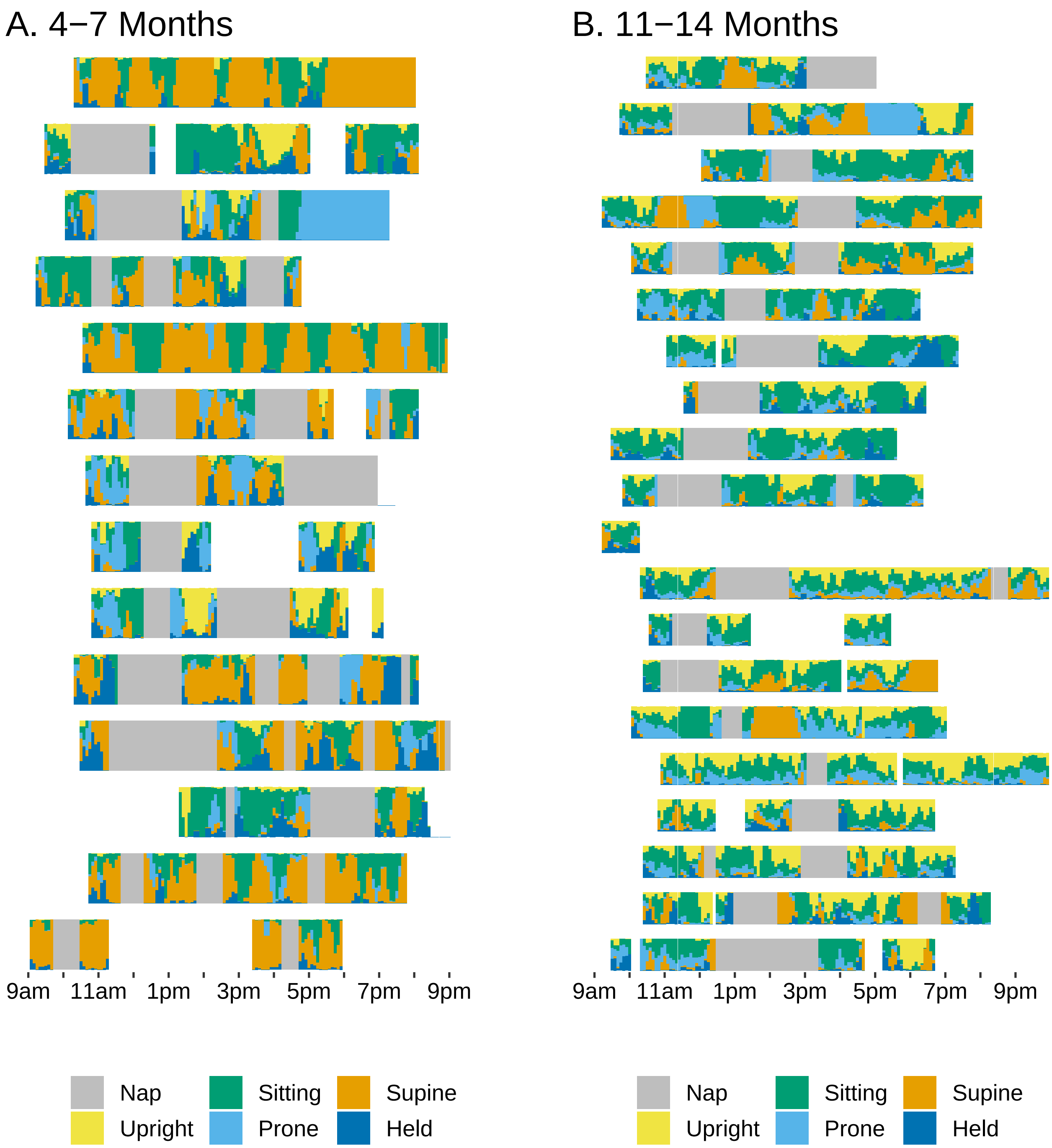

Figure 5. Timelines

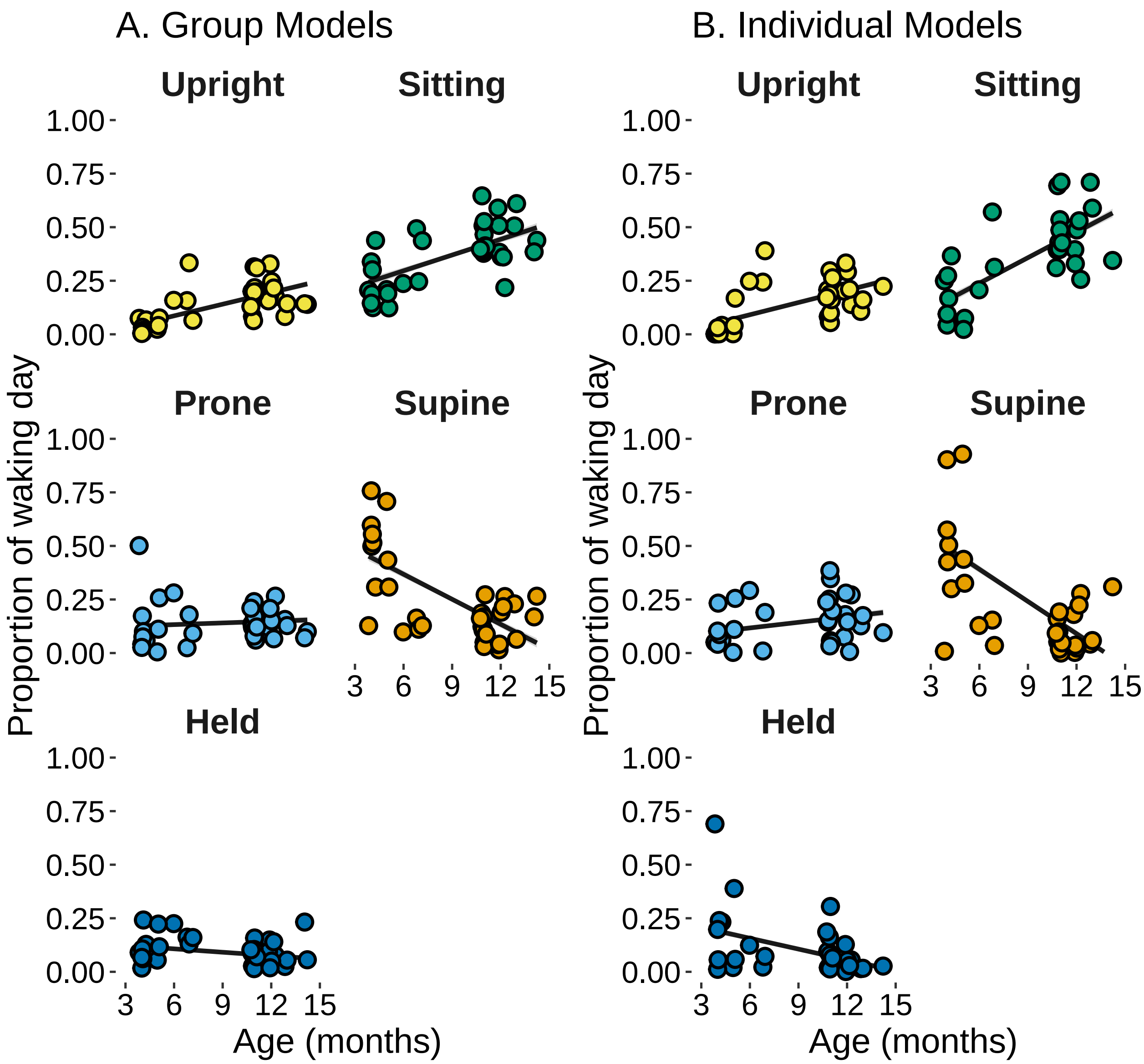

Figure 6. Age trends