Do we need expensive equipment to quantify infants' movement? A cross-validation study between computer vision methods and sensor data

David López Pérez
Institute of Psychology
Polish Academy of Sciences
Warsaw, Poland
0000-0002-1235-6376

Zuzanna Laudańska
Institute of Psychology
Polish Academy of Sciences
Warsaw, Poland
0000-0001-6790-9559

Alicja Radkowska
Institute of Psychology
Polish Academy of Sciences
Warsaw, Poland
0000-0002-0909-3368

Karolina Babis
Institute of Psychology
Polish Academy of Sciences
Warsaw, Poland
0000-0003-2002-5371

Agata Kozioł
Faculty of Psychology
University of Warsaw
Warsaw, Poland
0000-0002-6469-4107

Przemysław Tomalski
Institute of Psychology
Polish Academy of Sciences
Warsaw, Poland
0000-0002-0390-5759

***Abstract*—Recent progress in the study of infant motor development has been achieved by ground-breaking paradigm shifts combined with clever and innovative tasks that place the infant center stage as the acting subject. One of the challenges that developmental scientists are facing today is understanding the complexity of infants' spontaneous movements. Novel methods such as wearables and computer vision methods have the potential to revolutionize the measurement of infants' motor behavior in various situational and social contexts. However, a comparison between any computer method and wearables data has not been carried out so far in this age group for spontaneous behavior during social interactions with a caregiver. In this paper, we compare the results of Deeplabcut, an algorithm for tracking user-defined body parts, with simultaneously acquired wearable data and show that computer vision can be a good alternative to advanced wearable systems.**

Keywords—Computer Vision, IMUs, wearables, DeepLabcut, infant movement

I. Introduction

The development of human motor behavior involves learning through practice as infants improve their skills over time. Changes in behavior take place across long periods of time (days, weeks, or months). Being able to capture constituent movements in detail across longer timeframes in multiple situations may help us discover developmental trajectories of emerging motor skills [1]. Standardized assessments of motor milestones, which assume a specific order or progression and age of acquisition, cannot reliably capture the complexity of infant development. To achieve this goal, a large variety of sophisticated, lab-based methods and experimental designs have been used for many decades (e.g., instrumented floors, treadmills, force plates, video), but thanks to recent advances in technology, we are now able to characterize infants' spontaneous movements with much higher accuracy and less manual labor. New technologies, like small wearable sensors and computer vision for analyzing movement in video recordings may greatly facilitate the measurement of infants' natural activity.

Wearable motion trackers are now available in a wide variety of choices ranging from basic accelerometers to the more advanced Inertial Motion Units (IMUs), and they can be even found in the forms of miniature sensors embedded in baby suits (e.g., [2]). Simple accelerometers measure the 3-dimensional (3D) acceleration (the rate of change in velocity of an object) of a body part, so they can be used, for example, to explore infants' physical activity and sedentary behaviors across daytime hours (see review by [3]). More advanced IMUs typically combine accelerometers and gyroscopes with magnetometers that measure the strength and direction of the Earth-magnetic field. By using a biomechanical model of the human body (see more details in [4]) or additional information from magnetometers (e.g., [5]), we can even track the dynamic motion and estimate the orientation and position changes of body segments in 3D.

Wearable motion trackers can precisely detect subtle changes in infants' postures and body movements that are not easy to perceive with the naked age, but they do not yet offer a complete or easily deployable solutions for developmental research. They are expensive, the weight and size of the currently available ones may not be suitable for the youngest infants and newborns, they may be difficult to use with some clinical populations (e.g., infants with sensory difficulties who may be irritated by additional weight on limbs), and the majority of promising results come mostly from proof-of-concept studies (e.g., [6-7]). Thus, video-based measurement is still considered the gold standard in infant studies because it is the only tool able to capture the richness and complexity of

behaviors as well as the details of the surrounding context (e.g., [1]). Available for almost a century (e.g., [8]), it can be used to study infant behavior across long periods of time (days, weeks, or months), and it is increasingly available through multiple devices (e.g., smartphones, cameras). Video recordings reveal the physical and social context in which a behavior occurs; they are inexpensive, readily accessible and provide a potentially unobtrusive way of recording infants' behavior in their natural environment (e.g., [9]), and may even serve as a representation of an infant's field of view when head-mounted cameras are used (e.g., [10]). However, as technology improves, the amount of data also increases. The analysis of longer and more complex recordings of infants' natural activity requires laborious and costly manual coding. Unsurprisingly, this has recently led to the development of many new tools for automatic tracking of human posture (e.g., Deep Pose [11]; OpenPose [12]). One major advantage is that they enable the use of videos recorded for other purposes or available in open repositories to conduct large-scale analyses ([1]). Additionally, these tools are predominantly open-source and freely available, which makes them the most affordable of the methods for tracking infant movement. However, studies carried out so far used 2D previously recorded videos (e.g., [13-15]), where the infant was in a clear view, which leaves open (1) the accuracy of estimation in complex environments and (2) the utility of these methods in multi-person settings such as parent-infant interactions.

Developmental scientists are in need of cheap, reliable and unobtrusive methods for measuring the dynamics of body movement of young infants in naturalistic social interactions. Current wearable methods have limitations on their use across contexts and age groups. We aimed to test the reliability of computer vision methods as an alternative for studying spontaneous motor behavior of young infants. To this end, we compared the reliability of movement estimation between Deeplabcut (DLC) [16], a pose-estimation computer vision algorithm that predicts and tracks the location of a person, animal or object and allows the tracking of user-defined body parts, and IMU wearable data. The amount of computer vision methods to quantify movements has increased over the last decade [17,18], but we chose DLC due to its flexibility which allows the user to define what should be tracked and does not rely on a predefined skeleton. We show that video-based analyses can provide a good movement estimate of individual limbs of young infants in different interactive tasks, when compared against data collected with advanced IMUs.

II. MATERIALS AND METHODS

A. Participants

A total of 12 healthy, full-term (36 gestational weeks or more) infants at the age of 4 to 5 months (M = 4.49, SD = 0.22) contributed data for the analysis. Participants came from predominantly middle-class families living in the city with >1.5 million inhabitants. The study was approved by the local ethics committee and conformed to the standards of the Declaration of Helsinki. Prior to the testing, all parents gave written informed consent. For their participation, the families received a diploma and a small gift (a baby book).

B. Procedure

During lab visit, dyads of parents and infants participated in several semi-structured interactions with age-appropriate toys. Interactions were recorded in a laboratory room, in a carpeted play area, using three remote-controlled CCTV color cameras in HD quality with a 1920x1080 resolution at 25 Hz. During the interaction, one experimenter operated the cameras (this included zooming in and out as well as moving cameras vertically and horizontally) to ensure that at least one camera captured the infant's behavior and one camera captured the parent's behavior. In this paper, we compare data from two tasks varying in demands: play with rattles (Task 1) and free play (Task 2). Task 1 was a semi-structured play: the infant was placed in a baby bouncer (N = 9) or, in case of refusal, on the floor (N = 3). The dyad was provided with 4 rattles (2 for a baby and 2 for a parent) and asked to use them in a play with infant for 4-5 minutes. In task 2, the dyads were given a set of age-appropriate toys (stuffed teddy bear, puppet, plush fruits, children's books, rattles, teething toys, and colorful rubber blocks) and the caregiver was asked to play with their infant as they usually do at home for 10 min. Infants' position was not constrained with any positioning device in this task to provide a more naturalistic set-up.

In both tasks, the caregiver was asked to clap at the beginning to mark the start of the recording, which allowed for synchronization of wearable motion trackers with audio and video recording.

C. Sensor Data Acquisition

Infants' and caregivers' movements were recorded using 12 wearable motion trackers (MTw Awinda 3DOF, Xsens Technologies B.V., Enschede, Netherlands) at 40 or 60 Hz. The motion trackers were synchronized using Awinda Recording & Docking Station (Xsens Technologies B.V., Enschede, Netherlands) and operated with MT Manager Software (Version 4.6.0, Xsens Technologies B.V, Enschede, Netherlands) running on a computer with Windows 10 (Microsoft, Inc.). All sensors were synchronized within MT Manager Software. Motion trackers were placed in black textile pockets attached to soft black straps with Velcro at the ends. The length of each strap was adjusted to allow for comfortable positioning on infants' and caregivers' limbs, heads and torsos. For the current analyses, we focused on the infants' leg movements (1 sensor on each ankle); however, straps with sensors were also placed on the infants' arms (1 sensor just above each wrist), head (1 sensor on the side of the head) and torso (1 sensor in the middle, 1 on the left side). Caregivers had 1 sensor placed on each arm (just above each wrist), 1 on the head and 2 on the torso (1 sensor in the middle, 1 on the left side).

The data from IMUs was converted into two time series. First, we collapsed the three-dimensional information from the IMUs to a one-dimensional overall acceleration time series by calculating the magnitude of acceleration for each three-

dimensional data point. Second, since the IMUs contain magnetometers, we also calculated quaternions, which offer a robust estimation of changes in orientation.

D. Deeplabcut

To extract movement from the videos, we used an open-source toolbox called Deeplabcut (DLC) that builds on a state-of-the-art pose estimation algorithm (a computer vision technique that predicts and tracks the location of a person, animal, or object) to precisely track user-defined body parts [16]. DLC uses transfer learning (i.e., the ability to take a network that has been trained on one task to perform another) and an 'extremely deep neural network' [16] which leads to a relatively small amount of data required to train the model. We manually labelled the location where the sensors were placed on each participant (see section II.C and Figure 1 for an example). In those cases where the sensors were not visible, the same task was labelled again using a different camera view.

Two hundred frames were labelled for each video (~1% of the total number of frames) and the algorithm was trained for 600.000 using supercomputing resources (GPU Nvidia Tesla K40 XL at the Akademickie Centrum Komputerowe Cyfronet AG, Krakow, Poland). We used 200 frames because previous studies have shown that even a small number is sufficient to obtain good performance [17]. After the training was finished, the algorithm was used to estimate the coordinates for the remaining frames. Before the analysis, the data were filtered using a 1-dimensional median filter with a window size of 7 to avoid 1-time-point-outliers (e.g., errors in the tracking) for each coordinate. DLC returned a set of coordinates for each tracked body part. To quantify movement, we calculated the Euclidean distance between the x and y coordinates of consecutive frames for tracked body parts.

Figure 1. Example of DLC labelling. The user-defined body parts are represented with color dots of 8-pixel radius. The parent gave written consent for the publication of the picture.

E. Data Analysis

Prior to the analysis, the DLC and manually coded data were resampled to match the sensors' frequency. Since sensor and video data were not in synchrony, we used clapping at the beginning of each task to find the lag between video and wearable systems. First, we manually coded each clap in a frame-by-frame manner using ELAN software [19]. The times series were categorized as 1 during the duration of the clap and 0 otherwise. Second, we categorized the DLC and sensor movement time series using a median split, where values larger than the median were categorized as 1 and smaller as 0. Finally, we used cross-lag recurrence quantification analysis (CRQA, [13]) to find the delay between the coded clapping and the categorized sensor movement data. Negative delay values meant that the sensors were turned on earlier than the video acquisition. In this case, the categorized movement time series derived from videos (DLC and manually coded time series) were zero padded at the beginning for as many seconds as the delay represented and the equivalent length was removed at the end. However, if the delay was positive (i.e., video turned on earlier), the time series were zero-padded at the end, and the equivalent length was removed at the beginning of the time series.

Next, we computed the Pearson correlation coefficients between the time-aligned sensor and DLC time series. To test that the correlations do not arise by chance, we also calculate the cross-correlations between randomized time series. The correlation between two time series as a single number may be uncorrelated on short timescales due to noise but strongly correlated on larger wavelengths. Thus, we plotted the semblance, which is the correlation as a function of both time and wavelength and it varies both as a function of time, and of frequency [20]. In this plot, correlated values are plotted in red (correlation =1) and anticorrelated in blue (correlation = -1).

III. Results

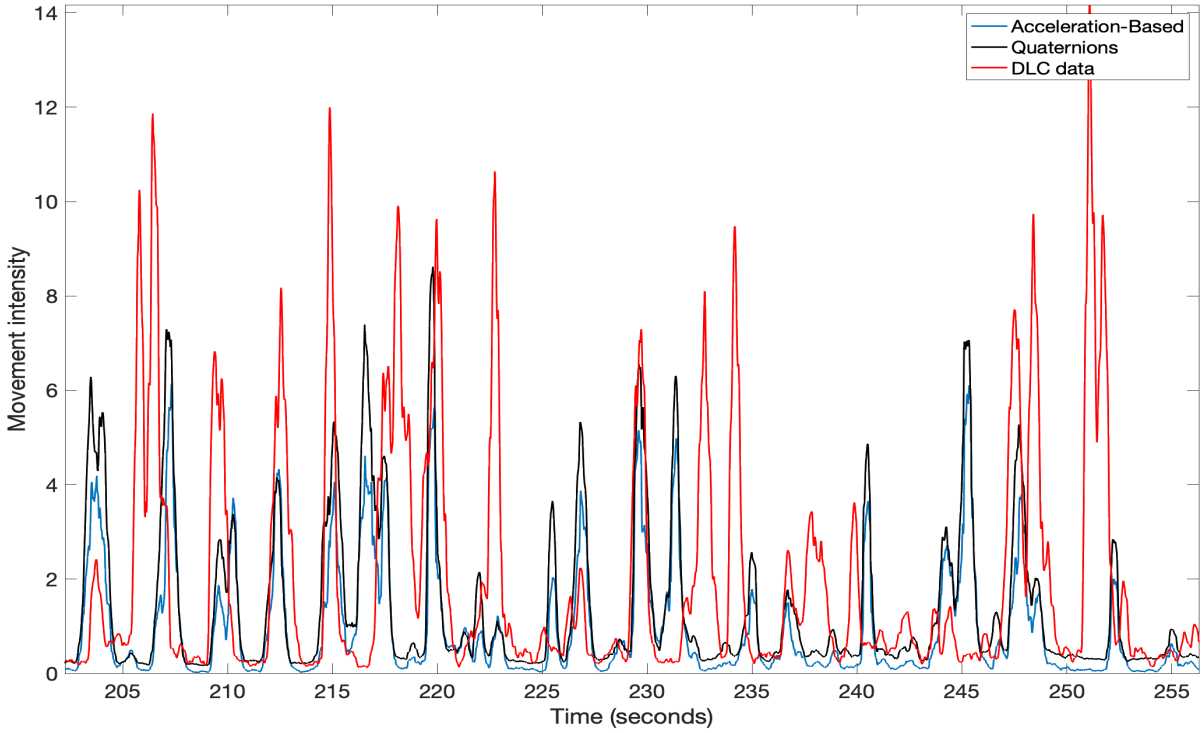
A. Deeplabcut

DLC was able to track the labelled body parts with high precision in all the videos with a small training error in both tasks, which highlights the high accuracy of the algorithm (see descriptives in Table 1). The higher error in Task 2 is probably due to the infants' position not being as restricted in movement as in Task 1, but given the high resolution of the videos in pixels, the efficiency of the algorithm is remarkably high.

TABLE I. DEEPLABCUT ERROR DESCRIPTIVES. THE VALUES ARE GIVEN IN NUMBER OF PIXELS.

	DeepLabCut Descriptives	
	Task 1	***Task 2***
Training Error	2.20 ± .12 [1.85-2.37]	2.26 ± .14 [2.04-2.53]
Test Error	4.54 ± 1.20 [2.73-6.85]	6.67 ± 1.77 [5.14-11.04]

Figure 2. Example of the extracted movement time series for Deeplabcut (red), Acceleration-based (blue) and quaternion (black) data. The time series has been scaled to facilitate visualization. This example presents selected 55 sec for left hand movements of one of the participants.



B. *Cross-Validation Analysis*

Figure 2 shows an example 55 sec of the obtained time series for the left hand of one of the participants. Visual inspection of the plots suggests that all the time series present good agreement in the movement estimation. To find out the similarity between them, we compared statistically the time series using Pearson correlation analysis performed for each participant. DLC time series significantly correlated for both legs (all *p*s < .001) with both the acceleration-based measures (mean Pearson coefficient for the left leg: r_{left} = .19 ± .09, [.12 – .36]; mean for the right leg: r_{right} = .18 ± .12, [.12 – .37]) and the quaternions (r_{left} = *.21* ± *.11*, [.12 - .40]; r_{right} = .18 ± .11, [.12 – .39]). For control purposes we calculated correlation coefficients for the randomized time series. No significant correlations were obtained (all *p*s > .05) either for the acceleration-based measures (mean for the left leg: r_{left} = .02 ± .02; [.00 – .07], mean for the right leg: r_{right} = .02 ± .02, [.01 – .06]) or for quaternions (r_{left} = .02 ± .02, [.00 – .07]; r_{right} = , = .02 ± .01, [.00 – .05]).

Next, we tested the possibility that the correlations were affected by the fact that the time series were not fully synchronised. A Matlab *xcorr* function was used to find the alignment between the DLC and sensor time series, which produces the maximum correlation coefficients. This analysis showed that there was on average a delay of ($delay_{left}$ = .143 ± .04, [.06 – .16]; $delay_{right}$,= .141 ± .05 [.01 - .166]) where the maximum correlation was found between DLC and both acceleration based measures (r_{left} = .43 ± .09, [0.36 – 0.67], r_{right} = .38 ± .10, [.31 – .64]) and quaternions (r_{left} = .45 ± .10, [.37 – .67], r_{right} =.,41 ± .10, [.33 – .67]). This suggests that even when the alignment was performed, a residual lag was still present between the DLC and sensor time series.

Finally, to better explain the effect of noise on our correlations, we estimated the change in the correlation in time and frequency for one sample time series. Figure 3 shows an example of a semblance plot for the time series for DLC and acceleration-based measures presented in Figure 2. Higher correlations were observed at higher frequencies, which suggests an additional influence of low frequency noise on the measurement of leg movements.

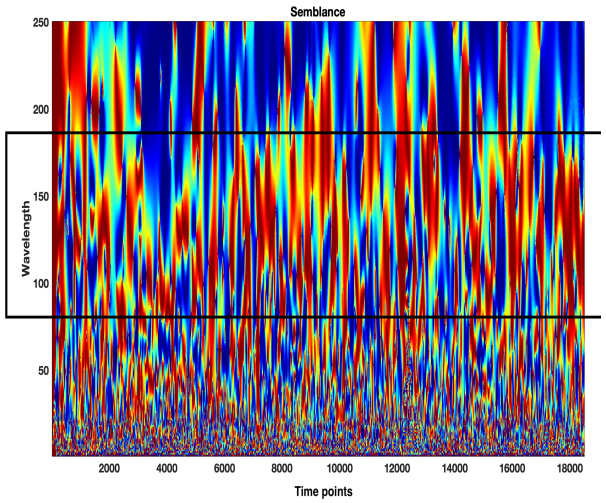
IV. Discussion

Recent progress in computer vision has led to the development of many excellent tools to analyze movement from the video (e.g., [11-12]). In the present paper we used Deeplabcut [16], which allows the tracking of user-defined body parts, to estimate movement from 2D videos and compared it with data extracted from wearables. Cross-correlation analysis showed comparable results among all time series, which suggests that computer vision might be a good alternative to the sometimes-expensive wearable systems.

It is worth emphasizing that the DLC showed remarkable accuracy in both tasks. This was achieved by labelling only 200 frames (~1% of the total number of frames in a video), which highlights the robustness of the method. Additionally, DLC is easy to set up, use and less likely to be subjective (see Manual in [17]), whereas methods such as manual coding require extensive training in the coding scheme to achieve high inter-rater reliability and time-consuming coding by multiple coders. Additionally, the DLC, as well as IMUs, are more precise in finding subtle changes in comparison to manual annotation, so they may offer a solution for identifying individual movements and estimating their onsets and offsets even for the shortest ones.

We tested the feasibility of DLC in two different infant-parent interactive situations: a structured play task with body movement partially restricted by seating (Task 1) and an unstructured free-play situation (Task 2). We have found a higher error in the free-play Task 2, which could be related to the fact that infants were moving freely, without constrains of the chair. However, even in this more naturalistic and unconstrained set-up the error was sufficiently low given the resolution of the video. Altogether, the DLC was robustly measuring infant leg movement across both tasks, although further analyses are needed to test it across a wider range of situational contexts, e.g., outdoor spaces.

Figure 3 Example of the semblance plot for the time series example shown in Figure 2. At higher frequencies (black box) the time series show strong correlations (red) which contrast with the variability showed at shorter frequencies. Anticorrelations are represented by a blue color. These results are consistent for all children.



When comparing the leg movements detected by the DLC and the IMUs, we found statistically significant, moderate correlations between the time series. Control analyses for randomized time series did not show any significant correlations. Although this suggests that the DLC and IMUs may differently measure the magnitude of movement, we consider these results promising. We discuss several reasons for finding only moderate correlations between different methods. First, video and sensor data were not automatically synchronized. We asked parents to clap at the beginning of each task, and we use cross-recurrence lag analysis to find the delay between both time series. However, although it worked sufficiently well (see Figure 2 for an example), it does not guarantee that both time series were perfectly aligned, which could introduce some discrepancies between time series. In fact, additional analyses showed larger correlations found at slightly different lags suggesting a lack of full synchrony among time series. Second, when missing data was present either in the DLC or the IMU data, a linear interpolation was performed to fill the gaps in every time series. If those missing data happen in both time series at once, some movement instances might not be present in any of the time series. As a result, the time series may look more dissimilar than they really are. Finally, the error in pixels that the DLC returns could also have influenced the accuracy of the estimation of movement. Larger errors would lead to higher variability in the estimation of movement and higher noise being present, therefore affecting our DLC-IMUs correlations (see Figure 3). However, we believe that despite these limitations, these results are very promising. Some of these issues could be solved with a higher number of labelled frames or by manually refining labels that were wrongly estimated to train the algorithm again [16]. Another route to improvement would be building a system where video and sensor data acquisitions are synchronised. Thus, further analyses are needed to systematically tackle this issue and test its effects on the final movement data.

However, there are also some aspects of the estimation of movement that cannot be easily assessed in automatic way using neither IMUs nor the DLC, and in these cases manual annotation could still provide the most reliable information. For example, episodes during physical contact between the infant and the caregiver are challenging to analyze automatically. For example, when the infant is being held, the data is affected both by infant-generated movements and passive movements of the infant induced by the caregiver (see [21] for an attempt to solve this problem using IMUs). Similarly, when the caregiver is moving the infant's limbs during play, it is difficult to disentangle between spontaneous and induced movements.

Our results should be interpreted with some caution, as we note several limitations. First, the number of participants in our preliminary study was low, which suggests a need for replication. Second, the infants included in the study were young (aged 4-5 months) and not yet very mobile, so further validation is needed in additional tasks and with older infants who have a larger motor repertoire. Third, the DLC requires strong computational power to run (i.e., GPUs), which not everybody may have access to, so it is not readily accessible for the use in low-resource settings. Fourth, we assessed the effectiveness of only one computer vision algorithm and additional computer vision algorithms (e.g., Open Pose) are needed to further validate the data. Finally, the DLC was applied on 2D video data while IMUs gather data in 3D space. This made the detection of orientation changes more challenging, and it is likely why DLC had better agreement with acceleration-based data than with quaternions. However, with prior knowledge of the calibration

parameters of the cameras, videos from several camera angles can also be combined to obtain a 3D representation of the movement of the entire body ([22]). Thus, further studies should focus on cross-validating 3D video data with 3D accelerometer data to determine to which extend computer vision can provide accurate movement estimates.

CONCLUSIONS

With the emergence and rapid evolution of various motion capture technologies, such as wearables and computer vision methods, we can observe and detect very subtle motor behaviors of infants and adults alike. Precise and unobtrusive methods that continuously record multiple behaviors are necessary to understand the emergence of those behaviors. In this study we showed that both computer vision and wearable sensors provide comparable quantitative data on infants' movements. Therefore, movement-based video analysis lends itself as a cheap, reliable and unobtrusive alternative for measuring the dynamics of body movement of young infants in naturalistic social interactions.

ACKNOWLEDGMENT

We thank all participating infants and parents for their generous contribution. This work was funded thanks to the Polish National Science Centre grants no. 2019/32/C/HS6/00199 to DLP and 2018/30/E/HS6/00214 to PT. This research was also supported in part by PLGrid Infrastructure.

REFERENCES

[1] K. E. Adolph. (2020). Oh, behave! Presidential address. *Infancy, 25*, 347-392.

[2] M. Airaksinen, O. Räsänen, E. Ilén, T. Häyrinen, A. Kivi, V. Marchi, A. Gallen, S. Blom, A. Varhe, N. Kaartinen, L. Haataja, & S. Vanhatalo (2020). Automatic Posture and Movement Tracking of Infants with Wearable Movement Sensors. *Scientific Reports*, 10(1). https://doi.org/10.1038/s41598-019-56862-5

[3] B. A. Bruijns, S. Truelove, A. M. Johnson, J. Gilliland & P. Tucker. (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*, Vol. 17. https://doi.org/10.1186/s12966-020-0912-4.

[4] D. Roetenberg, H. Luinge, & P. Slycke. (2009). Xsens MVN: full 6DOF human motion tracking using miniature inertial sensors. Xsens Motion Technologies B, 1–7. Retrieved from http://www.xsens.com/images/stories/PDF/MVN_white_paper.pdf

[5] F. Wittmann, O. Lambercy, & R. Gassert. (2019). Magnetometer-Based Drift Correction During Rest inIMU Arm Motion Tracking. Sensors (Basel, Switzerland), 19(6), 1312. https://doi.org/10.3390/s19061312

[6] M. Airaksinen, O. Räsänen, E. Ilén, T. Häyrinen, A. Kivi, V. Marchi, A. Gallen, S. Blom, A. Varhe, N. Kaartinen, L. Haataja, & S. Vanhatalo (2020). Automatic Posture and Movement Tracking of Infants with Wearable Movement Sensors. Scientific Reports, 10(1). https://doi.org/10.1038/s41598-019-56862-5

[7] J. Zhou, S. Y. Schaefer, & B. A. Smith. (2019). Quantifying caregiver movement when measuring infant movement across a full day: A case report. Sensors (Switzerland), 19(13). https://doi.org/10.3390/s19132886

[8] A. Gesell. (1935). Cinemanalysis: A method of behavior study. Journal of Genetic Psychology, 47, 3–16.

[9] J. Marencakova, C. Price, T. Maly, F. Zahalka & C. Nester. (2019). How do novice and improver walkers move in their home environments? An open-sourced infant's gait video analysis. PLOS ONE, 14(6), e0218665. https://doi.org/10.1371/journal.pone.0218665

[10] L. B. Smith, C. Yu, H. Yoshida, & C. M.F. (2015). Contributions of Head-Mounted Cameras to Studying the Visual Environments of Infants and Young Children. Journal of Cognition and Development, 16(3), 407–419. https://doi.org/10.1080/15248372.2014.933430

[11] Z. Cao, G. Hidalgo, T. Simon, S. E. Wei, & Y. Sheikh. (2019). OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields. IEEE transactions on pattern analysis and machine intelligence, 43(1), 172-186. https://doi.org/10.1109/TPAMI.2019.2929257.

[12] A. Toshev, & C. Szegedy. (2014). Deeppose: Human pose estimation via deep neural networks. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 1653-1660). https://doi.org/10.1109/CVPR.2014.214.

[13] D. López Pérez, D., G. Leonardi, A. Niedźwiecka, A., Radkowska, J. Rączaszek-Leonardi, & P. Tomalski. (2017). Combining Recurrence Analysis and Automatic Movement Extraction from Video Recordings to Study Behavioral Coupling in Face-to-Face Parent-Child Interactions. Frontiers in Psychology, 8(December), 1–14. https://doi.org/10.3389/fpsyg.2017.02228

[14] K. K., Yeh, W. Y. Liu, A. M. K Wong, & R. Lein. (2020). Validity of general movement assessment based on clinical and home videos. Pediatric Physical Therapy, 32(1), 35–43. https://doi.org/10.1097/PEP.0000000000000664

[15] O. Ossmy, R. O. Gilmore, K. E. Adolph. (2020) AutoViDev: A Computer-Vision Framework to Enhance and Accelerate Research in Human Development. In: Arai K., Kapoor S. (eds) Advances in Computer Vision. CVC 2019. Advances in Intelligent Systems and Computing, vol 944. Springer, Cham. https://doi.org/10.1007/978-3-030-17798-0_14

[16] A. Mathis, P. Mamidanna, K. M. Cury, T. Abe, V. N. Murthy, M. W. Mathis, & M. Bethge. (2018). DeepLabCut: markerless pose estimation of user-defined body parts with deep learning. *Nature Neuroscience*, 21(9), 1281–1289. https://doi.org/10.1038/s41593-018-0209-y

[17] Nath, T., Mathis, A., Chen, A.C. *et al.* Using DeepLabCut for 3D markerless pose estimation across species and behaviors. *Nat Protoc* **14,** 2152–2176 (2019). https://doi.org/10.1038/s41596-019-0176-0

[18] Delaherche, E., Chetouani, M., Mahdhaoui, A., Saint-Georges, C., Viaux, S., & Cohen, D. (2012). Interpersonal synchrony: A survey of evaluation methods across disciplines. IEEE Transactions on Affective Computing, 3(3), 349-365.

[19] ELAN (Version 5.9) [Computer software]. (2020). Nijmegen: Max Planck Institute for Psycholinguistics, The Language Archive. Retrieved from https://archive.mpi.nl/tla/elan M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[20] Gordon Cooper (2021). Comparing Time Series using Semblance Analysis (https://www.mathworks.com/matlabcentral/fileexchange/18409-comparing-time-series-using-semblance-analysis), MATLAB Central File Exchange. Retrieved March 15, 2021.

[21] P. Patel, Y. Shi, F. Hajiaghajani, S. Biswas & M. H. Lee (2019). A novel two-body sensor system to study spontaneous movements in infants during caregiver physical contact. Infant behavior & development, 57, 101383. https://doi.org/10.1016/j.infbeh.2019.101383

[22] Karashchuk, P., Rupp, K. L., Dickinson, E. S., Sanders, E., Azim, E., Brunton, B. W., & Tuthill, J. C. (2020). Anipose: a toolkit for robust markerless 3D pose estimation. BioRxiv, 2020.05.26.117325. https://doi.org/10.1101/2020.05.26.117325