AI-Enabled Big-Data Developmental Science: Longitudinal Analysis of Infant Sensorimotor Development

Matej Hoffmann
Department of Cybernetics, Faculty of Electrical Engineering, Czech Technical University in Prague
Email: matej.hoffmann@fel.cvut.cz

Abstract

We investigate infant sensorimotor development longitudinally across three contexts: spontaneous behavior, reaching to tactile stimuli on the body, and reaching to visually presented objects. Our goals are to identify early signs of active exploration and emerging goal-directedness in spontaneous movement, and to compare how reaching toward somatosensory versus visual targets develops over time. Because our dataset comprises dense recordings across hours, days, weeks, and months, manual annotation is impractical. We therefore use AI-based pipelines to extract 3D infant pose from video and map the resulting motion data onto robotic platforms for embodied analysis. This framework supports big-data developmental science and the study of emerging sensorimotor self-representation.

1 Introduction

1.1 Body representations

For successful embodied interaction with the world, body models of some form are needed. On the neural side, beyond the somatotopic representations (the "homunculi") in the primary motor and somatosensory cortices of primates (Leyton and Sherrington, 1917; Penfield and Boldrey, 1937), a large number of other somatotopically organized areas have been found in the brain, but their somatotopy is more fractured and less apparent (e.g., (Seelke et al., 2012)). On the cognitive side, concepts like "superficial and postural schema", "body schema", "body image", "corporeal schema" (see (Ataria et al., 2021) for an overview) have been proposed. One characteristic common to all these representations is their multimodal nature: they dynamically integrate information from different sensory modalities (tactile, proprioceptive, vestibular, visual), not excluding motor information. Somatosensory (tactile and proprioceptive) information constitute an important subset, perhaps most intimately tied to the body itself.

1.2 Development of body representations

Evidence from both psychology and neuroscience suggests that fragments of a functional body schema have been laid out before birth (Fagard et al., 2018; Zoia et al., 2007). Newborns are sensitive to some intersensory synchrony related to their own bodies (e.g., (Filippetti et al., 2013)). Infants' sensitivity to cross-modal perception of their own body, such as seeing the legs subject to different temporal or spatial manipulations (e.g., (Rochat, 1998)), has been studied. The development of tactile spatial remapping—younger infants code touch anatomically while older infants remap the stimulus into "space"—has also been investigated (Ali et al., 2015; Rigato et al., 2014). Such studies and others generally focus on the development of multisensory perception of own body alone or in relation to bodies of others.

1.3 Limitations of state of knowledge

The literature on the nature of body representations is not clear. Concepts like body schema and body image are umbrella notions for a range of observed phenomena rather than a result of identification of specific mechanisms. Regarding development, de Klerk et al. (2021) provide an associative learning account, which, however, needs to be further specified to be testable. Similarly, the Sensorimotor Contingency Theory has been applied to development (Jacquey et al., 2019) but still only as a high-level framework. Corbetta (2021) hypothesized how goal-reaching develops through links of visual and proprioceptive-tactile-motor spaces, with intrinsic motivation. A detailed theory specifying the underlying mechanisms is lacking.

2 Dense sampling of sensorimotor development

Infant development is difficult to study because babies are difficult experimental subjects that cannot be instructed to perform specific tasks repeatedly. Insights from spontaneous infant behavior often draw on small datasets and simplified manual scoring of video recordings (e.g., (DiMercurio et al., 2018; Thomas et al., 2015)). Behavioral data are sometimes complemented by brain imaging (see (Azhari et al., 2020) for a review). However, experimental paradigms often involve preferential looking times or habituation and deliberately avoid movement to reduce motion artifacts in brain

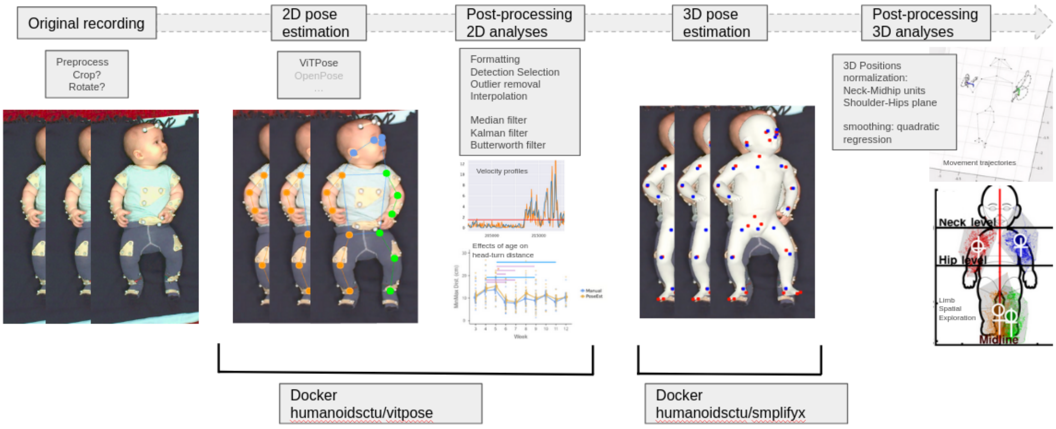

Fig. 1: Infant pose estimation pipeline. From (Gama, 2026).

scans. Daily spontaneous recordings of infants and their sharing will be needed (Adolph et al., 2017; Adolph and Robinson, 2011) to uncover developmental trajectories on multiple and nested time scales.

We longitudinally study infant sensorimotor development in three behavioral contexts: spontaneous behavior, reaching to tactile stimuli in the body, and reaching to visually presented objects. Using behavioral data collected in these scenarios, we aim to: (i) identify signatures of active exploration and emerging goal-directedness in spontaneous movements, and (ii) compare the developmental trajectories of reaching toward somatosensory and visual targets to understand the interaction between target localization and motor control. Our data set consists of dense longitudinal recordings that span hours, days, weeks, and months of infant behavior. Such a scale makes manual annotation infeasible. Therefore, we develop and evaluate AI-based analysis pipelines.

3 AI tools for developmental science

3.1 2D and 3D pose estimation from images and videos

Rapid progress in computer vision makes it possible to estimate the body configuration of humans from images or videos only. We use this technology to automatically estimate the pose of infants. Gama et al. (2025) compared the performance of seven state-of-the-art 2D human pose estimation networks on infant datasets, with ViTPose (Xu et al., 2022) performing the best overall. Khoury et al. (2022) performed 2D pose estimation (in image or pixel space) and then deployed SMPLify-X (Pavlakos et al., 2019) to get a 3D model. Our current pipeline is shown in Fig. 1.

3.2 Anonymizing infant videos through face swapping

The sharing of datasets, including raw video footage, is critical to making progress in psychological science (Adolph et al., 2017). For ethical reasons, this is not possible without anonymizing the videos. Simple methods like covering the face degrade the performance of downstream tasks like pose estimation. We harness modern large scale diffusion models to develop methods of anonymization that (a) change the identity significantly both for humans and face recognition engines, (b) keep gaze and facial expression unaltered, (c) remain visually consistent with the context without visual artifacts, (d) are temporally consistent for a video. See BLANKET (Hadera et al., 2025) and Fig. 2 for more details.

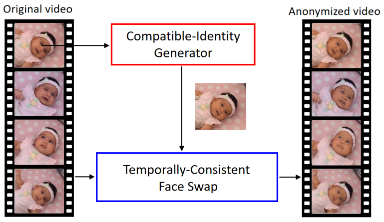

Fig. 2: Flowchart of the proposed video anonymization. First, an image replacing the original identity with a new compatible random identity from the first frame is created. The new face is generated by inpainting using Stable Diffusion (Rombach et al., 2021). Then, the new identity is swapped in every frame of the video. We propose to use FaceFusion (Facefusion, 2024), as it provides temporally consistent results while preserving original facial expressions. Figure and caption from (Hadera et al., 2025).

3.3 Motion retargeting: from infants to baby humanoids

The automatically extracted infant poses from the video recordings can then be mapped onto robotic platforms, allowing us to replay and analyze the motor, proprioceptive, visual, and tactile experiences of infants within robotic systems (López et al., 2026).

Fig. 3: Motion retargeting – infant to humanoids. The estimated infant pose or motion is retargeted onto a target robot or simulator platform. Here, the target is the iCub humanoid robot. The sensors of the robot can be used to emulate the first-person sensory experiences of the infant—vision and touch in this case. See (López et al., 2026) for details.

4 Beyond random motor babbling

Random motor or body babbling has been a typical developmental robotics method: a simulated or real robot randomly moves around its joints, thereby generating uncoordinated behavior and sensory input, and subsequently learns from this experience using reinforcement learning or another algorithm.

However, human infants hardly ever perform random movements. Instead, their movements are shaped by the biomechanical and neural constraints, and increasing voluntary control with age. Having the possibility of recreating the sensorimotor experience of actual infants at specific age and in specific contexts gives rise to much more ecologically valid multimodal datasets that can be used to test different learning algorithms (see also (Cusack et al., 2024)).

5 Conclusion

This approach establishes a framework for big-data developmental science, combining large-scale behavioral recordings, machine learning–based analysis, and embodied robotic modeling to advance our understanding of the emergence of sensorimotor self-representation.

Acknowledgment

This work was supported by the Czech Science Foundation (GA ČR), project no. 25-18113S.

References

Adolph, K., Gilmore, R. and Kennedy, J. (2017). Video data and documentation will improve psychological science. *Psychological Science Agenda*, 31(10).

Adolph, K. E. and Robinson, S. R. (2011). Sampling development. *Journal of Cognition and Development*, 12(4):411–423.

Ali, J. B., Spence, C. and Bremner, A. J. (2015). Human infants' ability to perceive touch in external space develops postnatally. *Current Biology*, 25(20):R978–R979.

Ataria, Y., Tanaka, S. and Gallagher, S. (2021). *Body schema and body image: New directions*. Oxford University Press.

Azhari, A., Truzzi, A., Neoh, M. J.-Y., Balagtas, J. P. M., Tan, H. H., Goh, P. P., Ang, X. A., Setoh, P., Rigo, P., Bornstein, M. H. et al. (2020). A decade of infant neuroimaging research: What have we learned and where are we going? *Infant Behavior and Development*, 58:101389.

Corbetta, D. (2021). Perception, action, and intrinsic motivation in infants' motor-skill development. *Current Directions in Psychological Science*, 30(5):418–424.

Cusack, R., Ranzato, M. and Charvet, C. J. (2024). Helpless infants are learning a foundation model. *Trends in Cognitive Sciences*, 28(8):726–738.

de Klerk, C. C., Filippetti, M. L. and Rigato, S. (2021). The development of body representations: An associative learning account. *Proceedings of the Royal Society B: Biological Sciences*, 288(1949).

DiMercurio, A., Connell, J. P., Clark, M. and Corbetta, D. (2018). A naturalistic observation of spontaneous touches to the body and environment in the first 2 months of life. *Frontiers in psychology*, 9:2613.

Facefusion (2024). Industry leading face manipulation platform. GitHub repository.

Fagard, J., Esseily, R., Jacquey, L., O'regan, K. and Somogyi, E. (2018). Fetal origin of sensorimotor behavior. *Frontiers in neurorobotics*, 12:23.

Filippetti, M. L., Johnson, M. H., Lloyd-Fox, S., Dragovic, D. and Farroni, T. (2013). Body perception in newborns. *Current biology*, 23(23):2413–2416.

Gama, F. (2026). *Learning body representations through self-touch: from babies to robots*. Dissertation, FaculTy of Electrical Engineering, Czech Technical University in Prague.

Gama, F., Mísař, M., Navara, L., T. Popescu, S. and Hoffmann, M. (2025). Automatic infant 2d pose estimation from videos: comparing seven deep neural network methods. *Behavior Research Methods*, 57(10):280.

Hadera, D., Cech, J., Purkrabek, M. and Hoffmann, M. (2025). Blanket: Anonymizing faces in infant video recordings. In *2025 IEEE International Conference on Development and Learning (ICDL)*, pp. 1–8. IEEE.

Jacquey, L., Baldassarre, G., Santucci, V. G. and O'regan, J. K. (2019). Sensorimotor contingencies as a key drive of development: from babies to robots. *Frontiers in neurorobotics*, 13:98.

Khoury, J., Popescu, S. T., Gama, F., Marcel, V. and Hoffmann, M. (2022). Self-touch and other spontaneous behavior patterns in early infancy. In *2022 IEEE International Conference on Development and Learning (ICDL)*, pp. 148–155. IEEE.

Leyton, A. S. and Sherrington, C. S. (1917). Observations on the excitable cortex of the chimpanzee, orang-utan, and gorilla. *Quarterly Journal of Experimental Physiology: Translation and Integration*, 11(2):135–222.

López, F. M., Kanazawa, H., Fiala, O., Balashov, Y., Marcel, V., Rustler, L., Lenz, M., Kim, D., Kuniyoshi, Y., Triesch, J. and Hoffmann, M. (2026). Simulating infant first-person sensorimotor experience via motion retargeting from babies to humanoids.

Pavlakos, G., Choutas, V., Ghorbani, N., Bolkart, T., Osman, A. A., Tzionas, D. and Black, M. J. (2019). Expressive body capture: 3d hands, face, and body from a single image. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10975–10985.

Penfield, W. and Boldrey, E. (1937). Somatic motor and sensory representation in the cerebral cortex of man as studied by electrical stimulation. *Brain*, 60(4):389–443.

Rigato, S., Ali, J. B., Van Velzen, J. and Bremner, A. J. (2014). The neural basis of somatosensory remapping develops in human infancy. *Current Biology*, 24(11):1222–1226.

Rochat, P. (1998). Self-perception and action in infancy. *Experimental brain research*, 123(1):102–109.

Rombach, R., Blattmann, A., Lorenz, D., Esser, P. and Ommer, B. (2021). High-resolution image synthesis with latent diffusion models.

Seelke, A. M., Padberg, J. J., Disbrow, E., Purnell, S. M., Recanzone, G. and Krubitzer, L. (2012). Topographic maps within brodmann's area 5 of macaque monkeys. *Cerebral Cortex*, 22(8):1834–1850.

Thomas, B. L., Karl, J. M. and Whishaw, I. Q. (2015). Independent development of the reach and the grasp in spontaneous self-touching by human infants in the first 6 months. *Frontiers in psychology*, 5:1526.

Xu, Y., Zhang, J., Zhang, Q. and Tao, D. (2022). Vitpose: Simple vision transformer baselines for human pose estimation. *Advances in neural information processing systems*, 35:38571–38584.

Zoia, S., Blason, L., D'Ottavio, G., Bulgheroni, M., Pezzetta, E., Scabar, A. and Castiello, U. (2007). Evidence of early development of action planning in the human foetus: a kinematic study. *Experimental brain research*, 176(2):217–226.