Modeling Social Interaction for Baby in Simulated Environment for Developmental Robotics

Md Ashaduzzaman Rubel Mondol **Aishwarya Pothula** **Deokgun Park**

Computer Science and Engineering
University of Texas at Arlington
Arlington, Texas USA
{mdashaduzzaman.mondol, aishwarya.pothula}@mavs.uta.edu, deokgun.park@uta.edu

Abstract

Task-specific AI agents are showing remarkable performance across different domains. But modeling generalized AI agents like human intelligence will require more than current datasets or only reward-based environments that don't include experiences that an infant gathers throughout its initial stages. In this paper, we present Simulated Environment for Developmental Robotics (SEDRo). It simulates the environments for a baby agent that a human baby experiences throughout the pre-born fetus stage to post-birth 12 months. SEDRo also includes a mother character to provide social interaction with the agent. To evaluate different developmental milestones of the agent, SEDRo incorporates some experiments from developmental psychology.

1 Introduction

To develop an Artificial Intelligence agent that can perform diverse tasks is still far from reality. Although task-specific AI agents are outperforming humans in many fields, there is none to perform all the tasks a single human can do. To build such an agent, combining different current task-specific models poses a great challenge as training time and dataset requirements grow exponentially as the number of tasks grows.

A human child is born with no experience of the world but over time learns to do many tasks that require complex sequential steps. The brain follows a universal algorithm to perform all the diverse tasks. To build a single agent to perform diversified tasks, we have to find out such kind of universal learning mechanism. Many researchers use a physical robot to study and test such a mechanism [6, 7, 11]. Building such an agent would require an environment that facilitates the necessary elements for longitudinal learning. Training that kind of agent in the real world is costly in terms of both time and money. Moreover, it may not be possible to reproduce all the scenarios. Computer simulated environments that can provide realistic experiences has become a common approach to diminish this problem.

The surrounding environment plays an important role in an infant's learning. Studies suggest that social interaction influences cognitive development. Right from birth, the infant's social interaction begins with its family members. The motherese or Infant Directed Speech (IDS) has a significant impact on the infants' cognitive development and language acquisition[15]. And mostly the mother plays a vital role.

Many researchers have been using computer-simulated environments to decode the different abilities of human baby like vision[5], motor skills[16] or in modeling Curiosity, Intrinsic Motivation [4, 17].

Figure 1: Screenshots from SEDRo environments. (a) Fetus environment. Almost dark space with no visual capability of the baby. (b) After birth house environment with a Mother character and some other toys. (c) Mother is feeding the baby. (d) Mother is showing a toy to the baby

There have been some simulated robot platforms also like iCub [19], Webot [12] including some works of Fetus environment[10, 13]. Simulated computer games like VizDoom[9], Obstacle Tower Challenge[8]. But these platforms don't provide the environments a newborn baby experiences throughout its first year including the social interaction with a mother or other family members.

In this paper, we present our ongoing work on building a Simulated Environment for Developmental Robotics (SEDRo) to facilitate the development of Generalized Intelligence of baby agent [14]. A mother character is added to interact with the baby agent. There will be different stages based on different ages of an infant after the birth as well as one stage that simulates the womb. Each stage will require the learning of previous stages to support incremental developments. Figure 1 shows the screenshots of the SEDRo.

2 Proposed Environment

SEDRo will simulate the minimal experience of a baby starting from the fetus stage to 12 months after birth. The key part of SEDRo will be a body of the baby agent and a surrounding environment for the agent to interact. Another important part will be a care-giving character, which will interact with the agent as part of social interaction. Also, there will be other interactive objects like furniture, toys, etc. The agent may interact with the surrounding objects in the room. A model of the agent can interact with the environment with the interface which is an extension of the OpenAI Gym API [2]. There will be four developmental stages with two different environments (Fetus and After-birth) to mimic different stages of the baby- 1) Fetus, 2) Immobile, 3) Crawling, and 4) Walking stage. Each one will provide a different experience for the agent and also unfold the new capabilities of the agent.

2.1 The Agent

The agent body has been developed with capabilities to crawl, walk, grasp objects, and follow the mother's attention. But these capabilities will unfold gradually in each subsequent stage. Initially, the maximum torque value for the joint motion is too small to support walking, but they will increase over time to enable it. The agent has 64 degree of freedom for the body part movements.

Vision　The agent has a binocular vision system with two eyes. Both eyes have a combined 3 degree of freedom; 1 for horizontal, 1 for vertical rotation of the eyeballs, and 1 to adjust focus. Each eye contains two cameras to simulate the central(8°) and peripheral(100°) vision of the human eye. There is one more camera on the head placed between the eyes of the agent to generate the combined visual image of both eyes. This is an optional camera provided for debugging purposes. Also, the depth of field effect has been applied to all the cameras for nearsighted focusing effects, since early infants cannot focus beyond arm's length. Figure 2 shows different visual inputs for the agent.

Tactile Sensitivity　As part of tactile sensitivity, the agent is equipped with touch sensors of varying density, since the perception of touch differs across the body [1]. There are a total of 2110 tactile sensors placed across the body. About half of the sensors are placed on the head. The sensors detect the touch, based on the collision detection mechanism. Each sensor generates a value of 1 when a 'touch' occurs or 0 otherwise. A sparse status vector is generated consisting of all sensor status and sent as part of observations.
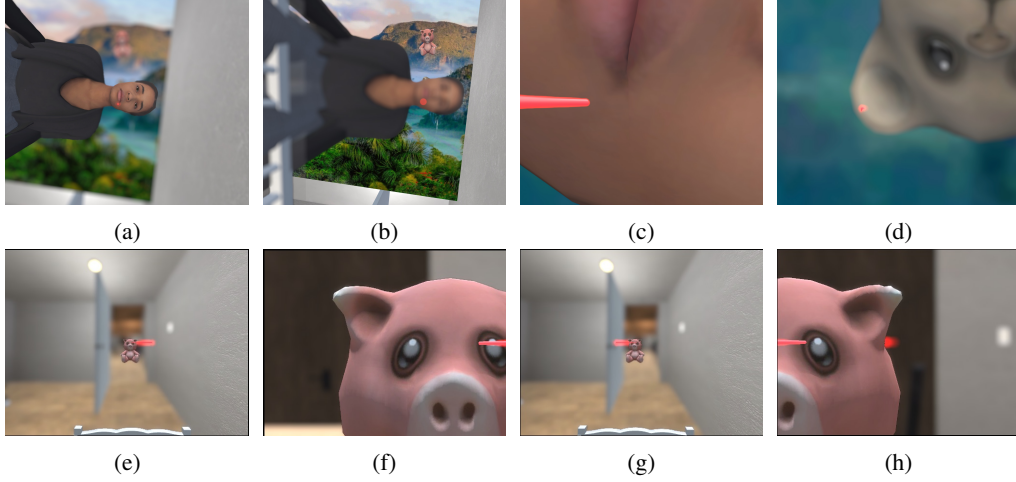
Figure 2: Baby Agent's vision. (a) Nearsighted peripheral vision. (b) Farsighted peripheral vision. (c) Central vision in-focus. (d) central vision out-of-focus. (e) Left peripheral vision. (f) Left central vision. (g) Right peripheral vision. (h) Right central vision

Proprioception   To learn the association of spatial locations and body parts movements, the baby will require its current joint positions along with visual information. In SEDRo, the current positions and rotations of all the body joints are given to the agent's observations. In total, 469 observations will be provided with continuous values ranging from -1 to 1. This vector will also include each joint's velocity and angular velocity to help understand the current body movements.

Interoception   The baby's stomach food level is also given with the observation vector as a body's internal sensitivity. It represents the current percentage of food available in the stomach. With time, the food level reduces. When the food level falls below a certain threshold representing the hunger, the baby needs to cry to get food. Whenever the baby cries, the mother will feed it and the stomach food level will increase.

2.2   Modeling the Motherese

We implemented a mother character for interacting with the baby. Both the baby and mother are inside a small house. The baby is placed in the crib or on the floor based on different situations and the baby's age.

Building the Mother   While facilitating the baby agent's intelligence development, one challenging part is building a mother that can interact with the baby. We tackle this challenge by limiting the experience up to the first year after birth because most interaction in this period does not require open-ended back-and-forth interaction. The mother has been programmed with some pre-programmed action capabilities to interact with the baby. We are building a library of mother's actions based on real-life mother-child interaction. Currently, we are manually building scenarios and we plan to analyze different video recordings from the houses with newborn babies. To make realistic behaviors, we will create the movements of the mother with pre-recorded motion captured(Mocap) animations based on those real footages. So, we will get a library with various responses of the mother for the same type of baby's actions.

Interaction with baby   As a concrete example of the social scenario, feeding the baby will be the mother's regular interaction. The mother will feed the baby at pre-scheduled times of the day. Also, the baby will cry when the stomach food level is below the threshold and the mother will begin feeding scenario. When feeding, it will move towards the baby and start the feeding. While walking and feeding, the mother can avoid different obstacles and also adjust body positions during the feeding animation based on the baby's current location.

Providing Infant Directed Speech (IDS) is another significant role played by mothers that helps with a child's development. For IDS in SEDRo, the mother character will be talking to the baby with
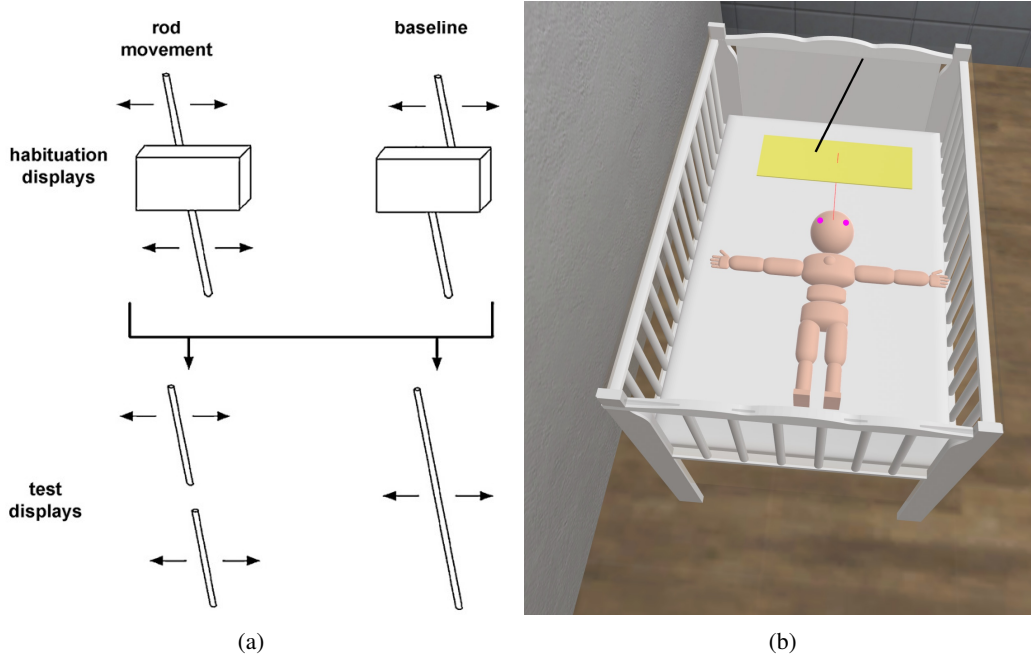
Figure 3: Evaluation experiments. (a) Paper rod experiment to evaluate unity perception[18]. (b) Paper rod experiment simulation in SEDRo

small words and physical expressions like nodding head while looking at the baby, moving arms. One limitation for implementing the mother's vocal is that sound can't be added directly as part of observation for the agent model yet. To overcome this limitation, in our initial version of SEDRo, we will use a one-hot encoded vector of length 26 to represent one English character at every time frame.

For joint attention, the mother will be holding different objects e.g. toys in front of the baby, looking at it, and describe its identity in small words. Also at a more developed stage, if the baby grabs or touches any object, the mother will describe it. Figure. 1c, 1d shows mother's interaction with baby.

3 Evaluation of Development

There have been many experiments to evaluate and track the progress of cognitive, motor skill, and visual developments of human babies in Developmental Psychologies [3]. Similarly, SEDRo will provide different experiments to evaluate the developmental progress of the agent. Figure 3a shows one such experiment where a moving rod occluded by a box is shown to the baby. Newborn babies under three months age think of it as two different rods but, older babies see it as a single rod[18]. This test is to evaluate the unity perception of babies. Similarly, more experiments will be provided in SEDRo from developmental psychologies.

4 Discussion

So far, we have presented our in-progress works. SEDRo is currently being implemented using the Unity 3D game engine. It will be improved further over time as we add more social interaction scenarios between the mother character and the agent.

In this version, we have simulated the motherese voice with character sequence observations. This can be further improved by adding the audio data directly to the observations. Representing voice as a text has a limitation. It's not possible to add variations in a speech this way, but generating prosodic voice requires variations in tone and length of the sounds. And a prosodic voice from motherese plays an important role in language development for the babies[15]. In the future version, we plan to include vocal audio data as observation for the agent model.

References

[1] R. Ackerley, I. Carlsson, H. Wester, H. Olausson, and H. B. Wasling. Touch perceptions across skin sites: differences between sensitivity, direction discrimination and pleasantness. *Frontiers in Behavioral Neuroscience*, 2 2014.

[2] G. Brockman, V. Cheung, L. Pettersson, J. Schneider, J. Schulman, J. Tang, and W. Zaremba. Openai gym. *arXiv preprint arXiv:1606.01540*, 2016.

[3] A. Cangelosi and M. Schlesinger. *Developmental robotics: From babies to robots*. MIT press, 2015.

[4] V. Fiore, F. Mannella, M. Mirolli, K. Gurney, and G. Baldassarre. Instrumental conditioning driven by apparently neutral stimuli: A model tested with a simulated robotic rat. 01 2008.

[5] S. Fuke, M. Ogino, and M. Asada. Body image constructed from motor and tactile images with visual information. *I. J. Humanoid Robotics*, 4:347–364, 06 2007.

[6] D. Gouaillier, V. Hugel, P. Blazevic, C. Kilner, J. Monceaux, P. Lafourcade, B. Marnier, J. Serre, and B. Maisonnier. Mechatronic design of nao humanoid. In *2009 IEEE International Conference on Robotics and Automation*, pages 769–774. IEEE, 2009.

[7] B. Johansson, T. A. Tjøstheim, and C. Balkenius. Epi: An open humanoid platform for developmental robotics. *International Journal of Advanced Robotic Systems*, 17(2):1729881420911498, 2020.

[8] A. Juliani, A. Khalifa, V.-P. Berges, J. Harper, E. Teng, H. Henry, A. Crespi, J. Togelius, and D. Lange. Obstacle tower: A generalization challenge in vision, control, and planning. pages 2684–2691, 08 2019.

[9] M. Kempka, M. Wydmuch, G. Runc, J. Toczek, and W. Jaśkowski. Vizdoom: A doom-based ai research platform for visual reinforcement learning. pages 1–8, 09 2016.

[10] Y. Kuniyoshi and S. Sangawa. Early motor development from partially ordered neural-body dynamics: Experiments with a cortico-spinal-musculo-skeletal model. *Biological cybernetics*, 95:589–605, 01 2006.

[11] G. Metta, G. Sandini, D. Vernon, L. Natale, and F. Nori. The icub humanoid robot: an open platform for research in embodied cognition. In *Proceedings of the 8th workshop on performance metrics for intelligent systems*, pages 50–56, 2008.

[12] O. Michel. Cyberbotics ltd. webots™: Professional mobile robot simulation. *International Journal of Advanced Robotic Systems*, 1(1):5, 2004.

[13] H. Mori and Y. Kuniyoshi. A human fetus development simulation: Self-organization of behaviors through tactile sensation. pages 82 – 87, 09 2010.

[14] A. Pothula, M. A. R. Mondol, S. Narasimhan, S. M. Islam, and D. Park. Sedro: A simulated environment for developmental robotics, 2020.

[15] C. Saint-Georges, M. Chetouani, R. Cassel, F. Apicella, A. Mahdhaoui, F. Muratori, M.-C. Laznik, and D. Cohen. Motherese in interaction: At the cross-road of emotion and cognition? (a systematic review). *PLoS One*, 10 2013.

[16] P. Savastano and S. Nolfi. Incremental learning in a 14 dof simulated icub robot: Modeling infant reach/grasp development. In T. J. Prescott, N. F. Lepora, A. Mura, and P. F. M. J. Verschure, editors, *Biomimetic and Biohybrid Systems*, pages 250–261, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[17] S. Singh, A. Barto, and N. Chentanez. Intrinsically motivated reinforcement learning. 01 2004.

[18] A. Slater, V. Morison, M. Somers, A. Mattock, E. Brown, and D. Taylor. Newborn and older infants' perception of partly occluded objects. *Infant Behavior and Development*, 13(1):33 – 49, 1990.

[19] V. Tikhanoff, P. Fitzpatrick, F. Nori, L. Natale, G. Metta, and A. Cangelosi. The icub humanoid robot simulator. *IROS Workshop on Robot Simulators*, 04 2012.