International Journal of Humanoid Robotics
Vol. 21, No. 1 (2024) 2350024 (8 pages)

DOI: 10.1142/S021984362350024X

Interplay Between Computational Neuroscience and Humanoid Robotics: Past and Future

Mitsuo Kawato

ATR Brain Information Communication Research Group
Hikaridai 2-2 Seika-cho, Soraku-gun, Kyoto, 619-0288, Japan
kawato@atr.jp

Accepted 31 October 2023
Published 30 November 2023

The interplay between computational neuroscience and humanoid robotics, both historically and looking forward, is presented, with a particular focus on two significant projects: the ERATO Kawato Dynamic Brain and the ICORP Computational Brain. Discussion revolves around pivotal brain regions, namely the cerebellum, basal ganglia, and cerebrum, and the associated computational models.

Keywords: Computational neuroscience; ERATO Kawato Dynamic Brain Project; ICORP Computational Brain Project; cerebellum; basal ganglia; cerebral cortex; internal model; reinforcement learning; hierarchical modular reinforcement learning.

Having specialized in computational neuroscience, a field with various possible definitions, my perspective is that computational neuroscience endeavors to understand the brain's functionality so completely that we could engineer a computational system or a machine to solve computational problems the same way the brain does. If we remove from this definition the latter limitation, this definition transitions into the realm of artificial intelligence and robotics. While neuroscience has made significant strides in understanding the brain's hardware, deciphering its computational principles remains a daunting task. Hence, replicating the brain's computational principles is immensely challenging. It is therefore natural that computational neuroscience and robotics, as closely related fields, should influence each other and develop together. This intersection is the crux of my contribution to the 20th-anniversary special issue of the International Journal of Humanoid Robotics. Regarding my credentials, I was a research director of the JST-funded ERATO Kawato Dynamic Brain Project (1996–2001; 3 group leaders; Stefan Schaal, Hiroshi Imamizu, Kenji Doya) as well as ICORP Computational Brain project (2004–2009; US co-research director; Chris Atkeson; 3 group leaders; Gordon Cheng, Jun

Morimoto, Jun Nakanishi). In the two projects, I truly enjoyed international collaborations between neuroscientists and roboticists. In this contribution, I trace the computational neuroscience — humanoid robotics interplay evolution from past to potential future, correlating three crucial brain regions, the cerebellum, basal ganglia, and cerebral cortex, with their respective computational principles. These principles address internal model learning, reward-based reinforcement learning, and the emergence of consciousness through a modular hierarchy.

Kenji Doya proposed distinct roles for different brain regions: the cerebellum for supervised learning, the basal ganglia for reinforcement learning, and the cerebral cortex for unsupervised Hebbian learning.[1] The fact that action potentials do not backpropagate from the cell body to dendritic synapse terminals in Purkinje cells, the only output cells in the cerebellar cortex, and the fact that synaptic plasticity requires an increase in calcium concentration associated with climbing fiber inputs, point toward supervised learning rather than Hebbian learning in the cerebellum.[2] In the 1980s, the author and collaborators proposed a feedback error learning model.[3] In this model, climbing fiber inputs transmit feedback motor commands to Purkinje cells, facilitating the acquisition of the inverse model in the cerebellar cortex through synaptic plasticity. Leveraging this model, the author collaborated with Hiroyuki Miyamoto, Masazumi Katayama, and Hiroaki Gomi at the Osaka University and ATR to achieve learning control of industrial manipulators, artificial-muscle Rubbertuator SoftArm, and SARCOS dexterous arms.[4] A significant discovery was the human ability to learn and control the arm's geometric impedance.[5] This provided the neuroscientific basis for successful learning control of artificial muscles that control mechanical compliance, e.g., artificial-muscle Rubbertuator SoftArm. In the ERATO project, a humanoid robot named DB was developed in collaboration with SARCOS Research Company. In the ICORP project, a humanoid robot named CB-i was developed again in collaboration with SARCOS. Stefan Schaal, Sethu Vijyayakumar, Jun Nakanishi and others demonstrated successful machine learning of internal models, and verified the stability and convergence of feedback error learning.[6] Neuroscientific evidence supporting the feedback error learning model includes electrical recordings of monkey Purkinje cell firing[7] and human fMRI experiments.[8]

In a recent study of the cerebellum, various neural elements within the cerebellum were found to represent various variables involved in reinforcement learning. Our recent studies[9–11] integrate Doya's theory with findings on reward-associated functions of the cerebellum. Tsutsumi and colleagues[12] Ca-imaged thousands of Purkinje cells in association-learning mice, which discriminate between high- and low-pitched sounds and make go/no-go decisions to lick or not lick to obtain liquid rewards. By deploying and combining sophisticated analytical methods, we estimated the precise timing of climbing fiber inputs to Purkinje cells. We discerned a wide variety of firing patterns of thousands of Purkinje cells over a wide spatial extent in the cerebellum, resulting in only four functional spatial modules. Regarding much-debated cerebellar hypotheses, one module related to the timing control hypothesis of Llinas,[13] while two modules connected to the cerebellar learning theory of Marr-Albus-Ito.[11,14–17]

The ERATO Dynamic Brain project was pivotal in fusing computational neuroscience concepts with robotic machine learning algorithms. However, designing algorithms that can learn in a reasonable time for systems with huge degrees of freedom remains a challenge.[18] In most of the ERATO and ICORP robotic projects, researchers manually reduced dimensionality and selected features, a common limitation of humanoid robotics. This status quo is unsatisfactory both for robotics and for neuroscience. Notably, Huu Hoang and colleagues[9] indicate that reaction–diffusion equations produced by electrical synapses and positive and negative neural feedback in the system consisting of the cerebellum and inferior olive nucleus self-organize functional modules as emergent patterns, and achieve overwhelming dimensional reduction and feature selection. We aim to explore this potential in both neuroscience and robotics.

The basal ganglia are widely believed to be the central hub for reinforcement learning. Numerous studies such as those mentioned below, support this theory. In macaques, dopamine cells embody reward prediction errors.[19] In humans, different regions of the basal ganglia represent different variables in reinforcement learning, as demonstrated in fMRI experiments.[20–22] As a humanoid robot demonstration, Morimoto and Doya[23] managed robot's stand-up from lying down position with hierarchical reinforcement learning using simple feedback: the robot was rewarded for raising its head and punished for failing to do so. Three critical components of this study are worth noting: the robot's internal model, the via-point target postures in the upper hierarchy, and offline reinforcement learning utilizing the internal model. The hierarchical structure enabled drastic dimension reduction in the upper hierarchy, allowing successful learning through approximately 1,000 offline and 100 live trials. Although the target posture was learned, the hierarchical structure, including the representation "target posture," was pre-configured by the researchers.

The ERATO & ICORP project's novelty stemmed from merging reinforcement learning in humanoid robots with imitation learning, i.e., learning by imitation or teaching by demonstration. Francesca Gandolfo, a graduate student from the Emilio Bizzi Lab at MIT, collaborated with ATR in 1993, initiating research on imitation learning. Imitation learning of kendama,[24] tennis serve,[25] gait,[26] air field hockey,[27] etc. was successfully demonstrated with SARCOS dexterous arm, DB-chan, and DB, etc.[28] The air hockey learning paper,[27] published in the first volume of the International Journal of Humanoid Robotics is original in its integration of imitation learning and hierarchical modular reinforcement learning. First, behavioral data from human–human air field hockey games are represented as hierarchical primitives and stored in a database. Next, control of the humanoid robot DB was also prepared hierarchically. Furthermore, how control refers to the human data is learned by an algorithm similar to the action-state value function: Q learning. Such foundational works underscore the importance of discerning how the brain autonomously acquires modules and hierarchy vital for robot efficient reinforcement learning and implementation in robotics. The author, in collaboration with Daniel Wolpert, Masahiko Haruno, Kenji Doya, Norikazu Sugimoto, and others, proposed learning control algorithms that autonomously acquire modularity and hierarchy. These algorithms segment the state space grounded in the predictive performance of multiple forward models, and forward models are paired with inverse models and reinforcement learning controllers. The proposed learning control algorithms include MOSAIC,[29,30] RL-MOSAIC,[31] and Hierarchical RL-MOSAIC,[32,33] for which forward models are paired with inverse models[29,30] and reinforcement learning controllers.[31–33] With this framework, humanoid robot CB-*i* can learn,[33] and two agents can collaborate through common symbols.[34] Algorithms like MOSAIC can autonomously develop multiple modules based on sensory-motor information that exists on the input-output interface sides for the brain, enabling hierarchical reinforcement learning and comprehension of higher cognitive functions.[35] Recent studies[9–11] demonstrate that the cerebrum, cerebellum, and basal ganglia collectively facilitate modular reinforcement learning. In this system, one functional module of Purkinje cells can be trained with supervised learning to acquire one actor of modular reinforcement learning.[10] The capacity of the cerebellum and inferior olive nucleus to self-organize functional modules, rooted in reaction-diffusion dynamics of the feedback neural network, is essential. This is particularly crucial for creating modules related to higher cognitive functions, which are challenging to base solely on sensorimotor information like MOSAIC.

Present-day AI largely operates as brute force intelligence, powered by vast training samples and high-speed computing. This approach is unsuitable for humanoid robot sensory-motor learning and real-time action selection.[18] The brain, capable of learning from a small number of samples, offers insights we must explore. A profound mystery in neuroscience is the origin of consciousness in the brain, despite its being a physicochemical system. Our computational model, the Cognitive Reality Monitoring Network (CRMN)[36] suggests that computational understanding of consciousness in neuroscience is intimately connected to challenges faced in humanoid robotics. David Marr proposed that three levels of brain research are essential: first, computational theory, second, representation and algorithms, and third, hardware.[37] This framework profoundly influenced subsequent computational neuroscience research. Yet, most consciousness theories focus primarily on hardware and some aspects of information representation, largely overlooking Marr's first and second levels. CRMN delineates the computational theory of consciousness. At its core, the computational objective is to control a nonlinear body with huge-degrees of freedom in a dynamic world, seeking to select a quasi-optimal action in a reasonable time. The input of the calculation is information from various sensory organs, and the final output is commands to motor organs. The principle that makes this computation possible is dividing the dynamics of the body and the environment to derive a quasi-optimal solution for subproblems.

The second level (algorithms and representations) is tightly constrained by the brain's hardware, at the third level. The brain's architecture is modular, organized by sensory, e.g., visual, auditory, somatosensory, and motor modalities, e.g., limbs and eyes. Topographic maps of visual, auditory, motor, and somatosensory areas are finer levels of this modularity. Within each sensory-motor module is a hierarchy that spans primary sensory/motor areas to the prefrontal cortex. Information representation transitions from being closely aligned with receptor and effector data near the input/output areas to becoming more abstract and dimensionality reduced closer to the prefrontal cortex. Regarding algorithms, feedback neural connections (from higher to lower hierarchies) between cerebral sensory areas provide generative models, e.g., forward optics model.[38] Conversely, feedforward neural connections (from lower to higher hierarchies) between sensory areas implement inference models, e.g., the inverse optics model.[38] For motor control areas, feedforward connections from higher to lower areas provide inference models, e.g., the inverse model of motor apparatus,[3] and feedback connections from lower to higher areas provide generative models, e.g., forward models of the motor apparatus.[39] At a certain hierarchy within a module, these generative and inference models function as inverse counterparts, preserving signals even when circulated through both. Within a module, these models cooperate to establish a coherent representation swiftly from lower to higher levels. For action selection and learning in specific environments, an agent selects the most appropriate module and hierarchy, and aligns it with the highest-level representation. This selection facilitates efficient learning from a small number of trials and samples by focusing on specific modules, hierarchies, and representations. This is central to effectively reducing the number of degrees of freedom of the learning machine, i.e., the brain. The selection of modules and hierarchies hinges on both the coherence of the representation between the generative and inference models and the accuracy of the value function estimation.[36] Parallel computations are executed across each module and hierarchy, with the highest level (prefrontal cortex) computing the predictive prior distribution. These algorithms, representations, and hardware together enable swift action selection and efficient learning. Within this framework, the highest-level representation metacognitively oversees the representation at lower levels. Consciousness emerges when the prior distribution at the top level aligns well with the posterior distribution, and a distinct peak is observed in the distribution.

Three decades ago, the interaction between computational neuroscience and humanoid robotics took root at ATR. The enthusiasm sparked by the inaugural issue of the International Journal of Humanoid Robotics two decades ago remains vibrant today. As we look ahead, the self-organization of cerebellar functional modules, hierarchical modular reinforcement learning centered on the basal ganglia, and computational theories of cerebral cortical networks generating consciousness are posed to further enrich the interplay between computational neuroscience and humanoid robotics.

Acknowledgments

The author would like to thank Huu Hoang and Aurelio Cortese for discussions on the topics covered in this paper. This work was supported by the Ikegaya Brain-AI

Hybrid ERATO grant (JPMJER1801) from the Japan Science and Technology Agency.

ORCID

Mitsuo Kawato https://orcid.org/0000-0001-8185-1197

References

1. K. Doya, Complementary roles of basal ganglia and cerebellum in learning and motor control, *Curr. Opin. Neurobiol.* **10**(6) (2000) 732–739.
2. M. Kawato, S. Kuroda and N. Schweighofer, Cerebellar supervised learning revisited: biophysical modeling and degrees-of-freedom control, *Curr. Opin. Neurobiol.* **21**(5) (2011) 791–800.
3. M. Kawato, K. Furukawa and R. Suzuki, A hierarchical neural-network model for control and learning of voluntary movement, *Biol. Cybern.* **57**(3) (1987) 169–185.
4. M. Kawato, From 'Understanding the brain by creating the brain' towards manipulative neuroscience, *Phil. Trans. R. Soc. B* **363** (2008) 2201–2214.
5. E. Burdet, R. Osu, D. W. Franklin, T. E. Milner and M. Kawato, The central nervous system stabilizes unstable dynamics by learning optimal impedance, *Nature* **414**(6862) (2001) 446–449.
6. J. Nakanishi and S. Schaal, Feedback error learning and nonlinear adaptive control, *Neural Netw.* **17**(10) (2004) 1453–1465.
7. M. Shidara, K. Kawano, H. Gomi and M. Kawato, Inverse-dynamics model eye movement control by Purkinje cells in the cerebellum, *Nature* **365** (1993) 50–52.
8. H. Imamizu, S. Miyauchi, T. Tamada, Y. Sasaki, R. Takino, B. Pützet, T. Yoshioka and M. Kawato, Human cerebellar activity reflecting an acquired internal model of a new tool, *Nature* **403** (2000) 192–195.
9. H. Hoang, S. Tsutsumi, M. Matsuzaki, M. Kano, M. Kawato, K. Kitamura and K. Toyama, Dynamic organization of cerebellar climbing fiber response and synchrony in multiple functional modules reduces dimensions for reinforcement learning, *bioRxiv* **12.05** (2022) 518634.
10. H. Hoang, S. Tsutsumi, M. Masanori, M. Kano, K. Toyama, K. Kitamura and M. Kawato, Negative reward-prediction errors of climbing fiber inputs for cerebellar reinforcement learning algorithm, *bioRxiv* **03.13** (2023) 532374.
11. H. Hoang, S. Tsutsumi, M. Matsuzaki, M. Kano, M. Kawato, K. Kitamura and K. Toyama, Dynamic organization of cerebellar climbing fiber response and synchrony in multiple functional components reduces dimensions for reinforcement learning, *eLife* **12** (2023) e86340.
12. S. Tsutsumi, N. Hidaka, Y. Isomura, M. Matsuzaki, K. Sakimura, M. Kano and K. Kitamura, Modular organization of cerebellar climbing fiber inputs during goal-directed behavior, *eLife* **8** (2019) e47021.
13. R. R. Llinás, The olivo-cerebellar system: A key to understanding the functional significance of intrinsic oscillatory brain properties, *Front. Neural Circuits* **7** (2014) 96.
14. J. S. Albus, A theory of cerebellar function, *Math. Biosci.* **10**(1–2) (1971) 25–61.
15. M. Ito, Neurophysiological aspects of the cerebellar motor control system, *Int. J. Neurol.* **7**(2) (1970) 162–176.
16. M. Ito, The molecular organization of cerebellar long-term depression, *Nat. Rev. Neurosci.* **3** (2002) 896–902.

17. D. Marr, A theory of cerebellar cortex, *J. Physiol.* **202** (1969) 437–470.
18. C. G. Atkeson *et al.*, What happened at the DARPA robotics challenge finals, in *The DARPA Robotics Challenge Finals: Humanoid Robots To The Rescue*, Springer Tracts in Advanced Robotics, Vol. 121 (Springer, Cham, 2018), pp. 667–684.
19. W. Schultz, P. Dayan and P. R. Montague, A neural substrate of prediction and reward, *Science* **275**(5306) (1997) 1593–1599.
20. M. Haruno and M. Kawato, Different neural correlates of reward expectation and reward expectation error in the putamen and caudate nucleus during stimulus-action-reward association learning, *J. Neurophysiol.* **95** (2006) 948–959.
21. M. Haruno and M. Kawato, Heterarchical reinforcement-learning model for integration of multiple cortico-striatal loops: fMRI examination in stimulus-action-reward association learning, *Neural Netw.* **19**(8) (2006) 1242–1254.
22. M. Haruno, T. Kuroda, K. Doya, K. Toyama, M. Kimura, K. Samejima, H. Imamizu and M. Kawato, A neural correlate of reward-based behavioral learning in caudate nucleus: A functional magnetic resonance imaging study of a stochastic decision task. *J. Neurosci.* **24**(7) (2004) 1660–1665.
23. J. Morimoto and K. Doya, Acquisition of stand-up behavior by a real robot using hierarchical reinforcement learning, *Rob. Auton. Syst.* **36**(1) (2001) 37–51.
24. H. Miyamoto, S. Schaal, F. Gandolfo, H. Gomi, Y. Koike, R. Osu, E. Nakano, Y. Wada and M. Kawato, A Kendama learning robot based on bi-directional theory, *Neural Netw.* **9**(8) (1996) 1281–1302.
25. H. Miyamoto and M. Kawato, A tennis serve and upswing learning robot based on bi-directional theory, *Neural Netw.* **11**(7–8) (1998) 1331–1344.
26. J. Nakanishi, J. Morimoto, G. Endo, G. Cheng, S. Schaal and M. Kawato, Learning from demonstration and adaptation of biped locomotion, *Rob. Auton. Syst.* **47**(2–3) (2004) 79–91.
27. D. C. Bentivegna, C. G. Atkeson, A. Ude and G. Cheng, Learning to act from observation and practice, *Int. J. Humanoid Robotics* **1** (2004) 585–611.
28. C. G. Atkeson, J. Hale, F. Pollick, M. Riley, S. Kotosaka, S. Schaal, T. Shibata, G. Tevatia, S. Vijayakumar, A. Ude and M. Kawato, Using humanoid robots to study human behavior, *IEEE Intell. Syst.* **15**(4) (2000) 46–56.
29. M. Haruno, D. M. Wolpert and M. Kawato, Mosaic model for sensorimotor learning and control, *Neural Comput.* **13**(10) (2001) 2201–2220.
30. D. M. Wolpert and M. Kawato, Multiple paired forward and inverse models for motor control, *Neural Netw.* **11**(7–8) (1998) 1317–1329.
31. K. Doya, K. Samejima, K. Katagiri and M. Kawato, Multiple model-based reinforcement learning, *Neural Comput.* **14**(6) (2002) 1347–1369.
32. N. Sugimoto, M. Haruno, K. Doya and M. Kawato, MOSAIC for multiple-reward environments, *Neural Comput.* **24**(3) (2012) 577–606.
33. N. Sugimoto, J. Morimoto, S. H. Hyon and M. Kawato, The eMOSAIC model for humanoid robot control, *Neural Netw.* **29–30** (2012) 8–19.
34. K. Samejima, K. Katagiri, K. Doya and M. Kawato, Symbolization and imitation learning of motion sequence using competitive modules, *Electron. Comm. Jpn. Pt. III.* **89** (2006) 42–53.
35. D. M. Wolpert, K. Doya and M. Kawato, A unifying computational framework for motor control and social interaction, *Philos. Trans. R. Soc. London Ser. B Biol. Sci.* **358**(1431) (2003) 593–602.
36. M. Kawato and A. Cortese, From internal models toward metacognitive AI, *Biol. Cybern.* **115** (2021) 415–430.

37. D. Marr, *Vision: A Computational Investigation into the Human Representation and Processing of Visual Information* (W.H. Freeman, San Francisco, CA, 1982).
38. M. Kawato, H. Hayakawa and T. Inui, A forward-inverse optics model of reciprocal connections between visual cortical areas, *Network* **4**(4) (1993) 415–422.
39. M. Kawato, T. Kuroda, H. Imamizu, E. Nakano, S. Miyauchi and T. Yoshioka, Internal forward models in the cerebellum: fMRI study on grip force and load force coupling, *Prog. Brain Res.* **142** (2003) 171–188.

Mitsuo Kawato received B.S. degree in physics from the Tokyo University in 1976, with M.E. and Ph.D. degrees in biophysical engineering from Osaka University in 1978 and 1981, respectively. In 1988, he moved to Advanced Telecommunications Research Institute International (ATR), and in 2010, he was appointed Director of the ATR Brain Information Communication Research Laboratory Group.

He has been working in computational neuroscience and neural network modeling, with publishing 300 papers, reviews and books. For the last 15 years, he has been involved in the research and development of neuroscience applications for the diagnosis and treatment of psychiatric disorders and has been developing new medical technology combining AI and brain science. He is also CEO of start-up XNef aiming at the implementation of these developments.