A system theory for the brain-like computer

Hiroshi Tsujino
Honda Research Institute Japan Co., Ltd.
8-1 Honcho, Wako-shi, Saitama, 351-0188, Japan.
tsujino@jp.honda-ri.com

Abstract --- Matsumoto [1] proposed two general principles, an output-driven operation and a memory-based architecture, giving the brain its auto-designing capacity and consequently its potential to acquire algorithms. We introduce these two principles while suggesting a design for this embodiment by introducing computational models, suggest brain-like computer characteristics and recommend future research directions.

1. Introduction

Living organisms exchange materials and energy, as well as information by means of its continuous existence. In such a complex system, this flow condition produces a non-equilibrium state. The emerging fluctuations determine the global outcome of the system. Thus, living organisms self-organize by spatial-temporal development, as it is a system that is both non-linear and in a state of non-equilibrium. A "brain" is a system behaving openly in a spatial-temporal dimension, attempting to generate a functional information process.

Matsumoto [1] hypothesized that two general principles govern the auto-designing ability of an artificial intelligent machine, the so-called “brain-computer.” These two general principles are the output-driven operation and the memory-based architecture.

The output-driven operation principle stipulates that input signals change the information structure within the system during an output operation. According to this principle, learning is output-driven. Throughout this process, input signals are utilized by the system to select an output most suitable to its surrounding, resulting in an enhanced learning efficacy. Input signals deposit their recording traces in the system according to the memory-based principle, irrespective of whether the stimulation is strong enough to drive an output. The recording traces are then consolidated during an output-driven operation. Thus, the combination of an output-driven operation and a memory-based architecture provides the brain-computer with the potential to acquire its learning algorithm with the property of predictability. The learning algorithm, once acquired according to the output-driven operation principle, is stored as a memory.

With the evolution of more complex and heterogeneous neurocomputational brain-like architectures, the two principles have become significant in allowing the brain-computer to self-develop robust, flexible and predictable functions [2].

2. Brain-like computer system theory

Key characteristics of the brain information processing system are its massiveness, parallel-distributed computation and efficiency. Tens of billions of neurons in a brain system act and decide system outputs in real-time while having little or no contact with each other. The comparison can be made with a neuron to a man living on the earth. It is suggested that world stability would still occur even if man would only communicate with clusters of 1000 humans, out of a possible 6 billion inhabitants. A brain is so mysterious and attractive system.

This type of system excels in its design flexibility and its innate ability to extend its own functionality. Theoretically, the system engineering aspect can be applied to a sociological model, as well. Therefore, our focus in this paper is to primarily concentrate on the discussion of the brain system theory with models derived from the theory generated in this article based upon our past progress in this field.

A system can be identified as a “creature” if it can construct by itself a non-linear state of non-equilibrium. A creature organizes configuration and mechanism by itself by taking the material/energy and the information selectively in the organism. The drive, which guides a selective flow of material/energy, we identify as a physiological drive, and the drive, which guides a selective flow of information as a linking drive. Based upon specific creature-made ethics, a creature can acquire in a self-organized approach the configuration pattern and mechanism as information in a direction filling up these drives.

We proposed two general principles for biological information processing, the output-driven operation and the memory-based architecture, both of which are carried through up to the evolved brain from an original creature at every hierarchy step [2] [14]. It is thought that a constitution characteristic of biological information processing system is a hierarchical structure – as a related neighbor, the brain system constitution is also a hierarchical structure when based upon this theory. In other words, (1) once the information processing algorithm is acquired and consolidated as a long-term memory, the memory is kept. New algorithms will be memorized, in a laminated way, on the existing memory as a hierarchically structured memory. (2) Acquisition method of information processing algorithm by output-driven operation is valid with every level of brain hierarchical structure (brain system, neural network, neural cell etc.).

In the following sub-sections, we describe and discuss our theoretical models for a brain system level and a neural network level.

2.1. System model

A multi-agent system approach is proposed in order to achieve a flexible intelligent system [3]. The brain is a large-scale multi-agent system composed of many neurons when considered as an information processing system. Although the mechanism of the brain system constitution has not been elucidated yet, we have many indications coming from brain science by investigating both the phylogenetic and ontogenetic development process of a brain.

Mammalian brain, with its dual structure of archicortex and neocortex, developed as a phylogenetic evolutionary process from the frog to the mammal stage. In other words, the mammalian brain added a new type of structure called the neocortex onto the old frog brain. The number of layers in the neocortex increases with each evolution from rodent to human [4]. The neocortex has strong connections with the archicortex in structure and a strong correlation in its functionality. Amygdala complex (AM), the center of vital human life support systems such as the emotional system / eating system / autonomy system / or visceral mechanism in the old brain, is thought to play, in particular , an important role in the build-up of a neocortex system.

There is a direct information pathway from the thalamus to the AM. This pathway contributes fast output decisions for biologically important stimulus [5] [6]. The speed of the decision is rapid. The response latency from stimuli in the right amygadala through to the superior collicullus and pulvinar nuclei has been measured to be lower than 40ms [7]. The superior collicullus (SC) is on an old visual information pathway obtaining visual sensory input from retinal ganglion cells, performing coarse and rapid information processing such as attention control in humans [8]. SC activates emergent behavior such as escape by this rapid decision making. AM is known as a biological evaluation system and sending information to the neocortex directly by projection [9]. The understanding of this projection method would be an important factor in the development of neocortex functions.

For example, the anterior part of the inferior-temporal cortex (IT) in the neocortex is known for processing visual object recognition. The result of visual object recognition in IT is sent to the AM, where emotional meaning is added [5]. Finally, the result is sent to the basal ganglia (BG) to select and activate a relevant behavior in the context [10]. However, an innate visual perception function would not present itself in the IT spontaneously, as it is an acquired developmental process.

It is thought that the brain is formed phylogenetically in an ontogenetic process by the old brain. By designing "projection from AM to IT" beforehand into the ontogenetic development process, AM, which is functionally developed earlier than the neocortex, can transmit coarsely interpreted signals to the IT. The signal constructed by AM could navigate the neocortex information processing structure development process. In fact, there is a projection from a large part of IT to AM in infancy. It is known that the projection confines itself to the anterior part of IT as a baby matures and develops. However, when the anterior part of IT is ablated in infancy, locus to project on AM becomes the posterior part of IT [11].

On this basis, we have proposed "the development model that organizes functions in a new brain based on the functions in an old brain" [12] [2] (Fig.1), "the processing model that provides a top-down semantic hypothesis from an old brain to a new brain" [13] (Fig.2), and "the learning model that performs simultaneous learning of perception and action based on a decision made by an old brain" [14] (Fig.3).

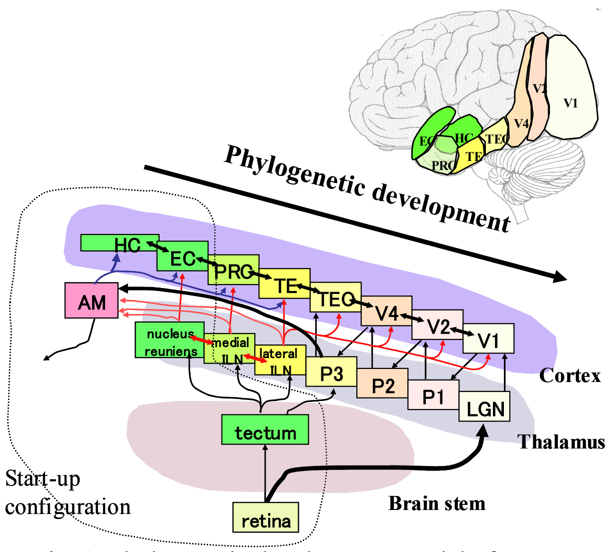

Fig. 1. Phylogenetic development model of cortex.

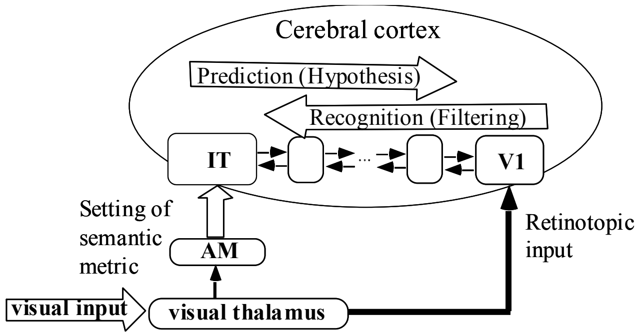

Fig. 2. Bi-directional processing system model.

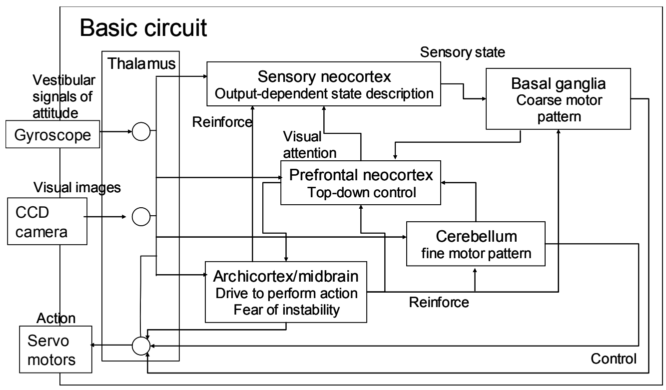

Fig. 3. Motor control learning system model.

2.2. Module-based computational architecture

The above-described system model connects learning and processing from each subsystem, and enables a more efficient processing in a whole system. It enables quick judgment and processing under incomplete information in a creature. On the other hand, we still have to include the facts that the brain information processing system is burdened by its massiveness and parallel-distributed computation problems. In this section, we argue for models of brain modules to determine these two characteristic.

The brain consists of different anatomical parts. It is presented as such because of the many different neurons and networks assembled as a subsystem hierarchically based on the phylogenetic development as in fig.1. However, the neuronal constitution and network are similar anatomically. An analogous network configuration becomes a cluster and constitutes a subsystem of each hierarchy. The smallest network configuration that can be built is called a "module" here.

If it is a neocortex, the module is a column configuration, if it is a lamella structure, it is the hippocampus, and if it is a lamina configuration, it is the cerebellum. Because this is such a hardware based architecture, it is suggested that a brain system does not merely perform parallel distributed processing. It is believed that a brain system executes the processing steps making use of its character embedded in a module.

We first introduce a neocortical computational module model proposed by [13] [15] in 2.2.1, then a basal ganglia computational module and neuron model proposed by [14] in 2.2.2.

2.2.1. Neocortical columnar module

The neocortex accomplishes a similar six layered configuration in all areas while the cytoarchitecture configuration varies to some extent by area. It is known that pyramidal neurons in layer II, III of the neocortex transmit a signal to higher level areas, while pyramidal neurons in layer V, VI transmit to lower area levels [16]. In addition, from the understanding of neocortex neural circuit functioning, it is observed that a signal from higher areas arrives at the apical dendrite of the neuron which subsequently sends information from the lower to the higher area [17]. We assumed the neocortex column configuration a module unit, based on the latest knowledge [13] (fig.4).

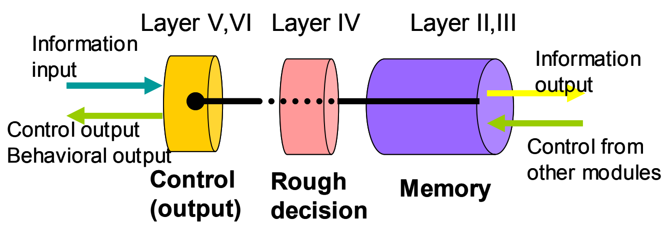

Fig. 4. Cortical columnar module model.

Neural network mechanism in a column module enables a symbolic description, since it has representation as a memory and is packed within a module. There is a description of the module itself in the memory of a module, gained from information and control inputs. The relationship between modules is learned by experience. A module receiving contradicting information becomes inhibited, a module behaving consistently with another activates mutual information flow, and when the input received is of a global control signal type, they become bound and synchronized. Modules are arranged on the neocortex like a two-dimensional array. They are mapped roughly by a control input given by a subsystem having a semantic space such as the AM.

When a neuron expresses the same temporal information, it is used as structural knowledge such as in the frame model [18]. On the other hand, it is used as logical knowledge when it expresses information with a different time period. Bi-directional reasoning that includes structure and logic functions in this way (deduction / induction).

A system can carry out the knowledge processing which, for example, was based on a hypothesis developed by the system theory in 2.1 describing the neural circuit module which can acquire this knowledge structure. We suggest that this process gives the neocortex a fast initial hypothesis when we put it forward from the AM since the AM output pathway sends a coarse semantic signal to the neocortex. Hierarchical processing has a major problem called combinatorial explosion. Higher modules have to combine results from the lower modules as a whole process of understanding. However, this problem is considerably reduced with our initial hypothesis, generation by AM driving, since the number of results linked, is limited by the initial hypothesis.

2.2.2. Basal ganglia computational module

The basal ganglia act as a rapidly learning system possessing a non-linear motor control selection. We propose a model based on our understanding of the mutually inhibited GABAergic connections and monoaminergic projections (fig. 5).

Although the structure of the basal ganglia could lead to unstable activity due to its open network architecture, its rapid learning properties would work to approach stable points for the initiating and switching processes between motivated behaviors. The behavioral quality, however, generally tends to be coarse due mostly to the animal body's poor performance in fine tuning its behavior compared to machines. This is because the rapid learning property of the basal ganglia should modify the memory in a shorter period, based and accumulated on its previous memories.

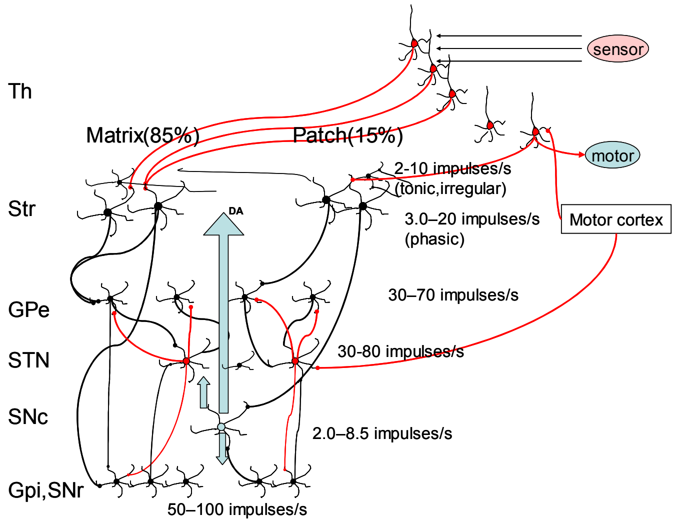

Fig. 5. Basal ganglia module model.

The correspondence between neural activities and rapid learning was solved by using the basal ganglia model. Here, we basically adopted the actor-critic system model [19] and realized it based on our proposed neuron model [14] (Fig.6).

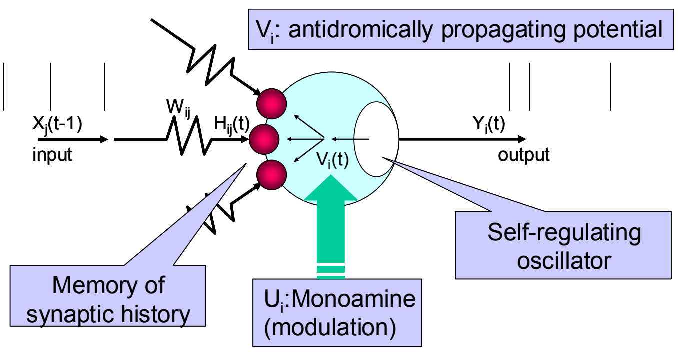

Fig. 6. Neuron model.

We included the modulation aspect in the neural model equation for the historical memory. The most novel idea on modulation in our model is the introduction of modulation fluctuation effects. Because the monoamine is a diffusing neural transmitter, it causes fluctuations in the neurons receiving this signal. The fluctuation triggers spiking behavior in the spatio-temporal topological distance between neurons, leading to the self-organization of the neural network connection. In the original actor-critic model [19], the authors introduced eligibility as an artificial factor in the learning temporal information process. Eligibility has a direct effect on the behavior of the predictive neuron in the critic subsystem, corresponding to that of a dopamine neuron in the basal ganglia [20]. Since eligibility is essential for TD-learning but biologically implausible in the brain architecture model, we introduced the new idea of a predictive neuron having the potential to learn temporal information based on synaptic history. Our new model is not only biologically plausible, but also demonstrates more selective learning because of its output-driven learning rule of synaptic efficacy.

3. Discussions

A brain system theory was presented including several derived models. It is evident from this discussion the brain remains complex and ambiguous. Essential concepts still to be completely elucidated are confined to the following: (1) learning control, (2) grounding of knowledge, and (3) processing method.

Learning control

We have proposed a system theory comprised of several modules performing different learning algorithms. Remaining still unclear is the method in which directional functionality could be imparted to these modules. Although we have introduced motivational signals propagating with a single purpose to each sub-system, this approach is only successful for a global direction and does not achieve a locally optimal direction. We suspect from these results that there should be more parameters controlling local learning profiles in a real brain system.

Grounding of knowledge

Symbol grounding is a major problem in artificial intelligence systems [21]. Our proposed model resolves some of the problems by introducing a bi-directional interaction between the environment and the system which also has bi-directional internal processing between its modules However, we cannot demonstrate that the developed knowledge is firmly grounded before producing a system performing a wide variety of functions using multi-modal sensory inputs. In this sense, a concentrated and focused effort will be required to establish a test bed such as a robot to achieve this purpose.

Processing method

Whether we follow the biological processing system such as the spiking neuron processing or equivalently, numerical processing remains a controversial discussion. Since the later is well understood and established in current computing architecture, we can easily demonstrate the desired functionality by designing relevant software. On the other hand, spiking processing has many unclear areas still to be determined. We believe an intermediate stage is required to establish a clear direction towards real brain-like computing.

References

[1] Matsumoto, G.. (1999) Brain Computing. Artif. Life and Robot. 3, 24-26.

[2] Koerner, E. and Matsumoto, G. (2002) Cortical architecture and self-referential control for brain-like computation: how the brain organizes computation. IEEE Eng. Med. Biol. 21, 121–133.

[3] Minsky, M. (1986). Society of Mind, Simon and Schuster, New York.

[4] Marín-Padilla, M. (1998). Cajal–Retzius cells and the development of the neocortex, Trends in Neurosciences, 21 : 64-71.

[5] LeDoux, JE. (1996). The emotional brain: the mysterious underpinnings of emotional life, Simon & Schuster, New York.

[6] Morris, JS,, Ohman, A,, Dolan, RJ,, (1999). A subcortical pathway to the right amygdala mediating "unseen" fear, PNAS, 96(4):1680-5.

[7] Morris, JS., Ohman, A,, Dolan, RJ,, (1998). Conscious and unconscious emotional learning in the human amygdala, Nature, 393(6684):467-70.

[8] Aizawa, H. (1996). Control of saccade and fixation by superior colliculus (in Japanese), Advances in Neurological Sciences,40(3):351-362.

[9] Iwai E., Yukie M,, Suyama H,, and Shirakawa S, (1987). Amygdala connections with middle and inferior temporal gyri of the monkey, Neuroscience Letter, 83(1-2):25-9.

[10] Parent A., and Hazrati L-N. (1995). Functional anatomy of the basal ganglia. I. The cortico-basal ganglia-thalamo-cortical loop. Brain Research Review, 20:91-127.

[11] Webster, MJ,, Ungerleider, LG,, Bachevalier, J, (1991). Lesions of inferior temporal area TE in infant monkeys alter cortico-amygdala projections, Neuroreport, 2(12):769-72.

[12] Koerner, E., Koerner, U. and Mastumoto, G. (1996). Top-down Selforganization of Semantic Constraints for Knowledge Representation in Autonomous Systems: A Model on the Role of an Emotional system in Brains, Bulletin of the Electrotechnical laboratory, 60:405-409.

[13] Koerner, E., Tsujino, H. and Masutani, T. (1997). A Cortical-type Modular Neural Network for Hypothetical Reasoning, Neural Networks, 10 : 791-814.

[14] Matsumoto, G. and Tsujino, H. (2003) Design of a brain computer using the novel principles of output-driven operation and memory-based architecture. In Cognition and Emotion in the Brain (T. Ono, G. Matsumoto, R. Llinas, A. Berthoz, R. Norgen, H. Nishijo, R. Tamura Eds.), 529-546.

[15] Tsujino, H., Körner, E. and Kondo, H. (2000). Hypothetical Reasoning and Brainware, Applied Mathematics and Computation, 111(2/3), 219-229.

[16] Pandya,D.N., Yeterian,E.H. (1985). Architecture and connections of cortical association areas, In: I. Association Cortices. Cerebral Cortex. 4, Association and Auditory Cortices, A.Peters & E.G. Jones eds., New York, Plenum Press : 3-62.

[17] Johnson, R.R. and Burkhalter, A. (1997). A Polysynaptic Feedback Circuit in Rat Visual Cortex. Journal of Neuroscience, 17: 7129-7140.

[18] Minsky, M. (1975). A Framework for Representing Knowledge. In P. Winston (Ed.), The Psychology of Computer Vision, McGraw-Hill: 211-277.

[19] Sutton RS and Barto AG. (1998). Reinforcement Learning: An introduction. MIT Press, Cambridge, MA.

[20] Schultz W. (1998). Predictive reward signal of dopamine neurons. J. Neurophysiol.; 80:1-27.

[21] Harnad, S. (1990). The symbol grounding problem, Physica D, 42: 335-346.