博士論文

Homeostasis by Action, Prediction, Selection in Embodied Neural Networks

(身体化されたニューラルネットワークにおける行動・予測・選択によるホメオスタシス)

Atsushi Masumori

升森 敦士

Contents

List of Figures

Abbreviations

LSA Learning by Stimulation Avoidance

LTD Long-Term Depression

LTP Long-Term Potentiation

PSA Principle of Stimulation Avoidance

SAA Stimulation Avoidance by Action

SAP Stimulation Avoidance by Prediction

SAS Stimulation Avoidance by Selection

SRP Stimulus Regulation Principle

STDP Spike-Timing Dependent Plasticity

STP Short Term Plasticity

Chapter 1

Introduction

1.1 Life as a Homeostatic System

What is life? Many researchers have been discussing this significant question, e.g., self-reproduction; dissipative structure; self-organization; minimizing entropy; autopoiesis are the important aspects of living organisms. It is difficult to propose a definition of life that convinces everyone. However, Pauling said: "In connection with the origin of life, I should like to say that it is sometimes easier to study a subject than to define it." [1]. It is worth examining necessary functions for living systems.

Living systems must maintain itself in order to exist. A property to stabilize internal states in the systems is called homeostasis [2]. For example, many variables of the body of living organisms are regulated to stabilize it within physiological limits (e.g., body temperature, glucose concentrations, pH, etc). Although homeostasis had been limited in physiological term, Ashby extended homeostasis to more general term, in a movement of cybernetics [3]. Ashby argued that the brain is an adaptive machine for maintaining homeostasis, and proposed a theory of ultrastability in which the system has two type of homeostasis, and if the first regular homeostasis is unstable and its essential variables exceed the limits then the second homeostasis works to rearrange the system dramatically. The system will reconstruct itself by trial and error until a stable homeostasis can be acquired. Ashby implemented this theory as a system called homeostat consisting of four modules. Ashby suggests that biological systems are ultrastable with these two types of homeostasis.

Homeostasis can involve not only internal mechanisms in systems but also behaviors that is output of systems. For example, the amount of glucose in the blood can be regulated by secretion of adrenaline or glucagon (i.e.,

internal mechanisms), and also by eating carbohydrate (i.e., behaviors). In this way, to maintain their own internal states, the systems have to adapt to their environment: the system must maintain the homeostasis through the interaction with the environment. Thus adaptation is regarded as a way, or an outcome of homeostasis like Ashby says "a form of behavior is adaptive if it maintains the essential variables within physiological limit." [4]

How do living systems adapt to environment for homeostasis? Dennett classifies living systems in four classes: Darwinian creatures, Skinnerian creatures, Popperian creatures, and Gregorian creatures [5]. Darwinian creatures are reactive systems designed thorough evolution. Darwinian creatures adapt to the environment by pre-designed behaviors. Skinnerian creatures have a phenotype that has plasticity by which, for example, the creatures can acquire sensory-motor coupling that is suitable for the current environment by trial and error. Although the creatures can adapt to the environment by learning new behaviors, the creatures apt to make fatal failures leading to the death in the process of the trial and error. Popperian creatures can simulate an outcome of their own action and environmental states in their brain, and select a good action based on the simulation. The creatures can avoid behaviors which apparently bring about such fatal consequences, and select behaviors which bring about good consequences. Thus they become more adaptable than Skinnerian. In addition to such simulation, Gregorian creatures use wisdom of ancestors. In other words, the ancestors leave their experience in a form of culture and the descendants use it to increase their adaptability. Dennett also argues that these types of creatures have been evolved in the order.

The ways of adaptation of these creatures can be roughly divided into two types: adaptation by hard-wired systems and adaptation by learning systems.

In the case of adaptation by hard-wired systems, systems must adapt to environment with pre-determined functions. In this case, the systems maintain the homeostasis by the pre-designed behaviors that can exploit the environmental state that is valuable for them (e.g., taking food reactively to specific sensor patterns). The systems can adapt to small environmental changes adjusting regulatory functions designed in advance (e.g., perspiration to temperature change). In this way, even if only the pre-determined functions can be used, the systems can adapt to the environment. However, such systems cannot adapt to large environmental changes (e.g., changes of a signal for foods) in an individual-level, and the systems adapt to such changes in a species-level through evolutionary process. Thus, this type of systems should have a shorter lifetime than the timescale of the large changes to adapt it. This strategy is effective to a situation that the environmental changes

are huge and require changing a design of the body rather than changing behavior, to survive there.

In the case of adaptation by learning systems, systems adapt to environment not only by pre-designed behaviors but changing behaviors. In this case, the systems can adapt to the large environmental change. By the learning ability, a prediction becomes possible and the systems can decide a behavior based on the prediction for the homeostasis in the future. This makes their adaptability higher in a situation in which the environmental change or the subsequence of action to the environment are predictable. In addition, a niche construction in a species-level (e.g., human society) becomes possible because the systems can learn from the remains constructed by the ancestors.

In this way, the way of adaptation does not simply indicate a level of progress, and effective ways vary depending on environment. Furthermore, the predictive systems need to be more complicated and larger than reactive systems to have a capability of prediction. This requires the systems to shift from microscopic world to more macroscopic world. For instance, Ashby advocated a Law of Requisite Variety [6]: in order to adapt to the diversity of the environment, the systems must have more diversity in itself; It is impossible to adapt the environment without having such diversity. Therefore, the adaptive learning systems require larger brain and body, and in a microscopic world, only simple adaptive hard-wired systems can exist and are effective.

In recent years many researchers have argued that prediction is most important for intelligence [7, 8]. Although it might be difficult to propose the definition of intelligence that convinces everyone, assuming that intelligence is adaptability in a macroscopic world which human can percept, prediction is essentially important for the adaptation since the outcomes of our action and the environmental states is somewhat predictable in the individual life timescale. (Evolution can be regarded as prediction in a species timescale: evolution imply prediction that descendants of survivors in current situations will adapt more in the future than that of non-survivors. Thus prediction might be essentially important regardless of timescale.)

To enhance the prediction, it is important to generalize phenomena as some sort of symbol (i.e., compressed information) and operate it. Consciousness might be evolved for simulating the world by treating such compressed information (e.g., counterfactual). For example, Aleksander proposed five axioms of consciousness (Here, A is an agent, S is sensorily-accessible world for A): Depiction: "A has perceptual states that depict parts of S.", Imagination: "A has internal imaginational states that recall parts of S or fabricate S-like sensations.", Attention: "A is capable of selecting which parts of S to depict or what to imagine.", Planning: "A has means of control over imaginational state sequences to plan actions.", Emotions: "A has

additional affective states that evaluate planned actions and determine the ensuing action." [9], although Aleksander does not argue the axioms are sufficient. All of these concern a process of determining action by simulating environment including the systems: depiction, imagination, attention and planning are required to simulate the environment and the system itself, and emotions is required to decide next action based on the simulation. After the emergence of the ability of generalization, by sharing the generalized symbols with others, languages considered to have emerged (i.e., evolution from Popperian creatures to Gregorian creatures). These ability make their adaptability higher in the macroscopic world.

In this way, living systems must maintain homeostasis taking suitable ways of adaptation to environment. If the systems live in the macroscopic world, the learning or predictive abilities become crucial, and as the consequence, learning or predictive system like brain have been evolved. In this thesis, we focus on such a homeostasis that is acquired by the systems with neural plasticity.

There are various theories and models of living systems based on such homeostasis.

In recent years, many researchers have argued that the most important function of the brain is prediction [7, 8] and the experimental results in vivo also support the argument [10, 11]. A modeling brain from the view point of prediction is called predictive coding and there are a lot of studies about it [12, 7]. In predictive coding, the networks reduce the error (prediction error) between input information and top-down prediction. This can be regarded as that the input information is suppressed by the prediction signal, and this directly leads to the stability of the internal state.

Friston has proposed an active inference in Free Energy Principle (FEP) [13, 14] based on Bayesian inference by extending the predictive coding to an action. For minimizing the prediction error (i.e., surprise) in FEP, the way of reconfiguring the internal model like predictive coding is called a perceptual Inference; on the other hand, reducing such surprise by actions is called an active inference. Both ways of reducing surprise should lead to maintain homeostasis. Indeed, Friston discusses the relationship between the homeostatic adaptive behavior of animals and the active inference [15, 16, 17]. There are some similar concepts to Friston's free energy minimization: a principle of redundancy reduction [18], maximization of Bayesian surprise [19], maximizing information transfer [20, 21].

There are also many pieces of research on the homeostatic adaptive system in the field of Artificial Life, that is a research field for studying on living systems by constructive approach. For example, Di Paolo and Iizuka reported that adaptive behavior is an indispensable outcome of homeostatic neural

dynamics [22, 23, 24], and Ikegami and Suzuki proposed homeodynamics where an autonomous self-moving state emerges from a homeostatic state [25].

In the next section, we introduce our framework and some more related researches.

1.2 Principle of Stimulation Avoidance

Some physiologists had long advocated theories on adaptive behaviors on the basis of animal experiments [26, 27]. Marom summarized such physiological theories and termed it Stimulation Regulation Principle (SRP) [28]. The principle consists of two properties: (i) a persistence of stimulation from outside drives the neural connectivity to explore the different topology (Modifiability); (ii) when removing the stimulus, there is no longer driving force to the connectivity and the network is stabilized in its last configuration (Stability). These two functions lead to that the system can acquire the behavior to avoid the stimulus.

Shahaf demonstrated that a dissociated neuronal culture can learn a desired behavior using a following protocol [29]. At first, an electrical stimulation with a fixed low frequency (e.g., 1-2Hz) is sent to a predefined input zone of the network; When the desired behavior appeared, the stimulation is removed; Repeating this protocol, the network learned to produce the desired behavior in response to the stimulation. In practice, the authors showed that the network learned to produce spikes at predefined output zone, in a predefined time window (within 40-60 ms after each stimulus) in response to the stimulation applied at the input zone. The results are promising in that behaviors for avoiding stimulation will be autonomously learned even without global reward. The authors argue that the results can be explained by SRP. However, the two properties which constitute SRP are macroscopic phenomenological explanations and these are not a concrete mechanism. Furthermore, assuming that modifiability is correct, learned configurations are destroyed at every stimulation; if stability is correct, it should not be necessary to repeat such cycle as in the experiment explained above.

In previous studies, we proposed a possible mechanism at a microscale of neural behavior similar to Shahaf's results, termed Learning by Stimulation Avoidance (LSA) [30, 31]. LSA states that spiking neural networks autonomously learn to avoid external stimuli by learning available behaviors. These behaviors emerge from spike-timing dependent plasticity (STDP) which has been found broadly in vitro [32, 33] and also in vivo [34, 35, 36], and the computational model has been proposed [37]. STDP causes changes

in synaptic weights between two firing neurons depending on the timing of their activity. Assuming there are presynaptic neuron i, postsynaptic neuron j, and synaptic connection from i to j (Fig. 1.1). If the presynaptic neuron i fires directly before the postsynaptic neuron j, the synaptic weight increases (long-term potentiation [LTP]), and if the presynaptic neuron A fires directly after the postsynaptic neuron B, the synaptic weight decreases (long-term depression [LTD]).

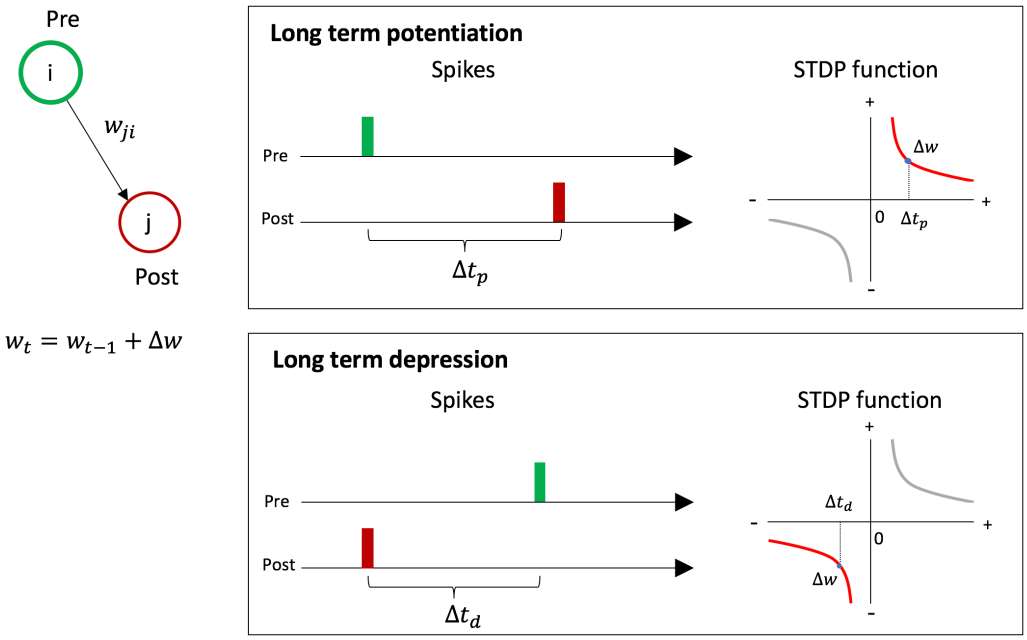

Figure 1.1: Spike-timing dependent plasticity in two neurons. There are presynaptic neuron i, postsynaptic neuron j, and synaptic connection from i to j (w_{ji}). Green circles represent presynaptic neuron; Red circles represent post synaptic neurons. Green bars represent spikes of synaptic neurons; Red bars represent spikes of post synaptic neurons. If the presynaptic neuron i fires directly before the postsynaptic neuron j, the synaptic weight increases with Δw (long-term potentiation). If the presynaptic neuron i fires directly after the postsynaptic neuron j, the synaptic weight decreases with Δw (long-term depression) .

In LSA, two dynamics for avoiding stimulation emerge based on STDP. To explain the two dynamics, here, we think about a minimal case with three neurons: an input neuron, an output neuron, a hidden neuron, and the connections from the input neuron to the other neurons (Fig. 1.2). The first dynamics is a reinforcement of behavior leading to a decrease in stimu-

lation by LTP (Fig. 1.2A). We assume an embodiment in which if the output neuron fires right after stimulation from its environment, then the stimulation is removed for a while. In that case, the connection strength from the input neuron to the output neuron increases because the the effect of LTP is stronger than LTD. On the other hand, the connection from the input neuron to the hidden neurons does not change much because the effect of LTP and LTD is almost same. Thus the behavior leading to the decrease in the stimulation is reinforced. The second one is a weakening of behaviors leading to an increase in stimulation by LTD (Fig. 1.2B). We assume an embodiment in which if the output neuron fires, then stimulation from environment to the input neuron starts. In that case, the connection strength from the input neuron to the output neuron decreases because the the effect of LTD is stronger than LTP. On the other hand, the connection from the input neuron to the hidden neurons does not change much because the effect of LTP and LTD is almost same. Thus the behavior leading to the increase in stimulation is weakened.

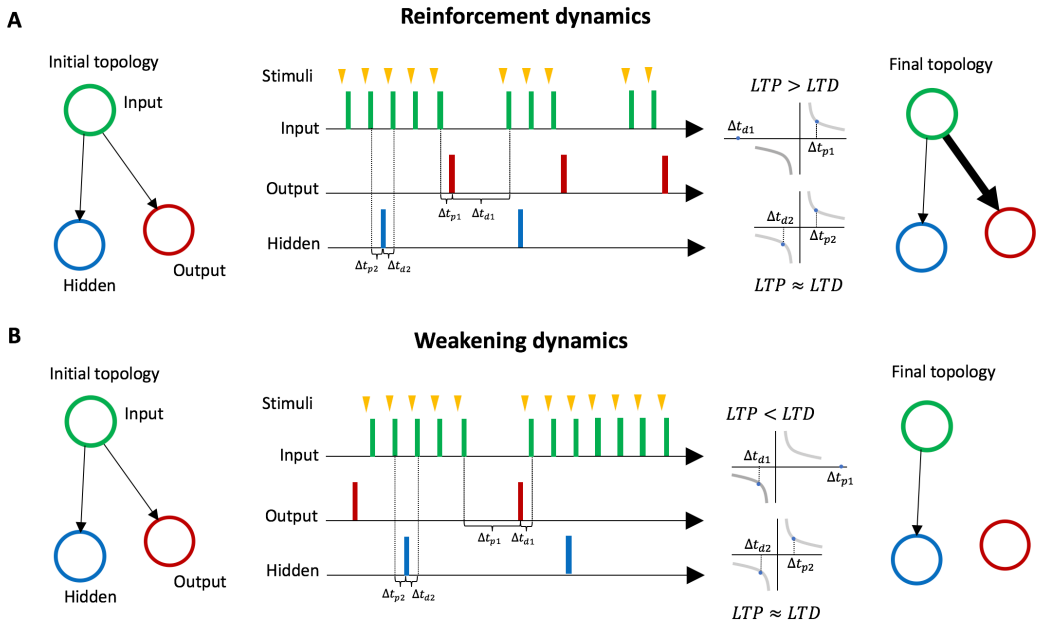

Figure 1.2: Learning by Stimulation Avoidance in three neurons: input neuron; output neuron; and hidden neuron. A: Reinforcement dynamics of LSA. Assuming an embodiment in which if the output neuron fires right after stimulation from its environment, then the stimulation is removed for a while. In that case, normally $\Delta t_{p1} < \Delta t_{d1}$, thus the effect of LTP to the connection from input to output is stronger than LTD, leading to reinforce the connection. On the other hand, normally $\Delta t_{p2} \approx \Delta t_{d2}$, thus the effect of LTP and LTD to the connection from input to hidden is almost same. B: Weakening dynamics of LSA. Assuming an embodiment in which if the output neuron fires, then stimulation to the input neuron starts. In that case, normally $\Delta t_{p1} > \Delta t_{d1}$, thus the the effect of LTD to the connection from input to output is stronger than LTP, leading to weaken the connection. On the other hand, normally $\Delta t_{p2} \approx \Delta t_{d2}$, thus the effect of LTP and LTD to the connection from input to hidden is almost same.

Although we considered about the minimal case, these dynamics should work in larger networks which can express more output patterns. Among those output patterns, output patterns which lead to decrease stimulation are reinforced by the first dynamics, and output patterns which lead to increase stimulation are weakened by the second dynamics. Actually, our previous study showed that LSA scales to large networks consists of 100 neurons [31]. Note that although the networks avoid the stimulation from the environment by LSA, the emerging behaviors are not limited to avoiding behaviors (e.g.,

object avoidance). What kind of behavior emerge depends on the embodiment of the agent (see Section 6.3).

In this way, LSA explains the property of trying to avoid stimulation by the two dynamics: reinforcement of behaviors leading to a decrease in stimulation; weakening of behaviors leading to an increase in stimulation. Unlike SRP, LSA is not a phenomenological explanation because LSA is based on STDP observed in vitro and in vivo. Nevertheless, LSA can also explain the experimental results by Shahaf. Particularly important in LSA is that behaviors for avoiding external stimulation emerge without any global rewards, if following two conditions are satisfied: the plasticity of the embodied neural network is driven by STDP; The network constitutes a closed-loop with the environment. This can be regarded as an intrinsic motivation emerged from the local dynamics of neurons (see Section 7.4).

The previous studies about LSA showed, embodied neural networks autonomously learn to avoid stimulation by "action". This can also be regarded as another neural implementation of Friston's active inference [14]. In addition, as introduced above, there are many experimental results supporting predictive coding in vivo. Predictive modules in the predictive coding can be regarded as a function which reduces an error between an input signal and a top-down prediction. If such the predictive module is regarded as an agent, such reducing error is equal to a reducing the influence of stimuli from the environment to the internal system. Thus it can be regarded as avoiding stimulation by suppressing input based on "prediction". Some studies showed that predictive coding can be implemented using spiking neural networks [38, 39, 40, 41], although the predictive networks in these studies require some well-designed structures or other functions than STDP. In this thesis, we show that spiking neural networks with STDP learn to avoid stimulation by prediction without preparing well-designed structures and a particular function other than STDP. It is suggestive that although Friston explains active inference and perceptual inference in the same framework FEP based on Bayesian inference, we show that such action and prediction emerge in a bottom-up manner based on STDP in neuronal cultures and spiking neural networks. In addition, we also show that another property to avoid stimulation by "selection" emerges in cultured neural networks and spiking neural networks: when an input neuron is uncontrollable (i.e., the networks cannot learn actions to avoid stimulation), the neuron is separated.

It means that at least three functions of avoiding stimulation emerge in embodied neural networks based on STDP: Stimulation Avoidance by Action (SAA); Stimulation Avoidance by Prediction (SAP); Stimulation Avoidance by Selection (SAS) (Fig. 1.3). We argue that these three functions can emerge from same neural network with STDP at the same time, and

which function emerges depends on a quality of stimuli: controllable inputs induce action, predictable inputs induce prediction and uncontrollable inputs (noise) induce selection. In other words, various functions emerge from stimulation avoidance as a principle of behavior. We integrate such mechanisms of stimulation avoidance in neural networks as Principle of Stimulation Avoidance (PSA). Stimulation avoidance can produce homeostasis: since unexpected stimulation to an agent from an environment represents environmental changes, avoiding the stimulation lead to decrease the influence of environmental changes to the internal state of the agent. It is interesting that some kinds of homeostasis directly emerges from synaptic dynamics, such as STDP. In this thesis, we present two approaches: analyzing cultured neural networks and modeling spiking neural networks, exploring the idea of PSA both theoretically and experimentally.

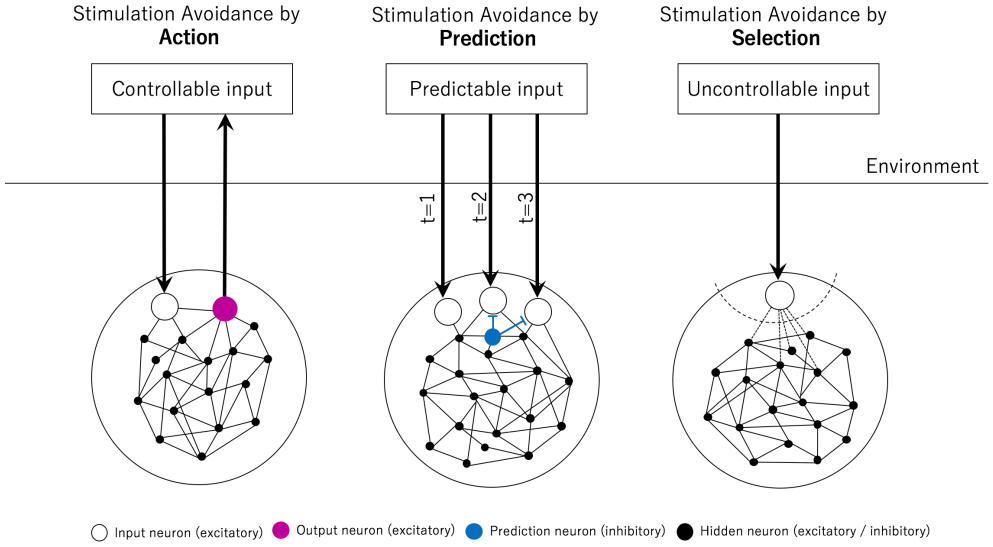

Figure 1.3: Principle of Stimulation Avoidance. Stimulation avoidance by action (SAA): when inputs are controllable, networks learn an action to avoid stimulation. Stimulation avoidance by prediction (SAP): when inputs are predictable, networks learn to predict target stimuli to suppress input neurons avoiding an effect of the stimulation to the internal network. Stimulation avoidance by selection (SAS): when inputs are uncontrollable (i.e., networks cannot learn actions or prediction to avoid stimulation), input neurons are separated avoiding an effect of the stimulation to the internal network.

1.3 Scope of This Thesis

In Chapter 2, we explain the common methods used in the following chapters. In Chapter 3, we focus on SAA by showing the results of the experiments in the embodied cultured neural networks. In Chapter 4, we focus on SAP by showing the results of the experiments in spiking neural networks on simple prediction that is possible to work in the dissociated neuronal cultures that we used in the experiments in Chapter 3. In Chapter 5, we find SAS in cultured neural networks, and we show computational model of spiking neural networks with asymmetric-STDP can reproduce the behavior. In Chapter 6, we discuss scalability and application of PSA (especially SAA) based on the experiments in the large spiking neural network models (Section6.1), estimation of the scalability on network size (Section 6.2), and the application on a humanoid robot experiments (Section 6.3) . In Chapter 7, we discuss some future problems regarding plasticity of neural networks and homeostatic properties.

Chapter 2

Methodology

2.1 Introduction

Biological neuronal cells cultured in vitro have been used to study neural systems [42, 43, 44, 45] because such neuronal cultures are easier to study than in vivo, they are composed of a relatively smaller number of neurons, and cultured in a more stable environment. Using neuronal cultures is also advantageous because potential unknown complex features in neuronal cells can be used, that are still difficult to implement in artificial neural networks. Although neuronal cultures are much simpler than real brains, they have essential properties, including spontaneous activity, various types and distribution of cells, high connectivity, and rich and complex controllability [46].

It has become easier and more popular to study the coupling between neuronal cultures and external systems [47, 48, 49]. In the previous studies, neuronal cultures were connected to an external system, such as a mobile robot in a real space. The sensory information coming from the external system was used to stimulate the neuronal cells, and the resulting neural activity controlled the external system. This change of external system provided feedback to the neuronal cell states, and this process could be repeated. This is called "closed-loop" and regard it as a model of primitive sensorimotor couplings. By studying such closed-loop systems, we can examine a learning or adaptability of a neural system with respect to embodiments.

Studies of closed-loop systems have been documented. For example, cultured neuronal cells were trained to achieve a desired behavior applying multiple stimulation patterns [45]. Another method were proposed for a cultured neural network to incrementally learn to respond in a particular way to a particular input [50]. One drawback of these studies is that they used a conventional microelectrode array as a recording device; this type of device

does not have sufficient spatial resolution so that it is difficult to stimulate and accurately detect a single neuronal state. In order to overcome the drawbacks, we used a recently developed device (high-density microelectrode array using complementary metal-oxide semiconductor [CMOS] technology) to detect activity of individual neurons with high precision. The details of this recording device are described in the following sections. The other drawback in those closed-loop studies is that an external evaluation function must be prepared and designed properly. A unique feature of our study is the examination of the self-development of such an evaluation function from the closed-loop system itself. We included this feature because we believe that self-development of an evaluation function is how adaptive behavior emerges spontaneously with most animals.

Such biological neuronal cells are often implemented and tested as a model in artificial neural networks to understand learning mechanisms [51]. Recently, the simulation of the more realistic artificial neural networks have become computationally efficient, by introducing models of spiking neurons [52, 53] and of synaptic plasticity [37, 54, 33]. These more realistic models can lead to theoretical understanding of biological neural networks. Many studies introduced previous chapter use abstract models, such an approach is thought to be important in understanding living systems, however, it is also meaningful to study using the more biologically plausible neural networks, since the brain functions is still not well known. There should be many things we can learn from the biological neural networks. Biologically plausible [55] or biologically inspired models [52] are not driven in a top-down manner, but driven by the local rules in a bottom-up manner. Such models also have the advantages that it is easier to implement as hardwares highly parallelized and processed asynchronously like brain, than such abstract models. Thus, a research on biologically plausible or biollogically inspired models would give important suggestions for the study of neuromorphic computing [56].

For these reasons, in this thesis we mainly used biological cultured neural networks and biologically inspired spiking neural networks for studying the homeostatic properties in embodied neural networks. In this section, the methods commonly used in the following chapters are described in the following section. Some additional methods or some modifications of following methods are described in each chapters.

2.2 Cultured Neural Networks

2.2.1 Neuronal Cell Cultures

The neuronal cultures were prepared from the cerebral cortex of E18 Wistar rats as previously reported [57, 58, 59]. The cortical area was trypsinized with 0.25% trypsin, and the dissociated cells were plated and cultured on a recording device (Fig. 2.1). The surfaces of the electrodes on the device were coated with 0.05% polyethylenimine and laminin to improve plating efficiency. The cells were cultured in Neurobasal Medium (Life Technologies) containing 10% L-Glutamine (Life Technologies) and 2% B27 supplement (Life Technologies) for the first 24 h. Half of the plating medium was replaced with growth medium (Dulbeccos modified Eagles medium (Life Technologies) containing 10% horse serum, 0.5 mM GlutaMAX (Life Technologies), and 1 mM sodium pyruvate) after the first 24 h. The cultures were placed in an incubator at 37°C with an H2O-saturated atmosphere consisting of 95% air and 5% CO2. During cell culturing, half of the medium was replaced once after several days with the growth medium. The all cultures used in our experiments consisted with about 100 cells and were sufficiently matured to show global burst synchronization [28].

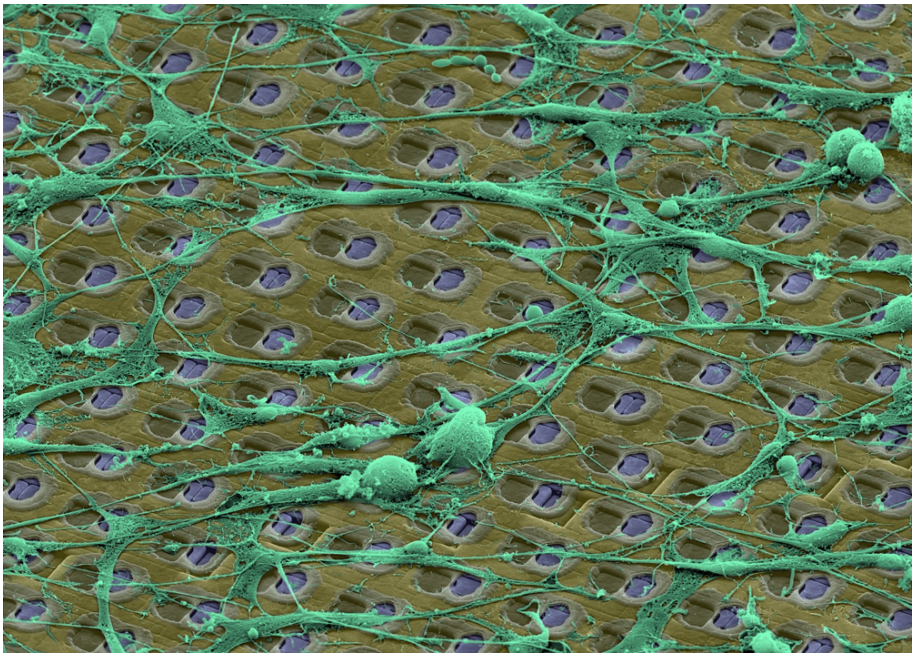

Figure 2.1: Dissociated neuronal cells cultured on recording device. (Copyright ETH Zürich)

2.2.2 High-Density CMOS Arrays

A high-density CMOS electrode array [60] was used to measure the extracellular electrophysiological activity of the cultured neurons (Fig. 2.2). This CMOS array is superior to the conventional multielectrode array (MEA) used previously [42, 43, 44] in that it has far higher spatio-temporal resolution. The number of electrodes in conventional MEAs is small, usually about 64, and the locations of the recording electrodes are predetermined with an inter-electrode distance of about 200 μm; thus, it is difficult to identify signals from an individual cell. In contrast, the CMOS arrays have 11,011 electrodes. The diameter of the electrode is 7 μm with an inter-electrode distance of 18 μm over an area of 1.8 mm $\times$ 1.8 mm. The device can record electrical activity on 126 electrodes at one time at a sampling rate of 20 kHz. Figure 2.3 shows a whole recording system including the CMOS arrays.

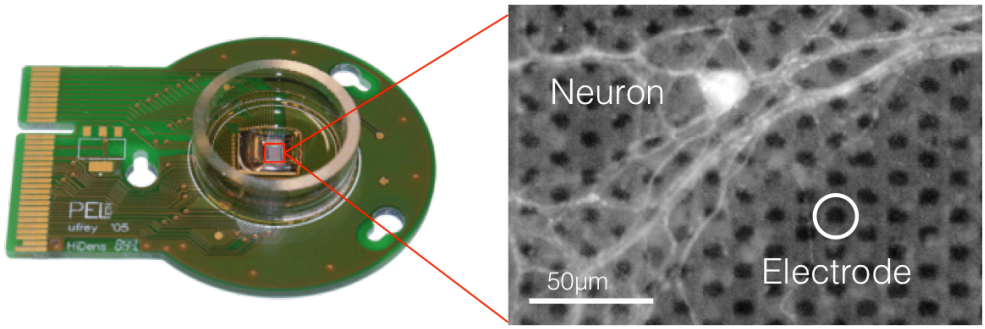

Figure 2.2: High-density CMOS electrode array. The device has 11,011 recording sites, a diameter of 7 μm, and an inter-electrode distance of 18 μm. The left figure shows the appearance of the device. The right figure shows partial enlargement of a part of the electrode array.

2.2.3 Estimation of Neuronal Somata Locations

Before recording the neural activity, we scanned almost all the 11,011 electrodes on the CMOS array to obtain an electrical activity map to estimate the locations of the neuronal somata. In each of the 95 recording sessions, the electrical activity were recorded for 60 sec with about 110-120 electrodes at the same time. In the recordings, sampling frequency set to 20 kHz and band path filter set to 0.5-20 kHz. An electrical activity map was obtained by averaging the height of the action potentials for each electrode. We applied a Gaussian filter to the map and assumed that the neuronal somata were located near the local peaks in the Gaussian-filtered map (Fig. 2.4).

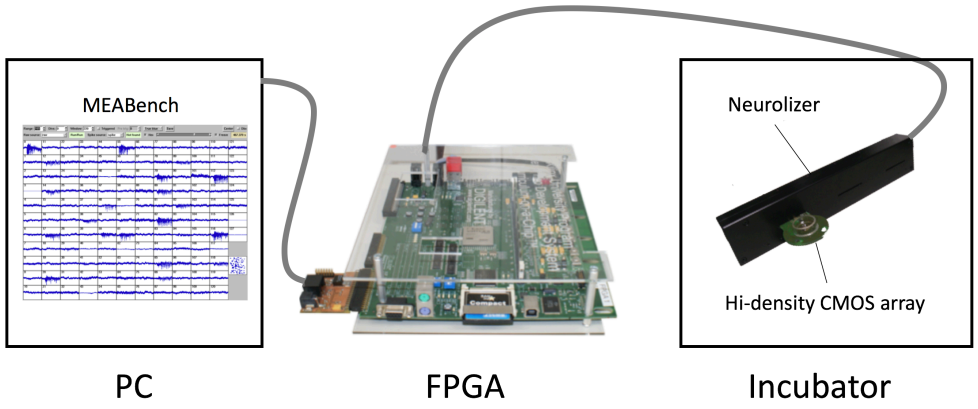

Figure 2.3: Overview of the whole recording system. Data of action potentials obtained in the CMOS arrays were processed: a frequency filtering, a signal amplifying and an analog-to-digital conversion. The processed data was sent to a recording software (MEABench) on PC via Neurolizer which take on roles: a power supplying to the CMOS arrays, sending data to and receiving data from the CMOS arrays. FPGA is a interface between the Neurolizer and the PC.

At most 126 of the all peaks were selected in descending order as the positions of neural cells, and the nearest electrodes to the peaks were selected for recording the neural activity. If the number of local peaks was fewer than 126, then all the peaks were used. By using the method, one electrode can ideally represent a single neural state, if cell size is less than 126.

2.2.4 Estimation of Excitatory and Inhibitory Synaptic Conductances

A type of neuronal cells are mainly divided to two types: excitatory neuron and inhibitory neuron. A spike of excitatory neuron increase the probability of firing of postsynaptic neuron of the excitatory neuron. On the other hand, a spike of inhibitory neuron decrease the probability of firing of postsynaptic neuron of the inhibitory neuron. In our experiments, identifying these type of neurons is important, since recording neural activity or stimulation was carried out for each neuron rather than group of neurons. A type of neuronal cell is classified as excitatory or inhibitory using the spike time series recorded for 10 min before the main experiment. Since the shapes of the action potential of these two neural types differ: In action potential of excitatory neuron, the distance between the maximum potential and the

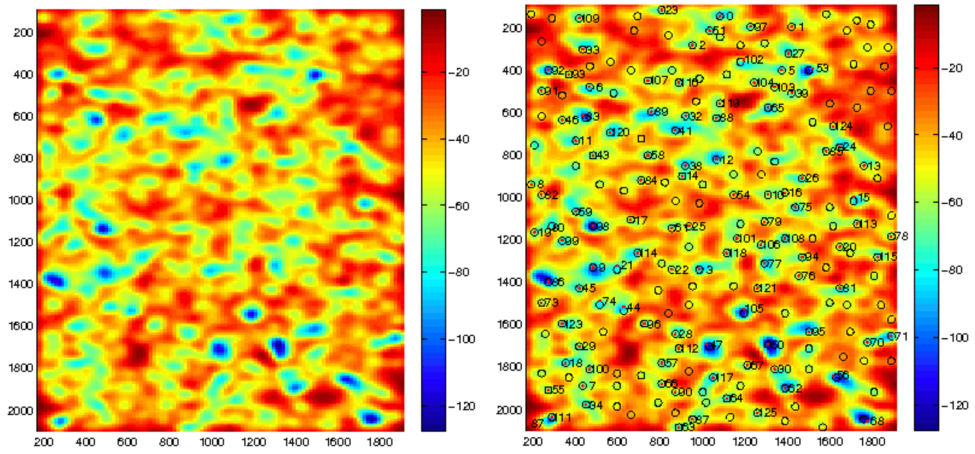

Figure 2.4: Estimation sample for location of neuronal somata. Left figure shows the results of applying Gaussian filter to the electrical activity map. Right figure shows selected peaks on the Gaussian filtered map. Red dots in a circle represent selected electrode and the number is the channel index.

minimum potential are longer than that of inhibitory neuron (Fig. 2.5), we classified the type of neuronal cell by using k-means clustering based on the difference of the shape.

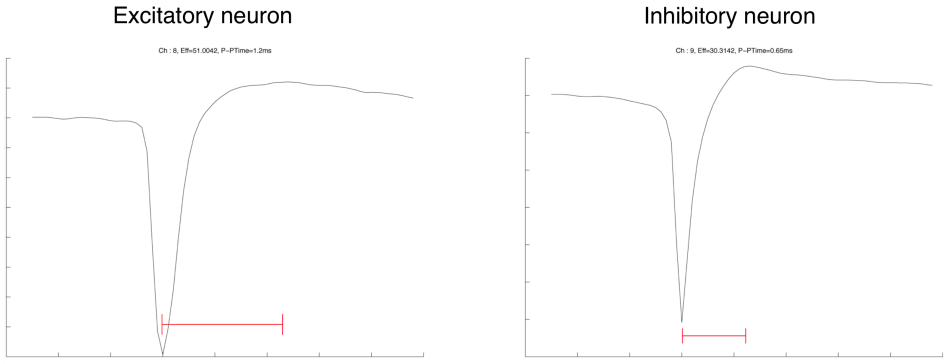

Figure 2.5: Estimation sample of neuron types. Left figure shows an action potential estimated as an excitatory neuron. Right figure shows an action potential estimated as an inhibitory neuron. The distances between the maximum potential and the minimum potential are depicted as red lines. A group of the longer distance was estimated as excitatory neurons. A group of the shorter distance was estimated as inhibitory neurons.

2.2.5 Recording and Preprocessing of Action Potentials

For detecting and recording the spikes of cultured neurons, we used the MEABench software developed by Wagenaar et al. (Fig. 2.6) [61]. All recordings were performed at a 20 kHz sampling rate using the real-time spike detection algorithm LimAda in MEABench. There are two type of recording data format: raw data and spike data (Fig. 2.7). Raw data format was used only in estimation for neuron type explained above. On the other hand, in all experiments and analysis in this thesis, spike data format was used. As the LimAda algorithm detects a spike that exceeds the threshold without distinction of positive and negative value, unexpected double detection of spikes can occur. These unexpected double-detected spikes were removed from the data before analysis: If the double detection occurred in 3.7 msec, second detected spike were removed.

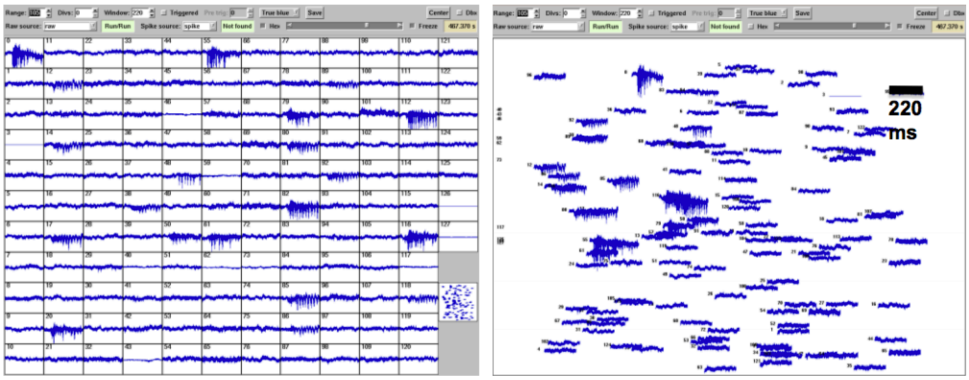

Figure 2.6: User Interface of MEABench. The neural activity can be observed in real time.

In experiments explained below, we need to detect the spike and stimulate the cultured neuronal cell at the same time. However, by sending the electrical stimuli to a neuronal cell through the electrodes, artifacts might occur. The Salpa filter in MEABench was used to remove the artifact in real time [62].

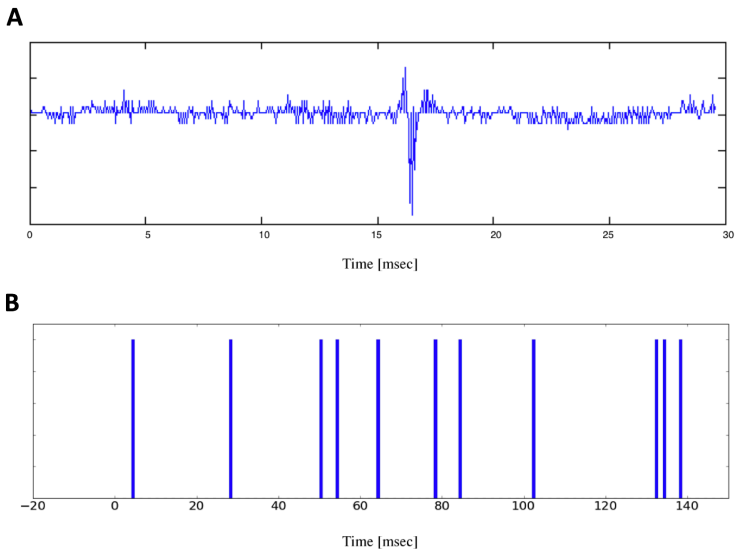

Figure 2.7: Example of recorded data. A: Raw data (30 msec). As the raw data format, time series of electric potentials of the electrodes is recorded. B: Spike data (170 msec). As the spike data format, time series of spike timing is recorded.

2.3 Spiking Neural Networks

2.3.1 Izhikevich neuron model

The model for spiking neuron proposed by Izhikevich [63] was used to simulate excitatory neurons and inhibitory neurons. This model is well know as it can be regulated to reproduce the dynamics of many variations of cortical neurons, and it is computationally efficient. The equations of the neural model are defined as:

$$\begin{aligned} \frac{dv}{dt} &= 0.04v^2 + 5v + 140 - u + I, \\ \frac{du}{dt} &= a(bv - u), \\ \text{if } v \geq 30 \text{ mV, then } &\begin{cases} v \leftarrow c \\ u \leftarrow u + d \end{cases} \end{aligned} \tag{2.1}$$

Here, v represents the membrane potential of the neuron, u represents a variable related to the repolarization of membrane, I represents the input current from outside of the neuron as explained in detail in the following

section, t is the time, and a, b, c, and d are other parameters [63] controlling the shape of the spike. The neuron is regarded as firing when the membrane potential $v \geq 30$ mV. The parameters for excitatory neurons (regular-spiking neuron) are set as: $a = 0.02,\ b = 0.2,\ c = -65$ mV, and $d = 8$, and for inhibitory neurons (fast-spiking neuron) are set as: $a = 0.1,\ b = 0.2,\ c = -65$ mV, and, $d = 2$ (Fig. 2.8). The simulation time step Δt is 1 ms.

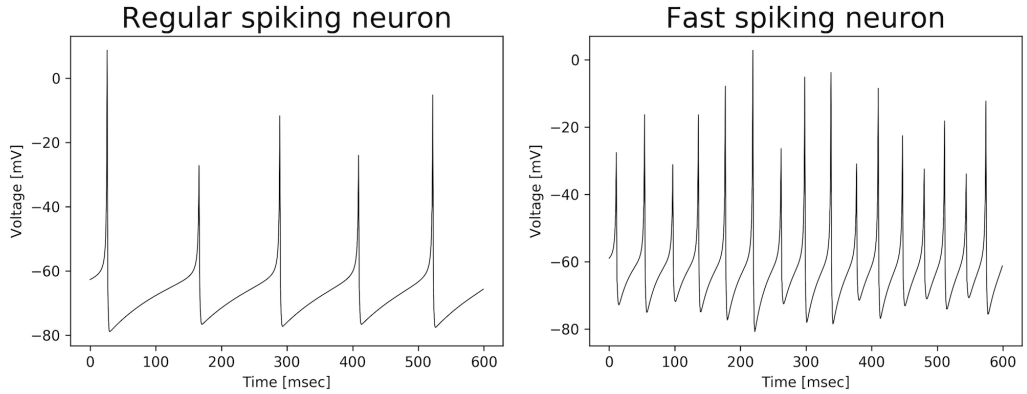

Figure 2.8: Dynamics of regular-spiking and fast-spiking neurons simulated using the Izhikevich model. Regular-spiking neurons are used as excitatory neurons and fast-spiking neurons are used as inhibitory neurons.

2.3.2 Spike-Timing Dependent Plasticity

STDP is a most important factor in our framework and used as a model for synaptic plasticity which changes the synaptic weight between two neurons depending on the timing of their spiking; when the presynaptic neuron fires right before the postsynaptic neuron, the synaptic weight increases, and when the presynaptic neuron fires right after the postsynaptic neuron, the synaptic weight decreases. The weight variation Δw is defined as:

$$\Delta w = \begin{cases} A_{LTP}(1 - \frac{1}{\tau_{LTP}})^{\Delta t}, & \text{if } \Delta t > 0 \\ -A_{LTD}(1 - \frac{1}{\tau_{LTD}})^{-\Delta t}, & \text{if } \Delta t < 0 \end{cases} \tag{2.2}$$

Here, Δt represents the relative spike timing between the presynaptic neuron a and the postsynaptic neuron b: $\Delta t = t_b - t_a$ (t_a represents the time of the spike of neuron a, and t_b represent the timing of the spike of neuron b). In this thesis, we change the parameters: $A_{LTP}, A_{LTD}, \tau_{LTP}, \tau_{LTD}$, depending on experiments to use various shapes of STDP. Figure 2.9 shows the variation of Δw in an example of rotating symmetry STDP depending

on Δt; Δw becomes negative when the postsynaptic neuron fires first, and is positive when the presynaptic neuron fires first.

The weight value w between excitatory neurons varies as:

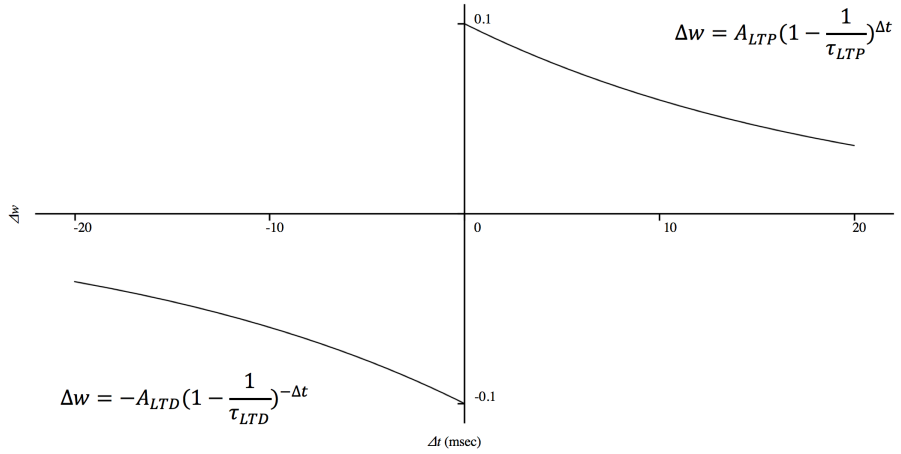

Figure 2.9: STDP function. The weight variation Δw of the synapse from neuron a to neuron b depends on the relative spike timing. The figure shows the values with the parameters: $A_{LTP}, A_{LTD} = 0.1$; $\tau_{LTP}, \tau_{LTD} = 20$ ms.

$$w_t = w_{t-1} + \Delta w \ . \tag{2.3}$$

The maximum possible weight is fixed to w_{max} (the values depends on experiments), and if $w > w_{max}$, w is reset to w_{max}. The minimum possible value of weight is fixed to $w_{min} = 0$, and if $w < w_{min}$, w is reset to w_{min}.

In addition to STDP, a weight decay function was also applied to the weights. The decay function is defined as:

$$w_{t+1} = (1 - \mu)w_t \tag{2.4}$$

The parameter μ was fixed as $\mu = 5 \times 10^{-7}$.

2.3.3 Short-Term Plasticity

The input current were added for each neuron n_i at each time step as:

$$
\begin{aligned}
I_i &= I_i^* + e_i + m_i \\
I_i^* &= \sum_{j=0}^{n} f_j w_{ji} s_j \\
f_j &= \begin{cases} 1, & \text{if neuron } j \text{ is firing} \\ 0, & \text{otherwise.} \end{cases}
\end{aligned}
\tag{2.5}
$$

Here, m represents Zero-mean Gaussian noise with a standard deviation $\sigma = 3$ mV that was delivered to each neuron at each time step as internal noise input; e represents external stimulation (conditions and frequency and strength of external stimulation depends on experiments); s represents short-term plasticity variables defined in (2.6). A phenomenological model of short-term plasticity (STP, [64]) was used, and s varies for each neuron n_j as:

$$
\begin{aligned}
s_j &= u_j x_j \\
\frac{dx}{dt} &= \frac{1 - x_j}{\tau_d} - u_j \; x_j \; f_j \\
\frac{du}{dt} &= \frac{U - u_j}{\tau_f} + U(1 - u_j) \; f_j
\end{aligned}
\tag{2.6}
$$

Here, x represents the amount of available resources, and u represents the resource used by each spike [64]. The parameters were set to $\tau_d = 200$ms, $\tau_f = 600$ms, and $U = 0.2$ mV.

STP is not necessarily required for LSA, but it is efficient to suppress global burst synchronization [31] and stabilizes the firing rate independent from the network size. Although LSA can be achieved only by a parameter tuning without STP (e.g., strength of noise input), it can be easily achieved with STP. We thus used the STP model in this study. Figure 2.10, 2.11 show the results of preliminary experiments for the STP function [65]. As adding an STP function, the burstiness index (an indicator for measuring global burst levels [61]) and firing rate was relatively stabilized regardless of the network size. Without STP, these variables significantly change with the network size.

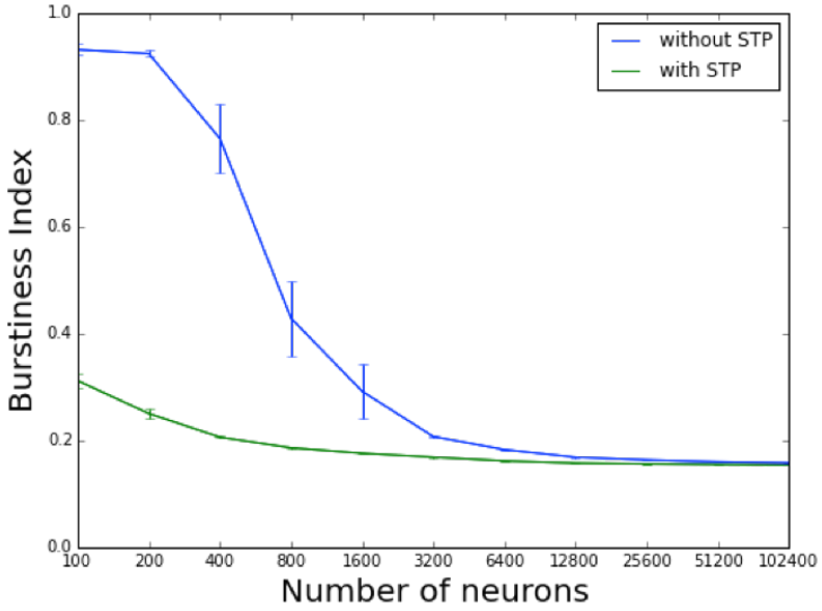

Figure 2.10: Dependence of burstiness index on network size. Error bars represent standard errors of the mean (n = 10). The burstiness index was not depending on network size with STP.

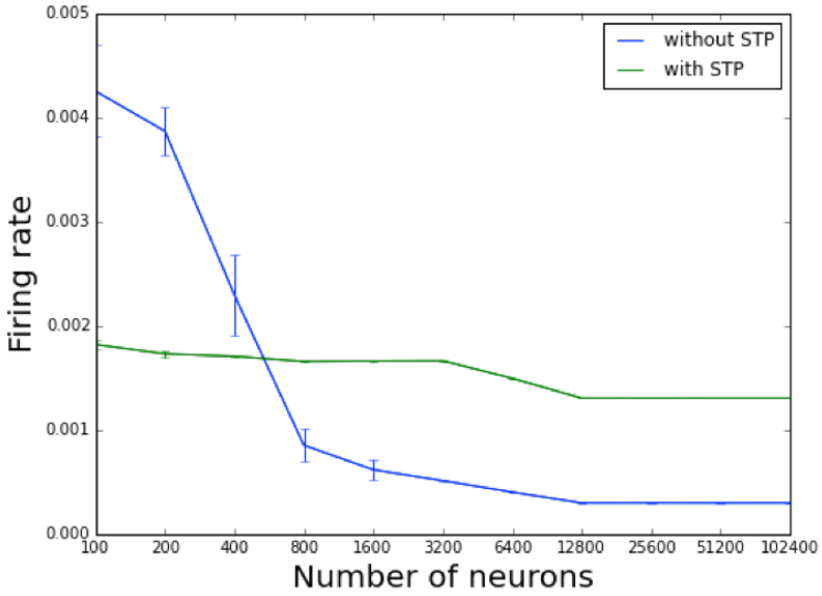

Figure 2.11: Dependence of firing rate on network size. Error bars represent standard errors of the mean (n = 10). The firing rate was not depending on network size with STP.

Chapter 3

Stimulation Avoidance by Action

3.1 Introduction

Autonomous agent have been a major research topic in the field of both artificial life and artificial intelligence from the beginning. For example, Braitenberg shown even various behaviors emerge from sensory motor couplings without brain; Brooks' subsumption architecture [66] was adopted to make "behavior-based" robots. Karl Sims proposed evolutionary approach to optimize the embodiment [67] and such an approach have been taken to evolutionary robotics [68]. In addition, learning agent have also been studied in many field, for example, bio-inspired reinforcement learning mechanism has been studied [69, 70], and in recent years, reinforcement learning with deep neural networks have been studied actively and received much attentions [71, 72]. These examples represent the adaptation by learning. However, a global utility function is still required to ensure the desired behaviors. For example, the neural architectures evolved according to a genetic algorithm and the agent's behaviors is reinforced based on the global rewards. In these approach, without using such a global utility function, we cannot expect the emergence of the desired behavior or other meaningful behaviors. However, in recent years, intrinsic motivation has attracted much attention in the context of autonomous agent, in particular in the field of reinforcement learning [73, 74]. Intrinsic motivation is a driving force to satisfy its own desire (internal reward). Applying intrinsic motivations, the agent can learn behaviors without global utility function. Usually intrinsic motivation is given to the agent in a top-down manner, thus it can also be regarded as indirect global utility function.

Shahaf and Marom demonstrated the interesting learning results where cortical cell cultures can learn a desired behavior as if it is motivated to avoid stimulation from environment, by following protocols [29, 28]. First, an electrical stimulation with a fixed low frequency (e.g., 1-2Hz) is delivered to the part of the network. When a desired behavior appeared, the stimulation is removed. Repeating these protocols, the network learned to produce the expected behavior in response to the stimulation. In practice, the authors showed that the networks learned to produce spikes at predefined output sections that was different from an input location in the network, and in a predefined time window (within 40-60 ms after each stimuli) in response to the stimulation applied at the input location.

They claimed that the cultured networks has two property: modifiability and stability, and the networks can learn a behavior to avoid the stimulation by these properties. However, these two properties seem to be impossible in small networks (e.g., 2 neurons). In addition, they showed almost no neural dynamics analysis, and since the learning task is very simple that can be achieved by just increasing whole firing rate of the networks, thus it is not clear whether the learning results are attained by merely increasing a whole firing rates of the networks or by changing the connectivity between input neurons and output neurons with a synaptic plasticity.

In a previous study, we proposed a possible mechanism at the micro scale of neural behavior similar to Shahaf and Marom's results, termed LSA [30, 31] (see Section 1.2). LSA is an emergent property of spiking networks coupled to STDP consisting closed-loop with environment. LSA states that the networks learn to avoid external stimuli by learning available behaviors. Such behaviors emerge based on STDP, which has been found in small networks both in vivo and in vitro (see Section 2.3.2). Thus LSA should work even in a small number of neuronal culture. In addition, LSA can be regarded as intrinsic motivation emerged from simple Hebbian rule such as STDP in a bottom-up manner. We discuss more about such stimulation avoidance behaviors as intrinsic motivation later (see Section 7.4).

This chapter focuses on LSA that cause Stimulation Avoidance by Action (SAA), in neuronal cultures. In following sections, we first show experiments on a simple embodied agent in a one dimensional virtual space using smaller size of neuronal cultures (approx. 100 neurons) than the previous study by Shahaf. Second we applied more complex closed-loop system to the small networks, where the networks connected to a mobile robot and placed on a two dimensional real space.

3.2 Autonomous Behavior in Neuronal Cell Cultures with Simple Embodiment

In this study, we first performed learning experiments using a smaller size of cultured neuronal cells than the previous works [29] to show such a learning mechanism scale from small to large cultured neural networks. In addition, we used high-density micro electrode arrays for recording neural activity, which can record each neuron's activity respectively, and analyzed the neural dynamics to examine whether neuronal cultures changed their network structure with synaptic weight to achieve a learning task.

3.2.1 Methods

We used neuronal cultures with a smaller number of neurons (approx. 100 neurons) and CMOS-electrode arrays for recording the neural activity and stimulation to the neurons (see Section 2.2); Using the CMOS-electrode arrays, we can record each neuronal activity respectively, if the number of neuron is less than 126.

Input and Output Channels

Input and output neurons were determined in the following way: input neurons were randomly chosen from channels that are classified as an excitatory neuron by the estimation method explained above (see Section 2.2). The number of input neurons depends on each experiment (2 or 10 neurons). Before starting learning experiments, 20 stimulations (1 Hz) were delivered to the input neurons and neural activity were recorded. Based on the recorded data, 10 output neurons were randomly chosen from excitatory neurons as to satisfy the following requirement: during the 20 stimulations, less than 5 neurons on average fired during the task window. This procedure was performed with 10 selected input neurons, and if there was no combination of such output neurons found, this procedure was performed with 2 selected input neurons.

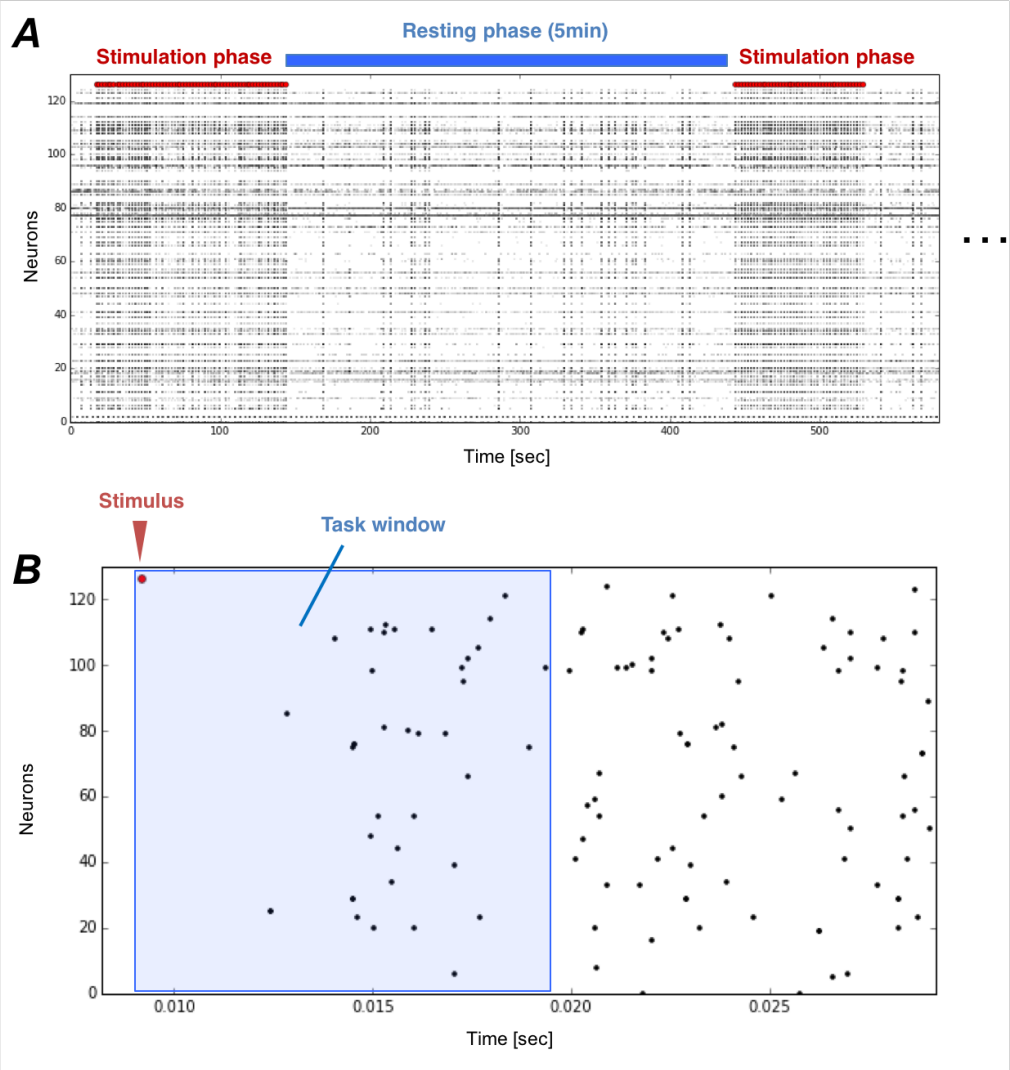

Figure 3.1: Illustration of the experimental setup. A: Stimulation phase and resting phase. The learning experiment consisted of the stimulation phase (in which stimulation was delivered to input neurons) and the resting phase (in which the stimulation was removed after 5 min when a learning task was achieved). B: Task window. The learning task was defined as more than 5 output neurons fired in the predefined task window (blue region) immediately after each stimulus.

Stimulation and Learning Task

Input currents as an external stimulation were delivered with a fixed frequency of 1 Hz to only input neurons (stimulation phase) and removed for 5 min (resting phase) when a learning task was achieved (Fig. 3.1A). After the resting phase, the stimulation starts again. This procedure was repeated 10 times per experiment. The learning task was defined as that more than 5 output neurons fired within a predefined time window (task window) immediately following a stimulus (20-40 ms; Fig. 3.1B). When the task was achieved, the stimulation would stopped. Control experiments were performed where the stimulus input stopped at random regardless of the neural activity of the networks (the other setup was same as the learning experiment).

This closed-loop between the cultured neurons and the environment is regarded as an embodied agent in a virtual one-dimensional space as follows: without the stimulation, the agent moves forward at a constant speed; if the robot approaches a wall of the space, the sensors stimulate the input neurons and if more than 5 out of 10 output neurons fire within 20 to 40 ms after the stimulation, the robot turns away from the wall by rotating at 180 degrees and there is no stimulation until the robot approaches the wall in the opposite direction (Fig. 3.2).

We performed six experiments with three different cultures (10-43 days in vivo) on the settings.

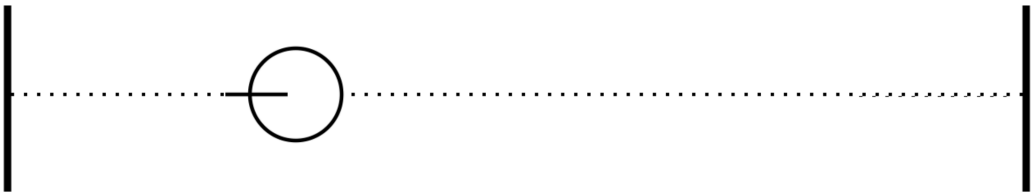

Figure 3.2: Conceptual diagram of the embodied agent in one-dimensional virtual space. In either end of the space, there is wall; When the robot contact with the wall, the robot will get stimulation. The robot moves from one side to the other side for 5 min.

3.2.2 Results

Evaluation of Wall Avoidance Behavior

We evaluated the learning results with a reaction time to the stimulation. The reaction time was defined as a time from start of stimulation to a time

of wall avoidance. Here, we defined a success of learning as that the reaction time is decreased by 30% or more. The average reaction time in the first 3 trials and the last 3 trials were used for calculating the success rate.

The success rate was 100% (6/6) in the learning experiment, while 16.7% (1/6) in control experiment (the success rate is not sensitive to the improvement rate with 30% above: e.g., the improvement rate with 20% or 40 % lead to same conclusion). Therefore, we found that the small cultured networks learn an action to avoid the stimulation well.

Figure 3.3 shows the average learning curves (n = 6), where a value on the learning curve represents the reaction time (lower value indicates higher learning ability). As shown in this figure, in the learning experiments (LSA), the number of stimulations rapidly lowered and stabled, indicating the higher learning ability. In the control experiments (Random), the number of stimulations did not stabilize at lower values and the variance was higher than LSA. This result is similar to previous results with large cultured neurons [29], showing that such learning behavior is scalable from small to large cultured neural networks.

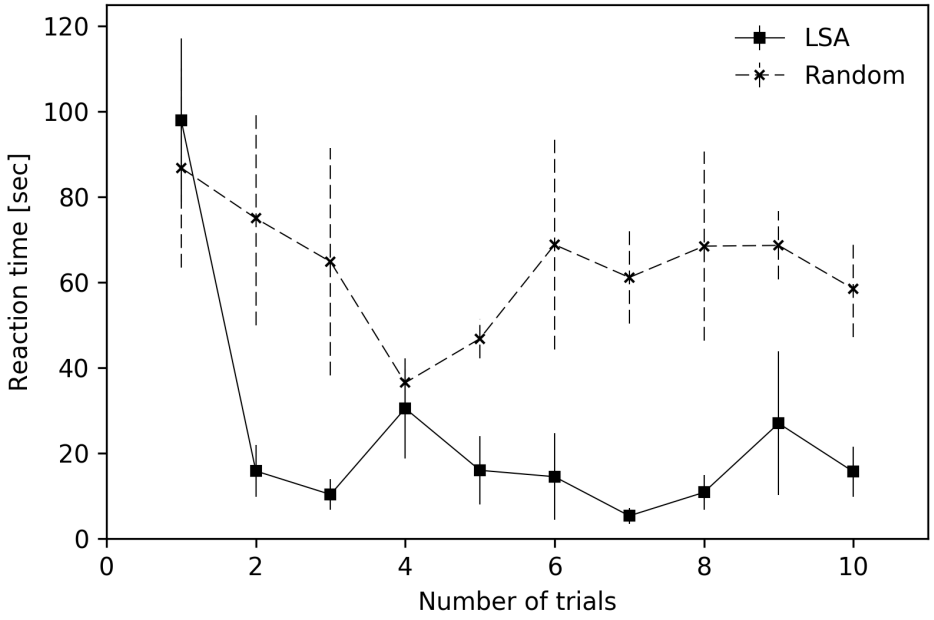

Figure 3.3: Learning curve in the learning experiment (LSA) and the control experiment (Random). The values of the learning curve represent the reaction time until the task was achieved (stimulation frequency was 1 Hz, thus the reaction time was equal to the number of stimuli to input neurons until task was achieved), so that lower value in the learning curves indicates higher learning ability. Error bars represent standard errors of the mean (n = 6).

In following sections, we show the analysis of the neural dynamics during the experiments.

Neural dynamics in stimulus-evoked spikes

This section focuses on analysis results of the stimulus-evoked spikes.

Figure 3.4 shows typical examples of an evoked firing pattern of output neurons. As shown in Fig. 3.4A, the number of spikes increased centering on the task-window and the spikes often decreased outside the task window, while the number of spikes in the control case did not increase (Fig. 3.4B).

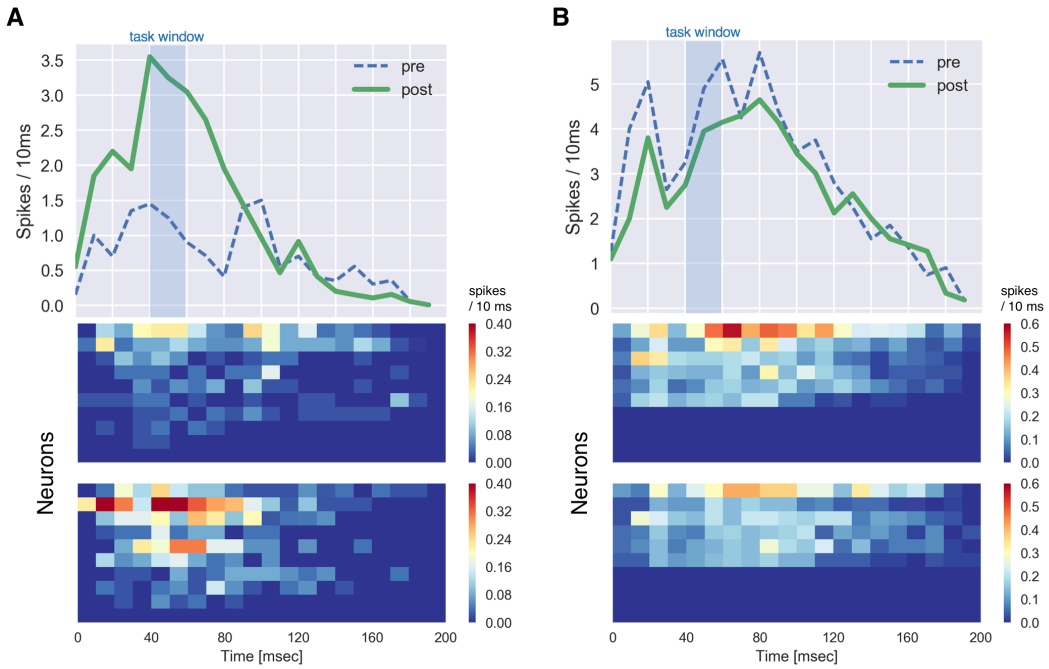

Figure 3.4: A: Evoked spikes of output neurons (200 ms) in LSA case. Upper panel is the time series of the evoked spikes of the output neurons. Lower panel is the heatmap illustration of evoked spikes of each output neurons. The masked regions represent the task windows. B: Evoked spikes of output neurons (200 ms) in the control case. Same conventions as in A.

Figure 3.5A shows mean spikes of the output neurons in the task-window. As shown in this figure, the mean spikes at the end of the experiment were significantly larger than those in the beginning of the experiment (Wilcoxon signed-rank test, $n = 6$, $p = 0.018$), while there was no significant difference in the controls (Wilcoxon signed-rank test, $n = 6$, $p = 0.866$). Figure 3.5B shows the mean spikes in not only the task window but the whole stimulation phase, and there was no significant difference between the mean spikes at the beginning of the experiment and those at the end in both LSA case and the controls. Therefore the results quantitatively show the number of spikes increased centering on the task-window.

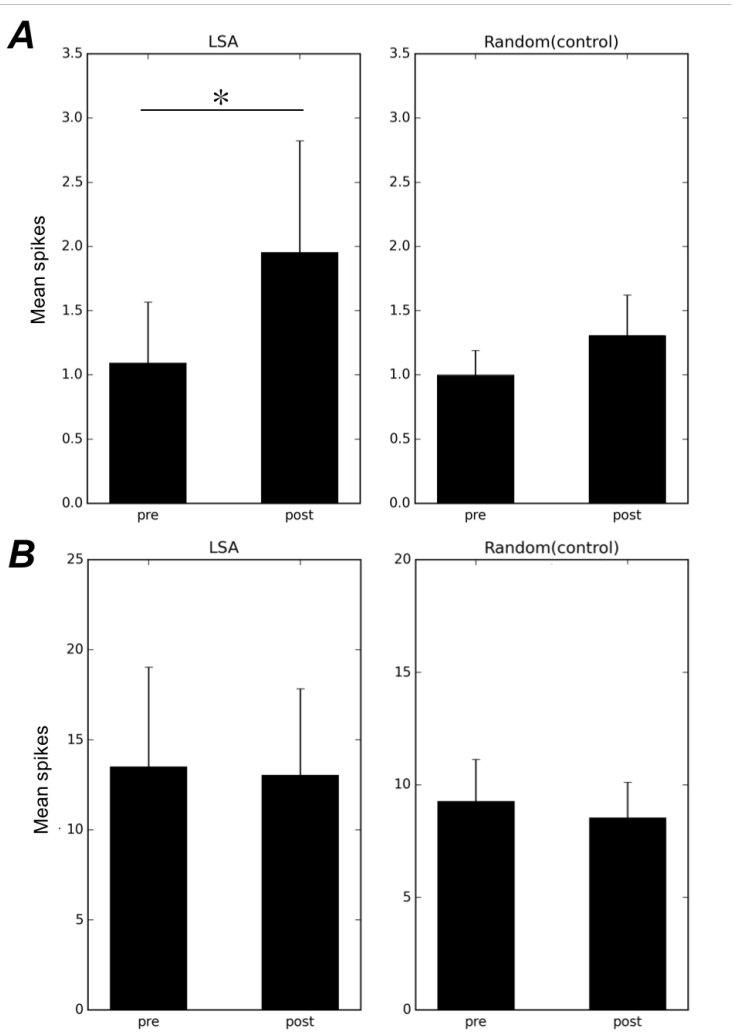

Figure 3.5: A: Mean spikes in task window. Error bars represent standard error of the mean (n = 6). In the learning experiment (LSA), the mean spikes in task window at the end of experiment (post) was significantly larger than at beginning of the experiment (pre) ($p<0.03$). B: Mean spikes in whole stimulation phase. Error bars represent standard error of the mean (n = 6). There was no significant difference in both the learning experiments (LSA) and the controls (Random).

Although task was very simple and thus the network can avoid the stimulation by just increasing the firing rate in a whole time, the network increased the firing rate only near the predefined task-window. This suggest that cultured networks can learn such temporal pattern to avoid electrical

stimulation.

Figure 3.6 shows a typical example of evoked firing pattern with all neurons. As shown qualitatively in this figure, the firing pattern at the end of the experiment (post) was different from the beginning of the experiment (pre 1 and pre 2).

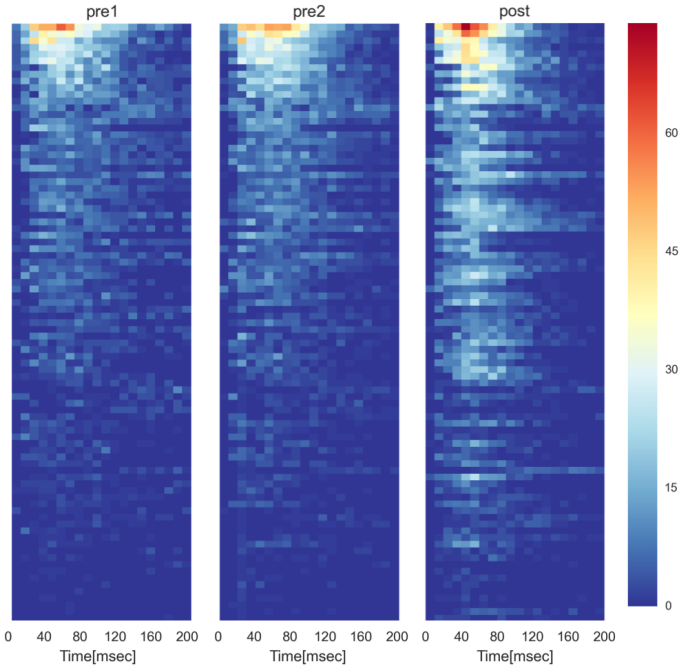

Figure 3.6: A typical example of evoked spikes of all neurons. pre1: the first 5 min in the experiment; pre2: the second 5 min in the experiment; post: the last 5 min in the experiment. The horizontal axis represents time and the vertical axis represents neurons; The order of neurons was sorted by the number of spikes in pre1 in a ascending order. The evoked spikes at the end of the experiment (post) differ from at the beginning of the experiments (pre 1 and pre 2).

To evaluate the change of the evoked spike pattern qualitatively, we used Jensen-Shannon divergence (JSD) to measure the similarity between the evoked spikes. JSD is a symmetrical variation of Kullback-Leibler divergence (KLD) [75], and defined as:

$$
\begin{aligned}
JSD(P||Q) &= \frac{1}{2}KLD(P||M) + \frac{1}{2}KLD(Q||M) \\
KLD(p||q) &= \sum_{i=0}^{n} p_i \log \frac{q_i}{p_i}
\end{aligned}
\tag{3.1}
$$

where $M = (P + Q)/2$.

Figure 3.7 shows that JSD between a baseline and a probability distribution of evoked firings; the baseline is a probability distribution of the evoked firings in the first 5 min in the experiment. This figure shows the probability distribution at the end of the experiment (post: the last 5 min) was significantly different from the value of the beginning of the experiment (pre: the second 5 min) in LSA case (Wilcoxon signed-rank test, n = 6, p = 0.027), and there was no significant difference in the controls (Wilcoxon signed-rank test, n = 6, p = 0.753). This quantitatively shows that not just firing rate but the firing pattern changed during the learning experiments.

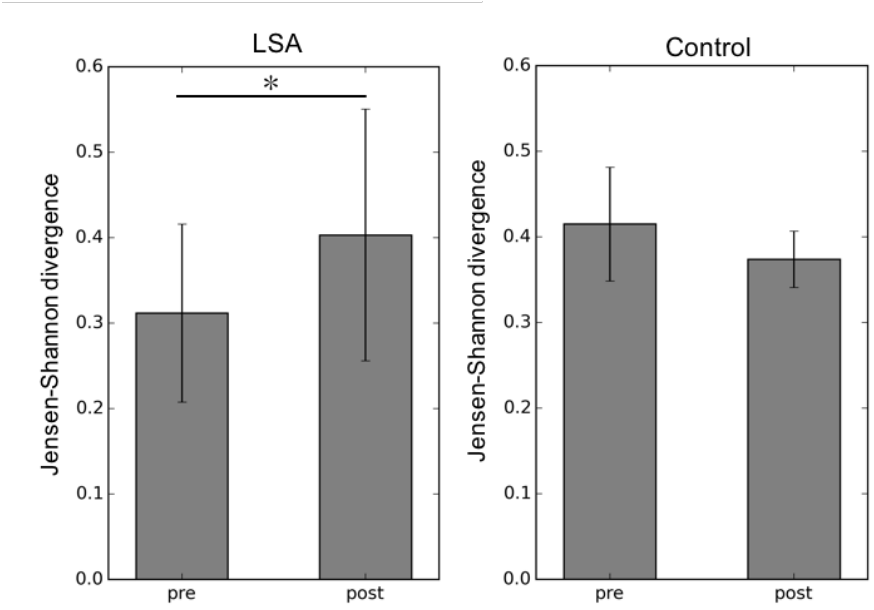

Figure 3.7: Mean JSD of the probability distributions of evoked firings with the baseline. Error bars represent standard error of the mean (n = 6). The value at the end of the experiment (post) was significantly different from the value at the beginning of the experiment (pre) ($p<0.03$).

In addition, in some of the control experiments, the firing rates in the evoked firing patterns at the end of the experiments (post) were considerably decreased, while such a transition did not occurred in the learning experiments. The study of this dynamics are described later (see Section 5.2).

Connectivity inferred with transfer entropy

This section focuses on the whole neural activity in the stimulation phase rather than the stimulus-evoked spikes.

We used transfer entropy to estimate the synaptic connectivity of the networks. Transfer entropy measures directed information transfer [76]. For instance, a higher transfer entropy from one neuron to another indicates that the former neuron strongly affects the latter. Thus, transfer entropy enables to discover the effective synaptic connectivity. We used an extended version of transfer entropy that is more suited to neural activity and called higher-order transfer entropy [77]. We applied it whole neural activity in the stimulation phase to see how the connectivity between input neurons and output neurons changed.

Higher-order transfer entropy (TE) from time series J to I is defined as

$$TE_{J,I}(d) = \sum p(i_{t+1}, i_t^{(k)}, j_{t+1-d}^{(l)}) log \frac{p(i_{t+1} \mid i_t^{(k)}, j_{t+1-d}^{(l)})}{p(i_{t+1} \mid i_t^{(k)})} \tag{3.2}$$

where i_t denotes the value at time t of I; j_t denotes the value at time t of J; i_{t+1} denotes the value at time $t+1$ of I, and these values could be either 1 or 0, indicating whether a neuron is fired or not. The parameters k and l give the order of TE, implying the number of time bins in the past that are used to calculate the histories of time series I and J and set to $k = l = 3$ here. The parameter d is time delay, and varies from 1 to 60. There is a synaptic time delay between neurons, and to account for such time delays, the time bin of J was shifted with d (1 to 60), thus TE was defined as a function of time delay d. After calculating TE with each d (1 to 60), the peak value was regarded as a connection strength between I and J.

Figure 3.8 shows a typical example of estimated functional networks with TE, where white points represent the input neurons, black points represents the output neurons. The neurons that have at least one connection, and the connections which have top ten percent of the TE value were depicted. This figure shows that at the beginning in the experiment (pre), there were strong connections only between the input neurons, however, at the end (post) there were many strong connections between the input and output neurons.

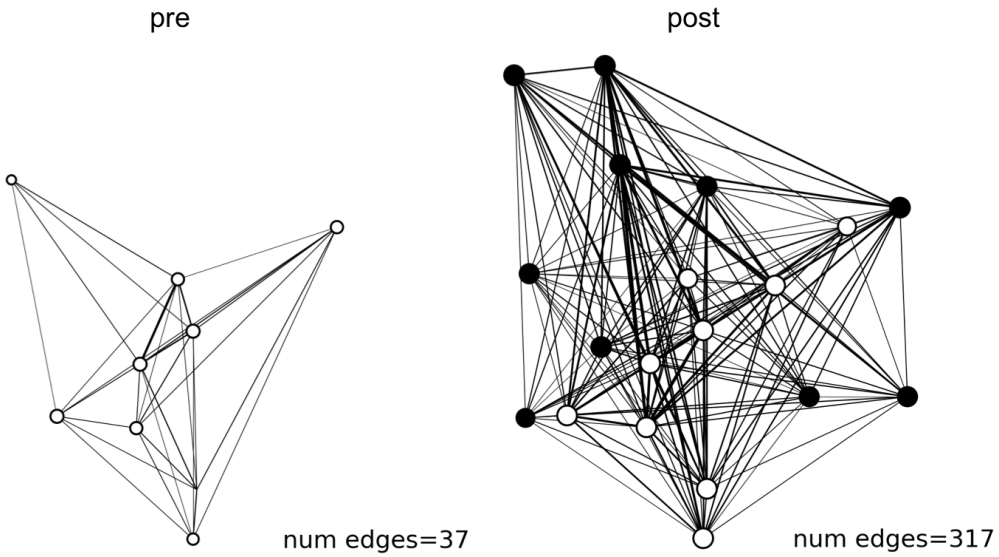

Figure 3.8: Typical example of functional networks inferred with transfer entropy between input and output neurons. White nodes represent input neurons and black nodes represents output, and only neurons which have a connection was depicted. Larger node indicates having higher inputs or outputs, and thicker line represents stronger connection. pre: the first 5 min in the experiment. post: the last 5 min.

Figure 3.9 shows mean number of connections inferred by TE between the input and output neurons. If the value of TE exceeds a threshold, there is considered to be a connection between the neurons. The threshold was determined to include connections that have top ten percent of TE value calculated with data in the first 5 min of the experiment. As shown in this figure, the number of connections at the end of the experiment (post) was significantly higher than the beginning (pre) in LSA case (Wilcoxon signed-rank test, $n = 6$, $p = 0.018$), although there was no significant difference in the control experiments (Wilcoxon signed-rank test, $n = 6$, $p = 0.753$).

These results indicate that information flow between the input and output neurons increased. Although these were the estimated functional networks, the results suggest that the network structure in the neuronal cultures changed with the synaptic plasticity.

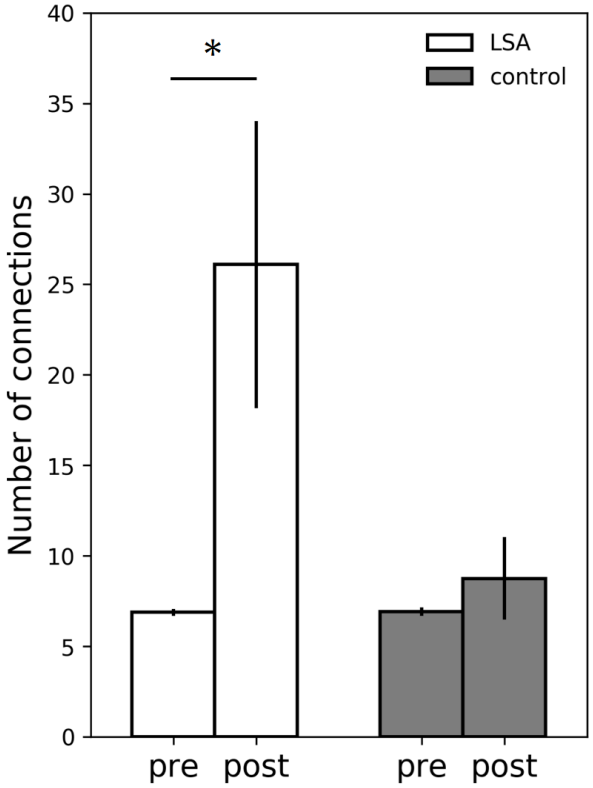

Figure 3.9: Mean number of connections inferred with TE between the input neurons and the output neurons. Error bars represent standard error of the mean (n = 6). In the learning experiment (LSA), the number of connections between input and output became significantly larger ($p<0.03$). There was no significant change in the control.

3.2.3 Conclusions

The learning experiment was performed with the smaller number of cultured neuronal cells (approx. 100 cells) than the previous work (approx. 10,000-50,000) [29], where the external stimulation was delivered to the input neurons with the fixed frequency of 1 Hz and the stimulation was removed for 5 min when expected neural behavior appeared; this closed-loop can be regarded as embodied agent in virtual one-dimensional space. We found that the smaller number of cultured neuronal cells can learn the simple behavior to avoid the stimulation as well as the case of larger number of neuronal cells in the previous work [29].

In this study, high-density micro electrode arrays capable of recording

each neuron's activity was used. The results of analyzing these neuronal activity revealed that learning results were attained by increasing firing rates only near the predefined task window and the evoked firing patterns gradually changed over time. In addition, observing connectivity between the input neurons and the output neurons, inferred by TE, revealed that the functional networks changed during the experiments. These results suggest that cultured neurons can learn the simple behavior by changing their network structures with synaptic plasticity rather than a state transition with response to the external stimulus (e.g., just increasing whole firing rate). These results suggest that learning such stimulus avoiding behaviors is caused by synaptic plasticity, such as STDP, thus LSA works in neuronal cultures.

3.3 Autonomous Behavior in Neuronal Cell Cultures with Complex Embodiment

In previous section, we demonstrated that embodied neural networks with the small number of cultured neurons can autonomously learned the simple behavior to avoid the stimulation. In this section, in order to investigate whether such learning system can scale to a more complicated embodiment and environment, we performed learning experiments coupling the cultured neurons to a mobile robot in a two-dimensional real space.

3.3.1 Methods

We used neuronal cultures and CMOS-electrode arrays for recording the neural activity and stimulation to the neurons (see Section 2.2).

Colsed-Loop System

We implemented a closed-loop system between an embodied cultured networks with the mobile robot and the environment. This system mainly composed of three components: a recording system monitoring the cultured networks, the robot as body of the cultured networks, and the interface connecting them. The system setup is depicted in Fig. 3.10.

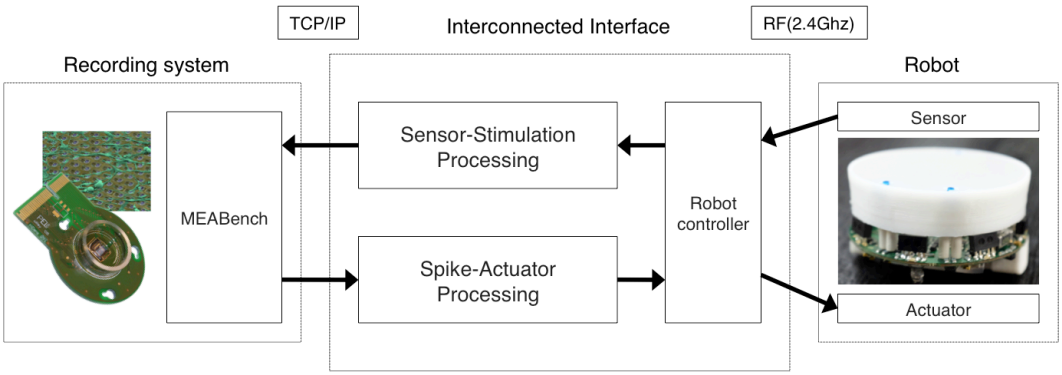

Figure 3.10: Overview of the closed-loop system composed of the high-density CMOS electrode array monitoring the neuronal cultures, a mobile robot, and the interface connecting them. The robot communicated with the interface by radio frequency, and the recording system communicated with the interface via TCP/IP.

The sensory information of the robot coming from environment were processed to stimulate the neuronal cells, and the resulting neural activity were processed to control motions of the robot. This change of movements of the robot provided feedback to the neuronal cell states, and this process could be repeated. Thus the system constitutes a closed-loop and regard it as a model of primitive sensorimotor couplings. In the recording system, the CMOS array and the MEABench software were used for recording and stimulating neuronal cells (see Section 2.2). Elisa-3 (GCtronic, Ticino, Switzerland) was used as the mobile robot (Fig. 3.11). Elisa-3 is a circular small robot of 5 cm diameter and has two independently controllable wheels. Only the front right and left distance sensors were used as sensory signals for stimulating the neuronal cells. The refresh rate of the robot was set at 10 fps.

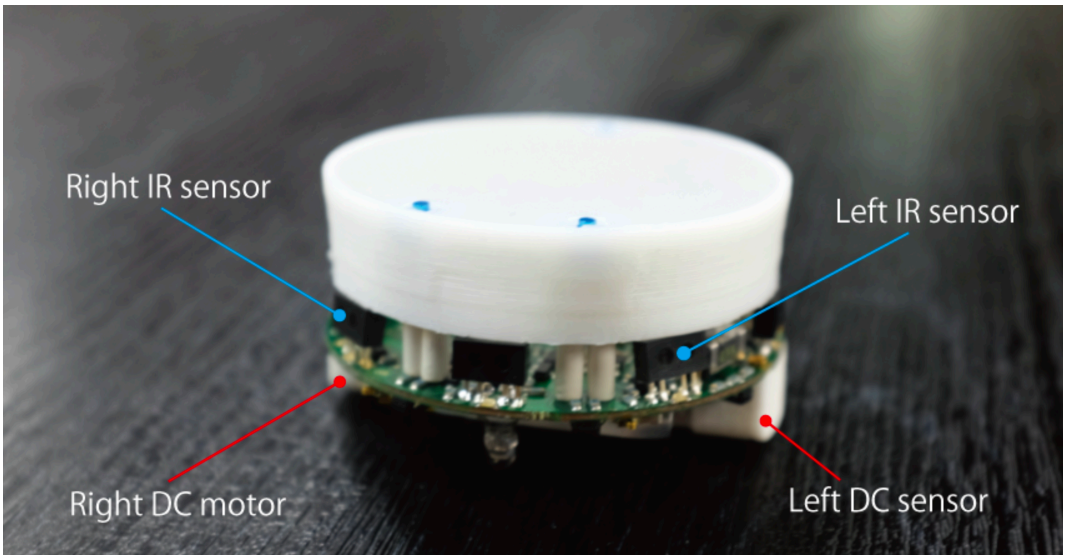

Figure 3.11: Appearance of mobile robot (Elisa3).

The interface played a role in receiving a sensor value from the robot and stimulating the neuronal cells based on the sensor value thorough the CMOS array. A maximum frequency of the stimulation is 10 Hz. The interface also played a role in receiving detected spike data from the CMOS array in real time and calculating a wheel speed based on the spike data and sending it to the robot. In this way, the robot and the neuronal cells form a closed-loop. More details of the sensorimotor mapping are described in the following section.

Sensorimotor mapping

A simple sensorimotor mapping was applied to the robot and the neuronal cells on the CMOS array (Fig. 3.12). We randomly selected two channels that were estimated as excitatory neurons as the left and right input neurons for sending the electrical stimuli. At given time intervals (100 ms), the probability $P_{L,R}$ for electrical stimulation to the input neuron was controlled by the sensory value of the mobile robot. Specifically, the probability was calculated as:

$$P_{L,R} = \begin{cases} 0 & (S_{L,R} < T) \\ S_{L,R}/S_{max} & (S_{L,R} \geq T) \end{cases}$$

If sensor value $S_{L,R}$ (0-950) was less than a threshold T, $P_{L,R}$ becomes zero. Otherwise $P_{L,R}$ is calculated by $S_{L,R}/S_{max}$. S_{max} denotes a maximum value of the sensor input (950). Whether the stimulus will be delivered to

the input neuron or not was determined with this probability every 100 ms. The threshold T was set to 100. According to this form, the distance from the robot to the wall was encoded as the stimulation frequency.

We also randomly selected 20 neurons that were estimated as excitatory neurons for calculating each left and right wheel speed; a half of them were used as left output neurons and the others were used as right output neurons. The wheel speeds were calculated based on the number of spikes of the output neurons that were integrated every 100 ms. The left and right wheel speeds $V_{L,R}$ was calculated as:

$$V_{L,R} = k \sum_{i \in N_{L,R}} v_i + C_{L,R}$$

These virtual neural states v_i take positive integers, which are equal to the number of spikes of the output neurons over a given time interval (100 ms), and sum them with the negative constant weight k and a positive constant C as a default wheel speed was added. N_L and N_R were set of channel number of left and right output neurons. Here, as k is a negative value and C is a positive value, the robot moves forward when the output neurons are not active. k was set to -0.3. The default values of $C_{L,R}$ are 12.5 and the values were adjusted respectively so that the robot go straight before experiment. As the activity of the output neurons increase, the speed of the forward movement decreases and finally the robot moves backwards. As the two wheels of the robot are independent, the robot can also turn, when the speeds are different.

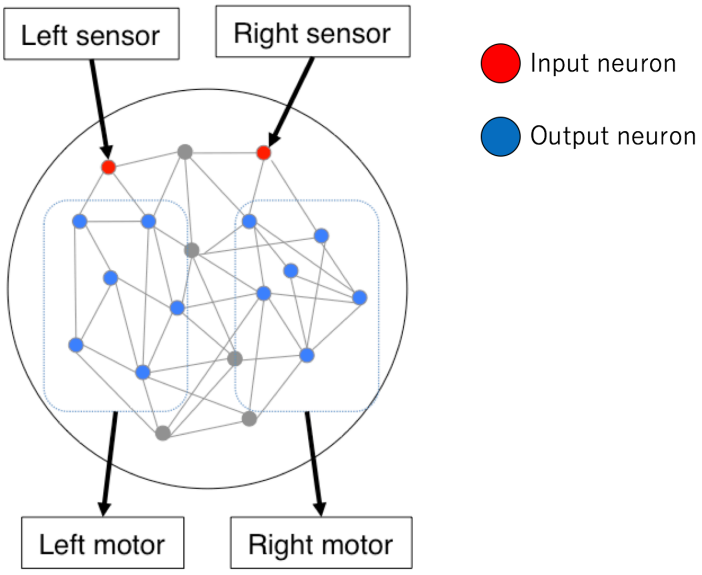

Figure 3.12: Sensorimotor mapping between the robot and the cultured networks. Two electrodes on the CMOS array were selected as input neurons; the input neurons got stimulation based on the value of distance sensors of the mobile robot. 20 electrodes were selected as output neurons: a half of them were used as left output neurons and the others were used as right output neurons. The spikes of the output neurons were used for calculating the left and right wheel speed of the mobile robot.

Experimental Setup

A robot (5 cm diameter) was placed in the 60 cm $\times$ 60 cm arena (Fig. 3.13), and the neural activity were recorded (1 h). We also recorded the right and left sensor input values of the robot, that indicate the distance between the robot and the wall on the environment.

We performed six experiments with three different cultures (26-61 days in vivo) (In one of the experiments, the technical issue occurred, thus the recorded data of the rest five experiments were analyzed.)

Figure 3.13: Experimental environment in real space. The robot was placed in the two dimensional square arena (60 cm × 60 cm).

3.3.2 Results

Evaluation of wall-avoidance behavior

This section focused on whether the mobile robot could improve wall-avoidance behavior autonomously.

We used a reaction time for evaluating the wall-avoidance behavior. The reaction time R was defined as:

$$R = t_2 - t_1$$

Here, t_1 denotes a time at which the sensory input value exceeds a threshold T and t_2 denotes a time at which the value is below T; T = 100 (the maximum sensor value is 950). When the robot collides with a wall or stands close to it, the sensor becomes activated, otherwise it receives a weaker signal. Therefore the lower reaction time indicates a higher performance of wall-avoidance. Here, the success was defined as: the reaction time is decreased by 30% or more. The average reaction time in the first 10 min and the last 10 min of the experiments were used for calculating the success rate.

The success rate of learning the wall avoidance behavior was 40% (2 out of 5 experiments; Fig. 3.14).

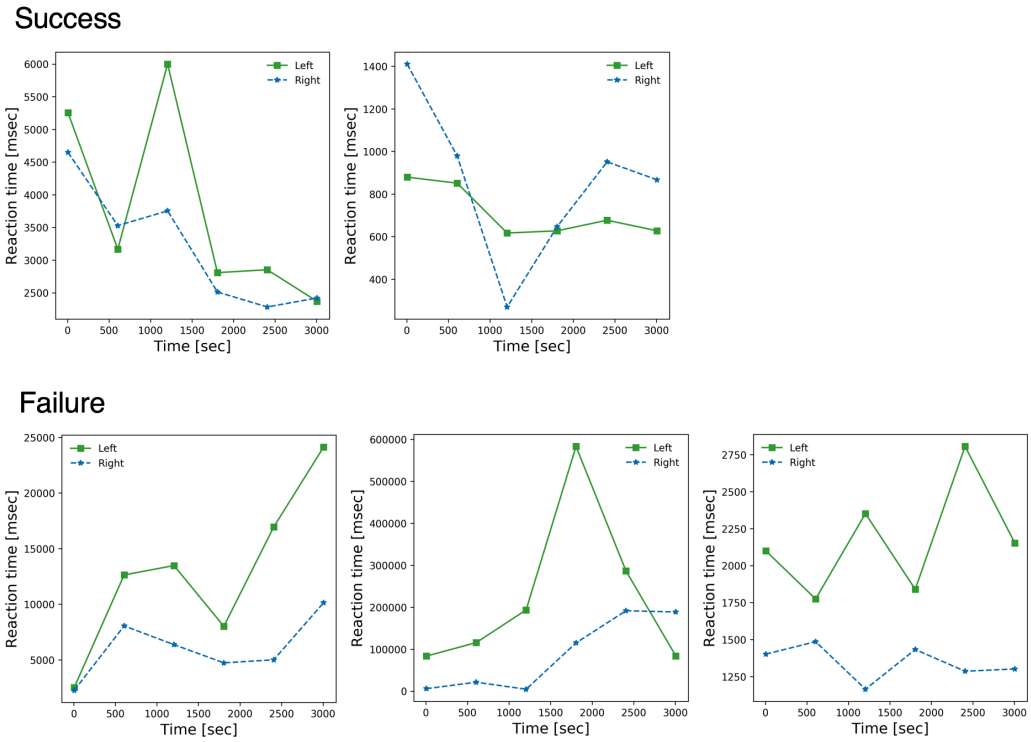

Figure 3.14: Time series of reaction time. The reaction time is defined as the duration between a time in which a sensor-input value exceeds a threshold (100) and the time in which the sensor-input value is below the threshold. Success was defined as: the reaction time is decreased by 30% or more. The average reaction time in the first 10 min and the last 10 min of the experiments were used for calculating the success rate.

We guess that one of the reasons of the lower success rate than one-dimensional virtual agent is that the learning of wall avoiding behavior in this section is more difficult than the wall-avoidance in the previous section, because the networks have to show the spatio-temporal pattern to avoid the wall (e.g., the left-wheel speed is much higher than the right to turn away from the wall).

In the previous simulation experiment with the similar experimental setup to the experiment here, the spiking networks can learn such a wall avoidance behavior well [31], and the previous study also shown that a burst suppression was required in such selective learning. Even in neuronal cultures, burst suppression is possible by inserting a noise input [61], however we did not apply the burst suppressing in this experiment, thus the bursting behavior

remained even during the stimulation phase (Fig. 3.15). In our future work, we need to examine whether learning such more complex behavior is improved by suppressing such bursting behavior.

In addition to the burst synchronization, the connectivity might be one of the causes of the failures. To evaluate the connectivity between the input neurons and the output neurons, we defined the connectivity measure as: the ratio of connections between input neurons and output neurons which have low time delay (40 ms). The time delay of each connection was estimated based on the argument of the maximum value of the cross-correlation between the spike time series of input neurons and output neurons in first 1 min in the experiments. Figure 3.16 shows a correlation between the connectivity measure and a success measure. The success measure was defined as improve rate of reaction time: $(R_0 - R_1)/R_0$ where R_0 denotes an average reaction time in the first 10 min of the experiment and R_1 denotes the last 10 min of the experiments. As shown in this figure the connectivity and the success measure have strong correlation. This suggest the connections with short time delay between input neurons and output neurons are required for this learning. Since the number of sample is small (n = 5), the p-value was a little bit high, thus additional experiments are required in our future research.

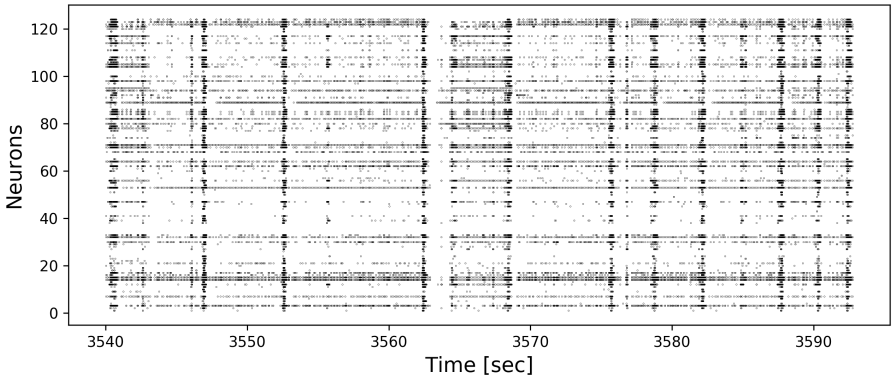

Figure 3.15: Burst synchronizations during the experiment. This figure shows a typical example of the experiments in the failure cases.

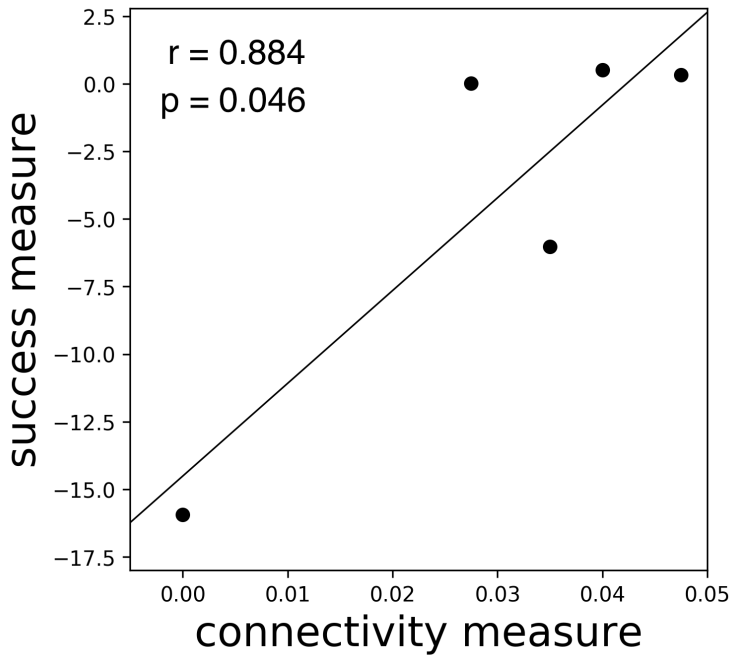

Figure 3.16: Correlation between the connectivity measure and the success measure. The correlation coefficient: $r = 0.884$ (p=0.046). Black line represents the linear regression.

Moreover, we found that the evoked-spikes tended to decrease in the all failure cases. This suggests that plasticity tends to work to separate uncontrollable inputs that are difficult to learn by LSA. We study more on this property in the sections below (see Section 5.2)

3.3.3 Conclusions

The learning experiment was performed more complex closed-loop than the previous section. The success rate of learning the wall avoidance behavior was 40%. As the success late of spatio-temporal task (selective learning) in the previous study is also low (about 50%) [29], it suggest that similar results are reproduced independently of the number of neurons. We hypothesized that there were two reasons of the low success rate: burst synchronization and connectivity.

3.4 Discussions

In this study, we demonstrated that SAA emerged in the embodied cultured neural networks even with the smaller number of neurons than the previous works [29]. We implemented the closed-loop system for connecting the cultured neuronal cells and the virtual or real robot, and conducted the experiment using this system. The results showed that the networks autonomously learned the simple behavior well in the virtual space (the success rate was 100%) and the more complex behavior in two dimensional real space partially (the success rate was 40%). The analysis of neural dynamics suggest that cultured neurons can learn a simple behavior by changing their network structures with the synaptic plasticity implying LSA based on STDP work in biological neural networks.

Shahaf and Marom showed that cultured networks can learn behaviors to stop stimuli from environment with large networks (10,000-50,000). In this chapter, we demonstrated the smaller network (about 100 neurons) learned the simple task to avoid the stimulation. These results suggest that SAA scales from small to large cultured networks. Although modifiability and stability proposed by Shahaf cannot emerge in very small number of neurons (e.g., 2 neurons), LSA scales from 2 neurons to 100 neurons [31] (In addition, in following sections, we also demonstrate that LSA scale to the larger spiking neural networks (see Section 6.1 and Section 6.2)). LSA is based on STDP which found in vitro and in vivo. Therefore we argue that LSA is a basic learning mechanism for SAA in neuronal cultures. To support the assertion, we need to conduct experiments with minimal size of neuronal cultures (e.g., 2 neurons).

In LSA, there must be some necessary conditions for structure of the closed-loop system. Here we discuss on the necessary conditions (Fig. 3.17). We identify one necessary condition for the network, Connectivity: Input neurons must be able to directly or indirectly transmit stimulation from the environment to the output neurons; additionally, there is a time constraint on this condition: the stimulation must reach the output neurons in less time than τ. τ is the time window during which the network can evaluate the consequences of a specific action that it took: for example, in a minimal network with 2 neurons, τ should be equal to the working time window of STDP (e.g., 20ms). The analysis of the correlation between the connectivity measure and the success measure (Fig. 3.16) supports that this conditions should be require for the neuronal cultures. We identify an additional condition for the environment, Controllability: There must be an output pattern from the network that can inhibit the stimulation through some action (e.g., turning away from the wall); there is also a time constraint on this condition: after

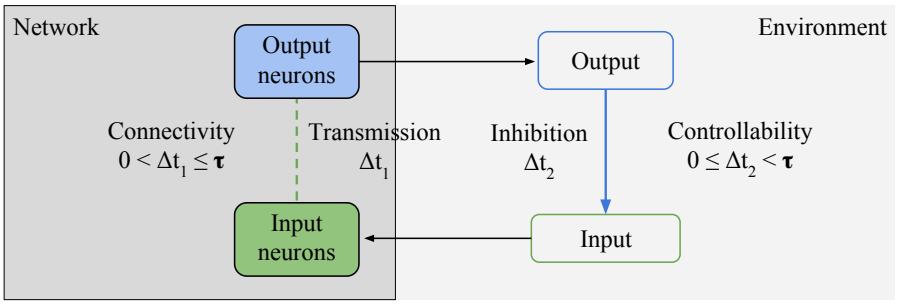

Figure 3.17: Necessary conditions for stimulation avoidance by action: Connectivity: Input neurons must be able to directly or indirectly transmit stimulation from the environment to the output neurons; the stimulation must reach the output neurons in less time than the time constant τ . Controllability: There must be an output pattern from the network that can inhibit the stimulation through some action; after the output is exhibited, the stimulation must stop in less time than τ.

the output is exhibited, the stimulation must stop in less time than the time constant τ. In the control experiment in Section 3.2 where Controllability is not respected (i.e., the stimulation is random and no output stop it), the networks did not learn the reactive behavior to avoid the stimulation. This result support this conditions should be require for the neuronal cultures. We argue that if these conditions are satisfied in closed-loop with embodied neural networks which have synaptic plasticity, such as STDP, then agents autonomously learn an action to avoid stimulation.

In this study, we focused on the wall-avoidance behavior. The networks learned the action to avoid the external stimulation based on LSA, however, emerging behaviors of agents should not limited to such avoiding behavior (e.g., object avoidance). Other adaptive behaviors can emerge from other couplings of embodiments and environment (see Section 6.3). For example, if an agent has sensors which has reverse dynamics of the distance sensors we used (i.e., if there is no object in front of the agent, stimulation start, and when an object appears in front of the agent, the stimulation stops.), the agent should learn object seeking behavior. Therefore, what kind of behaviors learned by agents to avoid stimulation depends on its embodiment. Coupling SAA with evolution of the embodiment, more various types of behavior should emerge. A drawback of the robot-neural platform here is that it uses the entire network of the neuronal cultures for making one style of behavior. This can be improved by using modular networks connecting multiple

neuronal cultures. We must study whether LSA works in such modularized networks in our future research.

Chapter 4

Stimulation Avoidance by Prediction

4.1 Minimal Predictive Networks in Spiking Neural Networks

4.1.1 Introduction

The previous studies showed that spiking neural networks can learn actions to avoid stimuli from environments [31]. In addition, the previous chapter showed that cultured networks learn actions in the same way. This behavior of the agents leads to maintain their homeostasis since the environmental changes can cause the changes of internal state of the agent. In the evolutionary perspective, next to such a learning reactive behaviors, prediction should be evolved. Because such a reactive agent who can only learn a reaction to avoid stimulation, cannot initiate an action before getting an actual stimulus. Thus, even if the agent can learn a reaction to an undesirable stimulus (e.g., damaging for the agent), the agent cannot avoid the stimulus before actually getting the stimuli. In such situations, if the agent can predict the incoming stimulus, the agent can initiate an avoiding behavior before getting damaged.

Prediction has recently been argued to be the definition of intelligence [7, 8] and predictive coding [12] attracts much attention. We consider that the predictive coding can be regarded as a function which reduces an error between an input signal and a top-down prediction. If such a predictive module is regarded as an agent, such reducing error is equal to a reducing the influence of stimuli from the environment to the internal system. Thus this can be regarded as Stimulation Avoidance by Prediction (SAP). At least,

the prediction is a powerful way to maintain homeostasis of an agent in a niche at macroscale where a lot of events in the environment are predictable within the individual lifetime (see Section 1.1).

There are some studies on prediction of temporal sequence in spiking neural networks [38, 39, 40, 41]. However, these predictive networks require some well-designed structures or other factors than STDP, thus it seems these predictive networks do not emerge in a primitive neural networks such as dissociated neuronal cultures we used in this thesis.

In this chapter, we focus on prediction in spiking neural networks based on STDP. We first demonstrate minimal predictive networks consists of 3-6 neurons can learn to predict some sequences of stimulation and finally demonstrate that even large random networks (100 neurons) without the well-designed structure can learn to predict a simple sequence of stimulation.

4.1.2 Methods

We performed some experiments using spiking neural networks. The model for spiking neuron proposed by Izhikevich [52] was used to simulate excitatory neurons and inhibitory neurons (see Section 2.3.1). The network size depends on the experiments (3-100). We used STDP as dynamics of neural plasticity (see Section 2.3.2). Note that, there was the neural plasticity by STDP not only for the connections between excitatory neurons but also for the connections to and from inhibitory neurons. Although STDP in inhibitory connections is still controversial, here we applied the same STDP function for all connections (A_{LTP}, A_{LTD}=1.0; τ_{LTP}, τ_{LTD}=20). In the most of the experiments, there was no synaptic time delay, however, in one of the experiments, there was a synaptic time delay between a timing of firing of presynaptic neuron and a timing of reaching the signal of the firing to postsynaptic neurons. In the experiments with the synaptic time delay, the function f_j which is the part of the functions for adding input current to each neuron (Eq 2.5; see Section 2.3.3) was modified for the synaptic time delay as:

$$f_j = \begin{cases} 1, & t - ts_j = td_{ij} \\ 0, & \text{otherwise.} \end{cases} \tag{4.1}$$

where t denotes the current simulation time, ts_j represents the timing of firing of presynaptic neuron j and td_{ij} represents the synaptic time delay between neuron i and neuron j. In the experiments with the synaptic delay, each pair of excitatory neurons and inhibitory neurons had 15 synapses (td of each synapse is varied from 1 to 15 ms).

Figure 4.1 shows the basic network topology for a minimal predictive network used in the following experiments. The network consisted of some excitatory neurons as input neurons and one inhibitory neuron, and the input neurons were not connected to each other, but all had an output and input weight with the same inhibitory neuron. The number of input neurons depended on the experiments (3-5). The weight values w between each neuron were initialized as 15.

We also used larger networks for evaluating its scalability. The network consisted of 80 excitatory neurons and 20 inhibitory neurons, and its topology was not well-designed and fully connected. The weight values w between each neuron were randomly initialized with uniform distributions as $0 < w < 5$ for connections from excitatory neurons, $-5 < w < 0$ for connections from inhibitory neurons. There was the synaptic plasticity in all connections except for the connections from the inhibitory neurons to the inhibitory neurons. There were three input neuron group (EG0-2), and 10 excitatory neurons were randomly selected for each input neuron group.

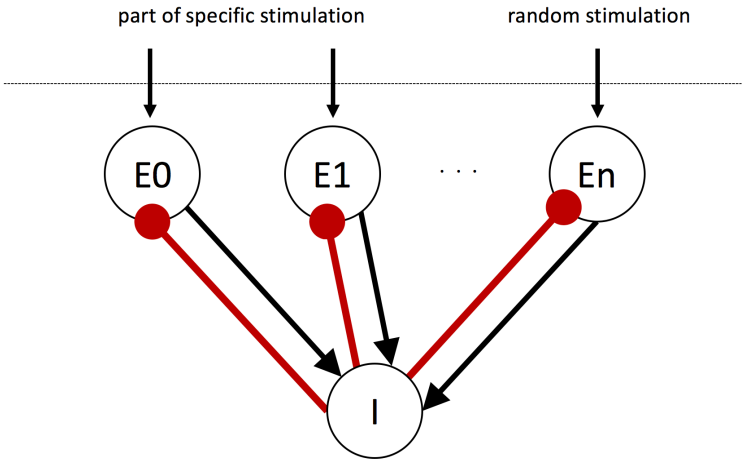

Figure 4.1: Basic topology for a minimal predictive network. The network consists of some excitatory neurons as input neurons and one inhibitory neuron. E represent the excitatory neurons. The number of E depends on experiments (3-5). E_n get random stimulus input as control; The other E_i get a part of specific stimulation sequence. I represents inhibitory neuron. The excitatory neurons were not connected to each other, but all had an output and input weight with the same inhibitory neuron. The black arrow represents excitatory synapses, The red arrows represents inhibitory synapses (the circles are the end of the arrows).

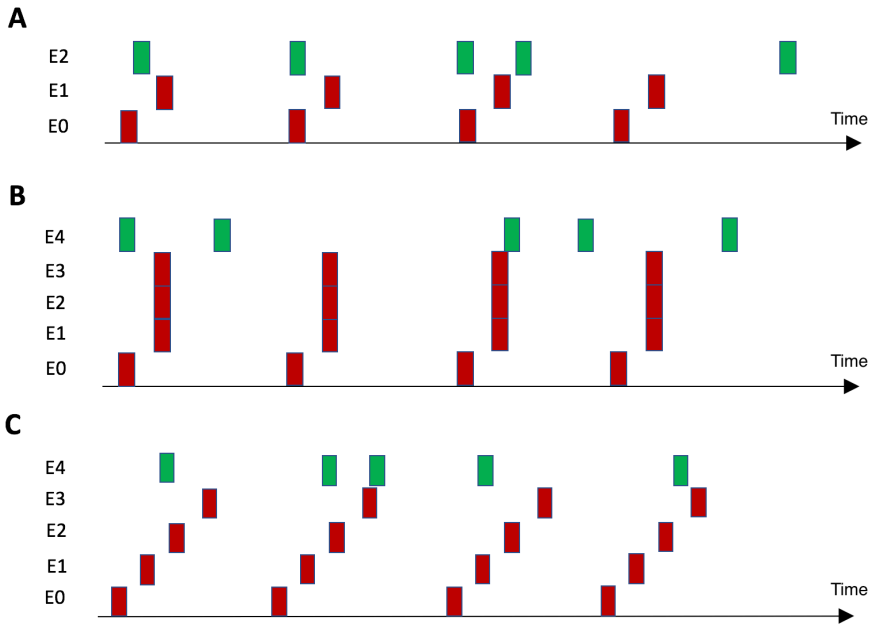

Figure 4.2: Stimulation sequence patterns of stimulation. The red squares represent stimuli that are part of the stimulation sequences. The green squares represent random stimuli. A: Minimal sequence. The sequence consisted of two stimuli (one anticipatory (first) stimulus and one target stimulus). B: Spatial sequence. The sequence consisted of four stimuli (one anticipatory (first) stimulus and three target stimuli). C: Temporal sequence. The sequence consisted of four stimuli (one anticipatory (first) stimulus and three target stimuli).

Three types of sequences of stimulation were applied to the networks: minimal sequence, spatial sequence, temporal sequence (Fig. 4.2). In the minimal sequence of stimulation, the sequence consisted of two stimuli (one anticipatory stimulus and one target stimulus): the anticipatory (first) stimulus was delivered to a specific input neuron and after a fixed time delay the target (second) stimulus was delivered to another specific input neuron (Fig. 4.2A). For the large networks, the minimal sequence consisted of two stimulus groups; Each stimulus group was delivered to the corresponding input neuron groups. In the spacial sequence of stimulation, the sequence consisted four stimuli (one anticipatory stimulus and three target stimuli): the anticipatory stimulus was delivered to a specific input neuron and after a fixed time delay the target stimuli were delivered to three other specific input neurons at the same time (Fig. 4.2B). In the temporal sequence of stimula-

tion, the sequence also consisted of four stimuli (one anticipatory stimulus and three target stimuli): the anticipatory stimulus was delivered to a specific input neuron and after a fixed time delay the first target stimulus was delivered to other specific input neuron and after a fixed time delay the second target stimulus was delivered to other specific input neuron and continued to the third target stimulus (Fig. 4.2C). For the every sequence patterns, the strength of each stimulus was 100 mV (10 mV for the large networks), the intervals of each stimulus was 10 ms that is sufficiently smaller than working time window of STDP (20ms), and the intervals of each sequence was 300 ms which is sufficiently larger than the working time window of STDP. In addition to those sequential stimulations, random stimulation with the same frequency to the sequential stimulations was delivered into another specific input neuron (or input neuron group for large networks) as the control.

4.1.3 Results

We first performed the experiment with the minimal network (3 neurons) with the minimal sequence of stimulation without synaptic time delay (n = 20). The networks consisted of three input neurons and one inhibitory neuron and the input neurons were not connected to each other, but all had an output and input weight with the same inhibitory neuron (Fig. 4.1). If the inputs are correctly predicted, the input neurons should be inhibited at the timing at which the stimulation is delivered. In Fig. 4.3, we look at the firing rates of all input neurons. We see that the firing rates of $E1$ which got the target stimulus decreased, whereas the firing rate of $E0$ which got the anticipatory stimulus and that of $E2$ which got the random stimulus did not change much: the networks gradually learned to predict the target stimuli, while the random stimulus was not predicted. This implies that the inhibitory neuron suppressed the $E1$ at the timing of the target stimuli, thus the network learned to predict the target stimuli.

Figure 4.4 and 4.5 show that the weight from the anticipatory neuron to the inhibitory neuron has increased, as did the weights from the inhibitory neuron to the target neurons. This path from $E0$ to I and from I to $E1$ is required to predict target stimuli to $E1$. In contrast, the weights to and from the random neuron did not change much.

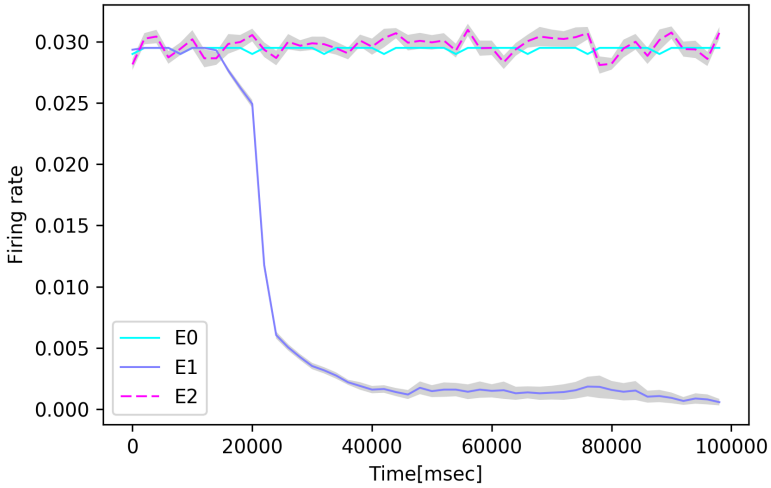

Figure 4.3: Time series of the firing rates of input neurons with the minimal sequence in the small network. The shaded regions represent standard errors of the mean (n = 20). The firing rate of $E1$ which got second stimulus of the minimal sequence decreased but $E0$ which got anticipatory stimulus and $E2$ which got random stimulus did not change much.

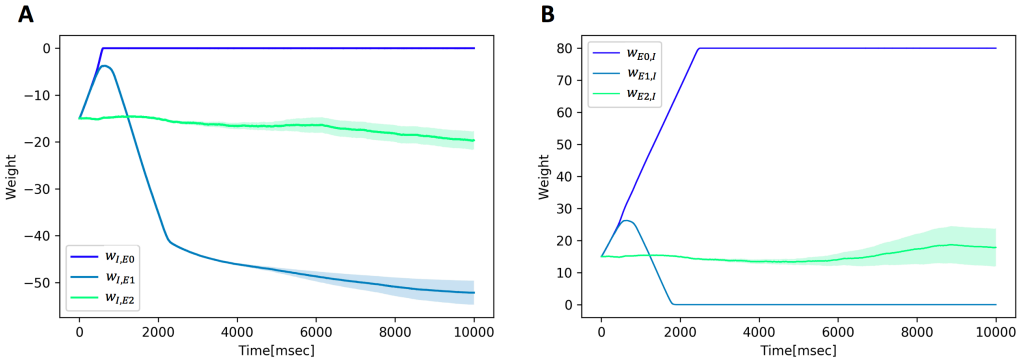

Figure 4.4: Time series of synaptic weights in small networks with minimal sequence. The shaded regions represent standard errors of the mean (n = 20). A: The weights from the inhibitory neurons to the excitatory neurons: $E0$, $E1$, $E2$. B: The weight from the excitatory neurons: $E0$, $E1$, $E2$, to the inhibitory neurons.

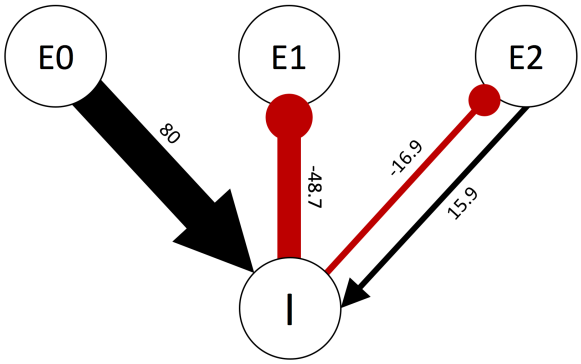

Figure 4.5: Obtained network in the small networks with the minimal sequence. A pathway from $E0$ to the inhibitory neuron and from the inhibitory neuron to $E1$ was strengthened. This path is required to predict target stimuli to $E1$. The black arrow represents excitatory synapses, The red arrows represents inhibitory synapses (the circles are the end of the arrows). The weight value of connections from inhibitory neurons is treated as the negative value for distinction from the excitatory synapse.

We evaluated whether the networks learn to predict more complex sequences: the spatial sequence and the temporal sequence. We first applied the spatial sequence of stimulation to the networks (n = 20) where two more input neurons were added to the minimal network above. Figure 4.6 shows that the firing rate of $E1$-3 which got the target stimuli of the spatial sequence decreased whereas E0 which got the anticipatory stimulus and $E4$ which got random stimulus did not change much: the networks gradually learned to predict the spatial sequence, while the random stimulus was not predicted. This implies the inhibitory neuron suppressed the $E1$-3 at the timing of the target stimuli and thus the network learned to predict the spatial sequence. On the other hand, the same networks did not learn to predict the temporal sequence, because there is only one inhibitory neuron, the network cannot learn to predict a longer temporal sequence than the minimal sequence consists of two stimuli.

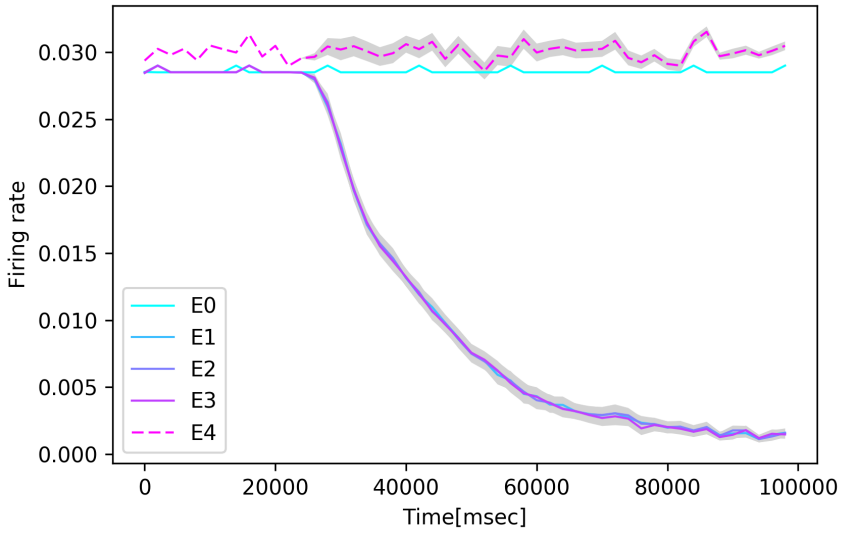

Figure 4.6: Time series of the firing rate of the input neurons with the spatial sequence. The shaded regions represent standard errors of the mean (n = 20). The firing rate of $E1$-3 which got the target stimuli of the spatial sequence decreased but $E0$ which got the anticipatory stimulus and $E4$ which got random stimulus did not change much.

For the networks to predict the longer temporal sequence, we hypothesized it must have more inhibitory neurons or more synapses between the input neurons and the inhibitory neuron which have different synaptic time delays, to encode information of such a longer temporal sequence. Therefore we applied 15 synapses between each input neuron and inhibitory neuron in the networks with different synaptic time delays (1-15ms).

Figure 4.7 shows that the firing rate of $E1$-3 which got the target stimuli of the temporal sequence decreased whereas $E0$ which got the anticipatory stimulus and $E4$ which got random stimulus did not change much: the networks gradually learned to predict the temporal sequence, while the random stimulus was not predicted. This implies inhibitory neuron suppressed the $E1$-3 at the timing of the target stimuli and thus the network with the synaptic time delay learned well to predict the temporal sequence.

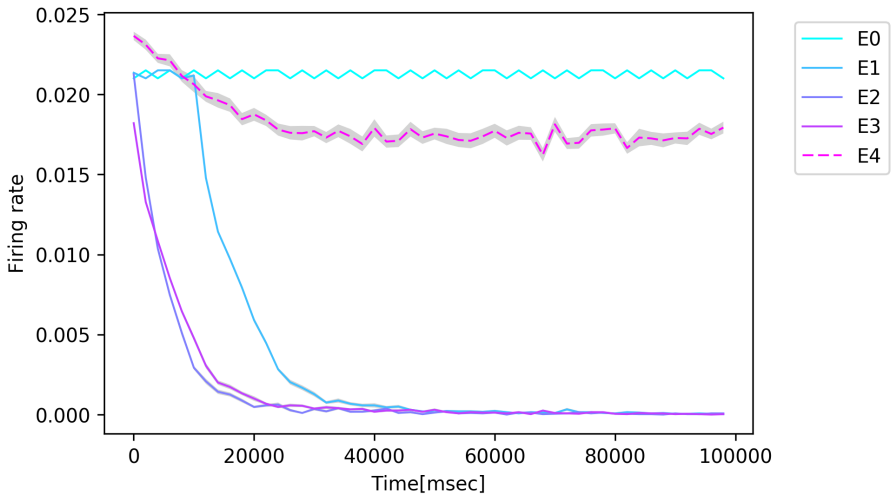

Figure 4.7: Time series of the firing rate of the input neurons with the temporal sequence. The shaded regions represent standard errors of the mean (n = 20). The firing rate of $E1$-3 which got the target stimuli of the temporal sequence decreased but $E0$ which got the anticipatory stimulus and $E4$ which got random stimulus did not change much.

We examined whether SAP scales to the large networks. We applied the minimal sequence of stimulation to the random networks consisted of 100 neurons (n = 20). Figure 4.8 shows that the networks gradually learned to predict the minimal sequence, while the random stimulation was not predicted. The green line represents the firing rate of hidden neurons that are the excitatory neurons exclude the input neurons. This can be regarded as the baseline of firing rate in the networks. The firing rate of $EG1$ (excitatory neuron group 1) which got target stimuli of minimal sequence gradually decreased to near the value of the hidden neurons, thus, the accuracy of the prediction was not bad.

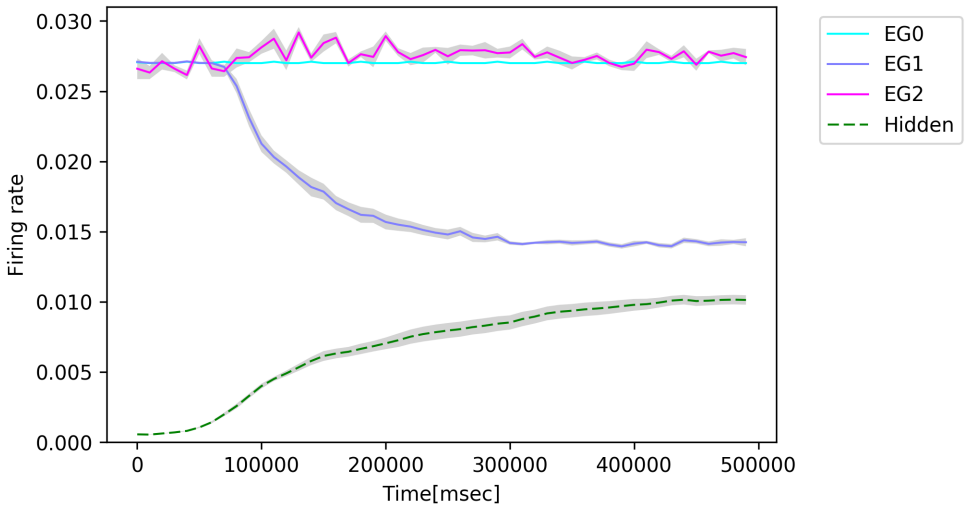

Figure 4.8: Time series of the firing rates of the input neurons with the minimal sequence in the large networks (100 neurons). The shaded regions represent standard errors of the mean (n = 20). The firing rate of $EG1$ which got the target stimuli decreased but $EG0$ which got anticipatory stimuli and $E2$ which got random stimuli did not change much. The firing rate of hidden neurons can be regarded as a baseline of firing rate in the network.

Figure 4.9 shows the typical example of the raster plot of spikes of each neuron in the large network. The figures show that in the first phase of the experiment, almost all neurons in $EG1$ fired at the timing they got the target stimuli, but in the last phase of the experiment, the neurons did not much fire at the timing and their firing patterns were almost same with hidden neurons. This implies that inhibitory neurons suppressed the firing of the neurons in $EG1$ at the timing of the target stimuli, suggesting the networks learned well to predict the minimal sequence of stimulation.

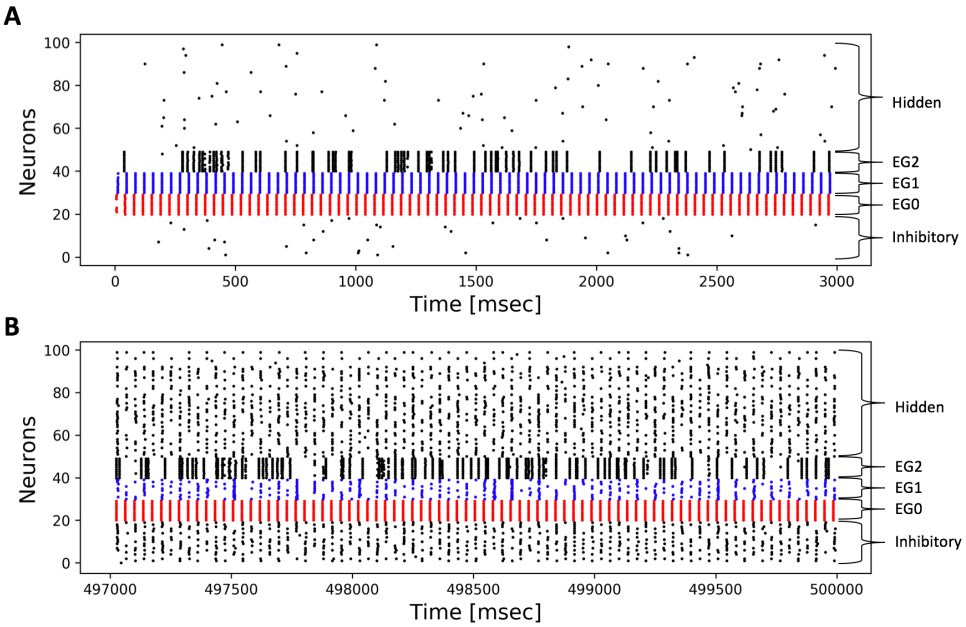

Figure 4.9: Raster plots of spikes in large network. Each dot represents one spike: red dots represent spikes of $EG0$ which got the anticipatory stimuli, blue dots represent spikes of $EG1$ which get the target stimuli, black dots represent spikes of the other groups ($EG2$, Hidden and Inhibitory). A: Spikes in the first 3,000 msec. Almost all neurons in $EG1$ fired at the timing they got the target stimuli. B: Spikes in the last 3,000 msec. The neurons in EG1 did not much fire at the timing and their firing patterns were almost the same as hidden neurons.

Figure 4.11 shows the time series of the synaptic weights between each neuron groups and its dynamics looks similar to that of the small networks. Thus this suggest that the mechanisms for prediction in the small networks scaled to the large networks.

Figure 4.10 shows the typical example of the initial weights and the last weights of the network, and Fig. 4.12 shows obtained topology of the network. Importantly, there was a strong pathway from EG0 to inhibitory neurons and from inhibitory neurons to $EG1$; this path is required for the prediction. In addition, there was a pathway from $EG0$ to hidden neurons, hidden neurons to inhibitory neurons and inhibitory neurons to $EG1$; this pathway might be required to adjust the timing of suppression of $EG1$.

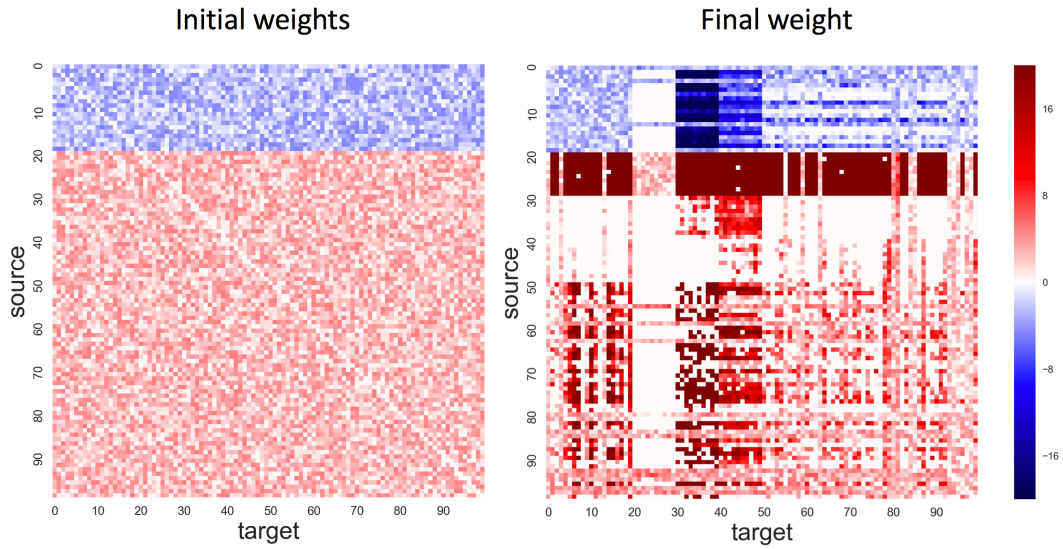

Figure 4.10: Initial weights and final weights of the large network. Color represents connection weight. The weight value from inhibitory neurons treated as negative value for distinction from excitatory synapse.

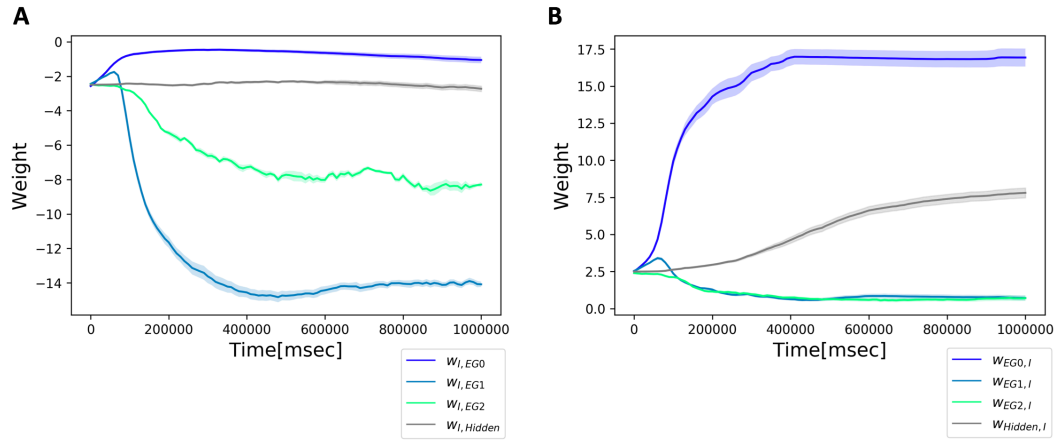

Figure 4.11: Time series of synaptic weights in large networks with minimal sequence. A: The weights from the inhibitory neurons to the excitatory neuron groups: $EG0$, $EG1$, $EG2$, and hidden neurons. B: The weight from the excitatory neuron groups: $EG0$, $EG1$, $EG2$, and hidden neurons to the inhibitory neurons.

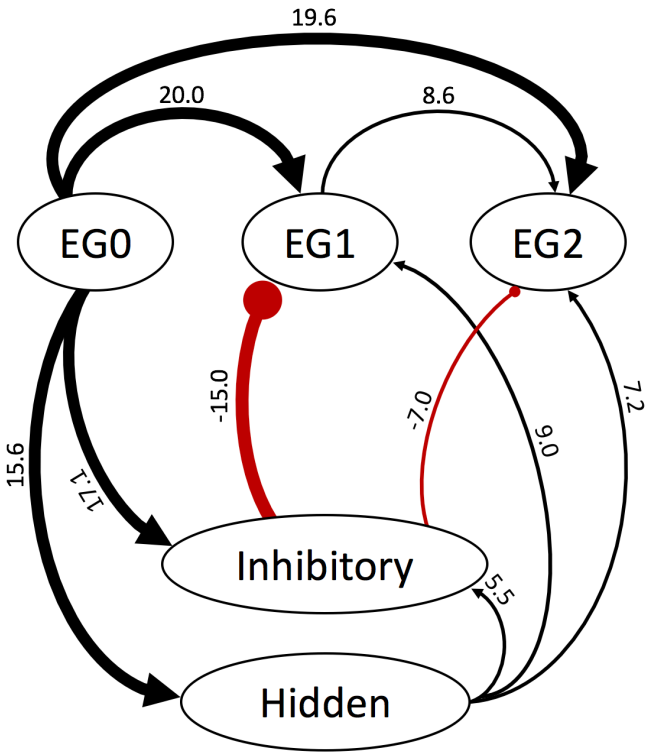

Figure 4.12: Obtained network from the large network. A connection between each group was depicted, if its absolute mean weight value is larger than 5.0: thick line represents larger weight value than 10.0, thin line represents less value than 10.0, black lines represent excitatory synapses and red lines represent inhibitory synapses. The weight values of connections from inhibitory neurons is represented with red arrows and treated as negative value for distinction from excitatory synapse.

These results show that spiking neural networks learn to predict some simple sequences of stimulation based on STDP.

4.1.4 Discussion

In this study, we demonstrated that the spiking neural networks learned to predict some causal inputs from the environment, based on STDP. In the prediction in this study, the input neurons which got the target stimuli was suppressed by the inhibitory neuron at the timing of receiving the stimulus. This leads to decreasing the influence of stimulation from the environment to the networks. Thus it can be regarded as SAP. We also showed this property

scales to the large random networks (100 neurons) without preparing well-designed structures and other particular functions than STDP. Therefore this suggests that SAP can also emerge in dissociated neuronal cultures like we used in the previous chapter. We must study further on the learning prediction in neuronal cultures in our future research.

As Stimulation Avoidance by Action (SAA) requires some necessary conditions for the structure of the closed-loop (see Section 3.4), SAP must have some necessary conditions. Here we first discuss the minimal conditions for the predictive network, then we also discuss how to integrate the predictive networks with the reactive agent (see Section 3.4) to a proactive agent.

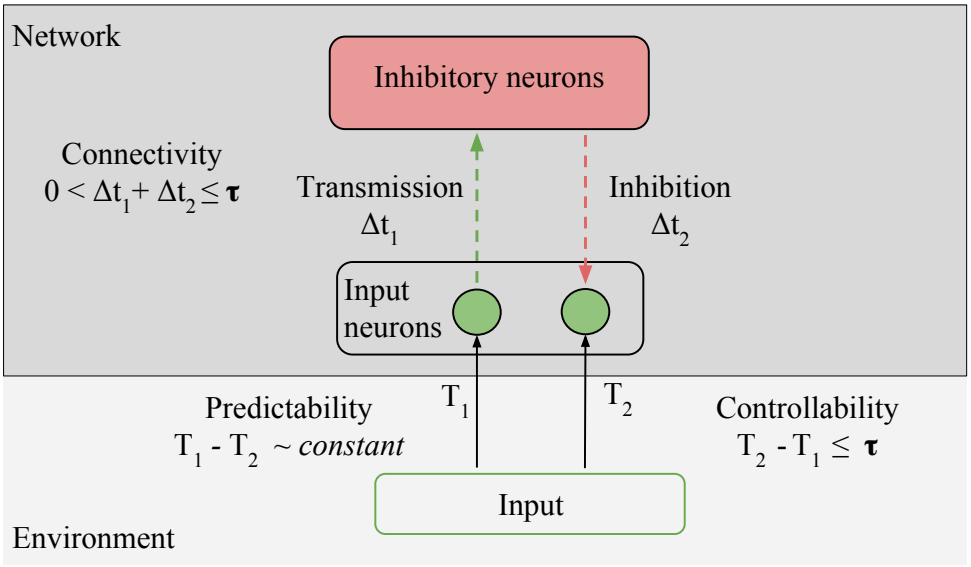

Figure 4.13: Conditions for predictive networks: Predictability: the time delay between the stimulus at T_1 (unpredictable anticipatory stimulus) and the stimulus at T_2 (predictable target stimulus) must be constant, i.e., the environment must provide predictable stimuli in order for predictions to be learned. Controllability: The time delay between the stimulus at T_1 and the stimulus at T_2 (predictable target stimulus) must be less than τ; τ is the time window during which the network can evaluate the consequences of a specific action that it took. Connectivity: Input neurons must be able to directly or indirectly transmit and receive stimulation from the inhibitory neurons.

For the minimal predictive networks, some modifications were required to the conditions for the learning action (see Section 3.4). In Fig. 4.13, compared to Fig. 3.17 the role of output has been shifted from the environment

into the network itself. Instead of output neurons acting upon the environment inhibiting external stimulation in SAA, now inhibitory neurons must act on the input neurons inhibiting external stimulation. This is a minimal shift in perspective, which only requires to add inhibitory neurons and its synaptic plasticity in the network. The dynamics of the prediction as follows: inhibitory neurons must learn to inhibit input neurons at time T_2; the input neurons must learn to control inhibitory neurons at time T_1 in order to make inhibition at T_2 possible. On the environment, we have two condition, Predictability: the time delay between the stimulus at T_1 (unpredictable anticipatory stimulus) and the stimulus at T_2 (predictable target stimulus) must be constant, i.e., the environment must provide predictable stimuli in order for predictions to be learned, Controllability: The time delay between the stimulus at T_1 and the stimulus at T_2 (predictable target stimulus) must be less than τ; τ is the time window during which the network can evaluate the consequences of a specific action that it took. On the network, we still have the Connectivity condition, this time involving inhibitory neurons: Input neurons must be able to directly or indirectly transmit and receive stimulation from the inhibitory neurons.

In a evolutionary perspective, it is difficult to discuss a prediction itself without an action. To enhance an adaptability by prediction for the reactive agent which learn the reactive behavior to stimulus discussed in the previous chapter, the structure of the reactive agent and the predictive network must be integrated to a single agent. Here, we discuss how the predictive agent and the reactive agent can be combined to have proactive behavior: actions that are based on predictions.

Figure 4.14 shows how a conceptual shift on the conditions for learning predictions (Fig. 4.13) can give the conditions for proactive behaviors. We insert output neurons between the input neurons and inhibitory neurons. Their position is similar as for the reactive agent (Fig. 3.17), but this time the "action on the environment that stops stimulation" is an action on inhibitory neurons that also ends up stopping stimulation, through the same prediction process as in Fig. 4.13. Here for clarity, the output and input neurons are separated, but in the simplest case, a neuron can act both as an input (by receiving external stimulation) and as an output (by outputting directly to an inhibitory neuron). More interesting cases are possible when we have several neurons: predictions can aggregate input from several input neurons, and predictions can be stacked hierarchically by aggregating several output neurons. As we have seen, input neurons receiving a predictable sequence of stimulation become inhibited by inhibitory neurons, except for the neurons receiving anticipatory stimuli. Therefore any proactive behavior (actions based on prediction) must rely on information from these anticipatory

neurons.

We need to study further to show such a proactive agent actually emerge with the proposed conditions here in the future works.

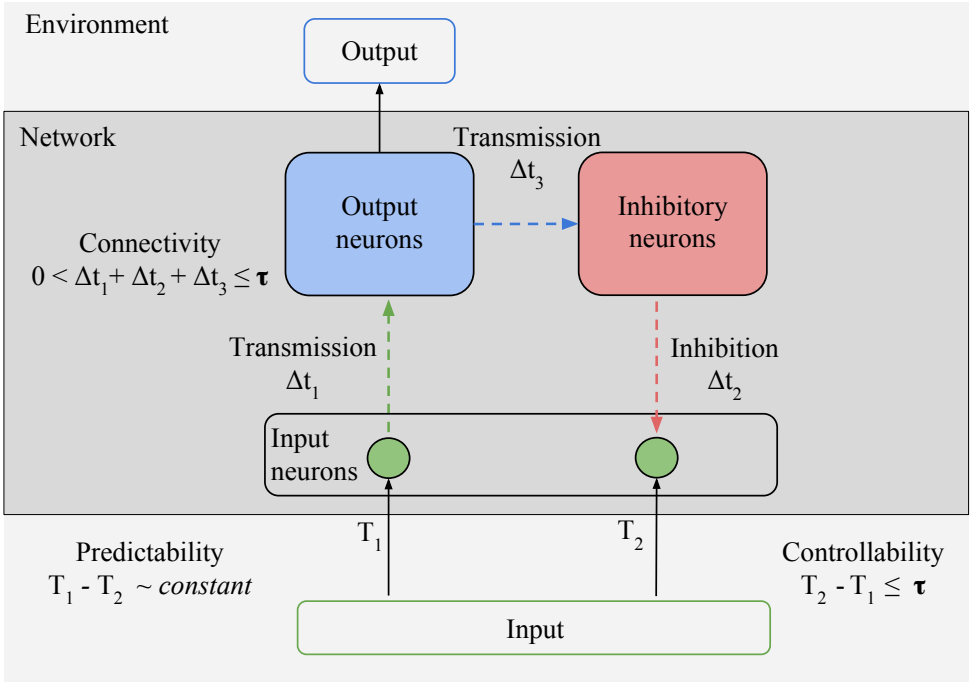

Figure 4.14: Conditions to learn proactive behavior: Predictability: the time delay between the stimulus at T_1 (unpredictable anticipatory stimulus) and the stimulus at T_2 (predictable target stimulus) must be constant, i.e., the environment must provide predictable stimuli in order for predictions to be learned. Controllability: The time delay between the stimulus at T_1 and the stimulus at T_2 (predictable target stimulus) must be less than τ; τ is the time window during which the network can evaluate the consequences of a specific action that it took. Connectivity: Transmission from input neurons to other input neurons through output neurons and inhibitory neurons must be less than τ.

Chapter 5

Stimulation Avoidance by Selection

5.1 Introduction

In the previous chapters, we showed that Stimulation Avoidance by Action (SAA) and Stimulation Avoidance by Prediction (SAP) can emerge in cultured neural networks and spiking neural networks. In this study, as the results of experiments using the cultured networks (see Chapter 3), we found that if the network cannot learn a behavior that stops an external stimulation, its responses to the stimulation are gradually suppressed as if it separates the uncontrollable input neurons. Such kind of response to repeated stimulation has been known as a neural adaptation or a sensory adaptation in neuroscience [78]. One of the possible mechanism is a synaptic fatigue [79]: Repeated stimulations cause exhaustion of neurotransmitters in synapse leading weakening the response to the stimulation. However, it seems that the observed behavior to the repeated stimulation in our experiment is not caused by the synaptic fatigue but synaptic plasticity, especially LTD. In addition, we demonstrate such kind of neural adaptation is reproduced by spiking neural networks without synaptic fatigue but with STDP. It means that in addition to SAA and SAP, there is the third property to avoid the stimulation, where the network avoids the stimulation by weakening the connection strength from the input neurons if it is hard to learn a behavior to avoid the stimulation.

In this Chapter, we focus on the Stimulation Avoidance by Selection (SAS) in both cultured neural networks and spiking neural networks. The selection behavior here means that weights from input neurons with controllable inputs is reinforced, and weights from input neurons with uncontrollable

inputs is depressed.

5.2 Synaptic Selection in Neuronal Cell Cultures

In this section, we focus on SAS in cultured neural networks showing the further analysis of the experimental data in Chapter 3.

5.2.1 Methods

We analyzed the neural activity in the experiment using the cultured neurons that we previously conducted (see Chapter 3).

In the analysis, we used two experimental data: experiment 1: SAA in one-dimensional virtual space (see Section 3.2); experiment 2: SAA in two-dimensional real space (see Section 3.3). In both cases, the stimulation was delivered as the sensor input when the robot contacted with the wall, and the sensor input stopped when the robot moves away from the wall.

In experiment 1, without stimulation the agent moves forward at a constant speed; If the agent approaches a wall, the sensors stimulate the input zone and if more than 5 out of 10 output neurons fire within 20 to 40 ms after the stimulation, the agent turns away from the wall by rotating at 180 degrees [80] (see Section 3.2). We also conducted the control experiment under the condition that the stimulus input stopped at random regardless of their neural activity and the other settings are the same as the experiments.

In experiment 2, the robot as the embodiment of the agent had two infrared sensors on the front left and right of the body, and two motor wheels on the left and the right side of its body. As the agent approaches the wall, the sensor values become higher and the stimulus input was delivered. The number of firings of output neurons was the coefficient of left and right motor speeds of the robot, thus in order to avoid the wall, the robot needs to control the motor speeds [81] (see Section 3.3). It is a more difficult condition compared with the experiment 1.

In order to investigate the neural behavior in conditions where it is difficult to learn a stimulus-avoiding behavior, we focused on the results of the control condition in experiment 1 in which the robot cannot learn the behavior and the failure cases of experiment 2 in which it is more difficult to learn the behavior than experiment 1.

5.2.2 Results

In Section 3.2, we showed that the embodied cultured neural networks learned to avoid the wall in virtual one-dimensional space. Interestingly, however, we found that the stimulus-evoked spikes decreased remarkably in some cases of the control experiments in which the stimulation was randomly removed regardless of their activity (Fig.5.1).

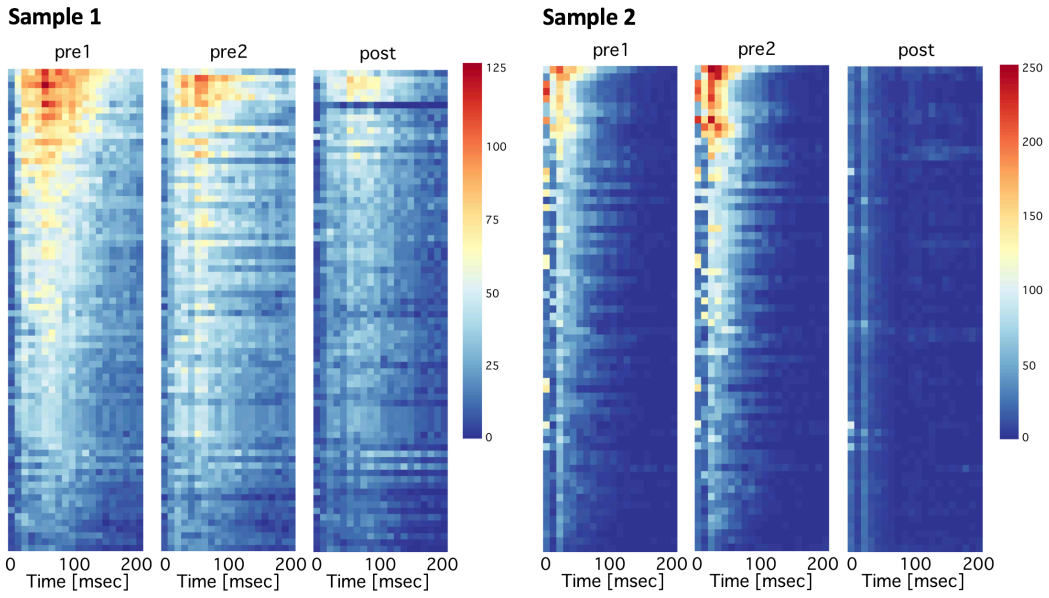

Figure 5.1: Evoked spikes of all neurons in the control experiments. The color indicates the number of spikes of each neuron. pre1: the first 5 min of the experiment; pre2: the second 5 min; post: the last 5 min. In both examples, the evoked spikes at the end of the experiment (post) were decreased.

Figure 5.2 shows mean evoked firing rate with standard errors. The evoked firing rate except for input neurons (Fig. 5.2A) consists of spikes in 200 msec after each stimulus. The evoked firing rate of input neurons (Fig. 5.2B) consists of spikes in 50 msec after each stimulus, since we want to focus on the evoked spikes by each stimulus rather than the evoked spikes by the feedback from other neurons. Besides the qualitative results above, the statistical results show that mean evoked firing rate of all neurons except for input neurons significantly decreased in the last 5 min of experiment relative to the first 5 min (Wilcoxon signed-rank test, $n = 6$, $p = 0.012$) (Fig. 5.2A), on the other hand, that of input neurons did not change significantly (Fig. 5.2B). These results imply that decreasing evoked spikes was

not caused by decreasing firing rate of input neurons.

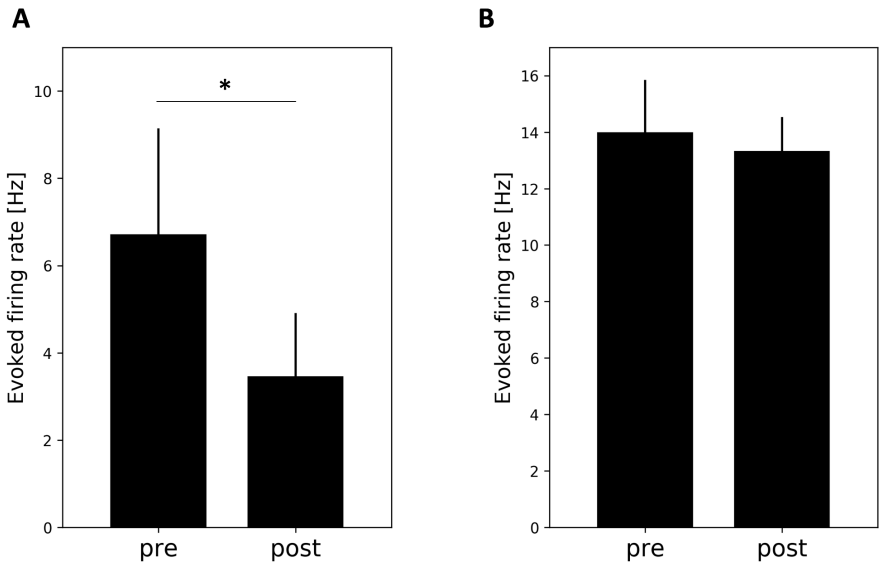

Figure 5.2: Mean evoked firing rates in the controls. Error bars represent standard errors of the mean (n = 6). pre: first 5 min in the experiment; post: the last 5 min. A: mean evoked firing rate of all neurons except for input neurons ($p<0.03$). B: mean evoked firing rate of input neurons.

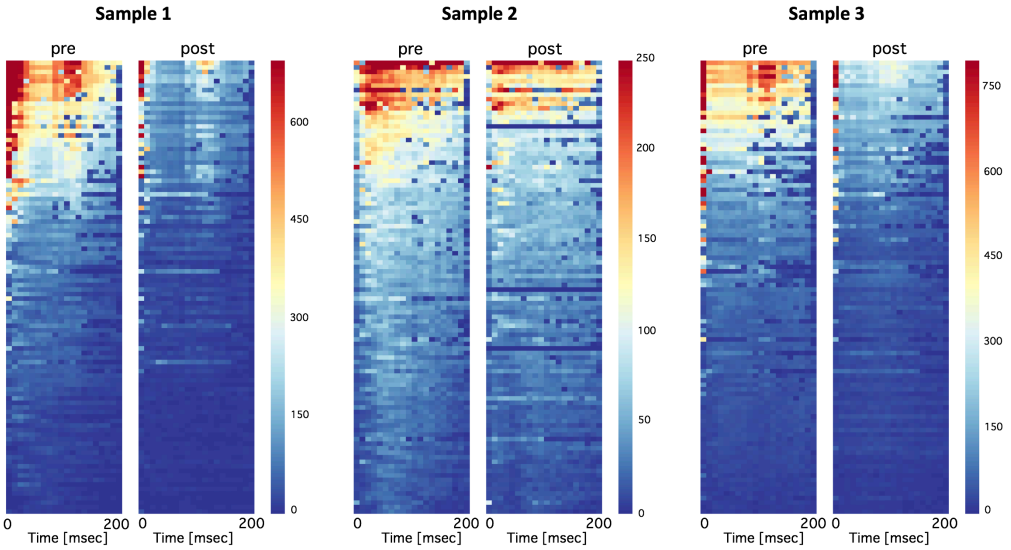

Figure 5.3: Evoked spikes of all neurons in failure cases in experiment 2. The color indicates the number of spikes of each neuron. pre: the first 5 min of the experiment; post: the last 5 min. In all examples, the evoked spikes at the end of the experiment (post) decreased.

In experiment 2, the success rate of the learning of wall avoidance behavior in two-dimensional real space was 40% (2 out of 5 experiments) (see Section 3.3). We focus on dynamics of the failure cases for examining whether similar dynamics with the control in experiment 1 are observed in this more difficult task case.

As shown in Fig. 5.3, the stimulus-evoked spikes decreased gradually in the failure cases. These results look similar to the result of the controls in experiment 1.

Figure 5.4 shows the evoked firing rates of input neurons and other neurons. The evoked firing rate of output neurons consisted of spikes in 200 msec after each stimulus. The evoked firing rate of input neurons consisted of spikes in 50 msec after each stimulus, since we want to focus on the evoked spikes by stimuli itself rather than the evoked spikes by feedback of the other neurons. As shown in Fig.5.4, the evoked firing rate of other neurons was gradually decreased, but the evoked firing rates of input neurons did not decrease.

Statistical results show that mean evoked firing rate of all neurons except for input neurons significantly decreased in the last 5 min of the experiment relative to the first 5 min (Wilcoxon signed-rank test, $n = 3$, $p = 0.023$) (Fig. 5.5A), on the other hand, mean evoked firing rate of input neurons did not change significantly between the first 5 min in the experiment and the

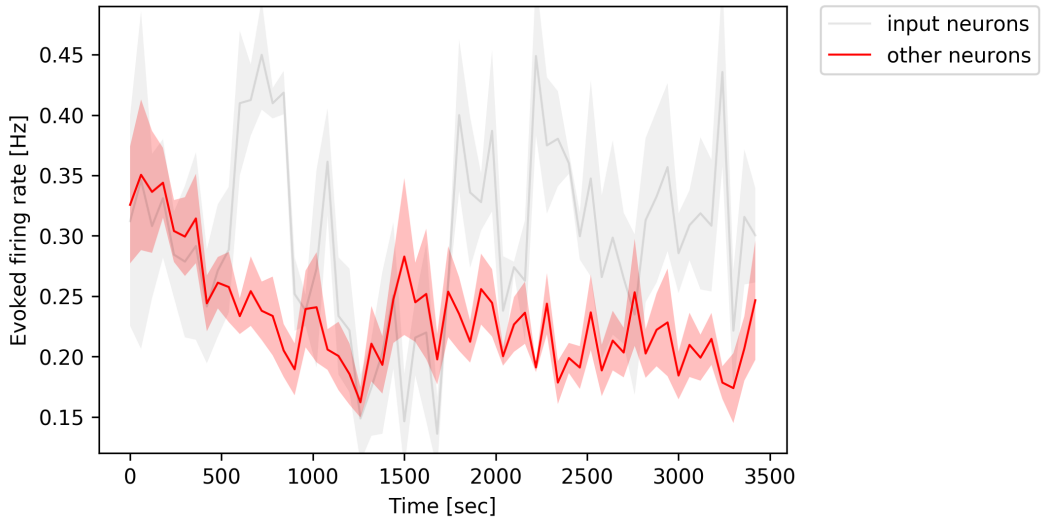

Figure 5.4: Time series of evoked firing rates in the failure cases. The values represent the evoked firing rates with standard errors (n = 3) in the specific time window after delivering a stimulus (50 ms for input neurons and 200 ms for other neurons).

last 5 min. (Fig. 5.5B).

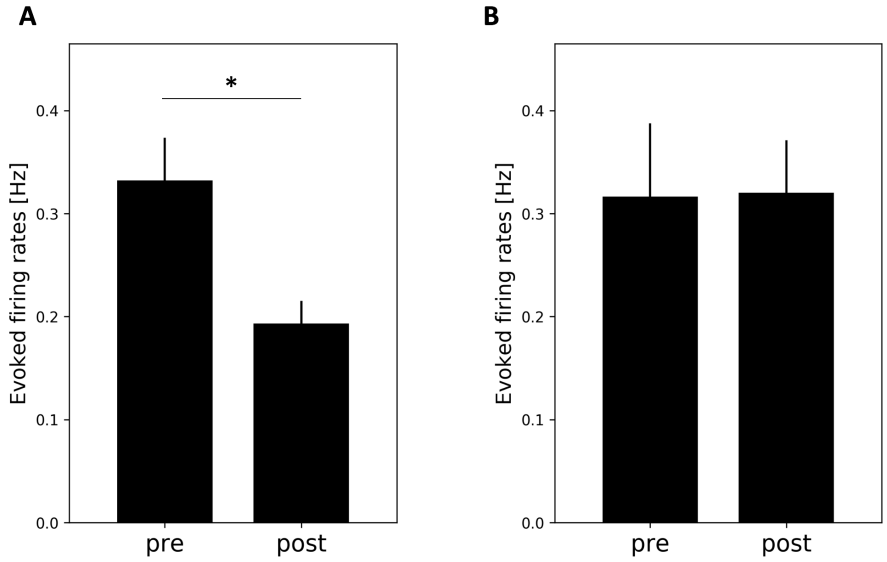

Figure 5.5: Mean evoked firing rates in the failure cases. Error bars represent standard errors of the mean (n = 3). pre: first 5 min of the experiment; post: the last 5 min. A: mean evoked firing rate of all neurons except for input neurons. ($p<0.03$). B: mean evoked firing rate of input neurons.

These results imply that the decreasing of firing rates of all neurons except for input neurons was not caused by decreasing of evoked spikes of the input neurons but decreasing the synaptic connection strength from the input neurons to the others. Therefore, it suggests that embodied cultured neural networks try to learn an action to avoid external stimulation, but if the learning is difficult, the networks tend to ignore the external stimulation weakening the connection strength from the input neurons.

5.2.3 Conclusions

In this study, we demonstrated that if it is impossible or difficult to learn the action to avoid stimulation, the response to the stimulation decreased in the cultured networks. It was not caused by decreasing firing rate of input neurons suggesting it was not caused by the synaptic fatigue but caused by decreasing synaptic weight by LTD. Therefore, the neural dynamics works to separate the input neurons which has uncontrollable input, as if it prohibits the rest of the network from getting information from the uncontrollable neurons.

5.3 Synaptic Selection in Spiking Neural Networks with Asymmetric-STDP

In this section, we focus on SAS in spiking neural networks. In previous chapters, although we used symmetric-STDP (LTP and LTD is rotational symmetry), STDP broadly found in vitro [32] is rotational asymmetric (e.g., the working time window of LTD is larger than LTP). Applying such asymmetric-STDP, we show the spiking neural networks reproduce the neural behaviors observed in the cultured networks in the previous section.

5.3.1 Methods

We performed a learning experiment using spiking neural networks, that has similar experimental settings to experiment 1 in the previous section, and the control experiments where the stimulation was randomly removed regardless of the output of the network (i.e., the network can not learn the behavior to avoid the stimulation in the control conditions).

The model for spiking neuron proposed by Izhikevich [52] was used to simulate excitatory neurons and inhibitory neurons (see Section 2.3.1). The network consisted of 80 excitatory neurons and 20 inhibitory neurons. This ratio of 20% inhibitory neurons is standard in simulations [63, 52] and similar to biological values [82]. The excitatory neurons were divided into three groups: Input (10 neurons), Output (10 neurons), Hidden (60 neurons). The networks were fully connected; The weight values w between each neuron were randomly initialized with uniform distributions as $0 < w < 5$ for excitatory neurons, $-5 < w < 0$ for inhibitory neurons. Only connections between excitatory neurons had synaptic plasticity based on STDP; The weight values of other connections did not change. Zero-mean Gaussian noise m with a standard deviation $\sigma = 3$ mV was delivered to each neuron at each time step for a spontaneous firing.

The number of input neurons was fixed to 10. The input neurons got the stimulation from the environment (frequency: 100 Hz; strength: 10 mV). If more than 5 out of 10 output neurons fired within 10 ms after the stimulation, the stimulation was removed for 1,000-2,000 ms (randomly chosen every time). In non-learnable condition as a control experiment, the stimulation was randomly removed regardless of the firing of output neurons, thus the network cannot learn any behaviors to avoid the stimulation.

Various shapes of STDP function were used for the synaptic dynamics (see Section 2.3.2). Here, as symmetric-STDP, the parameters of LTP and LTD that were symmetric (A_{LTP}, A_{LTD}=1.0; τ_{LTP}, τ_{LTP}=20) were used;

as asymmetric-STDP, the dynamics of LTP and LTD that were asymmetric (A_{LTP}=1.0, A_{LTD}=0.8-1.5; τ_{LTP}=20, τ_{LTD}=20-30) were used. A_{LTP} and A_{LTD} are the parameter for the strength of the effect of LTP and LTD in STDP. τ_{LTP} and τ_{LTD} are the parameter for the working time window of LTP and LTD in STDP. Figure 5.6 shows the curve of the symmetric-STDP and some examples of the asymmetric-STDP. In addition, we used STP and decay function for the synaptic dynamics explained above (see Section 2.3.3).

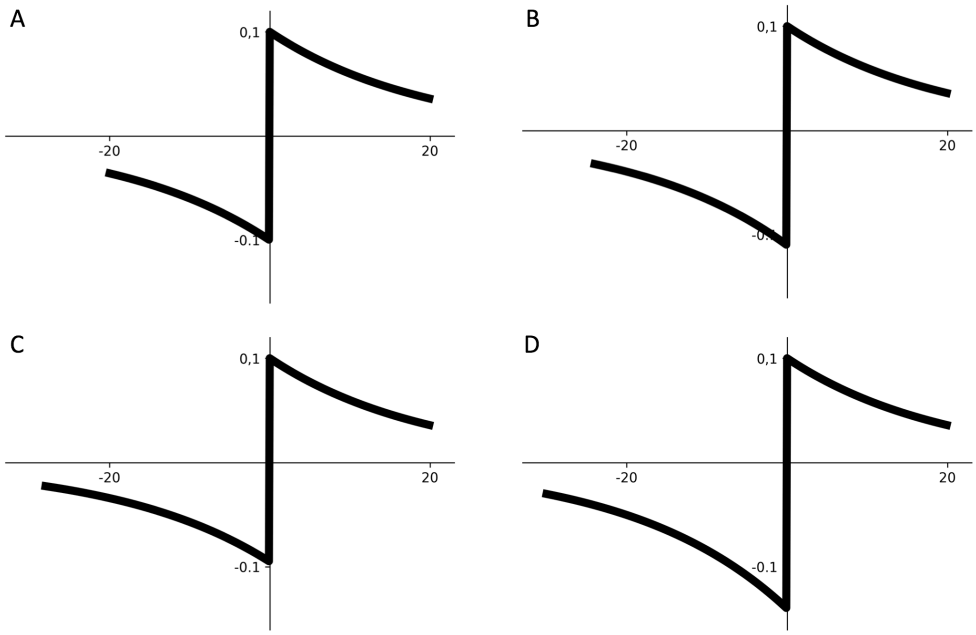

Figure 5.6: Parametric variations of STDP curve. A: Symmetric-STDP: A_{LTP}, A_{LTD}=1.0; τ_{LTP}, τ_{LTP}=20. B: Example of asymmetric-STDP: A_{LTP}=1.0, A_{LTD}=1.1; τ_{LTP}=20, τ_{LTP}=24. C: Example of asymmetric-STDP: A_{LTP}=1.0, A_{LTD}=0.95; τ_{LTP}=20, τ_{LTP}=28. D: Example of asymmetric-STDP: A_{LTP}=1.0, A_{LTD}=1.4; τ_{LTP}=20, τ_{LTP}=30.

5.3.2 Results

We performed the experiments for examining whether spiking neural networks with STDP reproduce SAS observed in the cultured networks.

With the symmetric-STDP (Fig. 5.6A), the synaptic weight from the input neurons increased in both the learning experiments and the control experiments (Fig. 5.7A). The results show there was no dynamics leading to separation of the uncontrollable neurons with the symmetric-STDP. On the other hand, with asymmetric-STDP (A_{LTP}=1.0, A_{LTD}=1.1; τ_{LTP}=20,

τ_{LTP}=24; Fig. 5.6B), the synaptic weights from the input neurons to the other neurons increased in the learning experiment (learnable), but it decreased in the control experiment (non-learnable; Fig. 5.7B). Therefore, we found that

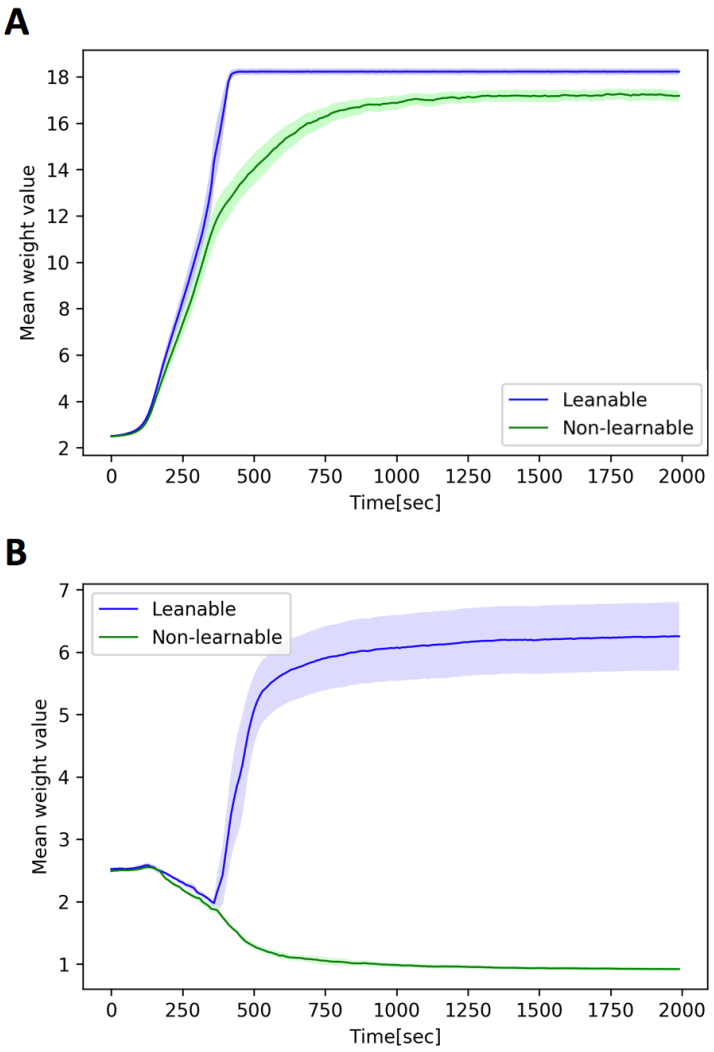

Figure 5.7: Time series of the mean connection strength from the input neurons to the other neurons. In the learnable case, the networks can learn the behavior to avoid the stimulation. In non-learnable case, stimulation were randomly removed, thus the networks cannot learn the behavior. The shaded regions represent standard errors of the mean (n = 20). A: The results of symmetric-STDP. B: The results of asymmetric-STDP: A_{LTP}=1.0, A_{LTD}=1.1; τ_{LTP}=20, τ_{LTP}=24.

spiking neural networks with asymmetric-STDP can reproduce the behavior to separate the uncontrollable input neurons observed in the experiments using the cultured networks.

For examining how this behavior works, we explored the parameter space changing A_{LTD} and τ_{LTD}. A_{LTD} represents the parameter for the strength of LTD. τ_{LTD} represents the working time window of LTD.

In Fig. 5.8A, the color indecates the value of selection indicator (SI) which is defined as:

$$SI = LW_{input} - NW_{input} \tag{5.1}$$

where LW_{input} denotes an average weights from input neurons to other neurons in the learnable cases and NW_{input} denotes that in the non-learnable cases. SI can be regarded as the indicator of selection behavior: the higher value indicates a higher selection tendency. The results of asymmetric-STDP with the parameters: A_{LTP}=1.0, A_{LTD}=1.1; τ_{LTP}=20, τ_{LTP}=24 (Fig. 5.6B), shows the maximum value of SI in this parameter space. With the parameter: A_{LTP}=1.0, A_{LTD}=0.95; τ_{LTP}=20, τ_{LTP}=28 (Fig. 5.6C), the shape of STDP function is more close to classical STDP function observed in vitro and in vivo [33]: the peak of LTD is lower than the peak of LTP, and the working time window of LTD is longer than that of LTP. The value of SI in this region was still positive implying it separate the uncontrollable input neurons. This suggests that in addition to SAA and SAP, SAS can also occur in biological neural networks.

Figure 5.8B shows the integral value of STDP function with those parameters. In blue color regions, LTD is stronger than LTP, thus random spikes in presynaptic neurons should cause decreasing weights from the neurons in theoretically. As shown in Fig. 5.8A, SAS occurred in those blue regions. However, if such decreasing effect was too strong (e.g., with the parameter: A_{LTP}=1.0, A_{LTD}=1.4; τ_{LTP}=20, τ_{LTP}=30; Fig. 5.6D), both SAA and SAS was not be observed and both LW_{input} and NW_{input} were decreased (Fig 5.9), because LTD was too strong compared to LTP thus the almost all weights decreased and the networks cannot learn anything. Therefore, we found that both SAA and SAS are only occurred at the same time in the balanced region where the integral value of LTD is stronger but not too much stronger than that of LTP.

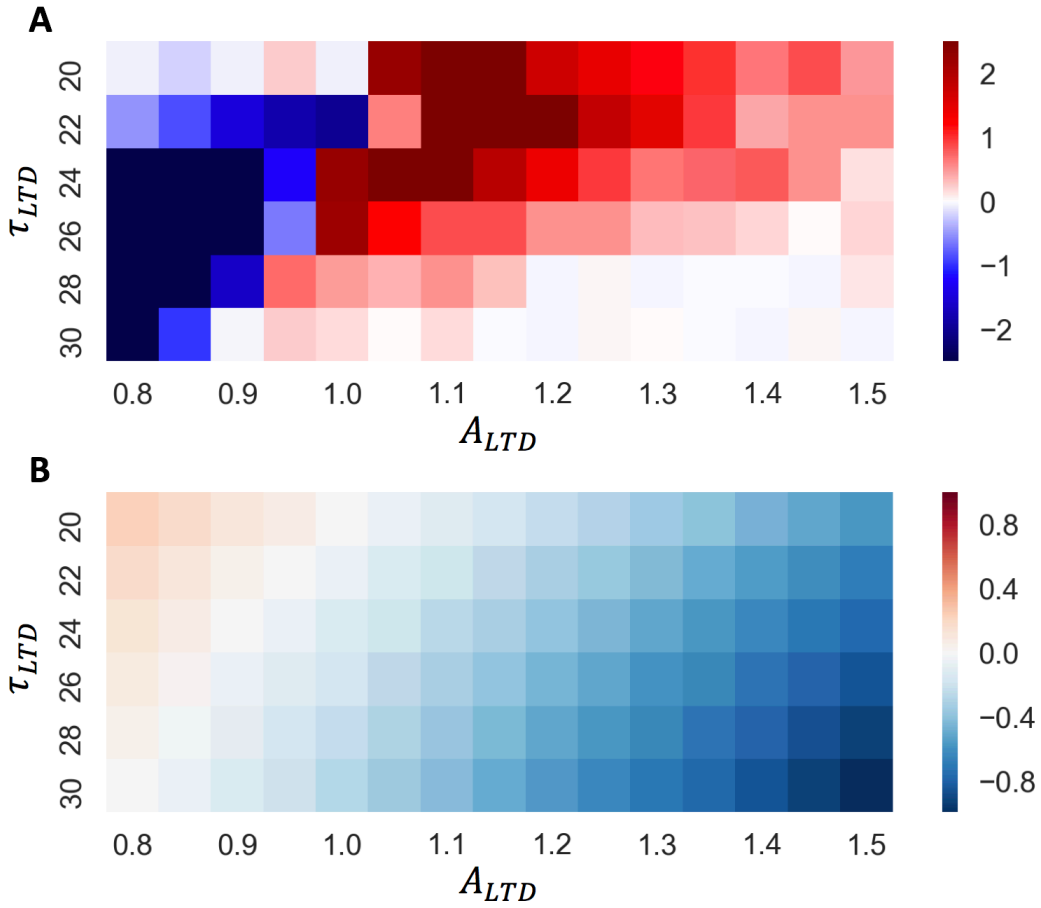

Figure 5.8: A: Dependence of the performance of SAS on A_{LTD} and τ_{LTD}. Color indicates the mean selection indicator (SI) with those parameters (n = 20). B: Dependence of the characteristic of STDP curve on A_{LTD} and τ_{LTD}. Color indicates the integral value of STDP function with those parameters.

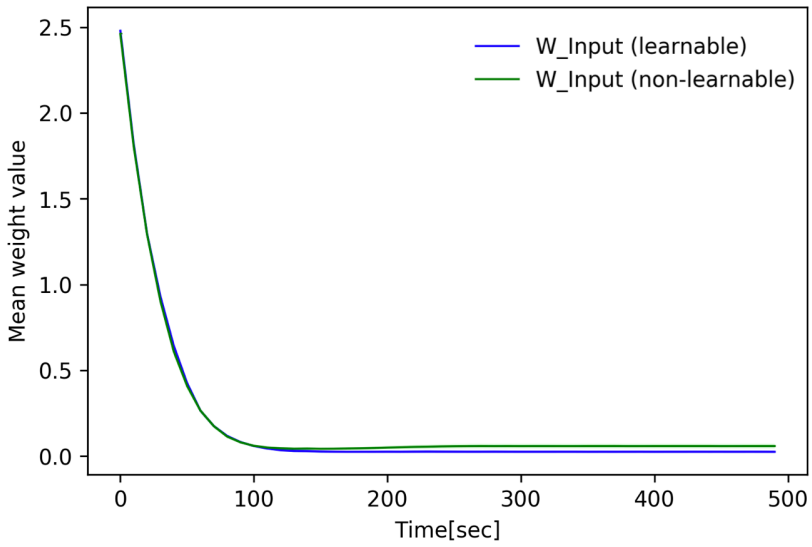

Figure 5.9: Time series of the mean connection strength from the input neurons to the other neurons in asymmetric-STDP: A_{LTP}=1.0, A_{LTD}=1.4; τ_{LTP}=20, τ_{LTP}=30. The shaded regions represent standard errors of the mean (n = 20).

5.3.3 Conclusions

In this study, we demonstrated that spiking neural networks with asymmetric-STDP reproduced the behavior observed in the cultured networks in which if it was impossible or difficult to learn a behavior that avoids stimulation, the response to the stimulation decreased. As the results of the parameter search with A_{LTD} and τ_{LTD}, we found that such behavior to select connections are limited in specific parameter space in which effect of LTD should be stronger but not too much stronger than the effect of LTP.

5.4 Discussion

In this study, we found that in both cultured networks and spiking networks, if it is difficult to learn an action to avoid a stimulus input, the plasticity works to suppress the influence of the uncontrollable stimulus input to the networks by weakening the connection strength from the input neurons.

In neuroscience, this kind of phenomena where constant sensor input gets ignored is known as neural adaptation, sensory adaptation or stimulus-specific adaptation, and this phenomena are observed in many regions in the brain (e.g., in the auditory system [83, 84, 85, 86]) and also *in vitro*

[78]. Such adaptation can be divided into two types: fast adaptation (less than one second) and slow adaptation (more than a few minutes) [87]. The mechanism of the slow adaptation is considered to be synaptic plasticity like LTD. While one of the possible mechanism of the fast adaptation is synaptic fatigue [79]: Repeated stimulation cause depletion of neurotransmitters in the synapse, weakening the response to the stimulation. Some studies suggest that synaptic fatigue is caused by high-frequency stimulation, and occurs in presynaptic neurons [79].

In our experiment with the neuronal cultures, low-frequency stimulation (1 Hz) was used, and the results showed that the evoked spikes of the presynaptic neuron (input neurons) did not decrease, and the evoked spikes of the all other neurons were decreasing for more than 20 minutes. Therefore our results with neuronal cultures suggest that the observed behavior was not fast adaptation caused by the synaptic fatigue, but slow adaptation caused by LTD. The spiking neural networks with the asymmetric-STDP reproduced the observed slow adaptation (although the networks have STP that has similar dynamics with synaptic fatigue, STP stabilize the firing rate within 1 sec thus the slow adaptation can not be caused by STP). Moreover, we found that the phenomena observed in some parameter spaces where the shape of STDP function was more similar to the shape broadly observed *in vitro*. Therefore, we argue that the same mechanism should be work in biological neural networks.

Here we discuss what kind of connections are weakened by this mechanism. To be simplified, we think about a minimal case in which there are 2 neuron and 1 connection (Fig. 5.10) with asymmetric-STDP comparing with the case with symmetric-STDP. In the asymmetric-STDP case, if $\Delta t_p \approx \Delta t_d$ and $\Delta t_p < \tau_{LTP}$, $\Delta t_d < \tau_{LTD}$, the connection decrease because with the asymmetric-STDP, LTD has stronger effect than LTP (Fig. 5.10A; τ_{LTP} and τ_{LTD} are working time windows of LTP and LTD in STDP function [see Section 2.3.2]). In the symmetric-STDP case, the connection does not change much because LTP and LTD effect equally to the connection with the symmetric-STDP (Fig. 5.10B). Thus, with the asymmetric-STDP, if the mean value of spike interval which cause LTP (Δtp) and the mean value of spike interval which cause LTD (Δtd) is close: $\sum_{t=1}^{N} \frac{\Delta tp_i}{N} \approx \sum_{t=0}^{N} \frac{\Delta td_i}{N}$, the connection will disappear. In the large networks, such situations are apt to be occurred if the presynaptic neuron fired independent from the other neurons with a high frequency. Therefore, connections from input neurons which get stimulation with high frequency should tend to decrease based on the dynamics of the asymmetric-STDP (e.g., more than 20 Hz, if $\tau_{LTP} + \tau_{LTD} = 50$ msec). There must be the conditions for the stimulation frequency F [Hz] in SAS: e.g., $F \geq k$ where $k \approx 10^3/(\tau_{LTP} + \tau_{LTD})$ (This is for minimal case

with 2 neurons, and this might be modified for larger networks.).

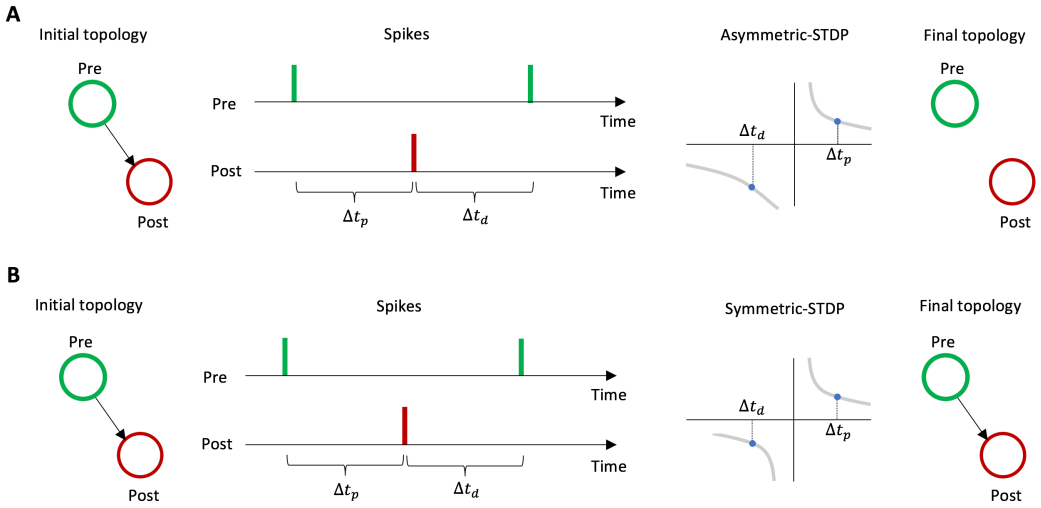

Figure 5.10: Stimulation avoidance by selection in two neurons: A presynaptic neuron, a post synaptic neuron and the connection between them. Green circles represent presynaptic neuron; Red circles represent post synaptic neurons. Green bars represent spikes of synaptic neurons; Red bars represent spikes of post synaptic neurons. A: Asymmetric-STDP case. If $\Delta t_p \approx \Delta t_d$ and $\Delta t_p < \tau_{LTP}$, $\Delta t_d < \tau_{LTD}$ (τ_{LTP} and τ_{LTD} are working time windows of LTP and LTD in STDP function), the connection decrease because with asymmetric-STDP, LTD has stronger effect than LTP. B: Symmetric-STDP case. The connection does not change much, even if the same conditions as A are satisfied, because LTP and LTD effect equally to the connection with the symmetric-STDP.

In our experiments, we focus on SA_S for input neurons, however, SA_S should also be work between other internal neurons. In the case of internal neurons, decreasing a synaptic weight from one of the internal neurons that has higher firing rate, leads to decrease the neurons' firing rate since there should be some feedback loop, thus, the synaptic weight should be decreased but less than completely separating like the case of input neurons. This effect should lead to stabilize the network, thus SA_S might work to regulate internal network to be stabilized especially at a developmental stage.

We demonstrated that even without the plasticity of connections to and from inhibitory neurons, the networks can select the controllable input neuron rather than uncontrollable input neuron. However, applying the plasticity to connections to and from inhibitory neurons the networks might be able to avoid the effect of the uncontrollable neurons by suppressing the firing by

the inhibitory neurons. We need to study further to show whether such other mechanism for SAS actually can realize or not.

Our results suggest that if SAA and SAP is not possible or difficult, then SAS work to maintain the homeostasis. This two-layered homeostatic principle looks similar to Ashby's theory of ultrastability in which the system has two type of homeostasis, and if the first regular homeostasis is unstable and its essential variables exceed the limits then the second homeostasis works to rearrange the system dramatically [4]. The system will reconstruct itself by trial and error until a stable homeostasis can be acquired. Ashby suggests that biological systems are ultrastable with these two types of homeostasis. Our results suggest such a behavior can emerge thanks to local dynamics of neurons in cultured networks and spiking neural networks with STDP.

Moreover, the simulation results showed that ignoring an uncontrollable constant stimuli is a strong feature of spiking neural networks with asymmetric-STDP. Almost all connections from the input neurons decreased to zero. On the other hand, the weights from input neurons with controllable input (sensor input that the agent can learn to avoid) increased. This suggests that the networks can isolate the input neurons which have uncontrollable stimulus input, and it can be regarded as dynamics trying to regulate self and non-self autonomously (Fig. 5.11). Here closed-loop of sensor and motor in which motor outputs control sensor stimulation like sensorimotor contingency [88] is regarded as self, on the other hand, open-loop of sensor and motor is regarded as non-self. The open-loop will be collapsed by isolating the sensor neuron. It is interesting that the dynamics emerge from just the simple local dynamics of neurons.

How to discriminate self from non-self reminds us a theory of autopoiesis proposed by [89]. In addition to the structural viewpoint of regulating self boundary above, here we discuss the result of self-regulating behavior from an autopoietic point of view.

In autopoiesis, the discrimination comes with the boundary between self and non-self. It is not a physical rigid boundary, but a dynamical one: it should be constantly produced and maintained by system's own processes.

In 1974, Varela et al. reported a simple mathematical model using artificial chemistry featuring autopoiesis [90]. Two metabolite particles (S) generate one boundary particle (L) catalyzed by a catalytic particle (C). Those boundary particles are connecting to each other to form a connected boundary, which encloses C and L. The boundary constantly decays and is being repaired by the free boundary particles L. This self-organizing process of encapsulating C and L defines self-discrimination. No single particle defines the self boundary. Self-entity only emerges at a certain collective level.

This picture becomes much clearer by taking the immune systems as an example. Vertebrates establish self/non-self discrimination by forming an idiotype network: antibody- antigen chain reaction exists among antibodies according to N.K. Jerne's hypothesis [91]. The current understanding of self/non-self discrimination has a molecular biological basis, however, acquired immunity still needs to be exploited. A candidate is the autopoietic picture. Each antibody can adaptively change the self-boundary. By self-organizing an idiotype network, self/non-self discrimination emerges as a result of the network reactions: the antigen-antibody reaction is suppressed locally for the self-antigens, but the reaction will be percolated for the non-self antigens. The reaction network determines the self/non-self discrimination similar to the Varela et al's simple artificial chemistry.

Coming back to the present study, we saw that each neural responses did not determine the self/non-self boundary. Only at the level of neural network of a certain size, a self/non-self boundary emerges. A neural network determines the self/non-self as well as the immune system does. The boundaries of self/non-self for immune systems and for neural systems are processed dynamically. It is not explicit for the network whether a certain firing pattern depends on what comes from outside or from inside. Like the immune system, a pattern which makes the network respond strongly to and causes structural change of the network is regarded as non-self here.

For example, the controllable input above is initially regarded as a pattern from outside (non-self). However, as the change of the network progresses and the network learns the behavior to control it, the input will no longer cause the large changes and will be regarded as a pattern from inside (self). In the case of the uncontrollable input, initially, it is regarded as a pattern from outside (non-self) like the case of controllable inputs, however, by weakening the connections from the sensor based on the dynamics proposed in this paper, explicit boundaries like Varela 's cellular boundary can be created, thus the inputs explicitly isolated from the inside and no longer affects to the internal network (self).

In addition, our previous results with a simulated minimal network consisting of three neurons (two excitatory neurons and a inhibitory neuron) with almost the same models in this paper could predict a simple causal sequence of stimuli [92]. In that case, like the controllable input, predictable input is initially regarded as an external pattern (non-self) that causes structural changes in the network, however, when the network would learn to predict the input, the input will be no longer affect to the network and is regarded as a pattern from inside (self).

Although we have discussed only the patterns from environment, the same dynamics should also work inside the network (although regulation by action

does not occur inside the network and requires a coupling with environment). Inside the network, the dynamics such as isolation of uncontrollable pattern and prediction of predictable pattern work to regulate the boundaries, and the network converge to stable states in which the network shows transitions of several patterns. In this way, the neural network can also be regarded as a system that acquires its own stability through the autopoietic process by means of action, prediction, selection.

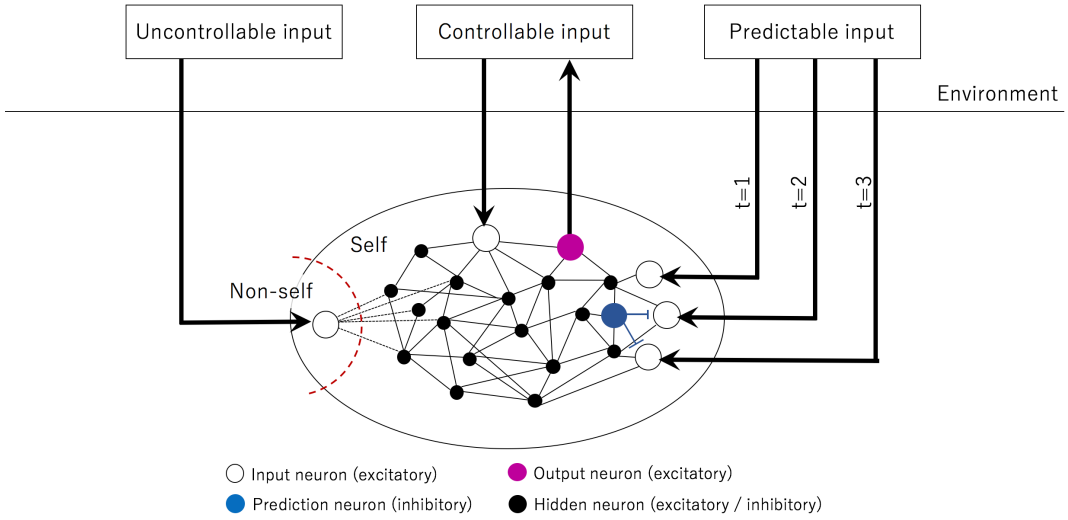

Figure 5.11: Autonomous regulation of self and non-self: When inputs are uncontrollable (i.e., networks cannot learn actions or prediction to avoid stimulation), input neurons are separated as non-self avoiding an effect of the stimulation to the internal network.

Chapter 6

Scalability and Application

In the previous chapters we described some studies on three ways of stimulation avoidance (Stimulation Avoidance by Action (SAA), Stimulation Avoidance by Prediction (SAP), and Stimulation Avoidance by Selection (SAS)) based on the synaptic plasticity, such as STDP in the cultured neural networks and the spiking neural networks. These studies suggest that such a principle for avoiding stimuli can emerge in both biological and artificial networks. However, in the previous chapters, we have not much discussed about its scalability and application. Therefore, in this chapter, we show some studies on the scalability and the application of SAA based on LSA. In the following sections, Section 6.1 shows the scalability of LSA against network size using spiking neural networks, Section 6.2 shows estimation of the scalability of LSA using the network analytical approach, Section 6.3 shows the application to autonomous humanoid robot and its experimental results.

6.1 Learning by Stimulation Avoidance Scales to Large Spiking Neural Networks

6.1.1 Introduction

This section focuses on the scalability of LSA that is a mechanism for SAA. In our studies using cultured neural networks, we demonstrated that even the cultured networks with small number of neurons (about 100 neurons) learned the desired behavior (see Chapter 3), although previous studies used much larger networks (10,000-50,000 neurons) [29]. These studies imply that such a homeostatic property scales from 100 neurons to 50,000 neurons in cultured neural networks.

Although our previous study shows that LSA in spiking neural networks

scales from 2 neurons to 100 neurons [31], some specific conditions for "selective learning task" where the networks must learn a spatial pattern to avoid the external stimulation, were required: (1) external stimuli were delivered to the input neurons every timestep (1,000 Hz), (2) other external stimulation were delivered in 10 ms to all neurons excluding inhibitory neurons and one of the output neuron groups, after n $\geq$ 1 neurons in the output group fired. These conditions are somewhat unnatural. In addition, it is still unclear whether LSA scales to larger networks (more than 100 neurons) or not.

In this section, using spiking neural networks, we first demonstrate LSA works well in networks consists of 100 neurons with more natural conditions. Second we demonstrate this learning behavior scales to larger networks (at least 3,000 neurons).

6.1.2 Methods

The model for spiking neuron proposed by Izhikevich [52] was used to simulate excitatory neurons and inhibitory neurons (see Section 2.3.1). Randomly connected networks of 100-3,000 neurons with 80% excitatory and 20% inhibitory neurons were simulated. This ratio of 20% inhibitory neurons is standard in simulations [63, 52] and similar to biological values [82]. The excitatory neurons were divided into four groups: Input, Output A, Output B, and Hidden (Fig. 6.5). Input group consisted of 20% neurons and external stimulation was delivered only to this group. Output A group and Output B group consisted of 10% neurons each and the desired behavior that the network learn was composed of these groups. Hidden group consisted of 40% neurons. The connection size of each neuron depended on the experiments (10-300). There were no direct connections between Input and Output group, and self-connections (e.g., from neuron i to neuron i) were forbidden. The weight values w between each neuron were randomly initialized with uniform distributions as $0 < w < 5$ for excitatory neurons, $-5 < w < 0$ for inhibitory neurons. The weights between excitatory neurons change based on STDP function (see Section 2.3.2).

There were three types of input to the neurons. (1) Zero-mean Gaussian noise m with a standard deviation $\sigma = 3$ mV was delivered into each neuron at each time step. This is required for spontaneous firing of the neurons. (2) External stimulation e was delivered to Input group. There were two conditions for the stimulation. Condition 1: input current (10 mV) was stimulated with a fixed frequency (100 Hz). When more than 40% of Output A fired and less than 40% of Output B fired, the external stimulation to the Input was removed in 1,000-2,000 ms (it was randomly determined every time). Condition 2: when neurons in Output B fire more than neurons in

Output A, input is stimulated once at 30 mV (with a maximum frequency of 100 Hz). These conditions imply that the goal for the network is to avoid external stimulation by increasing the firing rate of Output A compared to Output B. (3) Synaptic current from other neurons: when a neuron a spiked, the value of the weight $w_{a,b}$ was added as an input to neuron b without delay. All these input were added for each neuron (Eq. 2.5). Here we also applied STP to each synapse, thus the input from other neurons was weighted with STP term (see Section 2.3.3). Although STP is not indispensable for LSA to work, applying STP make firing rates and a burstiness index [61] stable against the network size (see Section 2.3.3).

In the control experiment, stimulations of condition 1 and 2 were randomly delivered to Input group and other experimental settings were same as the explained above.

To evaluate the whether LSA works well or not, we defined a measure of success M_S as

$$M_S = \frac{(F_A - F_B)}{N} \tag{6.1}$$

Here, F_A represents the mean number of spikes of Output A in the last 20,000 ms, F_B represents the mean number of spikes of Output B and N represents the total number of spikes of Output A and B. When Output A is fired more than Output B, M_S becomes positive, and in the opposite case, M_S becomes negative. To increase M_S, the firing rate of Output A should be increased and that of Output B should be decreased: the higher value of M_S indicates that the network learned to avoid the stimulation well in the both condition 1 and condition 2 explained above.

6.1.3 Results

We performed the learning experiments in small (100 neurons) and large networks (1000-3000 neurons) for examining scalability of LSA. First, the results of the small networks are shown before the results of larger size networks. Each experiment were performed with 20 networks.

Small Networks

The previous study showed the learning rate of LSA for more simple task changes depending on the connection size between neurons [31]. Therefore we first show the success measure varies with the connections size in the experimental setup here. Figure 6.1 shows the relationship between the success

measure and the connection size. As shown in this figure, when the connections size of each neuron was from 40 to 90, the success measures were significantly larger than the control experiments (random), and the success measure was maximized at 50 connection size. Below, we focus on the results of the experiments with 50 connection size.

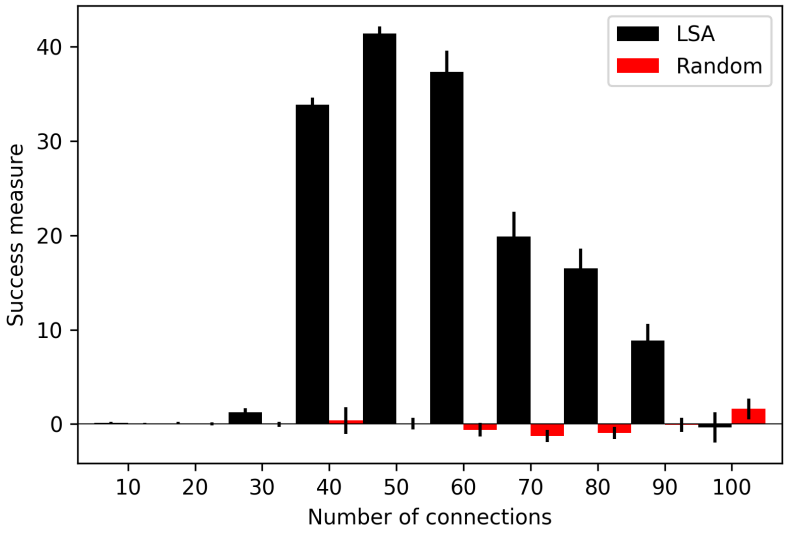

Figure 6.1: Dependence of success measure on connection size in the small networks. In learning experiment (LSA), the success measure varied depends on the network size. Error bars represent standard error of the mean (n = 20).

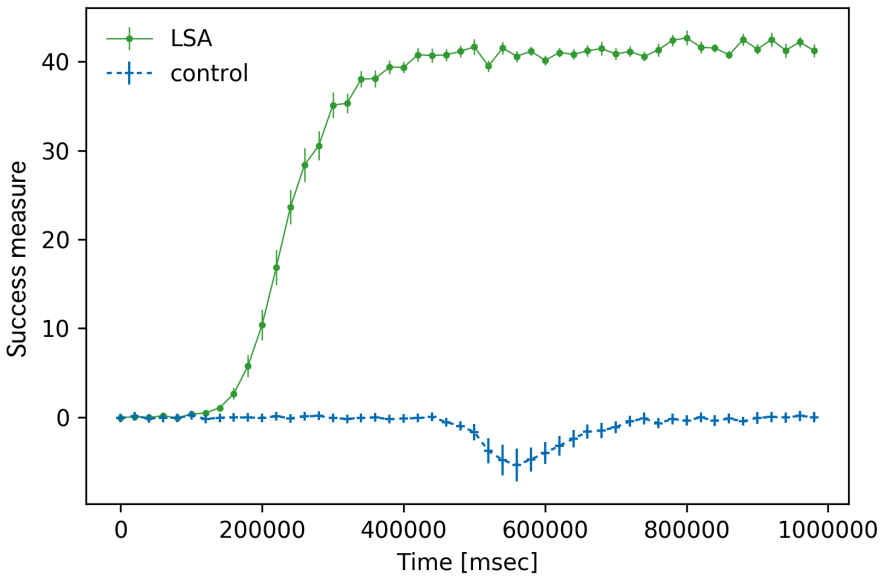

Figure 6.2: Time series of the success measure in the small networks (100 neurons with 50 connections per neuron). The success measure gradually increased and stabilized at the higher values than the control experiment (random). Error bars represent standard error of the mean (n = 10).

Figure 6.2 shows the time series of the success measure. The success measure in the learning experiments (LSA) gradually increased in 15,000-40,000ms and stabilized at the higher value, on the other hand, the success measure did not increased in the control experiments .

Figure 6.3 shows a typical example of a raster plot of spikes in the experiment with 50 connection size. At the beginning of the experiment, the spikes of Hidden gradually increased, then the spikes of Output A increased, but the spikes of Output B did not change much. This actual spiking pattern shows qualitatively that the firing rate of Output A became larger than that of Output B, which implies that the network had high success measure.

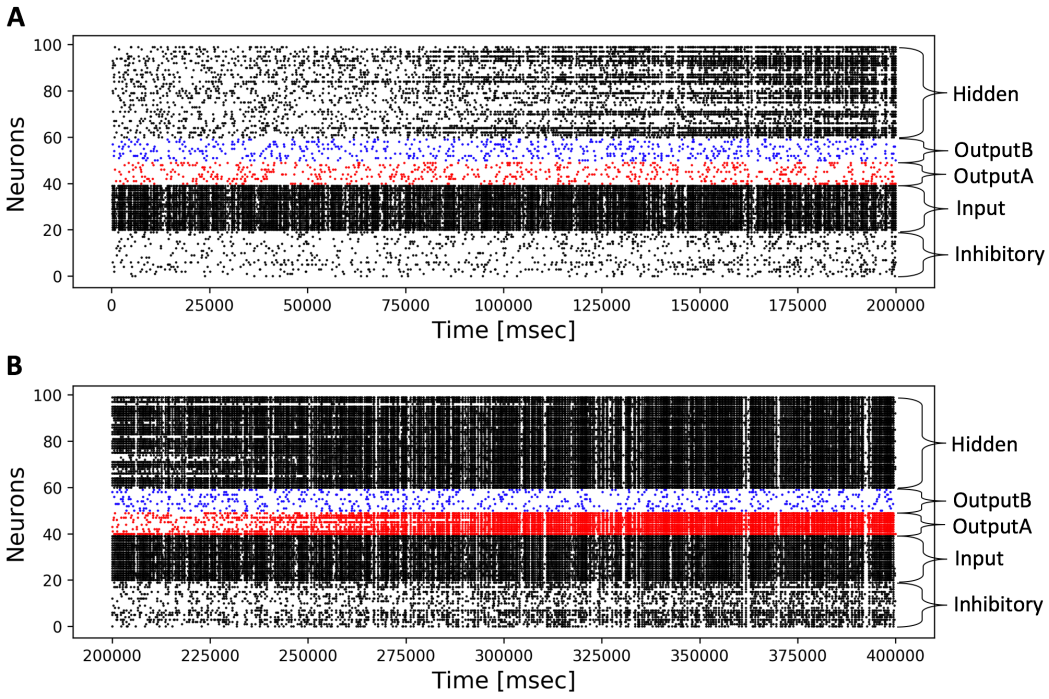

Figure 6.3: Raster plot of spikes in the small network (100 neurons with 50 connections per neuron). Each dot represents one spike: Black dots represent spike of inhibitory neurons, input neurons and hidden neurons; Red dots represent spikes of Output A; Blue dots represent spikes of Output B. A: First 200,000 ms. The spikes of Output A and Output B were almost same. The spikes of Hidden gradually increased. B: Second 200,000 ms. The spikes of Output A gradually became larger than that of Output B.

Figure 6.4 shows a typical example of the network diagram for the obtained network. The connections from Input to Hidden, and the connections from Hidden to Output A were larger than initial values thus information from Input can be transmitted to Output A via Hidden. On the other hand, there were almost no connections from other groups to Output B. The connections from Output B to other neurons were strengthened due to the counter to the fact that the connection to Output B from the others were weakened. The more abstract network diagram in Fig. 6.5 was drawn based on the network diagram. Since there are no direct connection from Input to Output A, to transmit information from Input to Output A, Hidden must mediate the transmission. Thus this network topology is a optimal for avoiding the external stimulations in this settings. This suggest that through LSA, the networks autonomously yielded the optimal sensory-motor coupling to avoid external stimulation in the environments.

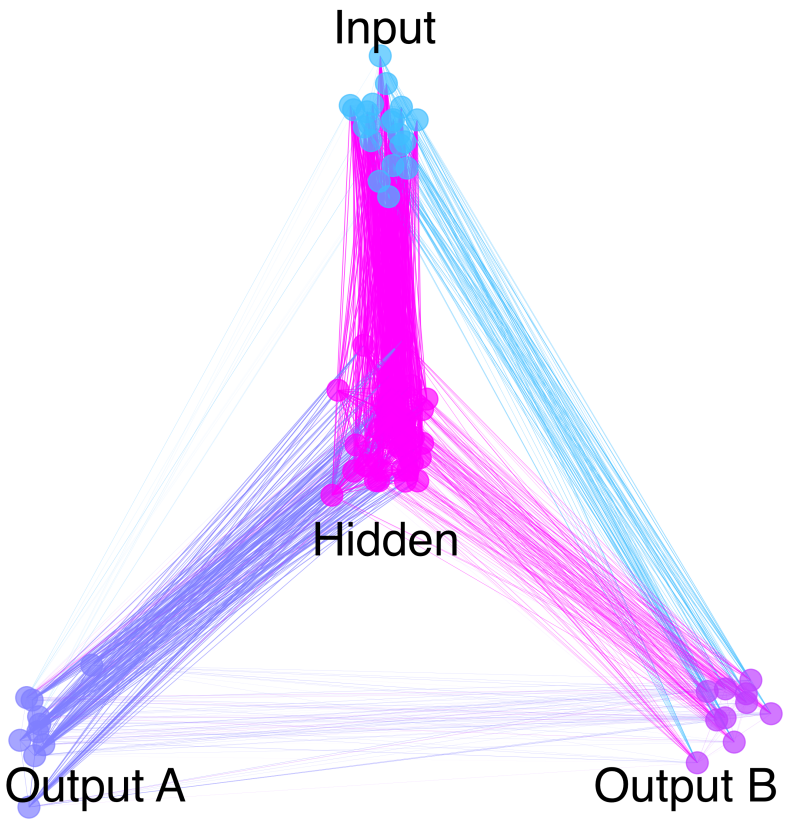

Figure 6.4: Typical example of a obtained network in the small network (100 neurons with 50 connections per neuron). Each node represent a neuron. Each edge represents a connection between two neurons. The width of the edge represents the weight value and the color of the edge represents the direction of the edge. The edges direct to node with same color (e.g., pink edge represents the edges to Hidden). There are many strong edges from Hidden to Output A, but almost no edge from Hidden to Output B.

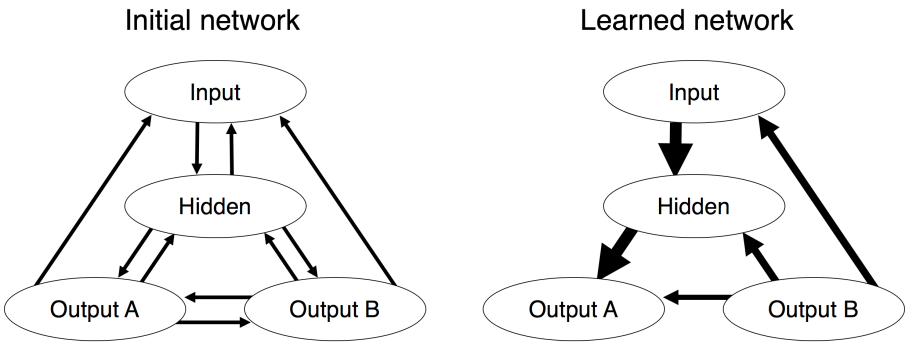

Figure 6.5: Network topology of the initial and the obtained networks in the small networks (100 neurons with 50 connections per neuron). In the obtained topology, There was a route from Input to Output A via Hidden, although, there was no route from Input to Output B.

In this way, we found that LSA worked well in 100 neurons with more natural settings than previous study [31]. The learning results was clear at both the neural activity level and the network structure level.

Large Networks

To examine whether LSA scales to larger networks, the learning experiments with larger networks were performed using the same experimental settings as the experiments above with 100 neuron except for the network size.

Figure 6.6 shows time series for the success measure of the large networks with the parameter: 50 connections. The success measure increased and stabilized at the higher values only for the networks with 1,000 neurons, although longer time were required to learn it compared to smaller networks (100 neurons).

However, we found that the larger networks than 1,000 neurons can increase the success measure with other parameters. For example, in the networks consists of 3,000 neurons with 300 connections (other settings are same with the settings above), the success measure gradually increased, although the deviation was larger than the smaller networks (Fig. 6.7).

These results suggest that when the network size is larger, the learning becomes more difficult compared to the smaller networks because it took more time to learn and the deviation was larger than smaller networks. In addition, it also suggest the parameter region for high success measure in the larger networks was different from the smaller networks.

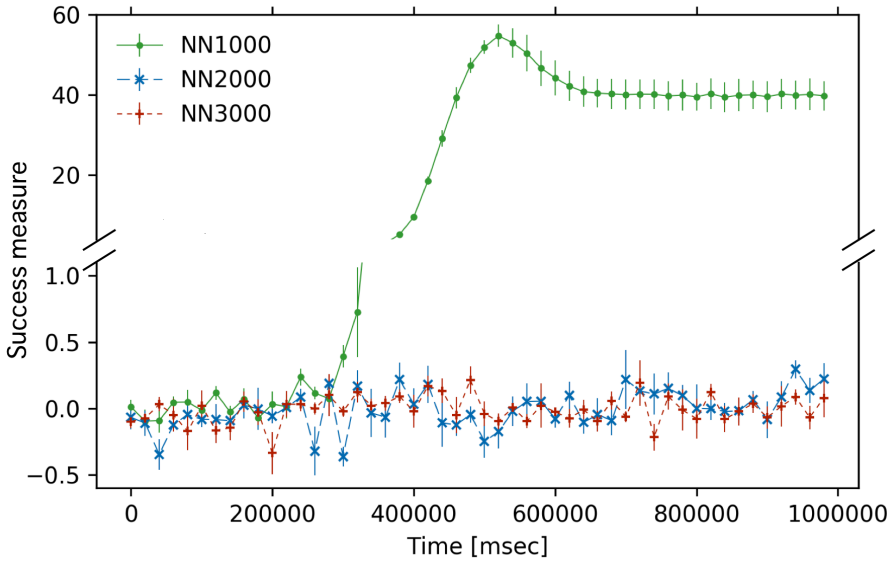

Figure 6.6: Time series of the success measures in the large networks (1,000-3,000 neurons with 50 connections per neurons). Only in the experiments with 1,000 neurons, the success measure gradually increased and stabilized at a high values. Error bars represent standard error of the mean (n = 10).

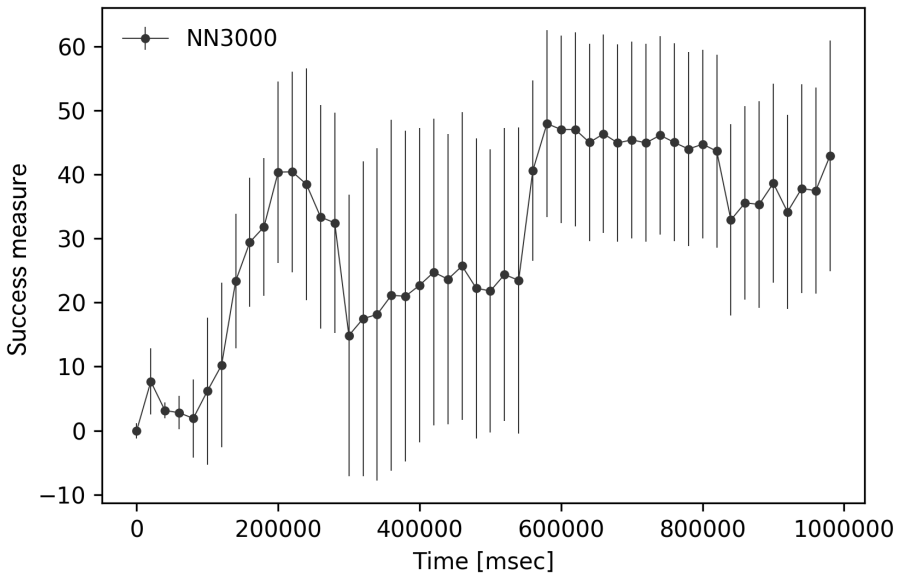

Figure 6.7: Time series of the success measure in the large networks (3,000 neurons with 300 connection per neurons). The success measure gradually increased, although the deviation was larger than in the smaller networks. Error bars represent standard error of the mean (n = 5).

We examined wether the parameter regions for high success measure actually differ depending on the network size using small networks (100-300 neurons). Figure 6.8 shows the dependence of the success measure on the connection size and the stimulation strength differ depending on their network size (100-300). Here, the stimulation intervals was set to 20 ms. Figure 6.9 shows that the dependence of the success measure on the stimulation interval and the stimulation strength differ depending on their network size (100-200). Here, the connection rate was set to 0.5. These results shows the parameter region for high success measure differ depending on the network size. In addition, Fig. 6.9 shows that there is upper limit of the stimulation intervals for the learning success, this supports the controllability condition we discussed previously (see Section 3.4).

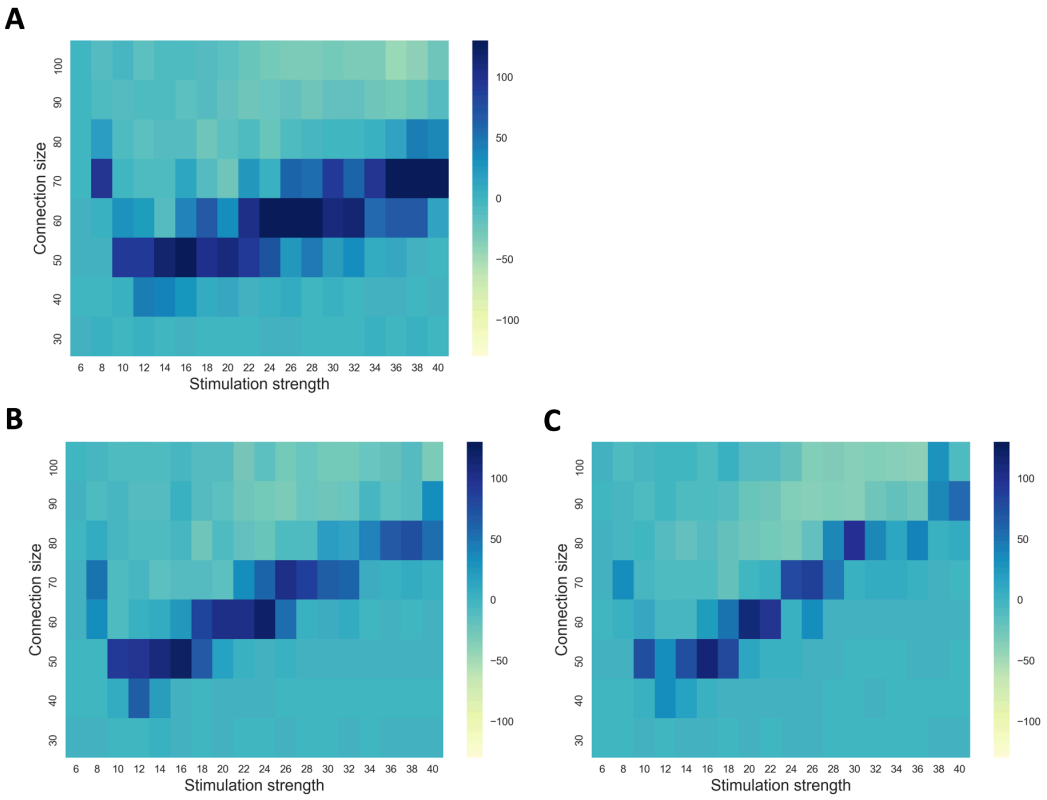

Figure 6.8: Dependence of the success measure on the connection size and the stimulation strength (100-300 neurons). The stimulation intervals was set to 20 ms. The color indicates the success measure. A: The result in the networks with 100 neurons. B: The result in the networks with 200 neurons. C: The result in the networks with 300 neurons.

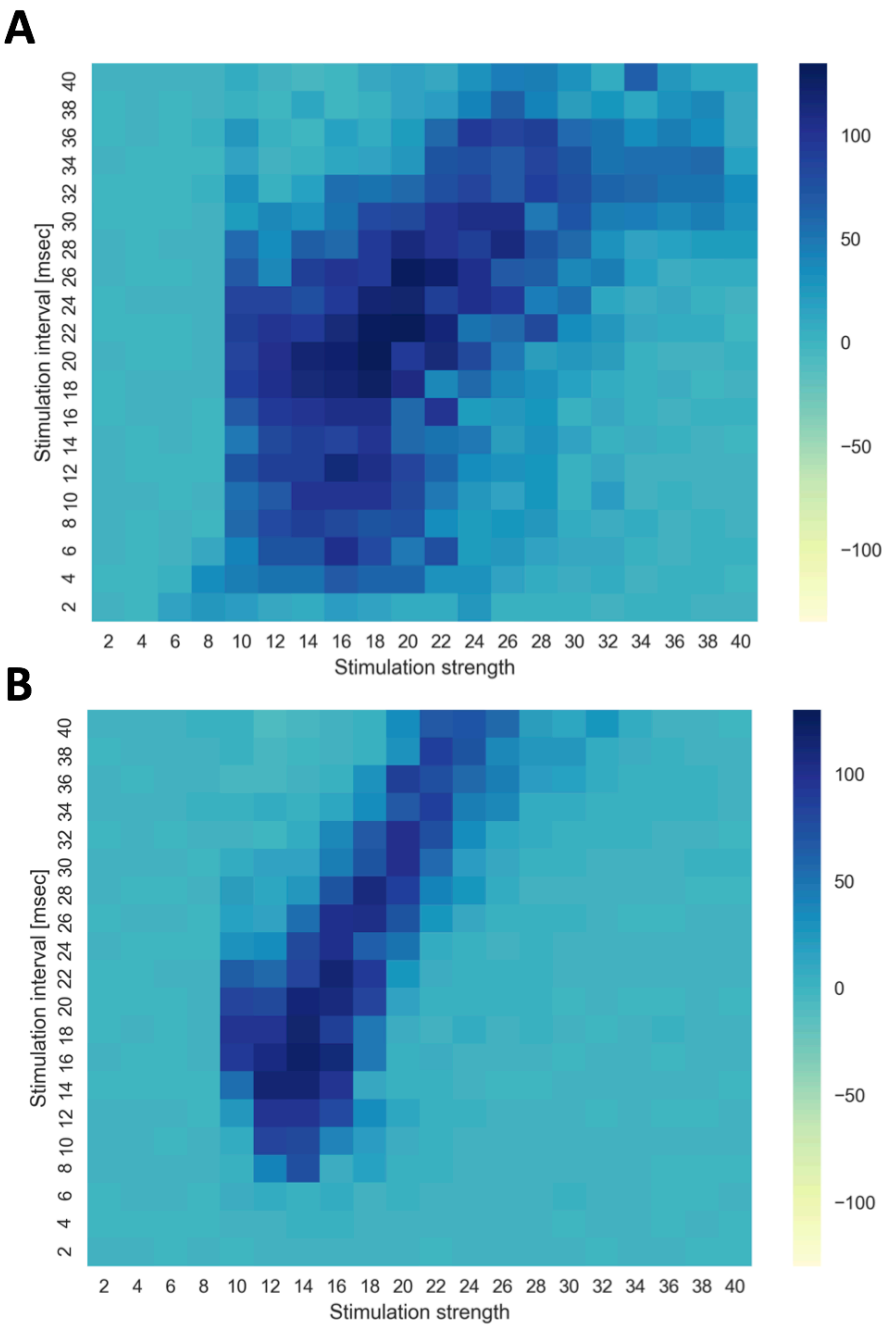

Figure 6.9: Dependence of the success measure on the stimulation intervals and the stimulation strength (100-200 neurons). The connections rate was set to 0.5. The color indicates the success measure. A: The result in the networks with 100 neurons. B: The result in the networks with 200 neurons.

6.1.4 Discussion

The learning experiments were performed based on LSA that cause SAA, and revealed that LSA can scale to the larger networks (at least 3,000 neurons) compared to the previous works (100 neurons) [31]. However, the parameter region where LSA works well depends on the network size and it seems that the learning difficulties increase with the network size. In the experimental setup, there were no direct connections between Input and Output so that if the network size becomes larger with same connection size per neuron, the average path length between Input and Output would becomes larger. We assumed that this larger path length made the learning in the large network more difficult (or impossible, because Δt_1 in Fig. 3.17 might exceed the threshold of the connectivity conditions; see Section 3.4) so that larger networks require the larger connection size which get the path length smaller, for high success measure.

The larger networks with the larger connections requires more computational resources to be simulated, and this make the parameter search more difficult, thus in next section, we estimate the scalability using a network analytical approach rather than actual simulation of the large networks.

6.2 Estimation of Scalability by Network Analysis Approach

6.2.1 Introduction

As shown in the previous section, the parameter region where LSA works well was limited, and its parameter region also varied depending on the network size. Since learning experiments searching for parameters in a large network require large computational resources, we take a network analytical approach to estimate the scalability of LSA focusing on a signal-to-noise ratio (SNR), rather than actual simulation for spiking neural networks. In the estimation, first we experimentally define the value of SNR at which LSA works well, then, to estimate the scalability, we generate random networks changing the network size and analyze the network topology to examine how large networks can maintain the SNR.

6.2.2 Methods

For determining the SNR, we performed learning experiments of LSA using a one-dimensional chain network consisted of an input neuron, an output neu-

ron, and hidden neurons. Stimulation (signal inputs) with fixed frequency were delivered to the input neuron. There were also noise inputs to other neurons except for the input neuron, which model input currents from other neurons which are not on the route from the input neuron to the output neuron (Fig. 6.10). We used Izhikevich model and STDP for the network (see Section 2.3), and the network consisted of only excitatory neuron for simplicity. In the learning experiments, the embodiment of the networks was simple: stimulation was delivered to the input neuron with a certain frequency, and if the output neuron fired within a certain time τ after the stimulation to the input neuron, then the stimulation to input was removed for a certain time (1,000-2,000 ms). Specifically, the stimulation frequency was 100 Hz and the interval of each stimulus was 10 ms. τ was set to 10 ms. Here, success of the learning was defined as that if all connection strengths that are directed from input to output are larger than the connection strengths in the opposite direction. This was defined in order to examine whether information from the input neuron can be reached to the output neuron.

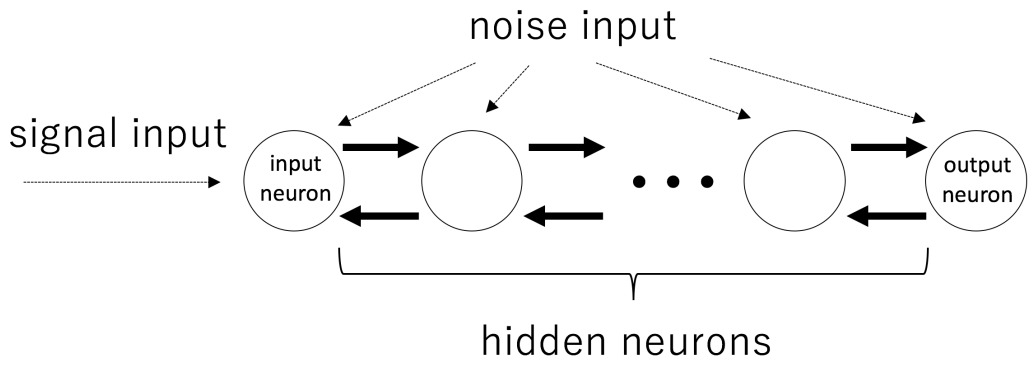

Figure 6.10: One-dimensional chain network consists of an input neuron, an output neuron and hidden neurons. The signal input was delivered to the input neuron, and the noise inputs were delivered to each neuron.

We calculated the SNR changing the number of hidden neurons and the intensity of the noise input. The SNR is defined as

$$SNR = \frac{I_{signal}}{I_{noise}} \tag{6.2}$$

Here, I_{signal} demotes a mean value of the initial connection strength connecting the path from the input neuron to the output neuron, and I_{noise} is a mean value of the noise inputs to all neurons. We defined lb-SNR as lower bound of the SNR at which the network with 10 more hidden neurons learn

well to avoid the stimulation: i.e., if the SNR is lower than the lb-SNR, it indicates the network cannot learn the behavior to avoid the stimulation. In other words, the lb-SNR is the lower limit of the SNR required to convey information from the input neuron to the output neuron.

Second, to estimate the scalability, we examined how the large networks can maintain the higher SNR than lb-SNR obtained by the above method. For calculating the SNR, we did not simulate for the dynamics of spiking neural networks, but estimated with the topology of the network in following ways: (1) we randomly generated a network, and (2) find paths that connect input to output through hidden neurons, (3) calculated the SNR regarding the hidden neurons on the path as a signal neuron, and the connection from the signal neurons as signal, and the number of connection from the other neurons as noise input (Figs. 6.11). When generating the network, each neuron has two uniformly distributed random number as its position (x and y), and the probability of connection between neurons is determined according to inversely proportion to the square of the distance between the neurons.

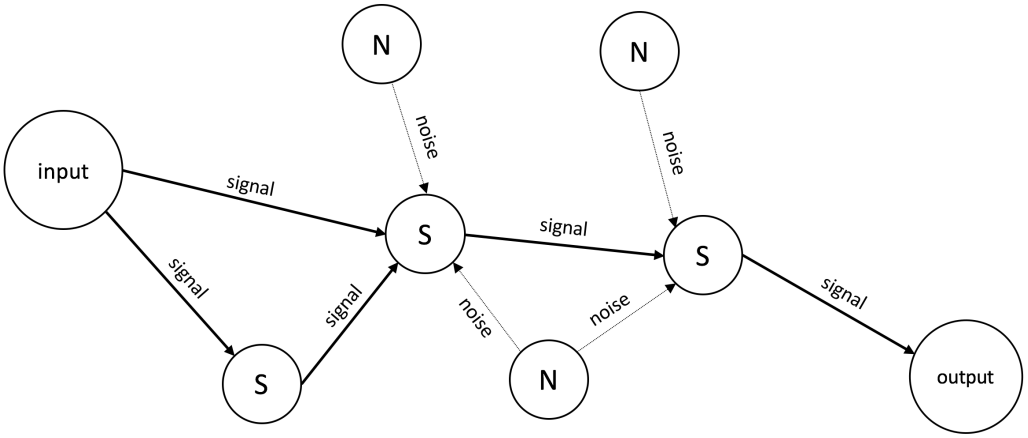

Figure 6.11: Conceptual diagram of a randomly generated network. Each nodes represents a neuron: one input neuron, one output neuron, and some hidden neurons (S and N). Each edges represents a connection between two neurons. Hidden neurons on a path directed from input to output, was defined as a signal neuron (S), and other neurons as noise neuron (N). The input from the signal neurons was defined as a signal, and the input from the other neurons as a noise.

6.2.3 Results

We first performed the simulation experiments to calculate the lb-SNR of the one-dimensional chain networks, changing the parameters: the noise input and the number of hidden neurons. Figure 6.12 shows success rates when strength of noise input and the number of hidden neurons are changed in one-dimensional chain networks. As shown in this figure, if the noise input became stronger, the network cannot learn with the large number of hidden neurons. The lb-SNR as the lower bound of the SNR at which the network that has 10 more hidden neurons can learn the behavior, was 0.4 (signal = 20, noise = 50).

We calculated the SNR of randomly generated network by the method above, changing the network size. Figure 6.13 shows the relation between the network size and the topologically estimated SNR of the randomly generated networks. Here, the dotted line represents the value of lb-SNR experimentally determined with the above methods. When the SNR is larger than this value, it is estimated that the learning will be success. Thus the result suggest that LSA scales to about 50,000 neurons.

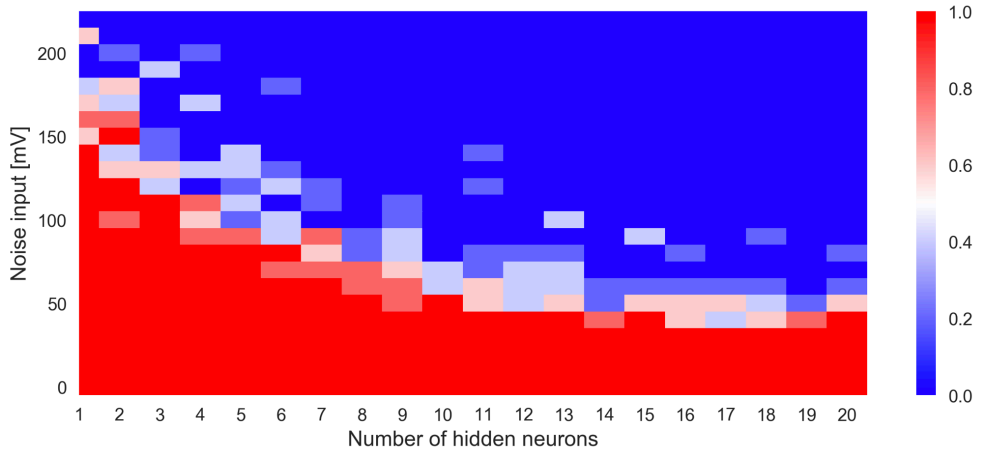

Figure 6.12: Dependence of the success rate on the noise input and the number of hidden neurons. The color indicates the mean success rate (n = 10). The lb-SNR was 0.4 (signal = 20, noise = 50).

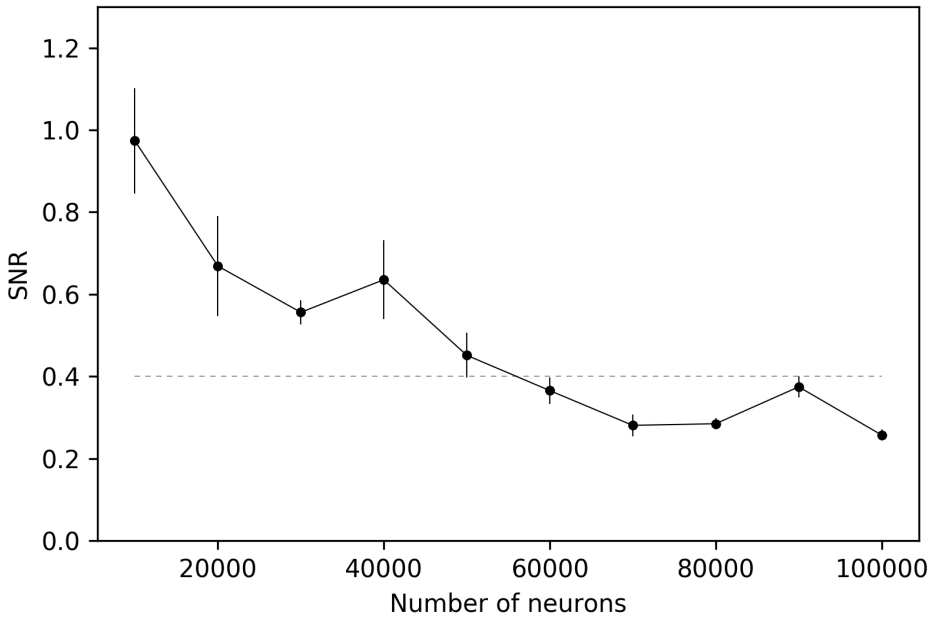

Figure 6.13: Dependence of estimated SNR on the number of neurons. The dotted line represents the value of lb-SNR. Error bars represent standard error of the mean (n = 20).

6.2.4 Discussion

In this study, we proposed a method to estimate the scalability of LSA from a network analysis approach focusing on the SNR, and estimated the scalability of LSA based on that method. The results suggested that the LSA scaled to larger network with 50,000-60,000 neurons.

In the learning experiments using cultured networks by Shahaf et al, they used the neuronal cultures with the cell number of about 10,000 to 50,000, and showed the network can learn such simple task [29]. The value estimated in this study is not inconsistent with the results. This support that the SAA in the cultured neural networks was obtained by LSA based on such local neural dynamics as STDP rather than the more macroscopic mechanism proposed in the previous studies; modifiability and stability [29, 28]. However, as this was estimated results, we need to simulate large spiking neural networks with more than 50,000 neurons to examine it in our future research.

The results in this section and previous sections also suggests that LSA

can work in animals that have the small number of neurons or the large number of neurons but modularized. Some animals with small nervous system such as *C. elegans*, *leech*, *Aplysia california*, and also cortical columns in some mammalian brain, have less than about 20,000 neurons [93, 94, 95], suggesting the LSA might be able to work in these systems. However, in LSA based on STDP, the timing of the input and the output should be closed in a specific time window (see Section3.4) since STDP can only work within specific short time window (e.g., less than 20 ms). Thus, whether LSA works depends not only on the network size but also on the network structure. We need to further study to examine whether the LSA works in such structured networks. In order to learn the relationship between input and output in the longer time window, it should be required other factors rather than STDP (e.g., short-term memory or dopamine system). In our future research, we need to discuss more about the relation between LSA and other functions observed in vivo rather than STDP.

6.3 Applying Learning by Stimulation Avoidance to Complex Embodiments

6.3.1 Introduction

In this section, we show our preliminary results in designing an autonomous agent applying LSA to a humanoid robot. As explained in the previous sections, LSA states that networks learn to avoid external stimuli by self-organizing the neural architecture (SAA). By adjusting the network architecture, the networks can actually remove the "cause" of the stimulus in the environment. In this sense, LSA is essentially embodied. It requires input from the environment, and the output from the network changes the environment. When the network removes the cause from the environment, the stimulus to the input stops and LTP in STDP works to reinforcement such behavior modifying the weight strength. On the other hand, when the network make the cause of the stimulus, the stimulus to inputs starts and LTD in STDP works to weakening such behavior modifying the weight strength. Thus the desired input-output relationship for avoiding stimulation is enhanced in the system. It is presumed that even if external input stimulation patterns change, the network learns new behavior to avoid the new stimulation. It is thus homeostatic adaptation [23, 24]. Applying LSA to a humanoid robot, the robot can autonomously learn an adaptive behavior without any external rewards system. In LSA, kind of intrinsic motivation for decreasing stimulation emerges from neural dynamics in a bottom-up manner and the

behavior of this autonomous robot is learned based on the intrinsic motivation (see the further discussion in Section 7.4)). We consider that this study is first step to build a autonomous humanoid robot.

In the previous sections, we did not apply various tasks to the agent. In this study, we applied LSA to the humanoid robot and conducted two learning experiments including more complex task than we done before, using spiking neural networks consists of 1,000 neurons. We demonstrate that LSA can be applicable to learn various tasks: i.e., the emerging behavior depends on the embodiment of the agents and environment.

We performed the experiments at the National Museum of Emerging Science and Innovation (Tokyo) between July 30 and August 6, 2016 [96] using humanoid robot named Alter (Fig. 6.14). In this "Exhibition of Alter," Alter was located in the middle of the exhibition room, where the audience could see its motions. We sampled the data collected from the motions of Alter, the spiking neural networks, and analyzed these data to understand how its behavior changed during experiments. Note that the research project for Alter is a collaboration work where our team developed the software and the other team developed the hardware, and in the exhibition, Alter consisted of not only spiking neural networks based on LSA but also coupled-phase oscillators as central pattern generator, and autonomous sensor networks as its sensory organs which make its internal dynamics based on the sensor input [97]. So far, although we were not only focused on LSA in the research project and the exhibition, in this chapter we focus on the learning experiments with only LSA, conducted independent of these other functions.

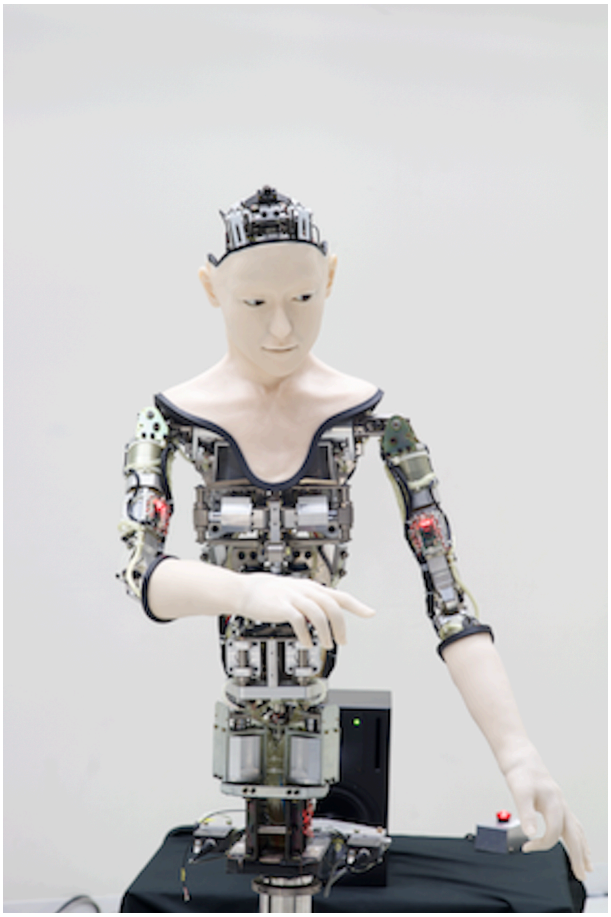

Figure 6.14: The appearance of the humanoid robot: Alter. The body was fixed to a pedestal. The total height is approximately 155 cm. (Photo by Kenshu Shintsubo.)

6.3.2 Methods

Alter's body had 42 movable air actuator axes, and its action production was controlled by an air compressor with a specially developed operating system for this humanoid (Geminoid server). Specifically, its action was controlled by special commands (e.g., MOVEAXIS AXISNAME (int 1-42) POSITION (int 1-255)). These commands were executed from an external computer using TCP/IP through a Geminoid server. Each axis had a potentiometer, which sends the actual value of the current axis's position to the computer. The refresh rate was every 50 msec.

In this study, we used the spiking neural networks 2.3 to generate the

motions of Alter's arms (six axes each). The networks consists of 1,000 neurons of which 80 % were excitatory neurons and 20% were inhibitory neurons. This ratio of 20% of inhibitory neurons is standard in simulations and close to real biological values [98]. Neurons in the network were fully connected; The initial weights were randomly chosen (uniform distribution: $0 < w < 5$ for excitatory neurons, $-5 < w < 0$ for inhibitory neurons). The only weights between excitatory neurons change based on STDP function (see Section 2.3.2). The all neurons received zero-mean Gaussian noise m with a standard deviation $\sigma = 3$ mV at each time step for a spontaneous spiking (see Section 2.3.3).

The network had input neurons and output neurons. The input neurons were chosen randomly from the excitatory neurons; 25 neurons per sensor input were chosen. They received stimuli from the distance sensors or the potentiometers; The input sources depended on the learning tasks. The output neurons were also chosen randomly from the excitatory neurons; 10 neurons per axis in the arms were chosen. The value of the axis was decided based on the number of firing neurons at each time step. Thus, Alter produces action by translating the micro-neural firing into macro-body motion.

We studied two tasks with Alter:

(i) "Raising hands" task: When Alter was surrounded by people, the input neurons were stimulated, and if Alter raised its hand, the stimulation removed for a while (1.65 sec). This can be interpreted as that Alter has a desire for greeting to surrounded people with raising its hands. The distance between Alter and a person was detected by distance sensor. The distance sensor value d was taken to determined the external stimuli e to each input neuron group at every time step as:

$$e = \begin{cases} d/2, & \text{if } d \geq 60 \\ 0, & \text{if } d < 60 \end{cases} \tag{6.3}$$

The success rate of this tasks S was measured as follows:

$$S = \frac{N_{success}}{N_{stimuli}} \tag{6.4}$$

where $N_{stimuli}$ represents the number of stimuli per minute and $N_{success}$ is the number of times that position of the hand exceed the threshold per minute.

(ii) "Smooth movement" task: Alter got stimulation, when the difference between the motion command and the actual value of potentiometers exceeded the threshold: i.e., if Alter moved its arm as it wanted, it receive no

stimulation. This can be regarded as that Alter has a desire for controlling its body as expected. We expected the neural network would learn to have smoother motions to avoid the internal stimulation. Since every axis was controlled by air pressure, both the real position (i.e., the value produced by the potentiometer) and the desired position sent by the computer was apt to be varied. The difference between these values was taken as external stimuli e to the input neuron group i at every time step as

$$e_i = \frac{k|C_i - P_i|}{255} \tag{6.5}$$

where C is the value sent from the computer (0-255) and P is the actual value of the potentiometer (0-255). Here, parameter k is set to 30.

The success rate of this tasks S was measured as follows:

$$S = \frac{N_{success}}{N_{stimuli}} \tag{6.6}$$

where $N_{stimuli}$ represents the number of stimuli per minute and $N_{success}$ represents the number of times that e fell below the threshold (0.1) per minute.

We performed the experiments at the National Museum of Emerging Science and Innovation (Tokyo) between July 30 and August 6, 2016. In this "Exhibition of Alter," Alter was located in the middle of the exhibition room, where the audience could see its motions (Fig. 6.15).

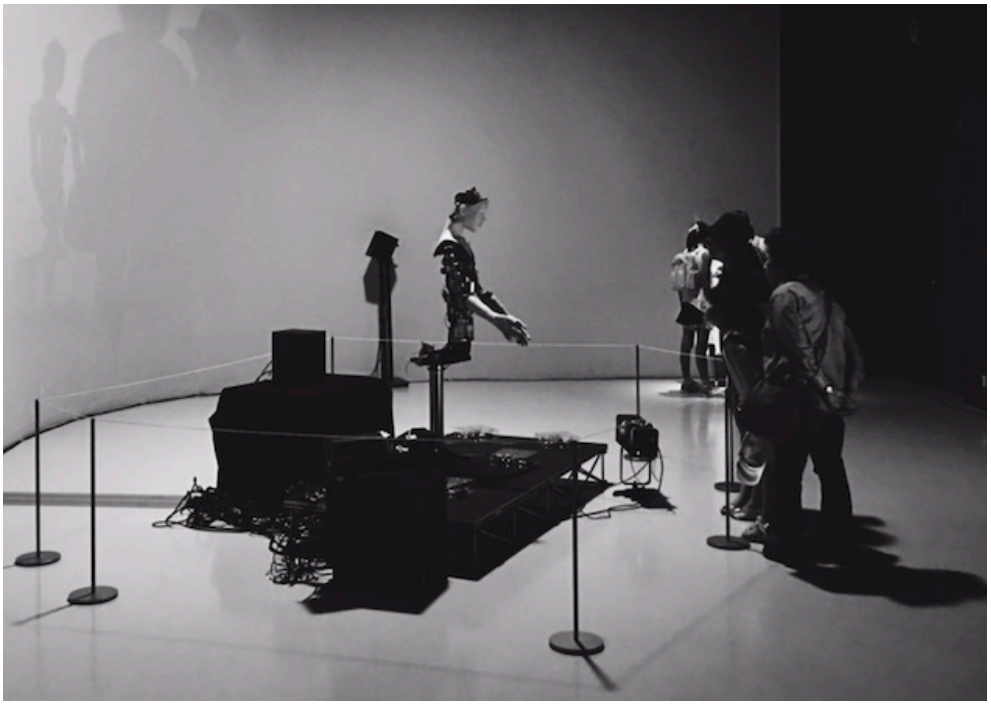

Figure 6.15: View of the Exhibition of Alter at the National Museum of Emerging Science and Innovation. The audience could clearly see Alter's motions. The distance sensors were installed at Alter's feet.

6.3.3 Results

Figure 6.16 shows the success rate of the "raising hands" task. The success rate was rapidly increased in the first 1000 sec and stabled at the higher value (near 70%). Therefore, Alter successfully learned to raise the hands to audience in front of Alter.

Figure 6.17 shows the success rate of the "smooth movement" task. The success rate was gradually increased to the higher value (near 75%) in the end; Over the course of the experiment, Alter learned and evolved so that it could move its hands much more smoothly in the end (i.e., matching the commands from the computer and the actual body motion). When Alter moved using pneumatic pressure, deviations occurred between the value sent from the computer (i.e., the brain) and the actual movement. In other words, there was a gap between Alter's intention and its movement. LSA worked well to resolve this discrepancy.

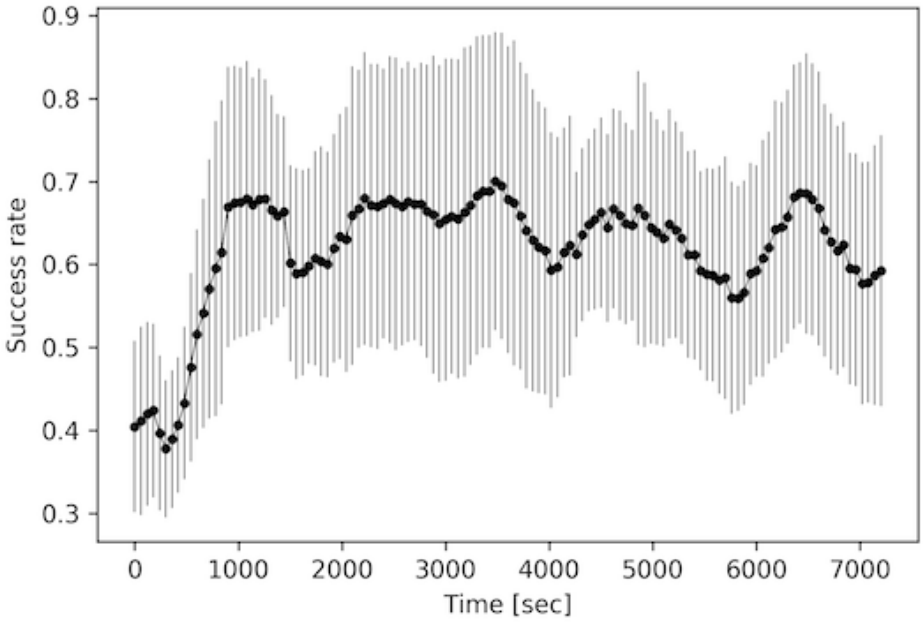

Figure 6.16: The time series of the success rate of the"raising hands" task. The horizontal axis represents time. The error bars represent the standard error.

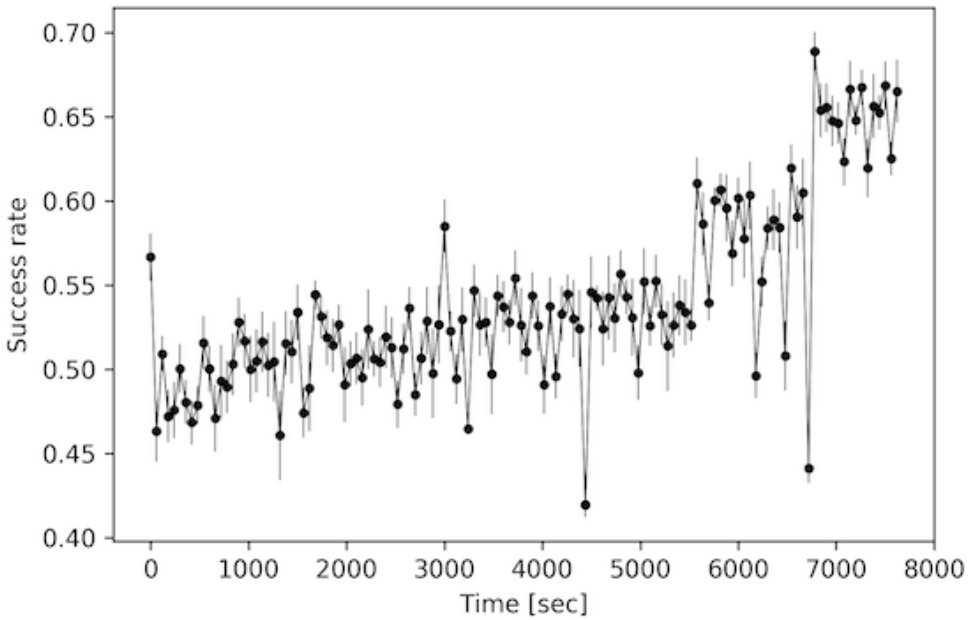

Figure 6.17: The time series success rate for the "smooth movement" task. The horizontal axis represents time . The error bars represent the standard error.

We also analyzed the time series for both neural activity and potentiometers in the experiment ("smooth movement" task). Figure 6.18 presents the results of the Principal Components Analysis (PCA) performed on every potentiometer. The colors represent the time course. As shown in this figure, Alter's body dynamics changed gradually over time. Figure 6.19 presents the time course of the neural activity in the PCA space. Their colors represent the time course and is consistent with those used in the previous figure. This figure also shows that the activity of the neural network changed gradually over time. In particular, the activity during the second half was more diverse than in the first half. On the other hands, the activity generated by the uniform random noise inputs (black dots) were widely spreading and there were no explicit time evolutions.

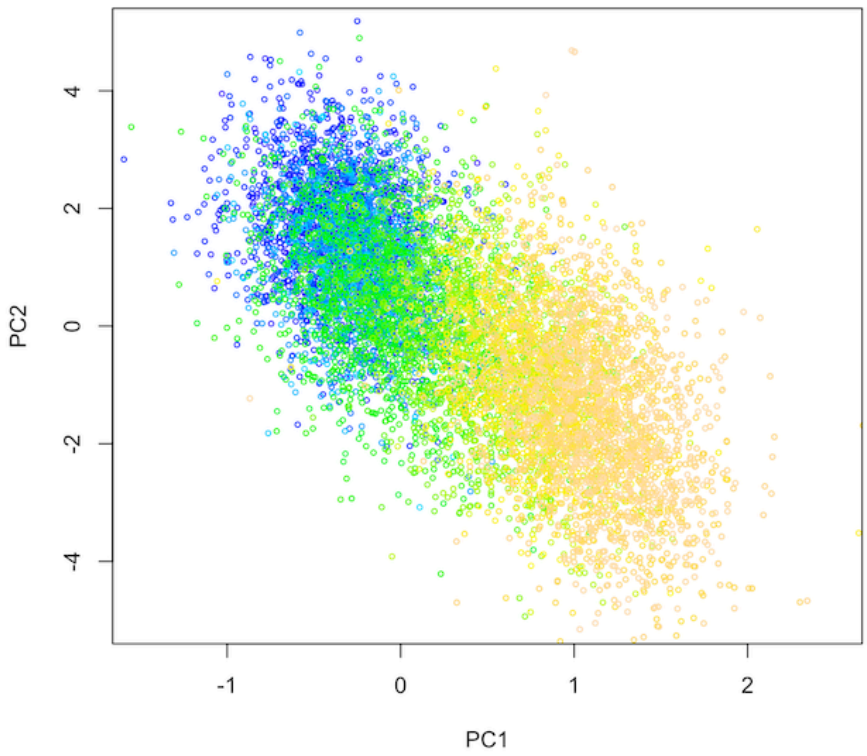

Figure 6.18: PCA of every potentiometer. Each point represents the average of 100 time steps of each potentiometer. The colors represent the time course (blue to yellow, through green) and the time interval is the same as that depicted in Fig 6.17.

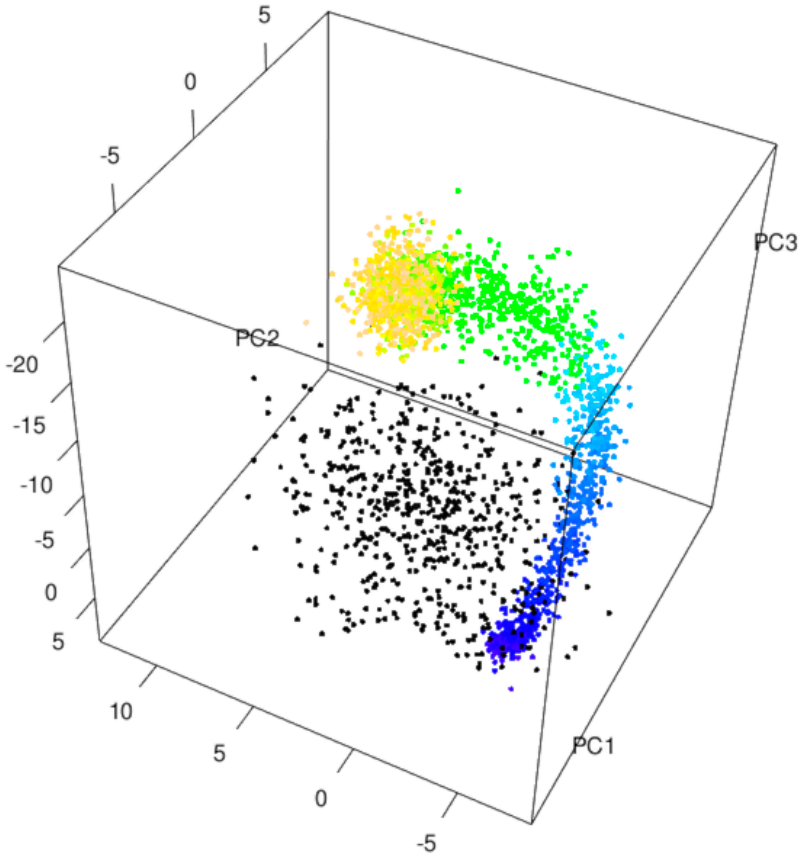

Figure 6.19: PCA of every neural activity. Each point describes the average of 100 time steps of every neuron. The colors describe the time course, and the time interval is the same as that depicted in Figure 6.17 (blue to yellow, through green). The black dots represent the pattern when the neural net received the uniform random noise (the range is the same as e (Eq. 6.5)).

6.3.4 Discussion

In this section, we applied LSA to the humanoid robot named Alter. As the results, we found that LSA works well with large network (1,000 neurons) for various tasks. Importantly, the networks should avoid the external stimulation based on LSA, however, the emergent behavior of the agent is varied depending on the design of the system (i.e., design of body, internal reward system, and environment). We need to further study about LSA coupled to the evolution of the embodiment in our future research.

During the exhibition, many people stopped and gazed at in front of Alter for a long time, and not a few people said that they did not lose interest in watching Alter. When we see living organism, we rarely lose interest in watching it, probably because it is an autonomous system and shows unexpected behavior occasionally. We believe that similar phenomena happened in Alter and this project is first step to build an autonomous humanoid robot.

Chapter 7

General Discussion

7.1 Summary of Thesis

In this thesis, we studied on homeostasis, particularly in regards to synaptic plasticity in both biological neuronal cultures and artificial spiking neural networks. Significantly, on the basis of synaptic plasticity, such as STDP, we found that three functions of avoiding stimulation (which we name Stimulation Avoidance by Action (SAA), Stimulation Avoidance by Prediction (SAP), Stimulation Avoidance by Selection (SAS)) which lead to homeostasis emerged in the embodied neural networks. We argue that these three functions can emerge in a same neural network (like Fig. 5.11), and which function emerges depends on a quality of stimuli: controllable inputs induce action, predictable inputs induce prediction and uncontrollable inputs (noise) induce selection. In other words, various functions emerge from stimulation avoidance as a principle of behavior. We termed these emerging properties of neural homeostasis as Principle of Stimulation Avoidance (PSA).

In each chapter, we studies these three aspects of stimulus avoidance. In Chapter 3, we showed that stimulation avoidance was self-organized by action (SAA) in embodied cultured neural networks even with a smaller number of neurons than the previous work [29]. The analysis of neural dynamics suggests that SAA in neuronal cultures is realized by changing their network structures with synaptic plasticity.

In Chapter 4 we demonstrated that SAP emerged in the spiking neural networks with STDP. The networks learned to predict sequence patterns of stimulation. We also demonstrated SAP scaled to the larger random networks consists of 100 neurons without preparing well-designed structures. This suggests that SAP can also emerge in primitive biological neural networks, such as dissociated neuronal cultures.

In Chapter 5, we found that neuronal cultures have a property of stimulation avoidance by selecting which external information is received or declined (SAS). We demonstrated that the spiking neural networks with asymmetric-STDP can reproduce such a behavior. Even if the shape of the asymmetric-STDP is similar to the shape broadly observed in vitro, SAS was also observed. Therefore, the same mechanism should work in biological neural networks, such as dissociated neuronal cultures.

In Chapter 6, we estimated LSA scales to about 50,000 neurons, and we applied LSA to humanoid robot to demonstrate that various behaviors emerged depending on different embodiments.

These three properties (SAA, SAP, SAS) emerge from a simple Hebbian rule (such as STDP), maintaining the homeostasis of the systems. As stimulation to an agent represents environmental changes, avoiding stimulation decreases the influence of environmental change on an agent's internal state. It is interesting that these three kinds of homeostasis directly emerge from the same synaptic plasticity, such as STDP in biological and biologically inspired neural networks.

In the following sections, we discuss these findings from some other perspectives.

7.2 Evolutionary Perspective

Action, prediction, and classification are three useful functions for embodied agents that must learn from their environment. In Chapter 4, we discussed the evolutionary pathway from a reactive agent (action) to a predictive agent and a proactive agent (prediction). In this section, we further extend this discussion to an inductive agent (classification).

Even without neural plasticity, an agent can act reactively, using hard-wired abilities. However, such hard-wired reactive behavior can have bad consequences if environmental changes happen in a short time scale (i.e., the scale of an individual life).. For example, a behavior resulting in a food reward might result in getting poisoned in the future. In such environments, learning reactive behaviors is an effective way to help an agent survive. When excitatory neurons and synaptic plasticity evolve, agents can learn their sensory-motor coupling to adapt to the environment. As shown in Chapter 3 and Chapter 5, SAA and SAS emerge from random networks comprising excitatory neurons with STDP; thus, those two type of homeostasis can be realized at that time frame. At this evolutionary stage, agents can avoid environmental changes to protect themselves, and the agents can also regulate their self-boundaries by selecting connections, such as separating

connections from uncontrollable neurons to stabilize the network.

As we discussed in Chapter 4, in addition to the learning actions, a prediction can lead to better adaptation of the agents. The reactive agent discussed above cannot initiate an action before receiving an actual stimulus. For example, even if the agent could learn a reaction to a undesirable stimulus (e.g., something damaging to the agent), the agent cannot avoid the damage before actually getting the stimuli. In such situations, if the agent can predict the incoming stimulus, the agent can initiate avoidance behavior before being damaged. To enhance the adaptability of a reactive agent, that agent and its predictive network must be integrated into a proactive agent. As shown in Chapter 4, random networks comprise excitatory neurons and inhibitory neurons learn to predict stimulus sequences based on STDP. Thus, to realize such a proactive agent, inhibitory neurons must be evolved in addition to excitatory neurons explained above. At this evolutionary stage, the proactive agents can predict some events and initiate a behavior before the events actually occur.

For proactive agents, generalization is an advanced way to improve adaptability. For example, the proactive agents without generalization can only make a prediction if its current inputs are exactly the same as previously learned inputs. Therefore, the ability to generalize is an advantage when the environment is variable or noisy. To realize this ability, the agent must learn to abstract relevant signals from variable inputs. This can be achieved by simple statistical grouping (classification) of inputs. A simple way to classify the various inputs is to convolute the inputs to a single output. At the upper layer of convolution, some similar input patterns can be treated as identical. A neural network can achieve the integration of inputs through neurons that take several inputs and have a single output. Therefore, in addition to the random network comprising excitatory neurons and inhibitory neurons, such a specific type of excitatory neuron is required to realize an inductive agent. At this evolutionary stage, with such a classification, the inductive agent becomes more adaptive. In our future work, we must conduct further studies to show that the proactive agent and the inductive agent also actually emerge with the proposed framework.

Although we have discussed the evolutionary pathway from reactive agent to inductive agent, our proposed minimal system is so primitive that agents can only execute simple actions, predictions, classifications, or combinations thereof. A link from primitive agent to higher-order cognitive functions is still missing.

What is necessary for higher-order cognitive functions, such as consciousness or mind? Although, we used a simple embodiment in our research, a more complex embodiments might be important. For example, Damasio

proposed a three-layered consciousness: proto self, core consciousness, and extended consciousness [99].

Proto self is still not a conscious state but a pre-conscious state and a base of consciousness. The proto self is regarded as a collection of neural patterns that represent the body's internal state. Thus, the primitive agent discussed above might be placed in this stage.

Core consciousness is the conscious state of living organisms that are only self-aware in the "here and now". This consciousness strongly relates to emotions and feeling. In Damasio's theory, emotion is regarded as an unconscious body reaction to external stimuli (e.g., cursing and crying); feeling is a state that senses the changes caused by emotions. With core consciousness, the organisms can be aware of the feeling of emotion. This consciousness is important for agents to decide their action based on the feeling of emotion. Emotion can be regarded as an evolutionarily designed, hard-coded signal of the body to evaluate the current situation, while feeling is modifiable based on experience with emotion. Core consciousness cannot be separated from emotion and feeling; thus, to evolve this type of consciousness requires at least coupling neural networks and body that cause emotions.

Extended consciousness is a conscious state that is self-aware not only in the "here and now" but also of the past and future, of imaginary states, and the act of inferring another's mind. This consciousness relates to working memory, long-term memory, and language; furthermore, it could not emerge without the predecessors: proto self and core consciousness. To evolve this type of consciousness, the neural networks must at least combine the complex embodiment described above with working memory and long-term memory.

This kind of higher-order cognitive function produces homeostasis effectively. In our future work, we must study about relation with our framework and such higher-order cognitive systems (e.g., evolutionary process coupling the neural network to such complex embodiment, and applying working memory and long-term memory to the systems).

7.3 Stimulation Avoidance in Brain

In Section 6.2, we estimated SAA based on LSA scales to about 50,000 neurons. Thus, LSA might work in animals with a central nervous system comprising less than 50,000 neurons—for exmaple, *C. elegans* (302 neurons) [93], *leech* (about 8,400 neurons) [94], and *Aplysia californica* (less than 10,000 neurons) [95]. Even in systems comprising more than 50,000 neurons, if there are small modular structures, such as functional columns in the human brain, LSA might work effectively. For SAP, it would be possible to realize a chain

of learning in which the output of one module can also be an input of the next module, provided that the second module predicts and suppresses the first module's input. For SAA, however, a duration from stimulus reception to output must be closed within a certain time frame since STDP can only work between neurons spiked in a short time (e.g., 20 ms). For example, preliminary results of the parameter search in the networks consist of 100 neurons show that the input and the output are must be closed within 40 ms to learn a behavior that avoids the an environmental stimulus. For SAA to work on a large network, it seems that other functions (e.g., short-term memory or dopaminergic systems) are required. further discuss the possibility of LSA in the brain, we must especially study the relation between LSA and well-structured networks, as well as the relation between LSA and functions other than than STDP.

SAA, SAP, and SAS, proposed in this thesis, are based on STDP. STDP has been found in vitro and in vivo [33, 100], though, as yet STDP can only be observed in limited conditions [101], and some experimental results for biological neurons cannot be explained by classic STDP (e.g., plasticity in the case of triplet of the spikes of a presynaptic neuron and a postsynaptic neuron, or that of quadruplet [102]). In recent years, memristor has attracted much attention as a voltage-based synaptic plasticity model. Memristor is the fourth passive element (after register, capacitor, and inductor), in which the conductance changes according to a potential difference between either end of memristor. Some studies shown that various types of synaptic dynamics, including STDP can be reproduced by a combination of wave forms connected to either ends of memristor [103]. Memristor can also reproduce the synaptic dynamics of triplets [104] which cannot be reproduced by classic STDP. These studies imply that memristor might be more versatile as a synaptic model than computational models of traditional interpretations of STDP are. In our future research, we must study whether homeostatic behaviors like SAA, SAP, and SAS can also emerge from the dynamics of models such as memristor, which reproduces a greater variety of behaviors for biological synaptic plasticity than classic STDP.

Although some forms of stimulation avoidance emerge from the classic STDP function we used in this thesis, other types should emerge from other synaptic dynamics. For example, if the curve of classic STDP functions are reflected on the x-axis, then the agent should have a characteristic utterly opposed to classic STDP: seeking stimulation from the environment. However, such shapes of STDP function have not been broadly found in vivo and in vitro. From an evolutionary perspective, this implies that stimulation avoidance (not stimulation seeking) behavior has contributed to maintaining homeostasis in agents. However, various type of synaptic dynamics rather

than classic STDP have been found in vivo and in vitro [33, 100]; therefore, we must study what kind of behavior emerges when combining these various types of synaptic dynamics in our future work.

7.4 Stimulation Avoidance as Intrinsic Motivation

LSA is a basic mechanism of an autonomous agent trying to maintain homeostasis by autonomously avoiding stimulation without global rewards. This can also be regarded as an intrinsic motivation that emerges from local dynamics, such as STDP, in a bottom-up manner.

In general, an agent's motivation is divided into two types of motivation: extrinsic motivation and intrinsic motivation. "Extrinsic motivation is a construct that pertains whenever an activity is done in order to attain some separable outcome" [105]. On the other hand, "intrinsic motivation is defined as the doing of an activity for its inherent satisfactions rather than for some separable consequence. [$\cdots$] for the fun or challenge entailed rather than because of external prods, pressures, or rewards" [105]. From an evolutionary perspective, intrinsic motivation should also lead to some separable consequence (not directly but indirectly) with greater causal distance. Furthermore, the separable consequences from intrinsic motivations are less specific than those of an extrinsic motivation (e.g., food), their generality leading to open-ended or lifelong learning.

In recent years, intrinsic motivations have attracted much attention in the field of Artificial Intelligence, with researchers trying to create autonomous agents, in particular in the field of reinforcement learning [73, 74]. Reinforcement learning is a biologically inspired learning model that reinforces actions that could receive rewards [70]. One of the critical problems of reinforcement learning is that until acquiring rewards, the systems cannot learn any behaviors (e.g., if the reward space is sparse, it is difficult to acquire rewards and learn actions without external instructions). To learn such difficult tasks, he agent must receive some manually designed curriculum in which the difficulty level of the task gradually increases.

Applying intrinsic motivations that drives to satisfy the agent's own desires (internal reward) can, to some extent, solve such problems in reinforcement learning (e.g., even if the reward is sparse, the agent can learn new behaviors according to its intrinsic motivations). Finally, the agent could be rewarded for using combinations of learned behaviors. In this way, intrinsic motivations seems to play important roles for autonomous agents. Ideally,

intrinsic motivation should lead to open-ended, lifelong learning (i.e., with an intrinsic motivation, the agent should continue to adapt well to the environment). Applying intrinsic motivations to reinforcement learning might lead to an autonomous agent that learns throughout its life. In our future works, we must further study how to integrate the PSA with such reward systems (e.g., dopaminergic system).

Many studies can be regarded as models of such intrinsic motivations [106], and the motivations is often given by manually in a top-down manner (e.g., empowerment maximization [107], free energy minimization [15, 16, 17] and novelty-seeking behavior[108]). In living organisms, it is impossible to given intrinsic motivation in a top-down manner, and it can only be hard-coded in the body to feel pleasure (or displeasure) in the form of emotions and the systems control (e.g., maximize or minimize) it. In that case, emotions can be regarded as a hard-wired signal evolved to evaluate situations.

Some studies take an evolutionary approach to optimize intrinsic motivations [109, 110]. Feeling emotions and generating actions based on the emotions are sophisticated processes, requiring higher-order cognitive systems, like a mind or consciousness (see Section 7.2). For more primitive systems without such higher-order cognition, intrinsic motivations can be regarded as a function of action generation, which does not directly lead to specific separable consequence (e.g., food) but can indirectly lead to a greater variety of separable consequences (e.g., open-ended learning). The PSA is not given in a top-down manner, and it can be regarded as a primitive intrinsic motivation, such as stimulation avoidance, which emerges from synaptic dynamics, such as STDP, in a bottom-up manner. The characteristic of stimulation avoidance can also be regarded as a kind of primitive emotions or origin of an agent's capacity to like or dislike something. It is interesting that such a primitive emotion can emerge with the dynamics of synaptic plasticity for even a system without an explicit emotion (e.g., caused by chemicals like dopamine or serotonin). In our future research, we must discuss in greater detail the evolutionary path from primitive emotions and motivations to more explicit emotions, motivations, and consciousness.

Bibliography

[1] Aleksandr Ivanovich Oparin. *Proceedings of International Symposium on the Origin of Life on the Earth*. Pergamon, New York, 1959.

[2] B Cannon. Organization for physiological homeostasis. *Physiological Reviews*, IX(3):399–431, 1929.

[3] Norbert Wiener. Cybernetics : or control and communication in the animal and the machine, 1961.

[4] W. Ross Ashby. *Design for a Brain: The Origin of Adaptive Behavior*. Chapman and Hall, 2 edition, 1960.

[5] Daniel Clement Dennett. *Darwin's Dangerous Idea: Evolution and the Meanings of Life*. Simon & Schuster, New York, 1995.

[6] W. Ross Ashby. *An Introduction to Cybernetics*. Chapman & Hall, London, 1956.

[7] Andre M Bastos, W Martin Usrey, Rick A Adams, George R Mangun, Pascal Fries, and Karl J Friston. Canonical Microcircuits for Predictive Coding, 2012.

[8] Andy Clark. Whatever next? predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3):181–204, 2013.

[9] Igor Aleksander. *World in My Mind, My Mind in the World: Key Mechanisms of Consciousness in People, Animals and Machines*. Imprint Academic, 2005.

[10] W. Schultz, P. Dayan, and P. R. Montague. A neural substrate of prediction and reward. *Science*, 275(5306):1593–1599, 1997.

[11] Edward N. Lorenz. Deterministic Nonperiodic Flow. *Journal of the Atmospheric Sciences*, 20(2):130–141, 1963.

[12] Rajesh P N Rao and Dana H Ballard. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience*, 2(1):79–87, 1999.

[13] Karl Friston, James Kilner, and Lee Harrison. A free energy principle for the brain. *Journal of Physiology Paris*, 100(1-3):70–87, 2006.

[14] Karl Friston. The free-energy principle: A unified brain theory?, feb 2010.

[15] Karl Friston, Jérémie Mattout, and James Kilner. Action understanding and active inference. *Biological cybernetics*, 104(1-2):137–60, feb 2011.

[16] Karl Friston. Active inference and agency. *Cognitive neuroscience*, 5(February 2015):119–21, 2014.

[17] Karl Friston, Thomas FitzGerald, Francesco Rigoli, Philipp Schwartenbeck, John O'Doherty, and Giovanni Pezzulo. Active inference and learning. *Neuroscience and Biobehavioral Reviews*, 68:862–879, 2016.

[18] H. B. BARLOW. Possible principles underlying the transformation of sensory messages. *Sensory communication*, 1961.

[19] Laurent Itti and Pierre Baldi. Bayesian surprise attracts human attention. *Vision Research*, 49(10):1295–1306, 2009.

[20] Ralph Linsker. Self-organization in a perceptual network. *Computer*, 21(3):105–117, mar 1988.

[21] Ralph Linsker. Perceptual neural organisation: some approaches based on network models and information theory. *Annu. Rev. Neurosci.*, 13:257–281, 1990.

[22] Ezequiel A. Di Paolo. Homeostatic adaptation to inversion of the visual field and other sensorimotor disruptions. *From Animals to Animats VI: Proceedings of the 6th International Conference on Simulation of Adaptive Behavior*, pages 440–449, 2000.

[23] H. Iizuka and E. a. Di Paolo. Toward Spinozist Robotics: Exploring the Minimal Dynamics of Behavioral Preference. *Adaptive Behavior*, 15(4):359–376, dec 2007.

[24] Ezequiel A. Di Paolo and Hiroyuki Iizuka. How (not) to model autonomous behaviour. *BioSystems*, 91:409–423, 2008.

[25] Takashi Ikegami and Keisuke Suzuki. From a homeostatic to a homeodynamic self. *Biosystems*, 91(2):388–400, feb 2008.

[26] Clark. L. Hull. *Principles of behavior: an introduction to behavior theory.* Appleton-Century Company, Oxford, England, 1943.

[27] Edwin R Guthrie and Christopher D Green. Psychological Facts and Psychological Theory. *Psychological Bulletin*, 43(1):1–20, 1946.

[28] Shimon Marom and Goded Shahaf. Development, learning and memory in large random networks of cortical neurons: lessons beyond anatomy. *Quarterly reviews of biophysics*, 35:63–87, 2002.

[29] G Shahaf and S Marom. Learning in networks of cortical neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 21(22):8782–8, nov 2001.

[30] Lana Sinapayen, Atsushi Masumori, Nathaniel Virgo, and Takashi Ikegami. Learning by Stimulation Avoidance as a Primary Principle of Spiking Neural Networks Dynamics. In *The 13th European Conference on Artificial Life (ECAL 2015)*, pages 175–182, 2015.

[31] Lana Sinapayen, Atsushi Masumori, and Takashi Ikegami. Learning by stimulation avoidance: A principle to control spiking neural networks dynamics. *PLOS ONE*, 12(2):e0170388, feb 2017.

[32] Guo-qiang Bi and Mu-ming Poo. Synaptic modifications in cultured hippocampal neurons: dependence on spike timing, synaptic strength, and postsynaptic cell type. *The Journal of neuroscience*, 18(24):10464–10472, 1998.

[33] Natalia Caporale and Yang Dan. Spike timing-dependent plasticity: a Hebbian learning rule. *Annual review of neuroscience*, 31:25–46, jan 2008.

[34] Li I Zhang, Hui-zhong W Tao, and Mu-ming Poo. Visual input induces long-term potentiation of developing retinotectal synapses. *Nature Neuroscience*, 3:708, jul 2000.

[35] Vincent Jacob, Daniel J Brasier, Irina Erchova, Dan Feldman, and Daniel E Shulz. Spike timing-dependent synaptic depression in the in vivo barrel cortex of the rat. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 27(6):1271–84, feb 2007.

[36] Blake Richards, Carlos Aizenman, and Colin Akerman. In vivo spike-timing-dependent plasticity in the optic tectum of xenopus laevis. *Frontiers in Synaptic Neuroscience*, 2:7, 2010.

[37] S Song, K D Miller, and L F Abbott. Competitive Hebbian learning through spike-timing-dependent synaptic plasticity. *Nature neuroscience*, 3(9):919–26, sep 2000.

[38] Dean V. Buonomano. Decoding temporal information: A model based on short-term synaptic plasticity. *Journal of Neuroscience*, 20(3):1129–1141, 2000.

[39] Rajesh P. N. Rao and Terrence J. Sejnowski. Spike-timing-dependent hebbian plasticity as temporal difference learning. *Neural Comput.*, 13(10):2221–2237, October 2001.

[40] Catherine Wacongne, J.-P. Changeux, and Stanislas Dehaene. A Neuronal Model of Predictive Coding Accounting for the Mismatch Negativity. *Journal of Neuroscience*, 32(11):3665–3678, 2012.

[41] Johanni Brea, Walter Senn, and Jean-Pascal Pfister. Matching recall and storage in sequence learning with spiking neural networks. *Journal of Neuroscience*, 33(23):9565–9575, 2013.

[42] Steve M Potter and Thomas B DeMarse. A new approach to neural cell culture for long-term studies. *Journal of Neuroscience Methods*, 110(1-2):17–24, sep 2001.

[43] Danny Eytan and Shimon Marom. Dynamics and effective topology underlying synchronization in networks of cortical neurons. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 26(33):8465–76, aug 2006.

[44] Radhika Madhavan, Zenas C Chao, and Steve M Potter. Plasticity of recurring spatiotemporal activity patterns in cortical networks. *Physical biology*, 4(3):181–93, sep 2007.

[45] Douglas J Bakkum, Zenas C Chao, and Steve M Potter. Spatio-temporal electrical stimuli shape behavior of an embodied cortical network in a goal-directed learning task. *Journal of neural engineering*, 5(3):310–23, sep 2008.

[46] Marco Canepari, Marco Bove, Eisaku Maeda, Marco Cappello, and Akio Kawana. Experimental analysis of neuronal dynamics in cultured

cortical networks and transitions between different patterns of activity. *Biological Cybernetics*, 77(2):153–162, aug 1997.

[47] T.B. DeMarse and K.P. Dockendorf. Adaptive flight control with living neuronal networks on microelectrode arrays. *Proceedings. 2005 IEEE International Joint Conference on Neural Networks, 2005.*, 3:1548–1551, 2005.

[48] A Novellino, P D'Angelo, L Cozzi, M Chiappalone, V Sanguineti, and S Martinoia. Connecting neurons to a mobile robot: an in vitro bidirectional neural interface. *Computational intelligence and neuroscience*, 2007:12725, jan 2007.

[49] R M R Pizzi, D Rossetti, G Cino, D Marino, A L Vescovi, and W Baer. A cultured human neural network operates a robotic actuator. *Bio Systems*, 95(2):137–44, feb 2009.

[50] Isao Hayashi, Minori Tokuda, Ai Kiyohara, Takahisa Taguchi, and Suguru N. Kudoh. Biomodeling System: Interactive Connection between Cultured Neuronal Network and Moving Robot Using Fuzzy Interface. *Journal of Japan Society for Fuzzy Theory and Intelligent Informatics*, 23(5):761–772, 2011.

[51] Terrence J Sejnowski, Christof Koch, and Patricia S Churchland. Computational Neuroscience. *Science, New Series*, 241(4871):1299–1306, 1988.

[52] Eugene M. Izhikevich. Which model to use for cortical spiking neurons? *IEEE Transactions on Neural Networks*, 15(5):1063–1070, 2004.

[53] Romain Brette, Michelle Rudolph, Ted Carnevale, Michael Hines, David Beeman, James M Bower, Markus Diesmann, Abigail Morrison, Philip H Goodman, Frederick C Harris, Milind Zirpe, Thomas Natschläger, Dejan Pecevski, Bard Ermentrout, Mikael Djurfeldt, Anders Lansner, Olivier Rochel, Thierry Vieville, Eilif Muller, Andrew P Davison, Sami El Boustani, and Alain Destexhe. Simulation of networks of spiking neurons: a review of tools and strategies. *Journal of computational neuroscience*, 23(3):349–98, dec 2007.

[54] Yang Dan and Mu-Ming Poo. Spike timing-dependent plasticity: from synapse to perception. *Physiological reviews*, 86(3):1033–48, jul 2006.

[55] A L Hodgkin and A F Huxley. A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4):500–544, aug 1952.

[56] C. Mead. Neuromorphic electronic systems. *Proceedings of the IEEE*, 78(10):1629–1636, Oct 1990.

[57] Douglas J. Bakkum, Urs Frey, Milos Radivojevic, Thomas L. Russell, Jan Müller, Michele Fiscella, Hirokazu Takahashi, and Andreas Hierlemann. Tracking axonal action potential propagation on a high-density microelectrode array across hundreds of sites. *Nature Communications*, 4:2181 EP –, 07 2013.

[58] Yuichiro Yada, Ryohei Kanzaki, and Hirokazu Takahashi. State-dependent propagation of neuronal sub-population in spontaneous synchronized bursts. *Frontiers in Systems Neuroscience*, 10:28, 2016.

[59] Yuichiro Yada, Takeshi Mita, Akihiro Sanada, Ryuichi Yano, Ryohei Kanzaki, Douglas J. Bakkum, Andreas Hierlemann, and Hirokazu Takahashi. Development of neural population activity toward self-organized criticality. *Neuroscience*, 343:55 – 65, 2017.

[60] Urs Frey, Jan Sedivy, Flavio Heer, Rene Pedron, Marco Ballini, Jan Mueller, Douglas Bakkum, Sadik Hafizovic, Francesca D. Faraci, Frauke Greve, Kay-Uwe Kirstein, and Andreas Hierlemann. Switch-Matrix-Based High-Density Microelectrode Array in CMOS Technology. *IEEE Journal of Solid-State Circuits*, 45(2):467–482, feb 2010.

[61] Daniel a Wagenaar, Radhika Madhavan, Jerome Pine, and Steve M Potter. Controlling bursting in cortical cultures with closed-loop multi-electrode stimulation. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 25(3):680–8, jan 2005.

[62] Daniel A Wagenaar and Steve M Potter. Real-time multi-channel stimulus artifact suppression by local curve fitting. *Journal of neuroscience methods*, 120(2):113–20, oct 2002.

[63] E.M. Izhikevich. Simple model of spiking neurons. *IEEE Transactions on Neural Networks*, 14(6):1569–1572, 2003.

[64] Gianluigi Mongillo, Omri Barak, and Misha Tsodyks. Synaptic theory of working memory. *Science (New York, N.Y.)*, 319(5869):1543–1546, 2008.

[65] Atsushi Masumori, Lana Sinapayen, and Takashi Ikegami. Learning by stimulation avoidance scales to large neural networks. In *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*, pages 275–282, Cambridge, MA, sep 2017. MIT Press.

[66] Rodney A Brooks. Intelligence without representation. *Artificial Intelligence*, 47(1-3):139–159, 1991.

[67] Karl Sims. Evolving Virtual Creatures. In *Proceedings of the 21st annual conference on Computer graphics and interactive techniques (SIGGRAPH '94)*, pages 15–22, 1994.

[68] Josh C Bongard. Evolutionary robotics. *Communications of the ACM*, 56(8):74, aug 2013.

[69] R.S. Sutton and A.G. Barto. Toward a modern theory of adaptive networks: expectation and prediction. *Psychol.Rev.*, 88(2):135–170, 1981.

[70] Richard S. Sutton and Andrew G Barto. *Reinforcement learning: an introduction.* MIT Press, Cambridge, MA, 1998.

[71] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, feb 2015.

[72] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, Yutian Chen, Timothy Lillicrap, Fan Hui, Laurent Sifre, George van den Driessche, Thore Graepel, and Demis Hassabis. Mastering the game of Go without human knowledge. *Nature*, 550(7676):354–359, oct 2017.

[73] Marc Bellemare, Sriram Srinivasan, Georg Ostrovski, Tom Schaul, David Saxton, and Remi Munos. Unifying count-based exploration and intrinsic motivation. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 1471–1479. Curran Associates, Inc., 2016.

[74] Deepak Pathak, Pulkit Agrawal, Alexei A. Efros, and Trevor Darrell. Curiosity-driven exploration by self-supervised prediction. In *ICML*, 2017.

[75] S. Kullback and R. A. Leibler. On information and sufficiency. *Ann. Math. Statist.*, 22(1):79–86, 03 1951.

[76] Thomas Schreiber. Measuring Information Transfer. *Physical Review Letters*, 85(2):461–464, jul 2000.

[77] Shinya Ito, Michael E. Hansen, Randy Heiland, Andrew Lumsdaine, Alan M. Litke, and John M. Beggs. Extending transfer entropy improves identification of effective connectivity in a spiking cortical network model. *PLOS ONE*, 6(11):1–13, 11 2011.

[78] Clarissa J. Whitmire and Garrett B. Stanley. Rapid Sensory Adaptation Redux: A Circuit Perspective. *Neuron*, 92(2):298–315, 2016.

[79] Nadine S. Simons-Weidenmaier, Maruschka Weber, Claudia F. Plappert, Peter KD Pilz, and Susanne Schmid. Synaptic depression and short-term habituation are located in the sensory part of the mammalian startle pathway. *BMC Neuroscience*, 7(1):38, May 2006.

[80] Atsushi Masumori, Norihiro Maruyama, Lana Sinapayen, Takeshi Mita, Urs Frey, Douglas Bakkum, Hirokazu Takahashi, and Takashi Ikegami. Learning by Stimulation Avoidance in Cultured Neuronal Cells. In *The 2nd International Symposium on Swarm Behavior and Bio-Inspired Robotics (SWARM2017)*, 2017.

[81] Atsushi Masumori, Norihiro Maruyama, Lana Sinapayen, Takeshi Mita, Urs Frey, Douglas Bakkum, Hirokazu Takahashi, and Takashi Ikegami. Emergence of Sense-Making Behavior by the Stimulus Avoidance Principle: Experiments on a Robot Behavior Controlled by Cultured Neuronal Cells. *Proc. of the European Conference on Artificial Life (ECAL) 2015*, pages 373–380, 2015.

[82] Stijn Cassenaer and Gilles Laurent. Hebbian STDP in mushroom bodies facilitates the synchronous flow of olfactory information in locusts. *Nature*, 448(7154):709–13, aug 2007.

[83] Nachum Ulanovsky, Liora Las, and Israel Nelken. Processing of low-probability sounds by cortical neurons. *Nature Neuroscience*, 6:391 EP –, 03 2003.

[84] Tomoyo Isoguchi Shiramatsu, Ryohei Kanzaki, and Hirokazu Takahashi. Cortical mapping of mismatch negativity with deviance detection property in rat. *PLOS ONE*, 8(12):1–10, 12 2013.

[85] Takahiro Noda, Ryohei Kanzaki, and Hirokazu Takahashi. Amplitude and phase-locking adaptation of neural oscillation in the rat auditory cortex in response to tone sequence. *Neuroscience Research*, 79:52 – 60, 2014.

[86] Gloria G. Parras, Javier Nieto-Diego, Guillermo V. Carbajal, Catalina Valdés-Baizabal, Carles Escera, and Manuel S. Malmierca. Neurons along the auditory pathway exhibit a hierarchical organization of prediction error. *Nature Communications*, 8(1):2148, 2017.

[87] Sooyoung Chung, Xiangrui Li, and Sacha B. Nelson. Short-term depression at thalamocortical synapses contributes to rapid adaptation of cortical sensory responses in vivo. *Neuron*, 34(3):437 – 446, 2002.

[88] J. Kevin O'Regan and Alva Noë. A sensorimotor account of vision and visual consciousness. *Behavioral and Brain Sciences*, 24(5):939–973, 2001.

[89] Humberto R. Maturana and Francisco J. Varela. *Autopoiesis and cognition : the realization of the living / Humberto R. Maturana and Francisco J. Varela ; with a preface to "Autopoiesis" by Sir Stafford Beer.* D. Reidel Pub. Co Dordrecht, Holland ; Boston, 1980.

[90] F.G. Varela, H.R. Maturana, and R. Uribe. Autopoiesis: The organization of living systems, its characterization and a model. *Biosystems*, 5(4):187 – 196, 1974.

[91] Niels K. Jerne. Towards a network theory of the immune system. *Annales d'immunologie*, 125C 1-2:373–89, 1974.

[92] Atsushi Masumori, Lana Sinapayen, Norihiro Maruyama, Takeshi Mita, Douglas Bakkum, Urs Frey, Hirokazu Takahashi, and Takashi Ikegami. Autonomous regulation of self and non-self by stimulation avoidance in embodied neural networks. *The 2018 Conference on Artificial Life: A Hybrid of the European Conference on Artificial Life (ECAL) and the International Conference on the Synthesis and Simulation of Living Systems (ALIFE)*, pages 163–170, 2018.

[93] J. G. White, E. Southgate, J. N. Thomson, and S. Brenner. The Structure of the Nervous System of the Nematode Caenorhabditis elegans.

Philosophical Transactions of the Royal Society B: Biological Sciences, 314(1165):1–340, nov 1986.

[94] E R Macagno. Number and distribution of neurons in leech segmental ganglia. *The Journal of Comparative Neurology*, 190(2):283–302, mar 1980.

[95] Dave Cash and Thomas J. Carew. A quantitative analysis of the development of the central nervous system in juvenile Aplysia californica. *Journal of Neurobiology*, 20(1):25–47, jan 1989.

[96] Itsuki Doi, Takashi Ikegami, Atsushi Masumori, Hiroki Kojima, Kohei Ogawa, and Hiroshi Ishiguro. A new design principle for an autonomous robot. In *Proceedings of the 14th European Conference on Artificial Life ECAL 2017*, pages 490–466, Cambridge, MA, sep 2017. MIT Press.

[97] Norihiro Maruyama, Mizuki Oka, and Takashi Ikegami. Creating spacetime affordances via an autonomous sensor network. In *IEEE Symposium on Artificial Life (ALIFE)*, pages 67–73, 2013.

[98] Douglas L. Meinecke and Alan Peters. GABA immunoreactive neurons in rat visual cortex. *The Journal of Comparative Neurology*, 261(3):388–404, jul 1987.

[99] A.R. Damasio. *The Feeling of what Happens: Body and Emotion in the Making of Consciousness*. Harvest book. Harcourt Incorporated, 1999.

[100] Daniel E. Feldman. The Spike-Timing Dependence of Plasticity. *Neuron*, 75(4):556–571, 2012.

[101] John Lisman and Nelson Spruston. Questions about STDP as a General Model of Synaptic Plasticity. *Frontiers in Synaptic Neuroscience*, 2:140, 2010.

[102] Jean-Pascal Pfister and Wulfram Gerstner. Triplets of Spikes in a Model of Spike Timing-Dependent Plasticity. *Journal of Neuroscience*, 26(38):9673–9682, 2006.

[103] T. Serrano-Gotarredona, T Masquelier, T Prodromakis, G Indiveri, and B. Linares-Barranco. STDP and sTDP variations with memristors for spiking neuromorphic learning systems. *Frontiers in Neuroscience*, 7(7 FEB):1–15, 2013.

[104] W. Cai, F. Ellinger, and R. Tetzlaff. Neuronal synapse as a memristor: Modeling pair- and triplet-based stdp rule. *IEEE Transactions on Biomedical Circuits and Systems*, 9(1):87–95, Feb 2015.

[105] Richard M Ryan and Edward L Deci. Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions. *Contemporary Educational Psychology*, 25(1):54–67, 2000.

[106] Martin Biehl, Christian Guckelsberger, Christoph Salge, Simón C. Smith, and Daniel Polani. Expanding the Active Inference Landscape: More Intrinsic Motivations in the Perception-Action Loop. *Frontiers in Neurorobotics*, 12(August):1–26, aug 2018.

[107] A. S. Klyubin, D. Polani, and C. L. Nehaniv. Empowerment: a universal agent-centric measure of control. In *2005 IEEE Congress on Evolutionary Computation*, volume 1, pages 128–135 Vol.1, Sept 2005.

[108] Laurent Orseau, Tor Lattimore, and Marcus Hutter. Universal Knowledge-Seeking Agents for Stochastic Environments. In Sanjay Jain, Rémi Munos, Frank Stephan, and Thomas Zeugmann, editors, *Algorithmic Learning Theory*, pages 158–172, Berlin, Heidelberg, 2013. Springer Berlin Heidelberg.

[109] David Ackley and Michael Littman. Interactions between learning and evolution. In C. G. Langton, C. Taylor, C. D. Farmer, and Rasmussen S., editors, *Artificial Life II, SFI Studies in the Sciences of Complexity*, volume X, pages 487–509. Addison-Wesley, Reading, MA, USA, 1992.

[110] Satinder Singh, Richard L Lewis, Andrew G Barto, and Jonathan Sorg. Intrinsically Motivated Reinforcement Learning: An Evolutionary Perspective. *IEEE Transactions on Autonomous Mental Development*, 2(2):70–82, 2010.

Acknowledgments

First of all, I would like to express deep gratitude to my supervisor Prof. Takashi Ikegami for his patient guidance and enthusiastic discussion. I would like to thank every members of Ikegami laboratory in the University of Tokyo. In particular, extensive discussion with Dr. Lana Sinapayen, Norihiro Maruyama, Itsuki Doi, and Hiroki Kojima, who are also my collaborators, have been insightful and indispensable for this thesis.

I would like to thank Assoc. Prof. Hirokazu Takahashi, Takeshi Mita and other members of Takahashi laboratory in the University of Tokyo, for their enormous support for culturing neuronal cells and valuable discussion at research meetings. I would like to thank Dr. Douglas Bakkum, Dr. Urs Frey and Prof. Andreas Hierlemann in ETH Zürich for providing tools they developed (e.g., high-density CMOS arrays) used in this thesis.

Last but not least, I would like to thank my family for their constant support.