自己動機づけ学習エージェント — 快楽と緊張により成長するネットワークを用いた 技能発達*

ジョアン リュウ[†]・アンジェイ ブラ[†]・ミハウ ヤヒムチャック[†]

Self-motivated Learning Agent

— Skill-development in a Growing Network Mediated by Pleasure and Tensions*

Juan LIU[†], Andrzej BULLER[†] and Michal JOACHIMCZAK[†]

We present a novel method of machine learning toward autonomous developmental systems. The method is based on a growing neural network that initially produces senseless signals but later associates rewarding signals and quasi-rewarding signals with recent perceptions and motor activities and, based on these associations, incorporates new cells and creates new connections, which results in more structured output patterns. The rewarding signals are produced in a device called “pleasure center”, while the quasi-rewarding signals (that represent pleasure expectation) are generated by the network itself. The network was tested using a simulated mobile robot equipped with a pair of motors, a speaker, a set of touch sensors, and a camera. Despite a lack of innate wiring for any purposeful behavior, the robot developed from scratch, without any external guidance (except hardwired perception-pleasure patterns), a set of perception-reaction patterns. The emerging patterns include obstacle avoidance, vocalization of interest, and approaching an object of interest, which are fundamental for creatures and usually handcrafted in traditional robotic systems.

1. Introduction

After decades of efforts aimed at building intelligent machines, researchers in the field of artificial intelligence (AI) have become aware that their most cherished approaches such as knowledge-based design, task-specific learning, and genetic searching, appear not powerful enough to yield machines that can demonstrate the complex, diverse, and highly integrated capabilities of a human brain [21]. Hence, the area that is recently gaining substantial attention in the AI community is developmental systems, which means a return to the very origins of AI, since in 1950 Alan Turing published in “Mind” his prophetic idea of education and self-education of a machine child.

The field of developmental systems includes developmental robotics [17], epigenetic robotics [22], and develops at the intersection of robotics, cognitive science, developmental psychology, and neuroscience. The developmental approach to AI stresses auton-

mous real-time interactions between the system and its environment, as well as an open-ended learning toward emergent behaviors resulting from coupled, context-dependent interactions [10].

Exploration of developmental mechanisms is closely tied to the study of emergence, the aim of which is to specify complicated phenomena via a small set of underlying laws. When a machine has neither innate wiring for producing purposeful signals nor predefined functional modules, emergent machine behavior is a scientific and engineering challenge. The mechanism for emergence must enable the pieces of information that are exchanged between machine and environment to be self-organized into a structure that is meaningful, and with which the structure and the meanings can develop autonomously [20].

We argue that two issues should be addressed when building developmental systems. One is a self-motivating and evaluating mechanism that enables the non-task-specific, cumulative, open-ended learning. The other is a unified structure that enables the representation and utilization of various cognitive abilities in an integrated system, other than *ad hoc* modules and learning algorithms. The psychodynamic architecture [5] and the growing network proposed in this paper are designed for this purpose and

* 原稿受付 2005年3月14日

[†] ATR ネットワーク情報学研究所 ATR Network Informatics Labs.; 2-2-2 Hikaridai, Keihanna Science City, Kyoto 619-0288, JAPAN

Key Words: developmental robotics, self-organization, neural network, emergent structure, reinforcement learning.

hopefully can be further developed to resolve the problems outlined above.

The key concepts in the psychodynamic paradigm are tension and pleasure [7]. The ability to experience pleasure is rarely an attribute of any of the contemporary robots. However, the term pleasure can be used as a label for a certain class of signals generated in an advanced controller as a self-awarding reinforcement of an unsupervised learning. In a psychodynamic system various kinds of tensions accumulate in dedicated devices without external reasons. Nonetheless their states are modified by signals that come either from perception or from other devices. An appropriately rapid discharge of one of these tensions results in the production of a *pleasure signal* [6]. We supplement this approach with the notion of a *quasi-pleasure signal* resulting from an event that is associated with a previously experienced pleasure signal. In the learning mechanism we propose, pleasure and quasi-pleasure signals induce the growth of a neural-like structure called *G-net*.

Although the idea of growth in a neural network is not new, unlike traditional solutions (see, e.g., [1,8,13,18]), in G-net adding nodes does not precede retraining of the network. On the contrary, in G-net a set of new cells pops out and gets connected as a result of accidentally gaining a pleasure or quasi-pleasure signal. The new cells and their connections modify the previous transition function in such a way that the reaction to the recent input pattern will be more probable in the future. Although this idea was inspired by the Time Sequence Processing Network [14,15,16], for G-net we worked out a novel cell transition function and rules of growth.

When comparing the subsumption architecture [3,4] with G-net, it can be noted that the latter solution does not use modules designed for defined behavioral primitives. A robot equipped with G-net learns everything from scratch and in the same single network. As for the motivation to do anything, instead of the popular idea of drives tending to maintain a homeostatic regime, such as that implemented in the M.I.T. Kismet [2: pp. 107-110], we adhere to the vision of a robot seeking pleasure for itself with the possibility that it may deliberately cause some sort of increase in a tension so as to gain pleasure later via a discharge of the tension.

We investigated G-net's performance using as a test-bed a simulated two-wheel differential-driven robot called Miao-V equipped with three proximity sensors, a camera, two drive motors, a sound-sequence generator, and a G-net-based controller (see Fig. 1). Experiments showed that G-net enables Miao-V to develop, from scratch, its ability to communicate with its caregiver, to approach and chase objects, as well as to avoid obstacles.

In Section 2 we describe the robotic controller called PreBrain, which consists of a fixed part *F-net*, a growing neural network *G-net*, and some supporting

Fig. 1 The simulated robot and its environment

devices, including the “pleasure center”. In Section 3 the construction of G-net is presented. We discuss results of experiments in Section 4, and Section 5 contains some concluding remarks.

2. Miao-V and Its PreBrain

Miao-V is a simulated mobile robot living in a simulated world: a rectangular area surrounded by a wall. For the first experiment the area was set to 640×480 pixels. The robot's diameter is set to 30 pixels and the maximum speed is 10 pixels per clock. It is equipped with a simulated camera, three proximity sensors, two motors, one speaker, and the *PreBrain* controller. As mentioned earlier, PreBrain has a fixed part called *F-net* which is responsible for simulating tensions and generating pleasure signals that drive the development of a growing part called *G-net*. In *F-net*, only the internal states of the cells can change, whereas its wiring defines the relationship between certain perceptual patterns and pleasure. In G-net, the number of inputs from or outputs to initially existing neurons can grow as new neurons are being established. Information within PreBrain is represented as signals that take the form of a series of spikes of fixed amplitude, where frequency can be modulated. Fig. 2 shows the scheme of Miao-V's control system.

A. Sensory System and Executive System

1) Sensory system

Miao-V has three proximity sensors, p_L , p_F , and p_R , that detect objects that appear at the robot's left side, front, or right side, respectively, and then produce a voltage. The voltage-to-frequency converter (VFC) processes the voltages into an appropriate signal. We define the detection of an object with one of the proximity sensors as touching, while others are detecting collision, and different styles of motion can be learned if touching is defined at the side or in the middle. The camera produces a series of video patterns that are directed to video-pattern recognizers. The current version of PreBrain processes signals from three recognizers, V_0 , V_1 , and V_2 , for red, green and yellow objects, respectively; however, the number of recognizers can be practically unlimited. Each

recognizer detects the presence of a defined object and sends a signal out through one of its five outlets depending on which of the five sectors of the visual field represents the current direction toward the object.

2) Executive system

The primary actuator of Miao-V is a sound-sequence generator (SSG) with a speaker (Sp). Based on a pair of received frequencies, SSG composes a request in the form of a sound sequence. The other actuators are two wheels propelled by dedicated motors driven by frequency-to-voltage converters (FVC). Each FVC makes the motor rotate clockwise or counterclockwise, depending on which of its two inputs receives the signal. The speed of the motor rotation depends on the frequency of the spiking.

B. F-net

1) Tension-accumulating device and Pleasure-measuring device

Tension and pleasure are key concepts in the architecture of a psychodynamic agent. The scheme of F-net defines what can change the tension level and generate a pleasure signal, and how this is done. Hence, F-net is equipped with a number of devices for processing tensions ($\theta_0, \theta_1, \theta_2, U_0, U_1, U_2, \delta_0, \delta_1, \delta_2$) as well as a *pleasure-measuring device* denoted as λ . The speed of accumulation/discharge of i th tension depends on parameters that can be set as defaults or defined by the designer. In Fig. 2, tension-accumulation devices are denoted as circles: inputs with arrows are accumulating signals, while those with solid circles are discharging signals. For a given i th tension, the parameters are:

$T_{S,i}$ – spontaneous increase/discharge time (in seconds); this number, if positive, shows for how many seconds the tension charges from 0 to 1 when no input signal (neither accumulating nor discharging signal) is provided; if the provided number is preceded by a “minus”, it denotes for how many seconds the state of the tension changes from 1 to 0 when no input signal is provided.

$T_{A,i}$ – speed of stimulated tension increase (in seconds); this number shows for how many seconds the tension charges from 0 to 1 when the influence of spontaneous charge/discharge is negligible and the tension accumulating signal $A_{i,t}$ is a spiketrain of 1 s only.

$T_{D,i}$ – speed of stimulated tension discharge (in seconds); this number shows for how many seconds the tension discharges from 1 to 0 when the influence of spontaneous charge/discharge is negligible and the tension discharging signal $D_{i,t}$ is a spiketrain of 1 s only.

In robotic practice, the influence of spontaneous charge/discharge is not negligible and the control signals $A_{i,t}$ and $D_{i,t}$ occur together. The level of i th tension, denoted $\Theta_{i,t}$, changes according to the following formula:

Fig. 2 The controller PreBrain. V_0, V_1, V_2 – video-pattern recognizers; PL, PF, PR – proximity sensors; VFC – voltage-to-frequency converter; M_L, M_R – motors; FVC – frequency-to-voltage converter; Sp – speaker; SSG – sound-sequence generator; F -net – fixed part of PreBrain; G -net – growing part of PreBrain, $\theta_0, \theta_1, \theta_2, U_0, U_1, U_2, \delta_0, \delta_1, \delta_2$ – tension-accumulating devices; D – desire modifier; λ – pleasure-measuring device; R_M, R_T, R_S, R_V – random-signal generators; S – group switch.

$$\Theta_{i,t+1} = F \left(\Theta_{i,t} + \frac{1}{T_{S,i}} + \frac{A_{i,t}}{T_{A,i}} - \frac{D_{i,t}}{T_{D,i}} \right),$$

where:

$$F(a) = \begin{cases} 0 & \text{if } a \leq 0 \\ a & \text{if } 0 < a < 1 \\ 1 & \text{otherwise} \end{cases}$$

t – time (in clocks)

$A_{i,t} = A_{i,0,t} + A_{i,1,t} + \dots + A_{i,n,t}$ – i th-tension-accumulating signal (Boolean)

$D_{i,t} = D_{i,0,t} + D_{i,1,t} + \dots + D_{i,n,t}$ – i th-tension-discharging signal (Boolean)

f – frequency of clocking (clocks per second)

According to psychodynamic theory of mind [9], pleasure is related to the dynamics of tension discharge. The greater the tension discharged in a given time and the faster the discharge of a given tension, the greater the pleasure recorded. The pleasure-measuring device calculates the pleasure record based on the following formula:

$$P_t = \begin{cases} P_{t-1} + \frac{\sum_{i=1}^N Q_i (\Theta_{i,t-1} - \Theta_{i,t})}{\sum_{i=1}^N Q_i}; \tau = 0 \\ \text{if } \exists i, (\Theta_{i,t-1} - \Theta_{i,t}) > 0, D_{i,t} > 0 \\ P_{t-1} e^{-a(\tau - T_{au})}; \tau = \tau + 1 & \text{otherwise} \end{cases}$$

where:

$P_t \in R$ – Pleasure level [0,1]

N – Number of tensions

$Q_i \in R$ – i th-tension-discharging importance [0,1]

τ – the elapsed time since the latest tension-discharging event

a, T_{au} – parameters (default values are 0.001 and 0, respectively)

2) F-net structure

In PreBrain, F-net contains the inborn part of the controller that defines the robot's physical needs and how tensions and pleasures are modulated. F-net is designed to drive the robot to act, to perceive, and to explore. In the present experiment, we assume the robot may have the desire to see or touch an object, and prefers to avoid collisions. What we are concerned with is how different skills could be consecutively learned in one computational model, i.e., how the system could go the other way, moving outward to make new kinds of goals [19:p.177]. Suppose the robot's goal is to reach a toy. This could lead to the sub-goal of learning how to move efficiently, which could lead to sub-subgoals of learning to avoid obstacles. The goal could also arouse the development of social skills to engage people's help in bringing it the toy. The process of goal shifting is a synergetic result of the robot's internal desires, past experiences and current status.

The "heart" of F-net is a collection of random-signal generators, R_V , R_T , R_S , and R_M , that represents a source of internal tensions. R_V and R_M send signals to G-net, which results in the production of random meaningless sounds or movements, respectively. R_S stimulates the accumulation of *primary tensions* in the tension-accumulating devices θ_0 , θ_1 ,

and θ_2 . θ_k discharges when a substantial number of spikes is sent within a defined time interval by any of the input neurons connected to V_k . Hence, tension in θ_k represents a desire for the k -th object that can be fulfilled simply by seeing the object. The signal also, through an OR-gate, stimulates R_V . The dynamics of the tension accumulated in θ_k is analyzed by λ_0 , a pleasure-generating device that is sensitive to a rapid collapse of incoming frequency.

R_T stimulates the accumulation of *secondary tensions* in the devices δ_0 , δ_1 , and δ_2 . Tension in δ_k represents a desire for the k -th object that can be fulfilled only through seeing the object in the straight-ahead direction and touching it at the same time.

The desire modifier D slowly accumulates tension when it is stimulated by λ_0 . If the tension in D is high, the device stimulates R_T and suppresses R_S . Owing to this, Miao-V initially asks its caregiver for a desired object with sound sequences and is satisfied when it sees the desired object (shown by the caregiver), but, as the tension in D increases, the robot starts desiring not only to see, but also to touch the object. This change in motivation forces Miao-V to start trying to get the desired object by itself, so it learns, unsupervised, how to approach/chase the object. The learning takes place in G-net and is driven by the signals that passed through the switch S and the signals produced by pleasure-generating device λ_1 . S selects video signals for sending to G-net's u_{12} , u_{13} , ..., u_{16} according to current tension δ_k .

Any high tension remaining in any δ_k for a substantial time (dozens of seconds) results in the accumulation of tension in the displeasure-accumulating device U_k , thus stimulating the related θ_k , which results in tension accumulation in θ_k . This mechanism works when Miao-V fails to get a desired object by using its motors and causes the robot resume to expressing its desire via sound sequences.

The robot's collision with an obstacle makes p_L or p_R stimulate one of the related input cells, which results in the production of a signal that is then directed to G-net's input u_{17} or u_{18} . Every collapse of both the u_{17} and the u_{18} signal is detected by λ_2 and a pleasure signal is directed to G-net's input u_2 . Based on signals u_{17} , u_{18} , and u_2 , G-net learns, unsupervised, how to avoid obstacles, a process that does not interfere with the simultaneous learning of approaching and chasing an object.

3. G-net: Growing network

G-net initially consists of a set of *peripheral cells* whose contents depend on the sensory-motor settings of the robot. Reinforced by pleasure signals that appear accidentally during the robot's encounters with its environment (including interactions with its caregiver), G-net grows, which means that some new *operational cells* pop out and some new connections are established. As time passes, the robot becomes an increasingly experienced and smarter "pleasure-seeker."

The following subsections provide a specific nomenclature for G-net and, based on the nomenclature, the description of mechanisms of G-net's growth.

A. Nomenclature

$C_i = \{j : \exists \alpha_{ji} > 0\}$ – set of indexes of cells receiving signal from i -th cell

E – set of indexes of executive cells

H – set of indexes of proximity-detecting cells

h, i, j, j', k, p – variables for cell identities

N_I – number of elements of set I

P – set of indexes of pleasure cells

Q – set of indexes of tension cells

R^S – set of indexes of random-sound cells

R^M – set of indexes of random-motion cells

$S_j = \{i : \exists \alpha_{ji} > 0\}$ – set of indexes of cells providing signals to j -th cell

t – time (in clocks)

u_t^j – external input signal to j -th cell at the clock t

V – set of indexes of video cells

y_t^i – state of i -th cell at the clock t

α_{ji} – strength of connection from i -th cell to j -th cell

β^j – decay coefficient for the state of j -th cell

$A(a) = \frac{\sum_{j \in C_a} \alpha_{1j}}{N_{C_a}}$ – average strength of connections to pleasure cell 1 from cells that receive signal from cell a

$\Omega = P \cup E \cup H \cup R^S \cup R^M \cup V \cup Q \cup \bigcup_{i \in V \cup H \cup Q} C_i \cup \bigcup_{j \in E} S_j$ – the whole set of cells in G-net

B. Neurogenesis

In the current system, the set of *peripheral cells* includes twenty-five cells: three “pleasure” cells, five video cells, two proximity-detecting cells, three tension cells, six executive cells, two random-sound cells, and four random-motion cells (Fig. 3). Each pleasure cell receives a signal from a dedicated pleasure-measuring device (λ_j) located in F-net, and each of the video cells receives a signal from a currently connected video-pattern recognizer (V_k). Each of the proximity-detection cells receives an appropriate signal from VFC, while each tension cell receives a signal from a related θ_k . Two of the executive cells send signals to the sound-sequence generator (SSG), while other executive cells send signals to dedicated inputs of the robot's two motors. Random-sound cells and random-motion cells send signals to dedicated executive cells. The strengths of the connections between random-sound/motion cells and executive cells are all equal to 1. A state decay β^j coefficient is set as 0.9 only for the random-motion cells; for all other cells the state-decay coefficient is equal to 0. Reinforced by pleasure signals appearing accidentally during the robot's more or less random motion, the network grows, which means that some new *operational cells* pop out and some new connections are established (Fig. 3). As time passes, the robot becomes more experienced and smarter.

Fig. 3 Initial configuration of *G-net*. The innate wiring allows the robot to make only random movements.

The initial configuration of *G-net* and related parameters can be formally described as follows:

$$\begin{aligned} P &= \{0, 1, 2\}, R^M = \{3, 4, 5, 6\}, R^S = \{7, 8\}, \\ Q &= \{9, 10, 11\}, V = \{12, 13, 14, 15, 16\}, \\ H &= \{17, 18\}, E = \{19, 20, 21, 22, 23, 24\}, \\ \alpha_{21,3} &= \alpha_{22,4} = \alpha_{23,5} = \alpha_{24,6} = \alpha_{19,7} = \alpha_{20,8} = 1 \\ \beta^3 &= \beta^4 = \beta^5 = \beta^6 = 0.9, \\ \text{for all } j \notin R^M, \beta^j &= 0; \\ \text{for all } j \notin R^M \cup R^S, C_j &= \{\}; \\ \text{for all } k \in E, N_{S_k} &= 1. \end{aligned}$$

The cell-update procedure is based on this formula:

$$y_t^j \leftarrow \begin{cases} u_t^j & \text{if } j \in P \\ F \left(\beta^j y_{t-1}^j + u_t^j + \frac{\sum_{i \in S_j} \alpha_{ji} y_t^i}{N_{S_j}} \right) & \text{otherwise} \end{cases}$$

The network-growth procedure is:

Step 1 Update the state y_t^i of cells $\{0, 1, \dots, 18\}$

Step 2 Apply Rules #1 and #2

Step 3 Update y_t^i of operational cells in ascending index order and then update y_t^i of executive cells

Step 4 Go to Step 1

where:

Rule #1:

For all $p \in P$, if $y_{t-1}^p = 0 \wedge y_t^p > 0$
 then for all $i \in Q \cup V \cup H$ such that $y_t^i > 0$
 $j \leftarrow N_\Omega$, $C_i \leftarrow C_i \cup \{j\}$,
 $\alpha_{ji} \leftarrow 1$, $\alpha_{pj} \leftarrow 1$,
 for all $k \in E$ such that $y_{t-1}^k > 0$
 $j' \leftarrow N_\Omega$, $S_k \leftarrow S_k \cup \{j'\}$,
 $\alpha_{j'j} \leftarrow y_{t-1}^k$, $\alpha_{kj'} \leftarrow 1$.

Rule #2:

If $\exists i \in V y_{t-1}^i = 0 \wedge y_t^i = 1$
 then for all $h \in V$ such that
 $h \neq i \wedge y_{t-1}^h = 1 \wedge A(i) > A(h) \geq 0$
 $j \leftarrow N_\Omega$, $C_h \leftarrow C_h \cup \{j\}$,

$$\begin{aligned}
&\alpha_{jh} \leftarrow 1, \alpha_{1j} \leftarrow 0.5\Lambda(i), \\
&\text{for all } k \in E \text{ such that } y_{t-1}^k > 0 \\
&\quad j' \leftarrow N_\Omega, S_k \leftarrow S_k \cup \{j'\}, \\
&\quad \alpha_{j'j} \leftarrow y_{t-1}^k, \alpha_{kj'} \leftarrow 1.
\end{aligned}$$

According to Rule #1, G-net reacts to pleasure signals by establishing new cells and interconnections in a manner dependent upon the current and previous states of related peripheral cells. Newly created cells will assist in recalling the current beneficial sensory-actuator pattern in the future. According to Rule #2, G-net reacts analogously to quasi-pleasure signals that come from certain video cells. A signal is treated as a quasi-pleasure one when it can cause an event potentially contributing to pleasure. The quasi-pleasure signals can be interpreted as the robot's expectation of pleasure given certain peripheral cells' states. After being triggered by Rule #1, the growth of G-net proceeds in a chain-reaction manner based on Rule #2. With every new set of produced cells, the robot acquires a new reaction useful in attaining pleasure.

The essence of the growing rules is to detect the beneficial changes of states and remember the actions that are carried out immediately before the changes as successful experiences. The same rule is applied to learn different behaviors and record them in one single network. In Fig. 4, it can be seen that operation cells are self-organized into clusters connected with corresponding peripheral cells.

Fig. 4 A resulting network capable of producing distinguishable sound sequences for three objects, avoiding walls, and approaching objects. The size of the cells shows their state: the more activated cells are bigger and brighter.

4. Experiments

A “newborn” Miao-V, like a newborn human baby, has no idea of how to purposefully use its actuators, so in the face of increasing tensions it can only produce random sounds and movements. Hearing sequences of sounds produced by Miao-V, its caregiver gives it the items she supposes it wants to have at the moment. If by accident the given item causes a discharge

of the tension accumulated in θ_k , the subsequently generated pleasure signal u_0 causes the creation of a cell taking a currently activated tension cell as its presynaptic cell, as well as the creation of two executive cells, which increases the strength of association between the tension and the recent vocal expression. This way, G-net gradually grows, which results in the robot's increasing ability to differentiate random vocal expressions into distinguishable sequences of sounds, each dedicated to a different desired object. At the same time the caregiver gradually learns to properly guess the robot's needs based on the sequences of sounds. However both the caregiver and the robot have no prior knowledge about what form the sound sequences should take. Their interaction gradually forms the signal. In other words, within the pair—Miao-V and its caregiver—a common, mutually understandable communication emerges.

As the state of desire modifier D becomes higher, Miao-V becomes more interested in touching the current desired object. G-net becomes motivated to try to use motors so that Miao-V will approach objects itself, while stopping its vocal requests. Initially, the motors obtain random signals from the 3rd, 4th, 5th, and 6th cells, which results in the robot producing meaningless movements. If we define the detecting of objects by the middle proximity sensor as touching, then when by accident (or owing to the caregiver's help) Miao-V touches a desired object that is already present in the central sector of the robot's visual field (signal u_{14}), the related tension δ_k rapidly discharges and λ_1 produces a pleasure signal u_1 . This event results in the creation of a handful of new cells, which memorize the fact that seeing the object of interest in the central sector of the visual field has a certain relationship with a potential pleasure. Therefore, a high frequency of u_{14} is also a desirable signal (quasi-pleasure). The process chain-reacts so that more and more quasi-rewarding signals are generated. Perceptions (which are visual sectors here) and actions are memorized and organized according to their potential contributions to the pleasure. Hence, the robot gradually gains knowledge about the relationship between the current contents of its visual sectors and beneficial motor actions.

If the touch sensor is defined to be the left proximity sensor, the system is able to adapt the change online without effecting changes to the rules. The resulting behavior is strategically different from the former chasing behavior in that the robot either directly approaches the object by its left side, or turns to the right when getting close enough to the object so that the left sensor can reach it. These two apparent different behaviors are actually generated by the same network. The interactions between cells as well as those between the robot and the environment produce the difference.

As seen in Fig. 5, the time interval between the appearance of the desired object in the visual field

and touching it systematically decreased. It can be observed that Miao-V did not obtain catching skills with the first couple of encounters with the object; indeed, the distance between the robot and the ball did not decrease dramatically before the object's disappearance from the visual field. However, after the first accidental success, the robot quickly acquired the knack of chasing the object.

While learning to approach/chase the object, the robot detects the wall with its left or right proximity sensor. If, by accident, the robot moves in such a way that neither sensor detects the wall, a related pleasure signal λ_2 becomes associated with the recent activities of the motor cells, resulting in the appearance of a handful of new cells in G-net. If this situation occurs repeatedly, G-net gradually learns, unsupervised, to make the robot efficiently avoid the wall, a process that does not interfere with simultaneously learning other skills.

Fig. 5 Record of the distance between the robot and an autonomously moving object of interest. The dark spots on the curve indicate the periods when the object is in the robot's visual field. The plot converges to the horizontal axis, which means that G-net has learnt to drive chasing the object of interest.

5. Concluding Remarks

It may seem that PreBrain is an excessively complicated controller system with respect to the relatively simple behaviors the presented robot has developed. However, our intention was not to develop a new method of chasing objects or language acquisition. The described behaviors were only examples illustrating the ability of a single G-net to develop different skills simultaneously in real time without a direct definition of target behaviors (i.e. based only on the knowledge about the sources of tension charges and discharges). The F-net, G-net, and experiments described in this paper refer to the early stage of a cumulative developmental agent. G-net seems to be able to scale up to dozen skills supported by hundreds of sensors/actuators. We believe that further research will confirm this supposition.

F-net's wiring defines how tensions and pleasures are modulated. It can be noted that the switches between the robot's sources of tensions induced during G-net's development by the desire modifier resemble the mechanism of switching gene activities inside a developing cell. This process appears crucial for exploring novel perceptual-motor mappings. In a more complex configuration of the agent and its environment, such coupled interactions provide a foundation for emergent behaviors.

Miao-V molds its own view of values according to rewards that come from its own "pleasure center" and then induces G-net to incorporate new cells and create new connections, which results in a gradual improvement of its performance. Even in this simple simulated environment, we are not able to predict the structure of G-net and the next action of the robot. Nonetheless, we may know that the robot will be able to learn behaviors that are helpful for gaining pleasure. G-net is composed of a number of neurons stemming from peripheral cells with specified properties. G-net's structure emerges driven by the interactions between the robot and the environment, as well as by the interactions between its cells. The key problem in the study of G-net is to find a set of essential growth laws that would result in efficient learning in even more complicated domains.

The proposed system follows the reinforcement learning paradigm, where F-net can be interpreted as the reward function for G-net. The connections between pleasure cells and other cells work as a value function, and the ensemble of cells in G-net is the policy for it, which is being continuously updated. The learning power of G-net compared with other reinforcement learning methods and neural network is still under investigation.

Although the mechanism underlying G-net's growth is believed to be very scalable, further research should bring accurate quantitative figures about the scalability of the presented method. There are also several possibilities to improve the paradigm: tuning the connection-updating mechanism, adding a mechanism to remove or reuse cells, integrating learned skills, acquiring new kinds of pleasures, differentiating cells, and/or evolving parameters of the mechanism that manages G-net's growth.

Acknowledgments

This research was supported in part by the National Institute of Information and Communications Technology (NiCT).

References

- [1] T. Ash: Dynamic node creation in backpropagation networks; *Connection Science*, Vol. 1, pp. 365–375 (1989) (cited in [8])
- [2] C.L. Breazeal: *Designing Sociable Robots*, A Bradford Book/The MIT Press (2002)

- [3] R. Brooks: A robust layered control system for a mobile robot; *Cambrian Intelligence: The Early History of the New AI*, A Bradford Books/The MIT Press, pp. 3–26 (1999)
- [4] R. Brooks: *Flesh and Machines: How Robots Will Change Us*, Pantheon Books (2002)
- [5] A. Buller: Psychodynamic robot; *Proceedings, 2002 FIRA Robot World Congress*, pp. 26–30 (2002)
- [6] A. Buller: From q-cell to artificial brain; *Artificial Life and Robotics*, Vol. 8, No. 1, pp. 89–94 (2004)
- [7] A. Buller, M. Joachimczak, J. Liu and K. Shimohara: ATR artificial brain project: 2004 progress report; *Artificial Life and Robotics*, Vol. 9, No. 4, pp. 197–201 (2005)
- [8] S.E. Fahlman and C. Lebiere: The cascade-correlation learning architecture; *Advances in Neural Information Processing Systems* (D.S. Touretzky (Ed.)), Vol. 2, Morgan Kaufmann, pp. 524–532 (1990) (cited in [10])
- [9] S. Freud: *An outline of psycho-analysis*, W.W. Norton, p. 15 (1940/1989)
- [10] J.H. Holland: *Emergence: from chaos to order*, Addison-Wesley (1998)
- [11] N. Jankowski: *Ontogeniczne sieci neuronowe – o sieciach zmieniających swoją strukturę*, EXIT (2003)
- [12] S. Kitamura, Y. Kakkuda and H. Tamaki: An approach to the emergent design theory and applications; *Artificial Life and Robotics*, Vol. 3, pp. 86–89 (1999)
- [13] C.P. Lam and D. Stork: Learning network topology; *The Handbook of Brain Theory and Neural Networks* (M.A. Arbib (Ed.)), 2nd edition, A Bradford Book/The MIT Press, pp. 628–631 (2003)
- [14] J. Liu: Toward machine intuition: A way-finding without maps or coordinates; *Proceedings, The 8th International Symposium on Artificial Life and Robotics*, pp. 520–531 (2003)
- [15] J. Liu, Z. Cai and X. Zou: A connectionist model for localization and route learning based on remembrance of perception and action; *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'02)*, pp. 655–660 (2002)
- [16] J. Liu, Z. Cai and C. Tu: Environmental feature extraction and mergence: Make the past serve the present; *The 9th International Conference on Neural Information Processing (ICONIP'02)* (2002)
- [17] M. Lungarella and G. Metta: Beyond gazing, pointing, and reaching – A survey of Developmental Robotics; *Proc. of the 3rd Int. Workshop on Epigenetic Robotics (EPIROB'03)*, pp. 81–89 (2003)
- [18] M. Mezard and J.P. Nadal: Learning in feedforward layered networks: the tiling algorithm; *Journal of physics*, Vol. 22, pp. 2191–2203 (1989) (cited in [10])
- [19] M. Minsky: *The society of mind*, Simon & Schuster (1986)
- [20] K. Shimohara: Human communication project: Towards full realization of the richness of natural human communication; *ATR UptoDate*, Spring, pp. 9–11 (2002)
- [21] J. Weng, J. McClelland, A. Pentland, O. Sporns, I. Stockman, M. Sur and E. Thelen: Autonomous mental development by robots and animals; *Science*, Vol. 291, No. 5504, pp. 599–600 (2000)
- [22] J. Zlatev and C. Balkenius: Introduction: Why ‘epigenetic robotics’?; *Proceedings of the First International Workshop on Epigenetic Robotics: Modeling Cognitive Development in Robotic Systems*, Lund University Cognitive Studies, Vol. 85, pp. 1–4 (2001)