Interpreting Two Psychological Functions by A Hierarchically Structured Neural Memory Model

Tetsuya Hoya[1]

Laboratory for Advanced Brain Signal Processing BSI-RIKEN, 2-1, Hirosawa, Wakoh-City, Saitama 351-0198 Japan

Abstract. In this paper, a novel neural memory model is proposed. The neural memory model proposed in this paper is based upon generalized regression neural networks (GRNNs) which are the paradigms of radial basis function neural networks (RBF-NNs). With the benefit of their quick learning capability and robustness, the application of GRNNs has been rapidly increased in many disciplines. Then, within the context of a newly proposed hierarchically arranged generalized regression neural network (HA-GRNN), two psychological functions, intuition and attention, are interpreted in terms of the evolution of the HA-GRNN. Within the framework of HA-GRNN, two types of memory, namely both the long and short term memory motivated from biological and cognitive studies, are considered and a dynamic learning system is thus proposed. In the simulation study, the effectiveness of the HA-GRNN in comparison with k-means clustering method is confirmed within the context of pattern classification tasks.

Keywords: Artificial neural networks, memory model, psychological functions, dynamic learning.

1 Introduction

Autonomous robotics can be ultimately defined as such robots that can 'think' and determine the next behavior by themselves without the instruction given by humans. Now, with the recent advancements in both biological and cognitive studies as well as computer technologies, one of which we wish to achieve in near future is develop, what is called, 'brain-style' computers (e.g., [1]), or truly autonomous robotics. It is said that one of the key approaches towards the development of brain-style computing is how to elucidate the mechanism of "intuition" in terms of artificial neural networks. On the other, modeling the notion of "consciousness" has recently been a topic of great interest in robotics [2,3]. Interpreting the notions related to emotional/psychological functions, however, has historically been a controversy among many disciplines from biology to philosophy. In this paper, it is addressed that such psychological functions, "intuition" and "attention", can be interpreted in terms of the evolution of an hierarchically arranged generalized regression neural network (HA-GRNN) model in which each sub-network has memory-based architecture. The evolution process is then justified within the framework of pattern classification tasks. The generalized regression neural networks (GRNNs) [4] fall in the category of radial basis function neural networks (RBF-NNs) [5], while, unlike ordinary RBF-NNs, having a special property that the weight vectors between the RBFs and output neurons are given identical to the target vectors. By exploiting this attractive property, a dynamic neural system can be modeled without any complex mathematical operations.

2 Configuration of a GRNN

A multilayered GRNNs (ML-GRNN) [5] with N_i input neurons, N_h radial basis functions (RBFs), and N_o output neurons is illustrated on the top of Fig. 1. In Fig. 1, each input neuron x_i $(i = 1, 2, \cdots, N_i)$ corresponds to the element in the input vector $\boldsymbol{x} = [x_1, x_2, \cdots, x_{N_i}]^T$ (T: vector transpose), h_j $(j = 1, 2, \cdots, N_h)$ is the j-th RBF (note that N_h is variable), $\| \cdots \|_2^2$ denotes the squared L_2 norm, and the output neuron o_k $(k = 1, 2, \cdots, N_o)$ is given as

$$o_k = \frac{1}{\delta} \sum_{j=1}^{N_h} w_{j,k} h_j, \tag{1}$$

where $\delta = \sum_{k=1}^{N_o} \sum_{j=1}^{N_h} w_{j,k} h_j$, $\boldsymbol{w}_j = [w_{j,1}, w_{j,2}, ..., w_{j,N_o}]^T$, and

$$h_j = f(\boldsymbol{x}, \boldsymbol{c}_j, \sigma_j) = exp(-\frac{\|\boldsymbol{x} - \boldsymbol{c}_j\|_2^2}{2\sigma_j^2}), \tag{2}$$

where $\boldsymbol{c}_j$ is called the centroid vector, σ_j is the radius, and $\boldsymbol{w}_j$ denotes the weight vector between the j-th RBF and the output neurons. As in Fig. 1 on the top, the structure of an ML-GRNN is similar to the well-known multilayered perceptron perceptron neural network (MLP-NN) except RBFs are used in the hidden layer and linear functions in the output layer. In Fig. 1, if the target vector $\boldsymbol{t}(\boldsymbol{x})$ corresponding to the input pattern vector $\boldsymbol{x}$ is given as

$$\begin{aligned} \boldsymbol{t}(\boldsymbol{x}) &= (\delta_1, \delta_2, ..., \delta_{N_o}), \\ \delta_j &= \begin{cases} 1 \text{ if } \boldsymbol{x} \text{ belongs to the class} \\ \quad \text{corresponding to } o_k \\ 0 \text{ otherwise} \end{cases} \end{aligned} \tag{3}$$

and if the centroid h_j is assigned for $\boldsymbol{x}$, $\boldsymbol{w}_j = \boldsymbol{t}(\boldsymbol{x})$, then the entire network becomes topologically equivalent to the one with a decision unit and N_o number of sub-nets as in the bottom of the figure [6]. In summary, the network configuration by means of an ML-GRNN is simply done in the following:

Network Growing: Set $\boldsymbol{c}_j = \boldsymbol{x}$ and fix σ_j, then add the term $w_{jk} h_j$ in (2). The target vector $\boldsymbol{t}(\boldsymbol{x})$ is used as a class 'label' indicating the sub-network number to which the RBF belongs.

Network Shrinking: Delete the term, $w_{jk} h_j$, from (2).

In the above, it is considered that, in hardware implementation, the network growing (learning) can be done very easily and the generalization performance is robust, while conventional neural networks such as multilayered perceptron neural networks (MLP-NNs) with the back-propagation algorithm [7] require iterative training scheme whenever the network configuration is changed and there is always a danger of being stuck in local minima [8].

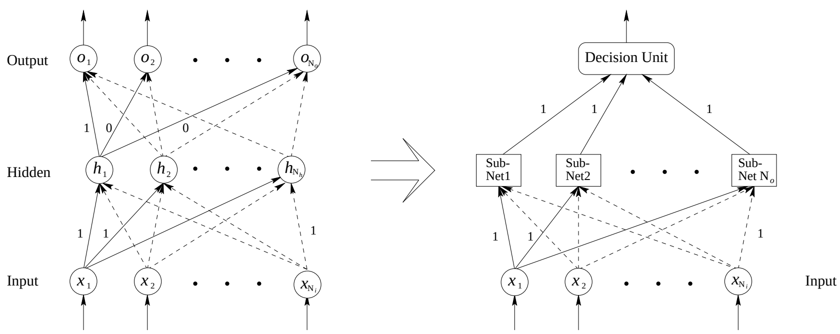

Fig. 1. Illustration of topological equivalence between the ML-GRNN with M hidden and N output units and the assembly of the N distinct sub-networks.

3 An Hierarchically Arranged Generalized Regression Neural Networks

The structure of an hierarchically arranged GRNN (HA-GRNN) is illustrated in Fig. 2. In the figure, a multiple of GRNNs representing long-term memory (LTM) networks (LTM Net (1 to L) in Fig. 2), a modified RBF network representing short-term memory (STM), and a decision unit are used. Moreover, the LTM nets can be subdivided into two parts; one for 'intuitive outputs' (denoted by Region 1 in a circle) and the others (denoted by Region 2). In the second part, each LTM Net (2 to L) has the same structure as in the bottom of Fig. 1, whereas both the STM and LTM Net 1 are given as modified RBF-NNs.

3.1 Structure of the STM Network

The output of the STM network $\boldsymbol{O}_{STM}$ is given in a vector form rather than a scalar value calculated as the sum. of the RBF outputs. The STM network, unlike the LTM nets described later, does not have any sub-nets, namely it is based upon a two-layered structure, with a maximum number of centroids M_{STM}, and functions as a temporal buffer to the LTM nets. The STM has, therefore, a structure similar to a queuing system. The learning of the STM network is summarized as follows:

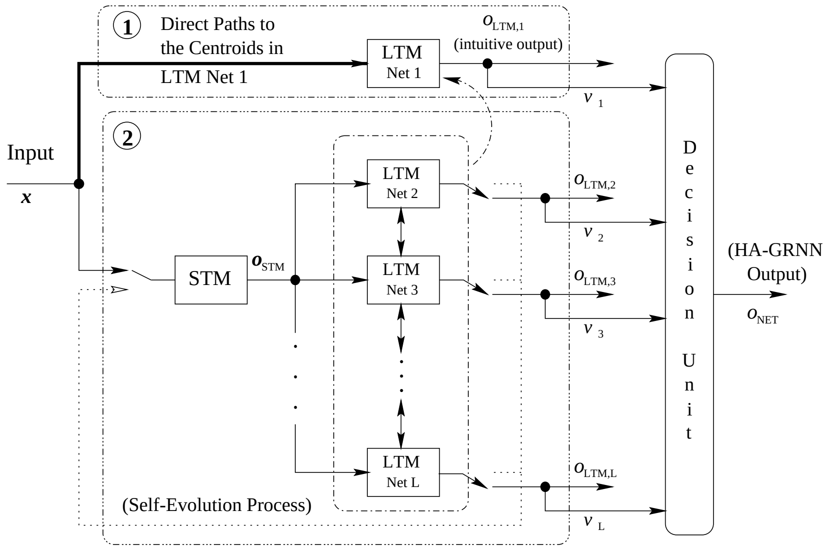

Fig. 2. Schematic representation of an hierarchically arranged GRNN

Step 1: If the number of the centroids is less than M_{STM}, add an RBF with h_i (given in (2)) and $\boldsymbol{c}_i = \boldsymbol{x}$ in the STM. Then, set $\boldsymbol{O}_{STM} = \boldsymbol{x}$.

Step 2: Otherwise,

1) If the activation of the least activated centroid (h_j, say) $h_j < th_{STM}$, replace it with a new one with $\boldsymbol{c}_j = \boldsymbol{x}$ and set $\boldsymbol{O}_{STM} = \boldsymbol{x}$.

2) Otherwise,

$$\boldsymbol{O}_{STM} = \lambda \boldsymbol{c}_k + (1 - \lambda)\boldsymbol{x} \tag{4}$$

where $\boldsymbol{c}_k$ is the centroid vector of the most activated centroid (k-th, say) h_k and λ is a *smoothing* factor ($0 \leq \lambda \leq 1$).

3.2 Structure of the LTM Networks

Similar to the STM, each LTM net in Fig. 2 has a maximum number of the centroids M_{LTM_i} $i = 1, 2, \cdots, L$. The LTM nets, except LTM Net 1, are, in turn, composed of GRNNs rather than RBF-NNs. Therefore, each LTM net is viewed as a collection of the sub-nets plus a decision unit as in the right side of Fig. 1 (except LTM Net 1). In contrast, LTM Net 1 consists of the centroids without a summing operation unit in the output. The output of LTM Net 1, $O_{LTM,1}$ is thus identical to the activation of the most activated centroid (l-th, say) h_l itself chosen by the 'winner-takes-all' strategy.

3.3 Evolution of the HA-GRNN

In the HA-GRNN, the role of the STM is to 'buffer' the incoming input pattern vectors, before storing them to the LTM nets. It is then hypothesized that long-term memory, in itself, has a layered structure representing an hierarchical classification system which is based on the 'significance' or 'attractiveness' of information. In this paper, such a classification system is modeled based on the activation of centroids. In summary, the construction of the HA-GRNN is divided into four phases:

Phase 1: STM (and LTM Net 2) formation ($t = 0$).
Phase 2: Formation of the LTM networks, LTM Net (2 to L).
Phase 3: Reconfiguration of the LTM Net (2 to L) (self-evolution) ($t = t_1$).
Phase 4: Formation of LTM Net 1 ($t = t_2$).

In Phase 1, the STM is formulated in the manner as described in Section 3.1, while LTM Net 2 is also formed by directly assigning the output vectors of the STM to the centroids in LTM Net 2. In the above, t denotes the t-th pattern presentation. The addition of the centroids in Sub-Net i ($i = 1, 2, ..., N_{cl}$, where N_{cl} is the number of classes) of LTM Net 2 is repeated until the total number of centroids in Sub-Net i reaches a maximum $M_{LTM_2,i}$. Otherwise, the least activated centroid in Sub-Net i is moved to LTM Net 3. This process corresponds to Phase 2. Then, Phase 2 is summarized as follows:

Step 1: Provided that $\boldsymbol{O}_{STM}$ belongs to Class i, then, for $j = 1$ to $L-1$, do the following:
If the num. of the centroids in Sub-Net i of LTM Net j reaches $M_{LTM_{j,i}}$, move the least activated centroid within Sub-Net i of LTM Net j to LTM Net $j+1$.

Step 2: If the num. of the centroids in Sub-Net i of LTM Net L reaches $M_{LTM_{L,i}}$ (i.e., all the i-th sub-nets within LTM Net from 2 to L are filled), there is no entry to store the newly coming vector $\boldsymbol{O}_{STM}$. Therefore, do the following:

Step 2.1: Discard the least activated centroid in Sub-Net i of LTM Net L.
Step 2.2: Shift all the least activated centroids in Sub-Net i of LTM Net from (L-1) down to 2 into LTM Net from L down to 3, respectively.
Step 2.3: Then, the new STM output vector is stored in Sub-Net i of LTM Net 2.

In Fig. 2, the output of the HA-GRNN O_{NET} is chosen as the largest value among the weighted LTM net outputs $O_{LTM,i}$ ($i = 1, 2, \cdots, L$):

$$O_{NET} = \max(v_1 \cdot O_{LTM,1},\ v_2 \cdot O_{LTM,2},\ \cdots, v_L \cdot O_{LTM,L}), \tag{5}$$

where $v_1 >> v_2 > v_3 > \cdots > v_L$. Note that the weight value v_1 for $O_{LTM,1}$ is given relatively larger than the others. This discrimination indicates the formation of the 'intuitive output' from the HA-GRNN. After the formation of the LTM nets, reconfiguration of the LTM nets is considered in Phase 3 in order to 'shape up' the pattern space spanned by the centroids in the LTM Net (2 to L). This process may be invoked at particular time. During the reconfiguration phase, presentation of any incoming input pattern vector is not allowed to process. In Phase 4, some of the centroids which keep relatively strong activation in a certain period in LTM Net (2 to L) are moved to LTM

Net 1. Each centroid newly assigned in LTM Net 1 eventually forms an RBF-NN and has a direct connection from the input vector $\boldsymbol{x}$.

4 Interpretation of Intuition and Attention

4.1 A Model of Intuition by HA-GRNN

In our daily life, we sometimes encounter such an occasion of which we feel the thing/matter is true but neither can we explain the reason why nor find the evidence or proof of it. This is referred to as the notion of, what is called, "intuition".

> Conjecture 1: In the HA-GRNN context, *intuition* can be interpreted such that, for a particular incoming input pattern vector there exists a certain set of centroids with *abnormally* strong activation within the LTM nets.

The above is drawn from the standpoint that the notion of intuition can be explained in terms of the information processing pertaining to a particular activity of neurons within brain (e.g., see [9]).
The evidence for referring to the output of LTM Net 1 as intuitive output is that LTM Net 1 is formed after a relatively long and iterative exposition of incoming input pattern vectors which results in strong excitation of some centroids in LTM Net (2 to L). In other words, the transition of the centroids from the STM to LTM Net (2 to L) is referred to as *normal* learning process, whereas that from LTM Net (2 to L) to LTM Net 1 gives the chances of generating "intuitive" HA-GRNN outputs.

4.2 Interpreting the Functionality of Attention by HA-GRNN

In the context of HA-GRNN, the model in [11] coincides with the evidence of having a 'hierarchical' structure for representing attention as a function of consciousness. In the HA-GRNN context, the following conjecture can be therefore drawn:

> Conjecture 2: The state of being 'attentive' to something is represented in terms of the centroids within the STM.

Accordingly, the following Phase 5 (at $t = t_3$) is appended to the evolution of an HA-GRNN:

[Phase 5: Formation of Attentive States]

Step 1: Collect m centroids of which number of activation count is the largest within all the LTM nets (2 to L) for particular classes.

Step 2: Add the copies of the m centroids back into the STM, where $M_{STM} - m$ most activated STM centroids are kept untouched. The m centroids so selected remain within the STM for a certain long period, without changing their centroid vectors but the radii.

It is also postulated that the ratio between the m centroids and the rest of the $M_{STM} - m$ in the STM explains the 'level of attention'. Therefore, the following conjecture can also be drawn;

> Conjecture 3: The level of attention can be determined by the ratio between the number of the m most activated centroids selected from the LTM nets and that of the remaining $M_{STM} - m$ in the STM.

Conjecture 3 is also related to the neurophysiological evidence of 'rehearsing' activity [10] in which the information acquired during learning would be gradually stored as a long-term memory after rehearsing. In the HA-GRNN context, an incoming input pattern vector $\boldsymbol{x}$ can be compared to the input information to the brain and are temporally stored within the STM. Then, during the evolution, the information represented by the STM centroids is selectively transferred to the LTM nets in Phases 1-3. In contrast, the centroids within the LTM nets may be transferred back to the STM, because the states being 'attentive' to certain classes is occurred at particular moments. In pattern classification tasks, one may limit the number of the classes to $N < N_{max}$ for representing attention in a way that "The HA-GRNN is particularly attentive to the N classes among a total of N_{max}" in order to compensate for the relatively 'weaker' area of the pattern space.

5 Simulation Study

In the simulation, an HA-GRNN is constructed using the data extracted from SFS database [12]. The data set used consists of a total of 900 utterances of the digits from /ZERO/ to /NINE/ recorded in English by nine different speakers (including even numbers of female and male speakers). The data set was then arbitrarily partitioned into two sets; one for constructing an HA-GRNN (i.e., the incoming pattern set) and the other for testing. The incoming pattern set contains a total of 540 speech samples, where 54 samples were chosen for each digit, while the testing consists of a total of 360 samples (36 samples per each digit). (The evolution within Phase 1 to 4 was therefore eventually stopped at $t = 540$.) Each utterance is sampled at 20kHz and was converted into the input vector of the HA-GRNN with a normalized set of 256 data points obtained by the well-known LPC-mel-cepstral analysis.

5.1 Parameter Setting

In the simulation study, the LTM parameters, $M_{LTM_1} = 5$, and $M_{LTM_2} = M_{LTM,3} = 40$, were used. For the STM, the choices, $M_{STM} = 30$ and $\lambda = 0.6$, were made to sparsely but reasonably cover all the ten classes during the construction. The number of sub-nets in LTM nets was equally fixed to 10 (i.e., for the ten digits). With this setting, the total number of centroids in LTM Net (1 to 3) $M_{LTM,Total}$ yields 85. Then, to give 'intuitive outputs' from LTM Net 1, v_1 was fixed to 2.0, while v_i ($i = 2, 3, \cdots, L$) were given by a linear decay $v_i = 0.8(1 - 0.05(i - 2))$. For the evolution, the parameters, $t_1 = 200$, $t_2 = 201$, and $t_3 = 300$, were used.

5.2 Simulation Results

To test the classification accuracy of the HA-GRNN, the generalization performance over the testing set was evaluated using only LTM Nets (1 to 3), since the pattern space is formed within LTM Nets (1 to 3). For comparison, a conventional GRNN a total of 85 centroids obtained by the well-known MacQueen's k-means clustering algorithm was also used, which yielded the overall generalization performance of 75.0% as shown in Table 1. During testing, 16 pattern vectors among 360 yielded the generation of the intuitive outputs from LTM Net 1 in which 13 out of 16 patterns were correctly classified. It was then found that the Euclidean distances between the 13 pattern vectors and the respective centroid vectors corresponding to their class IDs (i.e., digit number) are relatively small and close to the minimum. From this observation, it can therefore be said that intuitive outputs are likely to be generated when the incoming pattern vectors are very close to the respective centroid vectors in LTM Net 1. In the simulation, three cases, without any attentive states, with the attentive state of Digit /NINE/ only, and those of Digits /FIVE/ and /NINE/ (by following the procedure in Section 4.2), were considered. For the first case without the attentive states, the generalization performance of the HA-GRNN obtained was 84.4%, which outperforms that of the k-means. This indicates that the HA-GRNN can betterly construct the pattern space in comparison with the GRNN with k-means. For the latter two cases, 10 among the 30 centroids in the STM was fixed and used for representing attention, in order to compensate for the 'weaker' covering by the first setup. For the case of Digit /NINE/ only, the overall generalization performance was improved at 85.3%, while the case of Digits /FIVE/ and /NINE/ was further improved at 86.9% as in Table 2. In the table, it is considered that, since the performance for Digit /NINE/ was not improved more than expected, the pattern space for Digit /NINE/ is much harder to fully cover than other digits.

6 Conclusion and Future Direction

In this paper, the two psychological functions, intuition and attention, have been modeled using a newly proposed HA-GRNN. The HA-GRNN is based upon the GRNN model which is essentially easy to implement in hardware (i.e., only two parameters are required to adjust) and robust for learning (or even for forgetting) and yet can approximate any input-output combinations. The concept of the HA-GRNN and its evolution has been motivated from biological and cognitive studies. It has been justified that the two psychological

Digit	0	1	2	3	4	5	6	7	8	9	Total	Generalization Performance
0	35			1	1						34/36	94.4%
1		17				19					17/36	47.2%
2			28		8						28/36	77.8%
3			3	22	10	1					22/36	61.1%
4					36						36/36	100.0%
5						36					36/36	100.0%
6					1	1	34				34/36	94.4%
7	1		3		3	6		23			23/36	63.9%
8					2	1	1		32		32/36	88.9%
9			1			27				8	8/36	22.2%
Total											270/360	75.0%

Table 1. Confusion matrix obtained by the conventional GRNN using k-means clustering method

Digit	0	1	2	3	4	5	6	7	8	9	Total	Generalization Performance
0	30			1	3		2				30/36	83.3%
1		31			2	2				1	31/36	86.1%
2			31	1	3		1				31/36	86.1%
3				32	4						32/36	88.9%
4					36						36/36	100.0%
5		1			1	33				1	33/36	91.7%
6							32	2	2		32/36	88.9%
7			4					32			32/36	88.9%
8							1	1	34		34/36	94.4%
9		3		1		10				22	22/36	61.1%
Total											313/360	86.9%

Table 2. Confusion matrix obtained by the HA-GRNN after the evolution (with attentive states of digits 5 and 9)

functions, intuition and attention, can be interpreted within the framework of evolution of the HA-GRNN. In the simulation study, the models of both the psychological functions have been introduced to form an HA-GRNN using the data set for digit voice classification tasks. The effectiveness has been investigated and its superiority in comparison with a conventional GRNN using the k-means clustering method has also been confirmed.

Note that the proposed layered-memory concept is not limited to pattern classification oriented tasks but can be widely applicable where the tasks and the goals (or, more generally, 'aims') are appropriately known/given, for instance, various planning tasks for autonomous robotics. In such applications, the incoming input vectors are considered as a set of sequential data points, for example, sensory data to know the position of the robot or the internal states, and to interact with other robots, which can alternatively be represented by the attentive states, while the output sequence from HA-GRNN can directly/indirectly change/set parameters to actually control the movement of the robot or the response to other robots by built-in communication facilities, according to the current situations. With preserving the same attentive states in multiple robots, it is considered that the robots can cooperatively work together for a particular goal which is not able to be achieved by a single robot. Future work is therefore directed to the concrete implementation of the proposed neural architecture towards the development of such autonomous robotics.

References

1. G. Matsumoto, Y. Shigematsu, and M. Ichikawa, "The brain as a computer," in Proc. Int. Conf. Brain Processes, Theories and Models, MIT Press: Cambridge, MA, 1995.
2. I. Aleksander, "Impossible minds: my neurons, my consciousness," Imperial College Press, 1996.
3. T. Kitamura, Y. Otsuka, and Y. Nakao, "Imitation of animal behavior with use of a model of consciousness - behavior relation for a small robot," Proc. 4th IEEE Int. Workshop on Robot and Human Communication, pp. 313-316, Tokyo, 1995.
4. D. F. Specht, "A general regression neural network," IEEE Trans. Neural Networks, Vol. 2, No. 6, pp.568-576, Nov, 1991.
5. P. D. Wasserman, "Advanced methods in neural computing," Van Nostrand Reinhold, New York, 1993.
6. T. Hoya and J. A. Chambers, "Heuristic pattern correction scheme using adaptively trained generalized regression neural networks," IEEE Trans. Neural Networks, Vol. 12, No. 1, pp. 91-100, Jan. 2001.
7. D. E. Rumelhart, G. E. Hinton, and R. J. Williams, "Learning internal representations by error propagation," In Parallel Distributed Processing: Explorations in the Microstructure of Cognition (D.E. Rumelhart and J. L. McClelland eds), Vol. 1, Chapter 8, Cambridge, MA: MIT Press, 1986.
8. S. Haykin, "Neural Networks: A Comprehensive Foundation", Macmillan College Publishing Co. Inc., 1994.
9. M. Minsky, "Emotions and the Society of Mind," in Emotions and Psychopathology, Manfred Clynes and Jack Panksepp, eds., Plenum Press, N.Y., 1988.
10. O. Hikosaka, S. Miyachi, K. Miyashita, and M. K. Rand, "Procedural learning in monkeys - possible roles of the basal ganglia," in "Perception, memory and emotion: frontiers in neuroscience," pp. 403-420, eds. T. Ono, B. L. McNaughton, S. Molotchnikoff, E. T. Rolls, and H. Nishijo, Elsevier, 1996.
11. N. Matsumoto, "The brain and biophysics," Kyo-ritsu Shuppan Press, 1997 (in Japanese).
12. M. Huckvale, "Speech Filing System Vs3.0 – Computer Tools For Speech Research", University College London, Mar. 1996.