東北大医短部紀要　12(1)：35～41，2003

パルス結合RBFネットワークにおける音素認識

酒井正夫，杉山泰治*，本間経康**

東北大学大学院　情報科学研究科
*日本アイ・ビー・エム株式会社
**東北大学医療技術短期大学部　診療放射線技術学科

Phoneme Recognition using Pulse Coupled Neural Networks with a Radial Basis Function

Masao SAKAI, Taiji SUGIYAMA*, and Noriyasu HOMMA**

Graduate School of Information Sciences, Tohoku University
**IBM Japan, Ltd.*
***Department of Radiological Technology College of Medical Sciences, Tohoku University*

Key words: Pulse coupled neural networks, radial basis function, and sound pattern recognition

In this paper, we develop a novel pulse coupled neural network (PCNN) for phoneme recognition. One of advantages of the PCNN is in its *biological-based neural dynamic structure* using feedback connections. To recall the memorized pattern, a radial basis function (RBF) is incorporated into the proposed PCNN. Simulation results show that the PCNN with RBF can be useful for phoneme recognition.

はじめに

階層型フィードフォワードニューラルネットワーク（Multilayered Feedforward Neural Network: MFNN）や放射状基底関数（Radial Basis Function: RBF）ネットワークのような従来のニューラルネットワーク（以下NN）のほとんどにおいて，アナログ値信号が情報伝達の媒体として用いられてきた。一方，生体の中枢神経系において，情報伝達は神経インパルスによってなされている。パルス信号がデジタル通信システムに用いられているように，神経インパルスにのせられた情報は高いノイズ耐性をもち，そのため，遠く離れたニューロンへも情報を失うことなく伝達される[1]。このようなパルス信号を扱うモデルとして，パルス結合ニューラルネットワーク（Pulse Coupled Neural Network: PCNN）が提案されている[2]。

PCNNは，サルやネコの視覚皮質において，ニューロンが同期してパルスを発生している現象を基に開発されたNNであり，情報伝達媒体として，生体の脳と同様にインパルス信号を用いているのが特徴である。PCNNはその同期特性により，主として画像処理や，記憶の埋め込みに応用されている[3,4]。しかし，これらは，時間的に変化しない一定の入力に対して，その同期パターンの違いを利用するシステムであり，時間的に変化する入力に対しては適用できない。

本研究では，時間的に変化する音声情報を扱う新しいPCNN[5]を提案する。提案モデルはシナプ

ス演算と細胞体演算それぞれにフィードバックループをもち，そのダイナミクスを利用して入力情報を蓄積できる。また，記憶したパターンを呼び起こすために，放射状基底関数（RBF）を導入した。音素認識のシミュレーションを行い，提案PCNNの有効性を示す。

2　パルス結合ニューラルネットワーク

Fig. 1 にPCNNを構成するニューロンモデルを示す。一つのニューロンは複数の振幅情報を受け取り，軸索を通してインパルス信号を出力する。このモデルで重要な要素は，まず第一にダイナミクスをもつシナプスである。i 番目のシナプスを通して，シナプス前ニューロンからの入力インパルス列 $\boldsymbol{I}(t)=[I_1(t)\ I_2(t)\ \cdots\ I_n(t)]^T\in\Re^n$ は，次式のようにシナプス電位 $\boldsymbol{v}(t)=[v_1(t)\ v_2(t)\ \cdots\ v_n(t)]^T\in\Re^n$ という振幅情報に変換される。

$$v_i(t)=K_v e^{-t/\tau_v}\int_0^t \exp(\tau/\tau_v)I_i(\tau)d\tau,\ (i=1,2,\cdots,n). \tag{1}$$

ここで，n は入力数，K_v は定数，τ_v は $v_i(t)$ の減衰度合いを示す時定数である。シナプスがインパルスを受け取ると，シナプス電位は急激に上昇する。その後は τ_v で決められる割合で指数関数的に減少する。この減衰特性により，入力情報は蓄積されつつも過去の情報は自然に減衰されていく。

PCNNのニューロンは，シナプス結合を通してシナプス電位 $\boldsymbol{v}(t)$ に変換された入力信号を樹状突起で受け取り，次式のような内部電位 $u(t)$ を得る。

$$u(t)=g(\boldsymbol{v}(t)). \tag{2}$$

ここで，$g(.)$ は適当な統合関数であり，従来のPCNNでは次式のような簡単な代数和関数が用いられることが一般的である[2,3]。

$$g(\boldsymbol{v}(t))=\sum_{i=1}^{n} v_i(t). \tag{3}$$

また，この内部電位 $u(t)$ が動的に変化する閾値 $\theta(t)$ を上まわった場合，ニューロンの出力 $y(t)$ が励起される。この関係は，一般的に次のように定義できる。

$$y(t)=\begin{cases}1, & (u(t)>\theta(t)),\\ 0, & (\text{otherwise}),\end{cases} \tag{4}$$

$$\theta(t)=\begin{cases}K_\theta+\theta_0, & (u(t)>\theta(t)),\\ K_\theta\exp\left(-\dfrac{t-t_p}{\tau_\theta}\right)+\theta_0, & (\text{otherwise}).\end{cases} \tag{5}$$

ここで，K_θ は正の定数，τ_θ は時定数，θ_0 はバイアス，τ_p は最後にインパルスが発生した時刻である。内部電位 $u(t)$ が一定な場合の細胞体における動作をFig. 2 に示す。内部電位が閾値を上まわると，ニューロンがインパルス出力を発生し，閾値が $K_\theta+\theta_0$ という高い値に押し上げられる。このように閾値の上昇は，一度発火するとある期間は高い入力を受けても発火しないというニューロンの不応期も実現する[2]。すなわち，内部電位が閾

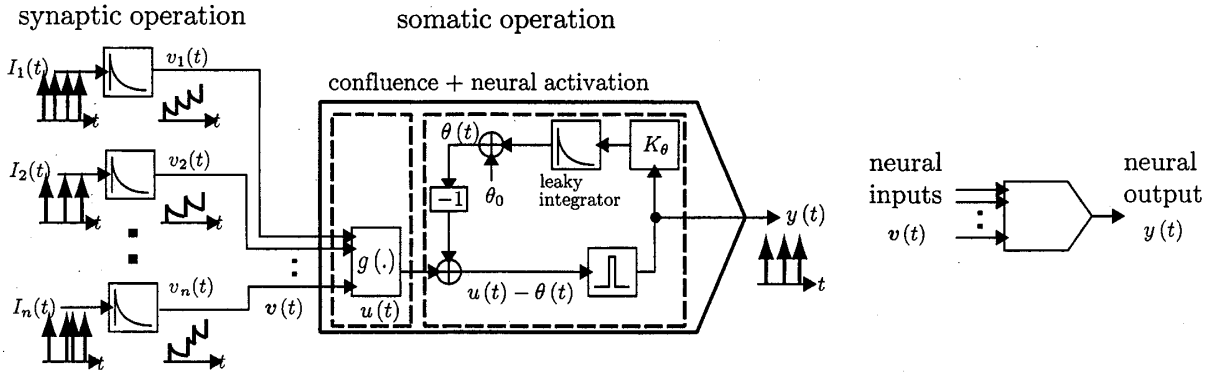

Figure 1. A neural unit in a PCNN.

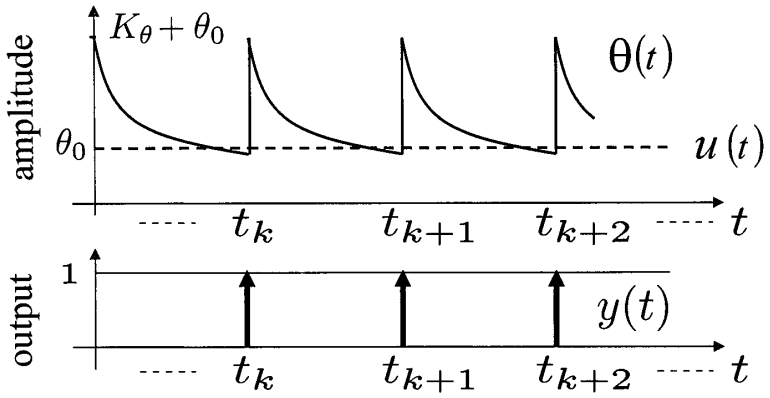

Figure 2. The dynamic threshold $\theta(t)$ and impulse generation for a constant-value input $u(t)$.

値を上まわる時刻を $t_k\,(k=1,2,...)$ とすると，ニューロンのインパルス出力はつぎのように定義できる。

$$y(t)=\frac{1}{\delta(0)}\sum_k\delta(t-t_k). \tag{6}$$

なお，簡単化のために本研究で扱う全ての時系列データは周期 T でサンプリングし離散データとして扱う。すなわち，連続的な時系列データ $p(t)$ $(p\in I, v, u, \theta, y)$ は，以後，離散データとして $p(\ell)\equiv p(\ell T)\,(\ell=1,2,...,L)$ のように表す。ここで，L は総サンプリングデータ数を表す。

PCNN は従来，生体の視神経系，とくにその同期特性を基にしていることから，目標探知等の画像処理に応用されてきた[3]。しかし，本研究では，PCNN の構成ニューロンがもつシナプスのダイナミクスに注目し，その情報蓄積機構を利用して，時間変化する音声情報を取り扱う。

3　音声パターン認識

3.1　認識機構

本研究では，R 個の音声パターンの時系列データ $S^r(\ell)\equiv S^r(\ell T)\,(r=1,2,...,R)$ を，出力数 $m(\geq R)$ の PCNN において認識させる問題を考える。

入力音声の前処理の参考として，人間の音声受容メカニズムを考える。音は外耳を通り，鼓膜を振動させ，振動エネルギーとして蝸牛に伝えられ，蝸牛内での基底膜の振動を通して神経活動に変換される[1]。このとき，基底膜の形状により，基底膜の振幅が大きくなる位置はその振動数毎に異なるため，ある時刻の入力音声情報は，基底膜のどの位置の振幅を大きくするかという地理的情報に変換される。このように，生体においては音は周波数成分毎に分解されてから聴覚野へ入力されている[1]。本研究ではこれにならい，PCNN への入力として，音声時系列をフーリエ変換して得られる周波数スペクトルを用いる。また，生体における周波数変換では時系列情報を蓄積する機能はもち合わせていないため，PCNN における音声パターン $S(\ell)$ のフーリエ変換も全ての時系列をまとめて行うのではなく，任意の間隔の部分時系列毎に行う。すなわち，離散フーリエ変換(DFT)を，Fig. 3 に示すように重複した窓をかけて行うことにより，次式のような周波数スペクトルの時系列を得る。

$$F_i(\ell)=\left|\sum_{w=\ell}^{\ell+W-1}S(w)\exp\left(-j\frac{2\pi(i-1)(w-\ell)}{W}\right)\right|,\quad(\ell=1,2,\cdots,L). \tag{7}$$

ここで，$W(\geq n)$ は DFT に用いられる部分時系列のデータ数，添え字 $i\,(i=1,2,\cdots,n)$ は周波数に対応した番号を意味する。

つぎに，PCNN における学習について考える。任意の r 番目の音声パターン $S^r(\ell)$ が PCNN に入力された場合，そのネットワーク出力 $\boldsymbol{y}^r(\ell)=[y_1^r(\ell)\ y_2^r(\ell)\ \cdots\ y_m^r(\ell)]^T$ のうち，r 番目の要素 $y_r^r(\ell)$ だけが応答するようにネットワークを学習

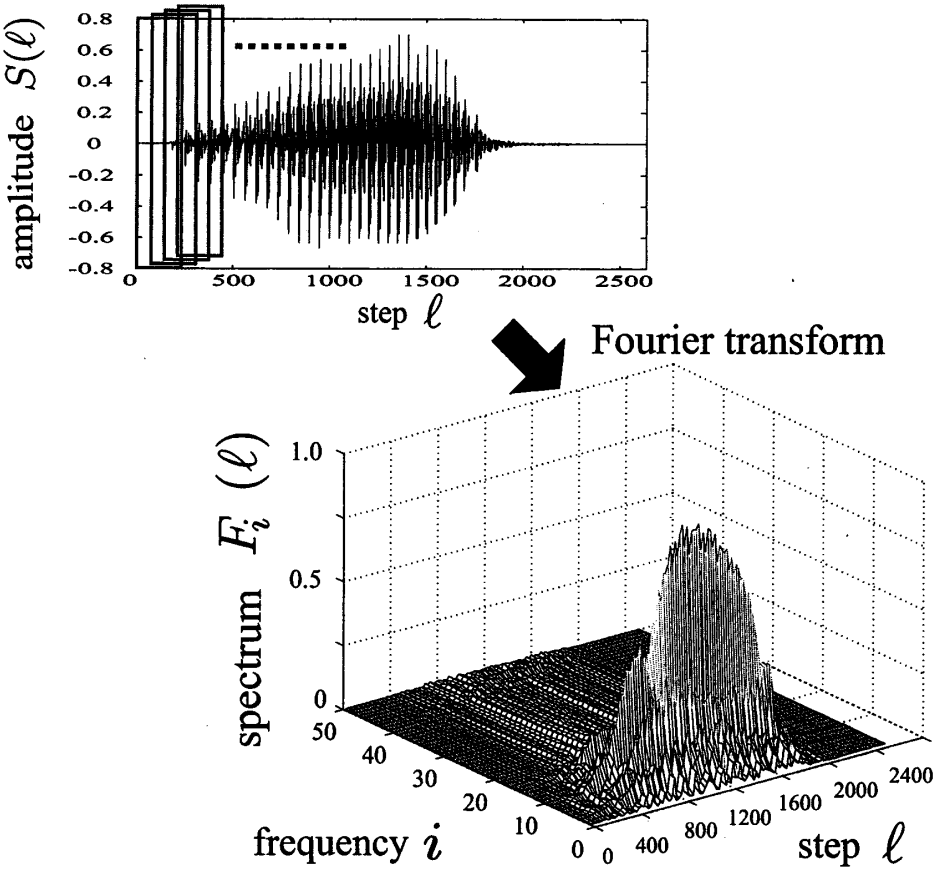

Figure 3. The transformation into frequency spectrum sequence $F_i(\ell)$.

させる。このとき，必要な教師信号 $\boldsymbol{y}^{rd}(\ell)=[y_1^{rd}(\ell)\ y_2^{rd}(\ell)\ \cdots\ y_m^{rd}(\ell)]^T$ は，次式のように定義できる。

$$y_k^{rd}(\ell)=\begin{cases}1, & (k=r),\\ 0, & (k\neq r),\end{cases}\quad (k=1,2,\cdots,m).$$

また，評価関数 E は次式のように定義できる。

$$E=\sum_{\ell=L_1}^{L_2}\sum_{r=1}^{R}\|\boldsymbol{y}^{rd}(\ell)-\boldsymbol{y}^{r}(\ell)\|^2,$$
$$=\sum_{\ell=L_1}^{L_2}\sum_{r=1}^{R}\sum_{k=1}^{m}(y_k^{rd}(\ell)-y_k^{r}(\ell))^2. \qquad (8)$$

ここで，L_1 と L_2 は評価開始と終了の時刻である。

本研究では，この評価関数 E を最小にすることを目的とする。しかし，Eckhorn が提案した前節で示しすようなニューロンで構成された PCNN には，記憶を埋め込むことができるモデルは提案されていない。そこで，本研究では，入力に対応した適切な出力を得るために，放射状基底関数 (RBF)[6] を用いて記憶を埋め込む手法を提案する。RBF は，従来の PCNN におけるシナプス電位 $\boldsymbol{v}(\ell)$ より各ニューロンの内部電位 $u(\ell)$ を求める統合関数 $g(.)$ に置き換えられる。すなわち，k 番目のニューロンの内部電位は，RBF を用い次式のように得られる。

$$u_k(\ell)=g_k(\boldsymbol{v}(\ell))=\exp\left(-\frac{\|\boldsymbol{v}(\ell)-\boldsymbol{w}_k\|^2}{2\sigma^2}\right). \qquad (9)$$

ここで，σ^2 は正規化パラメータ，$\boldsymbol{v}(\ell)\in\Re^n$ は時刻 ℓ におけるシナプス電位ベクトル，$\boldsymbol{w}_k\in\Re^n$ は RBF の中心であり，k 番目のニューロンに保持された記憶に相当する重みベクトルである。すなわち，ベクトル $\boldsymbol{v}(\ell)$ と $\boldsymbol{w}_k$ のユークリッド距離が小さい場合，内部電位 $u_k(\ell)$ が大きくなり k 番目の出力ニューロンが発火する。

3.2 ネットワーク構造

本研究では，パターン認識のために，Fig. 4 で表されるような新しい 2 層構造の PCNN を提案する。はじめに，入力層のダイナミクスを考える。入力層は，n 個のニューロンを持ち，ステップ ℓ における i 番目のニューロンは，周波数スペクトルの i 番目の要素を入力 $I_i^{(1)}(\ell)$ として受け取る。また，その入力はそのまま内部電位 $u_i^{(1)}(\ell) \equiv I_i^{(1)}(\ell)$ として用いられ，動的に変化する閾値 $\theta_i^{(1)}(\ell)$ との関係から，(4), (5) 式のようにニューロンの出力 $y_i^{(1)}(\ell)$ が求まる。

次に，出力層のダイナミクスを考える。入力層の出力 $y_i^{(1)}(\ell)$ は，そのまま出力層への入力 $I_i^{(2)}(\ell) \equiv y_i^{(1)}(\ell)$ として用いられ，(1) 式より出力層のシナプス電位 $\boldsymbol{v}(\ell) = [v_1(\ell)\ v_2(\ell)\ \cdots\ v_n(\ell)]^T \in \Re^n$ が求まる。これより，各ニューロンが保持する記憶 $\boldsymbol{w}_k$ との関係より，(9) 式のように内部電位 $u_k^{(2)}(\ell)$ が求まる。また，動的に変化する閾値 $\theta_k^{(2)}(\ell)$ との関係から，(4), (5) 式のようにネットワークの出力 $y_k^{(2)}(\ell)$ が求まる。

この提案ネットワークを用いて音声パターンを記憶するには，その音声パターンが入力された場合のシナプス電位ベクトル $\boldsymbol{v}(\ell), (L_2 \leq \ell \leq L_2)$ を，対応する出力ニューロンの重みベクトル $\boldsymbol{w}_k$ として記憶させる必要がある。そこで，音声パターンに対するシナプス電位 $\boldsymbol{v}(\ell)$ の時間変化に注目する。Fig. 5 に音素パターン「お」を入力した場合のシナプス電位 $\boldsymbol{v}(\ell)$ の時間変化を示す。この図より，$\boldsymbol{v}(\ell)$ のほぼ全ての成分が時間経過に対し山型になり飽和する区間があることがわかる。したがって，本来シナプス電位 $\boldsymbol{v}(\ell)$ は時刻 ℓ によって動的に変化するパラメータであるが，本研究では記憶ベクトル $\boldsymbol{w}_k$ を静的パラメータとし，その値は経験的に決定する。

4 シミュレーション

5 つの日本語の音素パターン「あ」，「い」，「う」，「え」，「お」を認識対象とし，提案した PCNN の性能を評価する。各音素パターンはサンプリング周波数 8[kHz] で，2,500 ステップ (0.3125 秒) からなる。本研究で用いたネットワークの入力層のニューロン数は $n=50$，出力層のニューロン数は $m=5$，また，各種定数を経験的に $K_v=1$, $\tau_s=150$, $K_\theta^{(1)}=5$, $\tau_0^{(1)}=0.5$, $\theta_0^{(1)}=0.1$, $K_0^{(2)}=5$, $\tau_\theta^{(2)}=2$, $\theta_0^{(2)}=0.5$, $\sigma^2=0.4$, $N=128$, $L_1=1{,}300$, $L_2=1{,}600$ と設定している。

音素パターン「お」を入力した際の認識結果を Fig. 6 に示す。この結果より，入力「お」に対応す

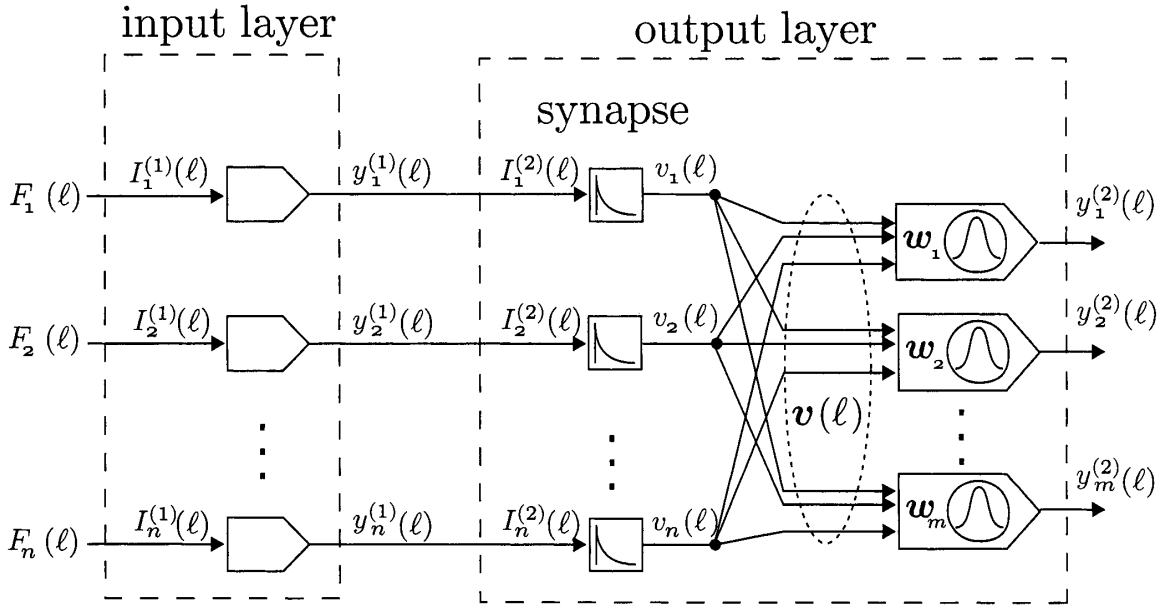

Figure 4. PCNN for phoneme recognition.

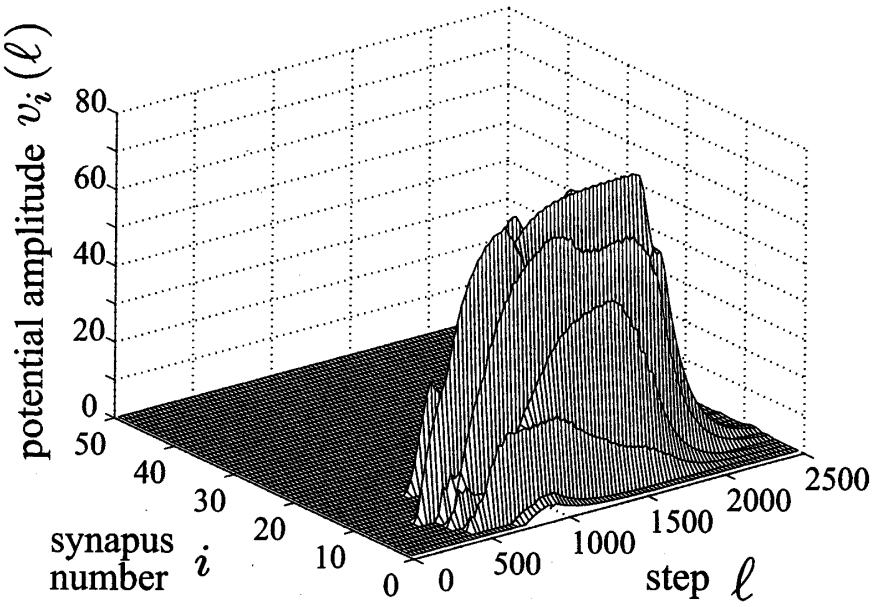

Figure 5. The synaptic potentials $v(\ell)$ for phoneme input "お".

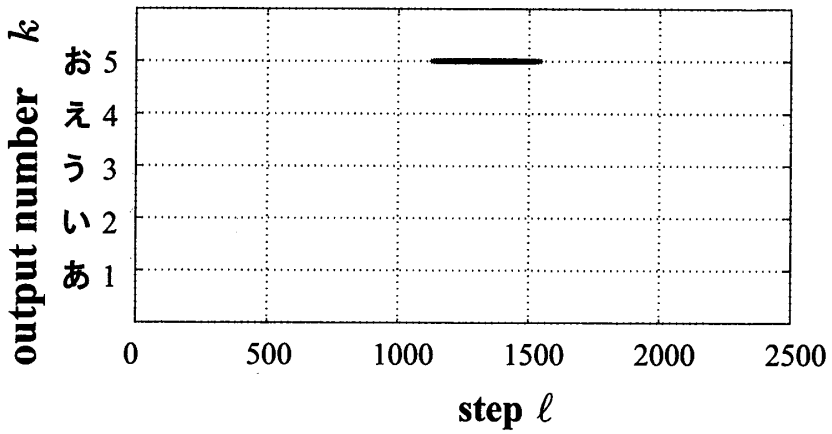

Figure 6. The recognition result for phoneme input "お".

る出力ニューロンのみが連続的にインパルスを出力していることがわかる。また，その他の音素パターンを入力した場合にも，Fig. 6 の結果と同様に正しい認識結果が得られることが確認されている。

ま と め

本研究では，PCNN を用いた認識ネットワークを提案した。提案モデルでは，ニューロンの入力の統合関数として，RBF を用いることにより，PCNN に記憶を埋め込むことを可能にした。このネットワークのもう一つの重要なメカニズムはシナプスにある動的な情報蓄積機構である。このシナプスのダイナミクスにより，入力情報は蓄積されつつ，古い情報は自然に減衰され，影響力は弱まっていく。このように，過去の情報を適当に考慮した認識を行うことは，生体においては極めて自然な作業であり，さらに，この減衰機能は時間遅れ要素を用いた離散ネットワークでは困難であった自然な認識区切りを実現できる可能性をも

つ。 PCNN は生理学的に観測されたパルスが基礎にある現象を計算機上で再現するのに有効なネットワークであり，他の認識問題に対してもフィードバック結合をもつパルスダイナミクスは有効であると考えられる。

文　献

1) Delcomyn, F.: ニューロンの生物学，南江堂，2000
2) Eckhorn, R., Reitboeck, H.J., Arndt, M., Dicke, P.: Feature linking via synchronization among distributed assemblies: Simulations of results from cat visual cortex, Neural Comput., **2**, 293-307, 1990
3) Broussard, R.D., Rogers, S.K., Oxley, M.E., Tarr, G.L.: Physilogically motivated image fusion for object detection using a pulse coupled neural network, IEEE trans. on Neural Networks, **10**-3, 554-563, 1999
4) Izhikevich, E.M.: Weakly pulse-coupled oscillators, FM interactions, synchronization, and oscillatory associative memory, IEEE trans. on Neural Networks, **10**-3, 508-526, 1999
5) Sugiyama, T., Homma, N., Abe, K.: Speech recognition using pulse coupled neural networks with a radial basis function, Proc. 7th AROB, **1**, 355-358, 2002
6) Possio, T., Girosi, F.: Networks for approximation and learning, IEEE Proc., **78**-9, 1481-1497, 1990