論文要旨

システム制御情報学会論文誌, 4 巻, 1 号, pp. 1~10 (1991)

可能性と必然性の概念に基づくニューラルネットの学習アルゴリズム

石渕 久生・田中 英夫

本研究では、2 群判別の問題を取り上げ、ニューラルネットによる可能性分析と必然性分析の方法を提案する。まず、可能性と必然性の概念に基づく判別分析の方法を示す。つぎに、可能性判別分析および必然性判別分析のためのニューラルネットの学習アルゴリズムを提案する。提案手法では、ニューラルネットの結合強度の学習は、重み付き 2 乗誤差の総和が最小になるように行なわれる。このとき、データ点ごとの重みを学習と共に徐々に変化させることにより、可能性あるいは必然性を考慮した判別分析を行なうことができることを示す。さらに、例題を用いて提案手法の説明を行なう。

システム制御情報学会論文誌, 4 巻, 1 号, pp. 11~20 (1991)

ニューラルネットを用いたリモートセンシングデータのパターン認識

吉田 知司・大松 繁・寺西 大

本論文はリモートセンシングデータのパターン認識問題へニューラルネットを用いる手法を提案している。これによって、誤認識の割合を減少させかつ従来の Bayes 法による手法と比較して識別誤差のバラツキの少ないデータ処理結果が得られることを実データに基づいて定量的に検証している。さらに、雲の影の補正を行なって分類する手法に対しても本論文の手法が有効であることを示す。

システム制御情報学会論文誌, 4 巻, 1 号, pp. 21~27 (1991)

階層型神経回路モデルによる時空間パターンの認識

伊藤 崇之

時空間パターンを認識できるニューラルネットワークモデルを提案する。モデルは特徴抽出部と統合部の二つの部分からなっている。前者は、入力信号の立ち上がり、立ち下がりなど、時間に固有の特徴を抽出する。後者は、 P 細胞と T 細胞の 2 種類の細胞からなっている。 P 細胞は特徴の統合を行なう。 T 細胞は、メモリ機能を持つ。これら 2 種類の細胞層を交互に継続的に接続し、位置不変性と変形の許容能力も実現している。コンピュータシミュレーションによって、モデルがパターンの時間伸縮や、位置ずれ、変形を許容できることを示した。

システム制御情報学会論文誌, 4 巻, 1 号, pp. 28~36 (1991)

チェビシェフネットワークによるパターン領域の識別

生天目 章・上田 延寿

コネクショニストモデルを用いて、2 次元領域のパターンを識別する問題について考察する。適当な数の中間ユニットを配置した従来の多層ネットワークによる識別ではパターン領域が複雑になるにつれてその識別能力は低下し、中間ユニットの数や学習例を増大させてもその識別性能は改善されないことを示す。また、一般に非線形の複雑な決定領域におけるパターン識別システムをコネクショニストモデルにより実現するためにはシグモイドのような単調な非線形関数のユニットだけでは不十分でチェビシェフ関数のような非単調関数を持つユニットとの組み合わせが必要であることを示す。従来のモデルに代わるものとしてチェビシェフネットワークという新しい多層ネットワークアーキテクチャを提案する。パターン領域の境界曲線をチェビシェフネットワークに学習させ、その出力値の論理結合により識別する方法および領域の境界曲線近傍における対象領域内外のいくつかのパターンをチェビシェフ関数により非線形変換し、それらを正負の学習例として学習・識別させる二つの方法について提案する。いずれの方法においても新しく提案したチェビシェフネットワークによる識別方法は、一般にパターン領域の複雑性に依存しない優れた識別性能を有することを示す。

システム制御情報学会論文誌, 4 巻, 1 号, pp. 37~47 (1991)

フィードバック誤差学習による閉ループシステムの学習制御

五味 裕章・川人 光男

本論文では神経回路モデルを非線形フィードバックコントローラに利用するときの学習法を提案する。学習後には制御対象の非線形性が補償され、参照モデルで規定される応答に従うようになる。

ここでは、この学習スキームの収束性を確率微分方程式の平均化とリアプノフの方法で証明し、有効性を倒立振り制御のシミュレーションの結果により示す。また、この学習スキームと動物の小脳における姿勢・歩行の学習制御との関連を議論する。

論 文

階層型神経回路モデルによる時空間パターンの認識*

伊 藤 崇 之**

Recognition of Spatio-temporal Patterns by a Multi-layered Neural Network Model*

Takayuki Ito**

In this paper, a neural network model is proposed which can recognize spatio-temporal patterns.

The model is composed of feature detection module and feature integration module. The former extracts time-specific features such as onsets and offsets of input signals. The latter consists of two types of neurons, a *P*-neuron and a *T*-neuron. A *P*-neuron integrates features. A *T*-neuron has a memory mechanism with input-dependent decay time. Pairs of these two types of neural layers are connected in cascade in the model. Combination of these two types of layers in hierarchical structure produces an ability of duration-independent recognition of spatio-temporal patterns.

In a computer simulation, it has been shown that the model can tolerate wide range of distortion of input patterns such as elongation and shrink along the temporal axis, shift along spatial axis and deformation in shape.

1. はじめに

最近、音声認識やパターン認識など人間の得意とする高度な情報処理能力を機械で実現するために、従来のような直列型計算機によるディジタル中心の処理ではなく、アナログ情報を並列的に処理するニューラルネットワークの有用性が注目を浴びている。ここでは、音声など情報が時系列で入力されるパターンの認識アルゴリズムを提案する。

聴覚系に関する生理学的知見については、すでによく知られているように、

- 蝸牛の基底膜において周波数分析されている。
- 基底膜以降では、各周波数成分の信号の立ち上がりに対応する細胞、立ち下がりに対応する細胞、さらには周波数に変化する部分に対応する細胞(FM細胞)などで処理がなされる。
- さらに上位の内側膝状態では特定の周波数の組み合わせに反応する細胞、ノイズに反応する細胞などが存在する。

などの機能が明らかになっており¹⁻³、その特徴抽出機構に関する神経回路モデルも提案されている^{4,5}。

しかしながら、これよりさらに上位の脳聴覚野以上についてはほとんど生理学的知見は得られていない。

ネコの場合、誘発電位法を用いた実験⁶で下位の聴覚系と同じ規則性で特徴周波数が分布している(tonotopic organization)ことが示されているものの、個々の細胞に対しては具体的にどのような入力が最適な刺激なのか、またどのような処理が行なわれているのかは明らかになっていない。

音声の認識過程を、時間領域の信号から種々の特徴を抽出し、異なる時刻で抽出されたそれらの特徴を統合するという二つのステップに分けて考えると、生理学的には後半のステップはほとんど明らかになっていないというのが現状である。

ところで、音声のような時系列信号を処理する場合、異なる時刻での情報をどのように統合するかが問題となる。上で述べた第2のステップの、情報を統合・認識するモデルとしては、

- ① 遅延回路などを用いて、あらかじめ時間軸の情報を空間軸に変換して扱うモデル。つまり問題を空間パターンの認識に置き換えたモデル^{7,8}。
- ② 記憶素子など情報を保持する機能をもつ素子を用い

* 原稿受付 1990. 5. 1

** NHK 放送技術研究所 視覚情報研究部

Key Words : neural network, spatio-temporal pattern, recognition, neocognition, memory.

て時系列パターンを認識するモデル。

③ 順序検出器等によって信号の系列を検出するモデル⁹。

などの枠組みが考えられる。

ここで述べるモデルは、この分類にしたがえば、②に属する機能を持つものである。本モデルは時間に固有の特徴を抽出する特徴抽出部と、入力パターンに適応して持続時間が変化する記憶機構を持つ特徴統合部からなり、時間軸方向の伸縮に強い認識能力を持つ。またネオコグニトロンに類似したネットワーク構造により、パターンの変形や位置ずれにも対応し得るモデルである。以下、ネットワークの機能を説明するとともに、その基本的な能力について単純な時空間パターンを用いて検証する。

2. モデル

2.1 モデルの構造

2.1.1 特徴抽出

モデルの構造を Fig. 1 に示す。入力として、音声ホルマントパターンのように、空間軸（周波数軸）と時間軸をあわせ持つ二次元の時系列パターンを想定している。

先にも述べたように音声認識の過程を二つに分け、前半で時系列信号に固有の特徴が抽出され、後半部分で特徴の統合・認識が行なわれると仮定する。特徴抽出部は、1で述べた聴覚神経系の末梢部に相当する部分で、筆者らが以前に構成した特徴抽出モデル⁵をそのまま用いることもできる。ここでは簡単化して、信号の立ち上がりを検出する細胞と立ち下がりを検出する細胞がそれぞれ

周波数軸上に並んだ2種類の細胞群、すなわち ON 細胞層と OFF 細胞層を用いている。空間軸上の細胞の番号を n ($n = 1, \dots, N$) で表わし、時間を一定間隔の離散時刻（間隔 Δt を単位時間とする）で表わすすると、時刻 t での、ON 細胞の出力 $u_{ON}(n, t)$ 、OFF 細胞の出力 $u_{OFF}(n, t)$ をそれぞれ、次のようにして求めている。

$$u_{ON}(n, t) = u_{IN}(n, t) - u_{IN}(n, t-1) \quad (1)$$

$$u_{OFF}(n, t) = u_{IN}(n, t-1) - u_{IN}(n, t) \quad (2)$$

ここに、 $u_{IN}(n, t)$ ($n = 1, \dots, N$) は時刻 t における入力パターンである。これらの細胞層の出力は二つの短時間の遅延素子群に入力される（図では“D”で示される二つの長方形で表わされている）。したがって、入力された時空間パターンは特徴抽出機構で各特徴に分解されたのち、狭い時間幅の観察窓（今回のシミュレーションでは、時間幅3にしてある）を通してつぎつぎとモデルに取り込まれることになる。

この遅延素子層を入力として、次の層で 3×3 の結合で特徴抽出を行なっているので、ON 細胞層と OFF 細胞層で得られた反応の組み合わせを特徴として捕らえるだけでなく、斜めの直線という形でホルマントパターンにおける FM 上昇音や FM 下降音に相当する特徴も検出していると言える。

2.1.2 P 層と T 層

処理の中心部分は P 層と呼ぶ層と T 層と呼ぶ層を複数段、縦続接続した階層構造を成す。Fig. 1 では、ひと

Fig. 1 Structure of the model. The numbers in the bottom row indicate the size of a plane and the number of planes in each layer used in the simulation (see section 3); (cells along spatial [vertical] axis) $\times$ (cells along temporal [horizontal] axis) $\times$ (the number of planes)

つの層を点線の長方形で表わしている。また、各層の上に示されている P または T はその層の種類を、数字は段の番号を表わす。 P 層、 T 層ともさらにいくつかの細胞群に機能分割される。ここではそれらを細胞面と呼ぶ (Fig. 1 において、実線で描かれたたて長の長方形がひとつの細胞面に対応する)。細胞面には細胞が一次元に並んでおり、入力面での空間的位置関係が各細胞面でも保たれている (tonotopic organization)。

P 層と T 層の機能は、視覚系の神経回路モデル、ネオコグニトロン¹⁰ の S 層と C 層の機能にはほぼ対応する。すなわち P 層は特徴の抽出・統合を行ない、 T 層は空間的、時間的な位置ずれを許容する。また入力に近い層では狭い範囲の特徴抽出、奥の層ほどより大局的な特徴抽出という関係を空間的にも時間的にも保たせている。

P 層の細胞と一段前の層との間の結合は可変である。ここでは最大値検出型¹⁰ の学習規則により、特定の入力パターンにたいして選択的に反応するように教師なしで強化される。ネオコグニトロンの場合と同様、学習した細胞と同じ結合がその細胞面の他の細胞にも複写されるメカニズムを仮定している、同じ特徴抽出能力を持った細胞が異なる位置にもできあがる。

T 層の各細胞面は直前の P 層の各細胞面と一対一対応している。 T 層のそれぞれの細胞は P 細胞面の対応する位置の細胞とその近傍の細胞からはほぼ等しい結合強度で入力を受け取っている。このように P 細胞層での反応を T 細胞層でぼかすことによって、 P 細胞で抽出された特徴の位置ずれを許容する¹¹。同時に、 T 細胞は一次遅れ特性を持っているので、入力側の P 細胞の反応が終わっても一定の定数で減衰しながら反応し続ける。すなわち T 細胞は、特定の特徴が抽出されたことを一定時間記憶する能力を持つ。

いま、いくつかの特徴が異なる時刻に検出される場合を考える。それらの特徴に反応する T 細胞群の出力は、それぞれ一定の定数で減衰しながら持続する。その反応が重なりあっている時刻に次の段の P 細胞がそれらを同時に観測して学習すれば、ある時間領域で検出されるいくつかの特徴をまとめた高次の特徴として統合することができる。ところが特徴間の間隔が離れすぎて先行特徴に対する反応がすでに減衰してしまっている場合には、同じ特徴の組み合わせとして検出されない。その結果、時間方向の特徴間隔の変動に対する能力が低下する。そこで本モデルでは、 T 細胞が出力を出しているときは、近傍領域の他の面からの入力も受け入れる機構を導入し、反応の持続時間が入力に依存して変化するようなしくみを組み込んでいる (Fig. 2)。このような機能を実現する神経機構としては、異シナプス性結合 (hetero-

Fig. 2 Schematic illustration of the mechanism to change the duration time of T-cell

synaptic connection) による促進作用、抑制細胞の組み合わせなどを考えることができる。

ただし、この機構の導入によって、さきに述べた P 細胞面と T 細胞面の一対一対応の関係がくずれることはない。 T 細胞の反応が他の P 細胞の反応によって強められるのは、その T 細胞が出力を出しているときだけという条件があり、あくまでも出力の大きさを変化させているだけである。

ここで、ある時刻 t における l 段目の P 層において、第 k 面内の位置 n における細胞の出力を $u_P^l(n, k, t)$ と表わす。ただし、 $l = 1, \dots, 4, k = 1, \dots, K_P^l$ であり、位置 $n (= (n_x, n_t))$ は、すべての層で統一的に表わすために D 層に射影した位置とする。同様に、 l 段目の T 層の第 κ 面の細胞の反応を $u_T^l(n, \kappa, t)$ と表わす。ただし、 $l = 0, \dots, 3$ で、 $u_0^l \equiv u_D$ 、また、 $\kappa = 1, \dots, K_T^l$ である。今回のシミュレーションにおける K_P^l, K_T^l の値を Fig. 1 に示しておく。 P 層と T 層の細胞の出力をこのように表わすと、 P 層の細胞出力は以下の式で求められる。

$$u_P^l(n, k, t) = r^l \cdot \varphi \left[\frac{g^l + \sum_{\nu \in A^l} a^l(\nu, \kappa, k) \cdot u_T^{l-1}(n+\nu, \kappa, t)}{g^l + \frac{r^l}{1+r^l} \cdot b^l(k) u_P^l(n, t)} - 1 \right] \quad (3)$$

$$\kappa = 1, \dots, K_T^{l-1}, \nu = (\nu_x, \nu_t) \in A^l,$$

$$k = 1, \dots, K_P^l, n = (n_x, n_t)$$

ここに関数 φ は

$$\varphi[x] = \max(x, 0) \quad (4)$$

である。 $a^l(\nu, \kappa, k), b^l(k)$ は、それぞれ興奮性および抑制性の可変入力結合の強度である。また A^l は興奮性結合の結合範囲を表わす。 r^l は抑制の強さを表わす正

定数, g^l は反応の飽和特性を制御する正定数である.

(3) 式中の $v_P^l(n, t)$ は P 細胞 $u_P^l(n, k, t)$ に抑制性信号を送っている抑制性細胞の反応の強さを表わす. この細胞は P 細胞が入力を受け取るのと同じ細胞群から入力を受け, その2乗平均値に等しい出力を出す. すなわち

$$v_P^l(n) = \sqrt{\sum_{\nu \in A^l} \sum_{\kappa \in A^l} c^l(\nu) \cdot \{u_T^{l-1}(n + \nu, \kappa, t)\}^2} \quad (5)$$

で表わされる. $c^l(\nu)$ はこの細胞への固定入力結合であり, ここではガウス関数を用いている, κ, ν の範囲は (3) 式と同じものである. これらの式で明らかなように, P 細胞は時間的な要素を持たない. このことは, P 細胞が非常に早い時定数で応答すると解釈することもできる.

一方, T 細胞の特性は以下のように表わされる. ある時刻 t における l 層, k 面の T 細胞の反応 $u_T^l(n, k, t)$ は

$$u_T^l(n, k, t) = \Theta[I_0 + I_1 + I_2 - \theta] \quad (6)$$

ただし

$$I_0 = \left(1 - \frac{1}{T^l}\right) u_T^l(n, k, t-1) \quad (7)$$

$$I_1 = \sum_{\nu \in D^l} d^l(\nu) u_P^l(n + \nu, k, t) \quad (8)$$

$$I_2 = \alpha \cdot \varphi \left[\sum_{\kappa \neq k} \sum_{\nu \in A^{l+1}} d^l(\nu) u_P^l(n + \nu, k, t) - \beta \cdot \varphi[1 - \gamma \cdot u_T^l(n, k, t-1)] \right] \quad (9)$$

である. また θ は

$$\Theta[x] = \min(\max(x, 0), 1) \quad (10)$$

と表わされる飽和関数である. θ はいき値を定める正定数であり, α, β, γ は正定数である. I_0 は細胞自身の内部電位による出力であり, I_1 はその細胞の属する k 番目の細胞面に一対一対応する P 層の細胞面からの, 現時刻での入力である. $d^l(\nu)$ は興奮性の固定結合であり, D^l はその結合範囲を表わす, 一方 I_2 は P 層の k 番目以外の面からの入力に出力制御用の項が加わったものであり, $d^l(\nu)$ も興奮性の固定結合である.

(6) 式において, I_2 がゼロ, すなわち他の面からの入力がない場合は, T 細胞は非線形特性を無視すれば時定数 T^l の一次遅れ要素としてふるまう. すなわち P 層の対応する面からの入力 I_1 がなくなると時定数 T^l で減衰する. ところで I_2 は, β, γ が1に比べて充分大きいと仮定すると,

$$I_2 \approx \begin{cases} \alpha I_1^l(n, k, t) & (u_T^l(n, k, t-1) > 0) \\ 0 & (u_T^l(n, k, t-1) = 0) \end{cases} \quad (11)$$

となる. すなわち自分自身が一時刻前に反応して ($u_T^l(n, k, t-1) > 0$), かつその近傍で P 層の他の特徴面が反応しているときは, 時定数 T^l で決まる持続時間をこえて反応が持続することになる. この近傍領域として, ここでは次の段の P 層の結合可能領域 (A^{l+1}) を用いている.

2.2 結合係数の学習

このモデルでは P 層の各細胞と一段前の T 層の細胞との間の結合を学習によって求めている. 結合係数の強化法には最大値検出型学習を用いる. すなわち, 各場所ごとにすべての細胞面の中で最大反応を示す細胞を選びだし, さらにその中から極大反応を示す (複数の) 細胞を代表として選ぶ. そのときの代表細胞への入力の強さに比例した強さで興奮性結合 $a^l(\nu, \kappa, k, t)$ を強化する. その細胞の属する細胞面の他の細胞にも同じ強化量を与える. 代表が $\hat{k}$ 面の位置 n_0 で選ばれたとすると, このときの興奮性結合の変化量 Δa^l は次式で表わされる.

$$\Delta a^l(\nu, \kappa, \hat{k}) = q^l \cdot c(\nu) u_T^{l-1}(n_0 + \nu, \kappa, t) \quad (12)$$

q^l は正定数で, 学習の速度を制御するパラメータである. また, 抑制性結合 b は毎回次式に従って計算するものとする¹².

$$b^l(\hat{k}) = \sqrt{\sum_{\kappa \in A^l} \sum_{\nu \in A^l} c(\nu) \{u_T^{l-1}(n_0 + \nu, \kappa, t)\}^2} \quad (13)$$

すなわち, 抑制性結合は興奮性結合の重み付きノルムに相当する. このように他のシナプス結合の強度を元に構成されるシナプスが実際に存在するかどうかについて生理学的には明らかでないが, (3) 式が学習パターンと認識時の入力パターンをベクトルとみなしたときの両ベクトルのなす角を基準に出力を出しているという数学的意味¹³ は, 従来の学習法¹⁰ に比べて明確になる.

モデルの動作を学習モードと認識モードに分けて説明する.

学習モードでは, 入力に近い方から一層ずつ学習を進める. 一つの層の学習について見てみると, ある時刻で最初に結合強化を受けた細胞面は, 次の時刻で似たような入力であれば, 引き続き最大出力を出し, さらに結合強化を受ける. かなり異なったパターンが入力されたとき, まだ強化を受けていない新たな細胞面が最大出力を出して強化を受けるよう, 特定の特徴に偏らない弱い初期結合を与えておく. この措置によって細胞面の機能分化がうながされる. 入力パターンを繰り返し呈示して, 新たな細胞面が形成されなくなるまで学習を進める.

最終層でも各時刻に結合強化が起こると仮定するなら, 一つのパターンの全体像を学習する細胞だけができあが

Fig. 3 Training patterns and the response of final layers after training

るとは言いきれず、パターンの一部分だけを学習してしまう細胞が生じることもある。そもそも入力パターンのどこからどこまでを一つのまとまりとみなすかという能力は、意味情報のフィードバックやすでに学習している内容からの類推など、かなり高度な機能のもとに成り立っているとも考えられるが、ここでは、入力からのフィードフォワードの情報として、パターンとパターンの間の切れ目が一定の時間続いたときに反応する細胞を想定して、この細胞から最終層への教師信号が送られると仮定している。

3. 計算機シミュレーション

このモデルについて、Fig. 1 の下部に示した回路規模で計算機シミュレーションを行なった。細胞面数は学習前から固定されているものではなく、学習の結果得られた数である。

学習パターンは、Fig. 3 上部に示す“A”から“E”までの5文字のアルファベットが時系列パターンとして、一定の速さで左から右に現れてくる。5個のパターンを連続的に呈示するのを1サイクルとして、これを2回繰り返した。学習は1サイクル目でほとんど終了した。

今回用いたパラメータの値をTable 1に示す。表の中で r (train), r (response) はそれぞれ学習モード、認識モードでの r の値を意味する。また、(9)式中の α は各層とも0.05とした。また β, γ は1に比べて十分大きな値を取ると考えて(9)式の変わりに(11)式を用いている。

Table 1 Parameters used in the simulation

	Layer 1	Layer 2	Layer 3	Layer 4
r (train)	6.0	5.0	5.0	3.0
r (response)	4.0	3.0	3.0	3.0
g	0.0002	0.0005	0.005	0.05
q	10.0	50.0	10.0	500.0
T	1.25	5.0	10.0	—

(5), (8), (9) 式で用いている固定結合 $c^l(\nu)$, $d^l(\nu)$, $d'(\nu)$ はいずれもガウス関数で、関数の広がりを表わす σ は各層固有の座標軸上で1.0としている。

Fig. 3 の下部に、最終層の5個の細胞の学習終了後の反応の時間経過を示す。5個の細胞がそれぞれ“A”から“E”のパターンに選択的に反応していることがわかる。

4. 検 討

4.1 パターンの部分に対する反応

このモデルでは、パターンの全体を同時に入力面で観測することはできないので、最終層の細胞が入力パターンの一部にしか反応していない可能性がある。そこでモデルがパターンのどの部分に反応しているかを確認するために、Fig. 4 に示すように、入力パターンの一部のみを呈示して、そのときの反応の強さを調べた。図はパターン“C”の場合である。図で明らかなようにパターン全体の80%以上呈示したとき初めて反応が得られており、モデルがほぼパターン全体を観察していると考えることができる。

特徴抽出部でパターンの時間的な変化部分（信号の立ち上がり、立ち下がり、FM成分）のみが抽出されるので、パターンの形によってこの割合は異なる。しかし、学習パターンの“C”のように時間軸に水平な部分（周

Fig. 4 Response property of the model to a part of the pattern “C”

波数一定部分)の両側に特徴がほぼ均等に存在する場合、特徴の出現のあとさきにかかわらず、均等に検出できていることを示している。

他のパターンについても、その後半部分のみを呈示しただけでは反応しないことを確認した。

4.2 パターンの時間軸方向の伸縮

音声認識の場合、時間軸方向に伸縮したパターンに対して高い認識能力が要求される。特に音声では非線形に時間伸縮するというのが特徴的である。このモデルでは、時間軸方向のパターン伸縮に対処するメカニズムとしては、

- ① 特徴抽出部における持続的な反応の制御と T 細胞の反応の持続時間可変の記憶機能
- ② $P1$ 層の FM 成分 (斜めの直線) の抽出能力

(a) with the mechanism of changing duration time

(b) without the mechanism ($\alpha = 0$).

Fig. 5 Response of the model to prolonged or shrunk patterns

に分けられるであろう。①はON層, OFF層で抽出された特徴間の間隔の変化(音声について言えば周波数の安定した母音の長さの変化)にどれほど対応できるか, ②については時間軸方向の伸縮によって生じた直線の傾き(子音および母音間の“わたり”の傾き)の違いを同じ特徴として抽出できるかどうかという問題に対処する方策である。

ここでは、第1の問題について検討した。Fig. 5はパターン“D”の時間軸に平行な部分のみを伸縮したパターンに対する反応で、同図(a)が持続時間可変機構を導入した場合、(b)が同機構がない($\alpha = 0$)場合の結果である。図より明らかなように、時間軸に平行な部分をおよそ3倍に伸ばしてもほぼ正しく反応している。一方、同機構が無い場合(他のパラメータは等しくして、同じ学習パターンで学習し直している)、パターンに対応する細胞が最大出力を出しているが、他のカテゴリーに対応する細胞も反応を示し、かつ伸びたパターンほど誤った出力が大きくなっている。したがって、本モデルで導入した持続時間可変機構は有効に機能していることが確かめられた。

4.3 変形や位置ずれパターンに対する反応

Fig. 6は変形やノイズの含まれたパターンに対する反応の例である。図でも明らかなように、かなり大きな変形を受けている場合には反応が弱くなっているものもあるが、学習パターンに反応した細胞と同じ細胞がそれぞれのカテゴリーに対して正しく反応していることがわかる。

4.4 特徴の出現順序

このモデルでは、複数の特徴に対する持続反応が重なっている時点で上位の層の細胞が学習することにより、時間軸方向の特徴統合を行なう。この場合、統合した細胞の反応は、統合されたそれぞれの特徴の反応強度には依存するが、出現順序に依存しない。その意味では、本モデル

Fig. 6 Response of the model to deformed patterns

ルは時系列信号の順序を検出しているのではなく、時系列を構成する各要素の存在を検出しているというべきであろう。

ただし、特徴統合部の1層目で直線検出を行なっているので、直線の傾き(上昇FM音か、下降FM音か)という形で局所的な因果関係を検出しているといえる。音声の場合は特に、FM音は前後の母音や先行子音、後続子音などの情報を含んでいると考えられる¹⁴⁾ので、FM音の検出だけでもかなりの前後関係を抽出していると考えることができる。

5. まとめ

時間に固有の特徴を抽出する特徴抽出部と、入力パターンに適応して持続時間を変化し得る記憶機構を組み込んだ階層的なニューラルネットワークを構成し、その基本的な特性について単純な時空間パターンを用いて検討した。その結果、本モデルが、(1)信号の立ち上がり、立ち下がりを検出する特徴抽出と持続時間可変のメカニズムにより、パターン時間軸方向の伸縮に強い認識能力を持つこと、(2)ネオコグニロンに類似したネットワーク構造により、パターンの変形や位置ずれにも強い認識能力を持っていることを示した。

本モデルは特徴系列の順序関係を検出してはいないが、検討の項でも述べたように、音声を認識対象とする場合にはそれほど問題はないと考えられる。むしろ特徴の順序を厳密に検出するモデルでは個々の特徴が正しく検出できない場合に、逆に認識能力のロバスト性を失う可能性が高い。音声を対象として、モデルの能力を検証することが今後の課題である。

謝辞 研究の機会を与えて頂いた、NHK放送技術研究所の大川昌彦前所長、杉本昌穂所長、および視覚情報研究部日下秀夫前部長、元本紀雄部長に感謝します。本研究を進めるに当たり、終始懇切なご指導を頂いた大阪大学福島邦彦教授(現在)に深謝します。また、日頃ご指導頂いている同部三宅副部長をはじめとする多くの方々に謝意を表わします。

参考文献

- 1) Y. Katsuki N. Suga & Y. Kanno : Neural Mechanism of the Peripheral and Central Auditory System in Monkeys ; J. of the Acoustical Society of America, Vol. 34, No. 8, pp. 1396~1410 (1962)
- 2) M. Nomoto : Discharge Pattern of the Primary Auditory Cortex in Cats ; Jpn. J. Physiol., Vol. 15, No. 5, pp. 427~447 (1980)
- 3) N. Suga : Classification of Inferior Collicular Neurons of Bats in Terms of Responses to Pure Tones, FM Sounds and Noise Bursts ; J. Physiol., Vol. 200, pp. 555~574 (1969)
- 4) 大串：聴覚計の情報処理機構のモデル；信学論(D), 45-D巻, 4号, pp. 332~339 (1971)
- 5) 伊藤, 福島：聴覚系の特徴抽出機構の神経回路モデル；信学論(D), J 70-D巻, 2号, pp. 451~462 (1987)
- 6) P. L. Knight : Representation of the Cochlea within the Anterior Auditory Field (AFF) of the Cat ; Brain Research, Vol. 130, pp. 447~467 (1977)
- 7) A. Waibel : Modular Construction of Time-Delay Neural Networks for Speech Recognition ; Neural Computation, Vol. 1, No. 1, pp. 39~46 (1989)
- 8) D. W. Tank & J. J. Hopfield : Neural Computation by Concentrating Information in Time ; Proc. Natl. Acad. Sci. USA, Vol. 84, pp. 1896~1900 (1987)
- 9) 二見, 星宮：相互結合型神経回路の状態遷移に基礎をおく時系列パターンの認識；信学論(D), J 71-D巻, 10号, pp. 2181~2190 (1988)
- 10) 福島：位置ずれに影響されないパターン認識機構の神経回路モデル-ネオコグニロン-；信学論(A), 62-A巻, 10号, pp. 658~665 (1979)
- 11) K. Fukushima : Analysis of the Process of Visual Pattern Recognition by the Neocognitron ; Neural Networks, Vol. 2, No. 6, pp. 413~420 (1989)
- 12) 有山, 塚田, 伊藤：誤り訂正型学習を導入したネオコグニロンによる英数字認識；信学技報, NC 89-91巻 (1990)
- 13) 福島：コグニロンのパターン分離能力の向上；信学論(A), 62-A巻, 10号, pp. 658~665 (1979)
- 14) 今泉, 桐谷, 高橋, 白井：速度可変型ホルムント合成規則の検討；信学技報, SP 89-64巻 (1989)