論　文

側頭葉短期記憶力学系の神経回路モデル

正　員　森田　昌彦†

A Neural Network Model of the Dynamics of a Short-Term Memory System in the Temporal Cortex

Masahiko MORITA †, *Member*

あらまし　サルの側頭葉TE野には，持続的な発火によって短期記憶を担っていると見られる一群のニューロンがあり，ある種の力学系の平衡状態として記憶情報が保持されていると考えられる．ところが，このニューロン群の振舞いは，従来の神経回路モデルによって実現される力学系ではうまく説明できない．本論文では，側頭葉の短期記憶系および連想記憶の力学について考察し，TE野と同様な力学的性質を実現する神経回路モデルを提案する．このモデルは，その構成単位が非単調な入出力特性をもつ点で従来のモデルと本質的に異なっており，脳の記憶回路の動作と機構を解明する上で重要な意味をもつ．

1. まえがき

相互結合のある神経回路網は，一種の力学系を構成する．そして，Hopfieldのモデル[1]をはじめとする連想記憶モデルの多くが，記憶すべきパターンを系のアトラクタ(エネルギーの極小状態)とすることによって連想記憶機能を実現している．しかし，神経回路網は非線形で多変数の力学系であり，一般的な回路網はもちろん，対称結合の回路網のように力学構造が比較的単純なものでも，その性質には不明な点が多い．古くから知られている自己相関型の連想記憶モデル[2]の場合でさえ，その奇妙な力学的性質が明らかになったのはごく最近のことである[3],[4]．

神経回路網の力学に関してもう一つ重要な点は，これまで用いられてきた力学系が連想記憶その他の情報処理を行うのに最適なものだという保証はどこにもない，ということである．実際，想起のダイナミックスを変えるだけで，従来の連想記憶モデルの能力が大幅に向上することがわかっている[4]．

一方，生理学の分野では，脳内のニューロンの活動が直接測定されるようになり，いくつかの重要な事実が明らかにされつつある．特に，宮下[5],[6]によって報告されている側頭葉TE野のニューロンの活動は，短期記憶がどのような形で情報表現されているかを示唆し，記憶のメカニズムを考える上で非常に興味深い．同時に，これは実際の脳の神経回路，特に記憶系の力学的構造を知るための貴重な手掛りである．

ところが，このニューロン群の振舞いには，従来型の連想記憶モデルでは説明できない部分がある．もう少し強く言うならば，側頭葉の短期記憶を保持する神経回路の力学系は，これまで考えられてきたような神経回路モデルでは実現することができない．つまり，従来の力学系は，脳のモデルという観点からも本質的な改善の余地があるのである．

これらの点から考えると，脳の記憶回路には，今までのモデルには欠けていた何らかの重要な原理や機構が存在するに違いない．その原理は何であり，それをどうモデル化すればよいのか，というのが本研究の主題である．

以下では，まずTE野ニューロン群の振舞いについて簡単に説明し，神経回路網の力学の観点から検討を加える．次いで連想記憶のダイナミックスとその改良法に関する考察をもとに，従来とは異なる力学的性質をもつできるだけ単純な神経回路モデルを構成する．また，シミュレーション実験によりその動作を調べ，TE野の力学系と比較する．更にこのモデルが予言する興味深い現象を示し，最後に本研究が示唆する事柄と

† 東京大学工学部計数工学科，東京都
Faculty of Engineering, The University of Tokyo, Tokyo, 113 Japan

今後の課題について述べる．

2. 側頭葉短期記憶回路の力学

2.1 TE野の短期記憶ニューロン

一瞬(0.2秒)提示されたフラクタル図形と，16秒の遅延期間の後画提示された図形とを比較するという課題(実際の実験手順はもっとち密である)をサルに課したところ，遅延期間中興奮し続ける一群のニューロンが下部側頭葉(主にTE野)に見つかったという[5],[6]．このニューロンには多くの特筆すべき性質があるが，以下の議論と直接関係するものを簡単にまとめると次のようになる(但し，これらには甘利ら[7]および筆者の解釈が含まれている)．

(1) 繰返し学習し見慣れた図形を提示した場合，一つのニューロンはごく少数の図形(100枚のうち2，3枚)に対してだけ強く反応する．それよりやや弱い反応を示す図形も数枚あるが，その他大多数の図形に対してほとんど反応を示さない．また，強い反応を引き起こす数枚の図形の間に，特に共通する特徴はない．

(2) 見慣れた図形に対する反応の再現性は高い．また，回転や拡大・縮小などの変換を施した図形を提示しても，ほとんど同じ反応を示す．

(3) 以前に見たことのない新奇な図形を提示したとき，強い反応を示すことはほとんどない．しかし，比較的多数の図形に対して弱く反応する．見慣れた図形の場合より反応の再現性が低く，時間的な変動も大きい．

以上は個々のニューロンに対する実験結果であるが，これからある図形に対するニューロン群全体の反応を次のように推測することができる．

(4) 見慣れた図形を提示したときにTE野に現れるのは，全体のごく少数だけが強く興奮しているパターン(「スパース」なパターン)である．但し，それよりやや弱い反応を示すニューロンや，ごく弱い興奮を持続するニューロンもある程度含んでいる．このパターンは提示される図形ごとに異なるが，多少の変形に対しては不変である．

(5) 新奇な図形を提示すると，見慣れた図形の場合よりもやや多数のニューロンが弱く興奮しているパターンが現れる．そのパターンはやはり図形によって異なるが，同じ図形を見せても同じパターンが現れるとは限らない．

2.2 系の力学的性質

さて，これらのニューロンは互いに結合し合って一つの力学系を構成しており，TE野ニューロン同士の相互作用によって短期記憶(どのような図形を提示されたかについて16秒間覚えている情報)を保持しているものと考えられる．短期記憶を担う別のニューロン群があって，そこから強い持続的入力を受けている可能性も残されているが，そのようなニューロン群は発見されていない．また仮にあったとしても，それを含めた全体の系に対して同じ議論が成り立つ．そこで，TE野はそれ自体が短期記憶を保持する力学系になっているとして，次にその性質について考察する．

まず，見慣れた図形はスパースな興奮パターンとして保持されるが，安定で再現性が高いこと，図形の多少の変形に影響されないことから，このパターンは力学系の強い(引込み領域が広い)アトラクタになっていると考えられる．また，各図形はその図形的特徴に関係なくコードされているから，こうした強いアトラクタはかなり一様に分布していることになる．

一方，新奇な図形もスパースなパターンとして表現されるが，強い興奮を持続するニューロンがほとんどないこと，安定性および再現性があまり高くないことから，このパターンはあまり安定でない平衡状態ないし不完全なアトラクタ(後述)になっていると考えられる．また，このような状態は非常にたくさんあり，初めて見る図形はそのうちのどれか一つにコードされると考えるのが自然であろう[7]~[9]．

従って，TE野の短期記憶回路は，少数の強いアトラクタとそうでない多数のアトラクタとが共存し，両者でニューロンの反応強度の分布が異なるような力学系であるということができる．

2.3 従来の力学系との比較

ところが，以上のような性質をもつ力学系を神経回路網で実現しようとすると，大きな問題があることに気が付く．それは，中くらいの強さの興奮を持続するニューロンの割合が大き過ぎるという点である．

これについて説明するために，まず各細胞が0から1のアナログ値をとるHopfieldモデル[10]を考えよう．i番目の細胞C_i $(i=1,2,\cdots,n)$の出力をx_i，平均膜電位をu_iとすると，そのダイナミックスは次式で与えられる．

$$\tau\frac{du_i}{dt}=-u_i+\sum_j w_{ij}x_j-h_i \qquad (1)$$

$$x_i=f(u_i) \qquad (2)$$

ここで，τは時定数，h_iはしきい値，w_{ij}はC_jからC_iへの結合荷重で，$w_{ij}=w_{ji}$，$w_{ii}=0$である．また，出力

関数 $f(u)$ は $u\to\pm\infty$ でそれぞれ 0，1 とる単調増加関数であり，

$$f(u)=\frac{1}{1+e^{-cu}} \qquad (3)$$

を用いることが多い（c は正の定数）．

よく知られているように，この回路網はエネルギー関数

$$E=-\sum_{i<j}w_{ij}x_ix_j+\sum_i h_ix_i+\sum_i g(x_i) \qquad (4)$$

が時間と共に減少するように動作し，系のアトラクタは E の極小点に対応する．ここで，

$$g(x)=\int_{1/2}^{x}f^{-1}(y)dy$$

$$=\frac{1}{c}(x\log x+(1-x)\log(1-x)+\log 2) \qquad (5)$$

であり，$g'(1/2)=0$ である．

今，式(3)の定数 c の値が十分大きいものとする．これは，どの細胞も他の細胞と十分強く結合していることと等価である．$g(x_i)$ が一定ならば $\partial E/\partial x_i$ は x_j $(j\neq i)$ の 1 次関数であるから，x_i が 1/2 付近の値をとるとき，$\partial E/\partial x_i=0$ となることはほとんどない．逆に，E が極小となるのは，ほとんどすべての x_i が 0 または 1 に非常に近い値をとる場合に限られる．従って，いくつかの細胞が 0 と 1 の中間的な値を出力するような状態は，強いアトラクタとなり得ない．一方，c の値が小さい（相互結合の強さが弱い）と，外部からの入力がなくなったとき，初期状態によらず系の状態はいつも同じになってしまう．また，一部の細胞が他と弱く結合している場合，常にそれらの細胞だけが中間的な出力を出すことになる．

このように，少なくとも Hopfield モデルのような対称結合の回路網では，いくつかの細胞が中間的な出力値をとるアトラクタの存在を説明できない．このような性質は，おそらく式(1)〜(3)で表される力学系の本質的な構造とかかわっており，結合の対称性を多少崩す程度では変化しない．また，非対称な神経回路網の力学に関するいくつかの理論的研究[11)〜13)]から考えて，たとえ対称性の条件を完全に取り払ったとしても，一様な構造（w_{ij} の分布が i，j に極端には依存しない）のモデルでは，前述のような TE 野ニューロン群の振舞いは説明できないと思われる．

結局，従来のダイナミックスを用いて側頭葉短期記憶回路の力学系のモデルを構成するためには，回路網に何らかの構造を導入しなければならない．では，どのような構造を考えればよいのだろうか．

確かに，大脳の神経回路には比較的整然とした構造があり，何種類ものニューロンからなる局所的な回路がある．しかし，この回路はかなり複雑であり，そのままモデル化することは困難であるし，できたとしてもあまり意味があるとは思えない．情報処理の原理を明らかにするには，もっと別のアプローチをとる必要があろう．

そこで，以下では連想記憶モデルのダイナミックスとその改良法について考察し，神経回路網の力学的性質を改善する上で何が本質的かを探ることにする．

3. 連想記憶のダイナミックスの改良

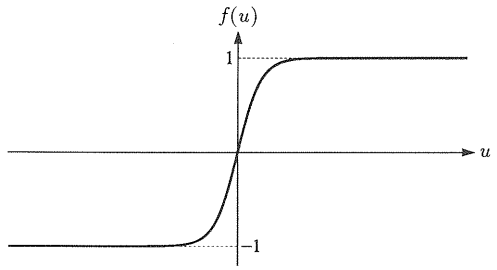
3.1 非単調出力関数の利用

神経回路網によって作られる力学系を用いて，連想記憶のモデルを構成することができる．例えば，前述の Hopfield 回路で荷重行列 $W=[w_{ij}]$ を適当に定め，与えられたパターン S がエネルギー極小状態になるようにすれば，S に十分近い初期状態が与えられたとき，系の状態 X は時間と共に S に近づき，うまく S が想起できると考えられる．

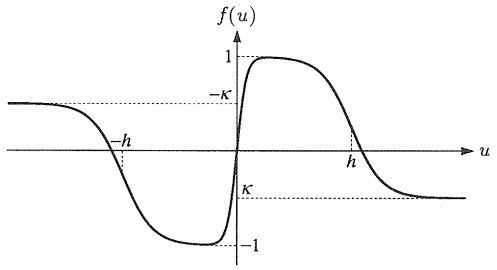
しかし，実際の想起の過程はそれほど単純ではなく，X が S にかなり近づきながらその後遠ざかってしまうことがしばしばある．特に自己相関モデルの場合，細胞数 n の約 15 % 以上のパターンを記憶させようとすると，必ずこのような現象が生じる．また，このとき，X は最終的に記憶したパターン以外の平衡状態（偽の記憶）に達するが，そのときの細胞の出力を見ただけでは，真の記憶パターンと区別がつかない点も大きな問題である（TE 野の場合，見慣れた図形と新奇な図形とでニューロンの反応が異なることに注意）．

こうした問題は，従来のようなエネルギーの最急降下というダイナミックスを使う限り，どのような荷重行列 W を用いても解決されない．その一方で，W として単純な自己相関行列を用いた場合でも，平衡状態において x_i ではなく u_i の分布を調べれば，正しく想起できたかどうかがわかる．つまり，従来のダイナミックスでは，W に含まれている情報の一部が使われていなかったわけである．逆に，この情報を利用するには，想起のダイナミックスを改良する以外に方法はない．

実際，例えば Hopfield の連想記憶モデル（連続時間で x_i が -1 から 1 の間の連続値をとるものと考える）で，式(2)の $f(u)$ を従来のシグモイド形単調増加関数（図 1(a)）から図 1(b)のような非単調関数に変えると，回



（a） Conventional sigmoid function



（b） Improved nonmonotonic function

図1 アナログ細胞モデルの出力関数の改良
Fig. 1 Improvement of the output function $f(u)$.

路網の連想記憶能力が大きく向上する[4],[14]．また，正しい想起ができなかったときには系の状態がいつまでも変化し続けるため，偽の記憶を想起してしまうことがなくなる．

このように，出力関数 $f(u)$ の形を変えるだけで力学系の性質が大幅に変化するが，その際に最も本質的なのは $f(u)$ の非単調性である．

3.2 スパースコーディング

連想記憶のダイナミックスを改良するもう一つの手段として，いわゆるスパースコーディング[15]を用いる方法がある．これは，記憶するパターンをスパースな(0,1パターンで考えたとき1の割合が非常に小さい)ものに限るという方法である．n が一定のとき，これにより1パターン当りの情報量が減るから，より多くの記憶パターンを蓄えることが可能になる．

但し，連想能力を高めるためには，想起の際に系の活動度(細胞の出力値の総和)をほぼ一定にする必要がある．つまり，記憶パターンの活動度(1をとる要素の数)に関する情報を利用して系のとり得る状態を制限することが，スパースコーディングによるダイナミックスの改良において本質的な役割を果たす．

活動度を低いレベルに保つという方法は，TE野を含めて現実の脳で広く行われていると考えられる．ところが，連想記憶モデルの場合，活動度を一定に保つ機構を自然な形で実現することは，それほど簡単ではない．単純に活動度に応じた抑制をフィードバック回路によって加える方法は，振動が起こりやすいなど，なかなかうまくいかない．しかし，もし入力がある程度大きくなると出力が減少に転じるような特性が個々の細胞に備わっていれば，活動度のコントロールはかなり容易になると考えられる．このことは，後述のシミュレーション実験でも確かめられる．

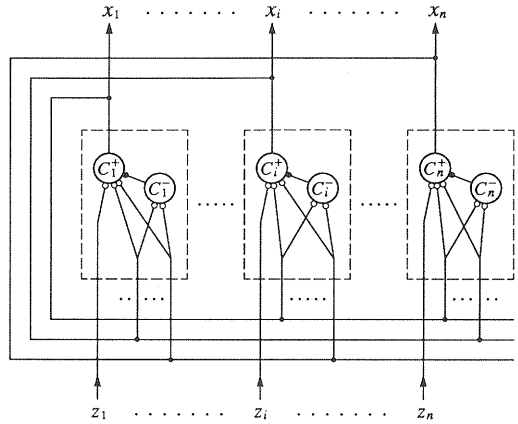


図2 モデルの構成
Fig. 2 Structure of the model.

4. モデルの構成

前章で述べたように，素子の非単調特性と連想記憶のダイナミックスの改良とは密接にかかわっている．しかしながら，現実のニューロンが単調な特性をもつことは事実であるから，モデルをなるべく自然なものにするために，複数の細胞の組合せを考えなければならない．そして，非単調な特性を得るには，細胞の出力ではなく入力に応じて抑制を加えること，すなわちフィードフォワード型の抑制が不可欠である．

以上のような考察に基づいて構成したのが，以下に示すモデルである．まず，全体の構成を図2に示す．

図の破線で囲まれた部分(ユニットと呼ぶ)がモデルの構成単位であり，従来の一つの細胞に対応する．i 番目のユニットは出力細胞 C_i^+ と抑制細胞 C_i^- の二つからなるが，前者はこのユニットの出力 x_i を出す細胞，後者は y_i を出力することによって前者に強い抑制を加える細胞である．他のユニットからの入力は C_i^+ と C_i^- の両方に入るが，系の外部からの刺激 z_i は C_i^+ だけに入力される．式で表すと

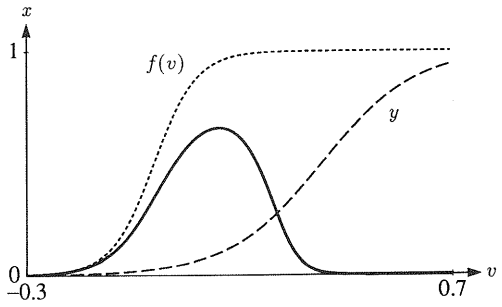


図3 ユニットの入出力特性
Fig. 3 Input-output characteristics of a unit.

$$y_i=f\left(\lambda\sum_{j\neq i}w_{ij}^{-}x_j-\theta\right) \tag{6}$$

$$\tau\frac{du_i}{dt}=-u_i+\sum_{j\neq i}w_{ij}^{+}x_j-w_Iy_i+z_i \tag{7}$$

$$x_i=f(u_i) \tag{8}$$

となる．ここで，w_{ij}^{+} と w_{ij}^{-} はそれぞれ j 番目のユニットから C_i^{+} および C_i^{-} への結合荷重，w_I は C^{-} から C^{+} (i によらない) への結合荷重，λ，θ および τ は正の定数である．また，出力関数 $f(u)$ として 0 から 1 の値をとるシグモイド関数 (式(3)) を用いるが，C^{-} は C^{+} よりも緩やかな反応特性をもつ ($\lambda<1$)．

今，すべての j について $w_{ij}^{+}=w_{ij}^{-}$ が成り立っているとしよう．このとき，C_i^{-} が受ける入力の荷重和は C_i^{+} への入力の荷重和

$$v_i=\sum_{j\neq i}w_{ij}^{+}x_j \tag{9}$$

と常に等しいから，このユニットの出力 x_i は v_i だけの関数となる．ここでパラメータを適当に選べば，$y_i=f(\lambda v_i-h)$ の非線形性のため，x_i は v_i の増加に対して単調に増加するのではなく，図3のようなベル型の特性を示す．

w_{ij}^{+} と w_{ij}^{-} とが等しくない場合にも，両者の相関が高ければ，やはり入力 v_i がある程度大きくなったところで出力が減少に転じる．但しこの場合には，v_i の大きさは同じでも，入力のパターン (どのユニットからの入力が大きいか) によって x_i の値が異なることになる．

ところで，図1(b)から考えると，入力が小さすぎるときに出力が増加するような特性を各ユニットにもたせる方が良いように思われる．しかし，そうするとモデルが必要以上に複雑になってしまうので，このような構成にした．また，ここに示したのはあくまでも TE 野の力学系のモデルであり，モデルの細胞と現実のニューロンの間に1対1の対応があるわけではない．むしろ，ある程度の数のニューロンの平均的動作が一つのユニットによって表されていると考えるべきであろう．但し，大脳の神経回路には局所的な抑制回路がごく一般的に見られ，その大部分がフィードフォワード型だという興味深い知見があることを指摘しておく．

5. モデルの動作

5.1 スパースなパターンの連想記憶

さて，上記の回路網に m 個のスパースなパターン S^1, S^2, …, S^m を記憶させるとしよう．ここで，$S^{\mu}=(s_1^{\mu},\cdots,s_n^{\mu})$ は，n 個の要素のうち l ($\ll n$) 個だけが1で残りは0であるようなパターンの中からランダムに選ばれるものとする．

ここでは，

$$w_{ij}=\frac{1}{l}\sum_{\mu=1}^{m}\left(s_i^{\mu}-\frac{l}{n}\right)\left(s_j^{\mu}-\frac{l}{n}\right) \tag{10}$$

で与えられる行列 $W=[w_{ij}]$ を利用する．$l=n/2$ のとき，この W は自己相関連想記憶モデルで用いられる荷重行列と等価である．但し，本モデルの場合，活動度に応じて全体に抑制を加えることが必要なので，結合荷重を

$$w_{ij}^{+}=w_{ij}-\alpha/l \tag{11}$$

$$w_{ij}^{-}=w_{ij} \tag{12}$$

とする．ここで α は正の定数であり，一様な相互抑制の大きさを表している．

記憶したパターンを想起する際は，適当な想起入力 $P=(p_1,\cdots,p_n)$ を

$$z_i=kp_i+z_0 \tag{13}$$

という形で入力する．ここで k は想起入力の大きさ，z_0 は全体の刺激レベルを表す．想起入力は十分な時間 (時定数 τ の数倍) 与え続ける必要がある．P が記憶パターンの一つ (S^1 とする) に十分近ければ，系の状態は S^1 をコードするアトラクタ (後述) に引き込まれ，k が0になった後もその状態を保持するものと考えられる．

実際にこのモデルについてシミュレーション実験を行い，その動作を調べた．実験は $n=1{,}000$，$m=400$，$l=100$ で行った．このとき，一つのユニットは，400個のパターンのうち平均40個をコードすることになる．その他のパラメータは次のように選んだ．

$c=50$，$w_I=1.0$，$\lambda=0.2$，$\theta=0.1$，

$\alpha=0.2$，$k=0.1$，$z_0=-0.1$

以下，実験結果について定性的に述べることにする．

図4は，$P=S^1$ を想起入力として与えた後，十分に時間が経過した時点での各ユニットの出力値 x_i の分布

を表すヒストグラムである．灰色の部分は S^1 をコードする（$s_i^1=1$ が成り立つ）100 個のユニットを表すが，これらが比較的大きな出力を出しているのに対して，その他のユニットはほとんど出力を出さないことがわかる．また，0.1～0.5 の中間的な値を出力するユニットもかなり存在する．このような分布は，見慣れた図形を提示したときの TE 野ニューロンの興奮強度の分布とよく一致する．

このとき系に多少の外乱を与えてもすぐにもとに戻るから，図の状態は安定な平衡状態である．また，P が S^1 と相当に異なっていても同じ平衡状態に達するので，強いアトラクタになっている．そして，出力パターン $X=(x_1,\cdots,x_n)$ は S^1 に最も近く，それ以外の記憶パターンとほとんど相関がないから，このアトラクタは S^1 をコードするものだといってよいであろう．なお，シミュレーションの結果を見る限りでは，一つのパターンをコードするアトラクタは一つしかない．但し，n および m が非常に大きい場合には，複数の平衡状態が狭い領域にかたまって存在している可能性もある．

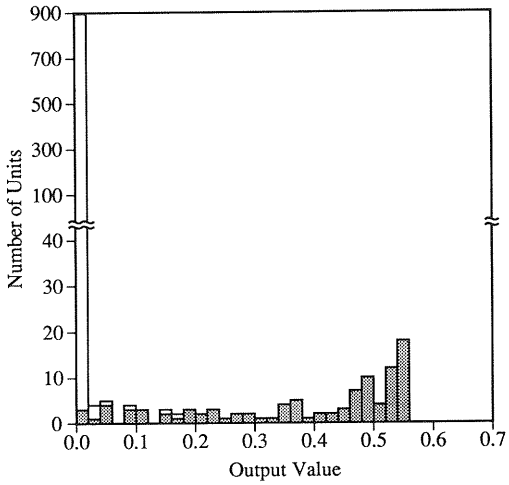
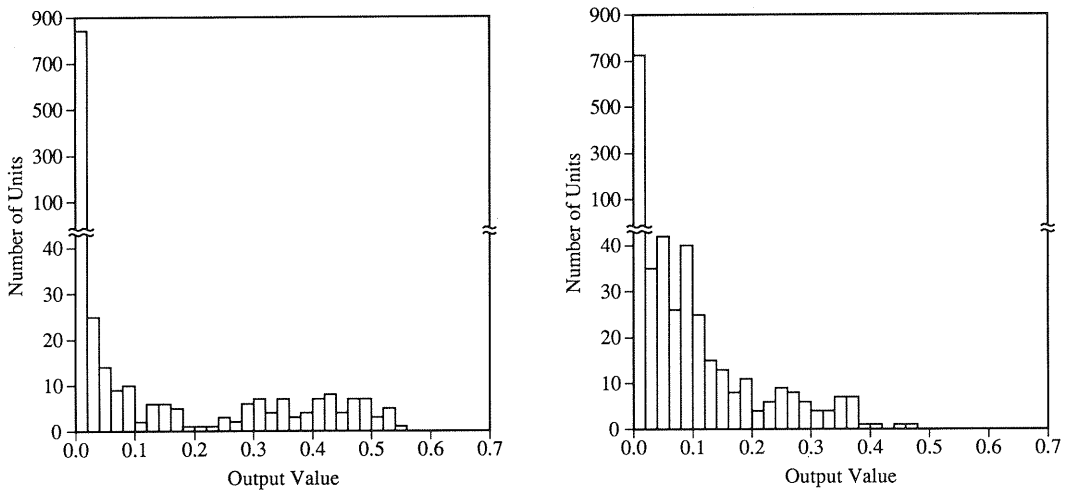


図 4　S^1 を入力したときの出力値の分布
Fig. 4　Distribution of x_i for $P \fallingdotseq S^1$.

これに対して，どの記憶パターンとも全く異なるランダムなパターンを入力すると，出力値の分布は図 5（a）のようになる．但し，これはある瞬間の出力値ではなく，$t=5\tau$ から $t=25\tau$ までの出力を平均した値の分布である．この間，系の状態はゆっくりと変化し続け，その後も半永久的に変化が続く．しかし，一度このような状態に達した後は，相当長い時間にわたってほぼ同じパターンが保たれ，多少の外乱を受けてもその付近の状態にとどまる．従って，このような状態（不完全なアトラクタと呼ぶことにする）は，短時間の情報保持に利用できると考えられる．

こうした不完全なアトラクタの数は極めて多く，想起入力が少し異なるだけで別のアトラクタに引き込まれる．そこで，図 5（a）で入力したパターン P の 1％（10 個）の要素を変えたものを五つ用意し，それぞれを入力したときの出力値を平均したところ，図 5（b）のよ



(a) Response for once　　(b) Average over 5 times

図 5　ランダムパターンを入力したときの反応
Fig. 5　Distribution of the time-averaged outputs to a random pattern.

うな分布を示した．反応の再現性が低く，大きな出力を出すユニットほど数が少ない点は，新奇な図形に対するTE野ニューロンの反応と符号する．

なお，不完全なアトラクタは，従来の連想記憶モデルにおける偽の記憶に相当するものだと考えられるから，その数は n の指数オーダで増加する．逆に，n や m が小さすぎると，荷重行列を変えて強いアトラクタの引込み領域を狭めない限り，不完全なアトラクタはほとんど存在しないことになる．

5.2 複数のパターンの保持

我々は一度に複数(7個程度までと言われる)の短期記憶を保持できるが，サルの場合もある程度これが可能である．では，複数の図形を覚えているとき，TE野の力学系はどのような状態にあるのだろうか．

この問題を考えるために，以下のような実験を行った．但し，前節の条件下では一度に一つのパターンしか保持できなかったので，式(10)の代わりに

$$W=\Sigma(\Sigma^T\Sigma)^{-1}\Sigma^T \qquad (14)$$

によって与えられる w_{ij} を用いた† ．ここで，Σ は s_i^μ を (i,μ) 要素とする $n\times m$ の行列であり，上添字 T は転置を表す．パラメータは $c=40$ とした以外，前節と同じであるが，刺激レベル z_0 は外部からコントロールできるものとする．

さて，刺激レベルを少し高くしておいて，二つの記憶パターン S^1，S^2 の和を想起入力として与えると，系の状態は S^1 や S^2 をコードするアトラクタ(それぞれ A^1，A^2 とする)とは別の平衡状態 $A^{1,2}$ に達する．まず，このときの x_i の分布を図6に示す．黒い部分は S^1 と S^2 の両方をコードするユニット(10個)を表しており，灰色はいずれか一方をコードするもの(190個)を表す．

この図からわかるように，二つのパターンの片方をコードするユニットが大きな出力を出すのに対し，両方をコードするユニットは比較的小さな出力しか出さない．これは，前述のように，他のユニットから受ける入力が強すぎると出力が小さくなる特性があるからである．実際，これらのユニットでは抑制性細胞の出力 y_i が大きい(図7)．

このような平衡状態は，安定で比較的強いアトラクタになっている．しかし，刺激レベルを下げると不安定になり，ほとんどの場合 A^1 か A^2 のどちらかに状態が遷移する．この際，パターン S^1 ないし S^2 が外部から入力されていると，それが非常に弱いものであっても，入力したパターンに対応するアトラクタに引き込まれる．このことから，状態 $A^{1,2}$ は二つのパターンを

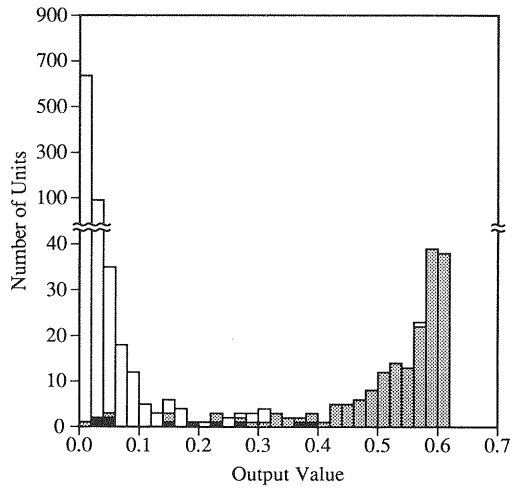


図6 二つのパターンを入力した際の x_i の分布
Fig. 6 Distribution of x_i for $P=S^1+S^2$.

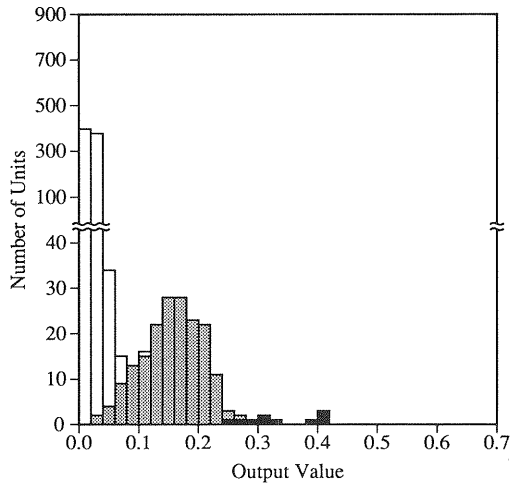


図7 抑制細胞の出力分布
Fig. 7 Distribution of y_i.

コードするものだと言うことができよう．

図8は，状態遷移の過程を具体的に示したものである．この例では，まず $z_0=0.1$ で $t=0$ から $t=2\tau$ まで S^1+S^2 を入力し($k=0.05$)，次いで $t=8\tau$ で刺激レベルを $z_0=0$ に下げると同時に $t=10\tau$ まで S^1 を弱く($k=0.005$)入力した．グラフの(a)と(b)は S^1 と S^2 の一方だけをコードするユニット，(c)は両方をコードするユニットを表し，それぞれの平均出力値がプロットさ

† このように W を選ぶのが必ずしも最良というわけではない[14]．また，m/n ないし l/n が小さければ，式(10)のままでも複数のパターンを保持できる．

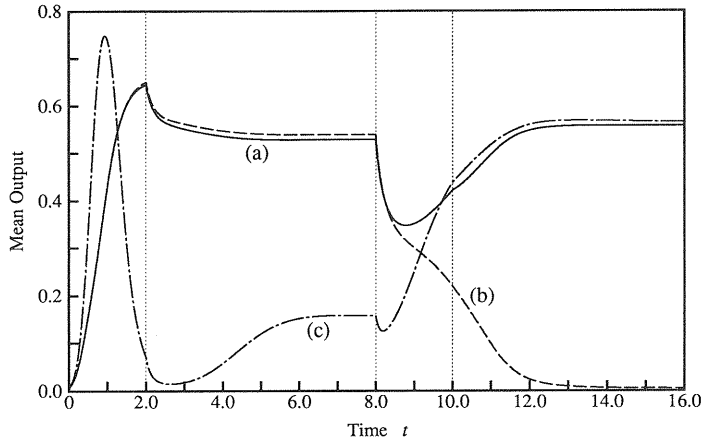


図8　各ユニットの出力値の時間変化
Fig. 8　Time course of change in the mean output of the units coding (a) only S^1, (b) only S^2, (c) both S^1 and S^2.

れている．

まず最初，(c)が急速に増大するが，(a)と(b)が大きくなるとすぐに減少する ($0<t<2\tau$)．外部入力がなくなると，(a)，(b)がやや小さく，それに応じて(c)が多少大きくなり，状態 $A^{1,2}$ に落ち着く ($2\tau<t<8\tau$)．刺激レベルが低くなると，(a)と(b)が小さくなって(c)がすばやく増加するが，このとき(a)と(b)との間に一種の競合が生じる．そして，外部入力の効果により(a)が(b)よりも大きくなると，最終的に状態 A^1 に達する ($t>8\tau$)．このように，特に両方のパターンをコードするユニットの振舞いには興味深いものがある．

ところで，以上のことをサルでの実験に対応させるならば，二つの図形のどちらにも強く反応するニューロンは，両方の図形を同時に提示したとき，非常に弱い反応しか示さないことになる．これは，このモデルが予言する最も興味深い現象の一つである†．

同様に，3個のパターンを同時に入力すると，そのうちの一つのみをコードするユニットが相対的に大きな出力を出し，すべてをコードするものはほとんど出力を出さない．但し，上述の実験よりも m/n か l/n を小さくしないと，そのような状態は平衡状態にはならない．多くの場合，より少ないパターンをコードする状態に変化してしまう．また，同時に入力するパターンの数が増えると，この系だけですべてを保持することは非常に困難になる．しかし，適当な間隔で想起入力を順次与えてやれば(心理学で言うところの短期記憶のリハーサルに相当する)，それらの情報を保ち続けることも可能である．

6. むすび

TE野のニューロン群と連想記憶の力学に関する考察から，非単調な入出力特性をもつユニットを考える必要性を論じ，フィードフォワード型の抑制回路を含む神経回路モデルを構成した．また，それが側頭葉短期記憶系と同様な力学的性質をもつことを示した．

ここで主張しておきたい点として，このモデルと本質的に異なる神経回路網ではTE野に見られるような力学系を実現できない，ということがある．この点に関する理論的な裏付けはまだ不十分だが，このような必然性はモデルから単なる可能性や示唆以上のもの(予想や予言)を引き出すのに不可欠である．

本モデルは，1種類の細胞からなる神経回路網の荷重行列 W に，ある非常に特殊な構造を導入したものとみなすこともできる．しかし，ユニットの概念は，系の動作を理解しやすくするだけでなく，荷重の学習を考える際に重要な意味をもつ (w_{ij}^+ と w_{ij}^- の間に高い相関があることなどから，抑制細胞 C_i^- の学習には，出力細胞 C_i^+ の出力 x_i が教師信号として関与していなければならないことがわかる．なお，このことは，おそらく大脳皮質のフィードフォワード型抑制回路の可塑性にも当てはまる)．

† 宮下の最近の実験によれば，TE野のニューロンは実際にこのような振舞いを示すと言う．従来のモデルを用いても複数のパターンの保持は実現できるが，こうした奇妙な現象は決して生じない．従って，この知見は本モデルの妥当性を示す決定的とも言える証拠である．

そのほか，本研究を通して数々の有益な示唆が得られるが，そのいくつかを挙げておく．

（1） 細胞数 n を一定にしたとき，記憶パターンがスパースになる（l/n が小さい）ほど，記憶容量が大きくなり，同時に保持できるパターンの数も増える．しかし，連想能力（アトラクタの引込み半径）は逆に小さくなる．両者のバランスを考えたとき，TE 野で用いられているような適度にスパースなコーディングは，非常に優れた方法だと言える．

（2） 大脳皮質の神経回路の構造はどこでもほぼ同じであり，TE 野だけが特殊なわけではない．従って，脳のその他の領域もこのモデルによって実現される力学系の一つになっている可能性が高い．領域によってニューロンの振舞いが異なるのは，情報表現の違いや結合のパラメータの違いが原因のように思われる．実際，**5.1** のモデルでパラメータを少し変える（主に刺激レベルを下げる）と，認識系のモデルとして興味深い動作をする．

（3） ある神経回路網を一つの力学系として見るとき，系全体の状態を考えずに個々の細胞を取り上げて論じることはあまり意味がない．本モデルのように，構成要素が非単調特性をもつ場合は特にそうである．そして，記憶系や高次の認識系の神経回路に対しては，そうした力学系としての見方が不可欠だと思われる．このことは，単一ニューロンの反応の測定をもとにして研究を進める際，十分に注意しなければならない．

このように，本モデルはかなり一般的な意味をもつものであり，より複雑なシステムを構成する際のモジュール[7]としても，重要な役割を果たすと考えられる．

その一方で，今後に残された課題も多い．まず第一に，力学系に保持されている短期記憶情報をどのように利用するのか（サルの場合で言えば，16 秒後に再提示された図形が最初の図形と一致しているか否かをどうやって判断しているのか），という問題がある．これに答えるためには，本モデルや TE 野の力学的性質を更に詳しく調べると共に，脳のその他の部位との関連を考慮する必要があろう．

もう一つの重要な問題は，ここで述べた力学系がどのようにして自己組織化するか，ということである．これは，どのようにして長期記憶が形成されるかという非常に大きな問題の一つだとも言える．

TE 野の場合，特に興味がもたれるのは「新奇な図形」がいかにして「見慣れた図形」へと変わっていくのかという点である．これに関連して，提示した順番が近い図形ほど似たパターンにコードされる傾向にある[6]という興味深い知見が得られているが，このことは，学習の過程で各図形をコードするアトラクタの性質が変わるだけでなく，アトラクタ間の関係（情報表現）も変化することを示唆する．また，数々の知見から，この過程には TE 野以外の領域，特に海馬が関係すると考えられる[5]が，海馬の機能や機構は明らかでなく，そのモデル[9]もまだ十分なものではない．

このように，学習の過程をモデル化し，更に海馬や側頭連合野を中心とする記憶システムのモデルを構成するには，数多くの問題を解決しなければならない．一方で，その基礎となる生理学的データ，特に記憶系の情報表現と力学的性質に関する知見が極めて不足していることも事実である．生理学的・実験的アプローチと工学的・理論的アプローチとがうまくかみ合い，刺激し合うことが，記憶のメカニズムの解明のために最も大切だと言えよう．

謝辞 日ごろ御指導と御討論を賜る東京大学工学部計数工学科の中野馨助教授および吉澤修治教授に感謝します．また，本研究のきっかけとなる重要な示唆を下さった同学科，甘利俊一教授および中村真理氏（現在，電子技術総合研究所）に感謝します．更に，貴重な御教示を賜った同学医学部の宮下保司教授と京都府立医大の外山敬介教授に感謝の意を表します．

文　献

（1） Hopfield J. J.: "Neural networks and physical systems with emergent collective computational abilities", Proc. Natl. Acad. Sci. USA, **79**, pp. 2554-2558 (1982).

（2） 中野　馨："アソシアトロンとその応用—連想記憶装置に関する研究"，信学インホメーション理論研資，**IT69**-27 (1969).

（3） Amari S. and Maginu K.: "Statistical neurodynamics of associative memory", Neural Networks, **1**, pp. 63-73 (1988).

（4） 森田昌彦，吉澤修治，中野　馨："自己相関連想記憶の想起過程とその改良"，信学論(D-II)，**J73-D-II**, 2, pp. 232-242 (1990-02).

（5） 宮下保司："視覚再認記憶のニューロン機構"，神経進歩，**32**, 4, pp. 553-565 (1988).

（6） Miyashita Y.: "Neuronal correlate of visual associative long-term memory in the primate temporal cortex", Nature, **335**, pp. 817-820 (1988).

（7） 甘利俊一，倉田耕治，赤穂昭太郎："短期および長期記憶の神経回路モデル"，信学技報，**MBE88**-143 (1989).

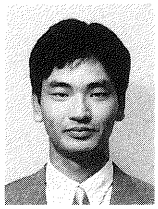
（8） 森田昌彦："連想記憶の海馬モデル"，信学論(D-II)，**J72-D-II**, 2, pp. 279-288 (1989-02).

（9） 中野　馨編："ニューロコンピュータの基礎"，コロナ社 (1990).

(10) Hopfield J. J.: "Neurons with graded response have collective computational properties like those of

two-state neurons", Proc. Natl. Acad. Sci. USA, **81**, pp. 3088-3092 (1984).
(11) Treves A. and Amit D. J.: "Metastable states in asymmetrically diluted Hopfield networks", J. Phys., **A21**, pp. 3155-3169 (1988).
(12) Crisanti A. and Sompolinsky H.: "Dynamics of spin systems with randomly asymmetric bonds: Ising spins and Glauber dynamics", Phys. Rev., **A37**, pp. 4865-4878 (1988).
(13) 浦浜喜一："ニューラルネットの局所安定性"，信学論(D-II)，**J72-D-II**, 9, pp. 1599-1600 (1989-09).
(14) Morita M., Yoshizawa S. and Nakano K.: "Memory of correlated patterns by associative neural networks with improved dynamics", Proc. INNC-90-Paris, pp. 868-871 (1990).
(15) Amari S.: "Characteristics of sparsely encoded associative memory", Neural Networks, **2**, pp. 451-457 (1989).

(平成2年6月26日受付)



森田　昌彦

昭61東大・工・計数卒．昭63同大大学院修士課程了．現在，同大学院博士課程に在学中．日本学術振興会特別研究員．生体情報工学，特に神経回路網の研究および脳の記憶機構の研究に従事．