第 2 章

対連合記憶の神経回路モデル

2.1 まえがき

大脳の下側頭葉皮質 (IT 野) およびその周辺部は，視覚情報の記憶と深く関わっていることが知られているが，そのメカニズムには不明な点が多い．特に，情報がどのように構造化されて長期記憶が形成されるのか，その記憶がどのようにして読み出され行動に利用されるのか，そしてこれらの過程がどのような神経機構によって実現されているのか，という重要な問題が未解明である．これを明らかにすることは，下側頭葉の記憶メカニズムだけでなく，脳における記憶の原理，更には脳の様々な高次機能の解明につながると思われる．

この問題を解く上で大きな手がかりとなるのが，Sakai と Miyashita [3] による生理実験である．彼らは遅延対連合課題を遂行中のサルの IT 野を調べ，pair-coding および pair-recall と呼ばれる興味深いニューロン活動を報告した．これらは，二つの記憶情報が連合されているとき，それがどのようにコードされ連想されるかについて重要な示唆を与える．しかしながら，このような機能を実現している神経機構はほとんどわかっていないし，実験データを説明する神経回路モデルも存在しない．

IT 野の記憶関連ニューロン群の活動分布は，Hopfield モデル [6] のような単純な連想神経回路では説明が困難であることが知られている [7]．これを説明するために改良したモデルがいくつか提案されている [7, 8, 9] が，記憶を点アトラクタに埋め込むという点には変わりがない．そのため，pair-recall ニューロンのように連想の過程で徐々に変化する活動は，これらのモデルによってもうまく説明されないのである．

本章では，計算論的な考察に基づいてこれらの問題を解決し，対連合記憶を形成し記憶課題を遂行する神経回路モデルを構築する．また，モデルの挙動と生理学的知見とを比較検討することによって，下側頭葉における記憶のメカニズムについて考察する．

2.2 モデルの背景

2.2.1 下側頭葉の対連合記憶関連ニューロン

まず，Sakai と Miyashita [3] が行った実験について簡単に説明する．

彼らは，コンピュータで生成した 24 個の図形を適当に組み合わせて 12 組の図形対を作成し，これを用いてサルに遅延対連合課題を行わせた．これは，図形対の一方を cue として短時間提示し，数秒の遅延期間の後に提示した図形が cue の対図形 (target) であるかどうかを判断させるというものである．課題の正答率が十分高くなるまで訓練した段階で，微小電極を用いて IT 野のニューロン活動が測定された．

この実験の主な結果は，以下のようにまとめられる．このうち 1. は図形を独立に記憶させた場合 [10, 11] と共通する性質であるが，2. と 3. は遅延対連合課題に特有である．

1. 図形に対する反応選択性のあるニューロンの多くは，複数の全く異なる図形に反応する．ただし，強い反応を引き起こす図形はその中の少数である．

2. 対となる二つの図形への反応の相関が全体的に高く，図形対の両方に反応するニューロン (pair-coding ニューロン) が比較的高い比率で観察される．これらの多くは，cue と target に対して強く反応するだけでなく，遅延期間中も持続的に活動する．

3. 最も強い反応を引き起こす図形 (最適刺激) の対図形を cue として提示したとき，cue に対してはほとんど反応を示さないのに，遅延期間中に徐々に活動を高めていくニューロン (pair-recall ニューロン) が見られる．これらは，逆に最適刺激を cue として提示すると，遅延期間中に活動が低下していく．

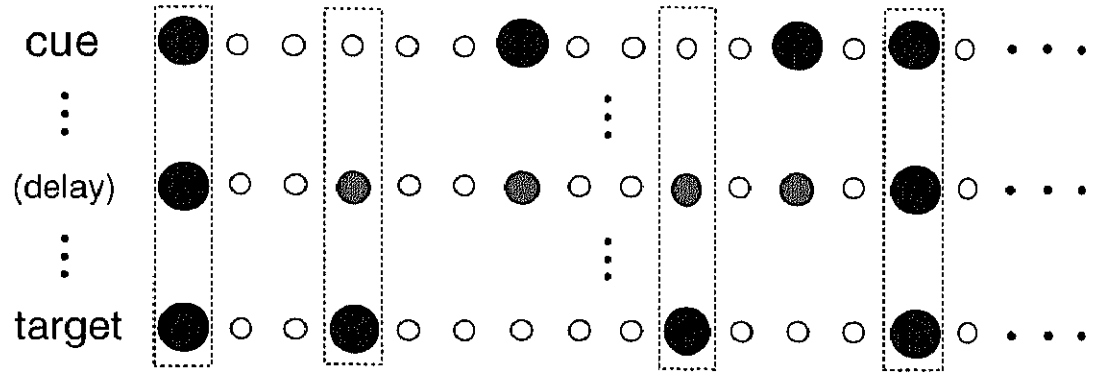


図 2.1: 想起過程における興奮パターンの推移

ここで，pair-coding および pair-recall ニューロンという呼び名から，これがニューロンの固定的な種別を表すと考えるのは，適当でない．例えば，ある図形対に限ると pair-coding ニューロンとして振る舞うニューロンが，別の図形対を提示すると pair-recall ニューロンと同様な活動を示すこともあり得るし，実際そのようなニューロンも文献 [3] 中に見受けられる．以下ではこの点を考慮し，「cue 入力時にはほとんどないが遅延期間中に増大し，target 入力時に最大となる活動」のことを，target が最適刺激の場合に限らずすべて pair-recall ニューロン活動と呼ぶことにする．

2.2.2 実験結果の解釈と問題点

上記の実験結果は，システム論的な立場から見たとき，次のように解釈するのが妥当であろう (図 2.1)．

1. 各図形は IT 野においてあるニューロン群の興奮パターンによって表現される．このパターン (図形のコード) は，興奮しているニューロンの割合が低いスパースなパターンであり，図形の部分的特徴とはほとんど無関係である．

2. 連合された二つの図形のコードは，興奮しているニューロンの一部が重複する．これが pair-coding ニューロンに対応する．

3. cue を提示した後の遅延期間中，ニューロン群の活動は cue をコードする状態から target をコードする状態へ徐々に変化していく．この過程において，いくつかのニューロンが pair-recall ニューロン活動を示す．

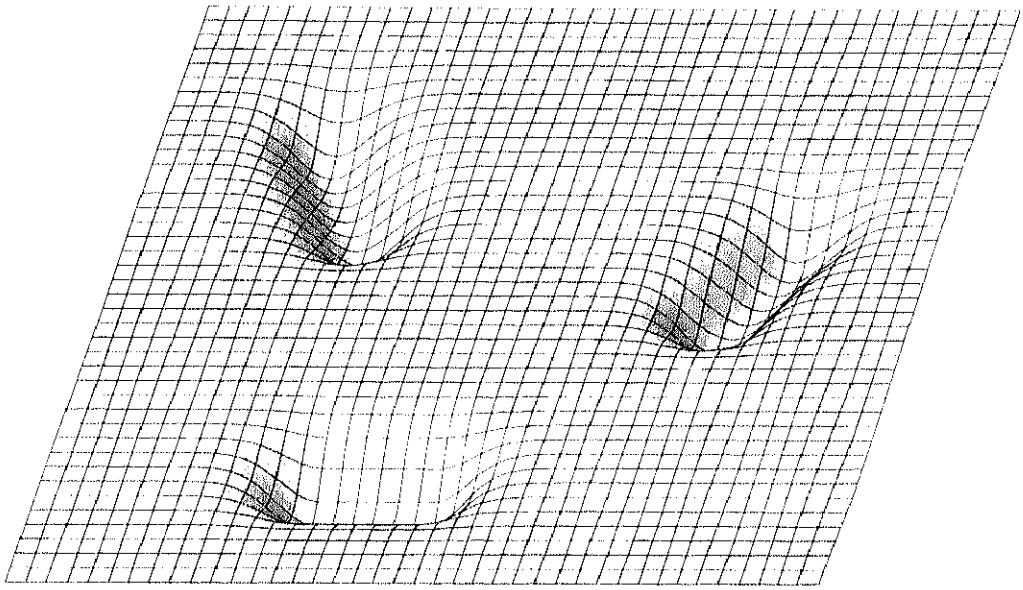


図 2.2: 対連合が形成された記憶回路網の仮想的なエネルギー地形

このようなニューロン群の活動，特に遅延期間中の興奮パターンの変化は，図 2.2 に示すような力学系が構成されていることを示唆する．この図は，系の各状態 (ニューロンの興奮パターン) の安定性を模式的に表現したもので，x–y 平面は状態空間 (興奮パターンの集合)，z 軸は仮想的なポテンシャルエネルギーを表している．図には 3 本の溝が描かれているが，それぞれが一つの対連合記憶に対応し，溝の両端が図形をコードする状態である．このように，cue や target をコードする状態だけでなく，その間を結ぶ経路全体が周辺の状態よりエネルギーが低いこと，またその経路沿いのエネルギー地形が滑らかであることが，連続的な状態遷移を安定に行うのに必要だと考えられる．

ところが，通常の神経回路モデルでは，一般にある状態を強いアトラクタにして安定性を高めようとすると，その周辺のエネルギー地形は凹凸が激しいものになる [12]．そのため，連続的な状態遷移を可能にするためには，IT 野のような分散表現ではなく局所表現を用いるか，記憶するパターンの数をごく少数に制限しなければならない．したがって，目的とする力学系を従来の神経回路モデルで構成するのは非常に困難である．

この問題を解消するためには，各細胞の入出力特性をある種の非単調増加関数 (図

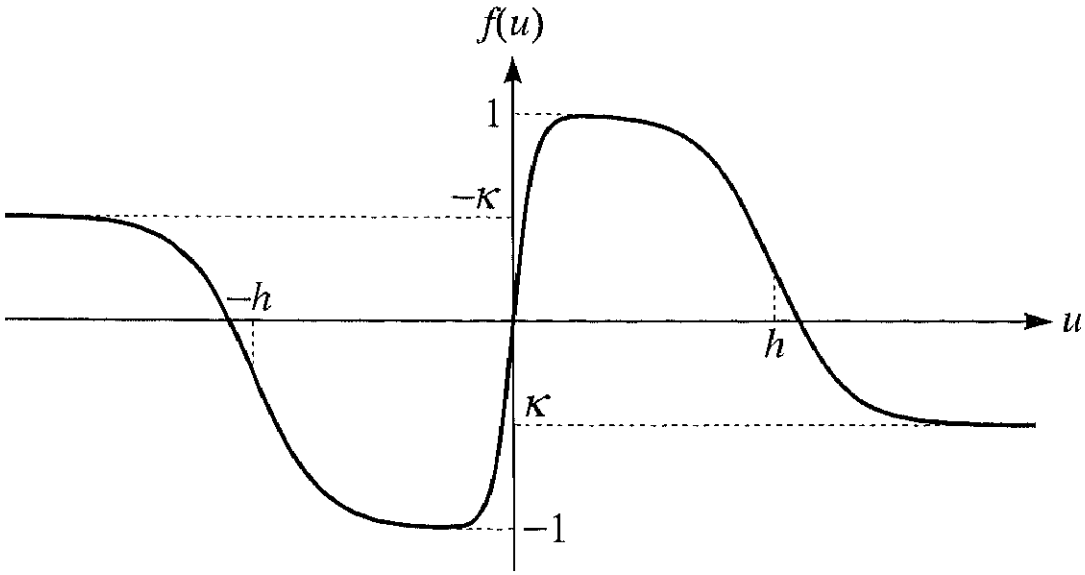


図 2.3: 非単調増加関数

2.3) にすればよいことが知られている [13, 14]. 実際, そのような細胞で構成された回路網 (非単調神経回路モデル) では, 軌道状のアトラクタに沿って連続的に状態遷移することが可能である [12, 15]. ただし, このモデルは生物学的な妥当性に欠け, そのままではIT野のニューロン活動のモデル化に適さない.

そこで, 本研究では, 以下に述べる feedforward 抑制型神経回路モデル [7, 16, 17] を用いる. これは, 単調な特性の細胞の組み合わせによって非単調モデルと同様なダイナミクスを実現する最も単純なモデルである. また, 回路網の活動度 (出力の総和) を低く一定に保つ働きがあるので, IT野に見られるようなスパースな興奮パターンを扱うのにも適している.

2.3 軌道アトラクタの形成と対パターンの想起

本節では, まず連続的に変化する学習信号が外部から与えられる理想的な状況を仮定した上で, feedforward 抑制型モデルによって図 2.2 のような力学系を形成しIT野の対連合ニューロン活動を説明することが可能かどうか検証する.

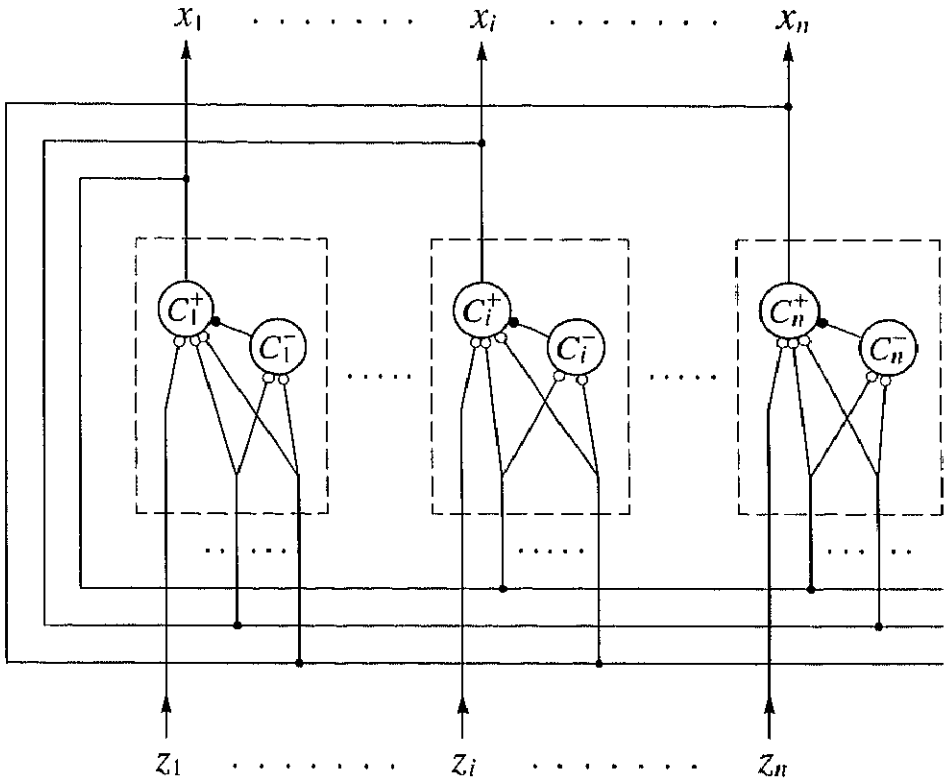


図 2.4: 記憶回路の構造

2.3.1 回路網の構造

このモデルは，興奮性細胞 C_i^+ と抑制性細胞 C_i^- からなるユニットが相互に結合した構造をしている (図 2.4)．C_i^+ は系の外部からの入力信号 z_i と他のユニットからの再帰的な入力を受けて x_i を出力し，それがユニットの出力となる．C_i^- は他のユニットからの入力に応じて y_i を出力し，C_i^+ を feedforward に強く抑制する働きがある．数式で示すと，

$$y_i = f\left(\sum_{j=1}^{n} w_{ij}^- x_j - \theta\right) \tag{2.1}$$

$$\tau\frac{du_i}{dt} = -u_i + \sum_{j=1}^{n} w_{ij}^+ x_j - w_i^* y_i + z_i \tag{2.2}$$

$$x_i = f(u_i) \tag{2.3}$$

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^- へのシナプス荷重，w_i^* は C_i^- から C_i^+ への抑制性シナプスの効率，u_i は C_i^+ の平均膜電位を表す．τ と θ は正の定数である．

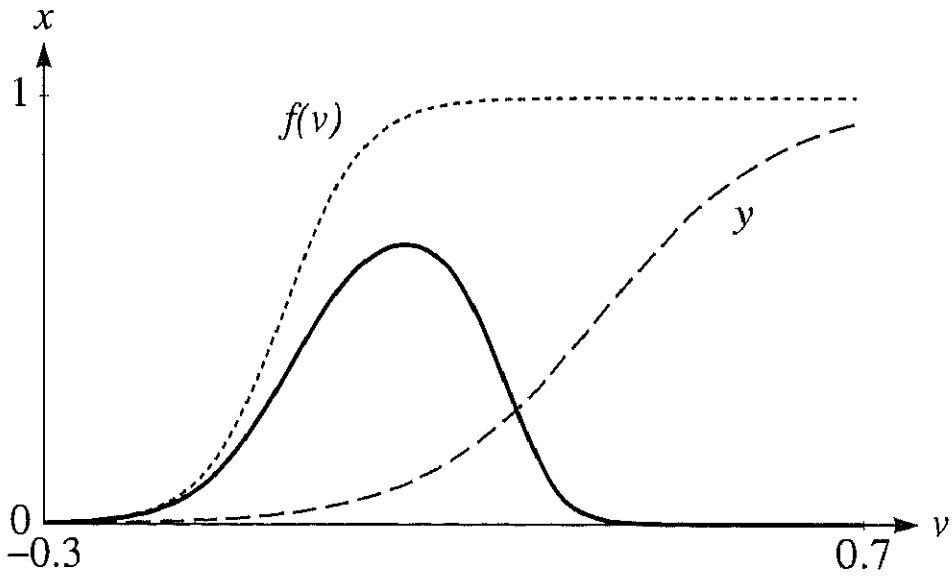


図 2.5: 各ユニットの入出力特性

各細胞の出力関数 $f(u)$ は，0 から 1 の値をとる単調なシグモイド関数

$$f(u) = \frac{1}{1+e^{-cu}} \tag{2.4}$$

である (c は正の定数)．しかし，ユニットの入出力特性は，適当な条件下において非単調となる (図 2.5)．すなわち，ユニットへの入力が小さいときには出力 x_i は入力とともに増加するが，入力がある値以上に大きくなると，抑制性細胞の出力 y_i の増大により x_i は減少する．

2.3.2 学習アルゴリズム

この回路網の学習 (図 2.6) は，成分が 0 または 1 をとる 2 値ベクトル $\boldsymbol{r} = (r_1, \ldots, r_n)$ を学習信号として与えることによって行われる [17, 18]．具体的には，$z_i = \lambda r_i$ の形で $\boldsymbol{r}$ を入力するとともに (λ は入力強度)，以下の式に従ってシナプス荷重を修正する．

$$\tau' \frac{dw_{ij}^{+}}{dt} = -w_{ij}^{+} + \alpha r_i x_j \tag{2.5}$$

$$\tau' \frac{dw_{ij}^{-}}{dt} = -w_{ij}^{-} - \beta_1 r_i x_j + \beta_2 x_i x_j + \gamma \tag{2.6}$$

ここで，τ' は τ に比べて十分に大きい時定数，α，β_1，β_2 は正の学習係数で $\beta_1 < \beta_2$ を満たす．γ は正の定数で，ユニット間の一様な側抑制を表す．なお，学習係数 α を

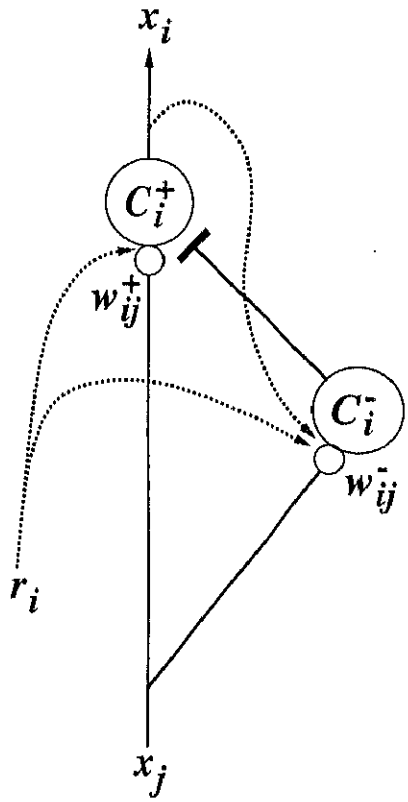


図 2.6: 各シナプスの学習則

x_i の減少関数にすることによって学習性能が高まるため，ここでは

$$\alpha = \begin{cases} \alpha'(\kappa - x_i) & (x_i < \kappa) \\ 0 & (x_i \geq \kappa) \end{cases} \tag{2.7}$$

とする ($\kappa \equiv \beta_1/\beta_2$，$\alpha'$ は正定数)．

$r_i = 1$ すなわち i 番目のユニットに学習信号が入力されたとき，出力 x_i が κ よりも小さいならば，x_j に応じてシナプス荷重 w_{ij}^+ は強化され，w_{ij}^- は抑圧される．その結果，このユニットの出力 x_i は増加するが，x_i が κ より大きくなると w_{ij}^- が強化されるので，x_i のそれ以上の増大は抑えられる．また，$r_i = 0$ のときには w_{ij}^- だけが強化され，x_i は減少する．

学習の過程を摸式的に示したのが図 2.7 である．図中の小球は回路網の現在状態 $\boldsymbol{x}$，矢印は現在の学習信号 $\boldsymbol{r}$ を表す．直観的に言うならば，この学習によって $\boldsymbol{r}$ が指示する状態およびその周囲のエネルギーが低くなる．したがって，もし $\boldsymbol{r}$ が時間的に一定ならば，一つの点アトラクタが形成される (a)．しかし，$\boldsymbol{r}$ がゆっくりと連続的に変化

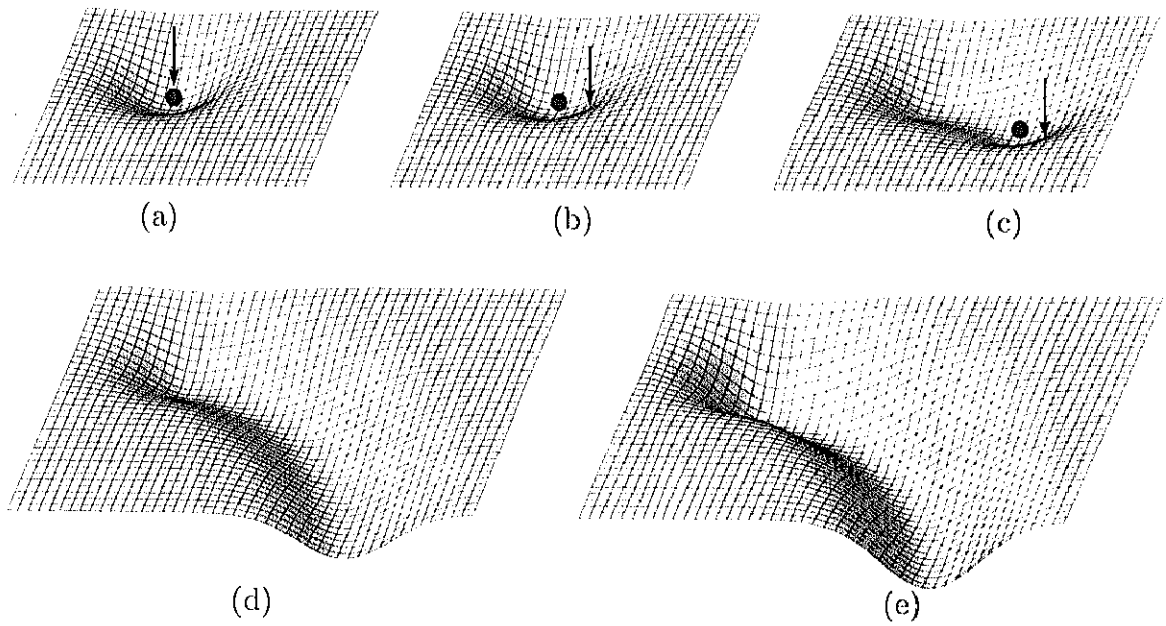


図 2.7: 学習過程の模式図

したならば，xは少し遅れてそれに追従するため，その軌跡に沿ってエネルギー地形に溝が刻まれる (b ~ d)．また，xからrの方向，すなわちrの進行方向に緩やかな流れが生じる．このようなその軌跡に沿ってエネルギーの溝が刻まれるとともに，溝の底にrの進行方向と同じ向きの緩やかな流れが作られる．このような学習をrの入力強度λを減らしながら数回繰り返すと，rの経路にほぼ沿う形で軌道アトラクタが形成される (e)．その結果，適当な初期状態を与えるだけで，学習した軌道に沿って回路網の状態が自動的に推移するようになる．

2.3.3 計算機シミュレーション

このモデルにcueパターンからtargetパターンを連想させる実験を，ユニット数$n = 1000$の回路網を用いて行った．

まず，cueおよびtargetパターンは成分の10%が1で残りが0であるスパースな1000次元ベクトルとし，それぞれ20個ずつランダムに作成した．次に，各cueパターンから対応するtargetパターンへ連続的に (一度に1bitずつ) 変化する時系列パターンを人工的に作成し，それを学習信号として上記の学習を行った．学習回数は10回，モ

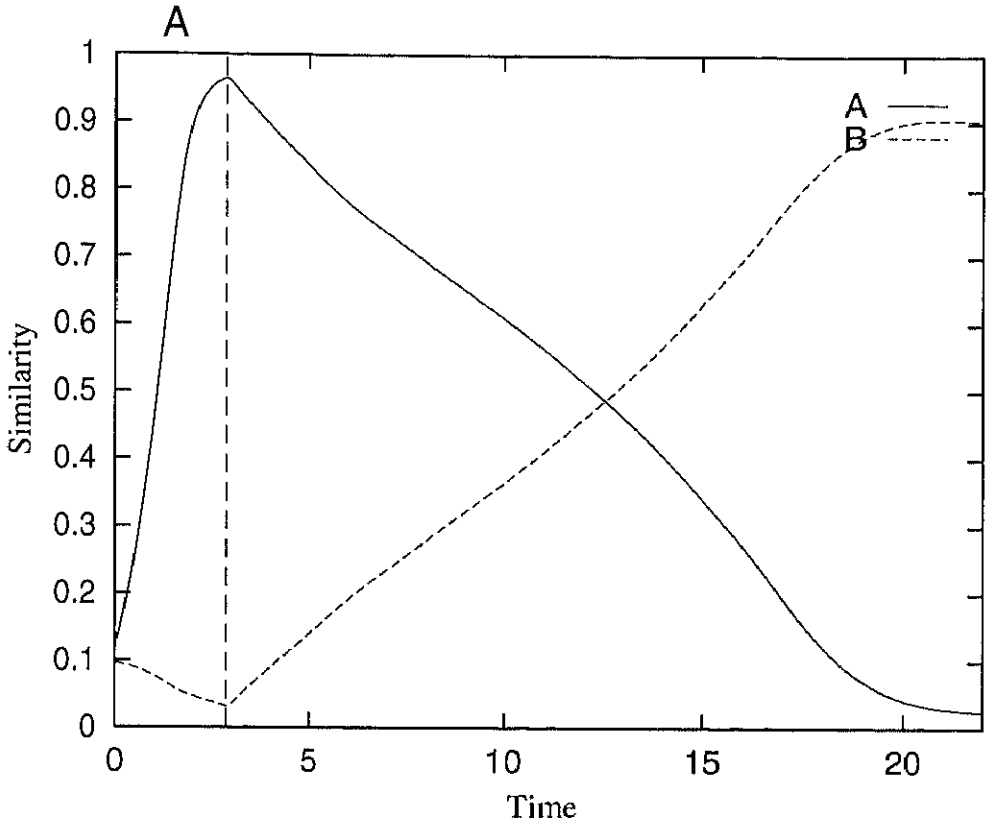


図 2.8: 想起の過程

デルのパラメータは以下のとおりである.

$$c = 10,\ \tau' = 50000\tau,\ \theta = 3,\ w_i^* = 10,\ \lambda = 0.3,$$

$$\alpha' = 50,\ \beta_1 = 25,\ \beta_2 = 50,\ \gamma = 0.05$$

学習後のモデルの挙動を図2.8に示す. これは, あるcueパターンAを短時間だけ入力し, その後何も入力しなかったときの回路網の状態変化を表したものである. 図の横軸は時定数τを単位とする時間, 縦軸は出力パターン$\boldsymbol{x} = (x_1, \ldots, x_n)$とAまたはtargetパターンBとの類似度である. ここで, $\boldsymbol{x}$とパターン$\boldsymbol{s} = (s_1, \ldots, s_n)$との類似度を$\sum_i x_i s_i / \sum_i x_i$と定義している.

この図から, Aの入力が$t = 3\tau$で打ち切られた後, $\boldsymbol{x}$とBとの類似度は, Aとの類似度が下がるとともに増加していき, 約0.9に達していることがわかる. このことは, 回路網の状態がAからB付近まで連続的に遷移したことを意味する.

同様に, 他のどのcueパターンを入力しても, 対応するtargetパターンが想起され

ることが確かめられた．また，このときの各ユニットの出力を観察すると，前述のIT野ニューロンの活動に似た時間変化を示すものが多数見られた(詳細は**2.4.4**参照)．

2.4 対連合記憶形成のモデル

前節の結果は，feedforward抑制型モデルによってIT野における対連合記憶の想起過程が説明できることを示唆している．しかしながら，実際の生理実験と比較したとき，このモデルには以下のような問題点がある．

1. 前述の遅延対連合課題では，図形対のどちらがcueになるかは試行ごとにランダムに選ばれる．つまり，サルは対図形を双方向に連想しなくてはならない．これに対して，前節のモデルでは一方向の連想しか扱っていない．仮にcueパターンとtargetパターンを入れ替えた学習を追加しても，双方向の学習信号の経路(A→BとB→A)が重なり，干渉によって学習がうまくできない．

2. 課題においてサルに提示されるのは，cueおよびtarget図形であって，cueからtargetにゆっくり連続的に変化する時系列パターンが与えられるわけではない．かといってAからBへ不連続もしくは急激に変化するパターンをモデルの学習信号として用いると，記憶が全くできないか，複数の独立したパターンが記憶されるだけでcueからtargetへの連想はできない．

3. 実際の課題では，遅延期間後に再提示した図形がtargetか否かを識別することが求められる一方で，targetの完全な再生は必ずしも必要でない．前節のモデルはtargetパターンを想起しているが，それが入力パターンと一致するかどうかの判断は行っていない．単純に考えると，両者を成分ごとに比較する回路を別に用意すればよいが，それが脳のモデルとして妥当かどうか疑問である．

本節では，以上の問題点を解決し，記憶形成の過程を含めたモデル化を行う．

2.4.1 モデルの再構成

上記の問題点1.は，図2.9に模式的に示すように，双方向の学習信号の経路がある程度離れるようにすれば解消できる[19]．すなわち，二つのパターンaとbをまっすぐ

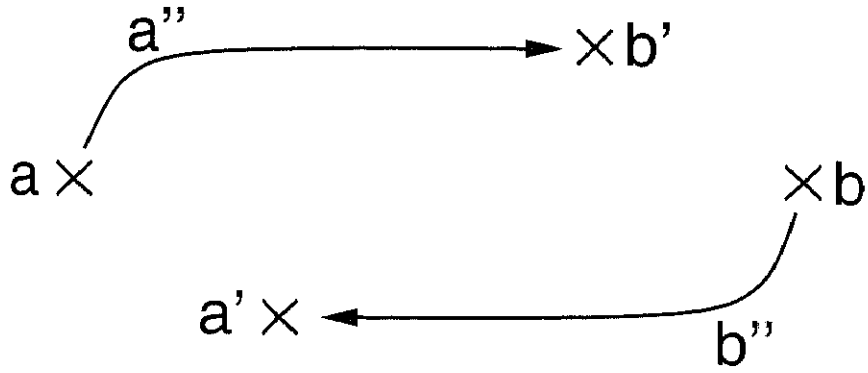


図 2.9: 学習信号の経路

結ぶ経路ではなく，そこから少し離れたところにある別のパターン a' および b' へ向かう経路を用いればよい．

ただし，こうすると学習後に a を cue として与えたとき b とは多少異なるパターンが想起される．しかし，問題点 3. で触れたように，target の識別さえできるならば，b を完全に想起する必要はないし，学習信号の始点 (a や b) が cue および target パターン (A や B) と同じである必要もない．実際，**2.4.4** でも示すように，target パターンを再生することなく識別を行うことが可能である [19].

したがって，cue パターンと target パターンが逐次的に入力されたとき，モデルの内部において図 2.9 のような連続的な学習信号を生成することができれば，上記の問題点はすべて解決されることになる．では，そのためにはモデルをどのように修正すればよいであろうか．

まず考えられるのは，前節の記憶回路 (N_1 とする) とは別の神経回路 N_2 を設け，そこで入力パターン $\boldsymbol{s}$ を学習信号 $\boldsymbol{r}$ に変換することである．しかし，実際に試みたところ，そのような回路を単独で構成するのは非常に難しいことがわかった．これは，cue パターンに関する情報を保持することと，target パターンの入力によって出力パターンを望ましい状態まで連続的に変化させることを，一つの回路網で両立させるのには無理があるからである．また，$\boldsymbol{r}$ には回路網 N_1 の状態遷移を先導する役割があるにもかかわらず，N_2 の出力は N_1 の状態とは無関係に変化する点も問題であった．

そこで，図 2.10 に示すように，N_1 の出力を N_2 にフィードバックしたところ，比較

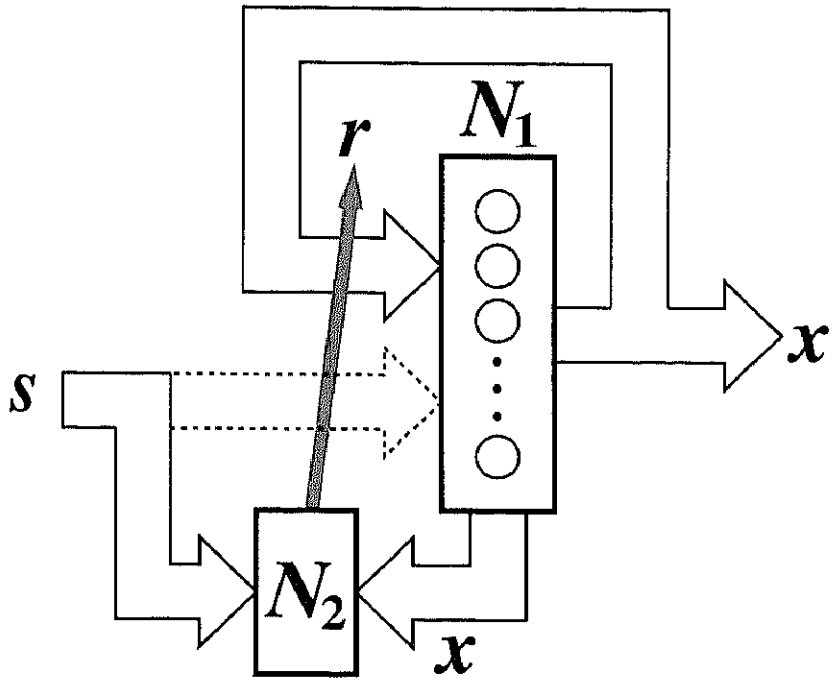


図 2.10: モデルのブロック図

的容易に目的の学習信号が生成され，N_1 の学習もうまくいくことがわかった [20]．回路の単純化をできる限り進め，最終的に構成したのが以下のモデルである．

2.4.2 学習信号生成回路の構造

学習信号生成回路 (回路網 N_2) の構造を図 2.11 に示す．この回路は回路網 N_1 のユニット数と同じ n 個の細胞からなる．i 番目の細胞 C_i は，入力パターン $s = (s_1, \ldots, s_n)$ をシナプス荷重 p_{ij} を介して受け，N_1 の i 番目のユニットへの学習信号 r_i を出力する．

この回路は基本的にランダム変換回路であり，入力シナプスの荷重 p_{ij} はランダムな値をとる．また，C_i は N_1 の全てのユニットからフィードバックを受けるが，そのシナプス荷重 q_{ij} もランダムである．

同時に，この回路は一種の競合系を構成しており，C_i は強度 ρ で他の細胞を側抑制するとともに，強度 σ の自己結合をもつ (ρ, σ は共に正の定数)．これにより，少数の細胞が大きな値を出力してそれ以外は 0 に近い値を出力するから，r はスパースなパ

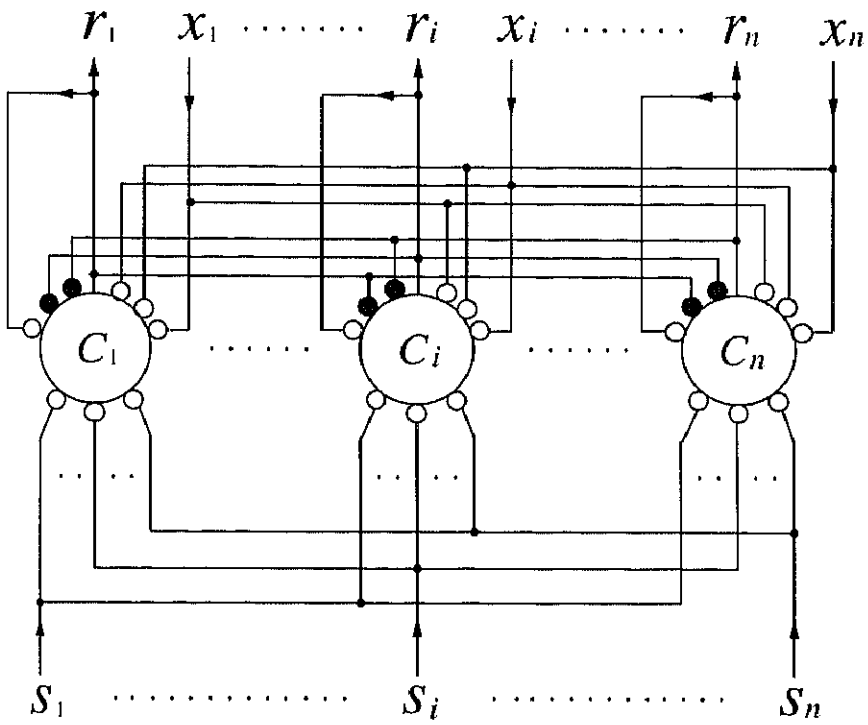


図 2.11: 学習信号生成回路の構造

ターンとなる.

以上を式で示すと,

$$\tau\frac{dv_i}{dt} = -v_i + \sum_{j=1}^{n} p_{ij}s_j + \sum_{j=1}^{n} q_{ij}x_j - \rho\sum_{j\neq i} r_j + \sigma r_i + h \tag{2.8}$$

$$r_i = f(v_i) \tag{2.9}$$

となる. ここで, v_i は C_i の平均膜電位, h はそのオフセットであり, $f(v)$ は式 (2.4) のシグモイド関数である.

2.4.3 モデルの動作

このモデルの学習時の動作を考える上で重要なのは, 回路網 N_1 と N_2 の相互作用である. N_2 の出力 $\boldsymbol{r}$ は N_1 の学習信号であると同時に $z_i = \lambda r_i$ の形で N_1 に入力されるから, 一般に $\boldsymbol{x}$ と $\boldsymbol{r}$ とは 1 に近い類似度をもつ. ただし, $\boldsymbol{r}$ が変化する際には $\boldsymbol{x}$ は $\boldsymbol{r}$ の少し後ろを追従し, 仮に $\boldsymbol{r}$ が不連続的に変化したとしても $\boldsymbol{x}$ はゆっくり連続的に変化する. 一方, N_1 から N_2 へのランダム結合は, $\boldsymbol{x}$ に依存して決まるある方向へ $\boldsymbol{r}$ を

動かそうとする作用をもつ．ただし，N_2 は競合系であり，一度大きな出力を出した細胞はそれを維持しようとするから，N_2 への入力が大きく変化しない限り $\boldsymbol{r}$ はほとんど変化しない．

これを踏まえた上で，すべての細胞の出力がほぼ 0 の静止状態において cue パターン A を入力し，遅延期間の後に target パターン B を入力した場合を考えよう．モデルのパラメータは適切に設定されているとする．

まず，A をある期間入力し $\boldsymbol{s}$ = A の状態をしばらく保つと，回路網 N_2 は A をランダム変換したパターン a を出力し，少し遅れて N_1 も a に近い (類似度が高い) パターンを出力する．N_1 の出力 $\boldsymbol{x}$ は N_2 にフィードバックされるが，A が入力されている間は $\boldsymbol{r}$ はほぼ一定である．

次に，A の入力を終えて $\boldsymbol{s} = 0$ とする (ただし，このとき $\sum_i r_i$ が減少しないようオフセット h の値は若干高くする) と，N_1 からのフィードバックの影響が相対的に強まり，$\boldsymbol{r}$ が動き出す．これが遅延期間中続くが，$\boldsymbol{r}$ の移動速度は次第に低下するので，a との類似度はある程度までしか下がらない．

ここで B を入力すると，$\boldsymbol{r}$ は，静止状態で B を入力したときの出力パターン b の方向へ移動していく．しかし，$\boldsymbol{r}$ は中立的な静止状態ではなく a に近く b とは離れた状態 a'' (図 2.9 参照) から出発するため，b までは到達せず，a'' 寄りの状態 b' で止まる．

同様に，B を cue，A を target として順に入力すると，図 2.9 に示したような b から b'' を経由して a' に至る学習信号が生成される．ここで，もし N_1 からのフィードバックがなければ，$\boldsymbol{r}$ の移動がうまく調節できないだけでなく，双方向の経路が非常に近接してしまうことに注意されたい．

このようにして学習信号を生成しつつ，同時に N_1 に関して **2.3.2** で述べた学習を行うと，図 2.9 の経路に沿って軌道アトラクタが形成される．その結果，N_1 に a を入力すると $\boldsymbol{x}$ は b' に達するが，その状態で b を入力したときの N_1 の反応は，通常よりも強いと考えられる．最初に b を入力し，遅延期間後に a を入力した場合も同様である．

ところで，実際にモデルに与えられるのは A や B などのパターンであるから，課題を実行する際にはこれを a や b に変換してから N_1 に入力する必要がある．そのための一つの方法として，$\boldsymbol{s}$ を直接 N_1 に入力し (図 2.10 の破線)，その入力シナプスを $\boldsymbol{r}$ を

用いた学習により修正することが考えられる．これによって N_1 単独で課題を実行できるようになる (実際にシミュレーションを行い，実行できることは確認している) が，ここではモデルの複雑化を避けるために，課題実行時にも N_2 を用いることにする．すなわち，入力パターン s が与えられたときには N_2 を動作させ，s を変換したパターンを N_1 に送る．ただし，それ以外 ($s = 0$) のときには N_2 は静止状態にあるものとする．

2.4.4 計算機シミュレーション

対連合課題に関するシミュレーション実験を行った [21, 22, 23]．記憶回路網 N_1 および 20 組のパターン対は **2.3.3** のシミュレーションと同じものを用いた．学習信号生成回路 N_2 に関するパラメータは，$\rho = 0.016$, $\sigma = 0.8$ であり，h は学習時の遅延期間中のみ 0.75，それ以外では 0 とした．シナプス荷重 p_{ij} および q_{ij} は，それぞれ平均 0.005・分散 0.05，平均 0.007・分散 0.08 の一様乱数によりランダムに設定した．

学習の手順は次のとおりである．まず，最初のパターン対の一方 (A とする) を 3τ の期間モデルに入力し，7τ の遅延期間の後にもう一方 (B) を 11τ の間入力する．次いでモデルをいったん静止状態にリセットしてから，今度は逆の順序 (B, A の順) で二つのパターンを入力する．再びモデルをリセットして，次のパターン対 (C, D) を 2 通りの順序で入力する．以下同様にすべてのパターン対を順に入力するが，これを 20 回繰り返した．

学習後，モデルに 3τ の間 cue パターンを入力し，17τ の遅延期間の後に test パターンを 3τ の間入力する．これを 1 試行とし，さまざまな cue および test パターンの組み合わせについてテストを行った．

結果の一部を図 2.12 に示す．これは，N_1 のユニットのうち A ～ D のいずれかをコードするものを適当に 20 個選び，それぞれの出力値の時間変化をプロットしたものである．最初の 4 試行は A を，次の 3 試行は B を，最後の 5 試行は C または D をそれぞれ cue とした場合であり，このうち第 2, 5, 10, 12 試行が test パターンが target に一致する match 試行である．

A を cue とした 4 試行を比較するとわかるように，test パターンとして B を入力したとき，それまで大きな出力を出していたユニットが出力を更に増大させるのに対し

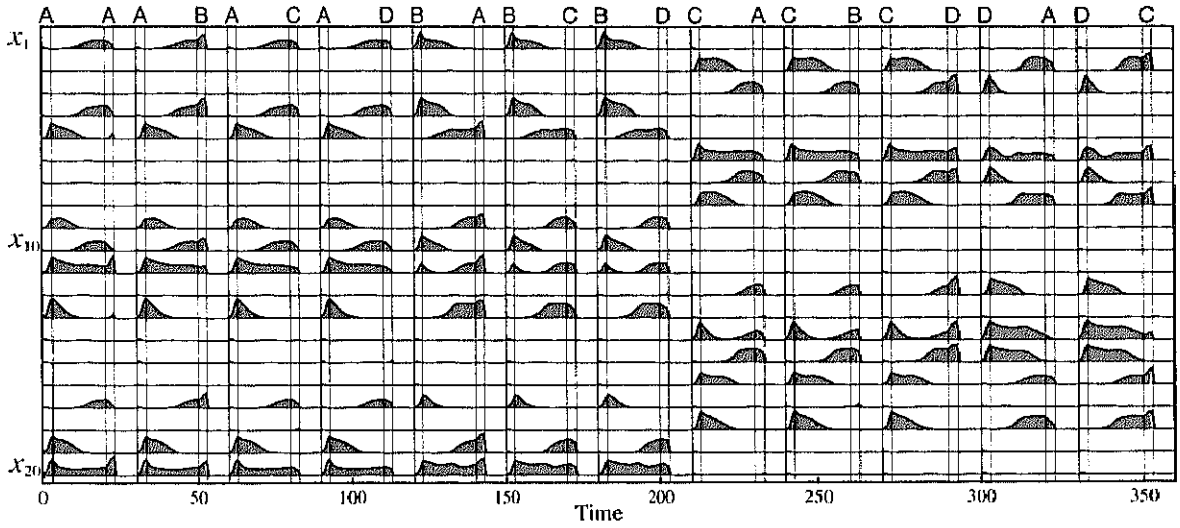


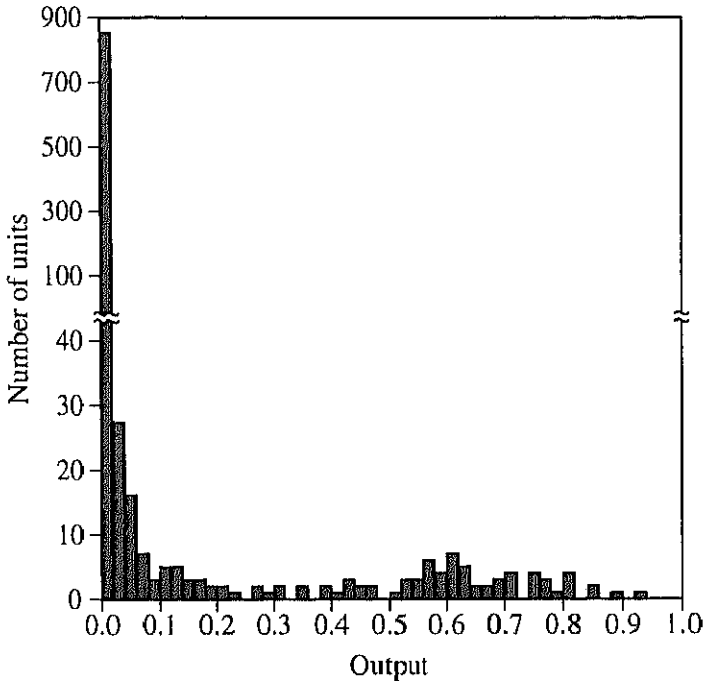
図 2.12: 学習後のモデルの挙動

て，CやDを入力したときにはそれらの多くが出力を低下させている．また，cueパターンAを再入力しても，Bを入力したときほど強い反応は見られない．逆にBをcueとした場合，Aの入力によって最も強い反応が生じている．CやDをcueとした場合についても同様である．

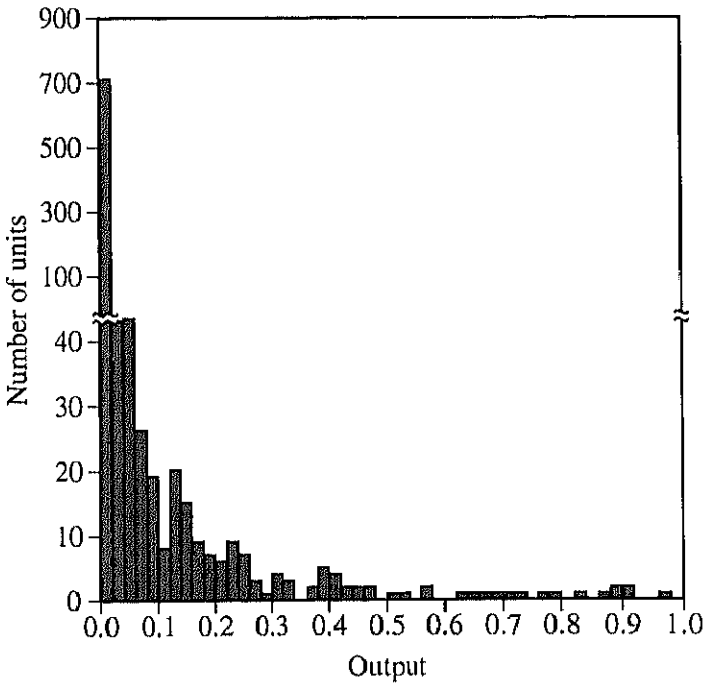
これに関連して，IT野ではmatchとなる刺激提示に対して活動増加が起こるという知見が得られている [24, 25]．これは，認知メカニズムに関連して非常に興味深いが，このモデルによって，そのようなIT野の活動変調を完全に説明できるわけではない．例えば，IT野ニューロンの多くが，match刺激に対してほとんど抑制された反応を示すことや，活動増加や抑制の効果が介在刺激の提示後でさえ持続することを説明できないからである．

図2.13は，testパターンの入力が終わる時点での全ユニットの出力値の分布を示したものである．(a)と(b)は，それぞれtargetとそれ以外のパターンに対する典型的な反応の例である．feedforward抑制型神経回路の性質により出力値の平均にはあまり差がないのであるが，分布には明確な違いがあることがわかる．例えば0.5以上の出力値をもつユニットの数で比較すると，(a)は(b)の約3倍であり，これにより両者を容易に判別できる．

他のすべてのパターン対についてもテストしたところ，0.5以上を出力するユニット



(a) target パターンを入力した場合



(b) target 以外のパターンを入力した場合

図 2.13: ユニットの出力値の分布

は，match 試行において最低でも 45 個あったのに対し，それ以外の試行では常にその半分以下であった．このことから，このモデルは target を正しく識別して課題を遂行する能力を備えていると言える．

更に，個々のユニットに着目すると，IT 野のニューロンと同様な活動が見られることがわかる．例えば，図 2.12 の 20 番目のユニットは，A と B の両方に反応し遅延期間中も比較的大きな出力を持続しているが，これは前述の pair-coding ニューロンの活動と定性的に一致する．また，cue に全く反応しないにもかかわらず遅延期間中に出力を増して target 入力時に強い反応を示すユニットが多数あるが，これは pair-recall ニューロン活動と合致する．これらのユニットは，target パターンを cue にした試行では遅延期間中に活動が低下しているが，この点も pair-recall ニューロンと同じである．このように，このモデルは IT 野で観察されるニューロン活動をかなりよく再現することができる．

2.5 考察

本モデルは，実際の脳の回路構造を直接モデル化したものではないから，モデルと脳の構造が一致するとは限らない．しかしながら，計算論的な要請に基づいて構築されていること，生物学的に無理のない動作原理に従っていること，他のモデルでは説明できなかった IT 野のニューロン活動を再現することなどから，著者らは本モデルと基本的に同じ原理が側頭葉の記憶システムでも用いられていると考えている．これが正しいならば，モデルに必要な機構は脳でも必要であるから，関係する脳の神経機構の少なくとも一部がモデルの構造に反映されているはずである．このような考えのもとに，本節ではモデルと脳との関係について考察する．

最初に，モデルの学習信号生成回路 N_2 に対応する脳の領域はどこか，という問題から考えよう．まず，N_2 は N_1 と双方向に結合しているから，対応する脳の領域は IT 野と双方向に強い線維連絡をもたねばならない．また，N_2 は対連合課題の学習に必要であるが，学習後の課題の実行には必ずしも不可欠ではない．

これらに該当する脳部位は，側頭葉内側部のいわゆる海馬系 [26]，中でも嗅皮質 (嗅内皮質および嗅周皮質) である．この領域は，主に嗅周皮質を介して IT 野と双方向に

密に連絡しており，嗅皮質の損傷によって対連合課題の学習が障害される [27].

更に，半側の嗅皮質を破壊したサルに遅延対連合課題を学習させたところ，破壊した半球の IT 野にも図形選択性を示すニューロンがあったが，対連合記憶関連ニューロンは見られなかった [28, 29]．モデルにおいて N_2 が機能しない場合，cue および target パターンを独立に学習することは可能であるが，両者を連合することはできない．このことからも，N_2 を嗅皮質に対応づけるのが妥当と考えられる．

なお，サルの海馬体 (歯状回，アンモン角，海馬台など) を切除しただけでは，対連合課題の学習能力はあまり低下しない [27]．このことから，今回のモデルの N_2 は海馬体には該当しないと考えられる．しかし，Eichenbaum ら [26] が指摘しているように，嗅皮質と海馬体とがそれぞれ別の形で対連合学習に関与している可能性が高い (前者の破壊により後者と IT 野の連絡も切断される点に注意)．したがって，海馬体の損傷で影響を受けるような課題をこのモデルに適用し，N_2 に加えるべき機能を検討することによって，海馬体の機能の一部をモデル化することが可能かもしれない．

最後に，このモデルによって説明されるその他の知見や予測される現象を列挙する．

1. 対連合課題において，刺激図形の中に類似したものがあるとき，それが cue である場合より target である場合の方が誤答率がはるかに高いことが知られている [30]．また，その際の誤答は大部分が target 以外の test 刺激を target と答えるものである．いま，モデルにおいて A と C が類似し，したがってそれらをコードする N_1 の状態 a と c が近かったとしよう．このとき，A を cue として与えれば N_1 の状態は c に影響されることなく図 2.9 の a から b' に到達するから，特に誤答は生じない．一方，B を cue とし N_1 の状態が a' に達した状態で C が test パターンとして入力されると，a' は c とも比較的近いため，多くのユニットが強く反応してしまう．ここで，a と a' は一致しないため，A をコードするユニットがすべて反応しなくても，ある程度多数のユニットが強く反応した場合には match と判定せざるを得ない．その結果，C を target と誤認識する可能性が生じる．このように，上記の知見はモデルによって説明されるとともに，脳内における a と a' の不一致を示唆する．

2. Murray ら [27] の実験において嗅皮質を切除したサルは，新しい刺激セット (図形対の集合) は何回学習しても正答できないのに対し，切除前に学習した刺激セットに関しては，再学習に要する時間が正常なサルより若干長いものの，正しく答えることができた．**2.4.3** の最後で述べたように，モデルの N_2 は，学習終了後は入力パターンを N_1 でのコードに変換するためだけに働き，しかもその機能は他の回路で代替可能である．したがって，この実験結果も N_2 を嗅皮質と見なすことによって説明できる．

3. 同じ Murray らの実験の中で，cue および target 図形を直接連合させる前に，中間的な図形 (両者を重ねたもの) との連合を訓練すると，対連合学習が大きく促進されることが示されている．モデルにおいても同様な誘導は効果的であるが，更にこれを拡張すると，「図形対を補間して徐々に変化する図形を連続的に提示すれば，嗅皮質を切除したサルでもある程度の学習が可能」という予測が成り立つ．ただし，このような強制的な誘導は，N_2 で生成する学習信号と干渉することにもなるから，正常なサルでは逆効果かもしれない．

4. N_1 から N_2 への信号は，パターン対を双方向に連合する際に特に重要である．したがって，仮に IT 野から嗅皮質へ向かう信号経路だけを切断したならば，常に一方の図形を cue とする単方向の連合に比べて，双方向の連合学習がより強い障害を受けると予測される．

5. **2.4.4** のシミュレーションにおいて，遅延期間中に出力が増加するユニットは多数あるが，増加が始まるタイミングはまちまちである．このことは N_1 の状態遷移の連続性を反映している．したがって IT 野の pair-recall ニューロンにも，cue 提示直後から活動しはじめるものや少ししてから活動を増すもの，かなり遅れて活動を開始するものなどがあると予測される．

6. 図 2.12 の第 2 試行と第 5 試行を比較するとわかるように，cue パターン A から B を連想する場合と B から A を連想する場合で，ユニットの活動が全く逆の時間経過をたどるわけではない．これは，図 2.9 に示したように状態遷移の経路が両方向で異なるからであり，IT 野の対連合記憶関連ニューロンの多くが同様な非対称性を示すと予測される．

7. 例えば図2.12の14番目のユニットは，CとDの両方をコードしているが，Cをcueとして与えても遅延期間中の活動は低下する．これは，N_1の状態がcからdへ直線的に遷移しないために生じる現象であり，IT野のpair-codingニューロンの一部に遅延期間中の持続的な活動を示さないものがあることを予測する．

8. Bをcueとして与えたとき，遅延期間中にAをコードするユニットの全部が出力を増すわけではない．また，大きな出力を出すユニットの一部は，testパターンとしてAを入力すると出力が低下する．これらは，1.でも触れたようにAをコードするN_1の状態aと遅延期間中に到達する状態a'とが一致しないことに起因し，IT野にも同様なニューロン活動があることを予測する．

9. 上記5.～8.はN_1の興奮性細胞に関する性質であるが，抑制性細胞はそれとはかなり異なる活動を示す．まず，入力パターンに応じて活動が大きく変化することはあまりない．また，遅延期間中の活動レベルは比較的低く，変化量も小さい．このためその活動にはあまり意味がないように見えるが，前述のように実は重要な機能を果たしている．IT野のニューロンには図形選択性をほとんど示さないものも多いが，少なくともその一部はモデルの抑制性細胞によって説明可能と思われる．