

自由再生実験における記憶の神経回路モデルについて

山崎 彰太[†] 西野 哲朗^{††}

認識や思考といった高次の脳機能は、記憶をその土台としている。したがって、記憶のメカニズムの解明は非常に重要であるが、その根本的なメカニズムの原理は未だ解明されていない。そこで、本研究では、計算論的アプローチにより神経回路モデルを構成し、そのメカニズムの解明を試みる。本論で提案するモデルは、人間の記憶のモデルとして広く知られている Atkinson と Shiffrin の二貯蔵庫モデルに基づいており、実際の脳内で発見されている抑制性の神経ネットワークを用いて表現できる。Winner-Takes-All(WTA) や Winners-Share-All(WSA) と呼ばれる競合を用いて構成される。また、人間の短期記憶の特徴を顕著に示した Murdock 等の自由再生実験を本モデル上でシミュレーションすることで、提案モデルの妥当性を考察する。

A Neural Circuit Model for Human Short-Term Memory

SHOTA YAMAZAKI[†] and TETSURO NISHINO^{††}

Higher level brain functions, such as recognition and thinking, are based on the human memory. Thus, it is very important to elucidate the memory mechanism, however, the principle of this fundamental mechanism is not known. In this paper, we construct a neural circuit model from computational point of view and try to elucidate the mechanism. Our model is based on the dual storage model of Atkinson and Shiffrin which is widely known as a model of human memory and is constructed by using the competition called Winner-Takes-All(WTA) and Winners-Share-All (WSA) which can be expressed by the inhibitory neural circuit which is actually discovered in the real brain. We also show the validity of our model by carrying out the simulation of the free recall experiment conducted by Murdock which showed the remarkable characteristics of human short-term memory.

1. はじめに

記憶は、短期記憶と長期記憶の大きく二種類に分けることができる。このうち短期記憶に関しては、Murdock¹⁾等によって行われた自由再生実験によって、著しい特徴があることが示されている。本論文は、この自由再生実験における記憶の神経回路モデルを提案する。その際、提案モデルを生物学的に妥当なものとするために、実際に脳の至るところで発見されている抑制性の神経ネットワークを用いて表現可能な、Winner-Takes-All(WTA) や Winners-Share-All(WSA) と呼ばれる競合に着目する。そして、この WTA と WSA を用いて、記憶の二貯蔵庫モデル²⁾に基づく神経回路モデルを提案し、その振る舞いを考察する。

2. 記憶と神経回路

2.1 記憶の二貯蔵庫モデル

短期記憶と長期記憶は、それぞれ異なった場所に存在するという考え方が二貯蔵庫モデルである。そのモデルの概略を図1に示す。外界からの情報は、まず感覚登録器に入り、選択的注意によって選り出されたものだけが、短期貯蔵庫に送られ、そこで一定期間保持される。しかし、短期貯蔵庫の容量は限られているため、新しい情報が入ってくると、古い情報は押し出され、忘れ去られてしまう。そこで、この忘却を防ぐために、情報をリフレッシュさせる、リハーサルと呼ばれる活動が行われる。短期貯蔵庫には、このために一定数のスロットを持つリハーサルバッファが想定されていて、スロットの数と同じか、それ以下の量の情報を繰り返し循環させることで、情報の保持が可能になる。この情報の保持機能に加えて、リハーサルには、情報を次の貯蔵構造である長期貯蔵庫に転送する機能もある。短期貯蔵庫内に滞在している時間が長いほど、その情報が長期貯蔵庫に転送される確率が高くなる。

[†] 電気通信大学電気通信学研究所

Graduate School of Electro-Communications, The University of Electro-Communications

^{††} 電気通信大学情報通信工学科

Department of Information and Communication Engineering, The University of Electro-Communications

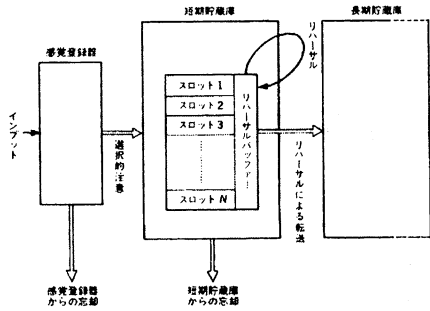


図1 Atkinson と Shiffrin の二貯蔵庫モデルの概略³⁾

2.2 自由再生実験

短期記憶の特徴を示す実験として、記憶項目をひとつずつ被験者に提示した後に、自由な順番で、できるだけ多くの項目を再生させる自由再生実験が Murdock¹⁾等によって行われている。この実験の結果、縦軸を再生率、横軸を項目の系列位置として得られる系列位置曲線と呼ばれるグラフに関して、以下の特徴が示されている(図5参照)。

- 最初に提示した数項目は、高い確率で再生される。これを初頭効果 (primacy effect) と呼ぶ。
- 最後に提示した数項目も、高い確率で再生される。これを親近効果 (recency effect) と呼ぶ。
- 中間の項目の再生率は、系列位置によらず一定の値をとる。
- リストが長くなるにつれて、初頭効果、中間項目の再生率が下がる。

2.3 WTA と WSA

脳は情報処理を行う際、個々の神経細胞が担っている情報を、必要性に応じて選択的活性化・非活性化を行っている。すなわち、神経細胞間に競合が起こっている。この競合を表現するのに、脳の至るところで発見されている抑制性の神経ネットワークを用いることができる。そのネットワークにおける神経細胞の競合ダイナミクスを表す方法として、Lotka-Volterra(LV)型競合システム(図2参照)がある。そのLV方程式は、

$$\tau \dot{z}_i = z_i(\gamma + W_i - k_s z_i - k_l \sum_{j=1, j \neq i}^N z_j) + \epsilon$$

で表される。ここでは、 N は細胞数、 z_i は i 番目の細胞の発火頻度、 W_i は i 番目の細胞への入力、 γ は全細胞にわたる共通の入力、 k_s および k_l はそれぞれ自己抑制、側抑制の強度、 τ は時定数、 ϵ は $z_i \neq 0$ を保障するための正の定数である。

このLV型競合システムの自己抑制強度 k_s と側抑

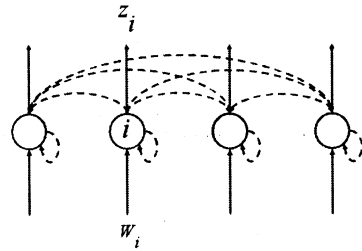


図2 LV型競合システム

制強度 k_l との関係により、以下の関係が成り立つことが知られている⁴⁾。

- (自己抑制強度 k_s) = (側抑制強度 k_l) のとき、WTA(Winner-Takes-ALL)を表現する。
- (自己抑制強度 k_s) > (側抑制強度 k_l) のとき、WSA(Winners-Share-ALL)を表現する。

WTAは、最大の入力を受けている細胞のみを生き残らせるという勝者競合ネットワークである。一方、WSAは、大きな入力を受けている細胞から、順にいくつかの細胞を生き残らせることが可能な多勝者競合ネットワークである。

3. 提案モデル

本論の短期記憶のモデルは、二貯蔵庫モデルを基に構成されており、感覚登録器部、短期貯蔵庫部、長期貯蔵庫部の3つのブロックから構成されている。

3.1 感覚登録器部

Murdockらによって行われた自由再生実験では、最初と最後に Ready 信号を言葉で提示する。そのため、最後に提示された項目が Ready 信号によって上書きされる可能性がある。そこで、全ての項目提示終了後に、ノイズとして10%の確率で Ready 信号を入力する。記憶項目は100%の確率で、短期貯蔵庫部へ転送する。

3.2 短期貯蔵庫部

短期貯蔵庫部は入力層、WTA層、メモリスロットによって構成されている。モデルの回路構成を図3に示す。

WTA層と入力層の間は全結線しており、結合荷重は0~1の実数値としてランダムに設定されている。WTA層の細胞数とメモリスロットの数は同数であり、WTA層の一つの細胞は一つのメモリスロットとつながっている。また、メモリスロット中の細胞数は、入力層の細胞数と同数であり、入力も受け取っている。また、WTA層と各メモリスロット間の結合荷重は全て1に設定されている。また、各メモリスロット内の細

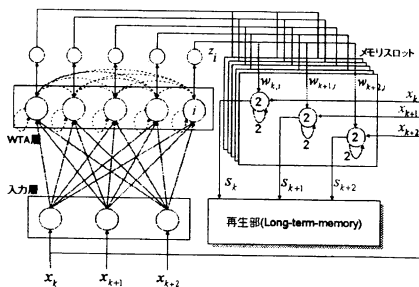


図3 短期貯蔵庫部の回路モデル

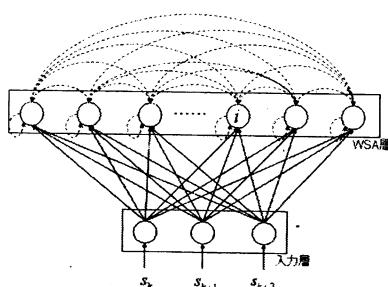


図4 長期貯蔵庫部の回路モデル

胞には興奮性の自己フィードバック結合があり、その結合荷重は全て2に設定されている。また、メモリスロット内の細胞は閾値を持っており、全ての細胞において閾値は $T = 2$ に設定されている。

短期記憶アルゴリズム

- (1) 感覚登録器部から入力層に入力が伝えられる。
- (2) WTA 層において、LV 型競合システムにより競合が起こり、勝者が一つ決定される。LV 型競合システムにおいて、 i 番目の細胞への入力 W_i は以下の式で表される。

$$W_i = \sum_{k=1}^n w_{i,k} \cdot x_k$$

ここで、 n は入力層の細胞の数、 x は入力を表す。決定された勝者を winner cell と呼び、その時の添字 i を win とする。また、winner cell とつながるメモリスロットを winner slot と呼ぶことにする。また、各細胞の上部に細胞があり、それによって出力活性度 z_{win} が1になる。

- (3) winner slot 内の細胞は、 $x_k + z_{win} \geq T = 2$ ならば出力活性度 $s_k = 1$ を出力する。
- (4) 興奮性の自己フィードバック結合により、一度発火すると上書きされるまで、持続的に発火する。また、winner cell との結合荷重を $w_{k,win}$ としたとき、シナプス可塑性により以下のように重み更新を行う。

$$\text{if } s_k = 1 \text{ then } w_{k,win} = -1$$

短期記憶貯蔵部からの忘却は、WTA により一度記憶したスロットが選択されたときに上書きが起こることにより、表現されている。また、興奮性の自己フィードバック回路の存在により、各メモリスロットに記憶されている情報は、上書きが起こらない限り保持され、この持続的な記憶の保持によって長期記憶貯蔵部に情報が転送されている。このことによってリハーサルを表現している。

3.3 長期貯蔵庫部

長期貯蔵庫部として WSA 層を用いる。長期記憶の回路モデルを図4に示す。入力層の各細胞と WSA 層の各細胞は全結線しており、その間の結合荷重は $0 \sim 1$ の実数値としてランダムに設定されている。

長期貯蔵庫部においては、WSA により勝者となった細胞の荷重ベクトルを以下のように変化させる。

$$w_{i(t+1)} = w_{i(t)} + \frac{z_i}{c} (x_{(t)} - w_{i(t)})$$

ここで、 i は WSA 層の細胞の番号、 w は荷重ベクトル、 x は入力ベクトル、 c は正の定数を表す。これは、荷重ベクトルを入力ベクトルに近づける学習をしていることに相当する。

ここで、荷重ベクトルを入力ベクトルに近づけることによって記憶項目を保持することにする。また、記憶された項目が、再生できるかどうかの判断は、入力された各単語の入力ベクトルと WSA 層の各細胞の荷重ベクトルとのユークリッド距離 $d_{(j)}(x, w)$ によって行うものとする。本実験では、以下の関係が成り立つときに再生可能と定義した。

$$d_{(j)}(x, w) = \sqrt{\sum_{i=1}^n (x_i - w_{j,i})^2} < r$$

ここで、 n は入力層の数、 j は WSA 層の細胞の番号である。 r は、誤り訂正符合の考え方により、各入力項目間のユークリッド距離 $\geq 2r$ になるように定める。学習の結果、この式を満たす荷重ベクトルを持つ細胞が存在すれば、再生可能となる。

WSA では、入力ベクトルと荷重ベクトルとの間で類似しているものから順に多勝者として生き残らせることができる。その性質から、後から入ってきた情報が先に蓄えられた記憶に影響を与える。例えば、記憶項目 A と B が入力される時を考えてみる。始めに、 A が入力され何個かの細胞が生き残り、その細胞の持つ荷重ベクトルは学習によって A の入力ベクトルに近

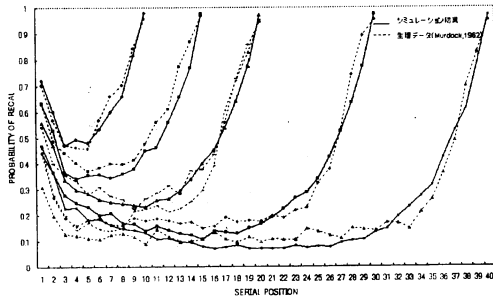


図5 自由再生実験における系列位置曲線

づけられる。その後、 B が入力され何個かの細胞が生き残るが、このとき A を記憶している細胞が勝ち残る可能性がある。そのとき、今まで A を記憶していた細胞の荷重ベクトルが、 B にいくらか近づけられるため、後に A を再生できなくなったり、 B に上書きされる可能性がある。そのため、十分な学習が行われていないと長期記憶から再生できなくなったり、忘却されてしまふと考えられる。

4. 計算機シミュレーション

自由再生実験を本モデル上でシミュレーションした。本研究では、Murdock 等の行ったような提示速度 (10 項目を 2 秒間隔で提示することを 10-2 と表す) は考慮していない。

以下のように各パラメータを設定した。

- 短期貯蔵庫部
メモリスロット数 = 5
- 長期貯蔵庫部
WSA 層の細胞数 = 300, 自己抑制強度 $k_s = 1.0$,
側抑制強度 $k_l = 0.8$, 定数 $c = 0.4$

入力は、各入力間に連想が起こらないように生成した。例えば、10 個の入力を生成するとき、入力層を 100 個とし、そのうち重複しないように 10 個ずつ発火させることで入力を生成した。そのため、長期貯蔵庫部からの再生可能かどうかの判断となる r は $\sqrt{5}$ となる。

シミュレーション回数は、Murdock 等の行った実験と同様に 1200 回行い、系列位置曲線を作成した (図 5

記憶項目提示数	平均再生数	
	シミュレーション	Murdock(1962)
10	6.36	6.40 (10-2)
15	7.54	8.19 (15-2)
20	8.12	6.83 (20-1)
		8.38 (20-2)
30	8.48	8.51 (30-1)
40	8.45	8.12 (40-1)

表 1 1 回当たりの平均再生数

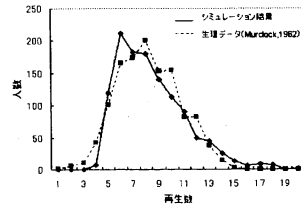


図6 40 個提示したときの再生数と人数の分布

参照)。また、表 1 には、シミュレーション 1 回当たりの平均再生数を、図 6 には、40 個の記憶項目を提示したときの再生数と人数の分布を示した。図 5、表 1、図 6 を見てわかるように、どれも心理学実験と似た値を示した。

5. おわりに

本論文では、自由再生実験における記憶の神経回路モデルを提案し、計算機実験によって、本モデルの妥当性を検証した。

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存在することが知られている⁵⁾。このことは、本研究で提案した短期貯蔵庫部のモデルにおいて、各メモリスロット内のニューロンが、持続的な発火により記憶項目の保持を行うということが、実際の脳でも起こり得ることを示唆していると考えられる。

今後の課題として、自由再生実験において提示速度を速めたときにも対応できるように、モデルを改良することがあげられる。

参考文献

- 1) Bennet B. Murdock, Jr. : Serial position effects of free recall, *Journal of Experimental Psychology*, 64, 482-488(1962).
- 2) 市川伸一, 伊東裕司, 渡邊正孝, 酒井邦嘉, 安西裕一郎: 岩波講座 認知科学 5 記憶と学習, 岩波書店, (1994).
- 3) 高野陽太郎 編: 認知心理学 2 記憶, 東京大学出版会 (1995).
- 4) Tomoki Fukai, Shigeru Tanaka : A Simple Neural Network Exhibiting Selective Activation of Neuronal Ensembles: From Winner-Takes-All to Winners-Share-All, *Neural Computation*, Vol.9 No.1, pp.77-97(1997).
- 5) Morita, M.: A neural network model of the dynamics of a short-term memory system in the temporal cortex, *Systems and Computers in Japan*, 23-4, pp.14-25(1992).