

TR-H-064

0020

人工ニューラルネットワークによる  
時空間パターン処理に関する研究

林 幸雄

1994. 3. 16

ATR 人間情報通信研究所

〒619-02 京都府相楽郡精華町光台2-2 ☎07749-5-1011

**ATR Human Information Processing Research Laboratories**

2-2, Hikaridai, Seika-cho, Soraku-gun, Kyoto 619-02 Japan

Telephone: +81-7749-5-1011

Facsimile: +81-7749-5-1008

人工ニューラルネットワークによる  
時空間パターン処理に関する研究

林 幸雄

# 目次

<b>1</b>	<b>序論</b>	<b>4</b>
1.1	研究の背景と目的	4
1.2	論文の構成	9
<b>2</b>	<b>静的ニューラルネットワークの情報処理能力</b>	<b>11</b>
2.1	はじめに	11
2.2	ニューラルネットワークモデル概説	12
2.2.1	生理学的ニューロン	12
2.2.2	ニューロンとニューラルネットワークの数理モデル	14
2.2.3	既存モデルの種々の情報処理能力向上法	17
2.3	フィードバック連想記憶モデル	19
2.3.1	能力向上法の関連性と改良モデルの提案	19
2.3.2	認識と学習処理に関する理論的考察	22
2.3.3	手書き文字による認識能力の実験的検証	25
2.3.4	誤認の傾向と静的な空間分割における問題点	31
2.4	おわりに	32
<b>3</b>	<b>発振ニューラルネットワークの動的挙動</b>	<b>34</b>
3.1	はじめに	34
3.2	発振ニューラルネットワーク	35
3.3	記憶点の不安定化による発振	37
3.3.1	記憶点の漸近安定性に関する摂動解析	37
3.3.2	単一の興奮-抑制性ペア素子の発振条件	39
3.3.3	単一のペア素子における挙動	41
3.4	ネットワークの挙動の数値解析	43
3.4.1	状態空間内の動的軌道	44

3.4.2	軌道のカオス性	46
3.4.3	入力バイアスによるアトラクター間のスイッチング例	48
3.5	記憶パターン数の増加に伴う遊走的軌道の性質の劇的な変化	50
3.5.1	記憶パターンへの引力の弱体化	50
3.5.2	入力バイアスによる遊走領域の制御	54
3.6	おわりに	56
<b>4</b>	<b>時空間パターン処理への応用</b>	<b>57</b>
4.1	はじめに	57
4.2	生理学的自律発振と認知心理学的連続変形操作	58
4.2.1	脳内の連続変形操作の存在可能性	59
4.2.2	連続変形パターンの軌道上への学習方法	60
4.3	軌道の多様性と可制御性を利用した顔パターンの生成	66
4.3.1	対話的処理による図形パターンの概念形成の支援	66
4.3.2	概念イメージに応じたパターン生成	67
4.4	状態空間構造の変化による識別領域の動的制御法	73
4.4.1	入力確信度に応じた揺らぎ出力を持つ連想記憶モデル	74
4.4.2	ベイスン拡大／縮小による状態空間構造の変化	77
4.4.3	単語からの予測による文字の識別領域の動的制御	79
4.5	おわりに	82
<b>5</b>	<b>ネットワークの結合形態と同期 / 非同期発振</b>	<b>83</b>
5.1	はじめに	83
5.2	パラメータ空間内の同期領域の定量解析	84
5.2.1	ネットワークの結合形態	84
5.2.2	数値解析結果	86
5.3	分岐ダイアグラムの利用方法	92
5.4	おわりに	94
<b>6</b>	<b>結論</b>	<b>95</b>
6.1	研究成果の要約	95
6.2	今後の研究課題	97
6.3	むすび	98

謝辞	99
参考文献	100
本論文に関する原著論文	110
A 付録： Lyapunov 安定性の証明	112
B 付録： 遊走軌道に対する最近接頂点の滞在率の分布	114
C 付録： 発振ニューラルネットワークの学習方程式の導出	123
D 付録： 顔パターンの特徴要素	126
D.0.1 顔の輪郭形状	127
D.0.2 顎の輪郭形状	128
D.0.3 目の特徴要素	129
D.0.4 眉毛の特徴要素	131
D.0.5 唇の特徴要素	132
D.0.6 顔の各要素の配置	133
E 付録： 漸近安定平衡点とリミットサイクルの共存条件	134
F 付録： 1次 Hopf 分岐点の近似解析	138

# 第 1 章

## 序論

### 1.1 研究の背景と目的

脳は140億ほどのニューロン（神経細胞）が複雑に結合した回路網で構成された巨大な情報処理装置である。その回路網上で、時間的にも空間的にも分散化された情報を並列的に処理することが、脳の情報処理における構造的特徴といえる。その機能的特徴としては、過去の失敗例や成功例の経験を記憶して利用出来るばかりか、外界の環境に適応した迅速な判断や行動を可能とするように自己の能力を高めていくという学習も可能な優れた能力を持っている。こうした脳の複雑な構造や機能に関して、これまでの生理学的または解剖学的研究で、単一のニューロンの基本動作や学習に関与する回路網内の結合の可塑性、脳のマクロな解剖学的構造などが明らかにされている。一方、実際の脳を調べることが難しい為、情報処理の立場からの数理的研究も活発に行われている。すなわち、生理学的知見に基づいた回路網のモデルや、特定の情報処理機能を実現する為の人工的モデルを作り、数学的に扱いやすいようにモデルを単純化して、その情報処理の本質を数理的に明らかにしようというわけである。脳の情報処理様式が明らかになれば、現在のコンピュータが不得意とするパターン情報を扱う認識や運動制御等のさまざまな問題に対して、その成果を工学的に応用することは大いに期待出来る。その際、応用に適した機能をモデルに実現するには、ネットワーク構造の設計が重要となる。なぜなら、もし特定構造のモデルがそのパラメータ値における広い範囲で要求される機能を持つならば、その実現は実際的に可能と考えられるからである。ここで、ネットワーク構造とは、モデルのネットワーク形態のみならず、ネットワークが有する結合重みなどに関する特徴（例えば、素子の活性化に対する正の結合や負の結合など）を含む性質をさすものとする。この反面、ネットワーク構造を特定することは計算能力を限定することになり、最悪の場合、どんな手続きによってもその機能を実現することが不可能となり得る。

脳内の回路網の構造と機能について、情報処理の立場からもう少し詳しく述べよう。脳は解剖学的に、大脳、小脳、脳幹と呼ばれる部分に大きく分けられ、このうち大脳には、視聴覚等の

五感を感覚器から受け取る感覚野、随意運動に関係する運動野、その他の言語機能等に関わる連合野などと呼ばれる、さまざまな機能的部位が存在する。さらに、各機能における情報伝達経路には、特異なネットワーク構造がみられることもわかっている（塚原, 1986; 小林他, 1987）。ネットワーク構造とその機能との間には密接な関係があると考えられるが、脳のような複雑なシステムにおいては、各部分の構造と機能とは必ずしも1対1に対応するとは限らない。そこで、1つの部分的なシステムが多機能を持つとする考えた方が指摘されている（Getting, 1989; 川原, 1992）。逆に、1つの機能には複数の実現方法があるということも考えられる。この多機能性の考えによれば、ある特定構造を持つ部分は、大まかには決まった機能を持つが、細かく見るといくつかの機能を内在するおかげで、状況に応じた機能を選択的に表出させることが出来る。これを支持する根拠は以下のようなものである。

脳の情報処理は、多数の要素間の非線形の相互作用による並列のダイナミクスに基づくと考えられる。ここで、ダイナミクスとは、システムの時間発展に関する（静止状態への収束や周期的発振などの）動的な挙動のことをさす。非線形システムでは外部入力やパラメータなどの値の変化によって、静止状態から突然に発振を始めるような、質的にまったく異なる挙動が生じ得る（こうした質的变化は分岐現象と呼ばれる）。情報処理の観点からは、静止状態と周期的発振のそれぞれに対応する機能は、例えば、下等動物のウミウシにおける引き込み反射の静止硬直状態と逃避の為の遊泳モードのように（Getting, 1989）、全く異なるものとなり得る。この様な構造と機能との対応付けに関する議論は、大脳皮質内のコラム構造と呼ばれる形態学的なモジュール構造についての機能単位の存在性に関してもなされている（丹治他, 1992; 清水他, 1992）。

以上の背景を踏まえて、本研究では、特定のネットワーク構造を持つ人工的なニューラルネットワークモデルのダイナミクスの解析を通じて、

- (1) ネットワーク構造と情報処理機能との密接な関係
- (2) 外部入力などの変化に従った情報処理機能のスイッチング

に関して情報処理の観点から議論する。モデルとしては、ネットワークが有する結合重み中に、複数の記憶パターンを分散的に保持した、内容番地方式の連想記憶の機能を持つものを考える。内容番地方式の連想記憶とは、多数の事項を同じ場所に重ね合わせる多重記憶方法の1つであり、アドレスを介さずに直接に関連事項からネットワークのダイナミクスを用いて、記憶事項の想起が並列に実行できる。このため、雑音などが交じった部分的な事項からも全体を想起できること、検索時間が短いことなどの長所を有する。モデル上で扱う対象としては、入力パターンは静的であっても、ダイナミクスの結果として生じる出力パターンは時間変化を伴う、時空間パターン処理を考える。例えば、ある回転角だけ異なる2つの図形を左右の目の視覚領域に提示した

時に、図形の一致度を判別する為に人間が頭の中で行う視覚イメージの心的回転操作（認知心理学では、メンタルローテーションと呼ばれる）のようなものが、時空間パターン処理の1例である。時空間パターン処理の重要性は、脳における時空間の情報処理やその時間情報のコード化方法に関する最近の議論（塚田他, 1992）からも伺える。以下では、本研究のより具体的な課題を述べる。

まず、(1)に関する研究として、これまでに提案された、記憶パターンの情報表現、ネットワーク構造、ネットワークの結合重み係数の設定方法や学習方法などの工夫による連想能力の向上法を説明し、これらが松岡によって提案された連想記憶モデル（松岡, 1989, 1990）において互いに関連付けられることを示す。しかしながら、このモデルは基本的には線形モデルなので、正しい記憶パターンの想起に悪影響を及ぼす（他の記憶パターンとの相関によって生じる）クロストークノイズに弱い。そこで、非線形処理を施しながら出力を入力にフィードバックすることを繰り返して、ノイズ除去を行う改良モデルを提案する。また、上記の向上法で得られるパターン認識能力を、本提案モデルの学習で獲得する方法を考え、線形主成分分析を近似的に実現する従来モデルの学習（Rumelhart, McClelland, & the PDP Research Group, 1986; Sanger, 1989; 船橋, 1990）との類似点及び相違点を示す。これらの類似性によって、パターン分離としての主成分分析の機能を近似的に実現するための、複数の実現方法の存在が具体的に示される。さらに、モデルの認識能力を理論的・実験的に検討する。この実験結果から誤認の原因を考察するとともに、このモデルを含んだ静的ニューラルネットによる情報処理が、ニューラルネットとは独立に研究されてきた統計的なパターン識別手法（長尾, 1983）における静的な空間分割の枠組を、概念的に越えていないことによる以下のような問題点を指摘する。ここで、静的ニューラルネット（又は、静的なモデル）とは、ネットワーク結合内にフィードバックループを持たない入力から出力への一方向の信号の流れで、入出力パターン間のパターン変換を行うものや、フィードバックループがあっても入力パターンである初期状態に従って、出力パターンがある静止状態へ収束するような、ネットワークモデルのことをさす。この場合、1つの出力パターンに対応付けられる入力パターンの集合によって、入力パターン全体の状態空間は分割される。しかし、その空間分割が決められた後は、学習や識別時の動的過程は情報処理には直接関与しない。すなわち、学習や識別時における時間変化を伴う状態の軌跡は過渡現象であり、最終的な落ち着き先の収束点（平衡点）のみが意味を持つからである。ゆえに、静的な空間分割においては、分割の境界付近のパターンを正しく識別することが困難であるばかりか、多義的に解釈可能な曖昧なパターンを識別する事が原理的に不可能となる。例えば、C /-\ T と T /-\ E における曖昧な文字 /-\ を、人間は前後関係からそれぞれ A と H に解釈出来る。一方、静的な空間分割においては、この同一文字 /-\ を同時に2つ以上の識別領域に対応付けることが出来ない為、必ず矛盾が

生じる。仮に、誤認領域の識別境界を学習等で修正したとしても、こうした矛盾領域が存在する限り、新たな誤認領域が発生して循環的なジレンマに陥ってしまう (Carpenter & Grossberg, 1988)。そればかりか、統計的データに基づいた膨大な時間を要する学習だけに頼っていたのでは、脳のような外部環境の変化に迅速に適応した情報処理は出来ないと考えられる。

上記の問題点を解決する為の1つの手段として、(2)に関する研究として、動的な出力軌道を情報処理機能に対応付け、外部入力などに従って迅速に変化する状態空間の分割構造を、情報処理機能のスイッチングに利用することを提案する。これは、非線形システムの分岐現象を伴った複雑な挙動の変化を、情報処理に積極的に利用しようという1つの試みである。通常、非線形システムは、アトラクターと呼ばれる時間発展しながら近付いていく状態の集合をいくつか持つ。アトラクターには、吸引されて静止状態となる漸近安定平衡点や、ある限られた範囲を循環して元に戻る周期的な軌道上に吸引されるリミットサイクル、さらに、ある限られた範囲内ではあるが複雑な挙動を示すカオスと呼ばれるものなどがある。各アトラクターへ収束する初期状態の範囲で定義される、吸引領域 (アトラクターのベイスンと呼ばれる) で全状態空間は分割されるが、それは静的な空間分割の場合とは以下の2つの点 (i) と (ii) において異なる。

(i) 第1に、静止状態への収束のみならず、時間変化を伴う様々な軌道を定常的なアトラクターとして生成出来ることから、もし状態空間内の軌道が多様性を持てば、それらの軌道に対応付られた種々な機能に関する時空間パターン処理を扱うことが可能となる。ここでの多様性とは、状態空間内のさまざまな点を通り過ぎて複雑に遊走するような軌道の性質のことをさす。特に、カオス軌道の初期状態鋭敏性 (微妙に異なる2つの初期状態からの軌道が指数オーダーで離れて行くという性質) は、多様なパターンを生成する情報生成源としての性質を備えている (Nicols, & Tsuda, 1985; Nicols, 1986; 津田, 1990)。この性質は、静的なモデルにはない新しい機能となり得る。

(ii) 第2に、静的なモデルにおいては、ベイスンによって分割された状態空間の分割構造を、外部入力に従って変えることは出来るが、時間的に変動する入力を入れられない限り、ダイナミクスの性質としては漸近安定平衡点への収束であることに何ら変わりがない。すなわち、空間分割に基づくパターン分離、あるいは入出力間のパターン変換の機能しか実現出来ない。一方、出力軌道にリミットサイクルやカオスなどを持つ動的なモデルにおいては、外部入力などの変化によって、状態空間内のベイスンの数や大きさを変えるのみならず、軌道の性質を変えるなどの分岐現象を伴った複雑な変化 (例えば、漸近安定平衡点からリミットサイクルの変化や、リミットサイクル上で軌道が戻るまでの回転数の変化、カオスへの変化など) が起こる。こうしたアトラクター上の動的な軌道の性質の変化によって、情報処理機能のスイッチングが可能となる。一般に、分岐現象を伴う状態空間構造の変化は非常に複雑なので、厳密な軌跡の制御は不可能と考え

られるが、動的軌道の遊走領域を大まかに制御することは可能と考えられる。本研究では、軌道の可制御性をこの意味で用いることとする。

これら (i) と (ii) の2つの特徴によって、外部入力などの変化に従って状態空間内のアトラクター間の遷移を引き起すことで、アトラクターに対応付られた情報処理機能を迅速にスイッチングすることが可能となる。例えば、記憶パターン付近のリミットサイクルによる記憶パターンの想起（振動は入力パターンの記憶パターンに対する差の度合としての曖昧さを表す揺らぎに相当）から、記憶パターン間のカオス的軌道による記憶パターンの探索という具合に、外部入力によって機能的なスイッチングが制御出来る。この概念を実現する為の1つの具体的なモデルとして、正と負の互いに反対符号の結合重み係数で2素子間を結合した興奮-抑制性ペア素子で構成される発振ニューラルネットワークを提案する。さらに、記憶パターンの想起と探索についての機能的側面から、状態空間内の軌道の多様性や可制御に関する特徴を検討する。

最後に、(1) と (2) の双方に関わる研究として、興奮-抑制性ペア素子で構成されるモデルにおけるネットワーク構造と同期/非同期発振（周波数ロック/周波数アンロック）の関係について検討する。動的なモデルは発振現象などの複雑な挙動を持つ為に、(1) に関するネットワーク構造と機能との関係が静的なモデルほど検討されていない。そこで、本研究では、2組のペア素子間の結合形態と（入力値や結合重み係数に関するパラメータ空間内での）同期発振の起こりやすさの関係に限定してこれらを検討する。同期発振の起こりやすさは外部入力にも大きく依存するので、同期/非同期発振に対応付られた機能の変化は、(2) に関係する研究課題でもある。

最近の生理学的実験から、入力に依存した同期発振にもとづいて、脳内では特徴要素の統合が動的になされる可能性が示唆されている (Eckhorn et al., 1990; Gray et al., 1990; 松本, 1992)。この仮説は、要素間の組合せ爆発を回避する1つの有力な手段と考えられる。例えば、赤い丸と緑の四角を見た時、脳内に存在すると考えられている、赤や緑の色特徴検出細胞（あるいは細胞群、以下同様）と、丸や四角の形状特徴検出細胞が同時に活性化（発火）したのでは、色と形の対応付けが出来ない。一方、色と形などの異なる要素間の同時発火をモニターする為の関連付け細胞があるとすれば（その存在性は定かでないが）、組合せ爆発を生じる事は明らかである。そこで、同期発振によって関連要素を動的に結びつけるとともに、非同期発振によって無関係な要素を空間的または時間的に分離出来れば、この問題は回避出来る。さらに、動的な特徴統合ならば物理的な結線を必要としないので、結線の空間的な制約がなく、遠くに離れた要素どうしを容易に関連付けられるという長所も有する。本研究の解析結果は、同期/非同期発振による特徴統合/分離問題に対するネットワークモデル設計に1つのガイドラインを与えるのみならず、前述の発振ニューラルネットワークが非同期的なカオスを起こしやすいネットワーク構造で

あることを示唆するものでもある。

## 1.2 論文の構成

本論文は、人工ニューラルネットワークにおけるネットワーク構造と情報処理機能との密接な関係や、外部入力などによる迅速な情報処理機能のスイッチングに関して論じたもので、序論と結論を含めて全体を6章で構成している。第1章の序論に続いて、まず第2章で、既存の静的ニューラルネットワークにおける種々の情報処理能力向上法の関連性を示し、静的な空間分割に基づく情報処理の問題点を述べる。第3章で、この問題点を解決する為の1つの手段として、動的な出力軌道を情報処理機能に対応付け、外部入力に従った状態空間構造の変化を情報処理機能のスイッチングに利用することを提案する。さらに、これを実現する1つの具体的なモデルとして発振ニューラルネットワークを提案し、その動的挙動の解析結果について述べる。第4章で、発振ニューラルネットワークの軌道の多様性と可制御性に関する特徴を積極的に利用した具体例から、時空間パターン処理への応用可能性を示す。第5章で、ネットワーク構造と同期発振の起こりやすさの関係について述べる。以下、各章の内容をやや詳しく述べる。

第2章では、まず、既存のニューラルネットワークの数理モデルについて概説する。その後、これまでに提案された種々の情報処理能力向上法の関連性を示し、これらの長所を継承しつつ、手書き文字などの変動成分の多いパターンの認識と学習に適した改良モデルを提案する。このモデルに関して、認識と学習処理に関する理論的考察を行い、手書きひらがな類似文字に対する実験でその認識能力を検証する。さらに、誤認結果の原因を考察するとともに、静的な空間分割に基づくパターン分離における本質的な問題点を指摘する。

第3章では、静的な空間分割に基づく情報処理の問題点を解決する為の1つの手段として、「外部入力に従った状態空間構造の劇的变化によって、アトラクター間の遷移を引き起こし、これを情報処理機能のスイッチングに利用する」ことを提案する。これを実現する為の具体的なモデルとして、興奮-抑制性ペア素子を構成要素とする発振ニューラルネットワークを考え、その動的軌道の多様性と可制御性についての特徴を、主に数値解析を通じて明らかにする。

第4章では、前述した軌道の多様性と可制御性に関する特徴を積極的に利用して、時空間パターン処理への応用可能性を検討する。まず、生理学的な脳内の発振現象と認知心理学的な図形パターンの連続変形操作との、自律処理の観点からの関連付けを試みる。ここでは、学習によってモデルのリミットサイクル上に文字パターンの連続変形が埋め込まれることを示す。次に、カオス的軌道の情報生成源としての性質を利用した、多様な顔パターンの生成を試みる。ユーザの概念レベルに適応した動的な探索領域の制御方法を提案するとともに、多様な顔パターンが生成出来ることを示す。さらに、単語からの文字予測で識別領域が迅速に変化出来るような、識別領

域の動的な制御方法を提案する。すなわち、外部入力的な予測に従って、ベイソンの拡大や縮小が起こり、その結果、ベイソンの集まりとしての識別領域が動的に変化することを示す。

第5章では、2組の興奮-抑制性ペア素子間の結合形態と同期発振の起こりやすさの関係を、2つの入力値と2組のペア素子間の結合重み係数についての3つのパラメータに対する定量的解析を行う。この結果は、同期発振と非同期発振で特徴要素の統合と分離をそれぞれ行う問題に対する、ネットワーク設計のガイドラインとして利用出来るのみならず、第3章で議論した発振ニューラルネットワークが広いパラメータ空間内で非同期のカオスの軌道を起こしやすいことを示唆するものでもある。

第6章では、結論として以上に関する成果のまとめを行った後、今後の課題について述べる。

## 第 2 章

### 静的ニューラルネットワークの情報処理能力

#### 2.1 はじめに

ニューラルネットワークモデルのネットワーク構造と、その情報処理機能との間には密接な関係があると考えられる。有限ステップもしくは連続時間の状態遷移を過渡的なものと見なし、情報処理には直接用いない静的なモデルにおいては、収束したモデルの各出力状態に対応付けられた入力空間の分割構造をどのように決めて、入力パターンを分類するかが重要となる。そこで、連想記憶モデルの記憶パターンの情報表現、パターン分類モデルにおけるネットワーク構造、連想記憶モデルのネットワークの結合重み係数の設定方法や学習方法という観点からの、種々の能力向上法がこれまでに提案されている。

本章では、これら 3 点に関する能力向上法が松岡によって提案された連想記憶モデル（松岡, 1989, 1990）によって互いに関連付けられると考え、これら 3 点からの高い能力を継承しつつ、手書き文字等のパターン認識に適した新しいモデルを提案する。まず、手書き文字等におけるさまざまな変動成分としてのノイズを吸収するために、基本的には線形な松岡のモデルに非線形フィードバックを付加する。次いで学習方程式を考えるために、前記の離散時間のモデルを連続時間モデルに拡張して、本提案モデルの認識処理と学習処理の挙動を理論的に考察する。また、手書きひらがな類似文字の認識実験より、その認識能力を評価する。さらに、線形主成分分析を近似的に実現する Sanger のモデルの学習（Sanger, 1989）や、恒等写像を近似する階層型のネットワーク（船橋, 1990）に対するバックプロパゲーション（BP）学習（Rumelhart, McClelland, & the PDP Research Group, 1986）と、本提案モデルの学習が定式化において類似することを示すとともに、これら 3 つの学習の相違点について、実験結果に基づいた検討を加える。これらの学習の類似性は、パターン分類としての主成分分析の機能をニューラルネットワークモデルで近似的に実現するための、複数の実現方法を示す具体例と考えられる。逆に、これら複数の方法で実現された機能は、本質的には主成分分析に基づくパターン分離であることが示される。

## 2.2 ニューラルネットワークモデル概説

2.2節では、本論文を通じての予備事項として、神経生理学が明らかにしたニューロンに関する知見（塚原, 1986; 小林他, 1987）を手短かに述べ、これを基にしたニューロンの数理モデルとネットワークモデルについて概説する。次に、静的なモデルの能力向上法として提案された、記憶パターンの情報表現、ネットワーク構造、ネットワークの結合重み係数の設定方法や学習方法を説明する。さらに2.3節では、これらの能力向上法が互いに関連性を持つ事を示す。

### 2.2.1 生理学的ニューロン

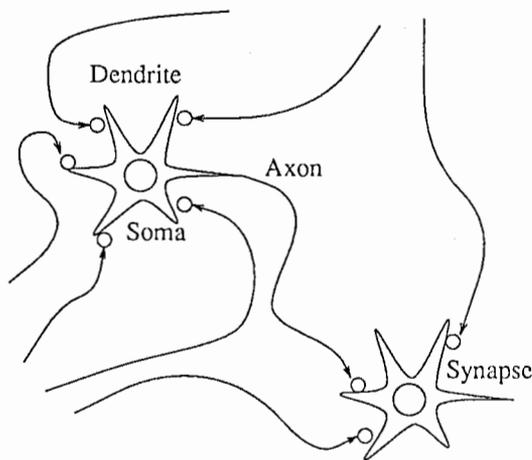
脳内のニューロンには、その大きさや形状が異なる様々な種類のものが存在するが、その構成要素や動作原理はどれもほぼ同じである。ニューロンは図 2.1(a) のように、核の存在する細胞体 (Soma) と呼ばれる本体部分、本体から樹状に突き出た多数の突起からなる樹状突起 (Dendrite) と呼ばれる部分、本体から伸びた一本の長い繊維の軸索 (Axon) と呼ばれる部分から構成される。軸索は途中で枝分けれして、その末端が他のニューロンの樹状突起とシナプス (Synapse) と呼ばれる部分で結合している (接触部分には約  $150 \sim 200 \text{ \AA}$  の間隙があり、軸索側とニューロン側の部分でそれぞれ、シナプス前膜とシナプス後膜と呼ばれている)。ニューロンの持つ機能には、(i) シナプスを介して他のニューロンから伝えられた入力信号の受容、(ii) 細胞体における複数の入力信号の統合と出力信号への変換、(iii) 軸索による出力信号の伝達がある。生理学的知見（塚原, 1986; 小林他, 1987）に基づくこれらの動作を以下に概説する。

まず、入力信号の受容について述べる。他のニューロンからの神経インパルス信号が、軸索を伝わって終末部に到達すると、シナプス前膜からシナプス間隙に化学伝達物質が放出され、これがシナプス後膜の受容体と結合すると、化学的な開閉チャンネルが開いて細胞内にイオンの流出が起こり、細胞体の膜電位を変化させる (ニューロンは脳脊髄液のプールに浸されていて、このような細胞膜を介した物質のやり取りに従って処理を行っている)。化学伝達物質はニューロンの種類によって異なり、結合した先の相手のニューロンの膜電位を高めようと作用するものと、低めようと作用するものがある。この化学伝達物質の性質に従って、パルスの送り手の細胞はそれぞれ、興奮性ニューロンと抑制性ニューロンと呼ばれるものに大きく分けられる。また、これらとは別にシナプス前抑制を仲介するニューロンもある。

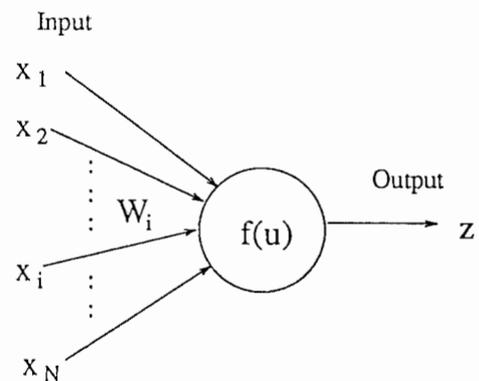
次に、細胞体における入力信号の統合と入出力の信号変換について述べる。シナプスに1つのパルスが到着してから膜電位が変化するまで、 $0.5\text{msec}$  程度の時間遅れが生じる。この膜電位の変化は樹状突起から細胞体へ伝わりながら、数 msec の時定数で減衰する。多数のパルスが同時に入力されると、各シナプスでの膜電位の変化は空間的に加算される。一方、複数のパルス入

力に時間遅れがある場合は時間的加算が行われる。すなわち、パルスによる膜電位の変化後、その減衰前に新しいパルスが到来すれば、残っている前の電位に新しい電位変化が加算される。細胞内の膜電位は外部と比べて  $-70\text{mV}$  ぐらいであるが、前述の様な空間的または時間的に加算された膜電位の変化によって、それがしきい値  $-55\text{mV}$  を越えると、出力信号としての神経インパルス（高さ  $100\text{mV}$ 、幅  $1\text{msec}$  程度）を放出させ、その後、膜電位は急速に落ち込み元に戻る。これをニューロンの興奮または発火という。但し、絶対不応期と呼ばれるパルス放出直後は、たとえ強い入力信号が来ても興奮出来ないばかりか、相対不応期と呼ばれるこの後のしばらくの間も、しきい値が高くなって興奮しにくくなる。ここで、脳内のシナプスは可塑性を持つことを特筆したい（もちろん、生後のある発達段階で固定化されるものもある）。すなわち、発生段階やその後の環境との相互作用（学習）を通じて、発芽等の形態学的変化や伝達物質の放出量の変化が起こり、（入力信号を膜電位変化に変換する）伝達効率が適応的に変わるのである。

最後に、軸索上の信号伝達について簡単に述べる。細胞体で生じた神経インパルスは、軸索を  $100\text{m/sec}$  程度のスピードで伝播する。これは、細胞体でのパルス発生と同様な、イオンチャネルの開閉とそれによる軸索内の膜電位変化が、軸索上で将棋倒しの様に次々と引き起こされる事によるものである。



(a) 生理学的ニューロンの形態



(b) ニューロンの数理モデル

図 2.1: 生理学的ニューロンとニューロンの数理モデル

## 2.2.2 ニューロンとニューラルネットワークの数理モデル

機能的には、ニューロンは図2.1(b)のような多入力-出力の情報処理素子と考えられる。その基本動作は、 $N$ 個のニューロンから受けた入力信号  $x_1, x_2, \dots, x_N$  が、 $i$ 番目の軸索の伝達効率  $w_i$  で膜電位の変化  $w_i x_i$  にそれぞれ変換され、その空間的 ( $\sum_i w_i x_i$ ) または時間的 ( $\int_{-\infty}^t w_i (t-s)x_i(t-s)ds$ ) な総和がしきい値  $h$  を越えた場合に、出力信号  $z$  を出す。ニューロンの数理モデルを考える時、生理学的な特徴を忠実に考慮することは必ずしも得策ではなく、動作の重要な本質的性質をそこなわない限りで単純化した方が良いと考えられる。そこで、数理モデルとしては、入出力信号をパルスの有無の様な離散情報とするか、ニューロン集団の平均的なパルス頻度の様な連続情報とするかで、さらに、その動作を離散時間で扱うか、連続時間で扱うかで、以下の4つのタイプに大きく分類出来る (甘利, 1978)。

**離散時間 - 離散情報モデル:** これは、McCulloch-Pitts の形式ニューロン、あるいは、しきい素子と呼ばれるもので、その動作は以下の式に従う。

$$z = 1\left(\sum_{i=1}^N w_i x_i - h\right). \quad (2.1)$$

ここで、 $\sum_{i=1}^N w_i x_i$  は膜電位の変化量の総和を表し、 $1(u)$  は  $u$  の値に従って2値状態  $1 (u > 0), 0 (u \leq 0)$  をとる図2.2(a)のような階段関数である (数学的取扱いの容易さから、 $\{-1, 1\}$  の2値状態のモデル  $z = \text{sgn}(\sum_{i=1}^N w_i x_i - h)$  を用いることもある)。

**離散時間 - 連続情報モデル:** 前述の離散時間 - 離散情報モデルは、単一のニューロンのみならず、時間的加算や不応期を考慮することでニューロン集団の動作を示すモデルともなる。しかし、これではモデルが非常に複雑になるので、これらの効果を自然に取り込められる、以下のような連続情報モデルが考えられている。

$$u(t) = \sum_{i=1}^N w_i x_i(t) - h, \quad (2.2)$$

$$z = f(u(t)). \quad (2.3)$$

ここで、非線形関数  $f(u)$  は図2.2(b)のような  $\{0, 1\}$  で飽和するS字型の関数であり、時間和作用、しきい値作用、不応期の効果が含まれている。

**連続時間 - 連続情報モデル:** 連続情報モデルを採用するとき、時間も連続的に取り扱う方が自然であることから、以下のような微分方程式で記述されたモデルが考えられている。

$$\tau \frac{u(t)}{dt} = -u(t) + \sum_{i=1}^N w_i x_i(t) - h, \quad (2.4)$$

$$z = f(u(t)). \quad (2.5)$$

その他のモデル: これら以外にも、 $\{1, 0\}$  の 2 値状態を確率的にとって、動作を確率的にとらえたモデルや、入力の重み付け線形和のみならず、シナプス前抑制等の効果に相当する 2 次以上の高次相関を考慮した  $\Sigma$ - $\Pi$  ユニット (Rumelhart, McClelland, & the PDP Research Group, 1986)、より忠実に生理学的な細胞膜の興奮を数式化した、Hodgkin-Huxley 方程式に関連するカオスニューロン (Aihara, Matsumoto, & Ikegaya, 1984; Aihara, 1990a, 1990b) などが考えられている。

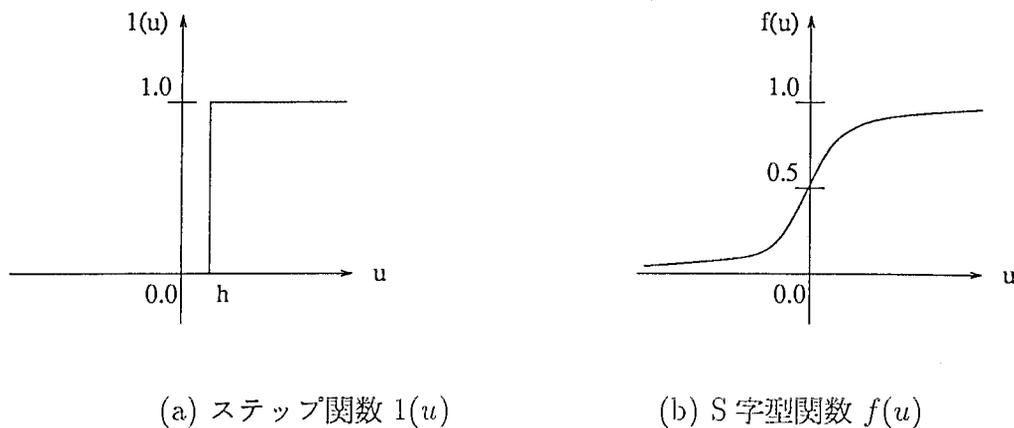


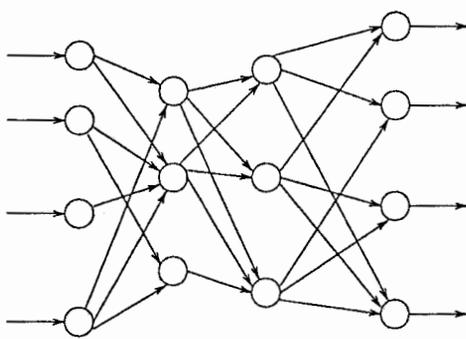
図 2.2: ニューロンモデルの非線形関数

以上、ニューロンの数理モデルについて説明した。以下では、これらの素子が結合したネットワークモデルについての、構造的な分類と原理的な計算能力を概説する。原理的な計算能力とは、そのネットワークで実現可能な能力の存在性を保証するものではあるが、その機能を実現する為の具体的手続きを与えるものではない。すなわち、パラメータ空間内の非常に限られたほんの 1 部分でしか、その機能が存在し得ないような極端な場合では、実際上それは実現不可能となる。ある具体的手続きによって実現可能な機能は、通常、ネットワーク構造や、結合重み係数の設定方法あるいは学習方法などに強く依存する。これらに関しては次の 2.2.3 で触れる。

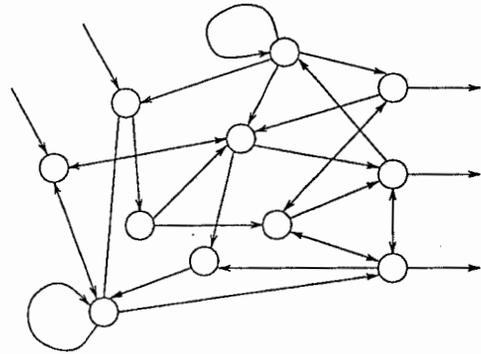
フィードフォワード階層型モデル: 図 2.3(a) のように入力側から出力側への一方向の信号の流れで、(離散時間の) 階層的な処理がなされるネットワークモデル。機能的には、特徴抽出やパターン分離等のフィルタ的な入出力間のパターン変換を行うものである (Lippmann, 1987; 麻生, 1988)。計算能力に関する理論研究として、離散情報のしきい素子回路の論理関数としての万能性 (室賀, 茨木, & 北橋, 1976) をはじめ、離散情報モデルと連続情報モデルのランダム結合の回路におけるパターン分離度 (甘利, 1978) や、連続情報モデルにおいて、十分多くの中間層素子を持つ 4 層ネットワークによる任意の連続写像の任意精度の近似実現可能性 (Funahashi, 1989) 等が明らかにされている。

リカレントネットワークモデル: 図 2.3(b) のようにフィードバックループを含むネットワークモデルで、一般に時空間パターンの変化を伴う複雑な挙動を示す。機能的には、閉回路等による発振現象が短期記憶 (小林他, 1987; 黒田, 1991) や動的な特徴統合 (松本, 1992) と関連することが示唆されているものの、上記の入出力パターン変換の静的な写像と比べて、動的な情報処理の枠組自体がまだはっきりしていないのが現状である。計算能力に関する理論研究も発展途上ではあるが、(離散、連続の情報と時間のそれぞれのモデルにおける) ランダムな相互結合ネットワークの全素子の平均活性度で定義された巨視的情報に関する、力学系の挙動の質的变化を示すカタストロフィー現象や、興奮 - 抑制性神経集団の発振現象 (甘利, 1978) をはじめ、連続情報 - 連続時間モデルにおける、有限時間の区間内での力学系の軌道の任意精度の近似的実現可能性 (Funahashi, & Nakamura, 1992) や、漸近安定平衡点を持つ力学系の軌道の (グローバルな時間区間での) 任意精度の近似的実現可能性 (中村, & 船橋, 1992) 等がある。

神経場のモデル: 大脳や小脳における一様な皮質構造に対応するモデルとして、可算個のニューロンが結合したネットワークを連続場で近似して扱ったもの。神経場のパターン力学の理論研究として、視覚系の錯視現象を説明する線形変換神経場、短期記憶のメカニズムと関連する、非線形神経場における興奮波の保持・伝播・再生 (甘利, 1978) 、皮質内で見られる地勢学的構造としての結合の対応関係を示す、トポグラフィックマッピングの自己形成 (甘利, & 竹内, 1979) 等がある。



(a) フィードフォワード階層型モデル



(b) リカレントネットワークモデル

図 2.3: ネットワークの結合形態による分類

### 2.2.3 既存モデルの種々の情報処理能力向上法

入出力間のパターン変換写像(フィードフォワード階層型モデル)や、平衡点への漸近安定収束性(対称相互結合型等のリカレントネットワークモデル)などの特徴を持つ静的ニューラルネットワークの連想記憶モデルにおいて高い連想能力を実現する為に、(i) 記憶パターンの情報表現、(ii) ネットワーク構造、(iii) 結合重み係数の設定方法や学習方法などに関する工夫が、これまでにいくつか提案されている。以下、これらについて順に説明する。

まず、高い連想能力を持つのに適した、記憶パターンの情報表現に関して説明する。発火と休止の2つの状態に対応する  $\{0, 1\}$  (あるいは、 $\{-1, 1\}$ ) の2値状態で、 $N$  個の各要素が符号表現された記憶パターン  $\{\xi_1^\mu, \xi_2^\mu, \dots, \xi_N^\mu\}$ , ( $\mu = 1 \sim M$ ) を考え、これらの記憶パターンの自己相関行列で結合重み係数  $\{W_{ij}\}$  を定義した、以下のような相関行列型連想記憶モデルを考える。

$$x_i(t+1) = 1(\sum_{j=1}^N W_{ij}x_j(t) - h_i), \quad (2.6)$$

$$W_{ij} \equiv \frac{1}{N} \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu. \quad (2.7)$$

この時、発火と休止の要素数に偏りがある符号表現の方が、発火と休止が同数の偏りのない符号表現より、安定に想起可能な記憶パターンの最大数を多く出来ることが報告されている (Palm, 1988; Amit et al., 1987a; Amari, 1989)。すなわち、 $N$  個の要素中  $\log N$  個という極端に少ない発火パターンで表現された、スパースな符号表現が記憶容量の面から有利となる。ここで、記憶容量とは、パターンの次元数が  $N$  の時、全ての記憶パターンが1ステップ後の想起で不変な平衡点となるような、埋め込み可能な記憶パターンの最大数  $M$  のことである。例えば、適当に素子のしきい値を調整すれば、 $\{-1, 1\}$  の時は記憶容量が  $0.14N$  から  $0.18N$  に向上し (Amit et al., 1987a)、 $\{0, 1\}$  の時はこれ以上に飛躍的に向上する (Amari, 1989)。生理学的にも、視覚的図形の再認記憶に関する側頭葉のニューロンの発火パターンが、スパースな符号表現を支持するという興味深い実験結果がある (Miyashita, & Chang, 1988)。

次に、連想能力を向上させる為のネットワーク構造について説明する。相関型連想記憶モデルでは、記憶パターン間の相関によるクロストークノイズが、正しい記憶パターンの想起に悪影響を及ぼす。例えば、式 (2.7) の定義により、 $\alpha$  番目の記憶パターン  $\{\xi_j^\alpha\}$  を入力した時は、

$$\sum_j W_{ij} \xi_j^\alpha = \xi_i^\alpha + \sum_{\mu \neq \alpha} \left( \frac{1}{N} \sum_k \xi_k^\mu \xi_k^\alpha \right) \xi_i^\mu. \quad (2.8)$$

という具合に、上記の第1項の信号成分に対して第2項がノイズ成分となる。この問題に対して、 $\{0, 1\}$  の相互相関型連想記憶モデルにおいてパターン分離機能を高めるように、入出力の中間に設けたパターン分離層で発火率を一定に制御するグローバルな抑制を施すモデルが提案さ

れ、その連想能力の理論解析がなされている（鳥岡, & 池田, 1989）。また、入出力2層の双方向連想記憶モデルの出力層に側方素子への抑制結合を持たせたモデルが提案されている（Kosko, 1989）。これは、出力層を中間層と見なして双方向の信号の流れを折り返してフィードフォワードと考えれば（入力 → 出力 → 入力）、上記の場合に相当する。但し、パターン分離機能を高める為の競合方法はそれぞれ異なる。この他、競合的な中間層を持つ3層ネットワークとその学習方法も提案されている（坂口, & 村田, 1990）。一般に、側抑制は境界部の鮮鋭化を行う高域強調周波数フィルターとしての機能を有し（塚原, 1986）、勝者が他の全てを抑圧するよう Winner-Take-All (WTA) 的にパターン分離機能を高める効果を持つ。

最後に、入出力2層の連想記憶モデルにおける、結合重み係数の設定方法や学習方法に関して説明する。入出力層間の結合重み係数の設定方法としては、記憶パターンの相関行列を用いるものと一般化逆行列を用いるものが提案されており、前者より後者の方が連想能力が高い（Kohonen, 1977; Personatz, Guyon, & Dreyfus, 1986）。相関行列型モデルでは、前述のクロストークノイズが記憶パターンの想起に悪影響を及ぼすが、一般化逆行列型モデルでは、見かけ上は記憶パターンが互いに直交化するのでこれがなくなる。例えば、対称相互結合型の連想記憶モデル（Hopfield, 1982）において、相関行列型モデルでは記憶容量の限界（ $\approx 0.14N$ ）を越えると  $O(\sqrt{N})$  のクロストークノイズの悪影響で疑似記憶パターン（記憶パターン以外の  $\{-1, 1\}$  の2値表現されたパターンで、漸近安定平衡なもの）にトラップされる。これに対して、一般化逆行列型モデルでは記憶パターンが互いに線形独立である限り、正しい想起が可能で記憶容量は1.0となる（Geszt, 1990）。一般化逆行列型モデルのこれらの利点の反面、非局所的な逆行列の計算を必要とすることや、記憶容量が多い分だけ記憶パターンと類似したものを引き込む引力圏（ベイスン）が縮退化してしまうという欠点もある。

一方、連想記憶モデルの結合重み係数を相関行列や一般化逆行列で設定するのではなく、反復的な学習によって獲得する方法も提案されていて、甘利（1978）によって統一的に扱われた学習理論がある。それは、結合重み係数  $W_{ij}$  が入力信号  $x_j$  と学習信号  $r_i$  の積に比例して増加するという Hebb の仮説に基本的には基づくものである。選ばれる学習サンプルの定常環境についてのアンサンブル平均（記号  $\langle \rangle$  で表す）を用いて、以下のような学習方程式を考える。この学習方程式の動作は、損失関数  $L_i(W)$  の確率的降下法に従うように定義されている。

$$\dot{W}_{ij} = -W_{ij} + C \langle r_i \cdot x_j \rangle, \quad (2.9)$$

$$\tau \dot{W}_{ij} = -\frac{\partial L_i}{\partial W_{ij}} = -\left\langle \frac{\partial R_i}{\partial W_{ij}} \right\rangle. \quad (2.10)$$

ここで、 $C$  は学習の効率を表す比例係数で、 $\dot{W}_{ij}$  は  $W_{ij}$  の時間微分を表す。また、学習のポテンシャル関数  $R_i(W(t), x(t), y(t))$  は、教師信号  $y_i$  と入力信号の荷重和  $v_i \equiv \sum_j W_{ij} x_j$  との関数

としての学習信号  $r_i(v_i, y_i)$  を用いた、

$$R_i \equiv \frac{1}{2} \sum_j W_{ij}^2 - C \int_0^{v_i} r_i(u, y_i) du. \quad (2.11)$$

である。この時、学習信号  $r_i$  を  $r_i \equiv y_i$  あるいは  $r_i \equiv \sum_j W_{ij} x_j - y_i$  とすれば、損失関数の最小値を与える結合重み係数  $\{W_{ij}\}$  の学習の収束先は、それぞれ相関行列と一般化逆行列になる。

## 2.3 フィードバック連想記憶モデル

まず、2.2節で説明した、情報の符号表現、ネットワーク構造、結合重み係数の設定方法や学習方法の3点に関する連想能力の向上法が、松岡によって提案された連想記憶モデルによって関連付けられることを2.3.1で示す。次に、松岡のモデルの問題点を明らかにし、これを改善する為の改良モデルを提案する。2.3.2では、学習方程式を考える為、離散時間処理に基づく改良モデルを連続時間モデルに拡張し、認識処理と学習処理における挙動の漸近安定収束性を示す。また、この学習処理の物理的解釈を示すとともに、学習によって本提案モデルに線形主成分分析の機能が近似的に実現されることを理論的に示す。2.3.3では、本提案モデルの認識能力を実験的に検証するとともに、本提案モデルの学習結果と主成分分析を近似的に実現する従来モデルの学習結果との相違点の比較を行う。2.3.4では、上記の実験結果における誤認の原因を考察するとともに、本提案モデルを含んだ静的なモデルにおける、静的な空間分割に基づくパターン分離の本質的問題点を指摘する。

### 2.3.1 能力向上法の関連性と改良モデルの提案

$M$  個の認識カテゴリーを代表する  $N$  次元の記憶パターンベクトル  $x^{(i)}$  にノイズの加わったパターン  $x \in R^N$  が入力された時、真の記憶パターンを連想出力する自己想起の問題を考える。

$$X = [x^{(1)}, x^{(2)}, \dots, x^{(M)}], \quad (2.12)$$

$$x^{(i)} = (x_1^{(i)}, x_2^{(i)}, \dots, x_N^{(i)})^T \in R^N. \quad (2.13)$$

ここで、 $X$  は  $N$  次元パターンベクトル  $x^{(i)}$  によって定義される行列で、 $T$  は転置行列を表す。Kohonen (1977) は、 $x^{(i)}$ , ( $i = 1 \sim M$ ) の張る部分空間  $\Pi_s$  を考え、 $x$  の  $\Pi_s$  上への直交射影  $u = XX^+x$  ( $+$  は一般化逆行列を表す) を出力する方法を提案した。これは、線形変換  $F$  のうち行列ノルム  $\|FX - X\|$  を最小にするものであることから、最適連想と呼ばれている。一方、上記の問題点である非局所的な逆行列の計算を避ける為、松岡 (1989, 1990) はハード化に有利な図2.4のような中間層に相互抑制結合を持つ3層モデルで、以下のように一般化逆行列を近似的に実現する方法を提案している。

$$\tau \dot{y}(t) = -X^T X h(y(t)) + X^T x, \quad (2.14)$$

$$z(t) = X h(y(t)), \quad (2.15)$$

$$h(y_i(t)) = \begin{cases} y_i(t) & (y_i(t) > 0) \\ 0 & (y_i(t) \leq 0) \end{cases}. \quad (2.16)$$

ここで、 $y(t)$ 、 $z(t)$  は中間層と出力層の出力パターンベクトルをそれぞれ表し、 $h(y)$  はベクトル  $y$  の各要素  $y_i$  ごとに施される準線形関数である。式 (2.14) の  $\dot{y}(\infty) = 0$  における  $h(y(\infty)) = (X^T X)^+ X^T x$ 、 $z = X h(y(\infty))$  より、このモデルでは中間層における競合処理で一般化逆行列を局所演算のみで求めている。さらに、この競合処理によって、中間層素子の出力パターンはスパースな符号表現（極端な場合、唯一の素子のみが発火）になると考えられる。このように、前節で紹介した種々の能力向上法は、松岡の連想記憶モデルによって互いに関連付けられる。

さて、 $m$  番目の記憶パターン  $x^{(m)}$  にノイズ  $n$  が混じった  $x^{(m)} + n$  を入力した場合を考えよう。この時、松岡のモデルにおける認識出力に相当する中間層の出力  $y(\infty)$  は、

$$y(\infty) \approx (X^T X)^+ X^T (x^{(m)} + n) = e^{(m)} + (X^T X)^+ X^T n. \quad (2.17)$$

となって、式 (2.17) の第2項で示されるノイズの直交射影成分が、想起出力に悪影響を及ぼす（ $e^{(m)}$  は第  $m$  成分のみが1で残りが0の  $M$  次元単位ベクトルを表す）。仮に、想起出力  $z$  を入力層にフィードバックしても、関数  $h(y_i)$  は中間層素子の出力の負の成分を抑える以外は基本的には線形関数なので、ノイズ成分はあいかわらず残ってしまう。このように、松岡のモデルは基本的に線形連想記憶モデルなので、ノイズ等の変動成分を含んだパターン識別には適していない。文献 (Matsuoka, 1990) では、階段関数を持つ中間層素子からなるフィードバック連想記憶モデルも提案されているが、許容出来るノイズ量を調節する機構がない。

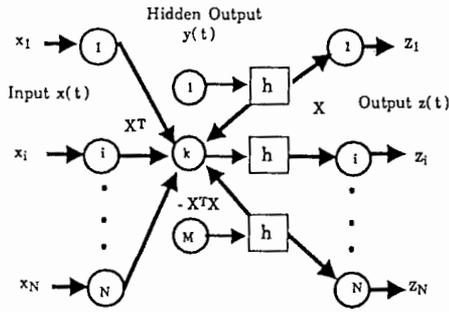
これらの問題点に対して、以下のような離散時間の改良モデルを提案する。但し、正しい認識出力は、そのカテゴリーを表す1つの中間層素子のみの発火状態（値が1）で表現すると仮定する。中間層素子数がカテゴリー数より多い場合 ( $K > M$ ) は、各カテゴリーを中間層素子の複数の発火状態で表現することが出来るが、本研究ではこの場合は考えない。

[離散時間のフィードバック連想記憶モデル]

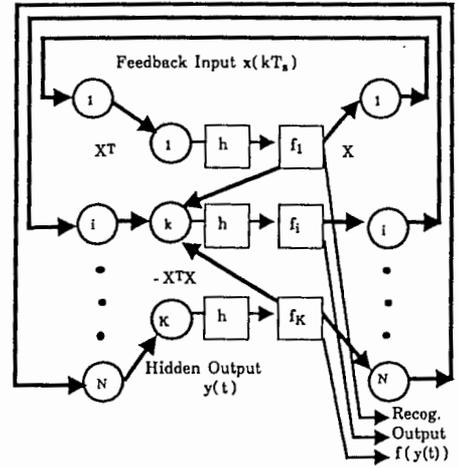
$$\tau \dot{y} = -X^T X h(y(t)) + X^T x(t), \quad (2.18)$$

$$x(t) = \begin{cases} X f(h(y(kT_s))) & (kT_s \leq t < (k+1)T_s, \quad k = 1, 2, 3, \dots) \\ x & (0 \leq t < T_s) \end{cases}, \quad (2.19)$$

$$y(0) = 0. \quad (2.20)$$



(a) 松岡の直交射影型連想記憶モデル



(b) フィードバック連想記憶モデル

図 2.4: 松岡の連想記憶モデルと本提案のフィードバック連想記憶モデル (但し、中間層素子間の結合は  $k$  番目の素子に関するものだけを図示した)

ここで、各中間層素子の非線形関数  $f_i(y_i)$  は図 2.5 のように、しきい値  $\theta_i$  を境にして値 0 と 1 が安定点となるよう、例えば以下のような関数を想定する。

$$f_i(y_i) = \begin{cases} 0 & (y_i \leq 0) \\ y_i^2/\theta_i & (0 < y_i \leq \theta_i) \\ -(y_i - 1)^2/(1 - \theta_i) + 1 & (\theta_i < y_i < 1) \\ 1 & (y_i \geq 1) \end{cases} \quad (2.21)$$

このモデルの動作は 2 段階である。すなわち、式 (2.18) の競合的な中間層の相互抑制による一般化逆行列の近似計算と、式 (2.19) の離散的なサンプル時間  $T_s (> \tau)$  ごとのフィードバックによる動的な連想を繰り返して入力パターンを再現する。このモデルにノイズが加わったパターン  $x^{(m)} + n$  を入力すると、

$$y(T_s) \approx e^{(m)} + (X^T X)^+ X^T n, \quad (2.22)$$

$$y((k+1)T_s) \approx (X^T X)^+ X^T x(t) \quad (2.23)$$

$$= (X^T X)^+ X^T X f(h(y(kT_s))) \quad (2.24)$$

$$= f(y(kT_s)). \quad (2.25)$$

となり、フィードバックによる動的な連想を繰り返すことで、図 2.5 のように、 $y_i(\infty)$  は値 0 か 1 のどちらかの安定平衡点に収束する。これによって、式 (2.19) に従った記憶パターンで再構成される再現パターンにおいては、もとの入力の記憶パターンに対するノイズ成分が段々除去されることになる。ノイズの許容範囲は、しきい値  $\theta_i$  で決められる。

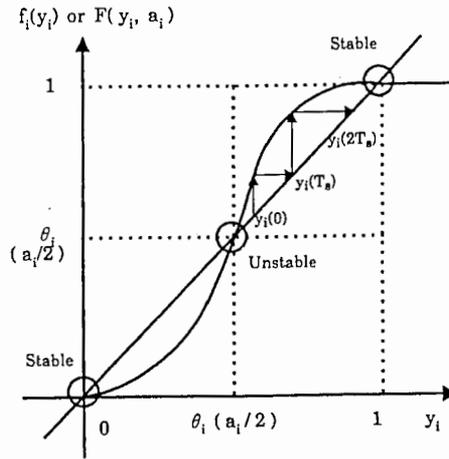


図 2.5: 中間層素子の非線形関数

### 2.3.2 認識と学習処理に関する理論的考察

上記の離散時間の改良モデルを連続時間モデルに拡張して、学習方程式を考える。さらに、以下の微分方程式で定義される認識処理と学習処理の挙動について、理論的な検討と物理的な解釈を加える。

[連続時間のフィードバック連想記憶モデル]

$$\tau \dot{y}(t) = -W^T W f(y(t)) + W^T g(x(t)) \quad (2.26)$$

$$= -f(y(t)) + R f(y(t)) + W^T g(x(t)), \quad (2.27)$$

$$\tau_x \dot{x}(t) = -x(t) + W f(y(t)), \quad (2.28)$$

$$R \equiv -W^T W + I_M. \quad (2.29)$$

ここで、 $g_i(x_i)$  と  $f_j(y_j)$  は、式 (2.21) で定義された関数や S 字関数のように有界で微分可能な単調増加関数である。また、 $\tau_x > \tau > 0$  は時定数、 $I_M$  は  $M$  次元単位行列、 $W$  はこれまで  $X$  と表記されてきた認識辞書に相当する結合重み行列を表す (以後は  $W$  と表記する)。すると、式 (2.18) と式 (2.19) から、しきい値と結合重みを固定した認識処理の挙動は、以下に定義される有界な Lyapunov 関数  $L(x, y)$  の極小値へ漸近安定収束する (付録 A 参照)。但し、収束先の平衡点は初期状態  $x(0), y(0)$  に依存する。

[Lyapunov 関数]

$$L(x, y) = -\sum_{i,j} W_{ij} g_i(x_i) f_j(y_j) - \frac{1}{2} \sum_{k,j} f_k(y_k) f_j(y_j) + \sum_i \int_0^{x_i} g'_i(\theta_i) \theta_i d\theta_i + \frac{1}{2} \sum_j f_j(y_j)^2. \quad (2.30)$$

学習方程式を考える為に、中間層素子の非線形関数  $f_j(y_j)$  を以下のような関数  $F(y_j, a_j)$  で再定義して、素子の内部状態の活性度に関する動作方程式を書き改める。

[動作方程式]

$$\tau \dot{y}(t) = -F(y(t), a(t)) + RF(y(t), a(t)) + W^T g(x(t)), \quad (2.31)$$

$$\tau_x \dot{x}(t) = -x(t) + WF(y(t), a(t)), \quad (2.32)$$

$$F(y_j, a_j) = \begin{cases} 0 & (y_j \leq 0) \\ 2y_j^2/a_j & (0 < y_j \leq a_j/2) \\ -2(y_j - 1)^2/(2 - a_j) + 1 & (a_j/2 < y_j < 1) \\ 1 & (y_j \geq 1) \end{cases} \quad (2.33)$$

ここで、 $a_j \equiv 2\theta_j$ , ( $0 < a_j < 2$ ) である。このとき、関数  $F(y_j, a_j)$  は次の性質を持つ。

性質 1 安定平衡点  $F(a_j, 0) = 0$ ,  $F(a_j, 1) = 1$  と、不安定平衡点  $F(a_j, a_j/2) = a_j/2$  を持ち、

性質 2  $a_j, y_j$  の両変数に対して連続微分可能で、 $F_y \equiv \frac{\partial F}{\partial y_j} \geq 0$ ,  $F_a \equiv \frac{\partial F}{\partial a_j} \leq 0$  となる。

[学習方程式]

$$\tau_a \dot{a}(t) = -F_a(y(t), a(t))(\tau \dot{y}(t)), \quad (2.34)$$

$$\tau_w \dot{W}(t) = -\gamma W(t) + g(x(t))F(y(t), a(t))^T - W(t)F(y(t), a(t))F(y(t), a(t))^T. \quad (2.35)$$

認識処理の場合と同様に、学習処理の挙動は以下の Lyapunov 関数  $\mathcal{L}(x, y, W, a)$  の極小値へ漸近安定収束する（付録 A 参照）。収束先は初期状態  $x(0), y(0), W(0), a(0)$  に依存する。

[Lyapunov 関数]

$$\begin{aligned} \mathcal{L}(x, y, W, a) = & -\sum_{i,j} W_{ij} g_i(x_i) F(y_j, a_j) - \frac{1}{2} \sum_{k,j} F(y_k, a_k) F(y_j, a_j) \\ & + \sum_i \int_0^{x_i} g'_i(\theta_i) \theta_i d\theta_i + \frac{1}{2} \sum_j F(y_j, a_j) + \frac{\gamma}{2} \sum_{i,j} W_{ij}^2. \end{aligned} \quad (2.36)$$

上記の学習方程式の物理的解釈は以下のようなものである。しきい値の学習方程式 (2.34) では、性質 2 の  $F_a(y_j, a_j) \leq 0$  によって、しきい値の時間的変化は中間層素子の内部状態  $y_j$  の時間的変化と反対方向に働く。すなわち、認識出力を大きくするように発火（内部状態が活性化）する素子では、そのしきい値が下がってより発火しやすくなり、発火が抑制される素子では、しきい値が上がってより発火しにくくなる。一方、結合重み係数の学習方程式 (2.35) では、第 1 項は結合重み係数（認識辞書）の忘却項、第 2 項は従来から基本的な学習規則と考えられている Hebb 型の学習項、第 3 項は異なる認識カテゴリーの素子間での競合を表す相互抑制項である。

以上は、ある 1 つの入力パターンに関する学習であるが、通常のパターン認識等の課題では複数のパターンを学習しなければならないことが多い。この場合も、上記の学習方程式 (2.34) と (2.35) で各入力パターンに対する結合重み係数やしきい値の修正を繰り返すと、確率的降下法に従って式 (2.36) の Lyapunov 関数の入力環境についてのアンサンブル平均  $\langle \mathcal{L}(x, y, W, a) \rangle$  が最小化（極小化）される。しかし、類似したカテゴリーにおける入力パターンを正しく認識す

る為には、モデル自身がパターンの分離基準を持つ上記のような自己組織化学習より、外部から強制的に類似パターンを分離させるように学習した方が良い。そこで、以下のように誤差最小化関数を式 (2.36) の Lyapunov 関数に付加して、各入力に対して少しずつ修正を行うと、以下の  $\langle \mathcal{L}^* \rangle$  を最小化 (極小化) するように学習が行われる。

$$\mathcal{L}^* = \mathcal{L} + \frac{\beta}{2} \sum_j \{T_j - F(y_j, a_j)\}^2. \quad (2.37)$$

但し、 $\beta > 0$  は強制学習の強さを表す係数で、 $T_j$  は正しい認識出力 (教師信号) とする。この時、上記の動作方程式 (2.31) (2.32) と学習方程式 (2.34) (2.35) のうち、式 (2.31) のみが以下のように変更される。

$$\tau \dot{y}_j(t) = -F(y_j, a_j) + \sum_i W_{ij} g_i(x_i) + \sum_k R_{kj} F(y_k, a_k) + \beta \{T_j - F(y_j, a_j)\}. \quad (2.38)$$

さて、本提案モデルの学習における非線形処理を線形化して、線形主成分分析との関連性を検討しよう。中間層素子が線形の場合、式 (2.36) の Lyapunov 関数の物理的意味は、

$$\mathcal{L}(x, y, W, a) = -\sum_{i,j} W_{ij} x_i y_j - \frac{1}{2} \sum_{k,j} y_k y_j + \sum_i \int_0^{x_i} \theta_i d\theta_i + \frac{1}{2} \sum_j y_j^2 + \frac{\gamma}{2} \sum_{i,j} W_{ij}^2 \quad (2.39)$$

$$= \frac{1}{2} \sum_i (x_i - \sum_j W_{ij} y_j)^2 + \frac{\gamma}{2} \sum_{i,j} W_{ij}^2. \quad (2.40)$$

から以下のように解釈出来る。上式の右辺第1項は入力パターンとそれによって想起される再現パターンとの自乗誤差項、第2項は結合重み係数の値の発散を防ぐ制約項で、過剰学習を防ぐ汎化作用の働きをするものと考えられる。

同様に、中間層素子が線形の場合、結合重み係数の学習方程式 (2.35) は、

$$\tau_w \dot{W}_{ij} = -\gamma W_{ij} + x_i y_j - \sum_s W_{is} y_s y_j \quad (2.41)$$

$$= -\gamma W_{ij} + [x_i - \sum_s W_{is} y_s] y_j. \quad (2.42)$$

となり、主成分を抽出する Sanger (1989) の線形モデルにおける学習方程式と非常に類似する。しかし、Sanger のモデルでは、式 (2.42) の右辺第3項の下三角行列成分のみを利用して Gram-Schmidt の直交化処理で固有値の大きい主成分から順に抽出しているのに対して、式 (2.42) による本提案モデルの学習にはこの直交化処理に相当する部分がない。

Sanger のモデルとは別に、入力パターンを出力層に再現する恒等写像を近似する3層ネットで、主成分分析が近似的に実現可能であることが理論的に示されている (船橋, 1990)。また、実際的にも階層型ネットワークの BP 学習で、中間層に主成分の線形結合が現れる。そこで、学習方程式 (2.42) と BP 学習との関連性を見てみる。式 (2.42) において、結合重み係数の忘却 (忘却係数  $\gamma$ ) は非常に小さいと考えられるので、結合重み係数の修正量  $\Delta W$  は、

$$\Delta W \approx \eta [x - Wy] y^T = \eta \delta y^T = -\partial \text{Error} / \partial W. \quad (2.43)$$

となり、これは3層ネットの中間-出力層間の結合重み係数に対するBP学習に他ならない。ここで、 $\eta$ は学習係数で、 $Error$ は入力と再現入力との自乗誤差： $Error \equiv \delta^T \delta / 2$  ( $\delta \equiv [x - Wy]$ )とする。

次に、 $\tau_x > \tau$ より、本提案モデルは近似的にフィードフォワードの3層ネットとして扱うことが出来る。この時、入力-中間層間の結合重み係数  $W1$  のBP学習による修正量  $\Delta W1$  は、 $y = W1x \approx (W^T W)^+ W^T x$ を用いて、

$$\Delta W1 = \eta W^T \delta x^T n = \eta W^T [x - Wy] x^T \quad (2.44)$$

$$\approx \eta W^T [x - W(W^T W)^+ W^T x] x^T \quad (2.45)$$

$$= \eta \{W^T - W^T W(W^T W)^+ W^T\} x x^T = 0. \quad (2.46)$$

となる。すなわち、BP学習としての入力-中間層間の結合重み係数の修正量は常にゼロとなる。ゆえに中間層素子が線形の場合は、本提案モデルの学習は恒等写像近似のBP学習と近似的に等価と考えられる。しかし、非線形の場合は次節に示すように、Sangerのモデルによる学習や3層ネットのBP学習と本提案モデルの学習とが異なった結果を示すようになる。

さて、上記の3つの学習が関連性を持つことの根拠を考えてみよう。恒等写像近似のBP学習は、入力に関する情報量を最大に保持してこれを再現するという課題を実現する為のものと考えられる。これに対して、本提案モデルやSangerのモデルの学習における競合処理で、入力情報源についての有効な情報のみが獲得されて冗長なものは除去されるとしたら、これらの関連性がうなずける。すなわち、主成分の固有ベクトルで張られる部分空間が、入力再現に有効な情報に対応し(冗長な情報はその直交補空間に対応)、また、この部分空間への直交射影が一般化逆行列で行こなわれると考えれば良いだろう。このような情報量最大化原理と主成分分析との関連性や競合処理によるその実現は、Linsker (1988)のモデルにも見られる。

### 2.3.3 手書き文字による認識能力の実験的検証

本実験で用いる文字データと実験方法を説明した後、これらのデータを用いた学習後の認識実験の結果から、本提案モデルの認識能力を調べる。さらに、Sangerのモデルによる学習や3層ネットのBP学習と、本提案モデルの学習との相違点や誤認の原因についても考察を行う。

<実験データ>

本実験で用いた文字データとして、電総研が作成した手書き文字データベース ETL9 から、「い」、「け」、「に」、「は」、「ほ」、「ゆ」の6つのカテゴリーの類似文字を選び、各文字について学習用と認識用にそれぞれ50文字ずつ用意した。モデルへの入力パターンには、各文字の2値画像(64 x 63)に大きさの正規化(64 x 64)の前処理を施した後、4 x 4画素内の

濃度を特徴量とするメッシュ特徴 ( $16 \times 16 = 256$  次元) を用いた。但し、各入力素子は線形素子:  $g(x) = x$  とした。また、中間層の素子数はカテゴリ数と等しく 6 個とし、各カテゴリに対応する 1 個の中間層素子のみが発火 (出力値 1) することで、各カテゴリの認識出力を表現した。

### <実験方法>

本提案モデルの学習の妥当性と認識能力を調べる為、手書きひらがな類似文字の学習と認識の実験を行った (実験 1)。検討項目は、(i) 強制学習の有効性の確認、(ii) しきい値の初期値と認識率の関係、(iii) 学習後の認識能力の評価とした。また、Sanger のモデルによる学習や 3 層ネットの BP 学習についても実験を行い、本提案モデルとそれらの学習後の結合重み係数についての比較を行った (実験 2)。以下、それぞれの実験手順を説明する。

**実験 1** 各入力文字の特徴パターンに対して、式 (2.32) (2.34) (2.35) (2.38) に従って、本提案モデルの結合重み係数としきい値を徐々に修正することを繰り返す。この時、各文字の入力層への提示ごとに、各中間層素子の状態値をゼロに初期化する。但し、学習中に正しい認識出力が得られたら、この文字に対する学習を止めて次の文字を提示する。結合重み係数の初期値は、一様乱数によってランダムに設定し、しきい値の初期値は、はじめから素子が誤って発火しないようにある程度高い値 (0.8 程度) に設定した。

**実験 2** 主成分を抽出するために、各特徴パターンを平均値分離 (各パターンから全パターンの平均を引く操作を施す) した入力パターンで、3 つのモデルの学習を行い、それぞれの学習後の結合重み係数を比較する。ここで、Sanger の線形モデルは 256-6 次元の入出力 2 層構造とし、各入力の提示ごとの逐次学習で結合重み係数を修正した。一方、BP 学習を行う 3 層ネットは、256-6-256 次元で中間層以外は線形素子とし (中間層ではシグモイド関数:  $f(x) = 1/(1 + \exp^{-x})$  を施す)、一括修正学習 (各入力による修正量の平均で一括修正する方法) で結合重み係数を修正した。

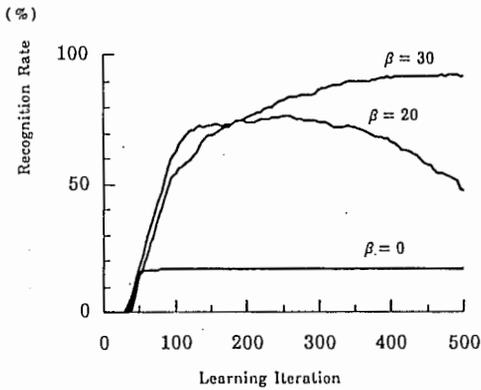
本提案モデルの学習に関する各パラメータは、 $\gamma = 0.01$ ,  $\tau_x = 125$ ,  $\tau = 100$ ,  $\tau_a = 100,000$ ,  $\tau_w = 10,000$  とした。これらのパラメータは、一般化逆行列を近似するための条件 ( $\tau_x \ll \tau_w$ ) や、認識率向上のため、識別境界の調整に大きく関与すると考えられるしきい値の変化を結合重み係数の変化より遅くする為の条件 ( $\tau_w < \tau_a$ ) を満足するものである。また、Sanger のモデルにおいて学習係数を  $\gamma_n = 0.01 \times (0.85)^n$  ( $n$  は学習の反復回数) とし、3 層ネットにおいて学習係数を  $\eta = 0.2$ 、学習の加速係数を  $\alpha = 0.9$  とした。

	い	け	に	は	ほ	ゆ	
学習用	92	96	100	100	80	98	94.3
評価用	82	74	92	82	70	96	82.7

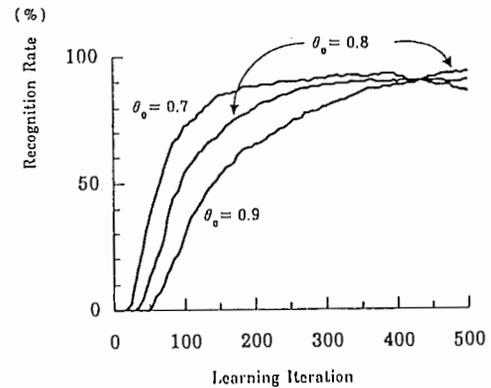
表 2.1: 本提案モデルの学習後の認識率 (%)

	い	け	に	は	ほ	ゆ	
学習用	98	84	86	46	90	100	84.0
評価用	92	64	82	30	78	98	74.0

表 2.2: 単純類似度による認識率 (%)



(a) 強制学習による違い ( $\theta_0 = 0.8$ )



(b) しきい値の初期値による違い ( $\beta = 30$ )

図 2.6: 学習の反復回数に対する認識率の変化

### <実験 1 の結果>

以下に、実験 1 の結果を示す。図 2.6(a)(b) は、学習の反復回数に対する 6 文字のカテゴリの平均認識率の変化を示すもので、それぞれ、強制学習の強さ ( $\beta$ ) としきい値の初期値 ( $\theta_0$ ) に関する比較を行う為のものである。

図 2.6(a) は、本提案モデルの強制学習によって誤学習が回避できることを示している。すなわち、強制学習を行わない場合 ( $\beta = 0$ ) は、入力パターンが互いに類似しているために全てのパターンを 1 つのカテゴリとして学習してしまい、認識率が向上しなかった。また、強制学習の影響はある程度大きくした方 ( $\beta = 30$ ) が良い結果となった。

図 2.6(b) は、最も高い認識率を得るには、しきい値の初期値を低くしすぎても、高くしすぎてもいけないことを示している。すなわち、しきい値の初期値を低くしすぎると誤ったカテゴリの素子が発火しやすくなって誤認を招き、高くしすぎると低い認識率から出発するために学習に時間がかかると考えられる。このため、適度な値 ( $\theta_0 = 0.8$ ) が最も良かったと言える。

さて、比較的高い認識率が得られたパラメータ値における学習後 (図 2.6(a) の  $\beta = 30$  で 500 回学習後) の認識結果 (表 2.3.3) と、従来の単純類似度による認識結果 (表 2.3.3) とを比較しよう。ここで、単純類似度による認識とは、入力パターンと各辞書パターンとのベクトル内積を両ベクトルの大きさで割った類似度のうち、最も高い類似度のカテゴリを認識結果とするものである。但し、類似度が次に高い第 2 候補との差がほとんどない場合 (0.05 以内) は、不確定な結果とした。表から、学習用と評価用の両データに対するほとんど全てのカテゴリで単純類似度の結果を上回り、カテゴリ間における認識率のばらつきも小さいことがわかる。

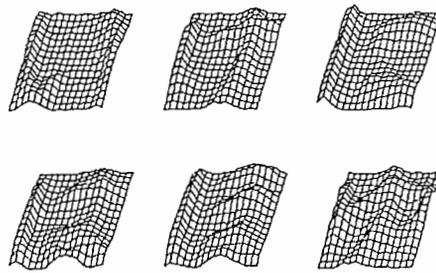
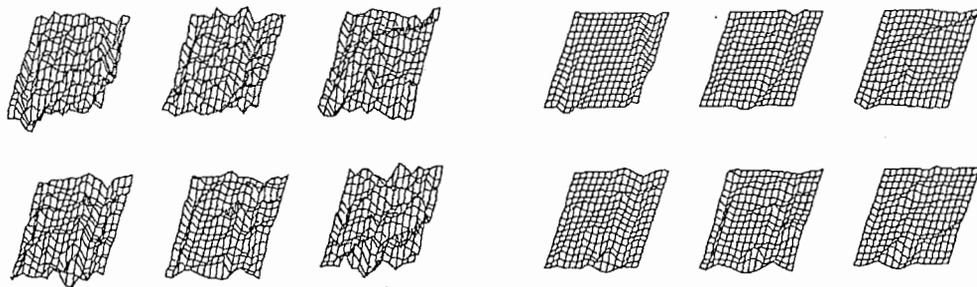


図 2.7: 学習後の各辞書パターン



(a) ランダム初期値

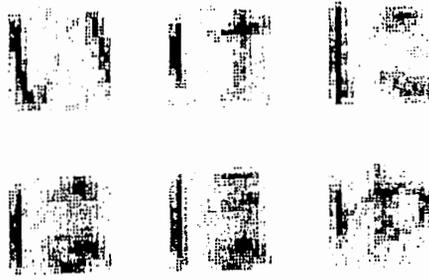
(b) 平均初期値

図 2.8: 学習後の辞書と各平均との差分パターン

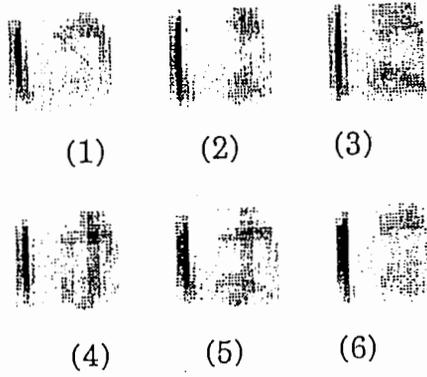
この時の、認識辞書パターンである結合重み係数行列  $W'$  の各列ベクトルを図 2.7 に示す。また、各カテゴリーの平均パターンと各辞書パターンとの差分パターンを、鳥瞰図的に図 2.8 に示す。各辞書パターンにおいて、平均パターンとの差分量は非常に小さく（差分を強調するために、図 2.8 では図 2.7 の 6 倍のスケールで表示している）、各辞書パターンは該当するカテゴリーの平均パターン（図 2.7）付近に収束しているのがわかる。結合重み係数をランダム初期値から始めた場合に得られた学習後の差分量（図 2.8(a)）と、各カテゴリーの平均パターンから始めた場合に得られた学習後の差分量（図 2.8(b)）を見ると、両者とも平均パターンとは異なった線分要素のストロークの位置や方向を強調する山や谷ができています。これは、誤認しやすい平均パターンとストロークの位置や方向が異なるようなパターンをより強調するように、平均から若干移動したパターンに収束したことを示している。

#### <実験 2 の結果>

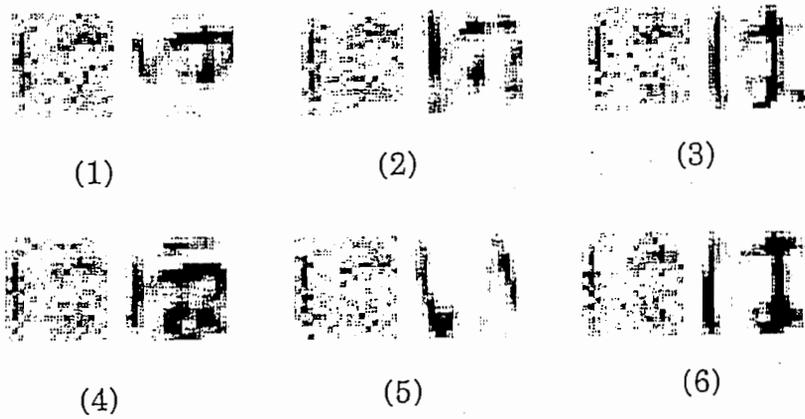
本提案モデルの学習後の実験結果と、Sanger のモデルの学習後や 3 層ネットの BP 学習後の実験結果との比較を行う。本提案モデルの学習後の辞書、Sanger のモデルの学習後の辞書、3 層ネットの BP 学習後の辞書を、それぞれ図 2.9(a)(b)(c) に示す。但し、実験 1 との対応付けから、平均値分離の入力パターンを扱ったことを考慮して、学習後の結合重み係数行列の各列ベクトルに学習用の全文字パターンの平均を加えたものを各カテゴリーの辞書と呼んでいる。図中の過去付き番号は抽出された 6 個の主成分を示すものである。また、BP 学習に関する図 2.9(c) の各番号における左右のものは、入力-中間層間の結合重み係数と中間-出力層間の結合重み係数をそれぞれ表している。図から、Sanger のモデルの学習と 3 層ネットの BP 学習では、複数のカテゴリーの成分が混じり合った分散的な辞書が得られたのに対して、本提案モデルの学習では、中間層の非線形関数と強制学習の影響で各カテゴリーごとの局所的な辞書が得られたという、両者の相違点が明らかとなった。なお、明示的に示していないが、本提案モデルで強制学習を行わない場合 ( $\beta = 0$ ) では、BP 学習と同様の分散的な辞書が実験で得られた。



(a) 本提案モデルの学習後



(b) Sanger のモデルの学習後



(c) 3層ネットのBP学習後

図 2.9: 平均値分離入力 of 学習後の辞書パターン

以上の実験結果から、本提案モデルの学習に関する以下の性質が明らかになった。

1. 類似性に基づく自己組織的な学習に、強制的なカテゴリー分離学習項を付加することで、誤学習を防ぐことが出来た
2. 学習で獲得された辞書は、各カテゴリーの平均パターン付近（変形の強いパターンを補うように若干ずれる）に収束した
3. 本提案モデルの学習と、これと類似した主成分抽出モデルの学習との比較で、前者はカテゴリーごとの局所的な辞書が、後者は複数カテゴリーによる分散的な辞書が得られた
4. 学習後の未学習文字の認識率は約 83%（学習文字では約 94%）であり、誤認の主要因は文字ストロークの位置ずれや傾きであった

#### 2.3.4 誤認の傾向と静的な空間分割における問題点

表 2.3.3における誤認文字を調べた結果、主な原因は以下の 3 つの場合に分けられた。

1. 極端に変形した入力文字のために、どのカテゴリーの辞書ともマッチせずに、すべての認識出力がゼロになってしまう場合
2. 文字ストロークの位置ずれや傾きが大きくて、誤ったカテゴリーの辞書と部分的によくマッチして誤認する場合
3. 非常によく似たカテゴリー間での、文字ストロークの部分的な一致によって、複数のカテゴリーを示す認識出力となってしまう場合

いずれにしても、入力文字の辞書からの変形が大きい為に、誤った認識出力となってしまう。これらの誤認に対する対策としては、文字ストロークの位置ずれや傾きが主要因であることから、重心分割による前処理で位置ずれを吸収したり、文字ストロークの方向性を考慮した特徴量を用いることが有効と考えられる。

以上の問題点において、安定平衡点に収束するまでの認識処理中で仮に誤認方向に変化しても、何らかの方法（例えば、前後関係から単語知識を用いて）で正しい方向に脱出できるメカニズムがあれば、誤認を回避できると考えられる。しかし、本提案モデルを含む静的なモデルでは、静的な空間分割に従ってパターン分離が行われるために、このような回避は不可能となる。例えば誤認を回避するために、入力パターンの分類を規定する辞書（結合重み係数）を変化させたとしても、新たな誤認領域が発生してしまい、安定な認識結果と辞書の学習との間の循環的なジレンマ（Carpenter, & Grossberg, 1988）を引き起こす。ゆえに、学習とは別のメカニズムで、外部入力などによって状態空間の構造を迅速に変化させることが必要となる。

これに対して、リミットサイクルなどの時間的な状態変化を伴うアトラクターを持つ動的なモデルでは、外部入力などで状態空間構造の迅速な変化を引き起こすことが出来る。しかし、動的なモデルにおいても、初期状態に従った静的な空間分割による入力パターンの分類を考えている限りでは、アトラクターの違いはカテゴリーを区別する為の指標となるのみで（平衡状態の値の違いと同様）、その軌道上の時間的な状態変化は情報処理に利用されないことになる。従って、状態空間構造の変化を考えるとともに、動的な出力軌道そのものを情報処理機能に対応付けることが重要となる。例えば、曖昧な入力に対して、複数のカテゴリーを含む大きなベイスン内を遊走する軌道を探索機能に、一方、確信度の高い入力に対して、前期の状態空間から構造変化を起こした新たな状態空間内の限られた範囲を動く軌道を出力の確定機能に対応付けられる。

一方、Hopfield 型連想記憶モデル (Hopfield, 1992, 1984) のような静的なモデルでは、平衡点に収束するまでの限られた軌跡上以外は探索機能がないので、上記のような機能的なスイッチングが出来ない。そればかりか、外部入力などで状態空間構造を変化させても、正しい出力にマッチする構造となるように、外部入力を微妙に調整する必要性が生じる。

## 2.4 おわりに

静的なモデルにおけるネットワーク構造と情報処理機能の関係に着目し、これまでに提案された情報表現やネットワーク構造、結合重み係数の設定方法や学習方法などに関する能力向上法が、主成分分析的なパターン分離機能の近似的実現の観点から、互いに関連付けられることを示した。また、上記の高い連想能力を継承しつつ、手書き文字等のパターン認識への応用性を考えて、文字の変動成分を吸収し柔軟な認識および学習を実現するための、新しいモデルを提案した。さらに、本提案モデルの認識処理と学習処理の妥当性を理論的に示した。本章で議論した異なるモデルの学習における類似性は、入力の情報量を最大限に保持するために、一般化逆行列によって主成分の固有ベクトルで張られる部分空間への射影が行われることに基づくと考えられる。その際、パターン分離としての主成分分析の機能を近似的に獲得する為の学習方法として、これらの複数の方法が存在すると言える（もちろん、この機能を実現出来るこれら以外の学習方法もあり得るだろう）。ネットワーク構造に関しては、一般化逆行列を入-出力2層の連想記憶モデルの結合重み係数に直接設定する方法や、3層ネットの中間層における競合処理で近似的に実現する方法、フィードフォワード型の3層ネットの結合重み係数中に（学習で獲得する / しなにかかわらず）設定された固有ベクトルの線形和で実現する方法などが、この機能に対する複数の実現方法と言える。いずれの方法においても、実現された機能は本質的には主成分分析によるパターン分離である。しかしながら、本章で指摘したように、こうしたパターン分離を支配する静的な空間分割は、誤認を柔軟に回避することが出来ないという本質的な問題点を持つ。

次章以降では、この問題点を解決する為の1つの手段として、動的な出力軌道そのものを情報処理機能に対応付けることで、軌道の多様性を時空間パターン処理に積極的に利用し、かつ、外部入力などに従った動的な出力軌道の性質の変化を、情報処理機能のスイッチングに利用することを提案する。さらに、これを実現するための具体的なモデルを提案し、そのダイナミクスの特徴について議論する。

## 第 3 章

### 発振ニューラルネットワークの動的挙動

#### 3.1 はじめに

静的なモデルの本質的機能は、静的な空間分割に従ったパターン分離機能である為に、変動成分の多い手書き文字の認識などの柔軟な情報処理を行うことが困難である。例えば、出力の誤り領域に対する分割構造（状態空間構造）を変えるために学習を行ったとしても、新たな誤り領域を生じてしまうことが多い。そればかりか、通常、学習は膨大な入力データの統計量に基づいて行われる為に、状態空間構造の変化を学習だけに頼っていたのでは、迅速な処理は望めない。

こうした静的な空間分割の問題点を解決する為の 1 つの手段として、本章以降では、リミットサイクルやカオスなどの時間変化を伴う出力軌道を持つ動的なモデルにおいて、

1. その動的な出力軌道を情報処理機能に対応付けて、軌道の多様性を時空間パターン処理に積極的に利用し、かつ、
2. 外部入力などに従った出力軌道の性質の変化を、情報処理機能のスイッチングに利用する

ことを考える。この考えを実現するには、動的なモデルが多様な軌道を持ち、その性質が外部からうまく制御出来ることが重要となる。

そこで、本章では、正と負の互いに反対符号の重み係数で結合された興奮 - 抑制性ペア素子からなる発振ニューラルネットワークを提案し、その挙動の特徴を明らかにする。まず、本提案モデルの特殊なネットワーク構造を利用して、限られた挙動についての理論解析を行う。これによって、結合重み係数の値に従った挙動として、記憶パターンに相当する  $N$  次元超立方体上の頂点付近の平衡点（以後、記憶点と呼ぶ）への漸近安定収束と、その記憶点が不安定化して発振を起し得ることが明らかにされる。さらに、本提案モデルの典型的な特徴である軌道の多様性と可制御性を数値解析によって示す。軌道の多様性としては以下の 3 点が示される。第 1 に、単一のペア素子の最大発振周波数を規定する結合重み係数を分岐パラメータとして、漸近安定平衡点、リミットサイクル、カオスの軌道などのさまざまなアトラクターを持つこと。第 2 に、素子

の初期状態に依存した複数のアトラクターが存在する、すなわち、状態空間はいくつかのベイスンで分割されること。第3に、本提案モデルの軌道が、状態空間内の頂点間を複雑に遊走することである。一方、軌道の可制御性として、記憶パターン付近のリミットサイクルから記憶パターン間を遊走するカオス的軌道などへ、軌道の遊走領域が外部入力に従って大まかに制御出来ることが示される。上記の2つの特徴的な軌道は、記憶パターンの想起（振動は想起の不確かさを示す揺らぎに相当）と、記憶パターンのカオス的な探索機能にそれぞれ対応付けられる。さらに、Hopfield型連想記憶モデル（Hopfield, 1982）における記憶点の引力圏の破綻（Amit, Gutfreund, & Sompolinsky, 1987a, 1987b; Geszti, 1990）と類似した現象として、記憶パターン数の増加に伴った遊走軌道の性質の変化も示される。次章では、これらの挙動の特徴を積極的に利用した具体例を通じて、本提案モデルの時空間パターン処理への応用可能性についての検討を行う。

### 3.2 発振ニューラルネットワーク

ニューラルネットワークで時空間パターンあるいは時系列パターンを生成するには、興奮性と抑制性の互いに反対符号の重み係数を持つ結合のような、非対称な結合が重要である（Amari, 1971, 1972a）。一方、これまでに提案された発振を起こすタイプのモデルは、以下の5つに大きく分けることが出来る。

1. 興奮 - 抑制性の相互結合を持つネットワーク（Amari, 1972a, 1972b; Baird, 1986; Lie, & Hopfield, 1989; Yao, & Freeman, 1990; Grossberg, & Somers, 1991）
2. 非対称結合を持つ（相互結合型などの）連想記憶モデル（Mori, Davis, & Nara, 1989; Wang, Pichler, & Ross, 1990; Renals, & Rohwer, 1990; Tsuda, 1992）
3. 動的なしきい値や順応としての時間遅れ（time delay）を持った、興奮 - 抑制性結合のネットワーク（Matsuoka, 1985, 1987; Eckhorn et al., 1990; Wang, Buhmann, & Malsburg, 1990; Horn, & Usher, 1991a; König, & Schillen, 1991）
4. 位相方程式による非線形結合振動子（Sompolinsky, Golomb, & Kleinfeld, 1991; Niebur, Kammen, & Koch, 1991; Kuramoto, 1991; Lummer, & Huberman, 1992）
5. Hodgkin-Huxley 方程式などに基づくカオスニューロンモデルのネットワーク（Aihara, Matsumoto, & Ikegaya, 1984; Aihara, 1990a, 1990b; Abbott, 1990）

これらのモデルは複雑な非線形ダイナミクスを持つため、その挙動の理論解析は一般的には困難である。しかしながら、甘利による初期の研究 (Amari, 1971, 1972a) は、興奮 - 抑制性ペア素子の相互作用による発振の創発を指摘し、離散情報モデルと連続情報モデルのそれぞれの発振条件を理論的に解析している。さらに、離散情報モデル (しきい素子のネットワーク) においては、平衡点および周期的な状態推移 (リミットサイクル) の安定性の理論解析も報告されている (Amari, 1972b)。

上記の示唆的研究に従って、以下では興奮 - 抑制性ペア素子で構成される発振ニューラルネットワークを考える。このネットワークは、それぞれ  $N$  個ずつの興奮性素子と抑制性素子を持つ。各興奮性素子はペアをなす1つの抑制性素子と結合するとともに、他の興奮性素子とも相互結合する。一方、各抑制性素子是对応する唯一の興奮性素子のみと結合する点が、このネットワークの特徴である。ここで、各素子の値は単一のニューロンの活性度というよりは、多数のニューロン集団の活性度の平均場近似 (Amari 1972a; Wilson, & Cowan, 1973; Schuster, & Wagner, 1990) と解釈できる。そのネットワーク構造を図 3.1に、動作方程式を以下に示す。

[発振ニューラルネットワーク]

$$\dot{x}_i = -x_i + G \left( \sum_{j=1}^N W_{ij} x_j - K_{EI}^{(i)} y_i + I_i \right), \quad (3.1)$$

$$\dot{y}_i = -y_i + G \left( K_{IE}^{(i)} x_i \right), \quad (3.2)$$

$$G(z) = \frac{2}{\pi} \arctan \left( \frac{z}{a} \right), \quad (3.3)$$

$$W_{ij} = \frac{1}{N} \sum_{\alpha=1}^M \xi_i^\alpha \xi_j^\alpha + \delta_{ij} \quad (N > M). \quad (3.4)$$

ここで、 $x_i$  と  $y_i$  はそれぞれ興奮性素子と抑制性素子の平均パルス密度を表し、 $G(z)$  は平均膜電位を平均パルス密度に変換するための非線形の S 字関数で、 $a$  はその急峻さを表すパラメータである。第  $i$  番目と  $j$  番目の興奮性素子間の相互結合の重み係数  $W_{ij}$  は、 $M$  個の記憶パターン  $\xi^\alpha$  ( $\alpha = 1 \sim M$ ) の自己相関行列である式 (3.4) で定義される。但し、記憶パターン  $\xi^\alpha$  の各要素は  $\pm 1$  の 2 値をとるものとし、 $\delta_{ij}$  は Kronecker のデルタを表す。自己相関行列の性質から、 $W_{ii} = (N + M)/N$ ,  $W_{ij} = W_{ji}$  となる。第  $i$  番目のペア素子における抑制性素子から興奮性素子への抑制性の結合重み係数を  $-K_{EI}^{(i)}$  で表し、興奮性素子から抑制性素子への興奮性の結合重み係数を  $K_{IE}^{(i)}$  で表す (肩の添え字  $i$  は素子番号を表す)。  $I_i$  は第  $i$  番目の興奮性素子への入力あるいは入力バイアスである。出力パターンは興奮性素子の活性値  $\{x_i(t)\}$  で与えられるものとする。

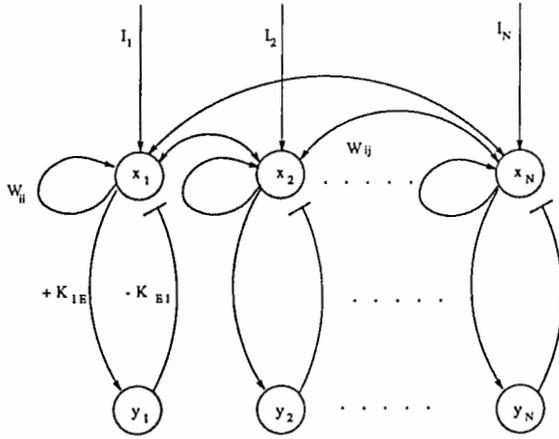


図 3.1: 発振ニューラルネットワーク (矢印  $\vdash$  と  $\rightarrow$  は興奮性結合と抑制性結合を表す)

本章の以下の議論では解析の単純さを考慮して、全てのペア素子が共通の結合重み係数  $K_{IE}$  と  $K_{EI}$  を持つものとして考える。しかしながら、4章の 4.2 節におけるネットワークの学習時では、結合重み係数  $K_{EI}^{(i)}$  を各素子ごとに異なる値を取り得るパラメータとする。

### 3.3 記憶点の不安定化による発振

前章で提案された発振ニューラルネットワークは比較的単純なネットワーク構造を持つが、それでも一般的な場合における理論解析は困難である。そこで、本節では状態空間内の記憶パターン付近の限られた挙動の理論解析を行い、次の 3.4 節と 3.5 節では 4 次の Runge-Kutta 法 (森, 名取, 鳥居, 1985) による数値計算から、本提案モデルの典型的な挙動を示す。

#### 3.3.1 記憶点の漸近安定性に関する摂動解析

ペア素子の結合を切って ( $K_{IE} = 0$ )、興奮性素子のみで構成されるネットワークを考えると、それは以下のような入力バイアスを持つ連想記憶モデルとなる。

$$\dot{x}_i = -x_i + G\left(\sum_{j=1}^N W_{ij}x_j + I_i\right) \equiv F_i(x_1, x_2, \dots, x_N). \quad (3.5)$$

このモデルが、記憶パターン  $\xi^\alpha$  付近に漸近安定平衡点  $\{\bar{x}_i\}$  を持つと仮定することは自然である。なぜなら、入力バイアス  $\{I_i\}$  をその記憶パターンに比例した大きな値で設定すれば、この仮定は常に成り立つ。この場合、平衡点  $\bar{x}_i \approx \pm 1$  における式 (3.5) の線形化方程式は負の実部の固有値を持つ。しかし、仮定における連想記憶モデルの平衡点の存在とその安定性を一般的な場合で解析することはそれ自体難しい問題で、それは、 $N$  個の多様体  $\dot{x}_i = 0$  の交わり方に依存する。この仮定のもとで結合重み係数  $K_{IE}$  の値が十分小さい時、発振ニューラルネットワークにも同様な漸近安定平衡点  $\bar{x}_i$  が存在することを以下の摂動解析によって示す。

もし、発振ニューラルネットワークに平衡点が存在すれば、それは Nullcline  $\dot{y}_i = 0$  上、すなわち、曲線  $y_i = G(K_{IE}x_i)$  上にある。十分小さい結合重み係数  $K_{IE} \equiv \varepsilon > 0$  より、これは、

$$y_i = G(K_{IE}x_i) = G(0) + \varepsilon x_i G'(0) + O(\varepsilon^2) = \frac{2x_i}{\pi a} \varepsilon + O(\varepsilon^2). \quad (3.6)$$

と表される。ここで、 $G'(z) = 2a/\pi(z^2 + a^2)$  は関数  $G(z)$  の  $z$  による空間微分である。Nullcline  $y_i \approx \frac{2x_i}{\pi a} \varepsilon$  においては、式 (3.1) は以下のように近似される。

$$\dot{x}_i = -x_i + G\left(\sum_{j=1}^N W_{ij}x_j - \frac{2K_{EI}\varepsilon}{\pi a}x_i + I_i\right) \equiv \tilde{F}_i(x_1, x_2, \dots, x_N). \quad (3.7)$$

これは式 (3.5) において、自己結合の重み係数  $W_{ii}$  をほんの少しだけ摂動させて、 $W'_{ii} = W_{ii} - \frac{2K_{EI}\varepsilon}{\pi a} + O(\varepsilon^2)$  としたものと等価である。さて、 $\tilde{F}_i(x_1, x_2, \dots, x_N) = 0$  の解である平衡点を考えよう。その平衡点  $\{\tilde{x}_i\}$  においては、以下の行列・ベクトル表記の摂動展開が得られる。

$$[\tilde{F}_i]_{x_k=\tilde{x}_k} = [F_i]_{x_k=\tilde{x}_k} + [\partial F_i/\partial x_j]_{x_k=\tilde{x}_k} \delta x + \text{diag}[\partial F_i/\partial W_{ii}] \delta W_{ii} + (\text{higher order}). \quad (3.8)$$

仮定によって、上記の右辺第1項  $[F_i]_{x_k=\tilde{x}_k}$  はゼロとなる。また、その Jacobi 行列が可逆であるとすれば、 $\delta x = -[\partial F_i/\partial x_j]_{x_k=\tilde{x}_k}^{-1} \text{diag}[\partial F_i/\partial W_{ii}]_{x_k=\tilde{x}_k} \delta W_{ii} \sim O(\varepsilon)$  だけ異なる平衡点  $\tilde{x} + \delta x$  が存在することになる。したがって、 $[\tilde{F}_i] = 0$  の解は存在し、それは元の式 (3.5) における平衡点をほんの少しだけ摂動した平衡点  $\tilde{x}_i = \bar{x}_i + \varepsilon x'_i + O(\varepsilon^2)$  となる。

次に、この平衡点  $\{\tilde{x}_i\}$  の安定性について議論する。発振ニューラルネットワークの平衡点  $\{\tilde{x}_i\}$  における線形化方程式の Jacobi 行列は、

$$\begin{bmatrix} -1 + W'_{ii}G'_{e1} & W_{12}G'_{e1} & \cdots & W_{1N}G'_{e1} \\ W_{21}G'_{e2} & -1 + W'_{ii}G'_{e2} & \cdots & W_{2N}G'_{e2} \\ \vdots & \vdots & \vdots & \vdots \\ W_{N1}G'_{eN} & W_{N2}G'_{eN} & \cdots & -1 + W'_{ii}G'_{eN} \end{bmatrix}. \quad (3.9)$$

となる。ここで、第  $i$  番目の素子の関数値  $G'_{ei}$  は、 $G'(W'_{ii}\tilde{x}_i + \sum_{j \neq i} W_{ij}\tilde{x}_j + I_i)$  である。さらに、

$$G'_{ei} = G'\left(\sum_{j=1}^N W_{ij}\tilde{x}_j + I_i\right) + \varepsilon\left(\sum_{j=1}^N W_{ij}x'_j - \frac{2K_{EI}\tilde{x}_i}{\pi a}\right) G''\left(\sum_{j=1}^N W_{ij}\tilde{x}_j + I_i\right) + O(\varepsilon^2), \quad (3.10)$$

より、Jacobi 行列は  $A + \varepsilon B + O(\varepsilon^2)$  という形となる。ここで、行列  $A$  はペア素子の結合を切った連想記憶モデルの平衡点  $\{\bar{x}_i\}$  における Jacobi 行列であり、仮定によってその固有値は負の実部を持つ。また、行列  $\varepsilon B$  は摂動行列である。

この時、 $0 < G''(z) \leq 1/\pi a^2$ ;  $-1 < \bar{x}_i, \tilde{x}_i < 1$ ;  $W_{ii} = (N + M)/N$ ;  $W_{ij} = O(1/N)$  によって、行列  $B$  は有界となるので、十分小さい結合重み係数  $K_{IE} \equiv \varepsilon$  の場合における平衡点  $\{\tilde{x}_i\}$  は漸近安定となる (山本, 1979)。以上の議論から、次の定理が得られる。

定理 1 ペア素子の結合を切った連想記憶モデルが記憶パターン付近の漸近安定平衡点を持ち、その *Jacobi* 行列が可逆である時、十分小さい結合重み係数  $K_{IE} \equiv \epsilon$  の場合の発振ニューラルネットワークも、同様な漸近安定平衡点を持つ。

### 3.3.2 単一の興奮 - 抑制性ペア素子の発振条件

さて、結合重み係数  $K_{IE} = \epsilon$  の値が大きくなると、平衡点は突然に不安定となって発振を引き起こす（パラメータの変化による漸近安定平衡点からの発振は、Hopf 分岐と呼ばれる）。発振を起こす要因の 1 つである、単一の興奮 - 抑制性ペア素子の Hopf 分岐の条件について議論しよう。もし、ある 1 つの興奮 - 抑制性ペア素子だけが Hopf 分岐を起こし、その分岐点では他のペア素子はまだ記憶パターン付近の平衡点で安定な場合は、この条件はペア素子を結合したネットワークの場合にも適用できる。上記は、発振を起こすための 1 つの特殊な場合であり、一般的な場合ではないが、3.4 節の数値解析ではこの現象に遭遇する。そこでは、発振が単一のペア素子の Hopf 分岐に基づくかどうかにかかわらず、結合重み係数  $K_{IE}$  の値の増加に伴って、ネットワークの挙動が記憶パターン付近の漸近安定平衡点への収束から発振へと変わることが示される。これを記憶点（記憶パターン付近の平衡点）の不安定化による発振と呼ぶ。

以下では、単一の興奮 - 抑制性ペア素子に限定した議論を行う。まず、唯一の平衡点  $(\bar{x}_i, \bar{y}_i)$  が存在すると仮定する。この時、結合重み係数  $K_{IE}$  の値の増加に伴う Hopf 分岐前後のようすを、図 3.2(a)(b) にそれぞれ示す。これより、図 3.2(a) の漸近安定平衡点の場合から、図 3.2(b) のように曲線  $\dot{y}_i = 0$  が立ち上がって、発振を引き起こすことが定性的に理解出来る。

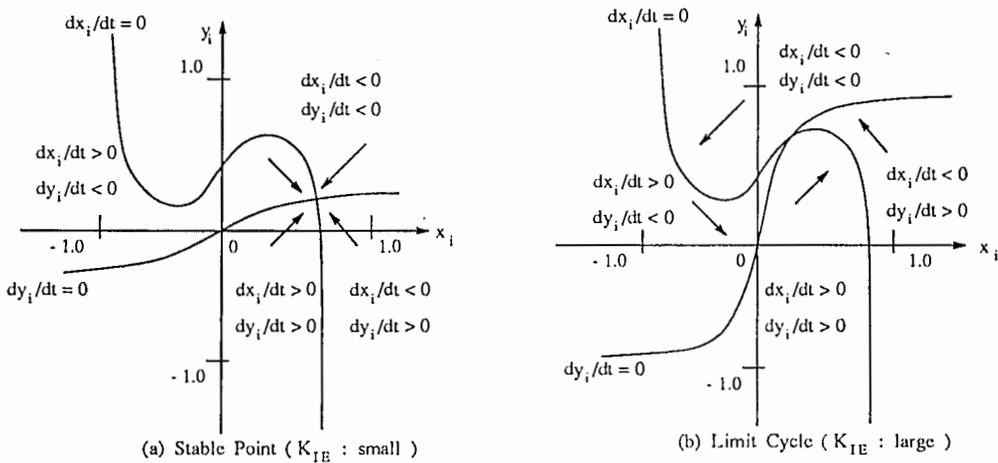


図 3.2: 単一のペア素子における平衡点の安定性

さて、平衡点  $(\bar{x}_i, \bar{y}_i)$  における線形化方程式は、

$$\begin{pmatrix} \dot{\delta x} \\ \dot{\delta y} \end{pmatrix} = \begin{pmatrix} -1 + W_{ii}G'(v_e) & -K_{EI}G'(v_e) \\ K_{IE}G'(v_i) & -1 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}. \quad (3.11)$$

となる。ここで、 $v_e \equiv W_{ii}\bar{x}_i - K_{EI}\bar{y}_i + I_i, v_i \equiv K_{IE}\bar{x}_i$  である。上記の行列の固有値  $\lambda$  は、その行列のトレース  $(Tr)$  と行列式  $(Det)$  を用いた、 $\lambda^2 - (Tr)\lambda + (Det) = 0$  の解であることから、

$$\lambda = [ (Tr) \pm \sqrt{(Tr)^2 - 4(Det)} ] / 2. \quad (3.12)$$

となる。Hopf分岐点においては、その2つの固有値はともに虚数となるので、その条件は  $(Det) > 0$  かつ  $(Tr) = 0$  となる (Guckenheimer, & Holmes, 1983)。

**定理 2** 単一の興奮 - 抑制性ペア素子における唯一の平衡点  $\{x_i^*\}$  が Hopf 分岐を起こすためには、条件  $K_{EI} > W_{ii}\pi x_i^*/2$  と、 $W_{ii} \geq \pi a$  がともに必要である。

(証明)

まず、 $(Tr) = 0$  についての条件を議論する。条件  $(Tr) = -2 + W_{ii}G'(v_e) = 0$  から、

$$\frac{2}{W_{ii}} = G'(v_e) = \frac{2a}{\pi(a^2 + v_e^2)}, \quad (3.13)$$

が得られる。これより、条件  $(Tr) = 0$  は  $\pi(a^2 + v_e^2) = aW_{ii}$  と等価となる。

$I_i > 0$ 、すなわち、 $v_e > 0$  の時 ( $I_i < 0$  の場合も同様の議論)、 $v_e = \sqrt{a(W_{ii} - \pi a)}/\pi$  から、

$$x_i^* = G(v_e) = \frac{2}{\pi} \arctan \sqrt{\frac{W_{ii} - \pi a}{\pi a}} = const. \quad (3.14)$$

となる。 $v_e = I_i + W_{ii}x_i^* - K_{EI}y_i^*$  を式 (3.13) に代入し、 $I_i$  に関して整理すれば、

$$-\pi I_i^2 - 2\pi(W_{ii}x_i^* - K_{EI}y_i^*)I_i - \pi(W_{ii}x_i^* - K_{EI}y_i^*)^2 + a(W_{ii} - \pi a) = 0. \quad (3.15)$$

が得られる。上式を満足するには  $I_i$  に関する2次方程式 (3.15) が実根を持たねばならないので、判別式  $D = 4\pi a(W_{ii} - \pi a) \geq 0$  から、 $W_{ii} \geq \pi a$  が必要となる (これは、 $\dot{x}_i = 0$  の曲線  $y_i = [I_i + W_{ii}x_i - G^{-1}(x_i)]/K_{IE}$  が変曲点を持つ為の必要条件  $W_{ii} \geq \pi a/2$  をも満たす)。

次に、 $(Det) > 0$  についての条件を、 $(Tr) = 0$  の場合で議論する。条件  $(Tr) = -2 + W_{ii}G'(v_e) = 0$  から、

$$(Det) = 1 + G'(v_e)[-W_{ii} + K_{IE}K_{EI}G'(v_i)] \quad (3.16)$$

$$= -1 + 2K_{IE}K_{EI}G'(v_i)/W_{ii} > 0, \quad (3.17)$$

は、条件  $W_{ii} < 2K_{IE}K_{EI}G'(v_i)$  と等価となる。

$v_i = K_{IE}x_i^*$  を不等式 (3.17) に代入し、 $K_{IE}$  に関して整理すれば以下のようになる。

$$W_{ii} < \frac{4aK_{EI}K_{IE}}{\pi(a^2 + (K_{IE}x_i^*)^2)}, \quad (3.18)$$

$$(W_{ii}\pi(x_i^*)^2)K_{IE}^2 - (4aK_{EI})K_{IE} + (W_{ii}\pi a^2) < 0. \quad (3.19)$$

$K_{IE}(> 0)$  がこの 2 次不等式 (3.19) を満足するには、その判別式が、

$$D = (4aK_{EI})^2 - 4(W_{ii}\pi(x_i^*)^2)(W_{ii}\pi a^2) = 16a^2[(K_{EI})^2 - (W_{ii}\pi x_i^*/2)^2] > 0, \quad (3.20)$$

となることが必要である。したがって、条件  $K_{EI} > W_{ii}\pi x_i^*/2$  が必要となる。(証明終り)

以上、記憶パターン付近の限られた挙動として、結合重み係数  $K_{IE}$  の値が十分小さい場合の記憶パターン付近の漸近安定平衡点の存在と、単一の興奮 - 抑制性ペア素子における Hopf 分岐の必要条件が、いくつかの仮定のもとで理論的に示された。

### 3.3.3 単一のペア素子における挙動

発振ニューラルネットワークの構成要素である単一のペア素子の挙動を調べることは、ネットワークの基本的な挙動を理解する上で重要と考えられる。以下では、数値解析に基づいた単一のペア素子の挙動を示す。ここで、定理 2 の Hopf 分岐の必要条件を満足するように、結合重み係数などのパラメータを  $W_{ii} = 1.0, K_{EI} = 2.0, a = 0.1$  と設定した。この時の、パラメータ  $I_i$  と  $K_{IE}$  に関する平衡点の安定性を図 3.3 に示す。図中の点線は、Hopf 分岐の候補点であるにもかかわらず、実際は発振が起こらなかった場合を示す。すなわち、候補点が  $(Tr) = 0, (Det) < 0$  のようなサドル点となり、もう 1 つの安定平衡点が存在した場合である。しかし、結合重み係数  $K_{IE}$  の値がより大きくなると、これら 2 つの平衡点が衝突して 1 つの不安定平衡点となり、発振を引き起こす。これは、定理 2 の Hopf 分岐とは異なるもので、生成 - 消滅分岐あるいはサドル点分岐と呼ばれるものである (Thompson, & Stewart, 1986; Wiggins, 1990; 合原, 1990b)。このように、Hopf 分岐の候補点で実際に発振が起こるかどうかは、入力バイアス値  $I_i$  に依存する。また、ペア素子の発振周波数の最大値は、結合重みパラメータ  $K_{IE}$  の値によって制限される (パラメータ  $K_{IE}$  と発振周波数  $\Omega$  の関係については、付録 E の標準形解析も参照されたい)。これに対して、結合重みパラメータ  $K_{IE}$  の値を固定した時は図 3.4 のように、入力バイアス値  $I_i$  が小さい程、発振周波数は高くなる。さらに、時間的に一定な入力バイアスとは別に、時間変動を持つ正弦波入力に対する共振現象も確認された。

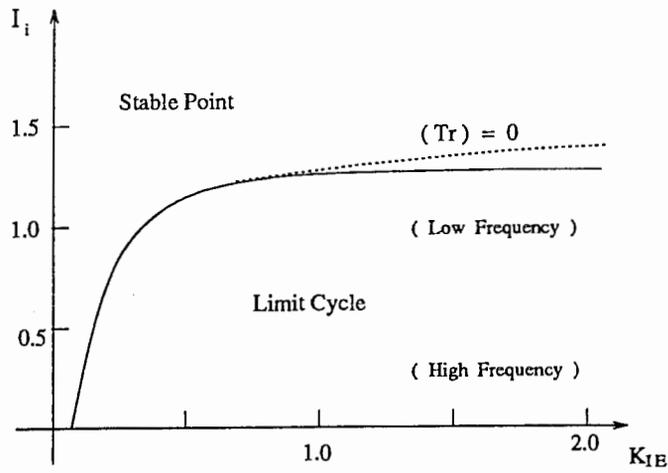
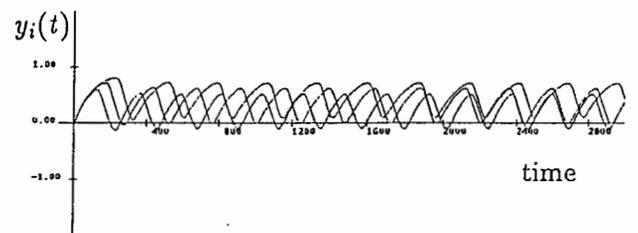
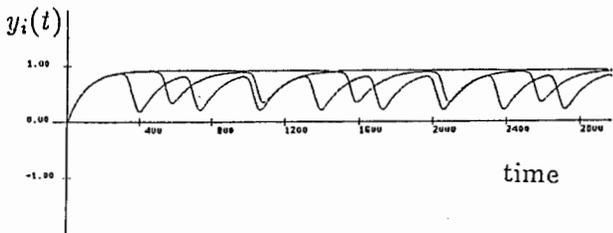
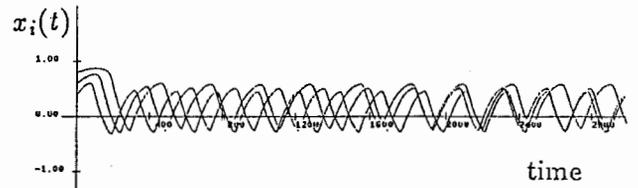
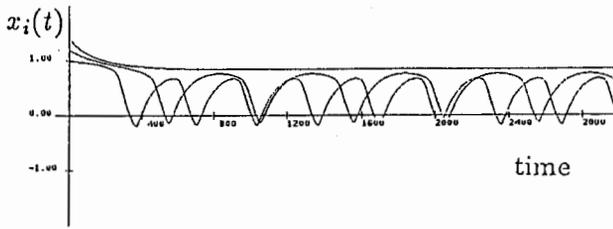


図 3.3: パラメータ  $I_i$ ,  $K_{IE}$  に対する平衡点の安定性マップ (実線は漸近安定平衡点への収束と発振の境界線を表し、点線は式 (3.15) を満たす Hopf 分岐の候補点を表す)



(a)  $I_i = 1.4 \sim 1.0$

(b)  $I_i = 0.8 \sim 0.4$

図 3.4: 単一ペア素子の発振周波数と入力バイアス値の関係

### 3.4 ネットワークの挙動の数値解析

前節の理論解析では、発振ニューラルネットワークの限られた挙動が明らかにされただけであった。本節では、それが複雑な軌道を持ちうることを議論した後、数値解析によってその典型的な挙動を示す。

3.3.3の単一のペア素子に関する実験結果で、結合重み係数  $K_{IE}$  がペア素子の最大発振周波数を規定し、入力バイアス  $I_i$  がその範囲内での実際の発振周波数を定めることが示された。これから、複数のペア素子を結合したネットワークにおいて、各素子に異なる入力バイアス値が与えられる場合は、さまざまな周波数成分を含んだ波形による複雑な軌道が生成されると考えられる。

一方、入力値が時間的に一定である限り、ネットワークは自律系のシステムとなるので、 $2N$ 次元の状態空間  $\{x_i, y_i\}$  には一般に複数のアトラクター（軌跡が交差することはない）が存在することになる（Hirsch, & Smale, 1974; Tompson, & Stewart, 1986）。すなわち、状態空間は初期状態  $\{x_i(0), y_i(0)\}$  に従ったベイスンによって分割される。ところで、ネットワークの特殊構造から、抑制性素子の活性値を表す  $y_i(t)$  は影の変数として、以下のような出力値（興奮性素子の活性値）を表す変数  $x_i(s)$  の過去の履歴 ( $s < t$ ) の関数に変換することが出来る。

$$y_i(t) = \exp^{-t} \left[ \int_0^t \exp^s G(K_{IE} x_i(s)) ds + y_i(0) \right]. \quad (3.21)$$

これによって、外部から可視化できない隠れ素子として、抑制性素子が興奮性素子の出力値の履歴情報を保持する働きを持つことがわかる。すなわち、動的な軌道上に情報を保持することが出来ると考えられる。この影の変数  $y_i(t)$  を式 (3.1) に代入すると、

$$\dot{x}_i(t) = -x_i + G \left( \sum_{j=1}^N W_{ij} x_j(t) - K_{EI} \exp^{-t} \int_0^t \exp^s G(K_{IE} x_i(s)) ds - K_{EI} \exp^{-t} y_i(0) + I_i \right). \quad (3.22)$$

が得られる。変化後の式 (3.22) は非自律系のシステムの動作方程式であるが、 $\{x_i\}$  の  $N$ 次元状態空間に射影されたアトラクターを考えることが出来る。但し、射影された  $N$ 次元状態空間内では、軌跡やアトラクター上の軌道が交差することがあり得る。ここで、式 (3.22) の第3項の抑制性素子の初期状態値  $y_i(0)$  は、指数的に減衰する外部入力と等価な効果を持つことに注意しよう（このことは4.4節で利用される）。

### 3.4.1 状態空間内の動的軌道

まず、3組のペア素子で構成される発振ニューラルネットワークを考える。ここで、3つの記憶パターン  $\xi^\mu$ ;  $\xi^1 = (1, 1, 1)$ ,  $\xi^2 = (1, -1, -1)$ ,  $\xi^3 = (-1, -1, 1)$  の自己相関行列によって結合重み係数  $W_{ij}$  を定義し、その他のパラメータは前節の定理2の条件を満足するように、 $K_{EI} = 2.0$ ,  $a = 0.1$ ,  $W_{ii} = 1.0$  とした。但し、結合重み係数  $K_{IE}$  は分岐パラメータとして変化させる。また、入力バイアス  $\{I_i\}$  は一様乱数で設定し、初期状態は  $x_i(0) = I_i$ ,  $y_i(0) = 0$  とした。この時、3つの興奮性素子の出力パターンを、 $\{x_i\}$  の3次元空間内の軌跡で表現する。各記憶パターンは3次元立方体 ( $-1 \leq x_i \leq 1$ ,  $i = 1 \sim 3$ ) の頂点となる。

無入力バイアスで結合重み係数  $K_{IE}$  の値が十分小さい時 ( $K_{IE} = 0.05$ ,  $I_i = 0$ )、状態空間内の各初期状態  $\{x_i(0)\}$  からそれぞれの記憶点 (記憶パターン付近の漸近安定平衡点) に収束する様子を図3.5 (a) に示す。一方、図3.5 (b) は、 $K_{IE} = 0.3$  とした時に3番目のペア素子  $x_3 - y_3$  が最初に Hopf 分岐を起こした直後の軌跡を示している。これらの挙動は、前節の定理1と2にそれぞれ対応するものである。

結合重み係数の値をさらに大きくした時 ( $K_{IE} = 0.5$ )、図3.5 (c) に示すような、リミットサイクルやカオスの軌道などのさまざまな軌道が生成される (図中、異なる3つの入力バイアス値に対する、アトラクター上の軌道を重ねて表示)。この時、入力バイアスに対して以下のような軌道の性質の変化が確認出来た。すなわち、記憶パターンに近い入力に対しては、その記憶パターン付近のリミットサイクルが出来るが、どの記憶パターンからも離れた入力に対しては、記憶パターン間を遊走するカオスの軌道が出来た。また、記憶パターンから離れた入力になる程、出力  $x_i(t)$  の波形は高周波成分を持って複雑な軌道となり、初期状態  $\{x_i(0)\}$  に依存した異なる軌道に収束するようになった。

さて、出力パターン  $x_i(t)$  と各記憶パターン  $\xi^\mu$  との距離の時間的変化を、以下のようなオーバーラップ関数  $m^\mu(t)$  で評価しよう (Horn & Usher, 1991a)。

$$m^\mu(t) = \frac{1}{N} \sum_{j=1}^N \xi_j^\mu x_j(t). \quad (3.23)$$

例として、図3.6 (a) は2つの記憶パターン  $\xi^2$  と  $\xi^3$  との間で周期的に変化する軌道を示し、図3.6 (b) は3つの記憶パターン間を複雑に遊走する軌道を示している。

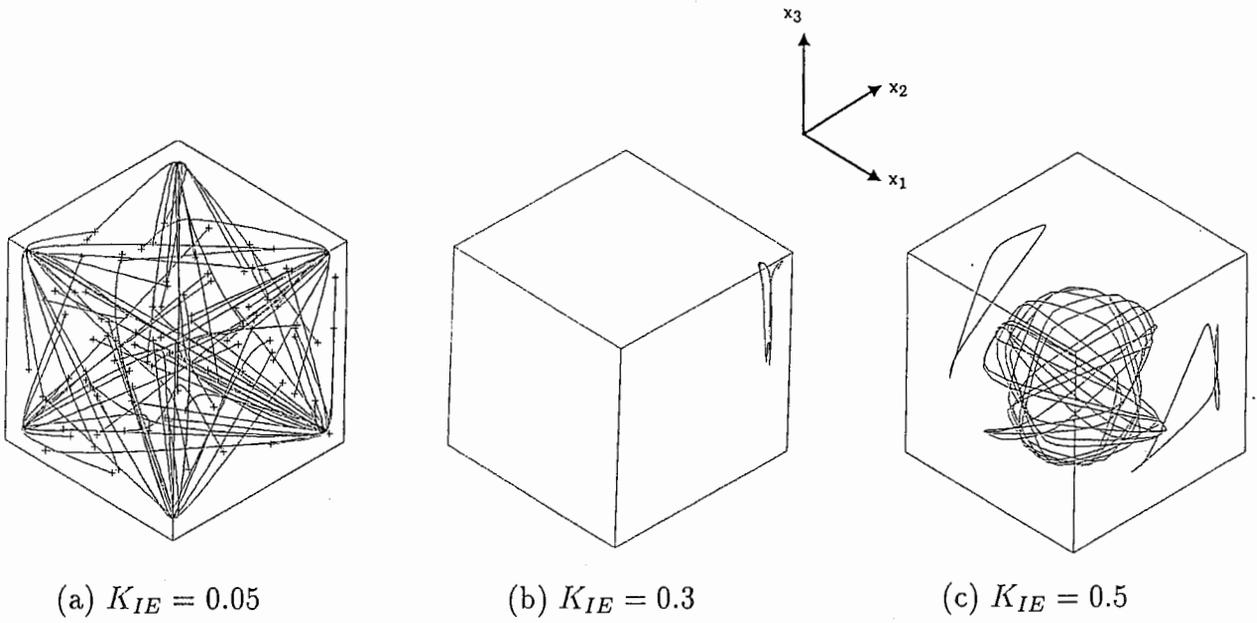


図 3.5: 結合重み係数  $K_{IE}$  に対する挙動の違い (+ 印の初期状態からの軌跡)

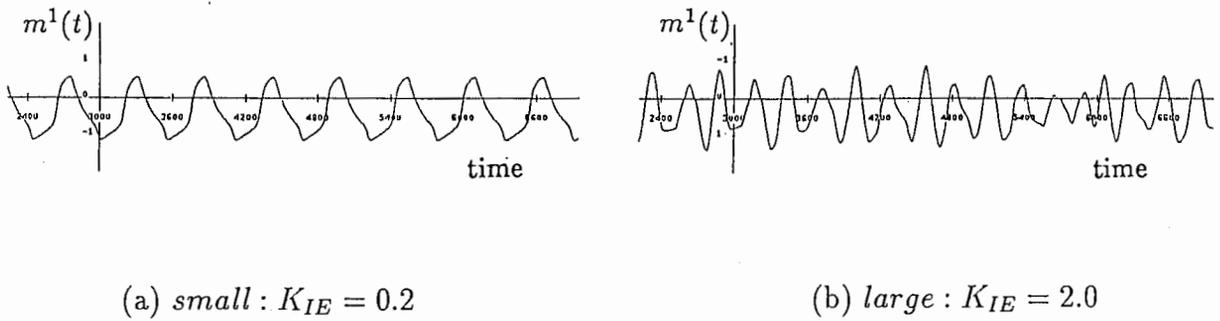
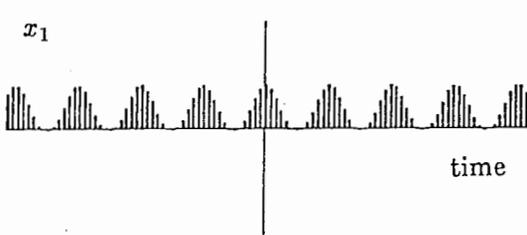


図 3.6: オーバーラップ関数  $m^1(t)$

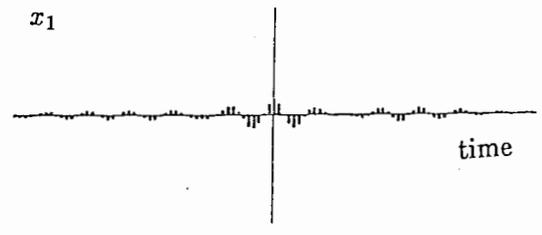
### 3.4.2 軌道のカオス性

これまで3次元の状態空間内の軌道から、その複雑なカオスの特徴を見てきた。以下では、出力波形の自己相関関数や周波数解析から、軌道のカオス性 (Bergé, Pomeau, & Vidal, 1984; 合原, 1990b) に関する検討をもう少し詳しく行う。図 3.7は、出力波形  $x_1(t)$  の自己相関関数  $\phi(s) \equiv \frac{1}{T} \sum_{k=1}^T x_1(k\Delta t) \times x_1(k\Delta t + s)$  のヒストグラムの例である。図 3.7 (a) の周期的なピークは、記憶パターンに近い入力に対する軌道がリミットサイクルであることを示すものである。一方、図 3.7 (b) において、離れた時間との相関をとる程、 $\phi(s)$  が減衰すること (未来の挙動の予測不能性) は、記憶パターンから離れた入力に対する軌道のカオス的な特徴を示している。他の素子に関する出力波形の自己相関関数や2素子の出力波形の間の相互相関関数についても、同様な結果を得た。

さらに、波形の周波数解析によっても、軌道のカオス性に関する検討を行った。図 3.8はオーバーラップ関数  $m^1(t)$  の波形のフーリエ解析結果を示すものである。図 3.8 (a) における、いくつかの周波数成分のピーク (周期波形であっても正弦波とは限らないので、こうした高周波成分を持つ) は、記憶パターンに近い入力に対する軌道の周期性を示している。一方、図 3.8 (b) における、連続的な周波数成分は、記憶パターンから離れた入力に対する軌道のカオス的な特徴を示している。ここで、図中の左端は周波数ゼロの直流成分を表している。他の記憶パターンのオーバーラップ関数  $m^2(t)$ ,  $m^3(t)$  についても同様な結果を得た。

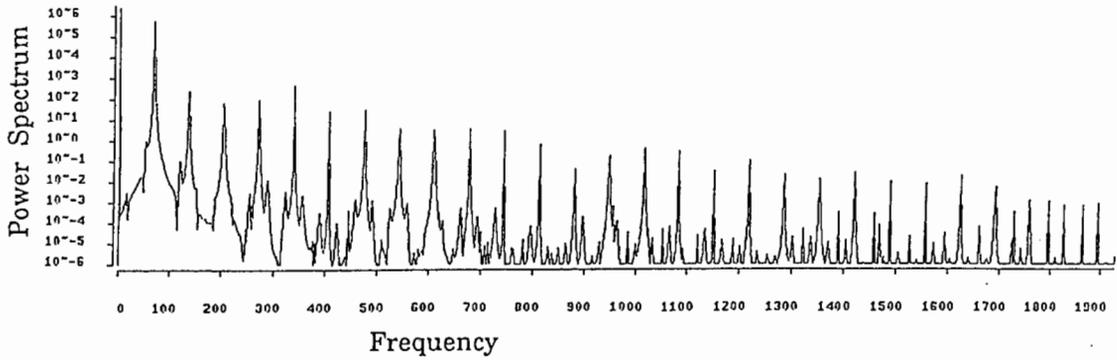


(a) 記憶パターンに近い入力

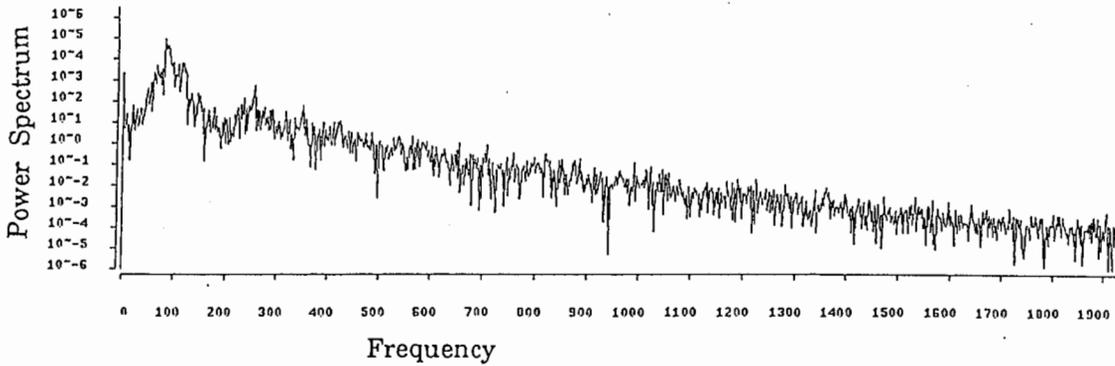


(b) 記憶パターンから離れた入力

図 3.7: 出力波形  $x_1(t)$  に対する自己相関関数  $\phi(s)$  のヒストグラム



(a) 記憶パターンに近い入力



(b) 記憶パターンから離れた入力

図 3.8: オーバーラップ関数  $m^1(t)$  の波形のパワースペクトル

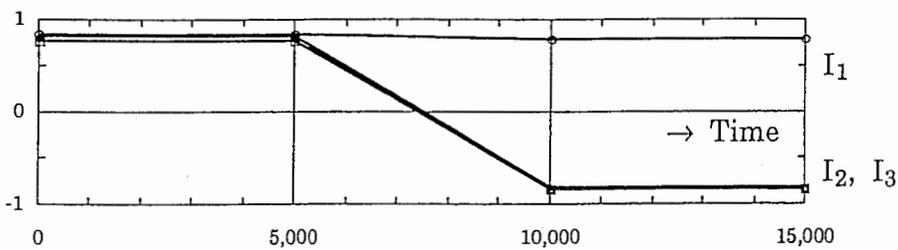
さて、カオスの存在可能性について検討しよう。Ruelle と Takens のシナリオ (Ruelle, & Takens, 1971; Newhouse, Ruelle, & Takens, 1978) によれば、3次元以上のトーラス上の準周期軌道 (周期軌道に近いが元の点に戻らず、トーラス曲面上を埋め尽くして閉じない軌道) からのカオスの発生が示唆されている。すなわち、3つ以上の自由度を持つ非線形システム (例えば、独立な固有振動数を持つ3つの自律振動子を弱結合したもの) においてはカオスは起こり得るのである。一方、発振ニューラルネットワークの各ペア素子は自律振動子であり、これら複数のペア素子は  $W_{ij} = O(1/N)$  の結合重み係数で相互に弱結合されている。ゆえに、発振ニューラルネットワークにおいても、入力バイアス  $I_i$  や結合重み係数  $K_{IE}$  が分岐パラメータとなって、トーラス上の準周期軌道からカオスへの分岐が起こると考えられる。

一方、以上の根拠と実験結果だけから、カオスの存在性が厳密に証明されたわけではない。通常、アトラクター上の軌道の線素ベクトルの各方向に対する伸縮の度合いを示すリアプノフ指数を数値解析によって求めて、正のリアプノフ指数を持つこと（軌道の発散性）からカオスの存在を検証することが多い（Bergé, Pomeau, & Vidal, 1984; 合原; 1990b）。しかしながら、このような解析は本論文の範囲を越えるものであり、今後の検討課題として残る。

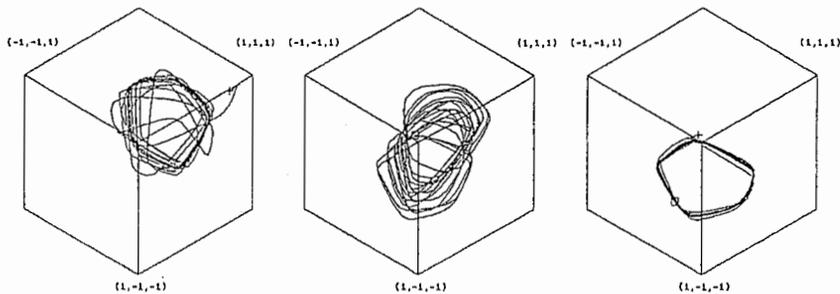
### 3.4.3 入力バイアスによるアトラクター間のスイッチング例

入力バイアスに従って、記憶パターン付近のリミットサイクルや記憶パターン間を遊走するカオスの軌道が生成されるという特徴は、入力バイアスの変化によってアトラクター間のスイッチングが引き起こされることを意味する。すなわち、入力バイアスで軌道の遊走領域の大まかな制御が出来ると考えられる。以下、前節と同様な3組のペア素子を結合したネットワークにおいて、入力バイアスの変化によるアトラクター間のスイッチングの例を示す。

図3.9は、上段(a)のように記憶パターン $\xi^1$ に近い入力から $\xi^2$ に近い入力に変化させた時の、3次元空間 $\{x_1, x_2, x_3\}$ における軌道の変化を下段(b)に示したものである。図の左右の軌道の変化は、入力に設定した2つの記憶パターン付近のリミットサイクル間のスイッチングを示している。また、図の中央は、入力を連続変化させた時に、軌道が比較的なめらかに連続変化していったことを示している。

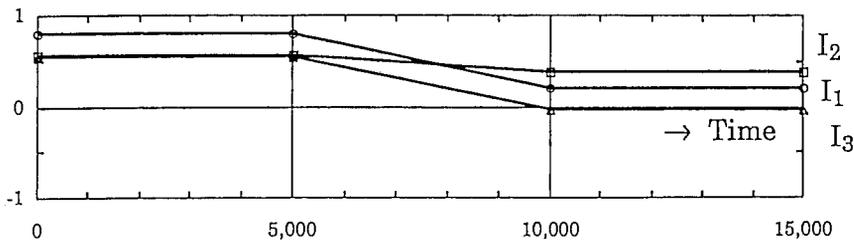


(a) 入力バイアスの変化 ( $y_i(0) \equiv 0$ )

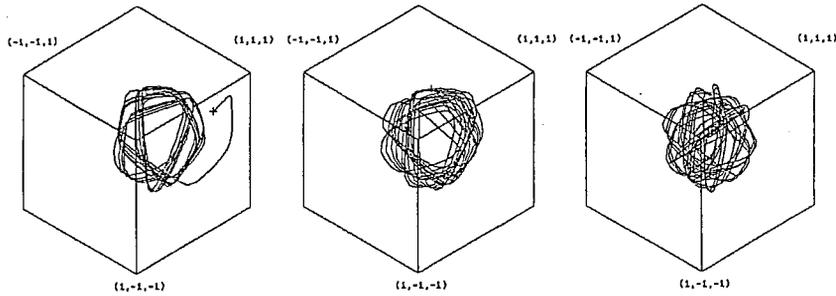


(b) 各区間に対応した軌道の変化

図 3.9: 2つのリミットサイクル間のスイッチング



(a) 入力バイアスの変化 ( $y_i(0) \equiv 0$ )



(b) 各区間に対応した軌道の変化

図 3.10: リミットサイクルからカオスの軌道へのスイッチング

上記のような2つのリミットサイクル間のスイッチングに対して、リミットサイクルからカオスの軌道へのスイッチングの例を図 3.10に示す。この例は、記憶パターン $\xi^1$ に近い入力から記憶パターンから離れた（原点近くの）入力に変化させた場合である。図の中央の、入力の連続変化に対する軌道の変化は、軌道がそれ自身に複雑に絡んでいくカオスへの分岐過程を示していると考えられる。

ところで、時間的に連続変化する入力を非線形システムに与えた場合、一般的にはアトラクターの存在が保証出来ない (Hirsch, & Smale, 1974; Tompson, & Stewart, 1986)。一方、図 3.9と図 3.10における2つの例は、入力の連続変化に対しても、軌道がアトラクターのようなものに引き込まれていることを示唆するものである。このような外部入力の変化による出力軌道（アトラクター）のスイッチングは、周期的な時系列パターンの入出力関係をリカレントネットワークに学習する場合にも指摘されている (Gouhara et al., 1992)。この場合、ネットワークを強制振動系として捉え、学習によって時系列入力に対応するアトラクターを状態空間に形成することが重要となる。また、外部入力や内部状態の変化に従ったアトラクター間のスイッチングを実験的に示した報告もある (朝日奈他, 1993)。より一般的な場合での、時間的に連続変化する入力に対するアトラクター間のスイッチング現象については、さらなる検討が必要となる。

以上、入力バイアスの変化によってアトラクター間のスイッチングを引き起こし、軌道の遊走領域を大まかに制御出来ることが示された。こうした入力バイアスに従った軌道の遊走領域の変化は、主にベースンの拡大や縮小に基づくものと考えられる。すなわち、状態空間内のバイアス点に近いアトラクターのベースンは拡大し、その点から離れたアトラクターのベースンは縮小または退化する。また、強い入力バイアスを設定すれば、ほとんどすべての初期状態を含むような大きなベースンを作ることも出来るだろう。もちろん、入力バイアスの変化に伴ってベースンの融合や分割が引き起こされ、その結果、アトラクターの数が変わるような分岐も起こり得る。こうした入力バイアスに従ったベースンの拡大や縮小は、外部入力（しきい値）によって連想記憶モデルの記憶容量が改善されるという現象（Amit, Gutfreund & Sompolinsky, 1987b）とも関係があると思われる。しかしながら、ベースンの形状は一般に非常に複雑と考えられる為（Amari, & Maginu, 1988）、状態空間のベースンによる分割構造を解析することは難しい問題となる。

### 3.5 記憶パターン数の増加に伴う遊走的軌道の性質の劇的な変化

前節までに、発振ニューラルネットワークの軌道の複雑さに関する特徴を調べてきた。本節ではさらに、 $N$ 次元超立方体内の状態空間で軌道がどのように遊走するかについての検討を以下の項目から行った。

- 静的な連想記憶モデルの記憶パターン想起の破綻に関する臨界値前後に相当する、 $M < 0.14N$  と  $M > 0.14N$  ( $M$  は記憶パターン数) とにおいて、遊走的軌道の性質の劇的な変化が起こるかどうかが
- 入力バイアスに従った、遊走領域の変化（軌道の最近接頂点における滞在率の分布のシフトの割合で評価する）
- 入力バイアスを記憶パターンに比例して設定した時と、記憶パターン以外（の頂点）に比例して設定した時との、遊走領域の変化の違い

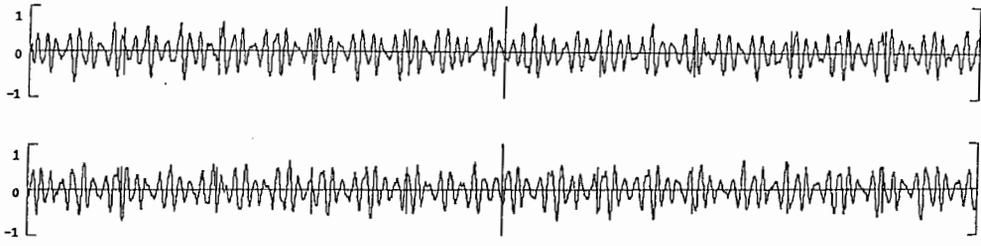
#### 3.5.1 記憶パターンへの引力の弱体化

本節では、以下のような2組のペア素子を結合したネットワークを考える。この次元数  $N = 22$  は、4.3節の顔パターンの生成における、目や口などの顔の構成要素に関する22個の特徴要素数に相当するものである（付録D参照）。一方、次元数  $N$  がより大きい場合 ( $N = 30, 50$ ) についても、以下に説明するものとほぼ同様な結果が得られたが、これらに関しては付録Bを参照されたい。

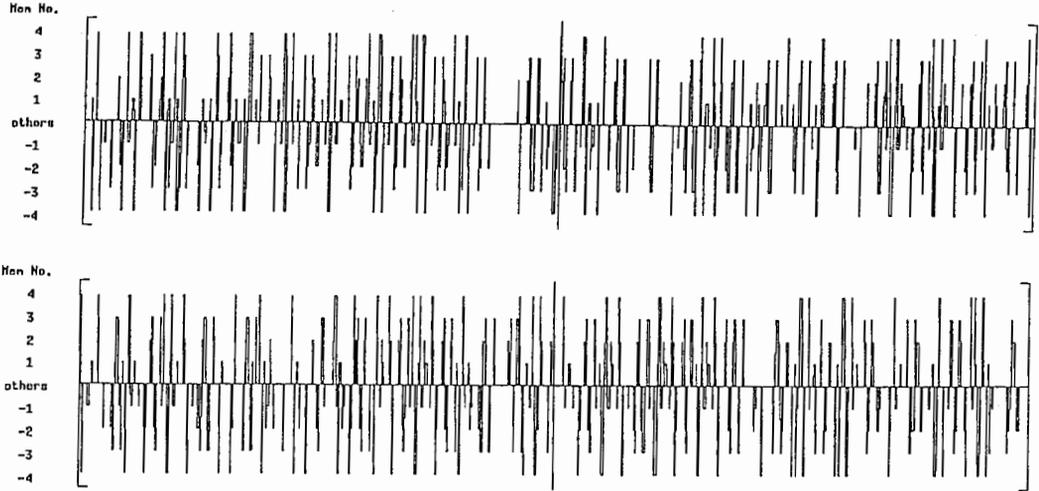
さて、結合重み係数  $W_{ij}$  を、ほとんど直交な  $M$  個の記憶パターン  $\xi^\mu$  ( $\mu = 1 \sim M$ ) の自己相関行列で設定しよう。すなわち、各記憶パターンは  $+1$  と  $-1$  の要素を同じ数だけ ( $1$  個ずつ) 持ち、記憶パターン間の相関がなるべく小さくなるよう、 $2 < |\sum_{i=1}^N \xi_i^\alpha \xi_i^\beta| < 6$  ( $\alpha \neq \beta$ ,  $\alpha, \beta = 1 \sim M$ ) となるものを考えた。これは、記憶パターンが互いに  $8 \sim 14$  ビット離れていることに相当する。他のパラメータは前節と同様な値にした。

$N$  次元超立方体内の遊走軌道の振る舞いを調べる為に、軌道の最近接頂点における滞在時間を計り、基準となる頂点からのビット数に従った滞在頻度の分布を表すヒストグラムを求めた。ここで、分布の形がアトラクターに落ち着く前の過渡期に依存しないように、4 次の Runge-Kutta 法 ( $\Delta t = 0.05$ ) による反復回数を 1 万回とした。また、分布に対する初期状態依存性を調べる為に、(i) 原点近く:  $|x_i(0)| \leq 0.2$ 、(ii) 頂点近く:  $0.8 \leq |x_i(0)| \leq 1.0$ 、(iii) 記憶パターン近く:  $x_i(0) \equiv \pm 0.8\xi_i^\mu$ 、(iv) これら以外の原点と頂点との中間:  $0.2 \leq |x_i(0)| \leq 0.8$  の 4 通りの場合を考え、それぞれ 50 個ずつの異なる初期状態 (一様乱数で生成) に対する分布を比較した。図 3.11 のように、軌道の最近接頂点の滞在順序や 1 回の滞在時間は初期状態に大きく依存した。一方、その平均的な滞在率を示す分布は初期状態にほとんど依存しないことが判明した。以下では、200 個 (上記の 4 通りの初期状態  $\times$  50 個) の初期状態に対する平均の分布に関して議論を行う。

無入力の場合 ( $I_i \equiv 0$ ) について、 $M < 0.14N$  と  $M > 0.14N$  における分布の違いを比較検討する。静的な連想記憶モデルがその記憶容量の臨界値  $M \approx 0.14N$  の前後において、記憶パターンへの収束から記憶パターンとは異なる頂点上の疑似記憶パターンへの収束に劇的に変化する (Amit, Gutfreund, & Sompolinsky, 1987a, 1987b; Geszti, 1990) のと同様な現象が、発振ニューラルネットワークにおいても起こると考えられる。すなわち、記憶パターン数の増加に伴って、記憶パターンの頂点方向に軌道を引きつける力が突然に弱体化することが予想される。この記憶パターンへの引力の弱体化によって、記憶パターン間に集中していた軌道の遊走範囲が、他の頂点間を含む大きな範囲に広がると考えられる。



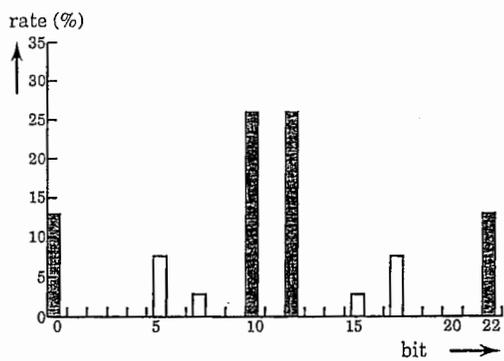
(a) オーバーラップ関数  $m^1(t)$



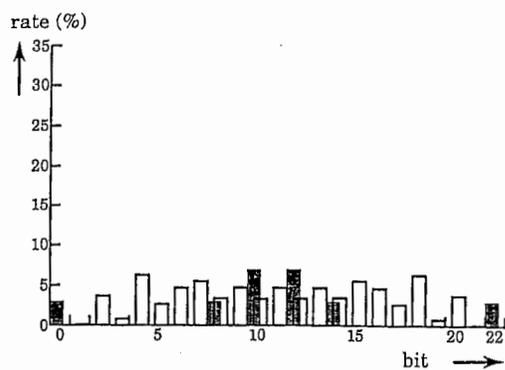
(b) 軌道の最近接頂点の時間的推移

図 3.11: 微妙に異なる 2 つの初期状態に対する遊走軌道の違い (2 つの初期状態における各要素の差は 0.02 以内とした時)

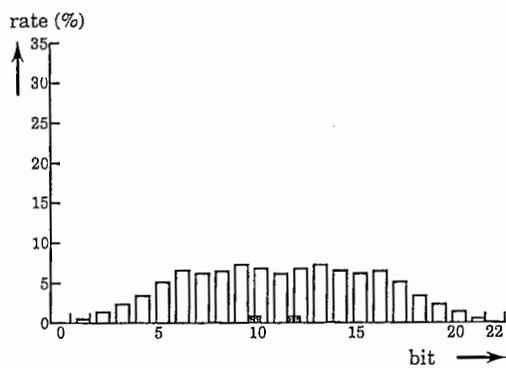
図 3.12(a) は、 $M = 3 < 0.14N$  において記憶パターン  $\xi^1$  を基準 (0 ビットの位置) とした場合の分布を示す。分布の特徴として、軌道の最近接頂点が記憶パターンの頂点に集中 (約 78%) していることが示されている。ここで、基準から  $k$  ビット離れた位置には、 ${}_N C_k$  個の頂点があることに注意しよう (もっとも、軌道の最近接頂点がこれら全ての頂点を含むとは限らないが)。一方、図 3.12(b), (c) は、 $M = 4, 5 > 0.14N$  における分布 (分布の基準は同じ  $\xi^1$ ) をそれぞれ示すもので、図 3.12(a) とは明らかに異なった分布となっている。すなわち、記憶パターンの頂点以外にも軌道が滞在するよう、劇的な変化が起きたことが伺える。



(a)  $M = 3, I_i = 0.0$



(b)  $M = 4, I_i = 0.0$



(c)  $M = 5, I_i = 0.0$

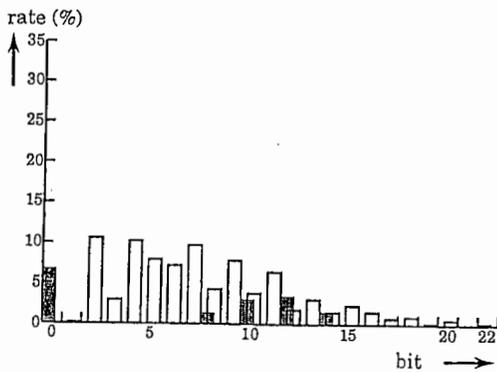
図 3.12: 軌道の最近接頂点における滞在率の分布

### 3.5.2 入力バイアスによる遊走領域の制御

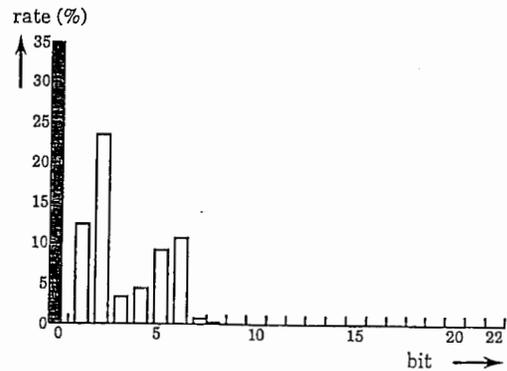
次に、 $M = 4$  において、頂点に比例した入力バイアスを設定した場合の分布のシフトについて検討する。図 3.12(b) の分布と比較して、図 3.13(a)(b) の結果は、記憶パターン  $\xi^1$  (0 ビットの基準点) に比例した入力バイアス  $I_i \equiv \alpha \xi_i^1$  に対する分布のシフトを示すものである (順に、 $\alpha = 0, 0.3, 0.6$  の場合を示す)。また、分布のシフトのみならず、図 3.13(b) では  $\xi^1$  以外の他の 7 つの記憶パターンに対応する頂点の滞在率が極端に小さくなっているのがわかる。

これに対して、記憶パターン以外の頂点  $\zeta \neq \pm \xi^\mu$  に比例した入力バイアス  $I_i \equiv \alpha \zeta_i$  ( $\alpha = 0.3, 0.6, 0.9$ ) を設定した場合、図 3.14(a)(b)(c) のように分布のシフト量が上記より少なくなる。但し、図 3.14 の分布は、バイアスをかけられた頂点  $\zeta$  を新たな 0 ビットの基準点としていることに注意されたい。このように、記憶点以外の頂点にたとえ強いバイアスをかけたとしても、その頂点が強い引力を持つことは難しいと言える。

以上の分布は、一様乱数で生成されたアナログパターンの最近接頂点を作る 2 項分布とは、明らかに異なるものである。この違いは、ネットワークのダイナミクスの特徴に従って、軌道が記憶パターンの頂点により引き付けられることによるものと考えられる。この場合、 $N$  次元超立方体の  $2^N$  個の全ての頂点を軌道の最近接頂点とするとは考えにくい。しかしながら、このことはまだ明らかではなく、今後の課題として残る。

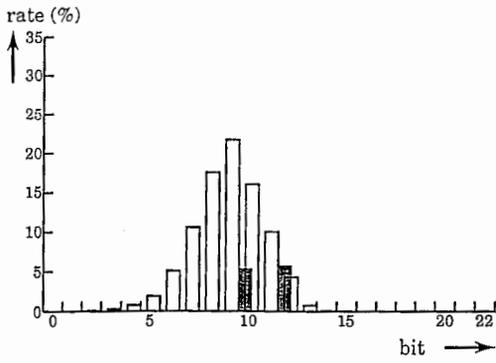


(a)  $M = 4, I_i = 0.3\xi_i^1$

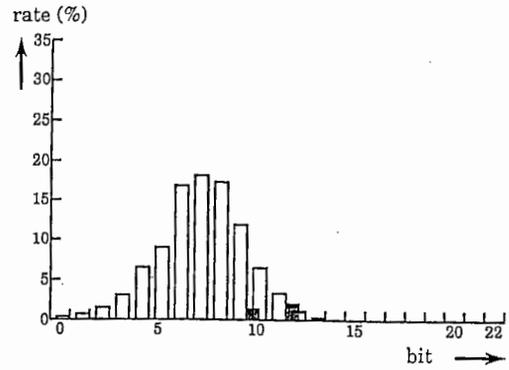


(b)  $M = 4, I_i = 0.6\xi_i^1$

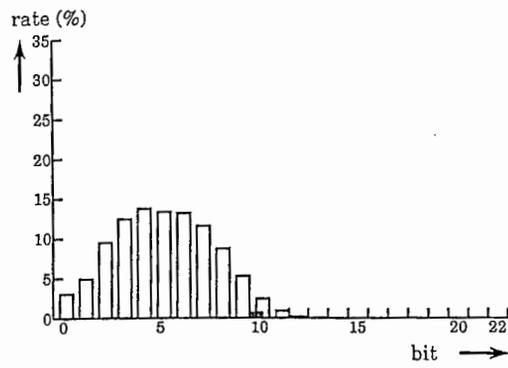
図 3.13: 記憶パターンに比例した入力バイアスに対する分布のシフト



(a)  $M = 4, I_i = 0.3\zeta_i$



(b)  $M = 4, I_i = 0.6\zeta_i$



(c)  $M = 4, I_i = 0.9\zeta_i$

図 3.14: 記憶パターン以外の頂点に比例した入力バイアスに対する分布のシフト

### 3.6 おわりに

静的な空間分割の問題点を解決する為の1つの手段として、時間変化を伴うアトラクターを持つ動的なモデルの出力軌道を情報処理機能に対応付け、外部入力などによる分岐現象を伴った出力軌道の変化を情報処理機能のスイッチングに利用することを考えた。そこで、この考えを実現する為の具体的なモデルの1つとして、興奮-抑制性ペア素子からなる発振ニューラルネットワークを提案し、その軌道の多様性と可制御性に関する特徴を、主に数値解析を通じて明らかにした。また、理論解析からも、記憶パターン近くの限られた挙動を明らかにした。軌道の多様性としては、リミットサイクルやカオス的軌道などのさまざまな複雑さを持った軌道の存在や、これらの軌道が記憶パターン間や他の頂点間を遊走することが示された。一方、軌道の可制御性としては、結合重み係数  $K_{IE}$  に従った軌道の複雑さの変化や、入力バイアスに従った軌道の遊走領域の変化、さらに、記憶パターン数の増加に伴う遊走軌道の性質の劇的な変化が示された。このような特徴的な挙動は、自己相関行列による記憶パターンの安定化作用（引き込み）と、これとは相反する、ペア素子による不安定化作用（遊走化）の共存によるものと考えられる。しかしながら、これらの解析結果から、モデルの複雑な挙動が完全に解明されたわけではない。

今後の課題としては、より厳密なカオスの存在性、最近接頂点に対する遊走軌道の距離や、より詳細な滞在頻度（滞在できる頂点が限定されるかどうかも含む）、さらに、状態空間内のベイソン構造や、入力バイアスや結合重み係数等の分岐パラメータに従った、その分割構造の変化などに関する検討項目がある。

## 第 4 章

### 時空間パターン処理への応用

#### 4.1 はじめに

動的なモデルの時空間パターン処理においては、素子間の相互作用のダイナミクスに基づいて、時間的情報を状態空間内の軌道上へ保持することが重要となる。この際、動的な軌道の情報処理機能への対応付けの仕方によって、モデルの複雑な挙動が有効に利用されるかどうかが決まる。また、ネットワーク内の素子間の階層関係や、出力の過去の履歴に対する時間的依存性を考慮することも応用上有効と考えられる。例えば、人間の単語知覚のモデル (McClelland, & Rumelhart, 1981) においては、素子間の興奮と抑制の相互作用に加えて、線分特徴レベルや文字レベル、単語レベルなどの素子間の階層関係が考えられている。これとは別の例として、音声データの単語認識の課題に対し、過去の音素の影響をさまざまな時間遅れフィルターを持った検出器で保持した、時間依存のエネルギー関数を持つ動的な連想記憶モデルがある (Tank, & Hopfield, 1987)。しかしながら、いずれのモデルにおいても出力の活性値の高さのみが注目されている為、その時間的変動の複雑さは情報処理に有効に利用されていない。これらに対して、前章で議論した発振ニューラルネットワークはより複雑な挙動を持つことから、時空間パターン処理への応用を考える際には、その動的な軌道と情報処理機能との対応付けがより重要な意味を持つ。ところで、カオスなどの複雑な軌道は情報処理に有用となり得るのであろうか？。

最近、カオスの情報処理に対する有用性が紹介されている (Ditto, & Pecorra, 1993)。例えば、レーザーの安定化による出力の増大、化学反応の振動の安定化、電子メッセージの暗号化などへの応用が考えられている。これら以外にも、動物が新しい臭いを学習する時に、嗅覚系の神経回路網でカオス的挙動がみられることから、カオスが学習に必要な触媒としての役割を持つことが主張されている (Sakarda, & Freeman, 1987)。また、カオスの別の役割として、大脳皮質のモデルにおける記憶パターン間の探索機能 (Tsuda, Koerner, & Shimizu, 1987; Tsuda, 1992) や、組み合わせ最適化問題に対する効率的な解の探索 (Nozawa, 1992) なども示唆されている。

これらに対して、奈良ら (1992, 1993a) は、モデルの内的なダイナミクスの特徴と、課題からの外的な要求との適合性が重要であることを強調している。この考えに基づいて、彼らの提案したモデルのカオス的挙動の特徴が、記憶された時系列パターンサイクルの効率的な探索に利用出来ることが報告されている (Nara et al., 1993a)。そこでは、モデルの出力が外部評価者にとって満足出来るかどうか依存して、モデルの挙動を収束 (解を確定) させたり発散 (解を探索) させたりするように、分岐パラメータの制御が対話的に行われる。さらに、彼らは複数の記憶パターンの合成パターンを発生することについても言及し (Nara et al., 1993b)、光学的な非線形システムでは、記憶パターン間の中間的状態を新しいアトラクターとして生成出来ることも示している (Davis, 1991)。これらには、初期状態への鋭敏性を持つカオスの情報生成源としての性質 (Nicols, & Tsuda, 1985; Nicols, 1986; 津田, 1990) が利用されている。以上のように、モデルの軌道が多様であるばかりか、外部からそれが制御可能であることが重要なのである。

本章では、時空間パターン処理に対する、発振ニューラルネットワークの動的な挙動の応用可能性を検討する。まず、4.2節では、認知心理学的な文字パターンの連続変形と仮定された軌道を、記憶パターン付近のリミットサイクル上に学習し、マクロな変形操作とミクロなニューロン集団の自律発振との関連付けを試みる。また、学習後のネットワークに、文字カテゴリーの許容範囲内の変形パターンの生成機能と、入力文字が未知である時の、文字カテゴリー間の自律的な探索機能が存在することを示す。次に、4.3節では、状態空間内の軌道の多様性と可制御性の特徴を多様なパターンの生成に利用することを考える。そこで、外部評価者であるユーザと情報生成源としてのモデルとの対話的処理による、ユーザの概念イメージに応じた探索領域の制御方法を提案する。また、ユーザの要求に応じて、候補パターンのサーチ機能と確定機能とのスイッチングが出来ることも示す。最後に、4.4節では、高次の文脈情報からの予測に従って低次のパターン情報に関する識別領域を変化させることを考える。このような動的な識別領域の制御に、ベイソンの拡大や縮小に基づいた状態空間構造の変化を利用することを提案する。

## 4.2 生理学的自律発振と認知心理学的連続変形操作

4.2.1では、人間の知覚過程における対象物体の連続的な変形操作の存在性を紹介する。一方、このようなマクロな操作も、脳内のミクロなニューロンの活動に基づくものなので、これらのマクロな操作とミクロな活動との間には何らかの関連があっても不思議ではない。そこで、連続変形操作の能動性の源が、ミクロなニューロン集団の自律的な発振現象 (Sakarda, & Freeman, 1987; Eckhorn et al., 1988; Gray, & Singer, 1989; 松本, 1992) にあるとする大胆な仮定を試みる。4.2.2では、こうした関連性に着目した1つの試みとして、文字パターンの連続的な変形サイクルを発振ニューラルネットワークの軌道上に、学習によって埋め込むことを考える。

#### 4.2.1 脳内の連続変形操作の存在可能性

人間が視聴覚等の感覚器官を通じて外界の物体を知覚する時、脳内で対象物体からの不変特徴の抽出が受動的に行われるだけでなく、対象物体の連続的な変形操作も能動的に行われている可能性が高い。例えば、ある回転角だけ異なる2つの図形を見た時、図形の一致度を判断する為に、人間が視覚イメージの内的な回転操作（メンタルローテーション）を行うことは、心理学的に広く支持されている。しかし、どんな場合でもメンタルローテーションが実行されるわけではなく、図形の構成要素や線分間の配置関係など（情報タイプと呼ばれる）が明らかに異なる場合は、これが実行されないことが報告されている（Takano, 1989）。このように、特徴抽出に基づく処理と連続的なパターンの変形操作は、互いに矛盾するわけではなく、その場の状況に応じて、競合的に選択されたり協調したりするものと考えられる。

さて、特定の感覚に限定しない広い意味での知覚において、内的な連続変形操作の存在性を指摘した研究として、発達心理学における Piaget の理論（Piaget, 1970; Sigel, & Cocking, 1977）がある。そこでは、

1. 7～8才頃を転換点として、外界の対象物の状態のみに着目した静的なイメージから、それを内的に操作することが可能となって、変形や運動を対象とした動的なイメージに変わること
2. 連続的な変形操作によって、外界の対象物の状態とその変化の関係が対応付けられ、予想されたイメージが確かになること
3. この段階以降、可逆性、相補性、分類、系列化などの操作を通じて、保存性の概念が確立されるとともに、知覚によってのみ、こうした機能的なイメージが形成されること

が述べられている。Piaget の理論と関連性を持つ実験結果として、例えば以下の2つのようなものがある。1つは、図形の類似性が、3 - 4才児では、いくつかの構成要素の特徴的な部分の類似性のみで判断されて、全体的な形状は無視されるのに対して、8 - 9才児では、操作による要素間の関係の類似性で判断されることを示す報告である（田中, 1969）。この結果は、特徴抽出後に識別を行うような通常のパターン認識の枠組では説明出来ないばかりか、脳内における変形操作の存在性をも示唆するものと考えられる。もう1つは、小学生から大学生までを対象とした、日本語のカタカナや英語のアルファベット文字の認識実験の報告である（横瀬, 1986）。ここでは、反応時間が標準文字と提示文字との差の割合に比例すること、認識率が年齢に比例することなどが示されている。この結果は、差の割合が大きいほど内的な変形操作を行うのに時間を要し、年齢の低い生徒ほど変形操作をうまく扱えない割合が多いことによると解釈出来る。

これらとは別の、パターン認知の変換構造説と呼ばれる仮説（伊藤, 1975; 今井, 1986）も、内的な変形操作の存在性を支持するものである。その仮説では、鏡映変換、原点の平行移動の変換、刺激の白黒系列の符号反転の変換、白黒系列の周波数の変換（拡大や縮小に相当）などの座標軸に関係するパターン変換と、これらの組合せからなる変換構造を考え、パターンの類似性の判断や良さの判断が、こうしたパターンの変換構造の選択やその操作の後に行われることが主張されている。上記のような変換構造の選択は、情報タイプ（図形の構成要素や線分間の配置関係など）に依存してメンタルローテーションという内的な変形操作の実行が選択されること（Takano, 1989）とも関連性があると思われる。

以上より、脳内のパターン認識過程における能動的な連続変形操作の存在性が示唆された。しかしながら、どんな対象に対して、どんな操作が（選択的に）実行されるかについては、ほとんど明らかでない。そこで、次節では、ある人工的な文字パターンの連続的な変形サイクルを、入力から想像された文字カテゴリ内の変形の許容範囲を示すような変形操作と仮定する。この操作は、入力に対応するカテゴリの候補パターンの生成や検索の機能に相当するものである。この仮定のもとで、マクロな変形操作とマイクロなニューロン集団の自律発振を関連付けることを試み、上記の連続変形パターンを発振ニューラルネットワークの軌道上に効率的に埋め込むことを検討する。

#### 4.2.2 連続変形パターンの軌道上への学習方法

手書き文字などにおいては、1つの文字カテゴリ内にさまざまな変形文字が存在する。しかし、その変形は主に、文字の全体的な傾き（シフト）と局所的な（非線形の）歪みに基づいている。そこで、極座標  $\{r, \theta\}$  において、以下のような局所的な歪み変換としての心臓形変換と、全体的な角度  $\phi$  のシフト変換との合成で定義される、人工的なパターンの連続変形を考える。

$$r_{new} = r (1 - k \sin \theta), \quad (4.1)$$

$$\theta_{new} = \theta. \quad (4.2)$$

上記の心臓形変換は、人の年齢に関連付けられる横顔の形状変換（Shaw, & Wilson, 1976）に使われたもので、 $k$  は歪み量に関するパラメータである。

このような変形によって生成された文字パターン系列を図 4.1 に示す。もちろん、脳内の文字認識過程に介在する操作が、このような変形であるという保証はないが、以下の議論ではこれを仮定して話を進める。また、処理の簡単さから、図 4.1 では文字パターンを画素の集まり（画像表現）で表した。一方、特徴抽出後の内部表現されたパターンを考えた方が生理学的にはより忠実であろうが、それが具体的にどんな表現かはまだ明らかでない（Miyashita, & Chang, 1988）。

このような連続的な変形パターンのサイクルを、発振ニューラルネットワークの軌道に学習によって埋め込むことを考える。すなわち、学習時に与えられた興奮性素子への教師信号である変形パターンのサイクルをネットワークで再現する。入力文字としては、電総研が作成した標準的な文字データベース ETL9 の手書きひらがな文字から、「あ」、「い」、「う」の3文字を選んだ。ここで、各文字についての2値パターン(64 × 63)に大ききの正規化(64 × 64)を施した後、16 × 16 (= 256 次元)のメッシュ内の平均濃度に変換したメッシュ特徴パターンを入力として用いた。これより、発振ニューラルネットワークは256組のペア素子を持つことになる。これら3文字の入力パターンを図4.2に示す(各要素の値を256階調の濃度で表現)。

さて、各文字パターンの連続変形サイクルの学習方法を検討しよう。これまでに、時間変化を伴う軌道に対するリカレントネットワークの結合重みパラメータの学習アルゴリズムが提案されている(Sato, 1990; Pearlmutter, 1989)。佐藤らは、この学習アルゴリズムを適用して、低次元のカオスであるローレンツアトラクター(Sato, Murakami, & Joe, 1990a)や、音声波形の揺らぎ(Sato, Joe, & Hirahara, 1990b)が軌道に埋め込めることを実験的に示した。しかしながら、このアルゴリズムを発振ニューラルネットワークに直接適用すると、分岐パラメータである結合重み係数  $K_{IE}^{(i)}$  の値が変更されるとともに、非効率的な時間逆方向きの処理が必要となってしまう(付録C参照)。分岐パラメータの変化は、ネットワークの挙動の質的变化を引き起こす為、学習時間を膨大にするばかりか、学習を妨げる深刻な問題をも引き起こす(Doya, 1992)。そこで、結合重み係数  $K_{IE}^{(i)}$  の値を全ての素子に共通な  $K_{IE}$  に固定して学習することを考える。この場合は、付録Cから時間逆方向きの処理が不要となって、学習の効率的な実行が期待出来る。但し、結合重みパラメータの値を固定することは、ネットワークの情報処理能力の限定を意味することに注意しなければならない。以下に、その改良アルゴリズムの学習手順を説明する。

まず始めに、3つの文字カテゴリーに対する記憶パターンを、上記の入力パターンの各要素の値を±1に2値化したもので定義する。これらの記憶パターンの自己相関行列で結合重み係数  $W_{ij}$  を設定し、その他の  $K_{IE}$  や  $K_{EI}^{(i)}$  などのパラメータは前章と同じ値に初期設定する( $K_{IE} = K_{EI}^{(i)} = 2.0$ ,  $a = 0.1$ )。結合重み係数  $W_{ij}$  のこのような設定の根拠は以下のようである。前章で示された発振ニューラルネットワークの軌道の性質から、記憶パターンに類似した入力パターンに対する軌道は、漸近安定平衡点やカオス軌道にはならず、その記憶パターン付近のリミットサイクルになりやすいと言える。ゆえに、このリミットサイクルの形や位置を修正すれば、教師信号の変形パターンのサイクルになるので、ネットワークの挙動を質的に変化させなくても学習でこれを実現出来ることが予想される。もちろん、これは、記憶パターン付近で変動する図4.1のような教師信号の変形パターンのサイクルを考えている場合の話で、記憶パターンから大きく離れて変動するサイクルの場合は、この方法で学習することは困難かもしれない。

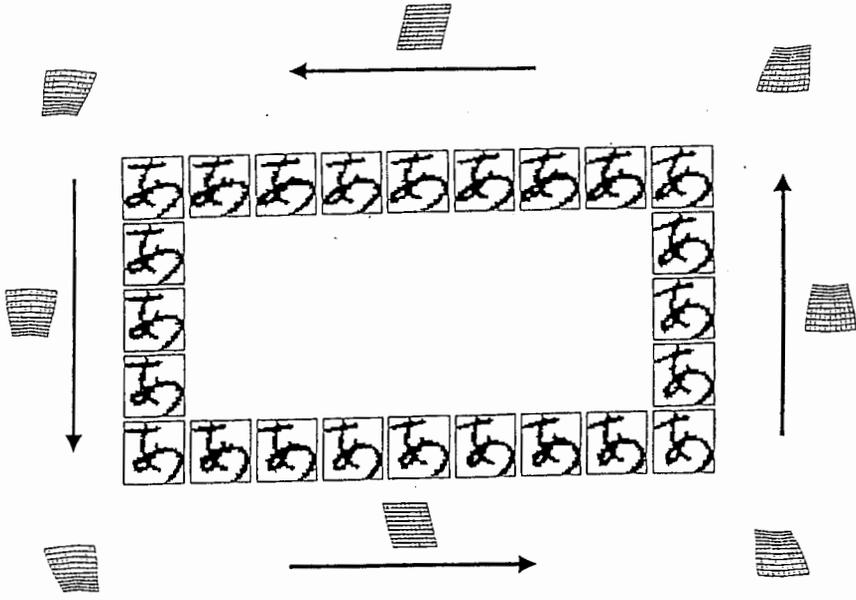


図 4.1: 文字の連続的な変形パターンのサイクル

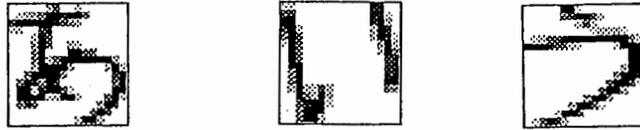


図 4.2: 3つの手書き文字パターン

次に、シフト変換と歪み変換で生成された図 4.1 のような教師信号の文字パターンのサイクルを、ネットワークの軌道上へ学習する方法を説明する。まず、各文字に対する教師信号の連続変形パターンのサイクルを、いくつかの区間で時間的に分割する。各区間の両端の時刻では、メッシュ特徴に変換前の 2 値画像にシフト変換と歪み変換を行った 2 つの文字パターンを考え、区間内の各時刻における連続的な変形パターンを、その 2 つのパターン間の線形補間（スプライン補間でも可）で定義する。時刻  $T$  から  $T+TB$  までの 1 つの区間では、上記の方法で定義された文字パターンをメッシュ特徴に変換した教師信号  $Q_i(t)$  で、各興奮性素子  $x_i(t)$  の値を固定する。この時、以下のような誤差関数  $E$  の最急降下法に従った結合重みパラメータの学習を考える。

$$E = \frac{1}{2} \int_T^{T+TB} \sum_{i=1}^N J_i(t)^2 dt, \quad (4.3)$$

$$J_i(t) = \dot{Q}_i(t) + Q_i(t) - G(v_e^{(i)}(t)), \quad (4.4)$$

$$v_e^{(i)}(t) = \sum_{i=1}^N W_{ij} Q_i(t) - K_{EI}^{(i)} y_i(t) + I_i, \quad (4.5)$$

$$\Delta W_{ij} = -\eta \frac{\partial E}{\partial W_{ij}}. \quad (4.6)$$

上記の強制入力  $J_i(t)$  は、各興奮性素子の活性値  $x_i(t)$  がその教師信号の値  $Q_i(t)$  に等しくなるように定義されている ( $J_i(t) \rightarrow 0$ ;  $x_i(t) \rightarrow Q_i(t)$ )。また、 $\eta$  は学習速度を規定する学習係数である。付録 C のように、佐藤 (1990) の学習アルゴリズムの導出に従って、以下の学習方程式が得られる ( $P_i(t)$  は Lagrange 乗数)。

$$P_i(t) = -J_i(t), \quad (4.7)$$

$$\Delta W_{ij} = -\eta \int_T^{T+TB} [P_i(t) G'(v_e^{(i)}(t)) x_j(t)] dt, \quad (4.8)$$

$$\Delta K_{EI}^{(i)} = \eta \int_T^{T+TB} [P_i(t) G'(v_e^{(i)}(t)) y_i(t)] dt. \quad (4.9)$$

1つの文字パターンのサイクルに対する学習は、そのサイクル中で分割された複数の区間における結合重みパラメータ値の修正の積み重ねとなる。こうした処理が各文字に対して、その誤差関数の値が十分小さくなるまで繰り返される。

図 4.3 と図 4.4 の左上から右下に、学習前の自己相関行列の結合重み係数  $W_{ij}$  を持つネットワークが生成した出力パターン系列を示す。図 4.3 は、記憶パターンに近い入力に対する出力パターン系列が、文字として読めないものを含んではいるが、それが周期的であることを示している。図 4.4 は、記憶パターンから離れたランダム入力に対する出力パターン系列が、3つの記憶パターンやそれらの白黒反転パターンの間を複雑に遊走したカオス的なものであることを示している。その遷移は、反転の「う」(27-29 番目) → 「う」(35-38 番目) → 反転の「い」(46-49 番目) → 「い」(55-58 番目) → 「あ」(60-62 番目) → 反転の「い」(65-67 番目) → 反転の「あ」(69-70 番目) となっている。

次に、各文字に対して結合重みパラメータの反復的な修正を 1,000 回行った時の、学習後の出力パターン系列を図 4.5 に示す (学習係数は  $\eta = 0.0001$  とした)。この時、学習の反復回数が 1,000 回を越える前に、3文字に対するサイクルについての誤差関数値の総和は収束した。図 4.1 は、記憶パターンに近い入力に対しては、学習によって埋め込まれた教師信号の変形パターンのサイクルが再現されていることを示している。他の 2 文字についても同様な結果となった。一方、記憶パターンから離れたランダム入力に対しては、学習前と同様な記憶パターン間を遊走するカオスの軌道が得られた。

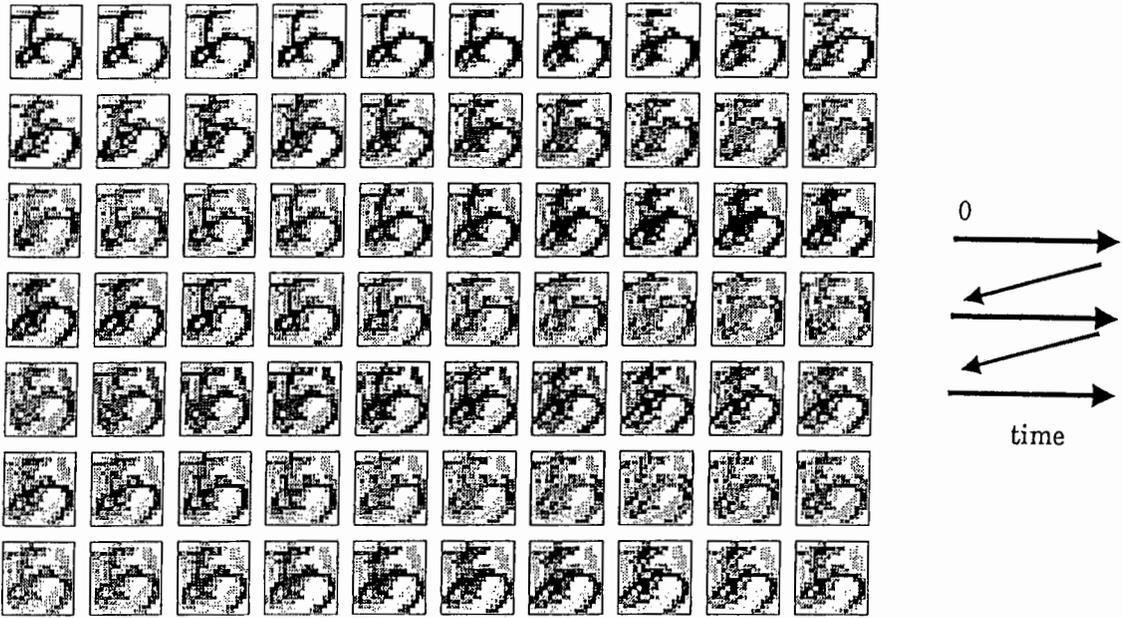


図 4.3: 記憶パターンに近い入力に対する学習前の出力パターン系列 (文字として読めないパターンを含む、記憶パターン付近のリミットサイクル)

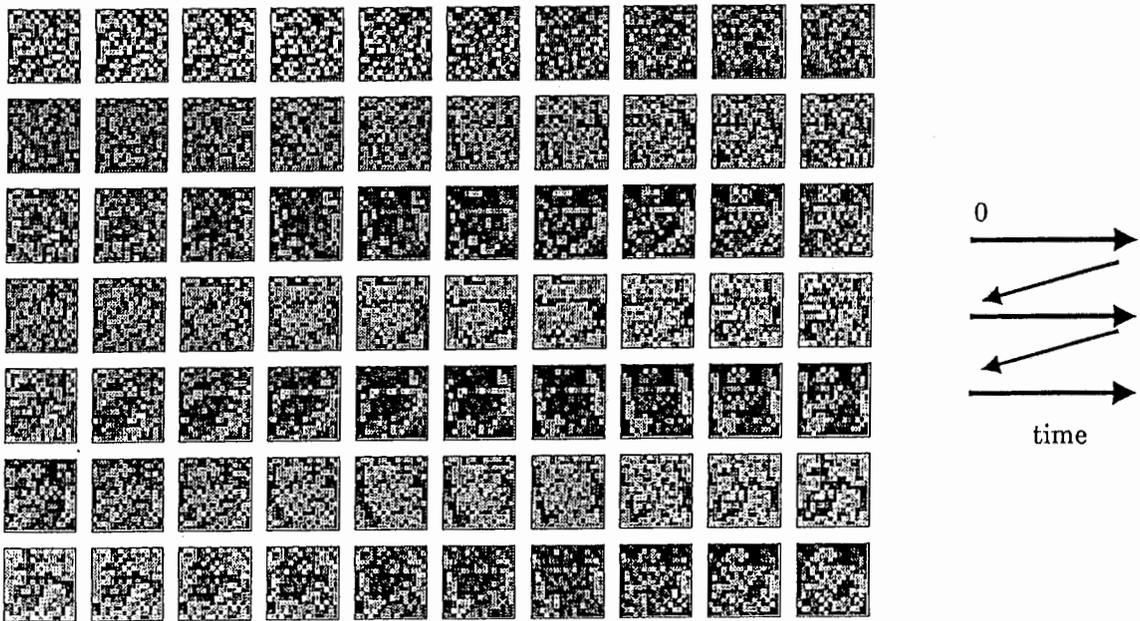


図 4.4: 記憶パターンから離れた入力に対する学習前の出力パターン系列 (記憶パターン間を遊走するカオスの軌道)

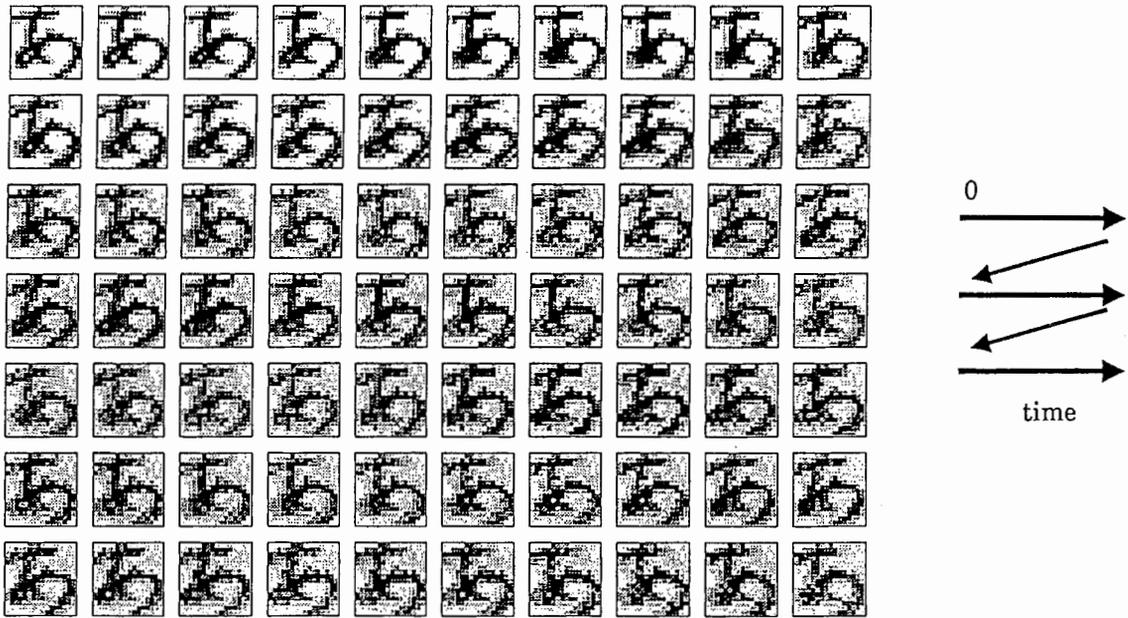


図 4.5: 記憶パターンに近い入力に対する学習後の出力パターン系列 (教師信号である文字の変形パターンのサイクルを再現)

以上のように、学習後のネットワークでは、入力文字に対するカテゴリー内の許容範囲を示すような変形サイクルを生成する機能が実現出来た。この変形サイクルの軌道上のパターンを、ネットワーク外部の識別系への認識候補として入力すればパターン認識に利用できる。さらに、このネットワークは、学習によって埋め込まれた軌道以外にも、記憶した文字とは異なる入力に対して、それらの文字パターン間をさまようカオス的軌道を生成出来る。このようなカオス的軌道は、1つのカテゴリー内の許容範囲を越えた変形パターンを示すものである。ゆえに、外部の識別系に対するその機能は、不確定な入力に対する自律的な探索、あるいは識別不能状態 (Sakarda, & Freeman, 1987) を示すものと考えられる。すなわち、記憶した文字パターンに対する入力パターンの類似性に従って、1つの文字カテゴリーの許容範囲を示すようなパターンの生成機能と、カテゴリー間の自律的な探索機能とのスイッチングが行われると言えよう。

### 4.3 軌道の多様性と可制御性を利用した顔パターンの生成

#### 4.3.1 対話的処理による図形パターンの概念形成の支援

3.4節に示した発振ニューラルネットワークの軌道の多様性と可制御性の特徴を、時空間パターン処理に積極的に利用した1つの応用例として、本節ではパターン生成の課題を考える。この課題は、図4.6のように外部評価者であるユーザと発振ニューラルネットワークのパターン生成器との対話的な処理によって実行される。すなわち、ユーザは自分の意図したパターンを生成させるよう、パターン生成器の出力パターン系列から候補を選んで、軌道のサーチ領域を制御する。一方、パターン生成器は、選ばれた候補がどの程度共通した確かなものであるかというユーザの概念レベルに応じながら、さまざまなパターンを生成する。こうした対話的な処理は、曖昧な初期イメージから最終的なイメージに段々と確定していくような、ユーザの概念形成過程 (Oda, 1991) を支援するものと考えられる。

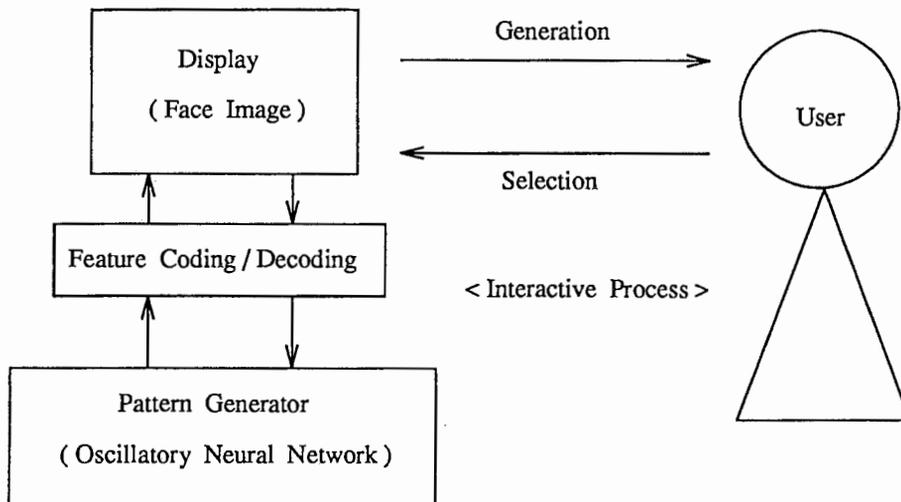


図 4.6: ユーザとパターン生成器との対話的処理

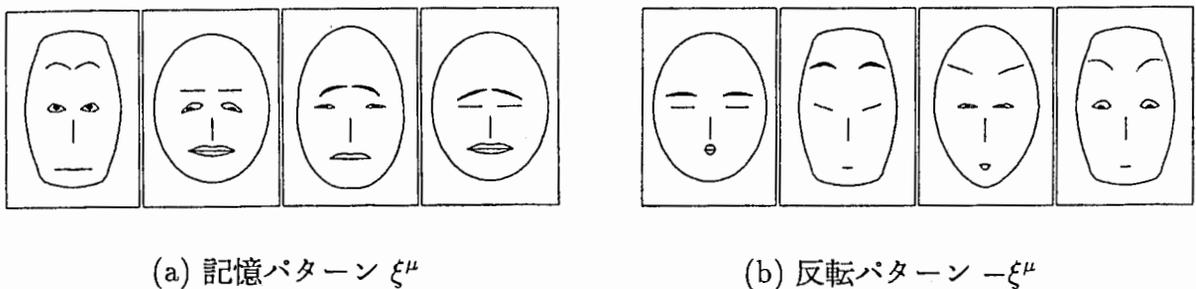


図 4.7: 各記憶パターンの顔イメージ ( $\mu = 1 \sim 4$ )

上記のようなパターン生成は、軌道の複雑さを多様なパターンの生成機能に対応付け、外部入力による軌道の性質の変化を探索領域の制御に利用したものである。また、結合重みパラメータ  $K_{IE}$  の値に従った、時間変化を伴う軌道上の候補パターンの探索機能から、頂点付近の漸近安定平衡点への収束による最終的なパターンの確定機能への機能的なスイッチングも可能となっている。さらに、3.5節で示された、記憶パターン間やそれら以外の頂点間における軌道の遊走性の特徴を多様なパターンの生成に利用している。すなわち、ネットワークが生成するパターンは、コンピュータグラフィックス等で通常用いられる2パターン間の補間 (Dewdney, 1986; Murakami et al., 1993) というよりは、記憶パターンやそれら以外の頂点のパターンを複雑な比率で混合したパターンとなるので、より多様なパターンが生成出来る。このような頂点間の中間的なパターンをも対象とすることは、頂点に位置する記憶パターンのみを対象としたパターン検索の課題 (Nara et al., 1992, 1993a, 1993b) とは異なる点である。以下では、顔パターンの生成を考えて、ユーザの3つの概念レベル、(i) 曖昧な初期イメージの段階、(ii) 中間的なイメージの段階、(iii) 最終的な確定イメージの段階についてのパターン生成例を示す。

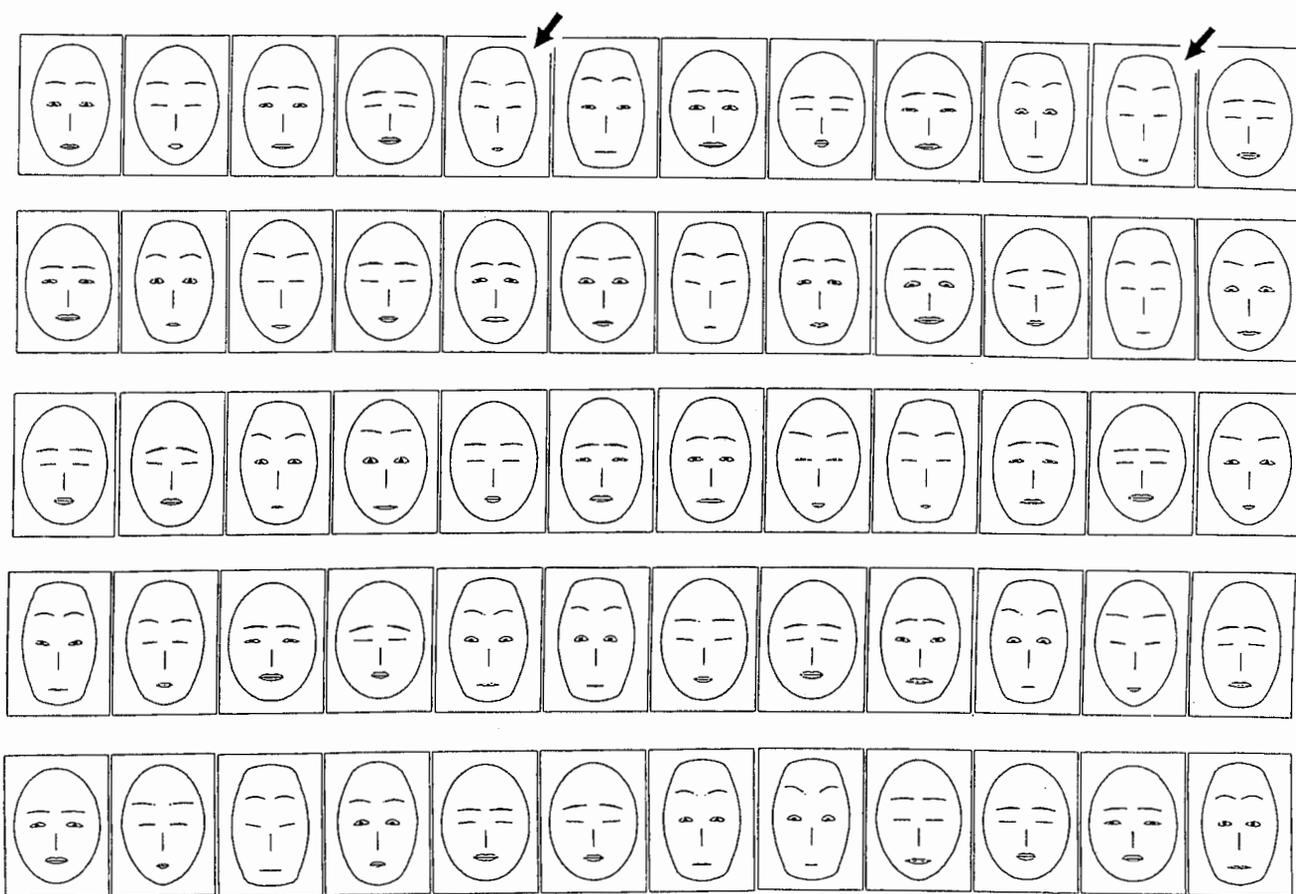
#### 4.3.2 概念イメージに応じたパターン生成

顔パターンの表現方法として、顔の識別にとって影響力の大きい目などの構成要素 (Hara, 1978) を重視した特徴要素による表現を考えた。ここで、目や口などの構成要素の標準的な配置関係を、成人なら年齢や性別に依存しない人間の頭部のカノン (Parramon, 1980) に従って決めることにした。これらの特徴要素は、発振ニューラルネットワークの22個の興奮性素子の入出力に対応するもので、その値に従って位置や大きさなどが変えられる (詳細は付録D参照)。このような特徴表現は、高次元の状態空間内のパターンの位置を視覚的に理解するのに適した表現でもある (Chernoff, 1973; Hara, 1978)。上記の特徴表現のおかげで、ネットワークが生成するアナログ的な出力パターンは、常に顔のイメージを表すことになる (但し、顔のイメージのみ)。一方、本提案モデルと類似した挙動を持つ奈良らのモデル (1993b) においては、顔パターンの直接的な画像表現 (ビットパターン) が用いられている為、記憶パターン間の中間的なパターンが顔を表現しないことがあり得る。さらに、彼らのモデルにおける2値化出力では、なめらかに連続変化するパターンの生成は不可能である。

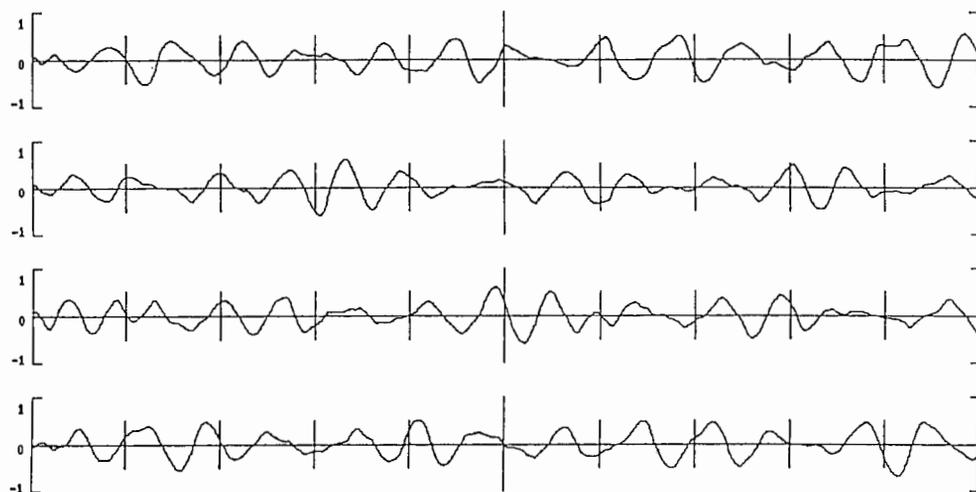
さて、4つの特徴的な顔を記憶パターンとして、その自己相関行列で結合重み係数  $W_{ij}$  を設定する。図4.7はこれらの記憶パターンの顔イメージを示すもので、(a) 記憶パターン  $\xi^1 \sim \xi^4$  と、(b) それらの反転パターン  $-\xi^1 \sim -\xi^4$  である。

曖昧な初期イメージの段階では、カオスの軌道上でさまざまな顔パターンが生成される。例として、図 4.8(a) に、1,500 回の反復中のサンプリング時間  $T_s = 25$  ごとの出力パターン系列を示す。また、図 4.8(b)(c) には、上記に対応するオーバーラップ関数  $m^\mu(t)$  ( $\mu = 1 \sim 4$ ) と、軌道の最近接頂点の遷移を示す。これらから、図 4.7 の記憶パターンやそれら以外の頂点間を、軌道が複雑に遊走していることが類推出来る。このような多様なパターンから、ユーザはいくつかの候補を選択し、それらの選択されたパターンの平均で入力バイアスを設定する。また、初期状態  $\{x_i\}$  に対する軌道の依存性を多様なパターンの生成に利用する為、パターン生成の試行ごとにその状態値を乱数などでリセットする。これと同時に、3.4 節の式 (3.22) における等価的な外部入力の影響をなくす為、状態  $\{y_i\}$  をゼロにリセットする。ある試行で例え候補が見つからなかったとしても、これらのリセット後に生成された前回とは異なった軌道上では、候補が含まれる可能性を持つ。上記の選択（入力バイアスの設定）後は、選択されたパターン近くに若干限定されたパターンが生成されるようになる。このような入力バイアスによる探索領域の制御は、奈良ら (1993a) のモデルにおける適応的なしきい値の強化学習と同様な効果を持つ。

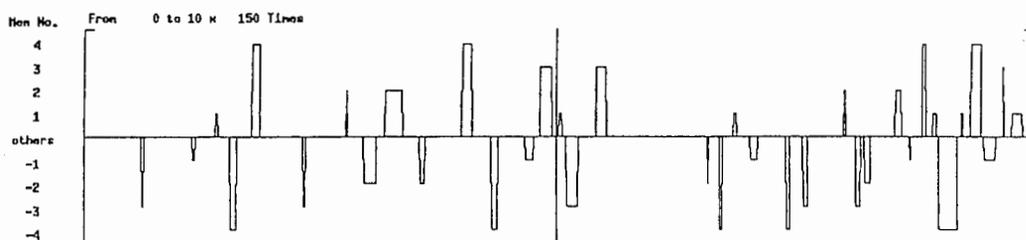
いくつかの対話的処理を通じた中間的な段階では、図 4.9(a) に示すような選択された候補パターン付近にある程度限定されたパターンが生成される。図 4.9(b)(c) に、この時のオーバーラップ関数  $m^\mu(t)$  ( $\mu = 1 \sim 4$ ) と、軌道の最近接頂点の遷移を示す。これらは、図 4.8(b)(c) に比べてより周期的になっている。入力バイアスが数少ない頂点のパターンに近付く程、軌道の探索領域はこれらのパターン付近に限定される傾向にあった。このように、限定された探索領域内での候補の選択を通じて、ユーザは自分の概念イメージを段々と確定したものにしていく。



(a) 時系列出力の顔パターン (矢印は選択されたパターンを示す)

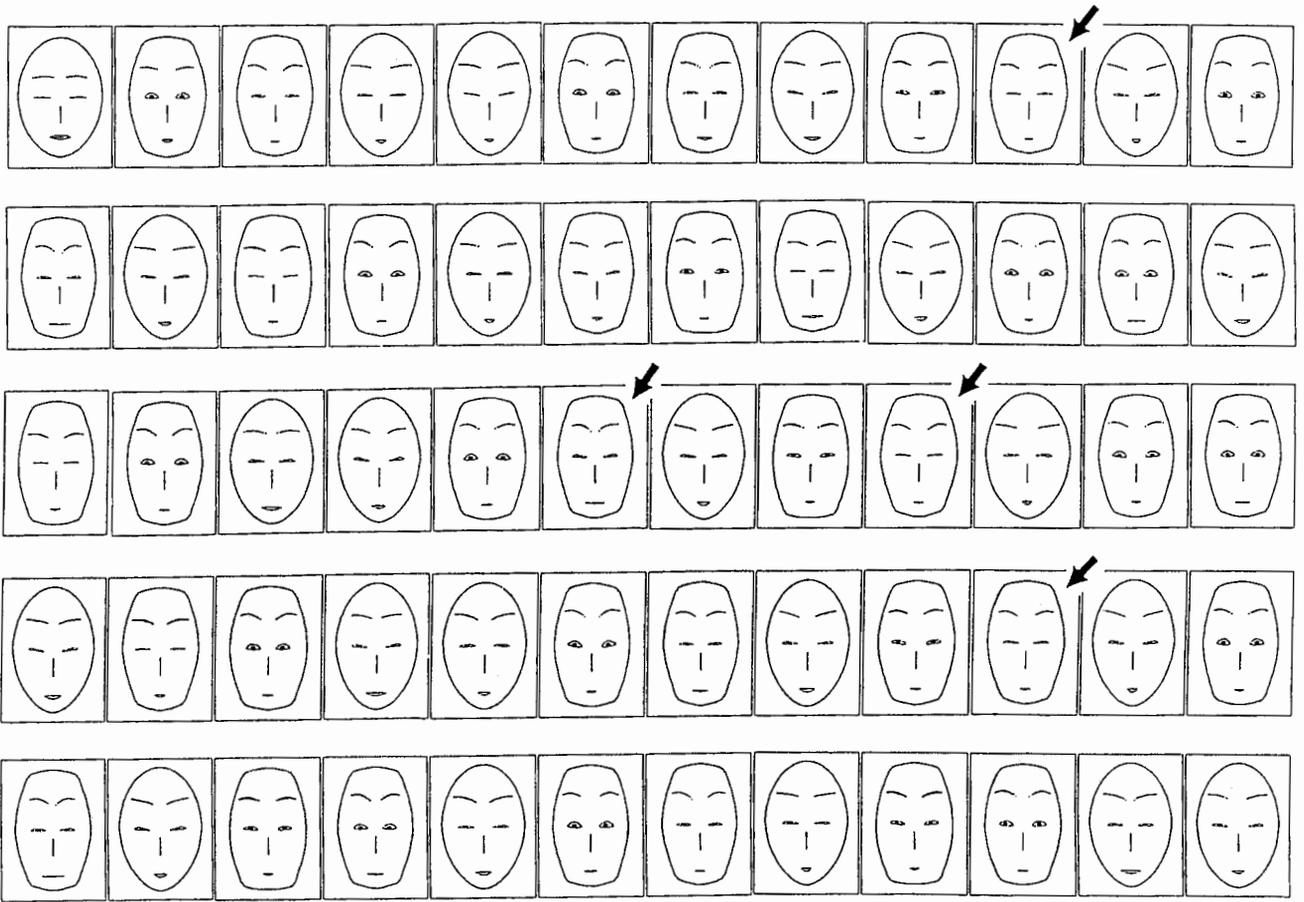


(b) オーバーラップ関数  $m^\mu(t)$

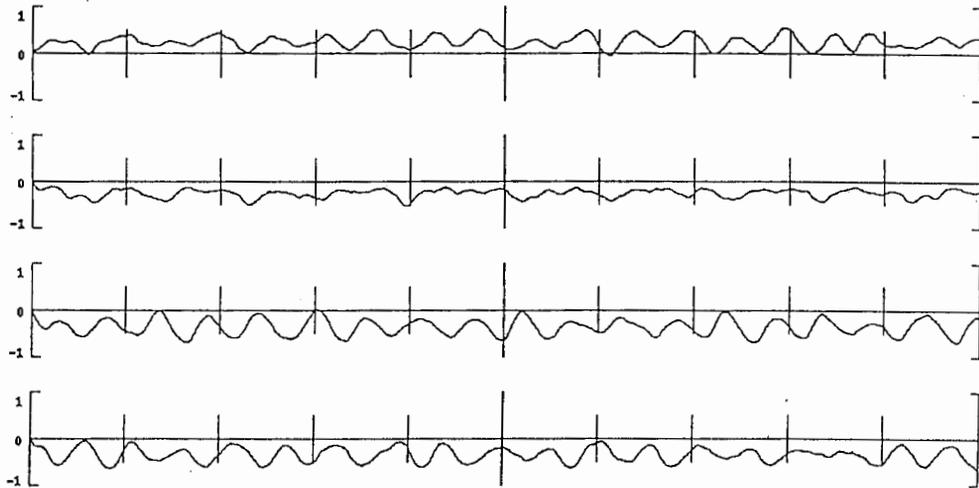


(c) 軌道の最近接頂点の遷移

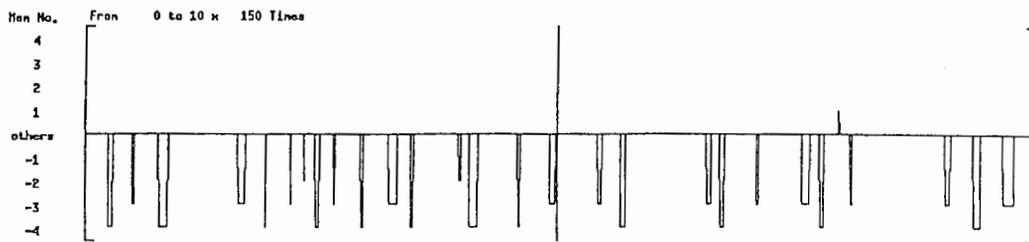
図 4.8: 曖昧な初期イメージ段階のパターン生成例



(a) 時系列出力の顔パターン (矢印は選択されたパターンを示す)



(b) オーバーラップ関数  $m^\mu(t)$

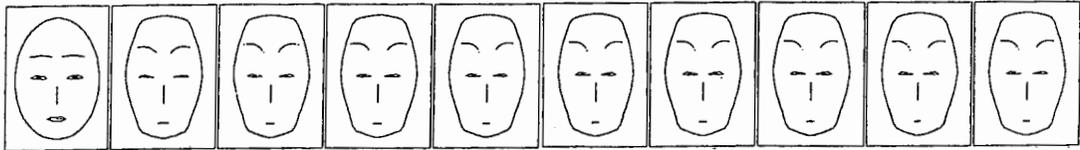


(c) 軌道の最近接頂点の遷移

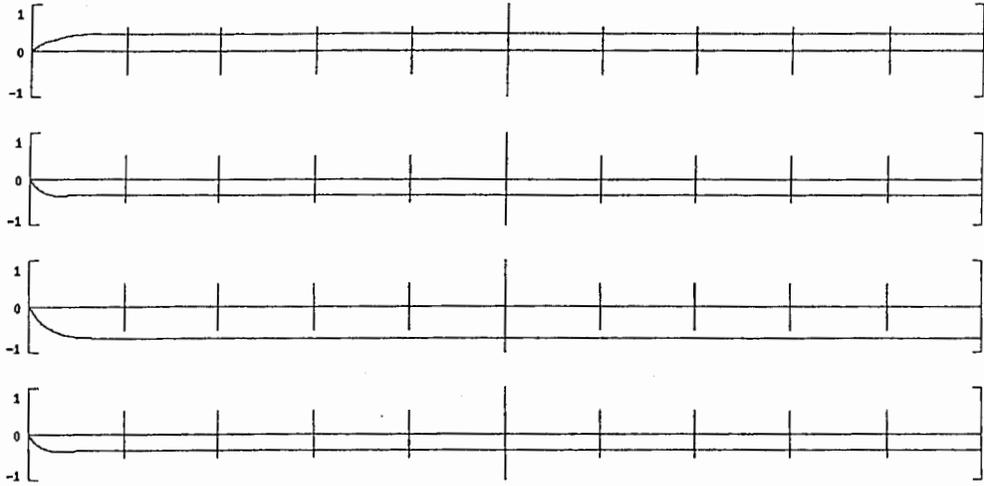
図 4.9: 中間的なイメージ段階のパターン生成例

概念イメージがほとんど確定された最終的な段階で、もし、ユーザがある記憶パターンもしくはそれら以外の頂点の1つのパターンを獲得したい場合は、結合重みパラメータ  $K_{IE}$  の値を小さくすれば良い。なぜなら、この時、軌道はその頂点近くのリミットサイクルや漸近安定平衡点に収束するからである。記憶パターンとは異なる頂点のパターン近くの漸近安定平衡点に収束した場合の、時系列出力の顔パターン、オーバーラップ関数、軌道の最近接頂点の遷移を図4.10の(a)(b)(c)にそれぞれ示す。もちろん、顔パターンのイメージを確定する前に、ワープロなどのイラストツールのように、ユーザが中間的な段階のパターンに部分的な編集を施すことも出来る。また、初期段階から顔の特徴要素に対して適当な値が分かっている場合は（例えば、大きな目、つり上がった眉毛、とがった細面など）、それらの値を入力バイアスに始めから設定した方が良いでしょう。

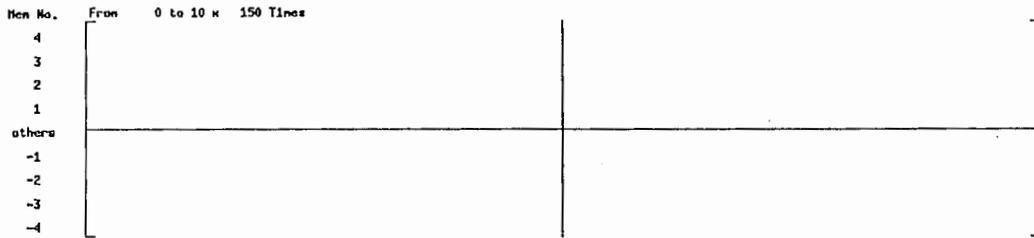
上記のようなパターン生成の有効性を評価するにあたっては、残念ながら適当な評価尺度がないのが現状である。その理由は、上記の処理過程が生成されたパターン系列に影響された概念形成 (Oda, 1991) を伴うことによる（それは、パターンの集合のみならず、系列の順序にも依存する）。すなわち、それがパターンの選択と生成との対話的な処理を通じた動的な過程である為に、最終的なイメージはあらかじめ予想できずに段々と変化するばかりか、特別な場合を除いて、そのイメージは有限個のパターン ( $2^N$  個の頂点) に限定出来されない中間的なパターンとなるからである。こうした評価尺度の問題と関連して、パターン生成を効率的に実行するのに適した記憶パターン数  $M$  も明らかでない。その最適数は、おそらく、ネットワークの内的なダイナミクスの特徴と、課題からの外的な要求との適合度に依存するものと思われる。



(a) 時系列出力の顔パターン



(b) オーバーラップ関数  $m^\mu(t)$



(c) 軌道の最近接頂点の遷移

図 4.10: 最終的なイメージ段階のパターン生成例

しかしながら、定性的には、ネットワークの内的なダイナミクスに基づくパターン生成はランダム生成に対する以下のような長所を持つと考えられる。第1に、内的なダイナミクスに基づく方法は軌道上を連続的に変化するパターンを生成するので、ユーザのイメージにとって自然であるが、ランダム生成による方法では状態空間内であちこち飛び飛びの不自然なパターン系列を生成してしまうこと。第2に、内的なダイナミクスに基づく方法で生成されたパターンは、ランダム生成のような一様なものではなく、記憶パターンに近いものがより多く生成されるなどのダイナミクスの特徴に応じて偏ったパターンが生成されることである。与えられた課題が記憶パターンを核とした探索を要求するものならば、このような記憶パターン近くに偏ったパターンを（探索経路を規定しなくても）生成する特徴がある方が有利となる。これに対して、ランダム生成で同様な探索を行うには、記憶パターン間の探索経路を何らかの方法で規定して、これにランダムノイズによる揺らぎを付加しなければならない。いずれにしろ、効率的なパターン生成にとっての、ランダム生成との定量的な比較は今後の検討課題である。

#### 4.4 状態空間構造の変化による識別領域の動的制御法

人間にとって、単独ではどちらとも判断できる曖昧な文字でも、高次の文脈情報である単語などの知識からそれを解釈できる場合がある（例えば  $/-\backslash$  は、 $C/-\backslash T$  なら  $A$ 、 $T/-\backslash E$  なら  $H$  と解釈可能）。この際、時には強引な拡大解釈もなされるが、いったんある解釈に落ち着けば、欠損や汚れなどの部分的な対応付けも可能なることが多い。もし、人工的なシステムで、高次の文脈情報（単語）からの予測に従って、低次（文字）のパターン空間内の識別領域を制御することが出来れば、このような文脈に依存した解釈は実現可能となる。そこで、前章で示した発振ニューラルネットワークの挙動の特徴を利用して、予測に相当する外部入力に従ったベイスンの拡大や縮小によって、識別領域を動的に制御する方法を提案する。しかしながら、上記のモデルでは、外部入力のわずかな変化で平衡点の位置や個数が変化し、その結果、状態空間内に存在するアトラクター数が膨大となってしまうので、ベイスンと識別領域との対応付けが難しくなる。この問題に対して、4.4.1では、外部入力を無くす代わりにこれと等価な効果を持つ抑制性素子状態を予測に従って変化させるとともに、状態空間内に限られた数のアトラクターしか持たない改良モデルを提案する。4.4.2では、このことを3組のペア素子のモデルにおいて実験的に確認する。さらに、4.4.3では、1つの例として、単語からの予測による文字パターンの識別領域の動的な制御を示す。

#### 4.4.1 入力確信度に応じた揺らぎ出力を持つ連想記憶モデル

前章で議論した発振ニューラルネットワークは、 $\{x_i, y_i\}$  の  $2N$  次元の状態空間内にアトラクターを持つが、一般にその位置や個数、形状などは外部入力によって変化する。一方、外部入力を無くせば、 $2N$  次元の状態空間内にアトラクターを固定することが出来る。幸いにも 3.4 節の式 (3.21) から、各抑制性素子の初期状態  $y_i(0)$  は、それとペアをなす興奮性素子  $x_i$  に対する外部入力と等価的な効果  $I_i(t) \equiv -K_{ei} \exp^{-t} y_i(0)$  を持つので、外部入力がない場合でも、この状態値によって  $N$  次元状態空間  $\{x_i\}$  におけるベイスン構造を変えることが出来る。この時、 $N$  個の超平面を規定する初期状態  $\{y_i(0)\}$  によって、 $N$  次元状態空間上に射影されたアトラクターの数は変化するが、状態空間内に射影されたそれらのアトラクターの位置は当然固定される。図 4.11 は、このような初期状態  $\{y_i(0)\}$  による  $N$  次元状態空間の構造変化を、超平面で輪切りにされた「木の年輪」のようなもので表した模式図である。

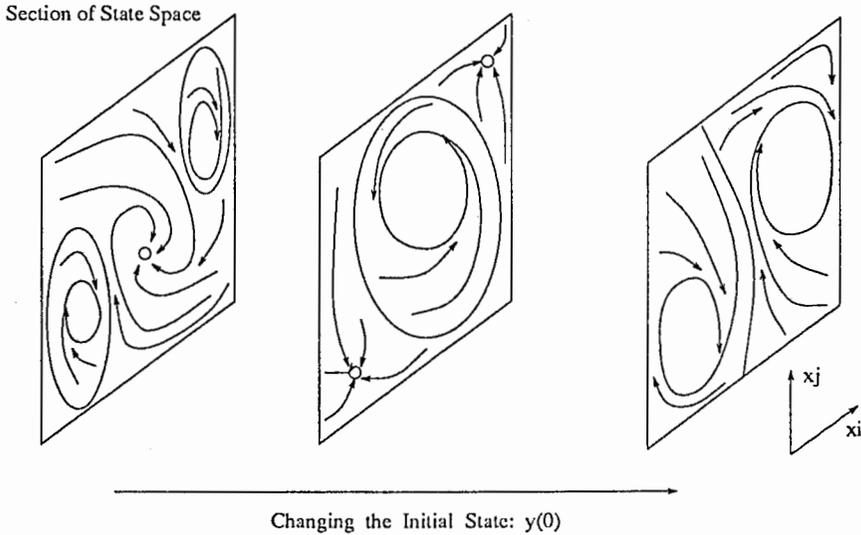


図 4.11: 制御パラメータに従った状態空間構造の変化

これらの議論に基づいて、以下のように改良された連想記憶モデルを考える。

$$\dot{x}_i(t) = -x_i(t) + G\left(\sum_{j=1}^N W_{ij} x_j(t) - K_{EI} y_i(t)\right), \quad (4.10)$$

$$\dot{y}_i(t) = -y_i(t) + G(K_{IE}(x_i(t)) x_i(t)), \quad (4.11)$$

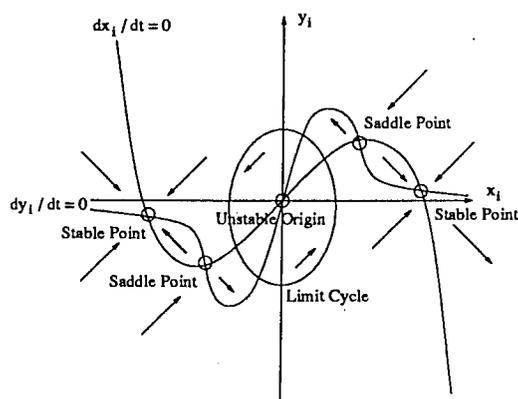
$$W_{ij} = \frac{1}{N} \sum_{\mu=1}^M \xi_i^\mu \xi_j^\mu, \quad (4.12)$$

$$G(z) = \frac{2}{\pi} \arctan\left(\frac{z}{a}\right), \quad (4.13)$$

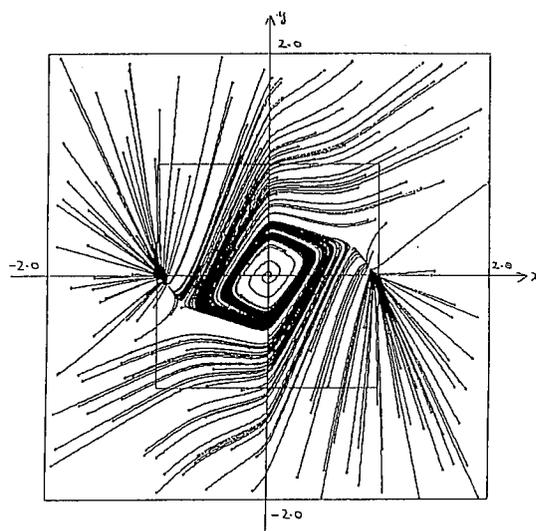
$$K_{IE}(z) = K_{max} \exp^{-\delta z^2}. \quad (4.14)$$

前章の発振ニューラルネットワークと上記のモデルとで異なる点は、結合重み係数  $K_{IE}$  を興奮性素子状態  $x_i$  についての関数としたことである ( $K_{max}, \delta$  は関数の非線形性に関するパラメータ)。また、自己結合が強くなりすぎないようにもしてある (式 (3.4) における  $\delta_{ij}$  が無い)。これらの変更は、外部入力が無くても、状態空間内に漸近安定平衡点やリミットサイクル、カオスの軌道などの複数のアトラクターが共存出来るようにする為に行った。このことは、以下の解釈から支持される。また、4.4.2の実験結果はこれを裏付ける1つの例となる。

単一のペア素子において、パラメータ  $K_{max}, \delta$  が付録 E の条件を満足する時、 $x_i - y_i$  平面には図 4.12 のような原点周辺のリミットサイクルと  $x_i = \pm 1$  付近の2つの漸近安定平衡点が共存すると考えられる。一方、複数のペア素子を結合した場合は、各ペア素子が自律振動子となり、これらが周波数ロッキングなどの複雑な引き込みの相互作用を起こす (Bergé, Pomeau, Vidal, 1984)。但し、単一のペア素子にとっては、他のペア素子からの入力信号で Nullcline  $\dot{x}_i = 0$  の曲線が ( $y_i$  軸方向の上下に) 摂動されると解釈出来るので、複数のペア素子を結合した場合でも、原点の不安定平衡点がなくなって漸近安定平衡点のみになることがあり得る。そこで、出来るだけこれを避けるよう、自己結合重み係数  $W_{ii}$  の値を小さくして、 $x_i = \pm 1$  付近の平衡点の安定性を弱めたのである。これらより、記憶パターン付近の漸近安定平衡点やリミットサイクル、原点を中心にして記憶パターン間を遊走するカオスの軌道などが状態空間内に存在することが期待出来る。この際、図 4.12(a) のようなサドル点は、ベイスンの分割に寄与すると考えられる。



(a) 平衡点の安定性



(b) 状態空間内の軌跡

図 4.12: 単一の興奮 - 抑制性ペア素子における挙動

さて、上記のモデルの時間変化を伴う出力軌道の多様さを情報処理に利用するため、以下のような対応付けを考えた。まず、出力の記憶パターンからの変動量を、入力に相当する興奮性素子の初期状態の記憶パターンに対する類似度（確信度）に応じた揺らぎと解釈する。すると、ネットワークの機能は入力の確信度に応じて、小さな揺らぎを持った記憶パターンの想起や、大きな揺らぎによる記憶パターン間の自律的なサーチ（または不確定状態）といった具合にスイッチング出来る。すなわち、入力の情報は軌道上の揺らぎという時間的な情報に保持されるのである。また、記憶パターン付近のアトラクターで定義された識別領域は、文脈情報からの予測に従ったベイスンの拡大や縮小で制御される。このような動的な制御方法では、高次の文脈情報と低次のパターン情報との相互作用（McClelland, & Rumelhart, 1981）という枠組を越えた、ダイナミクスの情報処理的な意味付けが出来る。

上記のようなベイスンの拡大や縮小によって識別領域を変化させるというアイデアは、失語症患者の症状を説明する人工的なニューラルネットワークにおいても指摘されている（Hinton, & Shallice, 1991; Hinton, Plaut, & Shallice, 1993）。ここで失語症とは、脳に損傷を受けた為に書いてある文字が正しく読めなくなるという症状で、それには視覚的に類似した単語に読み誤る場合と、意味的に関連のある単語に読み誤る場合との両方がある。彼らは、単語内の文字の位置を表す入力素子と単語の意味を表す出力素子を持つ3層のネットワークに、出力の単語の意味が正確に合致するよう調整を行う整理素子を設けて、この整理素子を介したフィードバックループを付加したモデルを考えた（但し、平衡点に収束するタイプ）。ここで、意味情報は単語に対する高次の文脈情報となることに注意しよう。このネットワークの学習後に、損傷に相当する素子間の結合の断線を施したモデルで、失語症患者の主な症状のみならず細かい特徴までが再現されたのである。さらに、この結果が意味空間におけるベイスンの大きさの変化（損傷が大きい時はベイスンの融合もある）で説明出来ることが主張されている。

これとは別の類似例として、外部入力に対する状態空間構造の変化によるアトラクターのスイッチングを、連続音声データからのキーワード検出に利用することが検討されている（長谷川, & 稲積, 1993; 佐藤, 久野, & 郷原, 1993）。そこでは、学習された時系列入力パターンに対応したアトラクターの形成とそれらのアトラクター間の遷移によって、話者の違いなどに対する非線形の時間的伸縮が実現できることが実験的に示されている。

#### 4.4.2 ベイソン拡大 / 縮小による状態空間構造の変化

連想記憶モデルのベイソンは入り組んだ複雑な形状 (Amari, & Maginu, 1988) と考えられる為、一般にその理論解析は困難となる。そこで、4.4.1の議論を検証する1つの例として、3組の興奮-抑制性ペア素子からなるモデルのベイソン構造を数値計算によって調べた。すなわち、3次元の状態空間  $\{x_1, x_2, x_3\}$  がどのようなアトラクターで分割され、初期状態  $\{y_i(0)\}$  に従ってそれらのベイソンがどう変化するかについて解析した。

まず、モデルのパラメータを前章と同じ値で設定する。但し、本章で付加したパラメータについては、付録 E の条件を満たすように  $K_{max} = 2.0, \delta = 6.0$  とした。高次の文脈情報からの予測に従って初期状態  $\{y_i(0)\}$  を制御する事を考え、仮に記憶パターン  $\xi^1$  が予測されたものとして、 $y_i(0) \equiv -\alpha \xi_i^1 / K_{EI}$  とした (予測の強さを規定するパラメータ  $\alpha$  は、0.0, 0.5, 1.0, 1.5, 2.0 の5段階の値とした)。各段階について、3次元立方体 ( $-1 \leq x_i \leq 1$ ) 内の各点の初期状態  $\{x_i(0)\}$  から、式 (4.10) と式 (4.11) の微分方程式を差分近似で反復計算した (刻み幅  $\Delta x_i, \Delta t$  はそれぞれ 0.1, 0.01 とした)。過渡状態を除いたアトラクターを抽出する為、5万回の反復計算中の最初の4万回を除いて、残りの1万回分の出力  $\{x_i(t)\}$  の軌道を調べた。この時、以下の4種類のアトラクターに分類出来た。

1. 各記憶パターン (または反転パターン) 近くの漸近安定平衡点
2. 記憶パターンと反転パターンとの間の直線振動型リミットサイクル
3. 各記憶パターン (または反転パターン) 付近の  $\infty$  の字型リミットサイクル
4. 記憶パターンや反転パターン点間を複雑にさまようカオスの軌道

図 4.13の左端に上記の4種類の各アトラクターの形状を示し、その右から順に5段階の  $\alpha$  の値に対する各アトラクターのベイソンの変化を示す。図 4.13の上から順に、(a1) 記憶パターン付近の漸近安定平衡点 (3つ重ねて表示)、(a2) 反転パターン付近の漸近安定平衡点 (3つ重ねて表示)、(b1) 記憶パターン  $\xi^1$  付近の直線振動型リミットサイクル (2つ重ねて表示)、(b2) それ以外の直線振動型リミットサイクル ( $\xi^2$  に関する物のみ表示)、(c1) 記憶パターン  $\xi^1$  付近の  $\infty$  の字型リミットサイクル (c2) それ以外の記憶パターン付近の  $\infty$  の字型リミットサイクル ( $\xi^2$  に関する物のみ表示)、(c3) 反転パターン  $-\xi^1$  付近の  $\infty$  の字型リミットサイクル、(c4) それ以外の反転パターン付近の  $\infty$  の字型リミットサイクル ( $-\xi^2, -\xi^3$  の2つに関する物を重ねて表示)、(d) カオスの軌道、に対する各ベイソンの変化をそれぞれ示す。さらに、この時のベイソンの体積変化を図 4.14に示す。図 4.14と図 4.13から、予測として  $\{y_i(0)\}$  に与えられた記憶パターン  $\xi^1$  に関するアトラクターのベイソンが拡大し、それ以外は縮小している事がわかる。

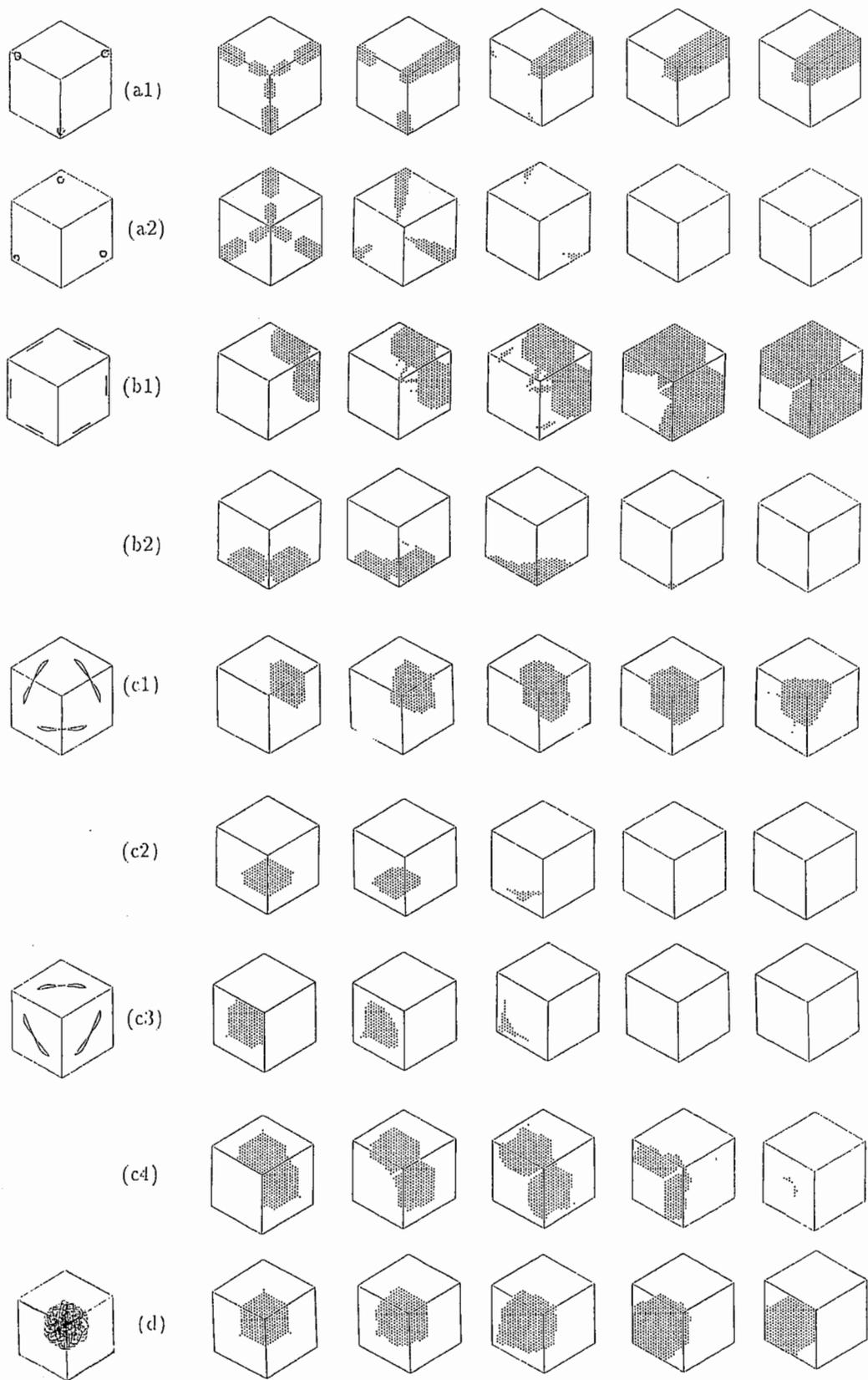


図 4.13: 4 種類のアトラクター (左) と各ベイスンの変化 (左から右に 5 段階の  $\alpha$  の値に対応)

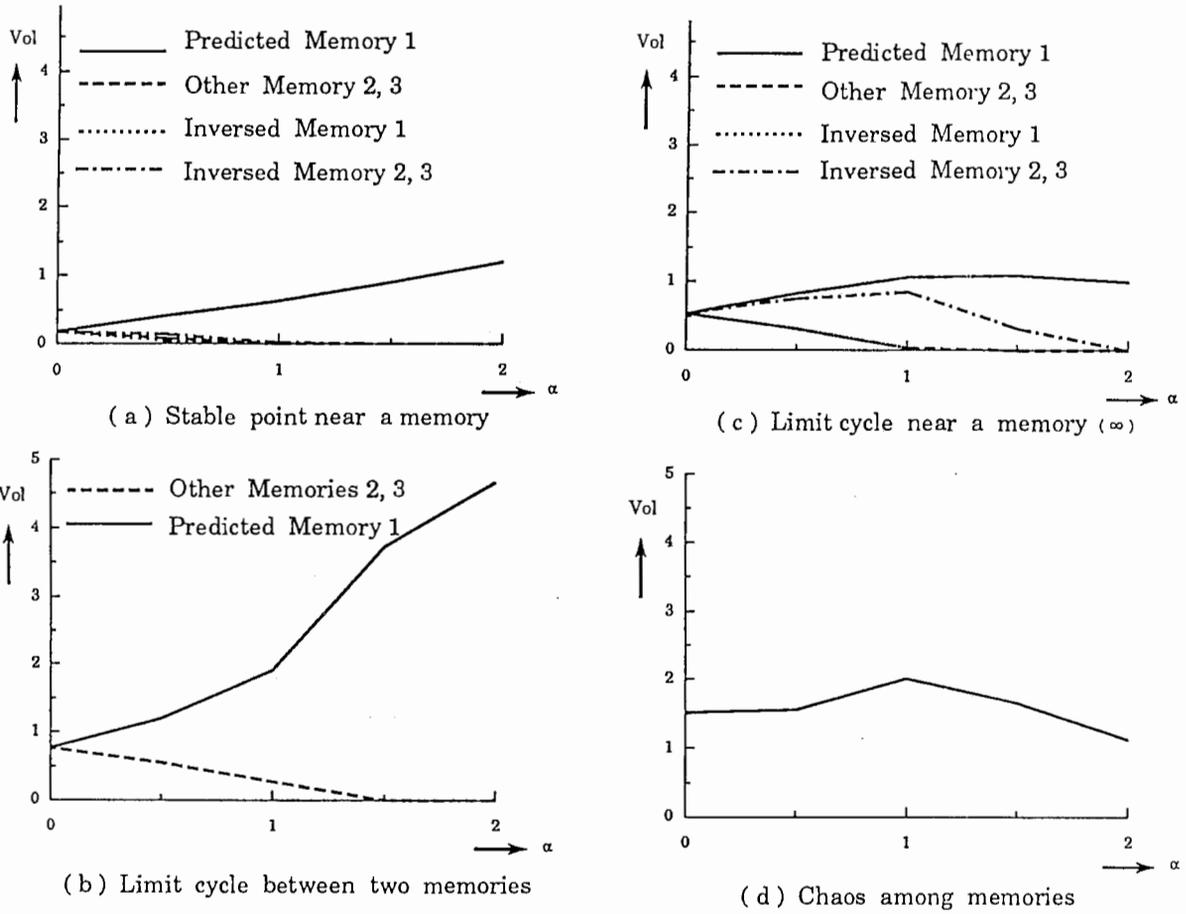
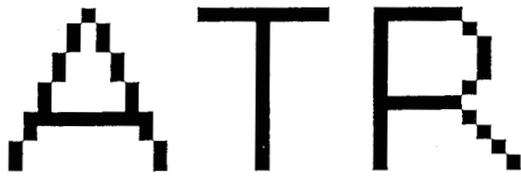


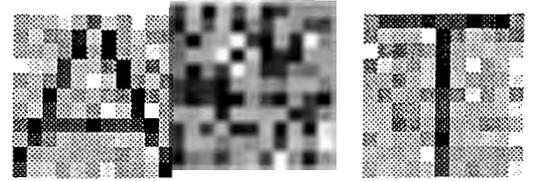
図 4.14: 予測の強さ  $\alpha$  とアトラクターのベイジンの体積 (但し、b1 と b2 のみは 2 つのアトラクターに対する体積和を示す)

#### 4.4.3 単語からの予測による文字の識別領域の動的制御

高次元のモデル ( $N > 3$ ) の状態空間においても、抑制性素子の初期状態に従った記憶パターン付近のベイジンの拡大や縮小が起こることが推測出来る。以下では、このようなベイジンの変化で識別領域を動的に制御する例として、単語からの予測に適応した文字パターンの識別について検討する。まず、3つのアルファベット文字: A, T, R で構成される 5 つの単語: ATR, ART, RAT, TAR, ATT を対象として考える。3文字の記憶パターンは図 4.15(a) のような、 $N = 11 \times 11 = 121$  次元のパターン (各要素の値は  $\pm 1$ ) とする。



(a) 記憶パターン



(b) 入力パターン系列

図 4.15: A, T, R の 3 文字のパターン

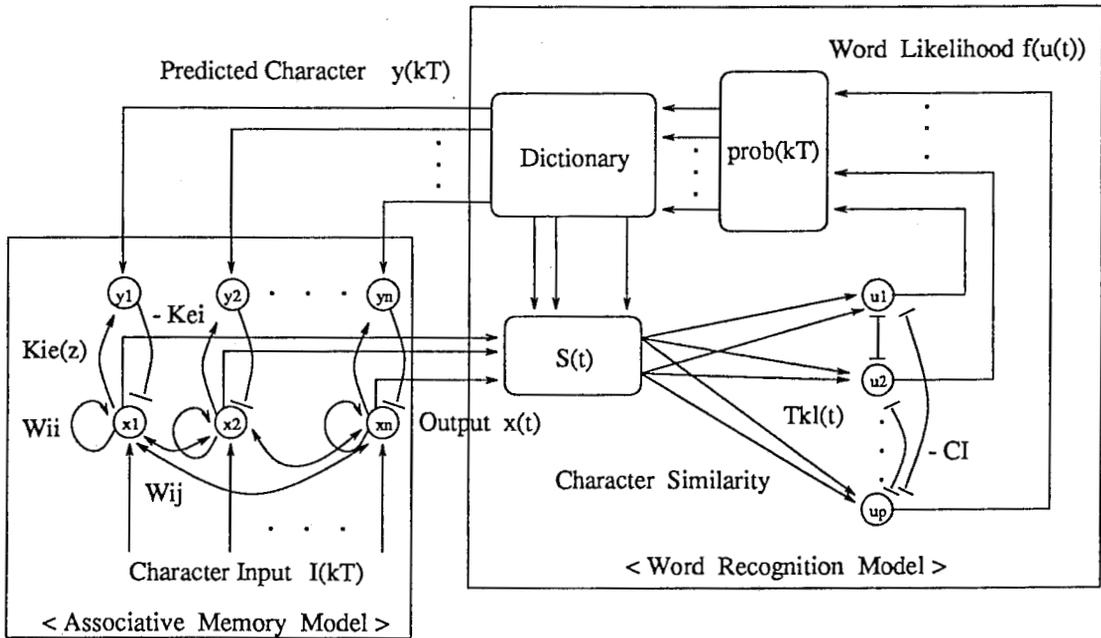


図 4.16: 文字-単語相互作用モデル

さて、 $N$ 次元の入力パターン（アナログ値）から、その入力の確信度に応じた揺らぎを持った記憶パターンを想起する問題を考えよう。この時、時系列の入力文字パターンに対して次々に想起された出力パターンは、単語処理に対する入力信号になるとともに、単語の知識から次の入力文字が予測されるとする。このような処理を行うモデルとして、図 4.16 のような 4.4.1 で提案された連想記憶モデルと単語認識モデルで構成されるものを考え、以下にその処理手順を説明する。もちろん、これは単なる 1 つの例であって、そのような計算手続きが必要であるわけではない。重要なのは、連想記憶モデルで想起された揺らぎ出力に基づいて、単語の知識から次の入力文字が予測され、これによって文字パターンに対する識別領域（対応するアトラクター上の出力パターン）が変化することである。

Step1 サンプリング時間  $t = kT_s$  ( $k = 0, 1, 2, \dots$ ) 毎に入力文字パターンを順に提示し、これを連想記憶モデルの興奮性素子の状態値  $\{x_i(kT_s)\}$  として設定する

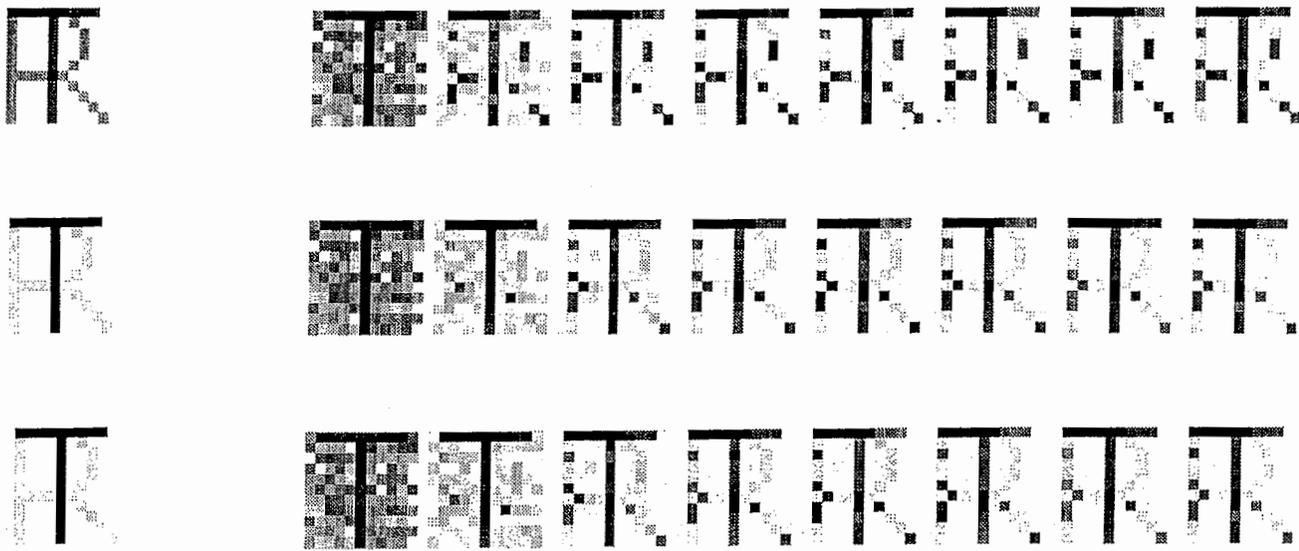
Step2 次の入力提示までの間 ( $kT_s \leq t < (k+1)T_s$ )、連想記憶モデルから揺らぎ出力が生成され、この出力パターンと3文字の記憶パターンとの類似度  $S^\mu(t) \equiv |x(t) \cdot \xi^\mu| / \|\xi^\mu\|^2$  ( $\mu = 1, 2, 3$ ) が計算される

Step3 単語認識モデルの相互抑制結合係数  $-C_I$  を持つネットワークで、類似度  $S^\mu(t)$  に基づいて5単語の確信度  $f(u_j(t))$  が、 $\tau \dot{u}_j(t) = -u_j(t) - C_I \sum_{i \neq j}^L g(u_i(t)) + \sum_{\mu=1}^M T_{j\mu}(t) \cdot S^\mu(t)$ ,  $g(z) \equiv 1/(1 + \exp^{-z/T})$  から競合的に計算される ( $T_{j\mu}(t)$  は、 $j$  番目の単語に対する入力系列中の文字が時刻  $t$  に  $\mu$  番目の文字となる時は1、そうでない時は0となる時間依存の結合重み係数、 $j = 1, 2, \dots, 5$ )

Step4 次の入力文字の提示時刻  $t = (k+1)T_s$  で、5単語の確信度  $f(u_j(t))$  による次文字候補の予想確率  $prob^j \equiv \max\{f(u_l) \mid NextChar(k, l) = j\} / \sum_{\mu=1}^M prob^\mu$  が計算され、 $prob^j$  による文字パターンの重み付け総和で状態  $\{y_i\}$  が、 $y_i(kT_s) = -\sum_{\mu=1}^M prob^\mu \cdot \xi_i^\mu / K_{ei}$  に更新される ( $NextChar(k, l)$  は、 $l$  番目の単語に対する入力系列で  $k+1$  番目のサンプリング時刻に提示されるべき文字の番号を表す)

Step5 もし、1つの単語の確信度  $f(u_j(t))$  の値が1.0付近に収束すれば処理を終了し、そうでなければ、Step 1 に戻って以上の処理を反復的に繰り返す

例として、図 4.15(b) のようにノイズの加わった文字パターン系列: ART を入力した場合を考える (順に 50, 90, 50 % のランダムノイズが記憶パターンに加わったもの)。ここで、単語認識モデルのパラメータは、 $T_s = 150$ ,  $C_I = 0.8$ ,  $\tau = 5.0$ ,  $T = 0.1$  とした。この時、50%のノイズを持つ入力文字 T に対して、(a) 各サンプリング時刻 (上から下の順) で予測された文字パターンと、(b) その間の時間における連想記憶モデルの揺らぎ出力の時系列パターンを図 4.17 に示す。これらの図から、入力の提示回数を増やすほど、予測文字パターンが確実になるとともに、予測文字パターンに従って連想記憶モデルの想起パターンも記憶パターンの1つに近付いていくのが読み取れる (他の文字についても同様)。すなわち、この結果は、その記憶パターンに近いアトラクターのベイソンが拡大したことを示すものと考えられる。但し、このような高次元の状態空間に存在するアトラクター数は不明であり、そのベイソンと識別領域との対応付けの仕方も今後の検討課題となる。



(a) 予測文字パターン

(b) 出力パターン系列 (左から右に時間的变化)

図 4.17: 50% のノイズを持つ文字 T に対する単語からの予測の効果 (上から下の各行は入力提示のサンプリング時刻  $kT_s$ ,  $k = 2, 5, 8$  に対応する)

#### 4.5 おわりに

発振ニューラルネットワークの軌道の多様性と可制御性の特徴を、時空間パターン処理に応用することについての検討を行った。まず、生理学的な自律発振との関連付けの観点から、認知的心理学的な文字パターンの連続変形を軌道上で再現することを考えた。学習後のネットワークでは、上記の軌道 (カテゴリー内の変形の許容範囲を示すもの) のみならず、新奇の文字に対するカオス的軌道 (カテゴリー間を遊走するもの) も生成されることが示された。次に、カオス的軌道の情報生成源としての性質を多様なパターンの生成に利用し、ユーザの概念レベルに応じた対話的な探索領域の制御方法を提案した。また、上記の処理が曖昧な初期イメージから最終的な確定イメージへの概念形成過程を支援することを指摘し、その過程中の 3 段階のレベルに対する具体例を示した。最後に、高次の文脈情報からの外部入力的な予測に従って、低次のパターンの識別領域をベイソンの拡大や縮小に基づいて動的に制御する方法を提案した。この方法を用いた単語と文字との相互作用の例では、予測に従った適応的なパターン認識が出来ることが示された。

これらの例は時空間パターン処理への応用可能性を示すものと考えられる。一方、より実際的な課題に対しては、動的軌道の情報処理機能への対応付けの仕方の工夫、モデルの内的ダイナミクスと課題からの外的要求との適合性、処理効率の評価基準や他の方法との比較などを検討する必要がある。また、まだ完全に解明されていない、ネットワーク規模を表す次元数  $N$  と記憶パターン数  $M$  に関連した挙動の変化や学習方法などの適用の仕方も今後の検討課題となる。

## 第 5 章

### ネットワークの結合形態と同期 / 非同期発振

#### 5.1 はじめに

脳の情報処理機能に関して、視覚系の経路では解剖学的にそれが比較的によくわかっているとは言え、それでも線分や動きの方向性、色の選択性などの分析的な処理が大部分である。一方、脳内に記憶された情報との照合や内的な情報の統合などの処理に関しては、ほとんど明らかでないのが現状である。最近、猫の視覚野で観測された入力刺激に依存した神経細胞集団の同期的な発振現象 (Eckhorn et al., 1988; Gray, & Singer, 1989) が、生理学的な興味のみならず脳における情報統合の基本原理としても注目されている (Eckhorn et al., 1990; Gray et al., 1990; 松本, 1992)。このような特徴統合仮説においては、関連する特徴要素間に動的で素早いリンクが同期的発振によって形成されるとともに、関係のない要素は非同期的発振によって、空間的 (Shimizu, & Yamaguchi, 1989; Sompolinsky, Golomb, & Kleinfeld, 1991) あるいは時間的 (Wang, Buhmann, & Malsburg, 1990; Horn, Sagi, & Usher, 1991; Malsburg, & Buhmann, 1992) に分離される。この際、同期的発振で情報統合を行うには、複数の振動子間の周波数ロッキングなどの分岐現象 (Bergé, Pomeau, & Vidal, 1984; 合原, 1990b) に対する、定性的解析のみならず定量的解析も要求される。なぜならば、そのような現象はネットワークの結合形態だけでなく、その結合重み係数や入力値などのパラメータにも依存すると考えられるからである。

しかし、複数の興奮 - 抑制性ペア素子を結合したモデルにおける一般的な場合についての同期現象の理論解析はほとんど不可能と考えられる。その困難さの理由の 1 つとして、外部入力に従って平衡点の位置や個数のみならず、各ペア素子の固有振動数までが変化し得ることが挙げられる。但し、入力バイアスがない場合の 2 組のペア素子においては、同位相と逆位相の同期発振に関する理論解析が既に行われている (Kawato, & Suzuki, 1980; Nagashino, & Kelso, 1991)。

一方、Khibnik ら (1992) は、2 組のペア素子に対する複数の結合形態を考え、それらの結合形態に対する分岐現象について、本章と類似した数値解析結果を報告している。本章で議論するモデルと比較して、彼らのモデルは 3.2 節で紹介したペア素子より複雑な Willson-Cowan 型振

動子 (Borisyuk, & Kirillov, 1992) で構成されている為に、複雑な分岐ルートとなっているが、Hopf 分岐 (漸近安定平衡点  $\leftrightarrow$  リミットサイクル  $\leftrightarrow$  トーラス) は、どの結合形態に対しても共通に存在している。但し、彼らは同一の入力値を持つペア素子の場合しか検討していない。

一方、入力やそれらの間の関係に依存した同期領域を動的に生成して特徴統合を行う為には、各ペア素子が異なる入力値を持つことが重要である。König ら (1991) は異なる入力値をもつペア素子間の結合形態として、同期的なタイプと非同期的なタイプの 2 つを紹介している。すなわち、各興奮性素子から最近接 (隣) の抑制性素子に結合する時は同期的で、各興奮性素子から次の最近接 (隣の隣) の興奮性素子に結合する時は非同期的な振動となることが報告されている。但し、それは実験例に基づいた定性的な議論である為、上記のような特徴統合 / 分離問題に対するネットワーク設計のガイドラインとするには、より定量的な議論が要求される。

本章では、2 組のペア素子間の結合形態と同期 / 非同期発振の起こりやすさの関係を数値解析によって定量的に明らかにする。その結果は、特徴統合 / 分離問題に対するネットワーク設計のガイドラインとして利用出来るのみならず、前章までで議論してきた発振ニューラルネットワークが非同期的なカオスを起こしやすいネットワーク構造であることを示唆するものでもある。

## 5.2 パラメータ空間内の同期領域の定量解析

### 5.2.1 ネットワークの結合形態

本章では、3.2 節に紹介した興奮 - 抑制性ペア素子で同種のもの (パラメータ値が同じもの) を 2 組考え、それらの間を対称に結合したネットワークについて議論する。この時、図 5.1 に示すような 4 種類の結合形態を考える。これら 4 種類の結合形態に対する動作方程式を以下に示す ( $i, j = 1 \text{ or } 2, j \neq i$ )。

$$\text{Case1: } \dot{x}_i = -x_i + G(W_{ii}x_i - K_{EI}y_i + I_i + W_{ij}x_j), \quad \dot{y}_i = -y_i + G(K_{IE}x_i), \quad (5.1)$$

$$\text{Case2: } \dot{x}_i = -x_i + G(W_{ii}x_i - K_{EI}y_i + I_i - K_{EI2}y_j), \quad \dot{y}_i = -y_i + G(K_{IE}x_i), \quad (5.2)$$

$$\text{Case3: } \dot{x}_i = -x_i + G(W_{ii}x_i - K_{EI}y_i + I_i), \quad \dot{y}_i = -y_i + G(K_{IE}x_i - IW_{ij}y_j), \quad (5.3)$$

$$\text{Case4: } \dot{x}_i = -x_i + G(W_{ii}x_i - K_{EI}y_i + I_i), \quad \dot{y}_i = -y_i + G(K_{IE}x_i - K_{IE2}x_j). \quad (5.4)$$

ここで、各記号の意味は 3.2 節と同様とし、結合重み係数などの値も定理 2 の条件を満足するように、 $W_{ii} = 1.0$ ,  $K_{IE} = K_{EI} = 2.0$ ,  $a = 0.1$  とした。但し、ペア素子間の対称な結合重みパラメータは、興奮性結合重み係数  $W_{ii}$  (Case 1),  $K_{IE2}$  (Case 4) と、抑制性結合重み係数  $-K_{EI2}$  (Case 2),  $-IW_{ij}$  (Case 3) である。また、入力バイアス  $I_i$  は 2 つのペア素子で異なる値 ( $I_1 \neq I_2$ ) を取るパラメータとした。すなわち、2 つの入力バイアスと 1 つの結合重み係数の合計 3 つのパラメータに対する 3 次元空間内で同期 (周波数ロッキング) 領域を調べた。

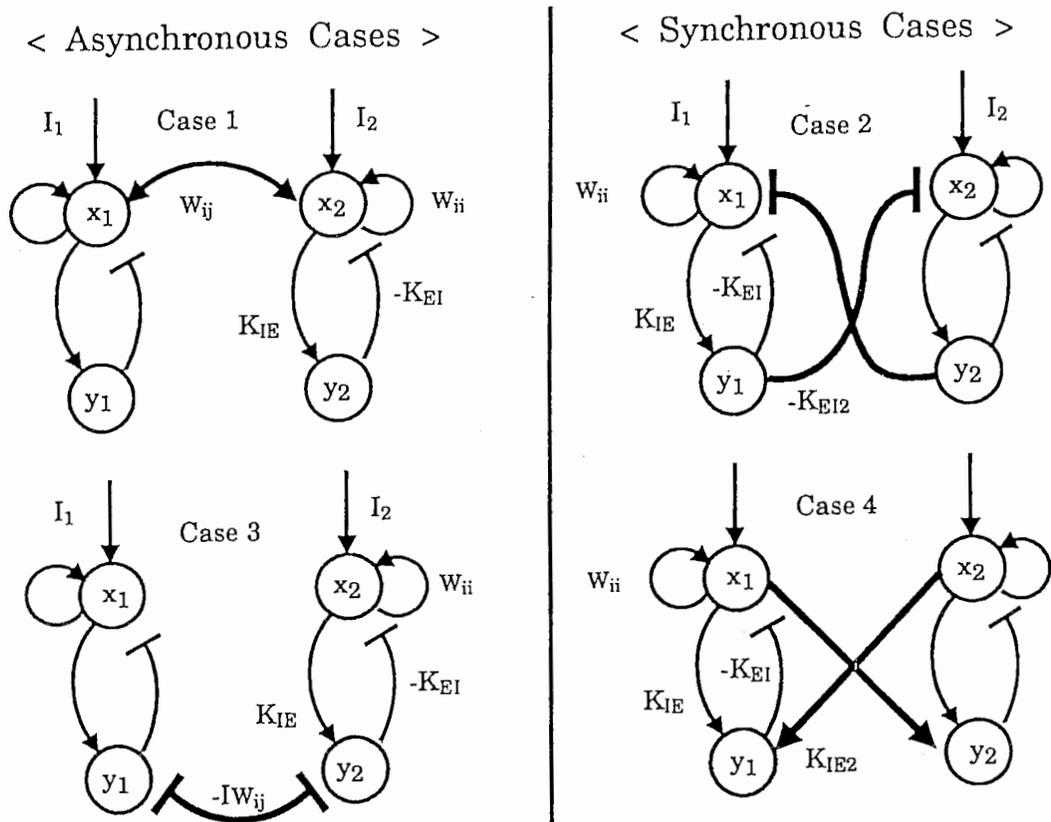


図 5.1: 2組のペア素子間の4種類の結合形態

さて、Königら (1991) のモデルと上記のモデルとの関係について議論しよう。彼らのモデルも複数の興奮 - 抑制性ペア素子を結合したものであるが、以下のように各素子が時間遅れ $\tau$ を持つ点が大きく異なっている (以下では、議論を簡単にする為に単一のペア素子を考える)。

$$\tau_0 \dot{x}(t) = -\alpha x(t) - K_{EI} F(x(t-\tau)) + I, \quad (5.5)$$

$$\tau_0 \dot{y}(t) = -\alpha y(t) + K_{IE} F(x(t-\tau)), \quad (5.6)$$

$$F(x) = 1/(1 + \exp^{\sigma(\theta-x)}). \quad (5.7)$$

ここで、 $\tau_0, \alpha$  は減衰係数、 $\theta$  はしきい値を表す。また、 $K_{EI} = K_{IE} \equiv K$  とする。

時間遅れ $\tau$ が小さい時は、 $x(t-\tau) \approx x(t) - \tau \dot{x}(t)$  を用いて、式 (5.5) ~ (5.7) の平衡点における線形化方程式は以下のように近似出来る。

$$(\tau_0^2 + \tau^2 K^2 F_e' F_i') \begin{bmatrix} \delta \dot{x} \\ \delta \dot{y} \end{bmatrix} = \begin{bmatrix} -\alpha \tau_0 + K^2 \tau F_e' F_i' & -(\tau_0 + \alpha \tau) K F_i' \\ -(\tau_0 + \alpha \tau) K F_e' & -\alpha \tau_0 + K^2 \tau F_e' F_i' \end{bmatrix} \begin{bmatrix} \delta x \\ \delta y \end{bmatrix}. \quad (5.8)$$

ここで、 $F_e', F_i'$  は各平衡点  $\bar{x}, \bar{y}$  における、それぞれの導関数  $F'(x)$  の値を表す。

上式右辺の Jacobi 行列の固有値は、

$$\lambda = -\alpha\tau_0 + K^2\tau F'_e F'_i \pm i(\tau_0 + \alpha\tau)K\sqrt{F'_e F'_i} \quad (5.9)$$

となる。固有値  $\lambda$  の実部が正の時、すなわち、 $K^2\tau F'_e F'_i > \alpha\tau_0$  ならば、不安定平衡点の周りにリミットサイクルが出来る。これより、時間遅れ  $\tau$  は、各素子に対する正の自己結合と等価的な働きををすると言える。

以上より、本章で議論するモデルのペア素子は、時間遅れ  $\tau$  が小さい時の König らのモデルのものと等価的と考えられる。しかしながら、時間遅れ  $\tau$  が大きい時は、両者のモデルが異なる挙動を示すと思われるが、これに関する議論は本論文の範囲を越えるものとなる。

### 5.2.2 数値解析結果

本節では、前記のネットワークの結合形態に依存した同期化の傾向として、(Cases 1, & 3) は広いパラメータ領域で非同期発振を起こしやすく、(Cases 2, & 4) は同期発振を起こしやすいことを数値解析によって示す。さらに、4つの結合形態に対する分岐構造の類似点や相違点についても検討を行う。

まず、2つの入力バイアスが  $I_1 > 0, I_1 > I_2$  であると仮定する。2組のペア素子を入れ換えれば、 $I_1 < I_2$  の場合と  $I_1 > I_2$  の場合の結果は明らかに等価となる。また、 $G(z)$  が奇関数であることを利用して変数変換  $\tilde{x}_i = -x_i, \tilde{y}_i = -y_i, \tilde{I}_i = -I_i$  を行えば、 $I_1 < 0$  と  $I_1 > 0$  の場合は等価となる。以下の実験では、各入力値を -1.0 から 1.0 まで 0.2 刻みで、また、結合重み係数を 0.01 から 2.6 まで 0.001 刻みで変化させた。これらの各組合せに対して、初期状態  $x_i(0) \equiv I_i, y_i(0) \equiv 0$  からの 4 次の Runge-Kutta 法 ( $\Delta t = 0.01$ ) を用いた反復計算中、最初の 30,000 回分を過渡期とした後の 30,000 回分の出力軌道を調べた。一般に、状態空間内には初期状態に依存した複数のアトラクター上の軌道が存在するが、以下では上記の初期状態に設定した場合だけを議論する。なぜなら、初期状態に依存した複数のアトラクターを考えることは、理論的観点からは興味深い、特徴統合 / 分離課題の応用上の観点からは、そのような分岐構造は複雑すぎて使えないものと考えられるからである。

各パラメータ値の組合せに対して反復計算を行った後、出力波形  $x_i(t)$  の周波数解析と、2次元の  $x_i - y_i$  平面上に射影された軌道の Poincaré 断面 ( $y_i$  軸との交点) から、周波数ロッキングした軌道に対するパラメータ空間内の同期領域を求めた。その結果、以下に説明するような分岐ダイアグラムが得られた。但し、周波数ロッキングしていない非同期的軌道に対して、カオスと準周期軌道 (周期軌道に近いが元の点に戻らず、トーラス曲面上を埋め尽くして閉じない軌道) との分類は行わなかった。

図 5.2 は、3つのパラメータの3次元空間における分岐ダイアグラムを示す。図中の屋根のようなものは、結合重みパラメータ値の増加に伴った平衡点への漸近安定収束の開始点（1次 Hopf 分岐点）を表す面で、その下の曲面は、準周期軌道などからリミットサイクルとなる周波数ロッキングの開始点を表す面である。図 5.3 は、 $I_1 = \text{const}$  における上記のダイアグラムの2次元断面を示す。図中の記号  $(\alpha, \beta)$  は、2平面  $x_1 - y_1$  と  $x_2 - y_2$  の上に射影された周期軌道（トーラス）のそれぞれの回転数を表す。これらの分岐図における（平衡点への漸近安定収束や周波数ロッキングを起こす）分岐点や同期発振の存在範囲から、Case 1 と Case 3 に対する分岐構造は定性的に同じであると言える（Case 2 と Case 4 についても同様）。定量的に異なる点は、高い回転数を持つトーラス上のリミットサイクルに対する HLC（Harmonic Limit Cycle）領域の広さや、1次 Hopf 分岐点を表す屋根の傾きなどである。

これに対して、Case 1 と Case 2 とでは、屋根の下の分岐構造が異なっている（Case 3 と Case 4 についても同様）。Case 1（or 3）では、屋根と曲面の間が SLC（1周して元に戻る Simple Limit Cycle）領域だけとなり、曲面の下には QPO（非同期的な準周期軌道やカオスなどの Quasi-Periodic Orbit）領域中にいくつかの HLC 領域が存在している。また、2つの入力値が非常に異なる時 ( $|I_1| \approx 1, |I_2| \approx 0$ )、もしくは、2つの入力値が小さい時 ( $|I_1| \approx |I_2| \approx 0$ ) には、曲面の下の領域はほとんど QPO 領域に支配されている。

一方、Case 2（or 4）では、屋根と曲面の間にさまざまな領域が存在するが、曲面の下は QPO 領域だけとなっている。また、同期的な軌道としては、(1, 2), (2, 2) の回転数を持つトーラス上の軌道と1周して戻るリミットサイクルが HLC 領域で共存している。さらに、入力値  $I_1$  が小さくなる程、QPO 領域と CM（図 5.4(b) のように複雑な軌道やリミットサイクルなどが混在する Complex Mix-mode）領域は狭くなっていることから、この場合は周波数ロッキングした同期発振が起こりやすいと言える。

これらより、ペア素子間を弱結合した場合（結合重みパラメータの値が0から1までの間）の、[Cases 1, & 3] と [Cases 2, & 4] との違いは、以下のようにまとめられる。

[ Cases 1, & 3 ] 広い QPO 領域中に非常に狭い HLC 領域が数少なく孤立して存在することから、2つの入力値が異なる程、より非同期発振となりやすい

[ Cases 2, & 4 ] SLC 領域と HLC 領域には含まれた QPO 領域は比較的狭いことから、同期発振となりやすい

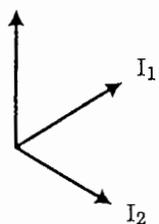
但し、上記の場合はペア素子間の相互作用を微小な摂動で近似出来ない為に、このような違いを理論的に解析することは出来なかった。この違いに関する検討は今後の課題となる。

ここで、図 5.3 における省略部分を述べなければならない。まず、Case 1 と Case 3 に対する左右の羽根状の QPO 領域には、図が対称性を持つことから HLC 領域を描かなかった。また、図に示していない Case 3 で入力  $I_1$  の値が 0.8 以下の時、回転数 (3, 4) を持つトーラス上のリミットサイクルが縮退してなくなったことも付記する。

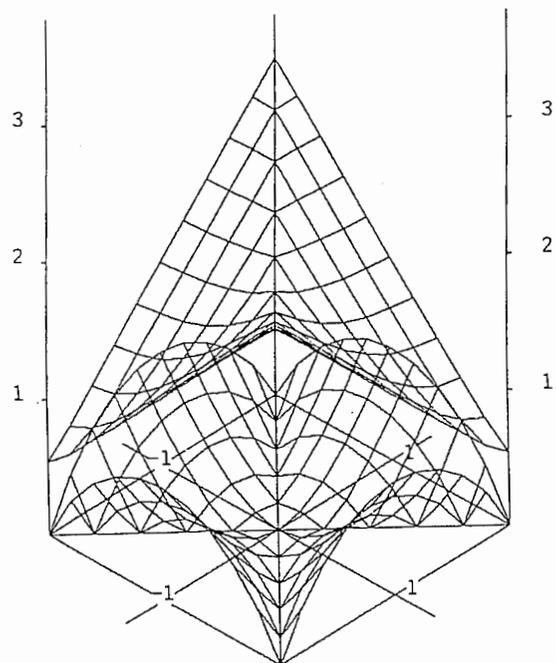
上記の結果以外にも、次の 2 つのことが判明した。1 つは 1 次 Hopf 分岐点に関することで、もう 1 つは分岐ルートに関することである。図 5.3 から、Case 1:  $W_{hopf}$  や Case 2:  $K_{hopf}$  では 1 次 Hopf 分岐点が近似的に直線上にのっているが、Case 3:  $IW_{hopf}$  や Case 4:  $K_{hopf}$  ではそれがほとんど一定値となっていることがわかる。このことは近似的な解析からも得られるが、その詳細については付録 F を参照されたい。

結合重みパラメータの増加に対する分岐ルートとしては、以下のようなものが各結合形態において共通に存在した (図 5.4(a) 参照)。その典型的なルートは、ほとんど独立な 2 つの軌道 → トーラス上のリミットサイクル (Poincaré 断面上のサドル点の生成 - 消滅分岐による周波数ロッキングやアンロッキング) ↔ 準周期軌道 (2 次 Hopf 分岐) やカオス ( $T^2$  トーラス上の軌道の崩壊) → 1 周で戻るリミットサイクル (1 次 Hopf 分岐) → 漸近安定平衡点、であった。但し、上記の分岐ルート以外にも、Case 2 と Case 4 においては図 5.4(b) に示すような複雑な分岐ルートも存在した。すなわち、この分岐ルートに対応した図 5.3 の CM 領域は、2 つのペア素子が同期しながら 1 周で戻るリミットサイクル、片方のペア素子のみの発振によってもう一方のペア素子を摂動した座標軸に平行な直線的振動、さらに、それら以下の複雑な軌道に対する領域を持つ。しかし、その複雑な軌道を理解するには、さらに詳細な検討が必要となる。

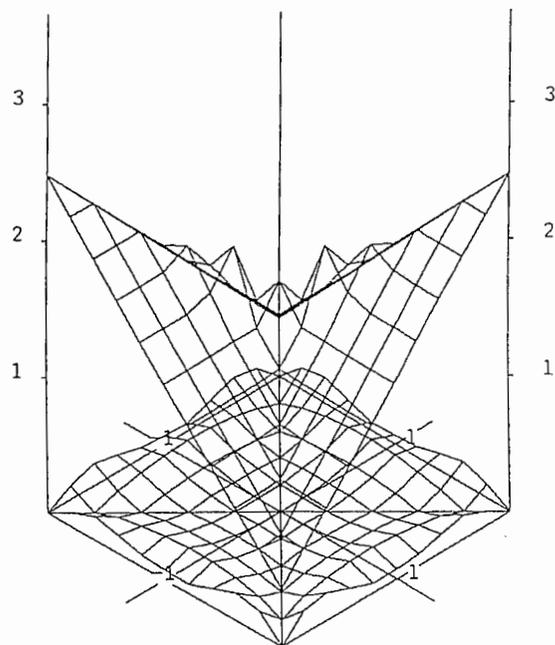
Connection Parameter



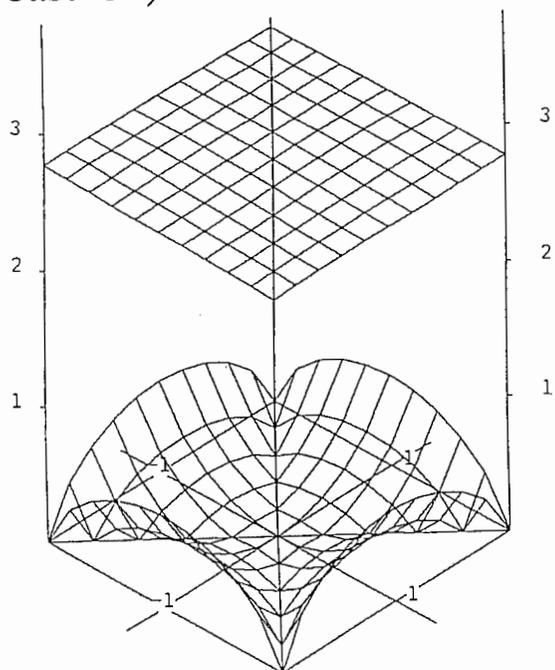
( Case 1 )



( Case 2 )



( Case 3 )



(Case 4 )

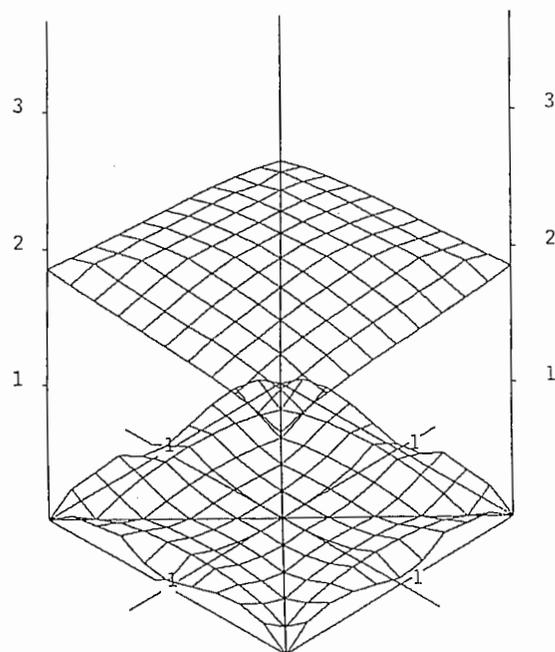
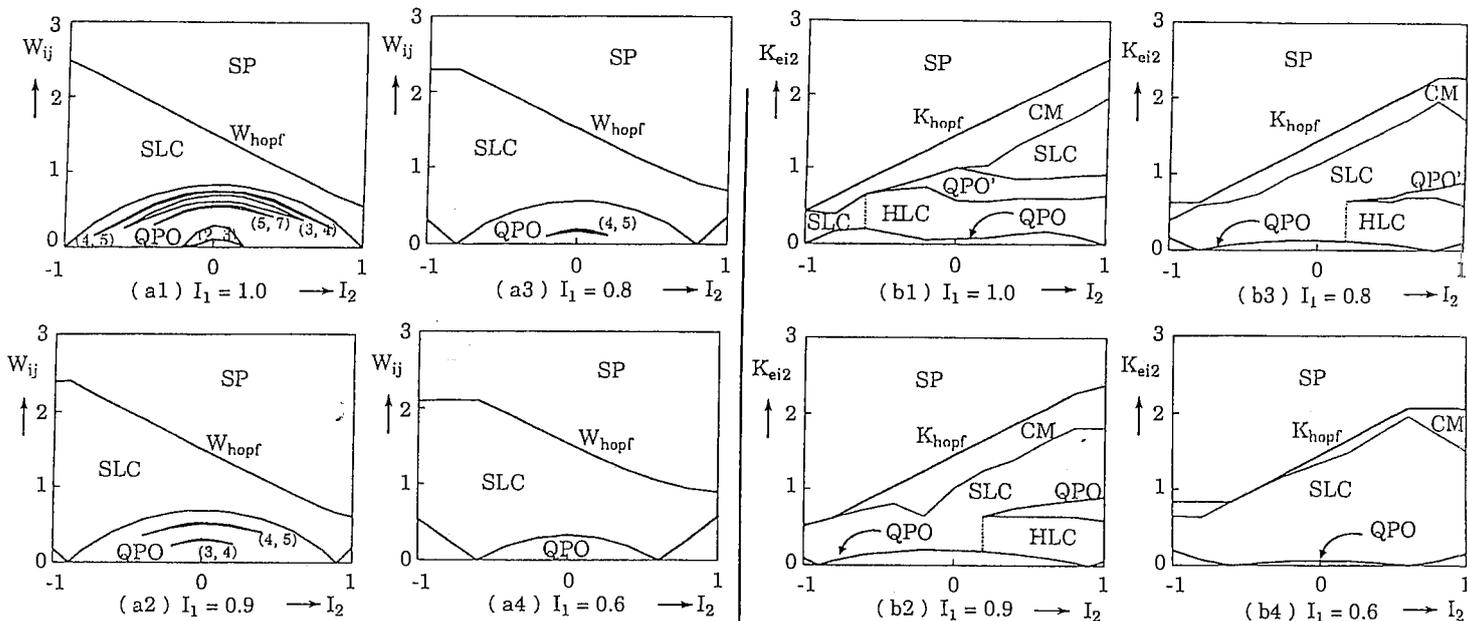


図 5.2: 4 種類の結合形態に対する 3 次元の分岐ダイアグラム (屋根: 漸近安定収束の開始点、  
曲面: 周波数ロッキングの開始点)

( Case 1 )

( Case 2 )



( Case 3 )

( Case 4 )

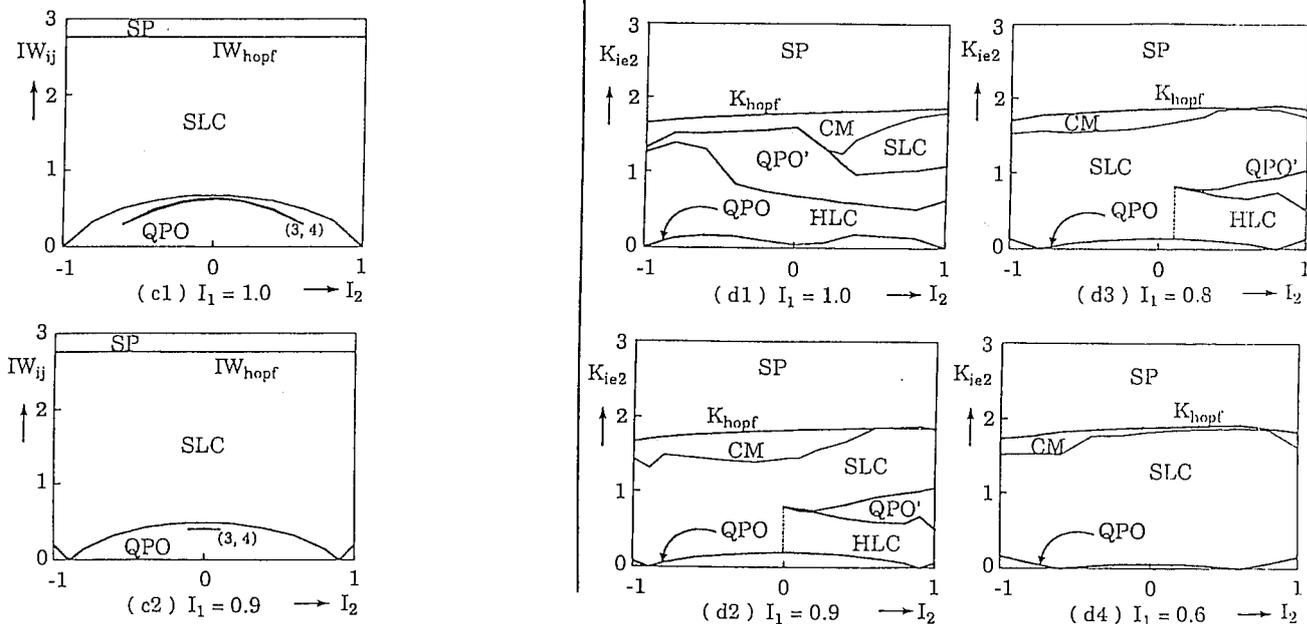


図 5.3: 各ダイアグラムの  $I_1 = const$  における 2次元断面 (SP: Stable Point, SLC: Simple Limit Cycle, HLC: Harmonic Limit Cycle, QPO: Quasi-Periodic Orbit from two independent cycles, QPO': QPO between SLC & HLC, CM: Complex Mix-mode)

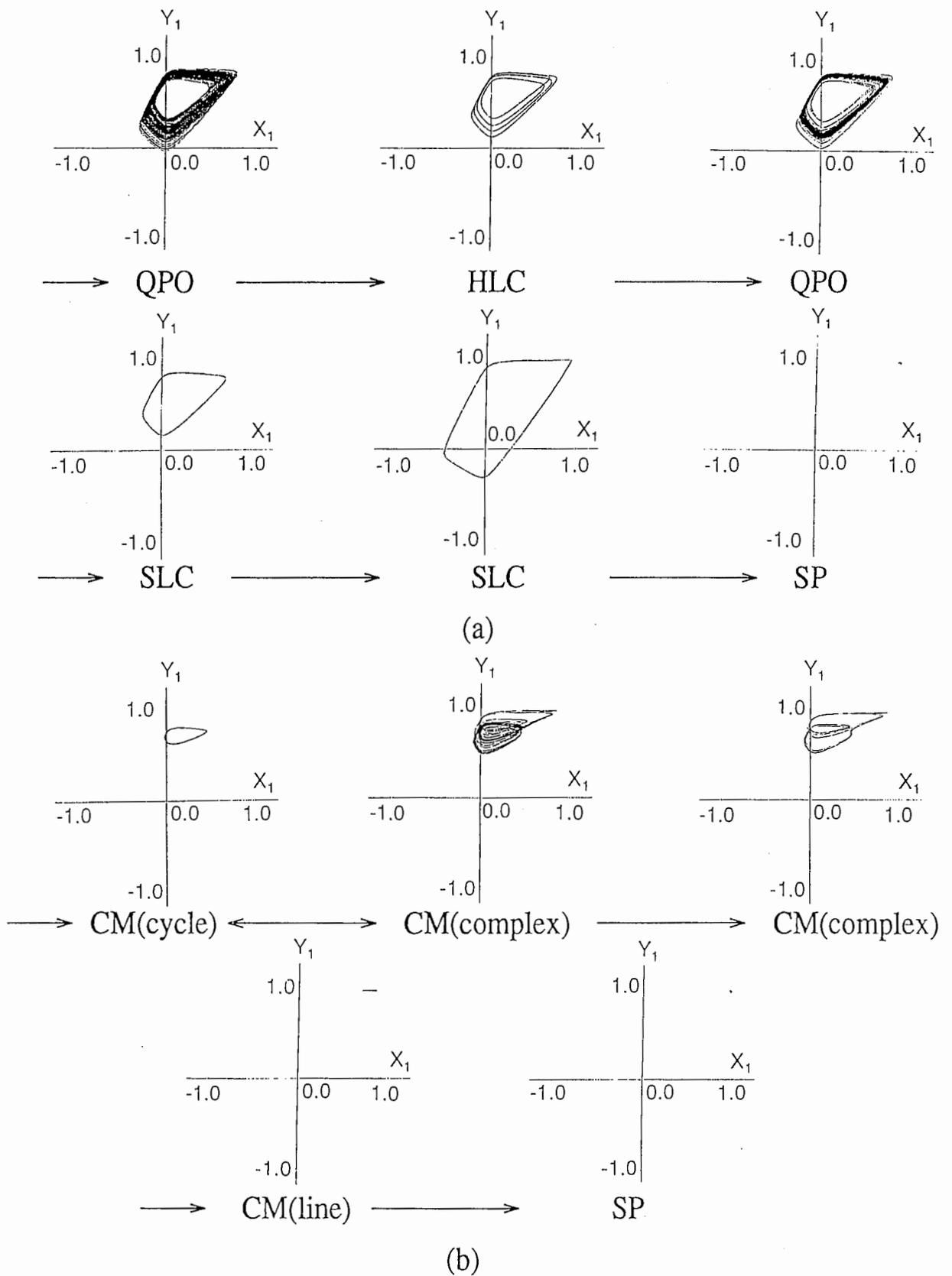


図 5.4:  $x_1 - y_1$  平面上に射影された軌道の変化 (a) 典型的な分岐ルート: QPO - locking  $\rightarrow$  HLC - unlocking  $\rightarrow$  QPO - 2ndHopf  $\rightarrow$  SLC - 1stHopf  $\rightarrow$  SP, (b) 複雑な分岐ルート: CM(simple limit cycle)  $\leftrightarrow$  CM(complex orbit)  $\rightarrow$  CM(limit cycle on line) - 1stHopf  $\rightarrow$  SP

### 5.3 分岐ダイアグラムの利用方法

同期 / 非同期発振による特徴統合 / 分離問題に対してネットワークを設計する時、前節の分岐ダイアグラムは有効なガイドラインとなる。このことを文字パターンからの方向線分特徴（水平、垂直、斜め $\pm 45^\circ$ の4方向）の抽出における例を用いて説明する。

興奮 - 抑制性ペア素子のような自律振動子を結合したネットワークにおける同期 / 非同期発振で、文字パターンから方向線分特徴を抽出したモデルは既に提案されている（Shimizu, & Yamaguchi, 1989）。そのモデルでは、低次のパターンと高次の記憶や概念情報との階層的な相互作用を通じた大域的な引き込みに従ってパターン認識が行われるが、入力や結合重み係数などのパラメータ値とモデルの同期化傾向との関係は明らかでない。そこで、以下の様なより単純なモデルを考えてこれに関して検討を行った。

まず、入力文字パターン（ $16 \times 16$ 画素、各画素は文字部分ほど高い値の $[0, 1]$ のアナログ値）を表す入力層を考え、その入力値を4つの方向に対応したそれぞれの特徴抽出層の各興奮性素子に直接与える。この時、入力の各画素はそれぞれの特徴抽出層の各ペア素子と1対1に対応付けられる。そのネットワーク構造は、単純化の理由で4つの特徴抽出層間の結合をなくしたのに対して、各特徴抽出層内では各興奮性素子が抽出すべき線分方向にある両隣の興奮性素子とだけ結合するようにした。このような結合は、前節のCase 1に相当する結合形態であるので、結合方向で類似した入力値を持つペア素子どうしは同期化し、異なった入力値を持つものどうしは非同期化することで、方向線分特徴が抽出できると考えられる。そこで、前節の分岐ダイアグラムを参考に、同期化させたい2つの入力値の差の範囲を想定して、それに適する結合重み係数の値を求めてモデルに設定した。

例として、結合重み係数 $W_{ij} = 0.2, 0.35, 0.5$ の3つの場合における同期化の強さを、各興奮性素子出力とその両隣の出力との相互相関関数の値で表示したものを図5.5に示す（図の左上は入力文字パターン）。図から、結合重み係数の値を大きくする程、相関関数の値が大きくなって、同期化領域が直接結合された両隣りを越えて広がったことが伺える。もちろん、画素間の同期化は入力値に大きく依存している。図には示していないが、他の結合形態についても前節の結果と同様な同期化傾向が確認出来た。これらの結果は、複数のペア素子を局所的に弱結合した場合は、2組のペア素子を結合した前節の結果と類似した同期化傾向を持つことを示している。もちろん、各ペア素子は非線形の自律振動子であるので、局所結合でない場合はそれほど単純に加法的とはならないと思われる。この問題に関しては、今後の検討課題となる。

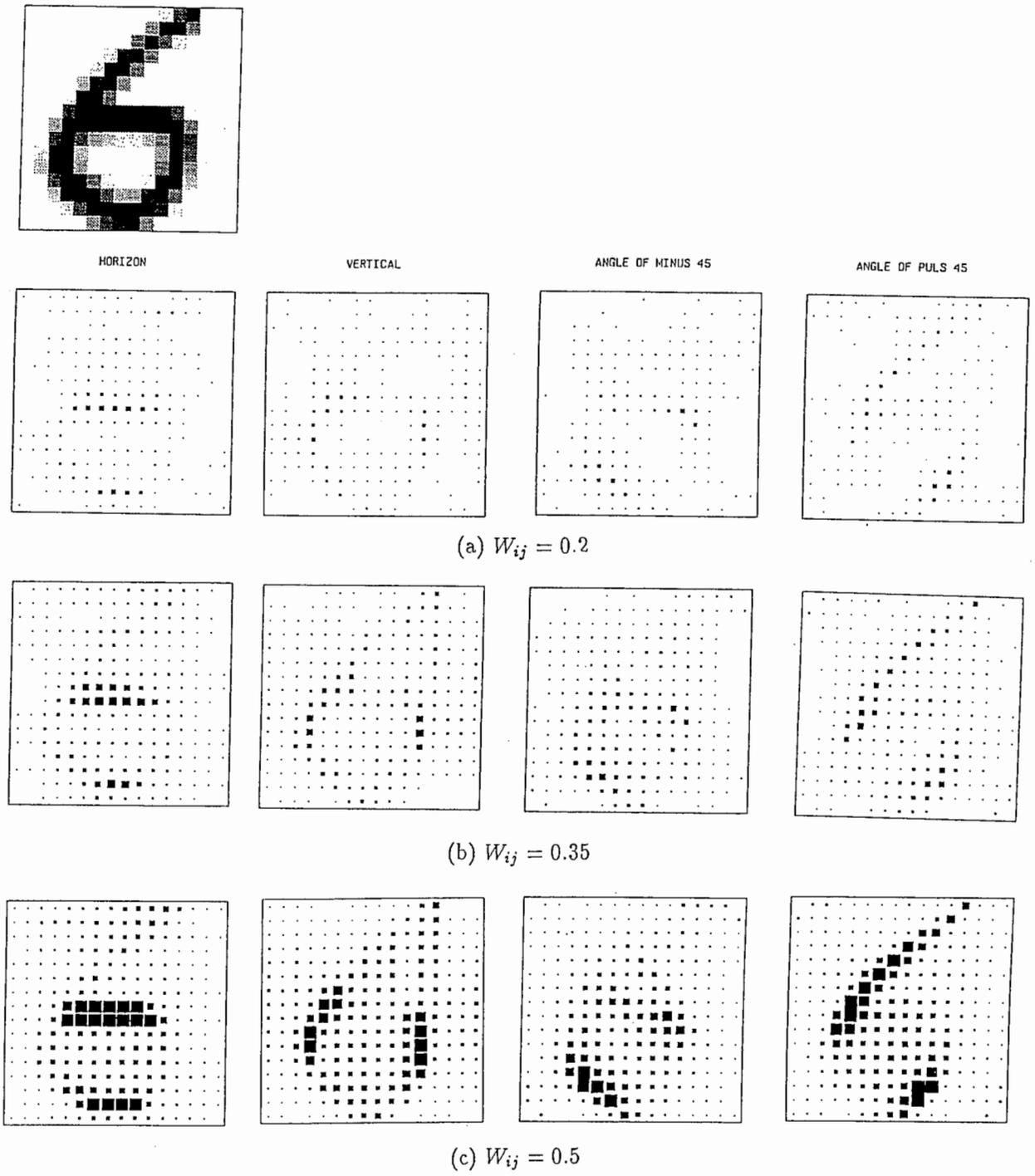


図 5.5: 結合重みパラメータに対する同期化領域の変化 (各画素の興奮性素子出力とその両隣の出力との相互相関関数の値を四角の大きさで表示)

## 5.4 おわりに

興奮-抑制性ペア素子を結合したネットワークの結合形態と同期発振の起こりやすさの関係について定量的な検討を行った。対称に結合された2組のペア素子についての数値解析結果から、興奮-興奮性結合や抑制-抑制性結合の場合は広いパラメータ領域で2つの出力が非同期化しやすく、興奮-抑制性結合や抑制-興奮性結合の場合はそれらが同期化しやすいことが判明した。これらの結果は、特徴統合 / 分離問題に対してネットワークを設計する際のガイドラインとなるのみならず、前章までで議論してきた発振ニューラルネットワークが非同期的なカオスを起こしやすいネットワーク構造であることを示唆するものでもある。

しかしながら、より詳細にその分岐構造を解明するには、カオスの存在性や分類不可能だった複雑な軌道についての検討などが要求される。さらに、応用上有益な複数のペア素子を結合した場合の同期化傾向についても、さらに検討を行う必要がある。

## 第 6 章

### 結論

#### 6.1 研究成果の要約

本研究は、時空間パターン処理への応用を念頭に、人工ニューラルネットワークにおけるネットワーク構造と情報処理機能の関係、及び、外部入力などに従った情報処理機能のスイッチングに関して、著者がこれまで行ってきたことをまとめたものである。以下にその成果を要約する。

第 2 章では、情報表現やネットワーク構造、結合重み係数の設定方法や学習方法などに関する既存の能力向上法の関連性に着目するとともに、これらの高い連想能力を継承しつつ、手書き文字などのパターン認識への応用の観点から新しいモデルを提案した。また、本提案モデルの認識処理と学習処理の妥当性を理論的に示し、手書き文字に対する実験から認識能力の評価も行った。さらに、本提案モデルの学習と類似する既存のモデルの学習との実験的比較を行い、認識辞書に相当する結合重み係数ベクトルがそれぞれ、各カテゴリ毎のプロトタイプパターンによる局所的な辞書とカテゴリ間のプロトタイプが混じった分散的な辞書という相違点を明らかにした。しかしながら、上記の異なったネットワーク構造を持つ複数のモデルで実現された機能は、本質的には主成分分析に基づくパターン分離機能である為、このような静的な空間分割では、誤認を柔軟に回避することが出来ないという問題点を指摘した。それは、学習などで識別境界の修正を行ったとしても新たな誤認領域が発生することが多いばかりか、こうした修正を膨大な統計データに基づく学習だけに頼っていたのでは、迅速な処理が望めないということであった。

第 3 章では、上記の問題点を解決する為の 1 つの手段として、動的な出力軌道を情報処理機能に対応付けることで、軌道の多様性を時空間パターン処理に積極的に利用し、外部入力などに従った軌道の性質の変化を情報処理機能のスイッチングに利用することを提案した。この考えを実現する為の具体的なモデルの 1 つとして、互いに反対符号の重み係数で結合された興奮-抑制性ペア素子で構成される発振ニューラルネットワークを提案し、その軌道の多様性と可制御性に関する特徴を主に数値解析を通じて示した。より具体的には、軌道の多様性として、リミットサイクルやカオスの軌道などのさまざまな複雑さを持つ軌道の存在と、それらの軌道が記憶パター

ンなどの  $N$  次元超立方体の頂点間を遊走することなどが、また、軌道の可制御性としては、入力バイアスに従った軌道の遊走領域の変化や、結合重み係数に従った軌道の複雑さの変化などが示された。さらに、記憶パターン付近の限られた挙動についての理論解析から、記憶パターン付近の漸近安定平衡点の存在や、モデルの1つの発振条件を明らかにした。

第4章では、上記の発振ニューラルネットワークの軌道の多様性と可制御性の特徴を積極的に利用した以下の3つの例題から、時空間パターン処理への応用可能性についての検討を行った。第1に、生理学的な自律発振と認知心理学的なパターンの連続変形操作との関連付けを試みた。そこで、認知心理学的な変形として仮定された文字カテゴリ内の変形の許容範囲を示すような連続変形パターンのサイクルを、モデルの軌道上に効率的に学習する方法を検討した。学習後のネットワークでは、上記の軌道を再現するパターン生成機能が実現されたのみならず、新奇の入力文字に対するカテゴリ間の探索機能（記憶パターン間を遊走するカオス的軌道）への自律的なスイッチングも示された。第2に、カオス的軌道の情報生成源としての性質を多様なパターンの生成に利用し、ユーザの概念レベルに応じた対話的な探索領域の制御方法を提案した。また、上記の処理が曖昧な初期イメージから最終的な確定イメージへの概念形成過程を支援することを指摘し、その過程での3段階のレベルに対する具体例を示した。さらに、分岐パラメータである結合重み係数に従った軌道の複雑さの変化が、パターンの探索機能と確定機能のスイッチングに利用できることも示した。第3に、高次の文脈情報からの外部输入的な予測に従って、低次のパターンの識別領域をベイソンの拡大や縮小に基づいて動的に制御する方法を提案した。しかし、上記のモデルにおいては入力バイアスに依存した膨大な数のアトラクターが存在する為に、ベイソンと識別領域の対応付けが困難であった。そこで、改良モデルを考えて状態空間内に数少ないアトラクターしか存在しないことを3次元の場合で検証した。さらに、この方法による単語と文字との相互作用の例から、予測に従った適応的なパターン認識が可能であることを示した。

第5章では、2組の興奮-抑制性ペア素子に関する結合形態と同期発振の起こりやすさの関係についての定量的な解析を行った。その数値解析結果から、興奮-興奮性結合や抑制-抑制性結合の場合は広いパラメータ領域で2つの出力が非同期化しやすく、興奮-抑制性結合や抑制-興奮性結合の場合はそれらが同期化しやすいことが判明した。同期/非同期発振の情報処理機能は、同期発振による動的な特徴要素の統合と非同期発振による特徴要素の分離を、入力に依存して選択的に行うことに対応付けられる。このような特徴統合/分離問題に対するネットワーク設計のガイドラインとして、上記の解析結果が利用出来ることが具体的に示された。さらに、その結果から、これまで議論してきた発振ニューラルネットワークが非同期的なカオスを起こしやすいネットワーク構造であることが示唆された。

## 6.2 今後の研究課題

人工ニューラルネットワークのネットワーク構造と情報処理機能の関係に着目した時空間パターン処理の研究は、モデルのパラメータに依存した複雑な挙動や、その動的な出力軌道の情報処理機能への対応付けの仕方に関連した、深い内容と広がりを持つ分野である。これに対して、本研究の成果はその緒についたに過ぎず、特に、モデルの内的なダイナミクスの特徴と応用課題からの外的な要求との適合性に関しては、さらに検討を行う必要がある。そこで、本研究が残した検討課題を以下の2つに分けて考え、それぞれに対する代表的な項目を挙げる。

(1) モデルの動的な挙動の解明

(2) 時空間パターン処理への応用

(1)に関する検討課題は、モデルの入力や結合重み係数などのパラメータ、及び、そのネットワークの結合形態に依存した、状態空間内のアトラクターの種類、ベイスンによる状態空間の分割構造、前記のようなパラメータ値の変化によるベイスン融合などに従った状態空間構造の変化(分岐構造)などをより詳細に解明することである。但し、これらに関する理論的解析はもとより、実験的解析も困難を伴うことが予想される為、応用上重要な挙動に関して優先的に検討すべきであると考えられる。具体的には、以下のような項目が挙げられる。

まず、3章で議論した発振ニューラルネットワークについての、軌道の複雑さを規定するペア素子の結合重み係数  $K_{IE}$  に対する分岐構造や、ネットワークの次元数(素子数)  $N$  と記憶パターン数  $M$  との比に関連した、 $N$ 次元超立方体の頂点間を遊走する軌道の性質の変化に関する検討が必要である。特に、記憶パターンやそれ以外の頂点に対する遊走軌道の距離や滞在頻度は、モデルの内的なダイナミクスの特徴を表すものであり、探索課題からの外的要求との適合性にとっても重要と考えられる。一方、軌道の探索領域とは別に複雑さの観点からは、このような遊走軌道が厳密な意味でのカオスかどうかを検討する必要がある。さらに、5章の解析結果から示唆されたことで、上記のモデルが非同期発振を起こしやすいこと(入力  $I_i$  やペア素子間の結合重み係数  $W_{ij}$  に対するパラメータ空間内の同期化領域が狭いこと)も、検討課題として残っている。もちろん、3組以上のペア素子に対する5章で紹介した上記以外の結合形態についても、それぞれの同期化傾向を明らかにすべきであろう。これらとは別の、4.4節で議論した連想記憶モデルについては、その3次元の実験例から類推される以下のような項目を検討する必要がある。それは、より高い次元数  $N > 3$  でも、状態空間内に数少ないアトラクターしか存在しないこと、及び、状態空間が各識別カテゴリーに対応付け可能な頂点近くの限定された範囲内のアトラクターと不確定状態を表す原点周辺のカオス的な遊走軌道とで分割されることである。

(2)に関する検討課題には、与えられた応用課題に対するモデルのダイナミクスの特徴の適合性と、モデルのダイナミクスの特徴に適した応用課題の発掘の2つがあるが、以下では本研究で紹介した応用例に対する前者についての項目を挙げる。上記の2つのいずれにおいても、モデルの動的な出力軌道と応用課題から要求される情報処理機能との対応付けの仕方が鍵となる。

まず、連続変形パターンの効率的な学習について、4.2節で議論した発振ニューラルネットワークのみならず、より一般的なリカレントネットワークに対しても検討を行う必要がある。一方、生理学的発振と認知心理学的変形操作との関連性を確かにするには、上記の問題以前に、脳内に存在すると思われる情報の内部表現とそれに対する変形操作がどんなものであるかを明らかにすべきであろう。4.3節のパターン生成については、その対話的な処理が曖昧な初期イメージから最終的な確定イメージへの概念形成過程を伴う為に、探索効率の評価基準が見当たらないこと、また、ランダム生成等の他の方法との定量的な比較などが検討項目として考えられる。また、4.4節で議論した高次の文脈情報からの予測に従った低次のパターンの識別領域の動的制御については、状態空間内のベイスンと各カテゴリーの識別領域との対応付けや、高次の文脈情報と低次のパターンとの相互作用における計算手続きなどについてさらに検討する必要がある。最後に、5章で紹介した同期 / 非同期発振による特徴統合 / 分離問題については、より実際的な応用例に対する複数のペア素子のネットワークの設計方法が今後の検討課題となる。もちろん、本研究で提案されたモデルによる方法が、上記のそれぞれの応用課題に対して有効であることを、より実際的な多くの事例で確認する必要があることは言うまでもない。

### 6.3 むすび

本研究は、非線形システムである人工ニューラルネットワークの複雑な挙動を、情報処理に積極的に利用することに関して1つの試みを行ったものである。この際、時間変化を伴う動的な軌道を情報処理機能に対応付け、かつ、外部入力などによって情報処理機能のスイッチングを行うことが重要であることを指摘した。そこで、これを実現する具体的モデルとして提案された発振ニューラルネットワークの挙動を部分的に解明し、その軌道の多様性と可制御性の特徴を利用した具体例によって、時空間パターン処理への応用可能性を示した。但し、上記のモデルの複雑な挙動の詳細や、より実際的な応用課題に対しては、まだ多くの課題が残されているが、このような方法論が有益となることを期待したい。

## 謝辞

本論文をまとめるにあたり京都大学工学部数理工学教室の茨木俊秀教授には懇切なる御指導と、御教示を賜りました。ここに、慎んで深謝の意を表します。

また、Neural Networks に投稿された発振ニューラルネットワークの理論解析について有益な御示唆を頂いた、東京大学工学部計数工学科の甘利俊一教授（当時 某雑誌の編集委員長）、ならびに、第5章で議論した2組のペア素子間の結合形態と同期発振の起こりやすさとの関係に類似した数値解析結果を示す Khibnik らの論文を紹介して頂いた、The Salk Institute の銅谷賢治博士に感謝致します。

本研究の大半は、著者が（株）ATR 視聴覚機構研究所および（株）ATR 人間情報通信研究所に在籍中に行われたものであり、研究の機会を与えて頂いた（株）ATR 視聴覚機構研究所 淀川英司社長（現 工学院大学教授）、同視覚研究室 上野圭一室長（現 NTT ヒューマンインターフェース研究所メディア応用システム研究部応用画像グループリーダー）、および（株）ATR 人間情報通研究所 東倉洋一社長、同第6研究室 下原勝憲室長に感謝致します。

本研究の途上において貴重な御討論および御助言を頂いた、同第6研究室の佐藤雅昭主任研究員をはじめ、同第2研究室の尾田政臣主幹研究員、同第3研究室の川人光男室長、同第1研究室の Norman D. Cook 客員研究員、および、同研究所の諸氏に心から感謝致します。

特に、研究所が異なるにもかかわらず、（株）ATR 光電波通信研究所の Peter Davis 客員研究員には、非線形力学系やカオスに関する貴重な御助言および御示唆を頂いたのみならず、著者の難解な英論文の更正にまで時間を裂いて頂きました。これらの彼の協力なしでは、本論文をまとめ上げることが出来なかったと言えます。ここに深く感謝致します。

なお、著者に出向の機会を与えて頂いた、（株）富士ゼロックス システム技術研究所の加藤守通所長（現 同社、技術開発センター長）をはじめ、著者の出向後の復帰先である同社システムコミュニケーション研究所の丹洋一所長、篠岡信主幹研究員、上林憲行主幹研究員、大住淳一副主任研究員、および（株）富士ゼロックスの皆様には多大な御支援、御尽力を頂きました。諸氏に心から御礼申し上げます。

## 参考文献

- Abbott, L. F. (1990). A network of oscillators. *Journal of Physics, A. Mathematical and General*, 3835-3859.
- Aihara, K., Matsumoto, G., & Ikegaya, Y. (1984). Periodic and Non-periodic Responses of a Periodically Forced Hodgkin-Huxley Oscillator. *Journal of Theoretical Biology*, 109, 249-269.
- Aihara, K. (1990a). Chaotic Neural Networks. In H. Kawakami (Ed.), *Bifurcation Phenomena in Nonlinear Systems and Theory of Dynamical Systems* (pp. 143-161). World Scientific, Singapore.
- 合原一幸 編著. (1990b). *カオス - カオス理論の基礎と応用 -*. サイエンス社.
- Amari, S. (1971). Characteristics of Randomly Connected Threshold-Element Networks and Network Systems. *Proceedings of the IEEE*, 59(1), 35-47.
- Amari, S. (1972a). Characteristics of Random Nets of Analog Neuron-Like Elements. *IEEE Transaction on System, Man and Cybernetics*, SMC-2(5), 643-657.
- Amari, S. (1972b). Learning Patterns and Pattern Sequences by Self-Organizing Nets of Threshold Elements. *IEEE Transaction on Computers*, c-12(11), 1197-1206.
- 甘利俊一. (1978). *神経回路網の数理*. 産業図書.
- 甘利俊一, & 竹内彰一. (1979). 神経回路網の自己組織理論. *電子情報通信学会論文誌 A*, J62-A, 12, 903-910.
- Amari, S., & Maginu, K. (1988). Statics and Dynamics of Associative Memory. *Neural Networks*, 1(1), 63-73.
- Amari, S. (1989). Characteristics of Sparsely Encoded Associative Memory. *Neural Networks*, 2(6), 451-457.
- Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1987a). Information storage in neural networks with low levels of activity. *Physical Review A*, 35(5), 2293-2303.

- Amit, D. J., Gutfreund, H., & Sompolinsky, H. (1987b). Statistical Mechanics of Neural Networks near Saturation. *Annals of Physics*, 173, 30-67.
- 朝日奈崇, 高木秀彦, 水野政司, & 郷原一寿. (1993). リカレントネットワークのアトラクターと状態遷移図. 信学技法, NC92-119, 191-198.
- 麻生秀樹. (1988). ニューラルネットワーク情報処理 - コネクショニズム入門、あるいは柔らかな記号に向けて -. 産業図書.
- Baird, B. (1986). Nonlinear dynamics of pattern formation and pattern recognition in the rabbit olfactory bulb. *Physica*, 22D, 150-175.
- Bergé, P., Pomeau, Y., & Vidal, Ch. (1984). カオスの中の秩序 - 乱流の理解に向けて -. 相沢洋二 邦訳, 産業図書.
- Borisyuk, R. M., & Kirillov, A. B. (1992). Bifurcation analysis of a neural network model. *Biological Cybernetics*, 66, 319-325.
- Carpenter, G. A., & Grossberg, S. (1988). The ART of Adaptive Pattern Recognition by a Self-Organizing Neural Network, *IEEE Computer*, 21(3) 77-88.
- Chernoff, H. (1973). The Use of Faces to Represent Points in k-Dimensional Space Graphically. *Journal of the American Statistical Association*, 68(342), 361-368.
- 千葉浩彦. (1993). 感情の変容と表情. 吉川左紀子, 益谷真, & 中村真編, 顔と心 - 顔の心理学 - (pp. 110-135). サイエンス社.
- Davis, P. (1991). Adaptive Search with Chaos in Nonlinear-Optical Systems. In N. B. Arbraham, E. M. Garmin, & P. Mandel (Eds.), *OSA Proceedings on Nonlinear Dynamics in Optical Systems*, vol.7 (pp. 562-566). Optical Society of America, Washington, DC.
- Dewdney, A. K. (1986). Computer recreation: computer caricature. *Science*, December, 160-165.
- Ditto, W. L., & Pecorra, L. M. (1993). Mastering Chaos. *Scientific American*, August, 62-68.
- Doya, K. (1992). Bifurcation in the learning of Recurrent Neural Networks. *Proceedings of 1992 IEEE International Symposium on Circuits and Systems*, 2777-2780.

- Eckhorn, R., Bauer, R., Jordan, W., Brosch, M., Kruse, W., Munk, M., & Reitboeck, H. J. (1988). Coherent oscillations: A mechanism of feature linking in the visual cortex? - multiple electrode and correlation analysis in the cat -. *Biological Cybernetics*, 60, 121-130.
- Eckhorn, R., Reitboeck, H. J., Arndt, M., & Dicke, P. (1990). Feature Linking via Synchronization Among Distributed Assemblies: Simulations of Results from Cat Visual Cortex. *Neural Computation*, 2(3), 293-307.
- Funahashi, K. (1989). On the Approximate Realization of Continuous Mapping by Neural Networks. *Neural Networks*, 2(3), 183-192.
- 船橋賢一. (1990). 3層ニューラルネットワークによる恒等写像の近似的実現性についての理論的考察. 電子情報通信学会論文誌 A, J73-A, 1, 139-145.
- Funahashi, K., & Nakamura, Y. (1992). Approximation of Dynamical Systems by Continuous Time Recurrent Neural Networks. *Neural Networks*, 6(6), 801-806.
- Geszti, T. (1990). ニューラルネットワークの物理モデル. 秋葉巴他 邦訳, 吉岡書店.
- Getting P. A. (1989). Emerging principles governing the operation of neural networks. *Annual Review of Neuroscience*, 12, 185-204.
- Gouhara, K., Takase, H., Uchikawa, Y., & Iwata, K. (1992). Attractor learning of recurrent neural networks. In I. Aleksander, & J. Taylor (Eds.), *Artificial Neural Networks 2* (pp. 371-374). Elsevier Science Publishers.
- Gray, C. M., & Singer, W. (1989). Stimulus-specific neuronal oscillations in orientation columns of cat visual cortex. *Proceedings of the National Academy of Science, USA*, 86, 1689-1702.
- Gray, C. M., König, P., Engel, A. K., & Singer, W. (1990). Synchronization of Oscillatory Responses in Visual Cortex: A Plausible Mechanism for Scene Segmentation. In H. Haken, & M. Stadler (Eds.), *Synergetics of Cognition*, volume 86 (pp. 82-98). Springer-Verlag.
- Grossberg, S., & Somers, D. (1991). Synchronized oscillations during cooperative feature linking in a cortical model of visual perception. *Neural Networks*, 4(4), 453-466.

- Guckenheimer, J., & Holmes, P. (1983). *Nonlinear Oscillations, Dynamical Systems, and Bifurcations of Vector Fields*. Springer-Verlag.
- 長谷川浩, & 稲積満広. (1993). 実音声の入力によるリカレントニューラルネットワークの振る舞いについて. *信学技法*, NC93-5, 33-40.
- 原文雄. (1978). 顔グラフによるコミュニケーション - 機械から人間への情報伝達 -. *自然*, 9月号, 26-35.
- Hinton, G. E., & Shallice, T (1991). Lesioning an Attractor Network: Investigations of Acquired Dyslexia. *Psychological Review*, 91(1), 74-95.
- Hinton, G. E., Plaut, D. C., & Shallice, T (1993). Simulating Brain Damage. *Scientific American*, October, 58-65.
- Hirsch, M. W., & Smale, S. (1974). *Differential Equations, Dynamical Systems, and Linear Algebra*. Academic Press, Inc.
- Hopfield, J. J. (1982). Neural networks and physical systems with emergent collective computational abilities. *Proceedings of the National Academy of Science, USA*, 79, 2554-2558.
- Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two state neurons. *Proceedings of the National Academy of Science, USA*, 81, 3088-3092.
- Horn, D., & Usher, M. (1991). Parallel activation of memories in an oscillatory neural network. *Neural Computation*, 3(1), 31-43.
- Horn, D., Sagi, D., & Usher, M. (1991). Segmentation, Binding, and Illusory Conjunctions. *Neural Computation*, 3(4), 510-525.
- 今井四郎. (1986). パターン認知の変換構造説. *心理学モノグラフ*, No. 17, 日本心理学会.
- 伊藤進. (1975). パターン間の変換構造の認知と類似性の評定. *心理学研究*, 46(1), 10-18.
- Kawato, M., & Suzuki, R. (1980). Two Coupled Neural Oscillators as a Model of the Circadian Pacemaker. *Journal of Theoretical Biology*, 86, 547-575.
- 川原茂敬. (1992). 神経回路の多機能性. *HOLONICS* ホロニック技術研究会誌, 3(3), 121-129.

- Khibnik, A. I., Borisyuk, R. M., & Roose, D. (1992). Numerical bifurcation analysis of a model of coupled neural oscillators. *International Series of Numerical Mathematics*, 86, 547-575.
- 小林繁, 熊倉鴻之助, 黒田洋一郎, & 畠中寛 共著. (1987). 絵ときブレインサイエンス入門 -脳と神経の科学-. オーム社.
- 小粥泉慧. (1985). 幸せをよぶ手相人相占い. 新星出版.
- Kohonen, T. (1977). システム論的連想記憶 -情報工学・心理学のために-. 中谷和夫邦訳, サイエンス社.
- Kosko, B. (1989). Unsupervised Learning in Noise. *Proceedings of the International Joint Conference on Neural Networks '89*, Washington D.C., Vol.I, 7-17.
- König, P., & Schillen, T. B. (1991). Stimulus-dependent assembly formation of oscillatory responses: I. synchronization and II. desynchronization. *Neural Computation*, 3(2), 155-177.
- Kuramoto, Y. (1991). Collective synchronization of pulse-coupled oscillators and excitable units. *Physica*, D50, 15-30.
- 黒田洋一郎. (1991). ヒトの記憶・学習のメカニズム -大脳皮質・海馬間閉回路と海馬の記憶強化・連合機能-. *Brain Medical*, 3(1), 13-20.
- Lie, Z., & Hopfield, J. J. (1989). Modeling the olfactory bulb and its neural oscillatory processings. *Biological Cybernetics*, 61, 379-392.
- Linsker, R. (1988). Self-Organization in a Perceptual Network. *IEEE Computer*, 21(3), 105-117.
- Lippmann, R. L. (1987). An Introduction to Computing with Neural Nets. *IEEE ASSP Magazine*, April, 4-22.
- Lummer, E., & Huberman, B. A. (1992). Binding Hierarchies: A Basis for Dynamic Perceptual Grouping. *Neural Computation*, 4(3), 341-355.
- Matsuoka, K. (1985). Sustained oscillations generated by mutually inhibiting neurons with adaptation. *Biological Cybernetics*, 52, 367-376.
- Matsuoka, K. (1987). Mechanisms of frequency and pattern control in the neural rhythm generators. *Biological Cybernetics*, 56, 345-353.

- Matsuoka, K. (1989). An associative network with cross inhibitory connections. *Biological Cybernetics*, 61, 393-399.
- 松岡清利. (1990). 直交射影型連想記憶回路の種々の構造について. 電子情報通信学会論文誌 *DII*, J73-DII, 4, 641-647.
- 松本修文. (1992). 40ヘルツ波による共鳴結合の仮説. 生体の科学, 43(1), 25-29.
- McClelland, J. L., & Rumelhart, D. E. (1981). An Interactive Activation Model of Context Effects in Letter Perception: Part 1. An Account of Basic Findings, *Psychological Review*, 88(5), 375-407.
- Miyashita, Y., & Chang, H. S. (1988). Neuronal correlate of pictorial short-term memory in the primate temporal cortex. *Nature*, 331, 68-70.
- Mori, Y., Davis, P., & Nara, S. (1989). Pattern retrieval in an asymmetric neural network with embedded limit cycles. *Journal of Physics*, A22, L525-L532.
- 森正武, 名取亮, & 鳥居達生 共著. (1985). 数値計算. 岩波講座 情報科学 18, 岩波書店.
- Murakami, K., Koshimizu, H., Nakayama, A., & Fukumura, T. (1993). Facial Caricaturing Based on Visual Illusion - A Mechanism to Evaluate Caricature in PICASSO System -. *IEICE Transaction on Information and Systems*, E76-D(4), 470-478.
- 室賀三郎, 茨木俊秀, & 北橋忠宏 共著. (1976). しきい論理. 産業図書.
- 中村雄一, & 船橋賢一. (1992). 漸近安定平衡点を持つ力学系の軌道のリカレントネットワークによる近似について. 信学技法, NC92-80, 1-8.
- 長尾真編. (1983). パターン認識と図形処理. 岩波講座 情報科学 21, 岩波書店.
- Nagashino, H., & Kelso, J. A. (1991). Bifurcation of Oscillatory Solutions in A Neural Oscillator Network Model. *Proceeding of The Second Symposium on Non-linear Theory and its Applications*, 119-122.
- Nara, S., & Davis, P. (1992). Chaotic Wandering and Search in a Cycle-Memory Neural Network. *Progress of Theoretical Physics*, 88(5), 845-855.
- Nara, S., Davis, P., & Totsuji, H. (1993a). Memory Search Using Complex Dynamics in a Recurrent Neural Network Model. *Neural Networks*, 6(7), 963-973.

- Nara, S., Davis, P., Kawachi, M., & Totsuji, H. (1993b). Memory Dynamics in a Recurrent Neural Network with Cycle Memories Embedded by Pseudo-Inverse Method. *Proceedings of 1993 International Joint Conference on Neural Networks*, III, 2383-2386.
- Newhouse, S., Ruelle, D., & Takens, F. (1978). Occurrence of Strange Axion A Attractors Near Quasi-Periodic Flows on  $T^m$ ,  $m \geq 3$ . *Communications in Mathematical Physics*, 64, 35-40.
- Nicols, J. S., & Tsuda, I. (1985). Chaotic dynamics of information processing: The "Magic number seven plus-minus two" Revisited. *Bulletin of Mathematical Biology*, 47(3), 343-365.
- Nicols, J. S. (1986). Chaotic dynamics applied to information processing. *Reports on Progress in Physics*, 49, 1109-1196.
- Niebur, E., Kammen, D. M., & Koch, C. (1991). Phase-Locking in 1-D and 2-D Networks of Oscillating Neurons. In H. G. Schuster (Ed.), *Nonlinear Dynamics and Neuronal Networks - Proceedings of the 63rd W. E. Heraeus Seminar Friedrichsdorf 1990* - (pp. 173-203). VCH.
- Nozawa, H. (1992). A Neural Network Model as a Globally Coupled Map and Applications based on Chaos. *Chaos*, 2(3), 377-386.
- Oda, M. (1991). Context Dependency Effect in the Formation of Image Concepts and its Application. *Proceedings of 1991 IEEE International Conference on Systems, Man, and Cybernetics*, 1673-1678.
- Palm, G. (1988). Local Synaptic Rilles with Maximal Information Storage Capacity. In H. Haken (Ed.), *Neural and Synergetic Computers*, volume (pp. 100-110). Springer-Verlag.
- Parramon, J. M. (1980). 肖像画の描法. 岸田杏子 邦訳, 絵画教室シリーズ -7, グラフィック社.
- Pearlmutter, B. A. (1989). Learning state space trajectories in recurrent neural network. *Neural Computation*, 1(2), 263-269.
- Personaz, L. I., Guyon, I., & Dreyfus, G. (1986). Collective Computation Properties of Neural Networks: New Learning Mechanisms. *Physical Review A*, 34(5), 4217-4228.

- Piaget, J. (1970). Piaget's theory. In P. Mussen (Ed.), *Carmichael's manual of child psychology*, volume 1 (pp. 703-732). New York: Wiley.
- Renals, S., & Rohwer, R. (1990). A study of network dynamics. *Journal of Statistical Physics*, 58, 825-848.
- Ruelle, D., & Takens, F. (1971). On the Nature of Turbulence. *Communications in Mathematical Physics*, 20, 167-192.
- Rumelhart, D. E., McClelland, J. L., & the PDP Research Group. (1986). *Parallel Distributed Processing - Explorations in the Microstructure of Cognition*. volume 1 & 2, The MIT Press.
- 坂口豊, & 村田昇. (1990). 出力信号の構造を反映した競合学習型神経回路モデル - 競合的な中間層を持つ多層神経回路による非線形関数の実現 -. 信学技法, NC89-54, 33-38.
- Sakarda, C. A., & Freeman, W. J. (1987). How brains make chaos in order to make sense of the world. *Behavioral and Brain Sciences*, 10, 161-195.
- Sanger, T. D. (1989). Optimal Unsupervised Learning in a Single-Layer Linear Feed-forward Neural Network. *Neural Networks*, 2(6), 459-473.
- Sato, M. (1990). A learning algorithm to teach spatiotemporal patterns to recurrent neural networks. *Biological Cybernetics*, 62, 259-263.
- Sato, M., Murakami, Y., & Joe K. (1990a). Chaotic dynamics by recurrent neural networks. *Proceedings of the International Conference on Fuzzy Logic and Neural Networks*, 601-604.
- Sato, M., Joe, K., & Hirahara T. (1990b). APOLONN brings us to the real world: Learning nonlinear dynamics and fluctuations in nature. *Proceedings of the International Joint Conference on Neural Networks, San Diego, I*, 581-587.
- 佐藤省三, 久野弘明, & 郷原一寿. (1993). リカレントネットワークによる音声波形の学習. 信学技法, NC93-56, 9-16.
- Schuster, H. G., & Wagner, P. (1990). A model for neuronal oscillations in the visual cortex - 1. mean-field theory and derivation of the phase equations -. *Biological Cybernetics*, 64, 77-82.

- Shaw, R., & Wilson, B. E. (1976). Abstract conceptual knowledge: How we know what we know. In D. Klahr (Ed.), *Cognition and instruction* (pp. 197-221). Lawrence Erlbaum.
- Shimizu, H., & Yamaguchi, Y. (1989). How Animals Understand the Meaning of Indefinite Information from Environments?. *Progress of Theoretical Physics Supplement*, 99, 404-424.
- 清水博他, 東大薬学部研究室各位. (1992). 箱根合宿・討論会「脳に機能単位は存在するか」 - その1, 2-. *HOLONICS* ホロニック技術研究会誌, 3(2), 83-93, 及び, 3(3), 131-138.
- Sigel, I. E., & Cocking, R. R. (1977). 認知の発達 - 乳児期から青年期まで -. 子安増生 邦訳, ライブラリ教育の心理学-3, サイエンス社.
- Sompolinsky, H., Golomb, D., & Kleinfeld, D. (1991). Phase Coherence and Computation in a Neural Network of Coupled Oscillators. In H. G. Schuster (Ed.), *Nonlinear Dynamics and Neuronal Networks - Proceedings of the 63rd W. E. Heraeus Seminar Friedrichsdorf 1990* - (pp. 113-130). VCH.
- Takano, Y. (1989). Perception of rotated forms: A theory of information types. *Cognitive Psychology*, 21, 1-59.
- 田中敏隆. (1969). 図形認知の発達の研究 (2) - 知覚的段階と思考的段階の検討 -. 心理学研究, 40(3), 146-150.
- 丹治順, 久保田競, 岩村吉晃, 小川尚, 藤田一郎, 辻下徹, 猪苗代盛, & 村岡洋一. (1992). 紙上討論「大脳皮質のコラム (モジュール) 構造」. 脳の高次機能の計算論的および実験的研究, 文部省重点領域研究ニュースレター No.11, 3-17.
- Tank, D. W., & Hopfield, J. J. (1987). Neural computation by concentrating information in time, *Proceedings of the National Academy of Science, USA*, 84, 1896-1900.
- Tompson, J. M. T., & Stewart, H. B. (1986). *Nonlinear Dynamics and Chaos, Geometrical Methods for Engineers and Scientists*. John Wiley and Sons Ltd.
- 鳥岡豊武, & 池田信彦. (1989). パターン分離層を有する連想記憶システムの連想能力. 電子情報通信学会論文誌 D-II, J72-DII, 11, 1945-1953.

- 塚田稔, 岩村吉晃, 清水博, 星宮望, 二見亮弘, 津田一郎, & 合原一幸. (1992). 紙上討論「脳は時間コードをどのように使っているか」. 脳の高次機能の計算論的および実験的研究, 文部省重点領域研究ニュースレター No.10, 5-12.
- 塚原仲晃 編. (1984). 脳の情報処理, 朝倉書店.
- Tsuda, I., Koerner, E., & Shimizu, H. (1987). Memory dynamics in asynchronous neural networks. *Progress of Theoretical Physics*, 78(1), 51-71.
- 津田一郎. (1990). カオスの脳観 - 脳の新しいモデルをめざして -. サイエンス社.
- Tsuda, I. (1992). Dynamic Link of Memory - Chaotic Memory Map in Nonequilibrium Neural Networks. *Neural Networks*, 5(2), 313-326.
- von der Malsburg, C. P., & Buhmann, J. (1992). Sensory segmentation with coupled neural oscillators. *Biological Cybernetics*, 67, 233-242.
- Wang, D., Buhmann, J., & von der Malsburg, C. (1990). Pattern segmentation in associative memory. *Neural Computation*, 2(1), 94-106.
- Wang, L., Pichler, E. E., & Ross, J. (1990). Oscillations and chaos in neural networks: An exactly solvable model. *Proceedings of the National Academy of Science, USA*, 87(1), 9464-9471.
- Wiggins, S. (1990). *Introduction to Applied Nonlinear Dynamical Systems and Chaos*. Springer Verlag.
- Wilson, H. R., & Cowan, J. D. (1973). A Mathematical Theory of the Functional Dynamics of Cortical and Thalamic Nervous Tissue. *Kybernetik*, 13, 55-80.
- 山本稔. (1979). 常微分方程式の安定性. 実教出版.
- Yao, Y., & Freeman, W. J. (1990). Model of biological pattern recognition with spatially chaotic dynamics. *Neural Networks*, 3(2), 153-170.
- 横瀬善正. (1986). 形の心理学. 名古屋大学出版会.

## 本論文に関する原著論文

### <学術論文>

- 林 幸雄. (1992). フィードバック連想記憶モデルによる手書き文字の認識学習. 電子情報通信学会論文誌 D-II, Vol.J75-D-II(5), 956-964.
- Yukio Hayashi. (1994). Oscillatory Neural Network and Learning of Continuously Transformed Patterns. (To appear in Neural Networks).
- Yukio Hayashi. (1994). Numerical Bifurcation Analysis of an Oscillatory Neural Network with Synchronous/Asynchronous Connections. (To appear in Neural Computation).
- Yukio Hayashi. (1994). Dynamical Search Domain Control According to the Input Bias of an Oscillatory Neural Network. (Submitted to Neural Networks).

### <国際会議>

- Yukio Hayashi. (1992). Learning of Continuously Transformed Pattern Cycles by an Oscillatory Neural Network. Proceedings of the International Joint Conference on Neural Networks '92, Baltimore, IV, 122-127.
- Yukio Hayashi. (1993). Switching the Vector Field According to the Input of an Oscillatory Neural Network. Proceeding of the International Conference on Artificial Neural Networks '93, Amsterdam. (Abstract, pp. 684).
- Yukio Hayashi. (1993). Dependency of the Connection Architectures of Oscillatory Neural Networks on Synchronization. Proceedings of the International Joint Conference on Neural Networks '93, Nagoya, II, 1434-1438.
- Yukio Hayashi. (1993). Ambiguously Fluctuating Associative Memory Model with Hysteresis Dependency. Proceedings of the International Joint Conference on Neural Networks '93, Nagoya, III, 2323-2326.

## <研究会、全国大会等>

- 林 幸雄. (1990). 直交射影型フィードバック連想モデルによるパターン認識生成について. 信学技法 NC90-36, 45-52.
- 林 幸雄. (1991). 記憶点の不安定化による発振ニューラルネット. 信学技法 NC91-39, 67-74.
- 林 幸雄. (1992). 発振ニューラルネットの結合パラメータに対する周期/準周期軌道の分岐. 信学技法 NC92-46, 13-20.
- 林 幸雄. (1993). 連想記憶モデルのベイズン拡大/縮小について. 信学技法 NC93-16, 49-56.
- 林 幸雄. (1993). 状態空間の構造変化による文脈効果を持つ連想記憶モデル. 神経回路学会第4回全国大会, 78-79.

## <社内技術報告書>

- Yukio Hayashi. (1993). Dynamical Control of Cluster Boundaries by Changing the State Space Structure of an Associative Memory Model. ATR Technical Report, TR-H-011.
- Yukio Hayashi. (1993). Oscillatory Neural Network and Learning of Continuously Transformed Patterns. ATR Technical Report, TR-H-012.
- Yukio Hayashi. (1993). Switching the Vector Field According to the Input of an Oscillatory Neural Network. ATR Technical Report, TR-H-013. (Extended version of the abstract in ICANN'93).
- Yukio Hayashi. (1993). Numerical Bifurcation Analysis of an Oscillatory Neural Network with Synchronous/Asynchronous Connections. ATR Technical Report, TR-H-014.
- 林 幸雄, & 大森 洋一. (1994). 発振ニューラルネットの遊走的な出力軌道の性質変化に関する実験的解析. ATR Technical Report, TR-H-063.

## 付録 A

### 付録：Lyapunov 安定性の証明

まず、以下のベクトル-行列表記の動作方程式で定義される、連続時間の連想記憶モデルの認識時の挙動に関して、Lyapunov 関数と呼ばれる有界な関数の極小点への漸近安定収束性を示す。Lyapunov 関数の厳密な数学的定義については、微分方程式の教科書 (Hirsch, & Samle, 1974; 山本, 1979; Wiggins, 1990) を参照されたい。

[動作方程式]

$$\tau \dot{y}(t) = f(y(t)) + Rf(y(t)) + W^T g(x(t)), \quad (\text{A.1})$$

$$\tau_x \dot{x}(t) = -x(t) + Wf(y(t)), \quad (\text{A.2})$$

$$R \equiv -W^T W + I_M. \quad (\text{A.3})$$

ここで、非線形関数  $f_j(y_j)$ ,  $g_i(x_i)$  は各要素  $y_j$ ,  $x_i$  ごとに施され、2.3節の図 2.5に示したような有界で微分可能な単調増加関数とする。また、 $\tau_x > \tau > 0$  は時定数、 $I_M$  は  $M$  次元単位行列、 $T$  は転置行列、 $\cdot$  (ドット) と  $'$  (ダッシュ) は、関数の時間微分と空間微分をそれぞれ表す。

さて、結合重み係数や非線形関数のしきい値を固定した認識時に、以下の有界な Lyapunov 関数を考える。

[Lyapunov 関数]

$$L(x, y) = -\sum_{i,j} W_{ij} g_i(x_i) f_j(y_j) - \frac{1}{2} \sum_{k,j} f_k(y_k) f_j(y_j) + \sum_i \int_0^{x_i} g'_i(\theta_i) \theta_i d\theta_i + \frac{1}{2} \sum_j f_j(y_j)^2. \quad (\text{A.4})$$

式 (A.1) と式 (A.2)、および、非線形関数  $f_j(y_j)$  と  $g_i(x_i)$  の単調性から、

$$\begin{aligned} \dot{L} &= -\sum_i \dot{x}_i g'_i(x_i) \{-x_i + \sum_j W_{ij} f_j(y_j)\} - \sum_j \dot{y}_j f'_j(y_j) \{-f_j(y_j) + \sum_i W_{ij} g_i(x_i)\}, \\ &= -\tau_x \sum_i \dot{x}_i^2 g'_i(x_i) - \tau \sum_j \dot{y}_j^2 f'_j(y_j) \leq 0. \end{aligned} \quad (\text{A.5})$$

よって、認識処理の挙動は、初期状態  $x_i(0), y_i(0)$  に従った Lyapunov 関数  $L(x, y)$  の極小解 ( $\dot{x}_i = 0, \dot{y}_i = 0$ ) に漸近安定収束する。

同様に、学習時の挙動の漸近安定収束性を示す為、以下のように再定義した動作方程式と学習方程式を考える。

[動作方程式]

$$\tau \dot{y}(t) = F(y(t), a(t)) + RF(y(t), a(t)) + W^T g(x(t)), \quad (\text{A.6})$$

$$\tau_x \dot{x}(t) = -x(t) + WF(y(t), a(t)). \quad (\text{A.7})$$

[学習方程式]

$$\tau_a \dot{a}(t) = F_a(y(t), a(t))(\tau \dot{y}), \quad (\text{A.8})$$

$$\tau_w \dot{W}(t) = -\gamma W(t) + g(x(t))F(y(t), a(t))^T - W(t)F(y(t), a(t))F(y(t), a(t))^T. \quad (\text{A.9})$$

ここで、 $\gamma > 0$  は結合重み値の忘却係数で、時定数  $\tau_w, \tau_a$  は、各素子の状態更新より学習による結合重み値  $W_{ij}$  やしきい値  $a_j (> 0)$  の変化が遅くなるように、 $\tau_w, \tau_a \gg \tau_x, \tau$  とし、非線形関数  $f_j(y_j)$  も以下のような  $F(y_j, a_j)$  に再定義する。この時、関数  $F(y_i, a_i)$  の変数  $y_j$  及び  $a_j$  による偏微分は、 $F_y(y, a) \equiv \partial F / \partial y_j \geq 0$  と  $F_a(y, a) \equiv \partial F / \partial a_j \leq 0$  となることに注意されたい。

$$F(y_j, a_j) = \begin{cases} 0 & (y_j \leq 0), \\ 2y_j^2/a_j & (0 < y_j \leq a_j/2), \\ -2(y_j - 1)^2/(2 - a_j) + 1 & (a_j/2 < y_j < 1), \\ 1 & (y_j \geq 1). \end{cases} \quad (\text{A.10})$$

[Lyapunov 関数]

$$\begin{aligned} \mathcal{L}(x, y, W, a) = & -\sum_{i,j} W_{ij} g_i(x_i) F(y_j, a_j) - \frac{1}{2} \sum_{k,j} F(y_k, a_k) F(y_j, a_j) \\ & + \sum_i \int_0^{x_i} g_i'(\theta_i) \theta_i d\theta_i + \frac{1}{2} \sum_j F(y_j, a_j) + \frac{\gamma}{2} \sum_{i,j} W_{ij}^2. \end{aligned} \quad (\text{A.11})$$

上記の新たに定義された Lyapunov 関数  $\mathcal{L}(x, y, W, a)$  について、式 (A.6) ~ (A.9) を用いて、

$$\dot{\mathcal{L}}(x, y, W, a) = -\tau_x \sum_i \dot{x}_i^2 g_i'(x_i) - \tau \sum_j \dot{y}_j^2 F_y(y_j, a_j) - \tau_a \sum_j \dot{a}_j^2 - \tau_w \sum_{i,j} \dot{W}_{ij}^2 \leq 0. \quad (\text{A.12})$$

よって、学習処理の挙動に関しても、入出力素子、結合重み係数、及び、しきい値の初期状態に従った、Lyapunov 関数  $\mathcal{L}(x, y, W, a)$  の極小解への漸近安定収束性が示された。

## 付録 B

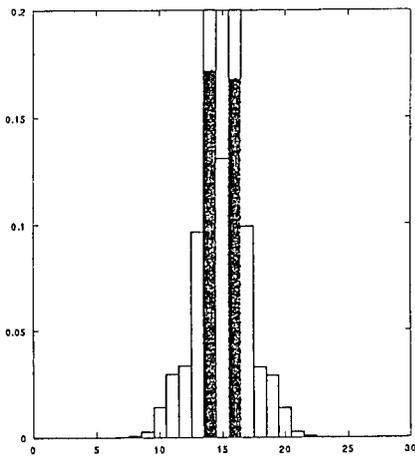
### 付録： 遊走軌道に対する最近接頂点の滞在率の分布

記憶パターン数の増加に伴う記憶パターンへの引力の弱体化による遊走軌道の性質変化について、次元数（出力素子数） $N$  が 30 と 50 の場合で、3.5節と同様な数値解析を行った。すなわち、軌道の最近接頂点に対する滞在率の分布と時間的遷移を、 $N = 30$  の記憶パターン数  $M = 4$  ( $< 0.14N$ ) と 5 ( $> 0.14N$ ) の場合で比較し、同様に  $N = 50$  の  $M = 6$  ( $< 0.14N$ ) と 8 ( $> 0.14N$ ) の場合でもこれを行った。但し、無入力バイアス  $I_i \equiv 0$  で、結合重み係数などのパラメータは、3.5節と同じ値に設定した。又、記憶パターン  $\xi^\mu$  ( $\mu = 1, 2, \dots, M$ ) は、同数の  $\pm 1$  要素をランダムに持つ  $N$  次元ベクトルで、互いに相関  $|\sum_{i=1}^N \xi_i^\alpha \xi_i^\beta|$  ( $\alpha \neq \beta$ ) が  $\sqrt{N}$  以下のものとした。

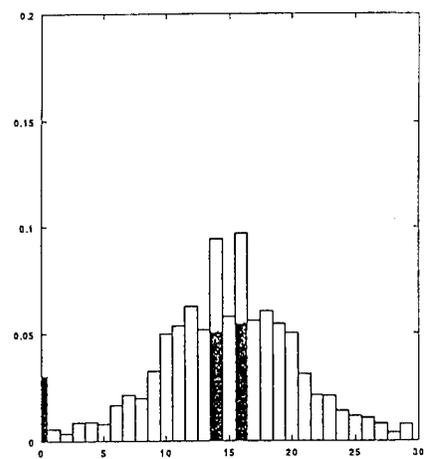
$N = 30$  で  $M = 4$  ( $< 0.14N$ ) と 5 ( $> 0.14N$ ) の場合の最近接頂点の分布（基準点は  $\xi^1$ ）を、図 B.1 と図 B.2 にそれぞれ示す。図中の記号 (a) から (e) は、5 つの異なる初期状態（但し、 $|x_i(0)| < 0.2$ ,  $y_i \equiv 0$  とした）に対する分布を示すもので、グラフ中の黒棒は記憶パターンの頂点の滞在率、白棒はその他の頂点の滞在率を表す。図 B.1 と図 B.2 を比較すると、3.5節の  $N = 22$  の場合と異なって分布の形が初期状態に依存した（定常分布に落ち着くまでの反復回数も 40 万回だった）ものの、臨界点  $M \approx 0.14N$  前後での記憶パターンの頂点への滞在率の低下は共通した傾向として読みとれる。また、図 B.3 と図 B.4 に示した、 $N = 50$  で  $M = 6$  ( $< 0.14N$ ) と 8 ( $> 0.14N$ ) の場合でも上記と同様なことが伺える。但し、臨界値  $M \approx 0.14N$  前後のその変化は  $N = 30$  の場合ほど顕著でなく、初期状態に対する分布の形の差もより大きくなっている。

次に、 $N$  と  $M$  のそれぞれの組合せに対する軌道の性質変化を、最近接頂点の時間的遷移から検討した。 $N = 30$  における図 B.5 と図 B.6 から、臨界点  $M \approx 0.14N$  前後での記憶パターンの頂点への滞在率（密度）の低下が読み取れる。 $N = 50$  における図 B.7 と図 B.8 から、同様な結果が読み取れるが、その変化の傾向は  $N = 30$  の場合より若干弱いことが伺える。

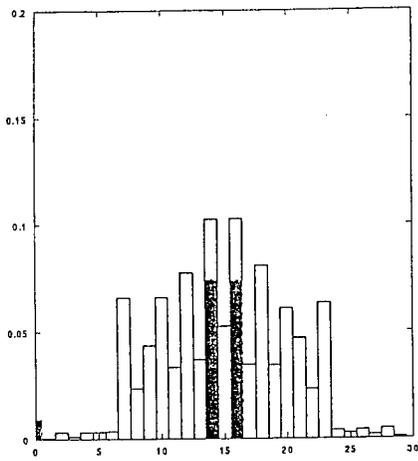
以上の結果を総合すると、3.5節と同様な記憶パターン数の増加に伴う軌道の性質変化が得られたものの、次元数が大きくなる程その傾向が顕著でなくなることや、軌道の最近接頂点の分布の形が初期状態に依存するという 3.5節とは異なる結果も得られた。これらの相違点に関しては、さらに詳細な検討を行う必要がある。



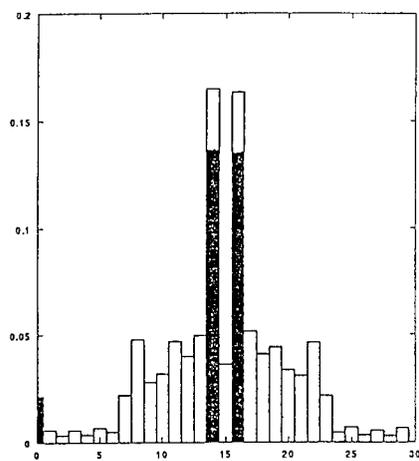
(a)



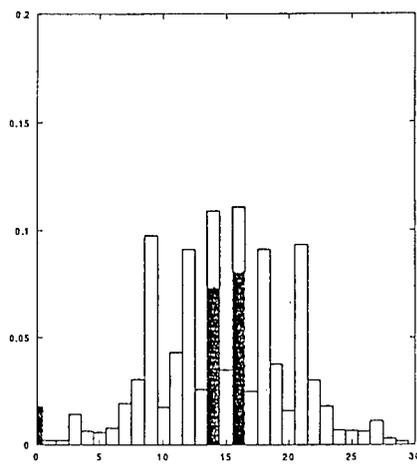
(b)



(c)

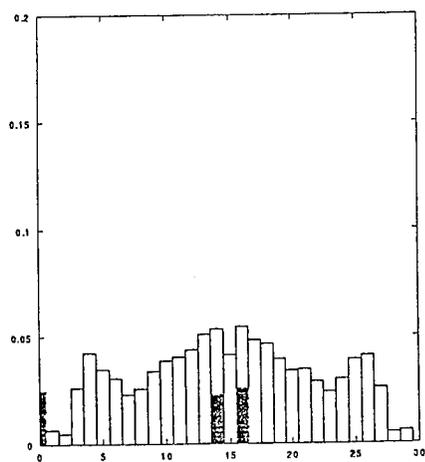


(d)

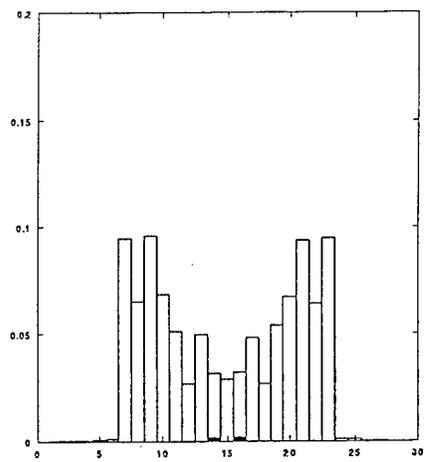


(e)

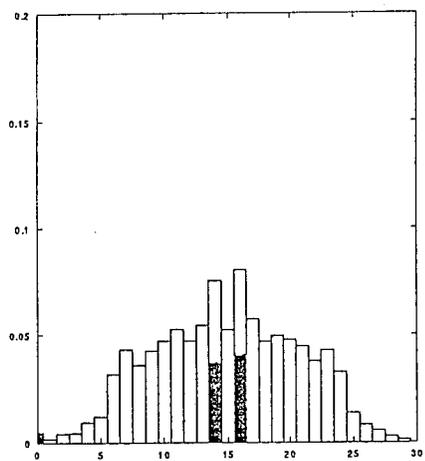
図 B.1: 最近接頂点の滞在率の分布 ( $N = 30, M = 4$ )



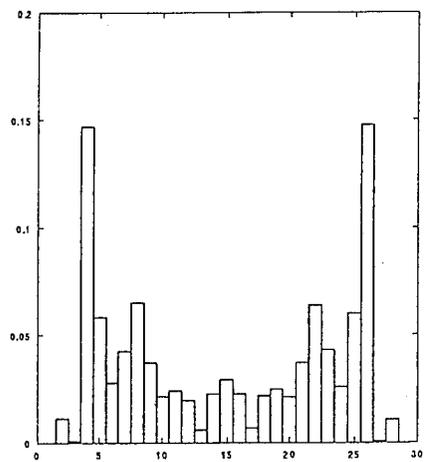
(a)



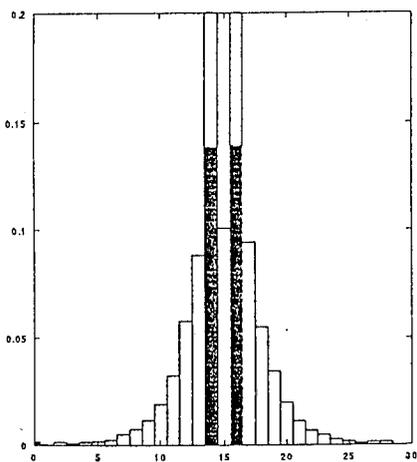
(b)



(c)

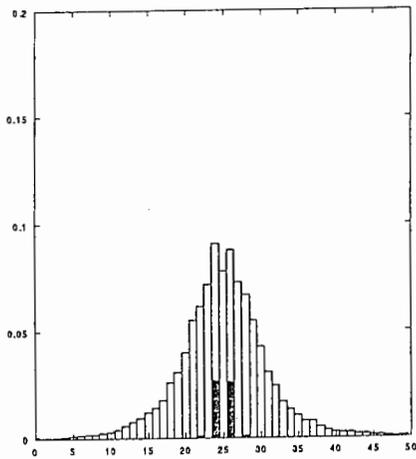


(d)

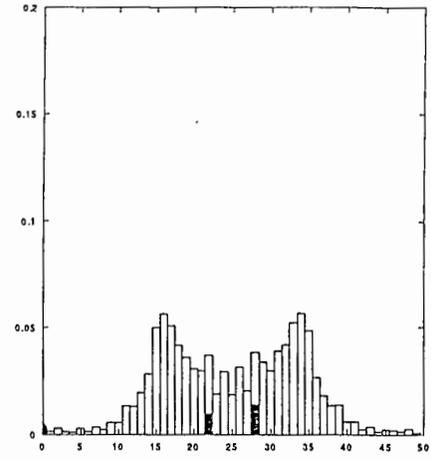


(e)

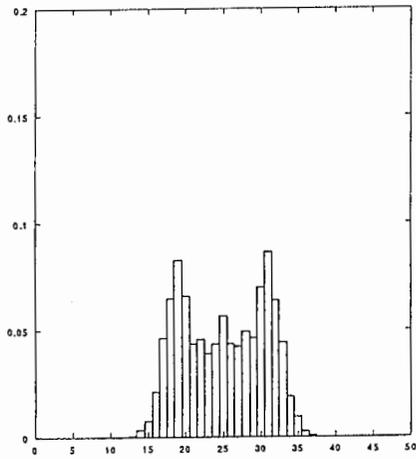
図 B.2: 最近接頂点の滞在率の分布 ( $N = 30, M = 5$ )



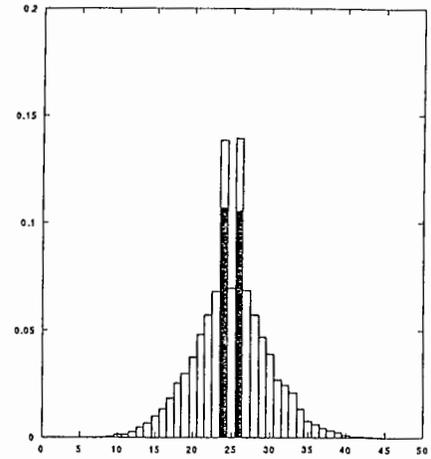
(a)



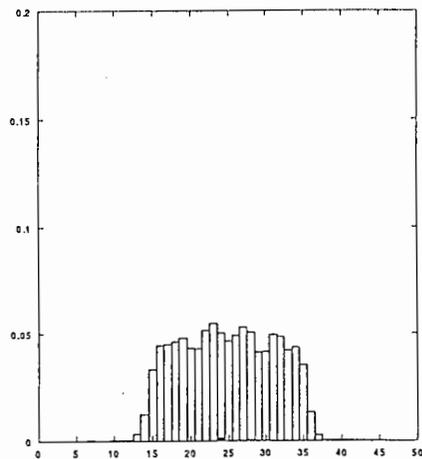
(b)



(c)

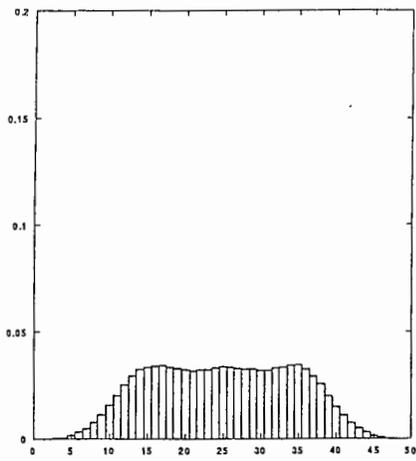


(d)

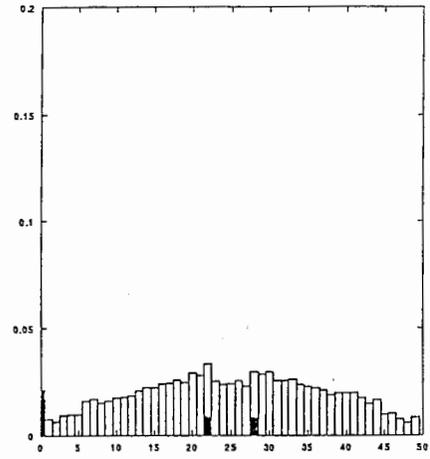


(e)

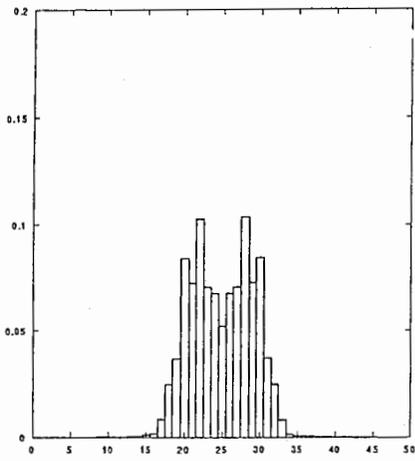
図 B.3: 最近接頂点の滞在率の分布 ( $N = 50, M = 6$ )



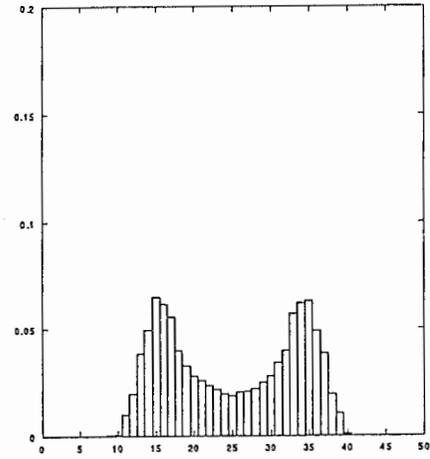
(a)



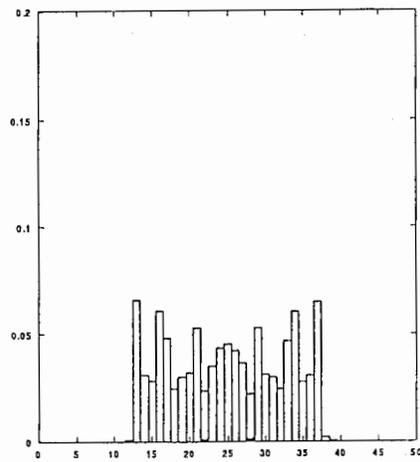
(b)



(c)

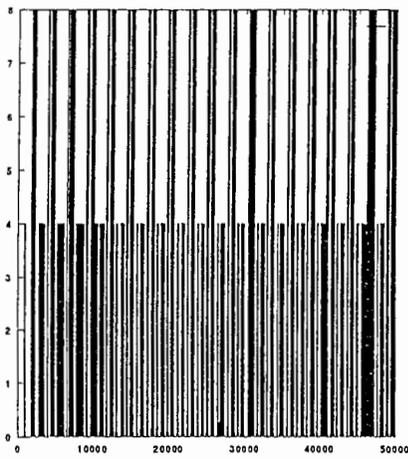


(d)

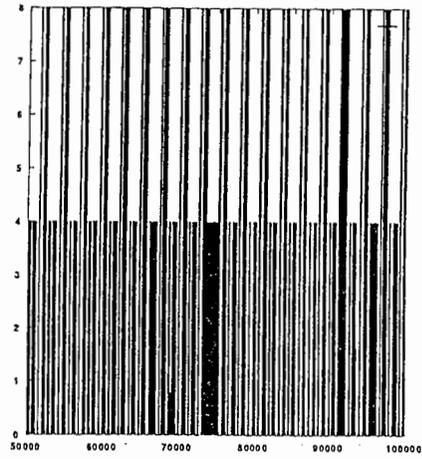


(e)

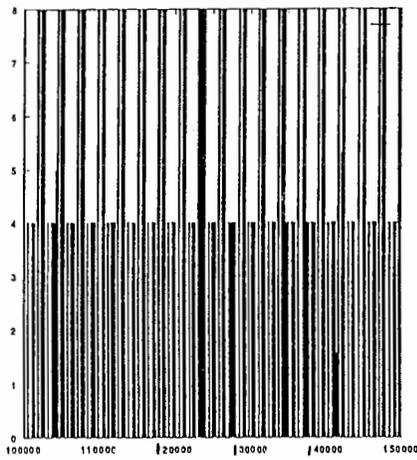
図 B.4: 最近接頂点の滞在率の分布 ( $N = 50, M = 8$ )



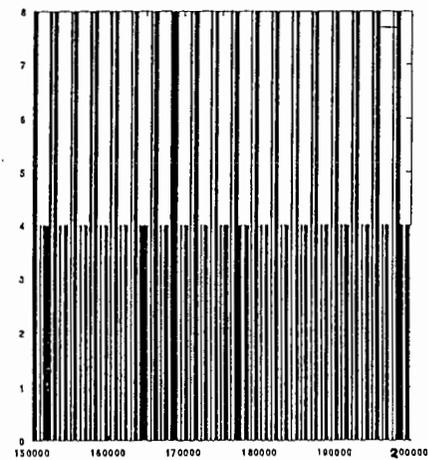
(a) 0 ~ 5万回



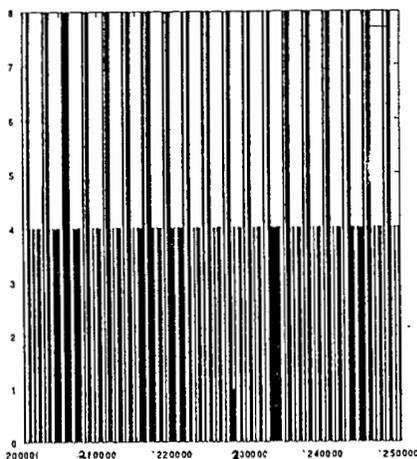
(b) 5万 ~ 10万回



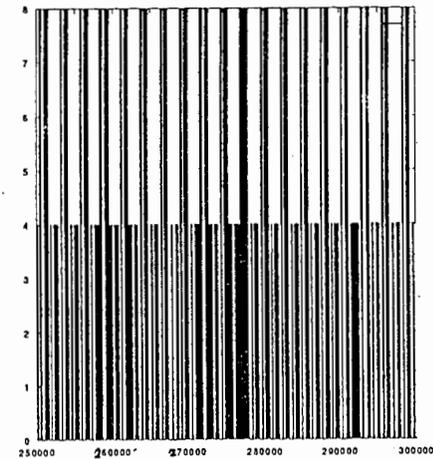
(c) 10 ~ 15万回



(d) 15 ~ 20万回

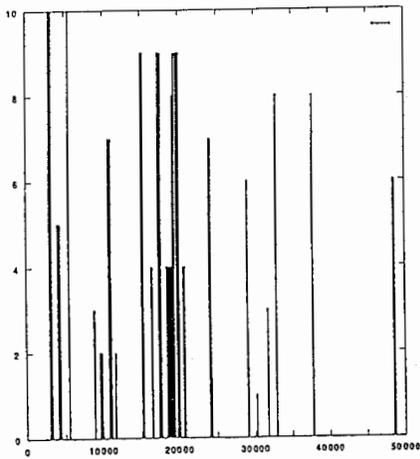


(e) 20 ~ 25万回

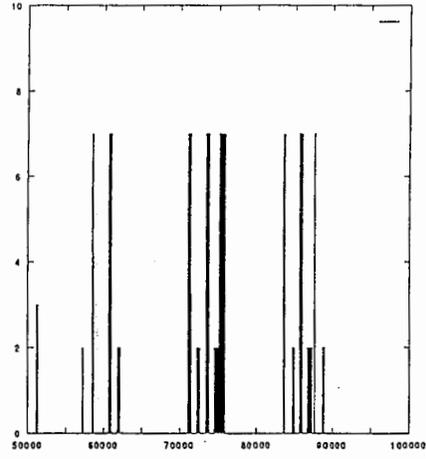


(f) 25 ~ 30万回

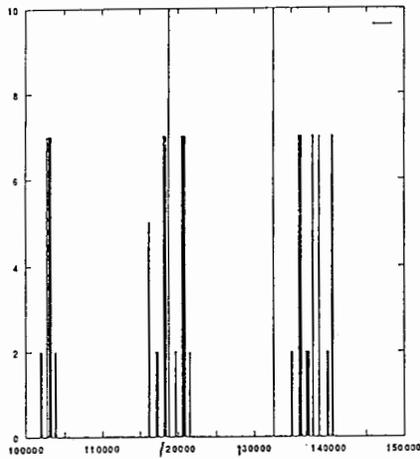
図 B.5: 最近接頂点の時間的遷移 ( $N = 30$ ,  $M = 4$ ) (縦軸: 記憶パターン番号で、 $\mu = 1 \sim M$  が  $\xi^\mu$  に、 $\mu = M + 1 \sim 2M$  が  $-\xi^{\mu-M}$  に対応し、0 はその他の頂点に対応、横軸: 反復回数)



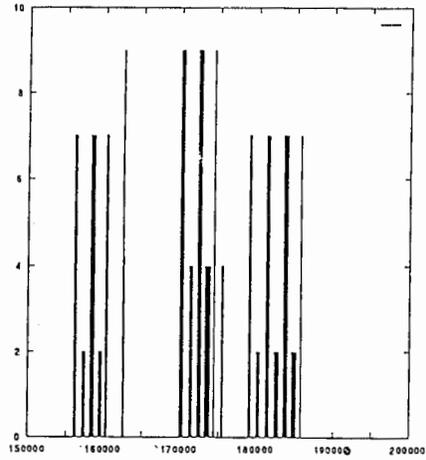
(a) 0 ~ 5万回



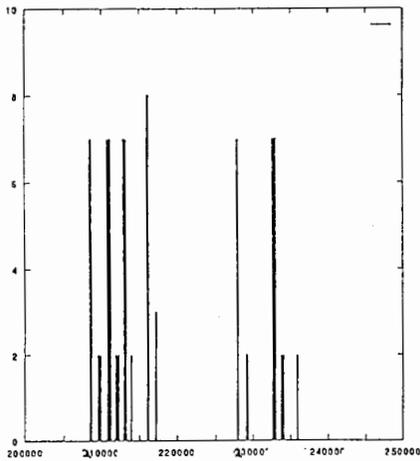
(b) 5万 ~ 10万回



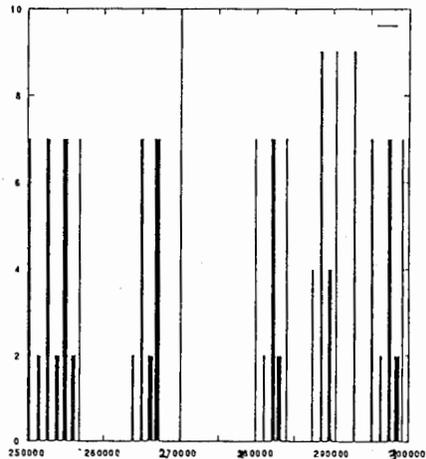
(c) 10 ~ 15万回



(d) 15 ~ 20万回

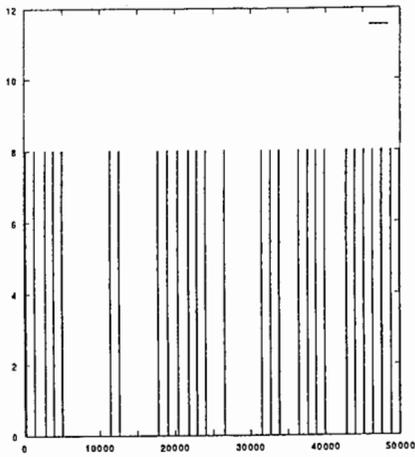


(e) 20 ~ 25万回

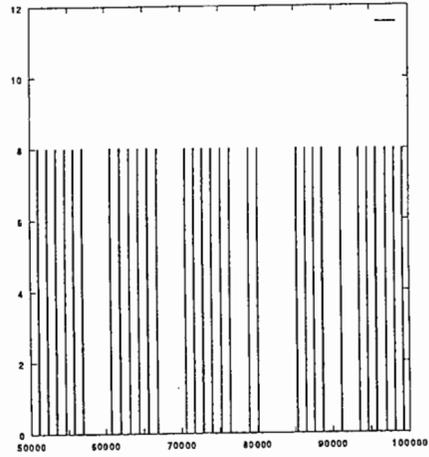


(f) 25 ~ 30万回

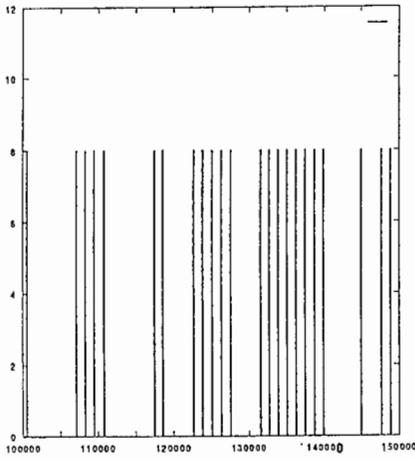
図 B.6: 最近接頂点の時間的遷移 ( $N = 30, M = 5$ )



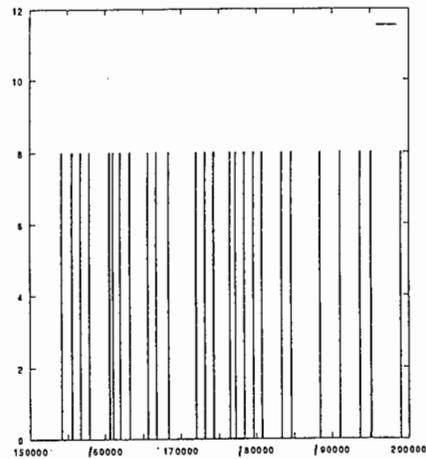
(a) 0 ~ 5万回



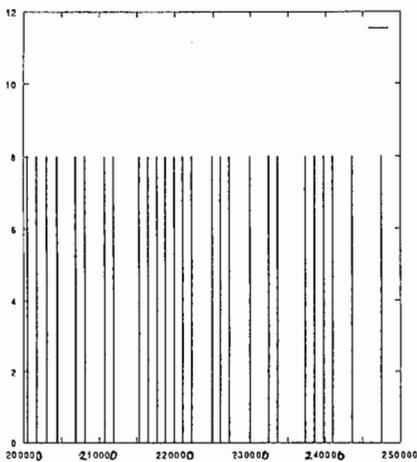
(b) 5万 ~ 10万回



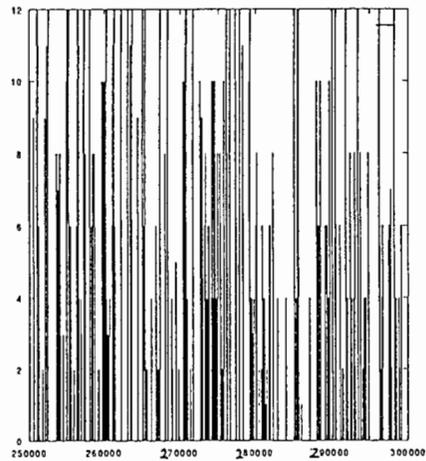
(c) 10 ~ 15万回



(d) 15 ~ 20万回

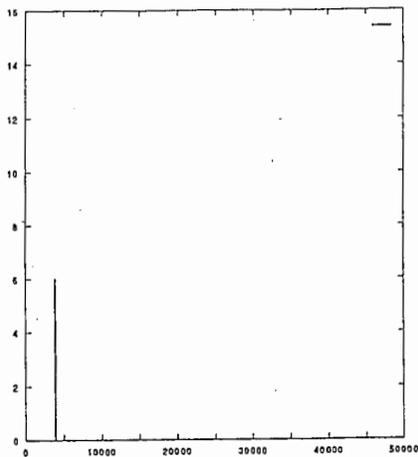


(e) 20 ~ 25万回

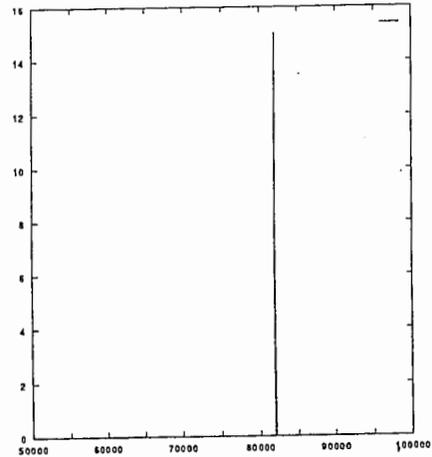


(f) 25 ~ 30万回

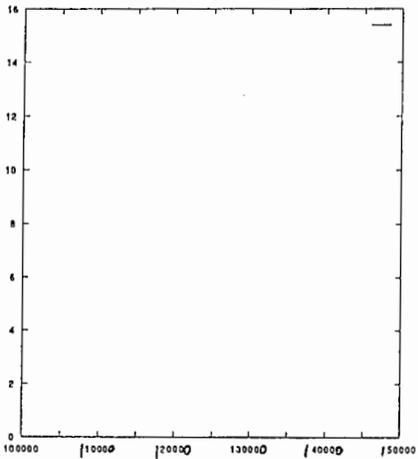
図 B.7: 最近接頂点の時間的遷移 ( $N = 50, M = 6$ )



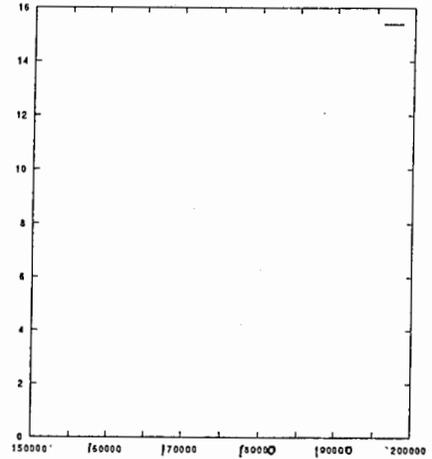
(a) 0 ~ 5万回



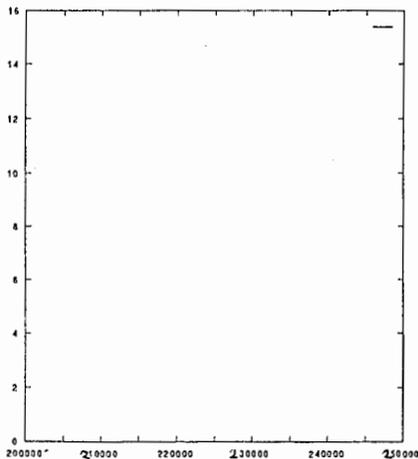
(b) 5万 ~ 10万回



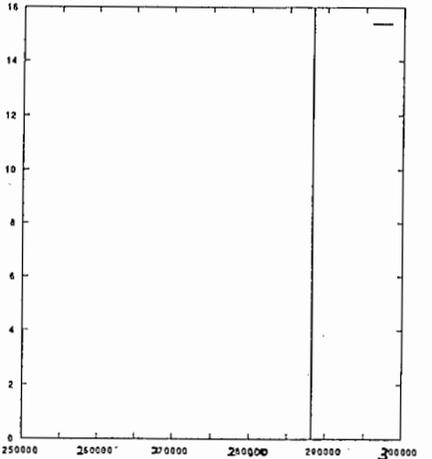
(c) 10 ~ 15万回



(d) 15 ~ 20万回



(e) 20 ~ 25万回



(f) 25 ~ 30万回

図 B.8: 最近接頂点の時間的遷移 ( $N = 50, M = 8$ )

## 付録 C

### 付録： 発振ニューラルネットワークの学習方程式の導出

まず、表記を単純化する為に、 $x_i(t)$  と  $y_i(t)$  の  $N$  個の変数をまとめて  $2N$  個の変数  $z_k(t)$  で表し、 $K_{IE}^{(i)}$  と  $K_{EI}^{(i)}$  を含めた結合重みパラメータ  $W_{ij}$  で動作方程式を書き改める。

$$\dot{z}_k(t) = -z_k(t) + G(u_k(t)), \quad (\text{C.1})$$

$$u_k(t) = \sum_{j=1}^{2N} W_{kj} z_j(t) \quad (k \notin V). \quad (\text{C.2})$$

$$z_k(t) = Q_k(t), \quad (\text{C.3})$$

$$J_k(t) = \dot{Q}_k(t) + Q_k(t) - G(u_k(t)), \quad (\text{C.4})$$

$$u_k(t) = \sum_{j=1}^{2N} W_{kj} z_j(t) + I_k \quad (k \in V). \quad (\text{C.5})$$

ここで、 $V$  は興奮性素子の教師信号に対する可視ユニットの集合を表す。学習時には、各可視ユニットの値は教師信号  $Q_k(t)$  に強制的に設定されるとする。以下では、佐藤によるリカレントネットワークの学習則の導出 (Sato, 1990) に従って、その1つの改良版としての発振ニューラルネットワークの学習方程式を導く。

さて、Lagrange 乗数  $P_k(t)$  を用いて、以下のような誤差関数  $E(W)$  を定義する。

$$E(W) = \int_{T_1}^{T_2} dt \left[ \frac{1}{2} \sum_{k \in V} J_k^2 - \sum_{k \in V} P_k \{ \dot{Q}_k + Q_k - G(u_k) - J_k \} - \sum_{k \notin V} P_k \{ \dot{z}_k + z_k - G(u_k) \} \right]. \quad (\text{C.6})$$

微小変化  $\delta W$  による変分  $\delta E(W)$  は、

$$\begin{aligned} \delta E(W) &= \sum_{k \in V} \frac{\partial E}{\partial J_k} \delta J_k + \sum_{k \notin V} \frac{\partial E}{\partial z_k} \delta z_k + \sum_k \frac{\partial E}{\partial P_k} \delta P_k + \sum_{kj} \frac{\partial E}{\partial W_{kj}} \delta W_{kj} \\ &= \int_{T_1}^{T_2} dt \left[ \sum_{k \in V} (J_k + P_k) \delta J_k \right. \\ &\quad - \sum_{k \notin V} \{ P_k (\delta \dot{z}_k + \delta z_k) - \delta z_k \sum_j P_j G'(u_j) W_{jk} \} \\ &\quad + \sum_{k \in V} \delta P_k \{ \dot{Q}_k + Q_k - G(u_k) - J_k \} + \sum_{k \notin V} \delta P_k \{ \dot{z}_k + z_k - G(u_k) \} \\ &\quad \left. + \sum_{kj} P_k G'(u_k) z_j \delta W_{kj} \right]. \end{aligned} \quad (\text{C.7})$$

となる。ここで、 $\delta W$  による各変数の変分は以下のものである。

$$\delta J_k(t | W) = J_k(t | W + \delta W) - J_k(t | W) \quad (k \in V), \quad (\text{C.8})$$

$$\delta z_k(t | W) = z_k(t | W + \delta W) - z_k(t | W) \quad (k \notin V), \quad (\text{C.9})$$

$$\delta P_k(t | W) = P_k(t | W + \delta W) - P_k(t | W). \quad (\text{C.10})$$

式 (C.1) と式 (C.4) によって、式 (C.7) の第3項は消去される。また、Lagrange 乗数  $P_k(t)$  を以下のように定義すれば、第1項も消去される。

$$\begin{aligned} P_k &= -J_k & (k \in V), \\ \dot{P}_k &= P_k - \sum_j P_j G'(u_j) W_{jk} & (k \notin V). \end{aligned} \quad (\text{C.11})$$

さらに、時刻  $T_1$  と  $T_2$  の境界条件によって、以下のように第2項も消去される。

$$\begin{aligned} [\text{the second term}] &= - \int_{T_1}^{T_2} dt \sum_{k \notin V} (P_k \delta \dot{z}_k + \dot{P}_k \delta z_k) \\ &= - \sum_{k \notin V} [P_k \delta z_k]_{T_1}^{T_2} \\ &= \sum_{k \notin V} \{ P_k(T_1 | W) \delta z_k(T_1 | W) - P_k(T_2 | W) \delta z_k(T_2 | W) \} = 0. \end{aligned} \quad (\text{C.12})$$

従って、第4項だけが残る結果、以下の学習方程式が得られる。

$$\delta E(W) = \int_{T_1}^{T_2} dt \sum_{kj} P_k G'(u_k) z_j \delta W_{kj}, \quad (\text{C.13})$$

$$\Delta W_{kj} = -\eta \frac{\partial E}{\partial W_{kj}} = -\eta \int_{T_1}^{T_2} dt [P_k G'(u_k) z_j]. \quad (\text{C.14})$$

この段階で、変数  $z_k(t)$  を元の変数  $x_i(t)$  と  $y_i(t)$  に戻して学習方程式を書き直そう。隠れ素子  $y_i$  に対応する  $\dot{P}_i(t)$  に関する式 (C.11) の時間逆方向の処理をなくせば、時間方向に後戻りを必要としない効率的な学習処理が期待出来る。一方、発振ニューラルネットワークのネットワーク構造の単純さから、結合重みパラメータ  $K_{IE}^{(i)}$  の値を固定すれば、これに対応する式 (C.14) の処理が不要になるので、時間逆方向の処理をなくすことが出来る。但し、結合重みパラメータ値の固定は、ネットワークの情報処理能力の (パラメータ空間全体内での) 部分的限定を意味することに注意しなければならない。このようにして、可視ユニットへ結合する  $W_{ij}$  と  $K_{EI}^{(i)}$  の結合重みパラメータに関する、以下の学習方程式が得られた。

$$P_i(t) = -J_i(t), \quad (\text{C.15})$$

$$\Delta W_{ij} = -\eta \int_{T_1}^{T_2} [ P_i(t) G'(v_e^{(i)}(t)) x_j(t) ] dt, \quad (\text{C.16})$$

$$\Delta K_{EI}^{(i)} = \eta \int_{T_1}^{T_2} [ P_i(t) G'(v_e^{(i)}(t)) y_i(t) ] dt. \quad (\text{C.17})$$

## 付録 D

### 付録： 顔パターンの特徴要素

4.3 節の発振ニューラルネットワークによる顔パターンの生成で用いた、ネットワークの次元数  $N$  に対応する 22 個の特徴要素について説明する。なるべく少ない特徴要素数で、バラエティーに富んだ顔パターンを生成出来るように、以下のようにして特徴要素を選んだ。

人の顔は大きく分けると、顔の輪郭、口、鼻、目と眉毛の 4 つの部分から成り立っている。このうち、目や眉毛に関する特徴要素（例えば、目の大きさや傾き等）は、（正面の）顔の差異に対する高い視覚感度を持つが、鼻に関する特徴要素の視覚感度は低い事が実験的に明らかにされている（原, 1978）。これより、我々は目や眉毛に関する特徴要素を多くした一方、鼻は単なる縦棒 1 本で省略表現した。また、顔の各特徴要素の標準的な配置関係を、図 D.1 に示すような人間の頭部のカノン（Parramon, 1980）に従って定めた。この配置関係は、成人なら年齢や性別にかかわらずあてはまるものである。さらに、人の顔においては、表情の全くない真顔における各部分特徴より、むしろ表情によってその人らしさが顕著に表われることが多い。表情は顔の筋肉に基づいて作られ、表情筋によって左右の目などの要素間の距離は変化すると考えられる（千葉, 1993）。そこで、顔の各部に関する特徴要素のみならず、このような要素間の距離を表す特徴要素をも選んだ。次節以降で、22 個の各特徴要素の説明をより詳しく行う。なお、各特徴要素の変数値の範囲は  $-1.0 \leq x_i \leq 1.0$  ( $i = 1 \sim N$ ) である。

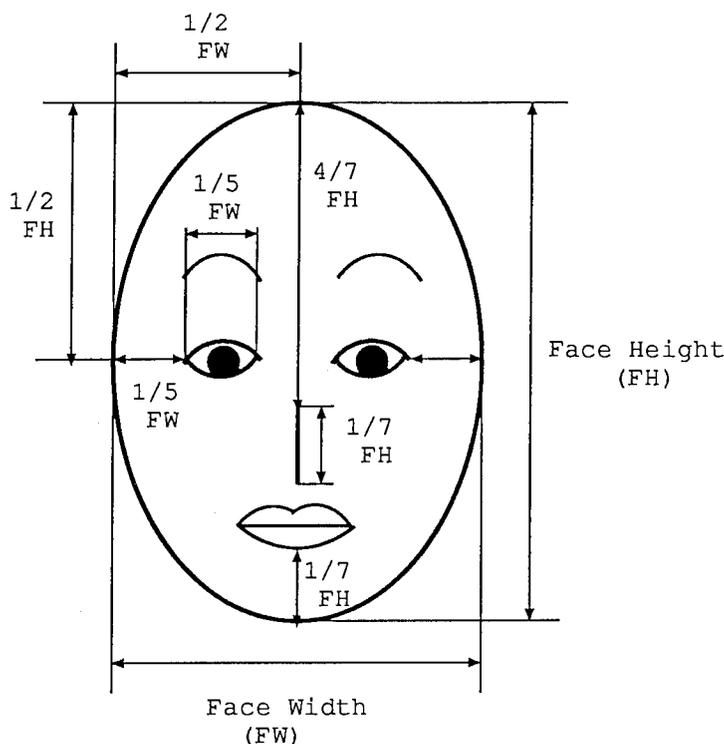


図 D.1: 顔の各部分の標準的な配置関係

### D.0.1 顔の輪郭形状

顔の輪郭形状は以下の *face\_form1* と *face\_form2* の 2 つの変数値 (変数名の左の番号 1. や 2. は、ペア素子  $x_i-y_i$  の添え字  $i$  に対応する) で決められ、丸っこい顔、卵型の顔、四角ばった顔のいずれかとなる。これらの輪郭形状は人相学における三型質に相当する (小粥, 1985)。ここで、1 つの特徴要素に 2 つの変数が当てはめられているのは、中間的な状態 (この場合では、卵型の楕円顔) を安定に生成させる為である (他の 1 特徴 2 変数表現の場合も同様)。なぜなら、我々のモデルでは、入力バイアス制御によって、 $N$  次元超立方体の頂点付近 ( $x_i \approx \pm 1$ ) にはリミットサイクルや漸近安定平衡点を持たせる事が出来るが、原点付近 ( $x_i \approx 0$ ) には記憶パターン間の遊走軌道しか持たせられず、1 変数では中間状態を安定に生成出来ないからである。

1. *face\_form1*:  $-1.0 \sim 1.0$  の変数値に比例して、 $face\_form2 \geq 0$  の時は円形の丸顔から卵型の楕円顔に、 $face\_form2 < 0$  の時は卵型の楕円顔から長方形の四角顔に変化させる。

2. *face\_form2*: 値の正負に従って、丸っこい顔と四角ばった顔にスイッチする為のフラグ。

具体的には、以下の定義式に従って、微小角  $\Delta\theta$  ごとの代表点  $(x(\theta_n), y(\theta_n))$  の折れ線補間で輪郭形状を作成した。ここで、 $NW$  は基準楕円の横幅で、*ElipsRate*, *CircleRate*, *RectangleRate*, *BrendRate* はそれぞれ、基準楕円の縦幅、丸顔の基準円、四角顔の長方形、楕円顔における基準円と基準楕円の混合比に関する係数である。これら 2 変数の値で定まる顔の輪郭形状によって、顔の幅  $FW$  と高さ  $FH$  が決まり、図 D.1 に従って各部分の標準的な配置が決まる。

( $face\_form2 \geq 0.0$ ) の場合、

$$\begin{aligned}
 \theta_n &= \theta_{n-1} + \Delta\theta, \\
 norm\_x(\theta_n) &= NW \times \cos(\theta_n), \\
 norm\_y(\theta_n) &= ElipsRate \times NW \times \sin(\theta_n), \\
 circle\_x(\theta_n) &= CircleRate \times NW \times \cos(\theta_n), \\
 circle\_y(\theta_n) &= CircleRate \times NW \times \sin(\theta_n), \\
 \alpha &= \frac{(1.0 + face\_form1)}{2} \times BrendRate, \\
 \beta &= 1.0 - \alpha, \\
 x(\theta_n) &= \alpha \times circle\_x(\theta_n) + \beta \times norm\_x(\theta_n), \\
 y(\theta_n) &= \alpha \times circle\_y(\theta_n) + \beta \times norm\_y(\theta_n).
 \end{aligned} \tag{D.1}$$

( $face\_form2 < 0.0$ ) の場合 (番号 1. ~ 4. は長方形の 4 辺に対応) 、

$$\begin{aligned}
 rect\_x(\theta_n) &= \begin{cases} 1. (\frac{\pi}{3.0} \leq \theta_n < \frac{2\pi}{3.0}) \text{ の場合} & RectangleRate \times \sqrt{3.0} \times \frac{NW}{\tan(\theta_n)} \\ 2. (\frac{2\pi}{3.0} \leq \theta_n < \frac{4\pi}{3.0}) \text{ の場合} & (-RectangleRate) \times NW \\ 3. (\frac{4\pi}{3.0} \leq \theta_n < \frac{5\pi}{3.0}) \text{ の場合} & (-RectangleRate) \times \sqrt{3.0} \times \frac{NW}{\tan(\theta_n)} \\ 4. \text{ その他の場合} & RectangleRate \times NW \end{cases}, \\
 rect\_y(\theta_n) &= \begin{cases} 1. (\frac{\pi}{3.0} \leq \theta_n < \frac{2\pi}{3.0}) \text{ の場合} & RectangleRate \times NW \times \sqrt{3.0} \\ 2. (\frac{2\pi}{3.0} \leq \theta_n < \frac{4\pi}{3.0}) \text{ の場合} & (-RectangleRate) \times NW \times \tan(\theta_n) \\ 3. (\frac{4\pi}{3.0} \leq \theta_n < \frac{5\pi}{3.0}) \text{ の場合} & (-RectangleRate) \times \sqrt{3.0} \times NW \\ 4. \text{ その他の場合} & RectangleRate \times NW \times \tan(\theta_n) \end{cases}, \\
 x(\theta_n) &= \alpha \times rect\_x(\theta_n) + \beta \times norm\_x(\theta_n), \\
 y(\theta_n) &= \alpha \times rect\_y(\theta_n) + \beta \times norm\_y(\theta_n).
 \end{aligned} \tag{D.2}$$

## D.0.2 顎の輪郭形状

顎の輪郭形状は、以下の  $chin\_form1$  と  $chin\_form2$  の 2 つの特徴要素変数で決められる。

3.  $chin\_form1$ :  $-1.0 \sim 1.0$  の変数値に比例して、 $chin\_form2 \geq 0$  の時は尖った顎から丸みを帯びた顎に、 $chin\_form2 < 0$  の時は丸みを帯びた顎から四角ばった顎に変化させる。
4.  $chin\_form2$ : 値の正負に従って、尖った顎と四角ばった顎にスイッチする為のフラグ。

具体的には、顔の輪郭形状 ( $face\_x, face\_y$ ) を基準にして、以下の定義式に従って顎の輪郭形状を作成した (但し、ダッシュは輪郭形状との区別の為の表記であって、微分記号ではない)。ここで、 $BaseRate, TriangleRate, RectangleRate$  はそれぞれ、基準とする顔の輪郭形状、尖った顎、四角ばった顎に関する係数である。

( $chin\_form2 \geq 0.0$ ) の場合、

$$\begin{aligned}
 \alpha' &= \frac{(1.0 + chin\_form1)}{2} \times BrendRate, \\
 \beta' &= 1.0 - \alpha', \\
 chin\_x(\theta_n) &= face\_x(\theta_n) \times (BaseRate - TriangleRate \times |\sin(2.0 \times \theta_n)|), \\
 chin\_y(\theta_n) &= face\_y(\theta_n) \times (BaseRate - TriangleRate \times |\sin(2.0 \times \theta_n)|), \\
 x'(\theta_n) &= \alpha' \times chin\_x(\theta_n) + \beta' \times face\_x(\theta_n), \\
 y'(\theta_n) &= \alpha' \times chin\_y(\theta_n) + \beta' \times face\_y(\theta_n).
 \end{aligned}
 \tag{D.3}$$

( $chin\_form2 < 0.0$ ) の場合、

$$\begin{aligned}
 \alpha' &= \frac{(1.0 + chin\_form1)}{2} \times BrendRate, \\
 \beta' &= 1.0 - \alpha', \\
 chin\_x(\theta_n) &= face\_x(\theta_n) \times (BaseRate + RectangleRate \times |\sin(2.0 \times \theta_n)|), \\
 chin\_y(\theta_n) &= face\_y(\theta_n) \times (BaseRate + RectangleRate \times |\sin(2.0 \times \theta_n)|), \\
 x'(\theta_n) &= \alpha' \times chin\_x(\theta_n) + \beta' \times face\_x(\theta_n), \\
 y'(\theta_n) &= \alpha' \times chin\_y(\theta_n) + \beta' \times face\_y(\theta_n).
 \end{aligned}
 \tag{D.4}$$

### D.0.3 目の特徴要素

目の特徴要素として、目の開閉度、目の傾き、目の形状、瞳の位置を考え、以下のような特徴変数でこれらを表現した (図 D.2参照)。

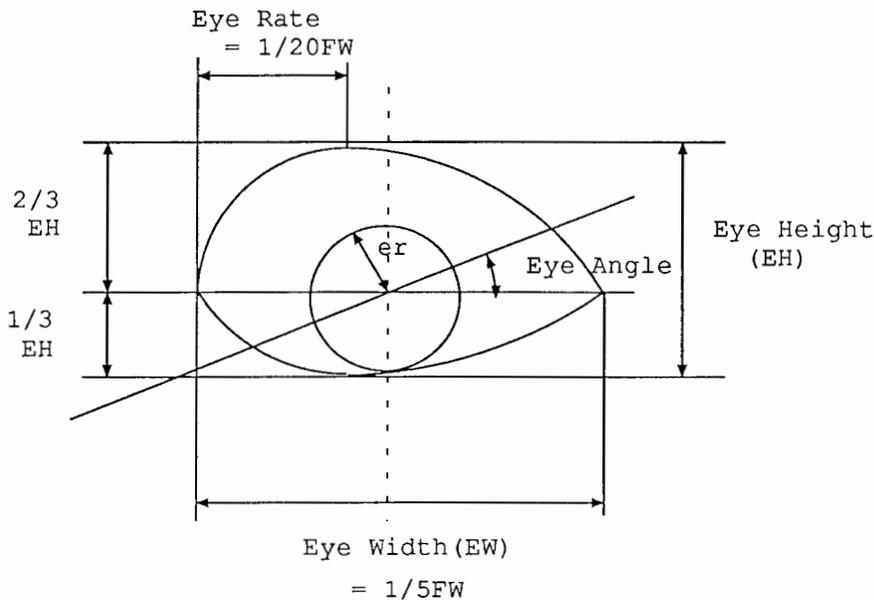


図 D.2: 目の構成

5. *eye\_open1*:  $-1.0 \sim 1.0$  の変数値に比例して、 $eye\_open2 \geq 0$  の時は標準的な目の大きさ ( $EH = \frac{3FH}{300}$ ) から十分開いた目 ( $EH = \frac{3FH}{50}$ ) に、 $eye\_open2 < 0$  の時は閉じた目 ( $EH = 0$ ) から標準的な目の大きさに変化させる。
6. *eye\_open2*: 値の正負に従って、開いた大きな目と閉じた細い目にスイッチする為のフラグ。
7. *eye\_angle1*:  $-1.0 \sim 1.0$  の変数値に比例して、目の傾きを  $eye\_angle2 \geq 0$  の時は水平から上がり目に、 $eye\_angle2 < 0$  の時は下がり目から水平に変化させる。目の傾き  $\theta \equiv \frac{\theta_1 - \theta_2}{2}$  は、 $\theta_1 = -22.5 \times (eye\_angle1 - 1.0)$ ,  $\theta_2 = -22.5 \times (eye\_angle2 - 1.0)$  より、 $-22.5^\circ \leq \theta \leq 22.5^\circ$  となる。
8. *eye\_angle2*: 値の正負に従って、上がり目と下がり目にスイッチする為のフラグ。
9. *eye\_rate*:  $-1.0 \sim 1.0$  の変数値に比例して、目の形状を表す曲線の変曲点の位置  $cx$  を、 $cx = \frac{EW}{2} \times eye\_rate \times \frac{FW}{20}$  によって変化させる。
10. *eye\_pos*:  $-1.0 \sim 1.0$  の変数値に比例して、瞳の位置を上下に  $cy = \frac{er}{2} \times eye\_rate$  ( $er \equiv \frac{3FW}{40}$ ) に従って変化させる。

瞳の位置の上中下はそれぞれ、人相学における、下三白眼（黒目のまわり左右と下の三方が白目の場合）、四白眼（黒目のまわりの四方が白目の場合）、上三白眼（黒目のまわり左右と上の三方が白目の場合）に相当するものである（小粥, 1985）。

#### D.0.4 眉毛の特徴要素

眉毛の特徴要素として、眉毛の太さ、眉毛の高さ、眉毛の形状、眉毛の傾きを考え、以下のような特徴変数でこれらを表現した（図 D.3参照）。

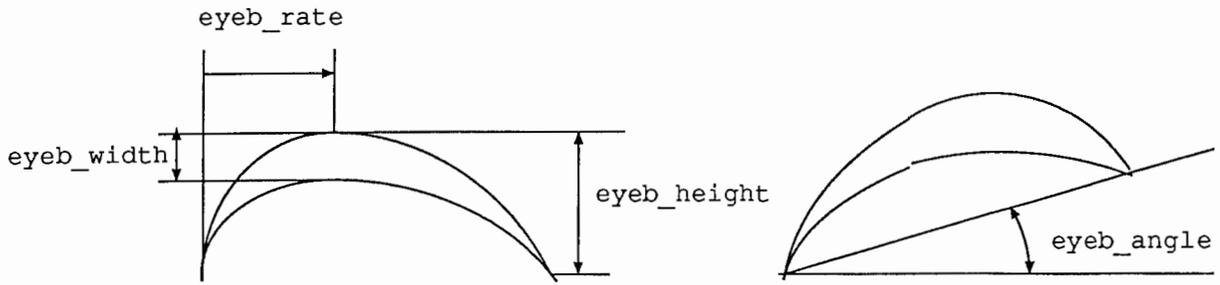


図 D.3: 眉毛の構成

12. `eyeb_width`:  $-1.0 \sim 1.0$  の変数値に比例して、眉毛を太く変化させる ( $2dots \sim 8dots$ )。
13. `eyeb_height`:  $-1.0 \sim 1.0$  の変数値に比例して、眉毛を高く湾曲させる ( $0 \sim \frac{FH}{20}$ )。
14. `eyeb_rate`: 目の特徴要素 `eye_rate` と全く同様。
15. `eyeb_angle1`: 目の特徴要素 `eye_angle1` と全く同様。
16. `eyeb_angle2`: 目の特徴要素 `eye_angle2` と全く同様。

### D.0.5 唇の特徴要素

唇の特徴要素として、唇の横幅、上唇の高さ、下唇の高さを考え、以下のような特徴変数でこれらを表した (図 D.4参照)。

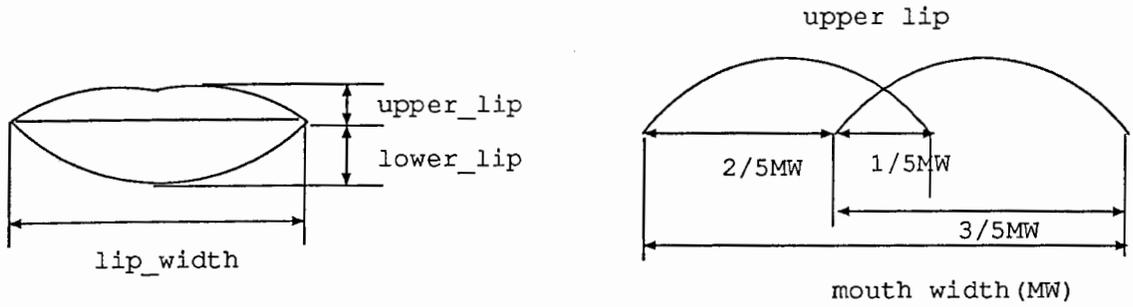


図 D.4: 唇の構成

19. lip\_w:  $-1.0 \sim 1.0$  の変数値に比例して、唇の横幅  $MW$  を変化させる ( $\frac{FW}{10} \sim \frac{2FW}{5}$ )。
20. upper\_lip:  $-1.0 \sim 1.0$  の変数値に比例して、上唇の高さを変化させる ( $1dots \sim \frac{FH}{28}$ )。
21. lower\_lip:  $-1.0 \sim 1.0$  の変数値に比例して、下唇の高さを変化させる ( $1dots \sim \frac{FH}{28}$ )。

## D.0.6 顔の各要素の配置

顔の各部の、目と目、眉毛と眉毛、目と眉毛、唇と顎の相対的な位置について、以下のような変数を考えた (図 D.5 参照)。これらの変数値に従って、各部の間の距離を図 D.1 に示された標準的な配置関係 (変数値が零の時) より、近づけたり遠ざけたりして変動させる。

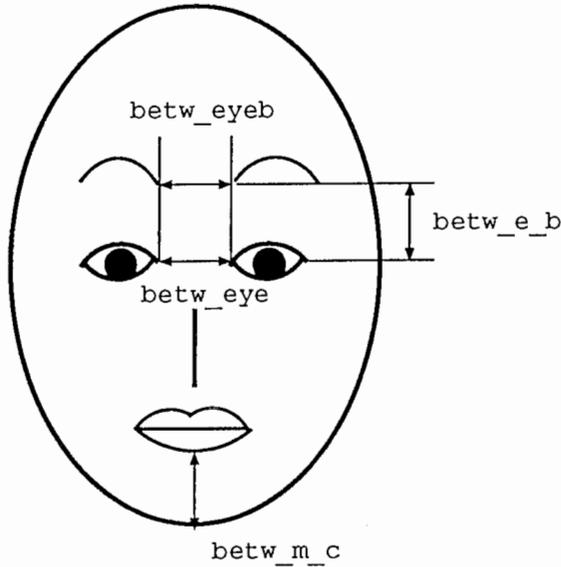


図 D.5: 顔の各要素間の距離

11. **betw\_eyes:**  $-1.0 \sim 1.0$  の変数値に比例して、左右の目の間の距離を標準より近く (最小  $\frac{FW}{10}$ ) から遠く (最大  $\frac{3FW}{10}$ ) に変化させる。
17. **betw\_eyeb:**  $-1.0 \sim 1.0$  の変数値に比例して、左右の眉毛の間の距離を標準より近く (最小  $\frac{FW}{10}$ ) から遠く (最大  $\frac{3FW}{10}$ ) に変化させる。
18. **betw\_e\_b:**  $-1.0 \sim 1.0$  の変数値に比例して、目と眉毛の間の距離を標準より近く (最小  $\frac{FH}{14}$ ) から遠く (最大  $\frac{3FH}{14}$ ) に変化させる。
22. **betw\_m\_c:**  $-1.0 \sim 1.0$  の変数値に比例して、顎の先と下唇の下限の間の距離を標準より近く (最小  $\frac{FH}{14}$ ) から遠く (最大  $\frac{3FH}{14}$ ) に変化させる。

## 付録 E

### 付録：漸近安定平衡点とリミットサイクルの共存条件

4.4節で議論した自律系の連想記憶モデルが、記憶パターン付近の漸近安定平衡点と原点周辺のカオスの軌道を持つには、モデルの構成要素である各興奮 - 抑制性ペア素子自身が、 $x_i \approx \pm 1$  に漸近安定平衡点を持つとともに、 $x_i$ - $y_i$  座標の原点周辺にリミットサイクルを持つ事が重要である。

以下では、単一の興奮 - 抑制性ペア素子において、 $x_i \approx \pm 1$  の2つの漸近安定平衡点と原点周辺のリミットサイクルが共存する為の、パラメータ  $\delta$  と  $K_{max}$  に関する必要条件を示す。

[自律系の連想記憶モデル]

$$\dot{x}_i = -x_i + G\left(\sum_{j=1}^N W_{ij}x_j - K_{EI}y_i\right). \quad (\text{E.1})$$

$$\dot{y}_i = -y_i + G(K_{IE}(x_i)x_i). \quad (\text{E.2})$$

さて、 $x_i = \pm 1$  と  $x_i = 0$  において、関数  $K_{IE}(x_i) \equiv K_{max} \exp^{-\delta x_i^2}$  の値は、最小値  $K_{max} \exp^{-\delta}$  と最大値  $K_{max}$  とにそれぞれなる。平衡点の近傍のみを議論の対象とするので、これら2つの場合に相当した固定結合パラメータ値  $K_{IE} = K_{max} \exp^{-\delta}$  or  $K_{max}$  を持つ以下のような興奮 - 抑制性ペア素子を考える（単一の興奮 - 抑制性ペア素子なので添え字  $i$  を省く）。

$$\dot{x} = -x + G(W_{ii}x - K_{EI}y), \quad (\text{E.3})$$

$$\dot{y} = -y + G(K_{IE}x). \quad (\text{E.4})$$

ここで、Nullclineの2曲線  $\dot{x} = 0$  と  $\dot{y} = 0$  とを、 $y = f_1(x) \equiv [W_{ii}x - G^{-1}(x)]/K_{EI}$  と、 $y = f_2(x) \equiv G(K_{IE}x)$  とでそれぞれ表記する。図 E.1を参考に、これら2曲線の交点である平衡点の存在性や安定性を議論する。以下の議論では、 $y = f_1(x)$  が変曲点を持つことを仮定しているので、3章の定理2の証明におけるその必要条件から、 $W_{ii}(= M/N) \geq \pi a/2$  とする。もちろん、 $y = f_1(x)$  が変曲点を持たない場合は、明らかに原点のみが不安定平衡点となる。

まず、 $K_{IE} \equiv K_{max} \exp^{-\delta}$  の場合について、 $x \approx \pm 1$  に漸近安定平衡点を持つ為の1つの条件を考える。さて、曲線  $y = f_1(x)$  の原点における接線の傾きが曲線  $y = f_2(x)$  のそれより大きい、すなわち、 $f'_1(0) > f'_2(0)$  ( $f'(x) \equiv \partial f(x) / \partial x$ ) でなければ、明らかに  $x \approx \pm 1$  に平衡点そのものが存在出来ない。

この条件は  $f_1(x)$  と  $f_2(x)$  の定義式から、

$$K_{IE} < [\pi a(2W_{ii} - \pi a)] / 4K_{EI}. \tag{E.5}$$

となる。

式 (E.5) に  $K_{IE} = K_{max} \exp^{-\delta}$  を代入し、 $\delta$  について整理すると以下の条件が得られる。

$$\delta > \log(4K_{EI}K_{max} / [\pi a(2W_{ii} - \pi a)]). \tag{E.6}$$

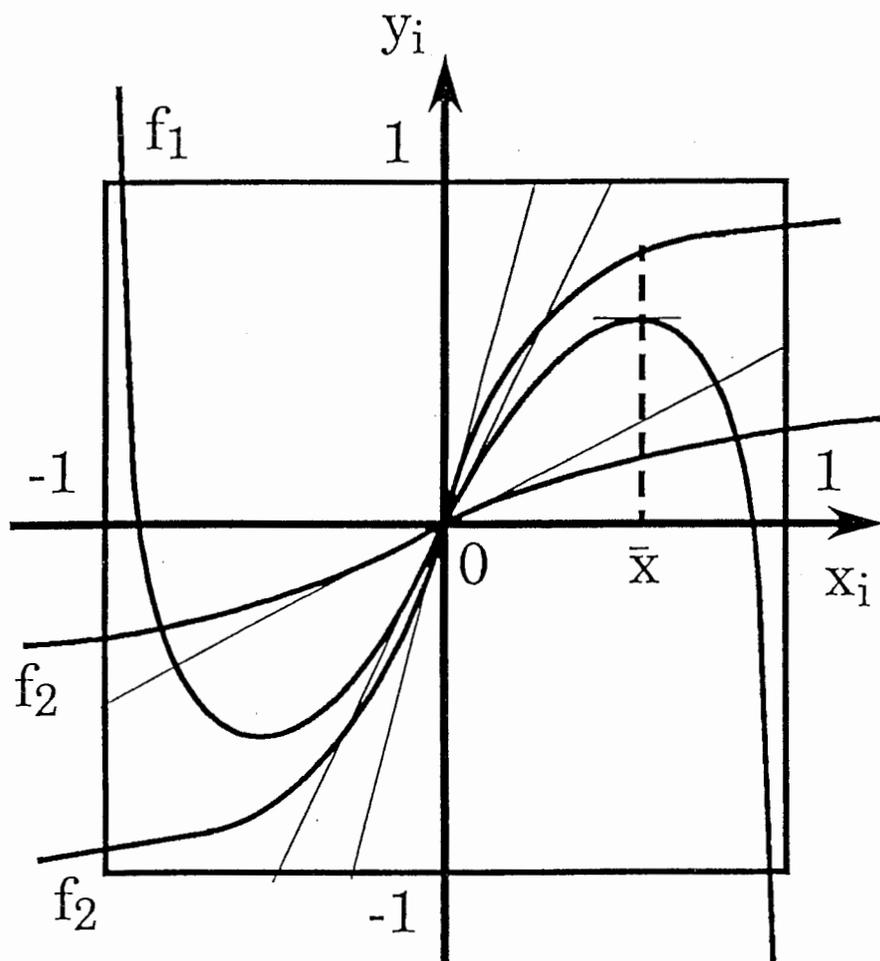


図 E.1: 2 曲線  $y = f_1(x)$  と  $y = f_2(x)$  (関数  $K_{IE}(x)$  の最大値と最小値の 2 つの場合は、 $f_2(x)$  で原点の傾きが大きいものと小さいものにそれぞれ対応する)。

次に、 $K_{IE} \equiv K_{max}$  の場合について、原点周辺にリミットサイクルが存在する為の必要条件を考える。もし原点が唯一の不安定平衡点だとしたら、ポアンカレ-ベンディクソンの定理よりリミットサイクルの存在性が保証される (Hirsch, & Smale, 1974; Wiggins, 1990)。一方、原点のみが平衡点となる為には、原点における接線の傾きが  $f_1'(0) < f_2'(0)$  となり、 $y = f_1(x)$  の変曲点  $\bar{x}$  において  $f_1(\bar{x}) < f_2(\bar{x})$  となる事が必要である。ここで、変曲点の座標値は  $f_1'(\bar{x}) = 0$  から解析的に求められて、 $\bar{x} = 2 \arccos(\sqrt{\pi a / 2W_{ii}}) / \pi$  となる。但し、 $\bar{x} > 0$  を仮定しているが、奇関数  $f_1(x), f_2(x)$  の原点对称性から、 $\bar{x} < 0$  の場合も以下と同様な議論となる。

$K_{IE} = K_{max}$  から条件  $f_1'(0) < f_2'(0)$  は、

$$K_{max} > [\pi a(2W_{ii} - \pi a)] / 4K_{EI}. \quad (E.7)$$

となる。また、有界性  $|f_2(x)| < 1$  から  $f_1(\bar{x}) < f_2(\bar{x}) < 1$  となり、関数  $f_1(x)$  の定義を用いて、

$$K_{max} > G^{-1}(f_1(\bar{x})) / \bar{x}. \quad (E.8)$$

が得られる。一方、 $f_1(\bar{x}) < 1$  から、パラメータ  $a, W_{ii}, K_{EI}$  に関する以下の条件も得られる。

$$(2W_{ii} \arccos \sqrt{\pi a / 2W_{ii}}) / \pi + a \sqrt{(2W_{ii} - \pi a) / \pi a} < K_{EI}. \quad (E.9)$$

さて、原点を不安定平衡点と仮定すると、原点における標準形解析 (Guckenheimer, & Holmes, 1983; Wiggins, 1990) から極座標表示で、

$$\dot{\theta} = \Omega + br^2 + O(r^5), \quad (E.10)$$

$$\dot{r} = \mu r + ar^3 + O(r^5). \quad (E.11)$$

が得られる。ここで、 $\mu \equiv -1 + W_{ii} / \pi a$ 、 $\Omega \equiv \sqrt{4K_{IE}K_{EI} - W_{ii}^2} / \pi a$  であり、係数  $a, b$  は関数  $G(x)$  の非線形性に依存して決まる定数である。原点の不安定性が成り立つのは、 $\Omega$  が実数である場合に限られるので、

$$K_{max} = K_{IE} > W_{ii}^2 / 4K_{EI}. \quad (E.12)$$

でなければならない。

以上、式 (E.7)、式 (E.8)、式 (E.12) を総合して、

$$K_{max} > \max \{ G^{-1}(f_1(\bar{x})) / \bar{x}, [\pi a(2W_{ii} - \pi a)] / 4K_{EI}, W_{ii}^2 / 4K_{EI} \}. \quad (E.13)$$

が得られる。

上記の条件 (E.6)、(E.9)、(E.13) を満足する例として、 $K_{max} = K_{EI} = 2.0$ ,  $a = 0.1$ ,  $\delta = 6.0$ ,  $W_{ii} = 1.0$  がある。これらのパラメータ値を持った単一の興奮-抑制性ペア素子 (関数型結合重み  $K_{IE}(x_i)$ ) における、Nullcline の 2 曲線  $\dot{x}_i = 0$ ,  $\dot{y}_i = 0$  とそれらの交点である平衡点 (矢印で平衡点の安定性を表示) を図 4.12(a) に、また、 $x_i$ - $y_i$  平面上の多数の初期状態からの軌跡を図 4.12(b) に示しているので、4.4 節に戻ってこれらを参照されたい。図 4.12 から、原点の不安定平衡点と 2 組の安定点-サドル点を持つ事が本質的に重要であることが伺える。但し、これらの条件は十分条件ではないので、関数  $K_{IE}(x)$  の最小値  $K_{max} \exp^{-\delta}$  を大きく設定するに従って、2 組の安定点-サドル点は段々近づいていき、ついには生成-消滅タイプの分岐点で重なって対消滅し (Thompson, & Stewart, 1986; Wiggins, 1990; 合原, 1990b)、リミットサイクルのみが残ることに注意しよう。

## 付録 F

### 付録： 1 次 Hopf 分岐点の近似解析

5章の 5.2.2における 2 組のペア素子についての数値解析結果では、ネットワークの挙動が漸近安  
平衡点への収束から片方のペア素子の Hopf 分岐による発振へ変化する 1 次 Hopf 分岐点が、入  
力バイアスと結合重み係数に関する 3 次元のパラメータ空間の平面（2 次元断面における直線）  
上にほぼ載っていた。以下では、この直線方程式を近似解析によって示す。

まず、3章の 3.3.2で示したように、片方のペア素子のみが Hopf 分岐を起こす時、その平衡  
点の座標値は一定となることに注意しよう。すなわち、もう一方のペア素子からの入力を一定値  
 $I$  または  $J$  とした時（但し、どちらか一方はゼロ）、単一のペア素子の平衡点  $(\bar{x}, \bar{y})$  における線  
形化方程式は、

$$\begin{pmatrix} \delta \dot{x} \\ \delta \dot{y} \end{pmatrix} = \begin{pmatrix} -1 + W_{ii}G'(v_e) & -K_{EI}G'(v_e) \\ K_{IE}G'(v_i) & -1 \end{pmatrix} \begin{pmatrix} \delta x \\ \delta y \end{pmatrix}. \quad (\text{F.1})$$

となり、Hopf 分岐点  $(x^*, y^*)$  では、その Jacobi 行列の固有値  $\lambda = [(Tr) \pm \sqrt{(Tr)^2 - 4(Det)}] / 2$   
が純虚数： $(Tr) = 0$ ,  $(Det) > 0$  となることより（Guckenheimer, & Holmes, 1993）、

$$x^* = \pm \frac{2}{\pi} \arctan \sqrt{\frac{W_{ii} - \pi a}{\pi a}} = const. \quad (\text{F.2})$$

であることが定理 2 の証明で示された。ここで、 $v_e \equiv W_{ii}\bar{x} - K_{EI}\bar{y} + I$ ,  $v_i \equiv K_{IE}\bar{x} + J$  で、 $(Tr)$   
と  $(Det)$  は上記の Jacobi 行列のトレースと行列式である。また、 $G'(z)$  は関数  $G(z) = \frac{2}{\pi} \arctan(\frac{z}{a})$   
の導関数である（以下の  $G^{-1}(x)$  は  $G(z)$  の逆関数）。

この時、

$$y^* = G(K_{IE}x^* + J) = const. \quad (\text{F.3})$$

である。また、 $y^* = G(K_{IE}x^* + J) = \{W_{ii}x_i^* - G^{-1}(x^*) + I^*\} / K_{EI}$  より、

$$I^* = -W_{ii}x^* + G^{-1}(x^*) + K_{EI}y^* = const. \quad (\text{F.4})$$

でもある。

	$\bar{x}_1$	$\bar{y}_1$	$\bar{x}_2(=x^*)$	$\bar{y}_2(=y^*)$
Case 1	正	正	正	正
Case 2	正	正	負	負
Case 3	正	正	正	負
Case 4	正	正	負	正

表 F.1:  $I_1 > I_2 > 0$  の時の平衡点の座標値の符号

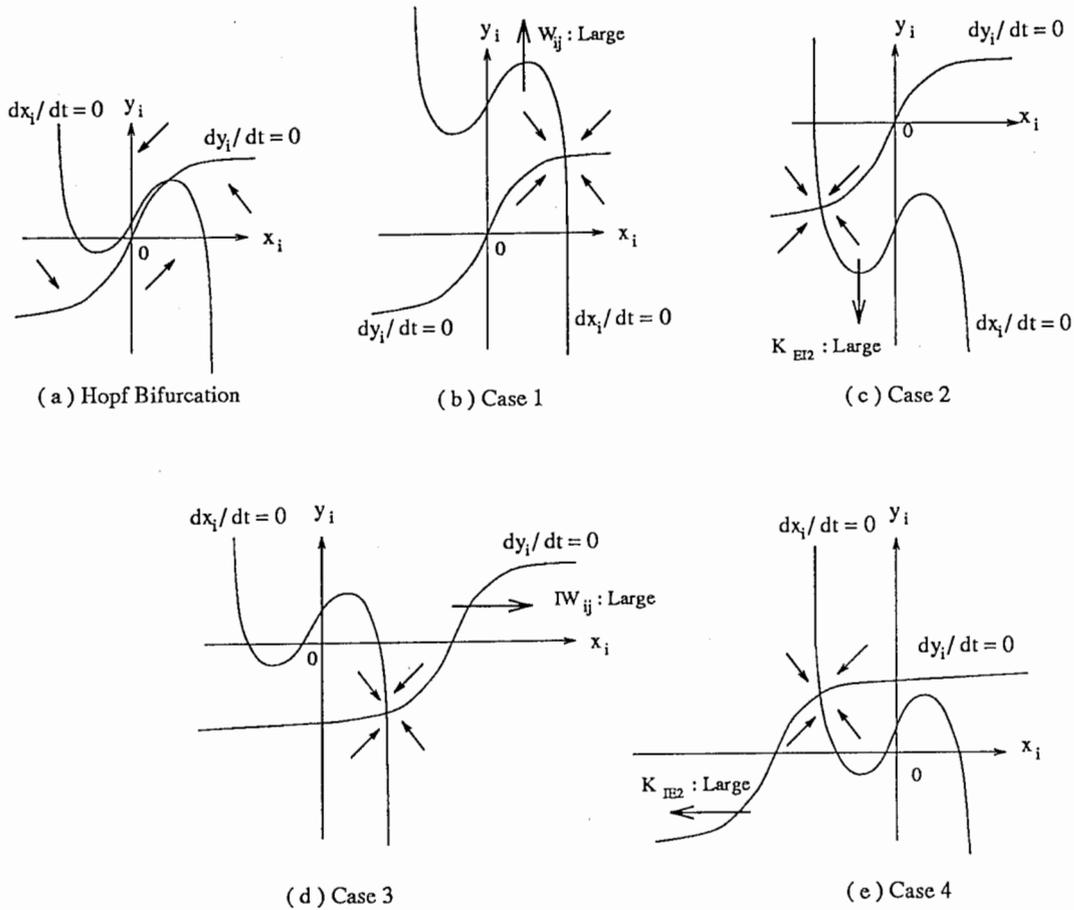


図 F.1:  $I_1 > I_2 > 0$  の時の結合重みパラメータ値の増加に対する Nullcline の変化の方向 (a) 分岐後の不安定平衡点の状態, (b)~(d) 分岐前の漸近安定平衡点の状態

5章の 5.2.2 では、 $I_1 > 0$ ,  $I_1 > I_2$  が仮定されていた。以下では、 $I_1 > I_2 > 0$  の場合のみを考えるが、他の場合もペア素子の入れ換え等による対称性から同様の議論となる。この時、入力バイアス値の絶対値が大きい方がより安定となるので、結合重みパラメータ値の増加に伴って  $x_2$ - $y_2$  が先に Hopf 分岐を起こす。各ペア素子の (分岐前の) 平衡点の座標値の符号を表 F.1 に、各結合形態に対する結合重みパラメータ値の増加に従った、 $x_2$ - $y_2$  平面上の Nullcline  $\dot{x}_2 = 0$  :  $y_2 = \{W_{ii}x_2 - G^{-1}(x_2) + I\}/K_{EI}$ ,  $\dot{y}_2 = 0$  :  $y_2 = G(K_{IE}x_2 + J)$  の変化の方向を図 F.1 に示す。

分岐前では図 F.1(b)~(d) のように、平衡点  $(\bar{x}_1, \bar{y}_1) \neq (0, 0)$  が Nullcline の 2 曲線の飽和部分にできるので、興奮性素子  $x_1$  (あるいは抑制性素子  $y_1$ ) への入力値の変化に対して、その座標値は近似的に一定と考えられる。この近似にもとづいて、それぞれの結合形態の場合に対する 1 次 Hopf 分岐点の直線方程式を導こう。

(Case 1)

式 (F.4) より、Hopf 分岐点では興奮性素子  $x_2$  への正味の入力値が一定値であること、すなわち、 $I_2 + W_{ij}\bar{x}_1 \approx I^* = const$  から、

$$W_{ij} \approx \frac{-I_2 + I^*}{\bar{x}_1} = -aI_2 + b. \quad (\text{F.5})$$

の直線方程式を得る。ここで、 $a \equiv 1/\bar{x}_1 > 0$ ,  $b \equiv I^*/\bar{x}_1 > 0$  は直線の傾きと  $W_{ij}$  軸に関する切片を表す係数である。

(Case 2)

平衡点  $(x^*, y^*)$  の座標値の符号を除いて同様に、 $I_2 - K_{EI2}\bar{y}_1 \approx -I^*$  より、

$$K_{EI2} \approx \frac{I_2 + I^*}{\bar{y}_1} = a'I_2 + b'. \quad (\text{F.6})$$

を得る。ここで、直線方程式の係数を  $a' \equiv 1/\bar{y}_1 > 0$ ,  $b' \equiv I^*/\bar{y}_1 > 0$  とした。

(Case 3)

上記の 2 つの場合とは異なって、他のペア素子から抑制性素子  $y_2$  に入力値が与えられる。しかし、この場合も式 (F.3) から、Hopf 分岐点における平衡点の座標値  $-y^*$  は一定でなければならない。ゆえに、 $K_{IE}x^* - IW_{ij}\bar{y}_1 \approx G^{-1}(-y^*)$  から、

$$IW_{ij} \approx \frac{K_{IE}x^* + G^{-1}(y^*)}{\bar{y}_1} = const. \quad (\text{F.7})$$

となる。

(Case 4)

興奮性素子  $x_1$  から入力を与えられる点を除いて同様に、 $-K_{IE}x^* + K_{IE2}\bar{x}_1 \approx G^{-1}(y^*)$  より、

$$K_{IE2} \approx \frac{G^{-1}(y^*) + K_{IE}x^*}{\bar{x}_1} = const. \quad (\text{F.8})$$

となる。

以上、Hopf 分岐を起こさない方のペア素子の平衡点の座標値  $(\bar{x}_1, \bar{y}_1)$  が一定であるという仮定のもとで、各結合形態に対する 1 次 Hopf 分岐点の直線方程式を近似的に示した。尚、5章の図 5.3において、この分岐点が直線から外れている部分については、上記の粗い近似が成り立たなかったことを示すものと考えられる。