数理解析研究所講究録
第1827巻 2013年 140-153

ニューラルネットワークの入力による分岐構造と学習による変化

東京大学総合文化研究科　栗川知己

Tomoki Kurikawa
Graduate School of Arts and Sciences
University of Tokyo

1．イントロダクション

神経科学の1つの目標として、神経系がどのように外部環境をコーディングしているのか、という問題がある。この問題に対してこれまで膨大な研究がなされてきており、視覚刺激のコーディング原理などが完全とは言えないまでも明らかにされてきた(Ben-Yishai, Bar-Or, & Sompolinsky, 1995; Hubel & Wiesel, 1962)。これらの研究では何らかの外部刺激、例えば視覚刺激などの感覚刺激、あるいは先に提示されたものと同一のものを同定する認知課題などが与えられ、その神経系応答を測定することで、与えられた刺激の処理される部位、あるいはどのような応答パタンを示すかを明らかにしてきた。ここでは、何らかの刺激を与えその応答をみるという一般的な方法が取られてきた。そして、刺激が与えられていない自発的な神経活動は意味のないノイジーな活動だと考えられてきた。

しかし近年の観測精度の向上、多点計測の確立などにより今までノイズだと思われてきた、刺激がない状況での活動が実は時空間パタンをもった振る舞いをすることが明らかになってきた(Kenet, Bibitchkov, Tsodyks, Grinvald, & Arieli, 2003; Luczak, Bartho, & Harris, 2009)。さらにこのような自発的パタンと外部刺激に対する応答性には関連があることも同時に明らかにされつつある

(Worgotter, Suder, & Zhao, 1998)。このように、今までの刺激-応答の枠組みでは無視されてきた自発的な活動が、刺激-応答に関与しているという事がわかってきた。応答は単に刺激によって定まるのではなく、内部の活動と関連しながら形成されていくのである。

では自発活動と誘起活動にはどのような関係があるのだろうか？

自発活動での神経系の発火パタンは、入力を引火した時のパタンと似たパタンを次々と時間的に遷移しながら示すことが観察されている(Kenet, et al., 2003; Luczak, et al., 2009)。また活動の相関距離が大きな自発活動が、入力が印加されることで相関距離が小さくなることも観察されている(Smith & Kohn, 2008)。これらのことから様々なパタンが現れ消えていく自発活動が、入力印加によりその内のパタンの１つにローカライズすると考えられる。このような入力印加による活動の変化は自発活動のダイナミクスに依存していると考えられるが、この自発活動のダイナミクスとその応答の関係はよくわかっていない。

本研究は、入力印加に対する系の応答が高次元の自発活動が分岐していく過程と捉える事で、自発活動のダイナミクスとその応答がどのように関係するのかを理解するための試みである。そのために新しい記憶モデルである memories as bifurcations(第二節)を提案し、それに基づいた学習モデルを構築することで、自発活動とその応答の関係の解析(第三節)を行った。最後にこの結果が示唆する点に関して第四節で議論する。

２．　モデル

2-1 Memories as Bifurcations

本小節では最近、著者らが提案している”memories-as-bifurcations”というあたらしい枠組みを説明する(Kurikawa & Kaneko, 2012)。

本研究の動機として自発活動とその応答の関係を明らかにする、という点がある。しかし、従来のメモリの view point(Hopfield, 1984)では、入力は初期条件として与えられその収束先のアトラクタが入力にたいする出力・応答であり(Memories as Attractors), 入力のない自発活動は初期条件が与えられない状況に対応するため自発活動とその応答の解析が難しい。これまでランダムネット

ワークなどを用いて、ネットワークトポロジー・結合強度分布とネットワークのダイナミクスの関係が明らかになってきた(Amit & Brunel, 1997; van Vreeswijk & Sompolinsky, 1996)が、入力とダイナミクスの関係にはあまり焦点が当てられてこなかった。それに対し、今回提案した Memories as Bifurcations では入力を分岐パラメタ、出力は分岐先のダイナミクスとみなす(但し今回は分かりやすいように分岐先のアトラクタをターゲットとしている)。このような見方をすることにより、自発活動の性質とその分岐過程の解析を通して、自発活動とその応答活動の関係を解析することができる。

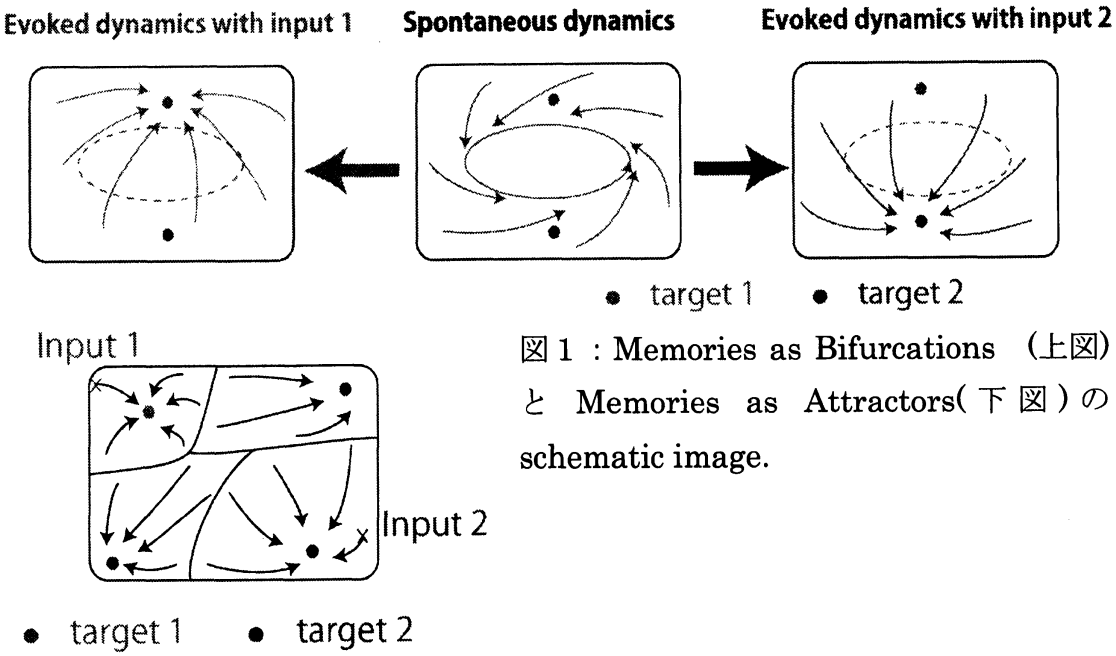

図1：Memories as Bifurcations (上図)と Memories as Attractors(下図)の schematic image.

2-2 Neural Dynamics

本研究では、上の Memories as Bifurcations に基づき Input/Output マッピングの学習を行うモデルを構築し、学習によりどのような自発活動が形成されるか、そしてそれがどのように入力により分岐されるか、自発活動とその応答(分岐)の関係を解析する。ここでは一般的に抽象モデル採用される連続発火率$\{x_i\}$をもつニューロンモデル

$$\dot{x_i} = \tanh\left(\beta\left(\sum_{j\neq i}^{N} J_{ij}x_j + \gamma\eta_i^{\nu}\right)\right) - x_i$$

を採用する。*Jij* は neuron *j* から neuron *i* への結合強度、η は±1をとる2値の *Nbit* 定数で入力パタン、γ は入力強度を示す。以下に述べる learning dynamics により入力パタン η^{μ} を印加することで系が変化し、ターゲット出力パタン ξ^{μ} がアトラクタとして生成されるように学習を行う。ここでターゲット出力パタン ξ も η と同様に±1の2値をとる *N*bit の定数である。以下では *N*=100,β=4 でのシミュレーション結果を示す。

2-3 Synaptic Dynamics (Learning Process)

入力パタン η とターゲット出力パタン ξ を学習するために次のダイナミクスに従い *Jij* を変化させる。

$$\dot{J_{ij}} = \alpha(\xi_i^{\mu} - x_i)x_j,$$

ここで α は学習パラメタと呼ばれるシナプス変化の時間変化の逆数である。この学習則はパーセプトロン学習に類似した形をしており、シナプス強度の微小変化 ΔJij は neural activity $\{x_i\}$ がターゲット出力パタンに近づくような変化を引き起こす(ただし neural activity も上で説明した独自のダイナミクスで変化しているので、単純にターゲットに近づかない)。この学習則は neural activity がターゲットとなると右辺がゼロとなるために、外部から恣意的に学習時間を決めることなく、自発的に学習が停止する。

以上のダイナミクスで1つのマッピングの学習が終了する。その後に新しいマッピングの学習を同様に行い、これを繰り返す。これは逐次的な学習でありパリンプセスト学習(Parisi, 1986)とよばれ、以前の学習したマッピングは、新しい学習により上書きされていく。また本モデルでは学習時に入力強度 γ_{lrn} で neural dynamics と synaptic dynamics により $\{x_i\}$, $\{Jij\}$ がともに変化していく。一方学習終了後は、どのような力学系が形成されたのかを解析するために synaptic dynamics は入れず、つまり $\{Jij\}$ は学習で形成されたものを使い、neural dynamics のみを動かす。自発活動は入力 γ_{rcl}=0 での系の挙動で、誘起活動は $\gamma_{rcl}=\gamma_{lrn}$ での系の挙動である。ここで γ_{lrn} は学習時に印加した入力強度

であり γ_{rcl} は学習後に誘起活動を解析するときに使用するパラメタである。これらは独立にとれるが、以下では入力強度による分岐解析以外では $\gamma=\gamma_{rcl}=\gamma_{lrn}$ を用いて行う。

2-4 Definition of Memory

本研究では自発活動とその応答の解析を行うために Memories as Bifurcations という新しい記憶の見方に従い学習を行う。従って記憶できたかどうかの定義も従来の Memories as Attractors View のように単に学習パタンの安定性ではなく、新しいものが必要である。ここでは以下のよう想起と記憶を定義する。ここで想起とは1試行に関して定義されるものであり、記憶とはネットワークに関して定義されるものである。

想起: 入力を印加することで系は変化するが、その変化した系において neural activity(発火パタン)とターゲット出力パタンとの重なり $x\xi/N$ が十分に大きければ想起ができたとする。ここで十分大きいとは、他のパタンとの重なりよりターゲットとの重なりが大きい事を指す[1]。また印加する入力強度は分岐パラメタであり任意の値がとれるが、ここでは学習時に使用した強度と同じ値を持つ強度で測定する。

記憶: ターゲットとの重なりの相空間(初期値)平均$<x\xi/N>$が他のパタンとの重なりの相空間平均より大きいか否かで定義される。例えばターゲットの basin が入力を印加した系の相空間全体の半分以上をしめる場合、相空間平均をとるとターゲットとの重なりが他のより大きくなるので、このような力学系をもつネットワークは記憶をできたという事になる。実際はさらに学習プロセス(学習するパタン)の違いにより同じパラメタセットに対してもネットワーク毎にばらつきができるので、それに対して平均をとりそのパラメタにおける平均的な挙動とする。ここで注意すべきことは、これらの定義は入力が印加されている系に対して定められていつ事である。従って自発活動(入力のない系)ではターゲットの安定性がどのようになっていても構わない。

本研究での目的は、学習により形成されたネットワークに対して記憶が出来ているかどうかを調べることで、記憶が出来る領域・できない領域があるか? そ

---

[1] 本モデルでは、印加している入力パタンとの重なりが一番大きのでいかでは、入力パタンとの重なり $x\eta/N$ とターゲットとの重なり $x\xi/N$ を比較することで想起の成否を決めている。

れに対する応答、分岐がどうなっているか？を通して自発活動とその応答の解析を行う事である。

3． 結果

以上のような学習則により形成されたダイナミクスは二つの挙動に大きく分けられる。１ R相：自発活動はカオスを示し、入力に対して応答しターゲットをアトラクタとしてもつ系に分岐できる相。２ NR相：自発活動において学習したターゲットが固定点となっている。そして入力に対して系はほとんど変化せず、学習したターゲットがアトラクタとして残っている相。学習したターゲットが入力を印加してもアトラクタとして残る。
まず１マッピングの学習後の２相の挙動を解析後、複数の場合の解析を行う。

3-1 １マッピング学習後の挙動

図２にR相とNR相での入力印加による挙動の変化(想起過程と呼ぶ)を示す。上の記憶の定義で述べたように、入力印加によりターゲットとの重なりが大きくなると想起ができ、大きくないと想起失敗となる。R 相では自発活動ではカオティックに振るまい、入力を印加することでその挙動が変化する。そして変化した結果、出力すべきターゲットがアトラクタとして生成される。一方でNR相ではターゲットが自発活動において固定点として存在しており、入力を印加しても変化していないことがわかる。

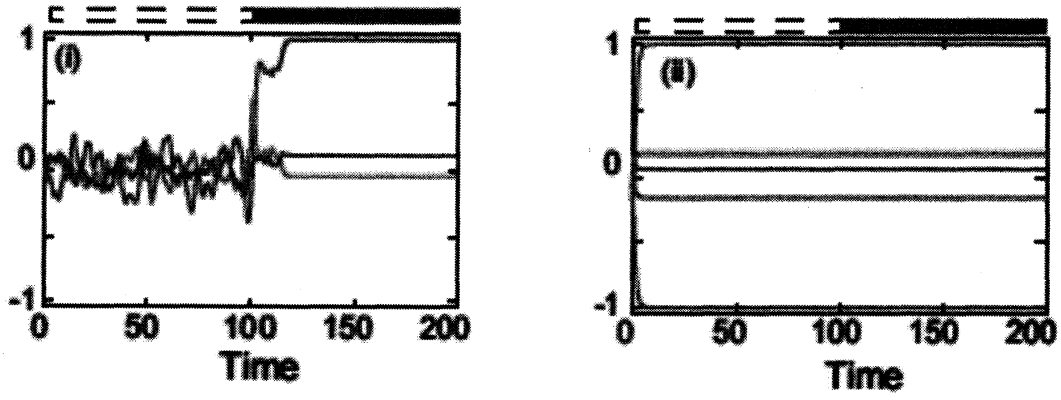

図２：　R 相(左図)と NR 相(右図)の想起プロセス例
0<t<100 が自発活動、100<t<200 が入力１が印加された系の挙動。赤：ターゲット１との重なり、緑：入力１との重なりを示す。

これらの相のα、γに対する相図を描くために、自発活動についてはダイナミクスとの重なりの時系列$x(t)\xi/N$のゆらぎの大きさ(標準偏差 SD)を、誘起活動に関しては単純に重なりの時間・ネットワーク平均値を各α、γに対して計算した(図３)。SD な大きいとき、自発活動での発火パタンはターゲットにしばしば近づく事を示しており[2]、誘起活動の重なりの平均値が大きいほどターゲットが正確に想起されていることを示す。
図３にあるように、α(学習パラメタ)が大きい、またγ(入力強度)が小さい領域で NR 相が存在している。これは直感的には、学習パラメタが大きいほど最新のターゲットに過学習になり、また入力強度が小さいほど自発活動と誘起活動が異ならないために、誘起活動での学習過程により自発活動もどうように変化してしまうせいだと考えられる。

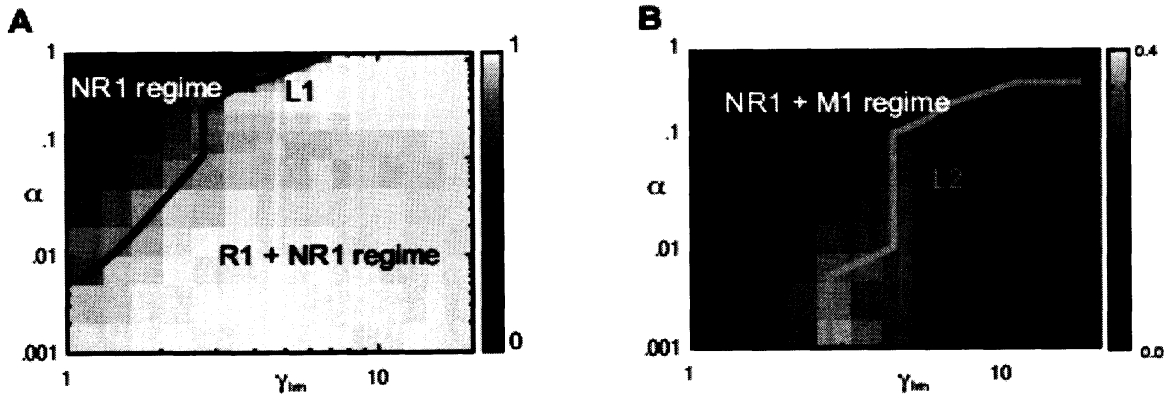

図３：誘起活動での平均的重なり(左図)と自発活動での SD(右図)のα、γに対する相図

[2] 本モデルでは自発活動はターゲットとの重なりを測定した時に原点周りで揺らぐ(R 相)か、ターゲット固定点をしめすかなので、時系列の SD が大きい時は原点周りで大きく揺らいでいる。

3-2 複数のマッピング学習後の挙動

次に先の１マッピング学習の結果を踏まえて、複数のマッピング学習後の解析を行う。複数の場合も１つの場合と同様に、R・NR 相が形成される。R 相では、カオティックな自発活動が形成され、過去に学習した複数の入力に対して正しい出力を行う。一方で NR 相では一番最近に学習したターゲットが自発活動においても固定点となり、どのような入力に対しても応答しない。以下で詳しく解析する。

a) R 相

まずR相に関して例として１つのネットワークの学習後の挙動を示す(図 4.A,C)。先に述べたように R 相では自発活動は原点を中心としたカオティックな挙動を示し、入力印加によりダイナミクスが変化しターゲット近くのアトラクタが生成される。ここでは応答の学習順序依存性μに着目する。μが１のマッピングが一番最近学習が行われたマッピングであり、μが大きくなるほど昔に学習が行われたマッピングである。図ではμ=１、５，３０の入力に対する応答を示してい

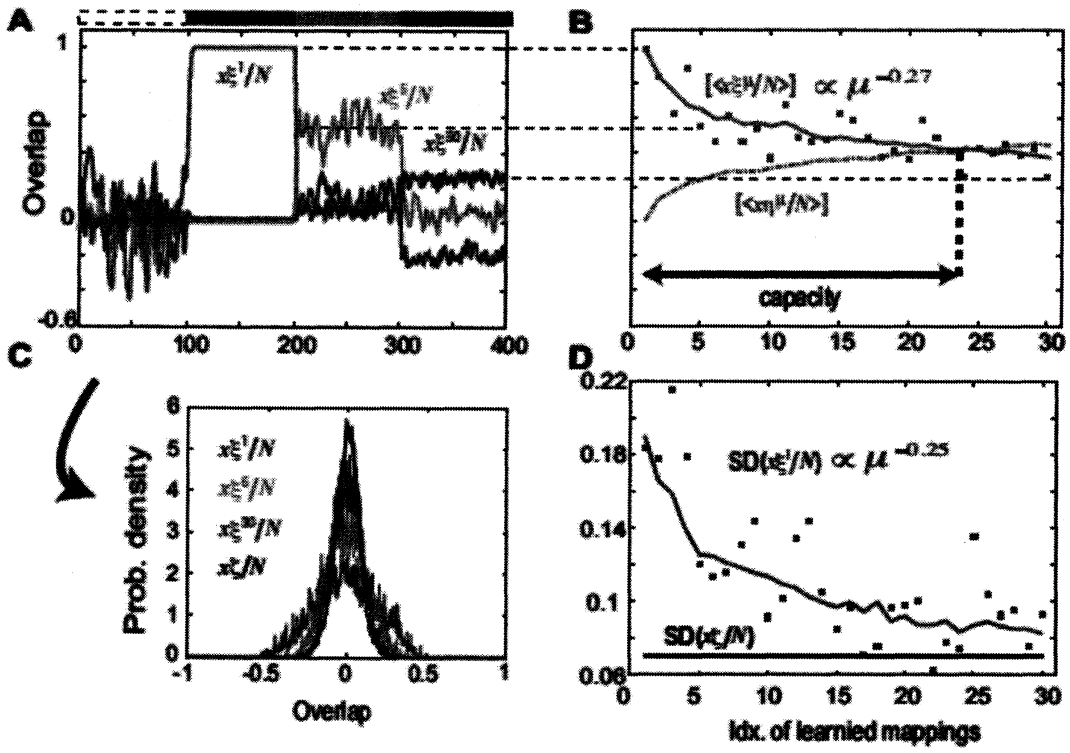

図４：４０マッピング学習後の R 相での挙動

A. 自 発 活 動 (0<t<100) 、 μ=1 (100<=t<200) μ=5 (200<=t<300) μ=30

(300<=t<400)の入力を印加した誘起活動。赤:ターゲットパタン1との重なり、緑:ターゲットパタン5との重なり, 青:ターゲットパタン30との重なりをそれぞれ表す。B.各μの入力パタンを印加したときの系の平均的挙動。ターゲットパタンμとの重なり $x\xi^{\mu}/N$と印加した入力パタンそのものとの重なり $x\eta^{\mu}/N$の時間・ネットワーク平均。 C. A で自発活動の確率分布。D. 各μ対する自発活動のSD。

る。μ=1の入力を印加すると系が変化し出力すべきターゲットとの重なりが1、すなわちターゲットそのものが固定点として生成される。それより少し古いμ=5の入力を印加した場合でも、ターゲット近くのアトラクタが生成されるが、固定点ではない。この場合は、ターゲットとの重なりが入力パタンとの重なりより大きく、定義から想起出来ていることがわかる。さらに古いμ=30を印加した場合では入力に対すて系は応答するもののその変化は小さく、ターゲットとの重なりは入力とのそれよりも小さい。この場合は想起に失敗している。μ依存性を詳しく解析するために、各μに関して重なりの時間・ネットワーク平均をとった[<xξ^μ/N>]を図4Bに示した。μが増加するに伴い徐々にターゲットとの重なりが減少する。一方で入力パタンの重なりはμに関して連続的に上昇する。従ってあるところ(=*Mc*)でターゲットの重なりと入力との重なりのクロスオーバーが起きる。*Mc* より小さいμではターゲットの重なりの方が大きいので想起は成功し、*Mc* より大きいμではインプットとの重なりの方が大きいので想起は失敗している。従って、*Mc* が想起できる最大のターゲット数を表しており記憶容量（ここでは N=100 に対して 20 程度）である。

次に自発活動に関して解析する。R 相において自発活動は原点周りでカオティックな挙動を取る。ここでは特徴量としてターゲットとの重なりの分布をある時間にわたってサンプルし、その標準偏差 SD を各μ毎に測った(図4C,D)。SD が大きいμは自発活動において、頻繁にターゲットμの重なりが大きくなるので、系の状態がターゲットにしばしば近づく事を示している。μが小さいところでは SD が大きく、μが大きくなるに連れ SD が減少していく。新しく覚えたパタンほど自発活動においてより強く想起されることがわかる。μが小さいと入力に対する応答も大きいことを考慮すると、R 相において応答が大きいパタンは自発活動でもより強く想起されるといえる。

入力強度 0 の自発活動から入力強度を徐々に大きくした時の分岐過程を解析す

る。図5はR相での$x\xi^{\mu}/N$ (μ=1,5,30)の分岐図と正のリアプノフ指数の数である。μ=1 のときはリミットサイクルを経ながら最終的にターゲットが固定点として分岐する。μ=5,30 では分岐図を見る限りこのようなクリアな分岐が起きず連続的に変化しているように見える。本モデルではN=100と高次元の力学系を扱っているために発火パタンとターゲットパタンとの重なりという一次元の軸だけでは、解析が著しく不十分なのでリアプノフスペクトルを測定し、その入力強度を上げた時の変化を解析した。図5B には正のリアプノフ指数の数をプロットした。ここでわかるように連続的に変化しているように見えたμ=5,30 の分岐過程においても正のリアプノフ指数の数はほぼ単調に減少している。従って初め高次元にわたるアトラクタが、安定性・不安定性が反転するという分岐を断続的に起こしながら、低次元のダイナミクスに変化している可能性を示唆している。ただしこの分岐過程の理解にはさらなる解析が必要である。

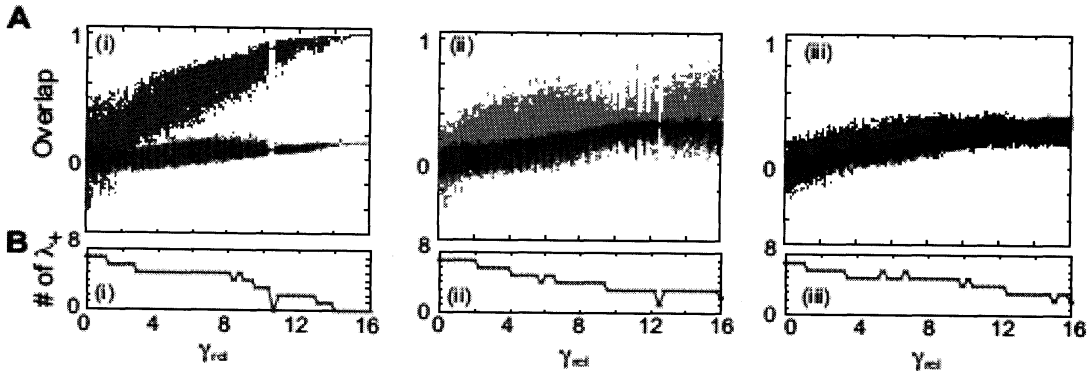

図5：入力パタンη^{μ}の入力強度による分岐
$x\xi^{\mu}/N$ による分岐図(A)と正のリアプノフ指数の数(B)。それぞれμ=1 (i),5 (ii),30 (iii)。

b) NR 相

NR 相での挙動を解析する。NR 相では自発活動は最新の(μ=1)のターゲット(とその原点に対しての対称解)が固定点アトラクタとして存在している(図6)。そしてこれらに入力を印加してもこのアトラクタは変化しない。従って最新のターゲット以外は想起できず記憶容量は1である。

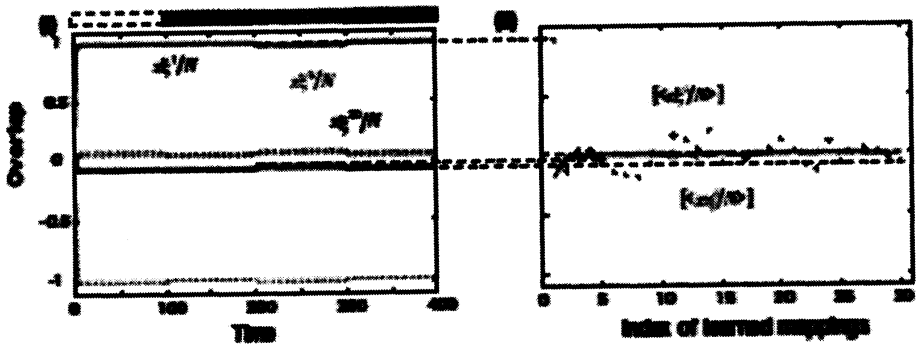

図６：　４０マッピング学習後の NR 相での挙動
A. 自発活動 (0<t<100)、 μ=1 (100<=t<200) μ=5 (200<=t<300) μ=30 (300<=t<400)の入力を印加した誘起活動。B.各μの入力パタンを印加したときの系の平均的挙動。ターゲットμパタンとの重なりと印加した入力パタンそのものの重なりの時間・ネットワーク平均。

c) α、γに対する相図
R,NR 相のα、γ依存性を解析した。図７AB はそれぞれ記憶容量 *Mc*、自発活動の平均 SD の二つの物理量をα、γに関して数値シミュレーションにより計算したカラープロットしたもの。記憶容量は誘起活動の、SD は自発活動の特徴量である。ともにαが大きく、γが小さい領域で小さい値をとり、αが減少しγが増大するとともに値が大きくなる。αが大きくγが小さい領域が NR 相で逆の領域が R 相に対応する。
記憶容量は上で述べたようにターゲットとの重なりと入力パタンとの重なりの交点でありターゲットとの重なりのμに対する減少率を表している。大きな記憶容量をもつネットワークは重なりの減少が小さく、小さな記憶容量では急速に減少する。従ってαが小さくγが大きなパラメタ領域では、記憶容量が大きく重なりの減少が少ない。αが大きくγが小さくなるにつれて、減少率が大きくなり最終的に記憶容量が１の NR 相となる。また記憶容量は 20 程度で頭打ちになる。ここでは詳しく述べないが N を大きくした時に R 相の最大記憶容量は 0.2N 程度と N に比例する。図では有限サイズ効果のために境界がぼやけて見えるが、熱力学極限において R 相では N に比例した記憶容量をもつ一方で、NR 相では定数なので、R-NR 相の境界はクリアな転移として出てくることが期待

されるが、この解析は今後の課題である。

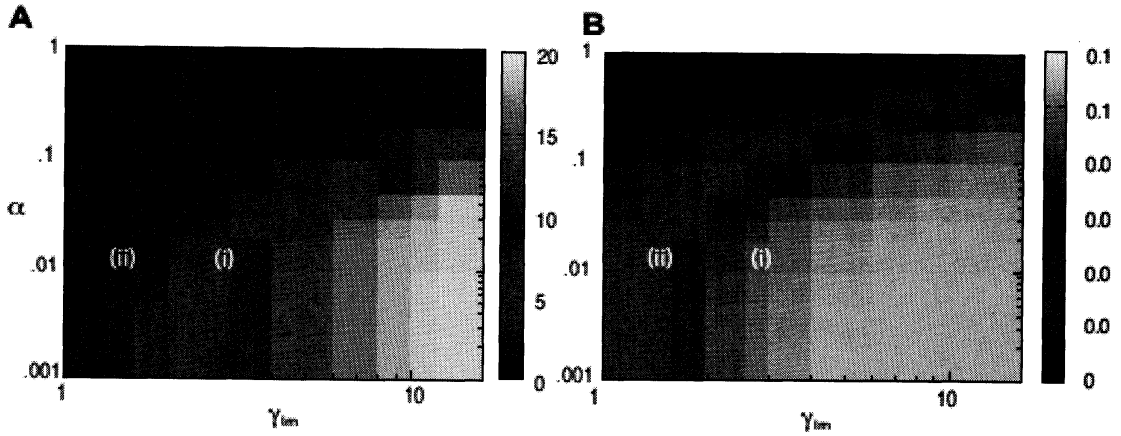

図７：入力による誘起活動と自発活動のα、γに対する相図
記憶容量 Mc(図 A)と平均 SD＝ΣμSD(μ)(図B)

平均 SD も記憶容量と同様に SD のμに対する減少率を表している。平均 SD が大きいと SD の減少が小さく、古いターゲットに対しても最新のターゲットと同じように近づく。一方で平均 SD が小さいと、減少率が大きい。αが大きくγが小さい領域では平均 SD は０であり固定点となっている。αが小さくγが大きくなるにつれて平均 SD は増大する。すなわち最新のターゲットとの重なりの SD だけが大きく最新ターゲットにのみ近づくような挙動が生成され領域から、αが小さくγが大きくなるにつれて古いターゲットにも近づくような自発活動の領域になる。平均 SD も熱力学極限は、NR 相では固定点、R 相ではカオスというようなクリアの転移があるとかんがえられるが、これについても今後の課題である。

4．　まとめ・議論

本研究では、近年着目されている自発的神経活動のダイナミクスの性質とその応答の関係について、新しい記憶・学習モデルを構築することで解析を行った。その結果、自発活動で固定点が存在し入力に対してロバストなため、応答ができない NR 相とカオティックな自発活動を示し入力に対しても適切に応答でき

るR相の2相が形成された。そしてこのR相の自発活動は単なるカオスではなく、学習されたパタンはすでに埋め込まれていて、発火パタンが学習されたパタンに次々と近づく挙動を示している。これらの結果は学習したパタンを経巡る自発活動の存在が適切な入力応答を促進していることを示唆している。また分岐解析により高次元の自発活動が、入力印加により実効的な次元を下げ低次元の“集団運動”のようなダイナミクスになる事も示された。
興味深いことに、過去のパタンを経巡る自発活動に類似した自発的神経活動は近年実験で盛んに観察されている。例えばネコの視覚野において方向特異性をもった視覚刺激に特有の活動パタンが知られているが、視覚刺激が存在しない状況では、その特有の応答活動パタンを次々に変遷するという事が観察されている(Kenet et al., 2003)。刺激特有の活動パタンが生後経験依存的に形成される(広義の学習とみなせる)事を考慮すると学習されたパタンを経巡る神経活動と解釈できる。本研究はこのように、近年観察されつつある自発的神経活動のダイナミクスに関して、入力応答という観点から新しい見方を提供している。

参考文献

Amit, D., & Brunel, N. (1997). Dynamics of a recurrent network of spiking neurons before and following learning. *Network: Computation in Neural Systems*, *8*(4), 373–404. doi:10.1088/0954-898X/8/4/003

Ben-Yishai, R., Bar-Or, R. L., & Sompolinsky, H. (1995). Theory of orientation tuning in visual cortex. *Proceedings of the National Academy of Sciences of the United States of America*, *92*(9), 3844–3848.

Hopfield, J. J. (1984). Neurons with graded response have collective computational properties like those of two-state neurons. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(10), 3088–3092.

Hubel, D. H., & Wiesel, T. N. (1962). Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. *The Journal of physiology*, *160*, 106–54.

Kenet, T., Bibitchkov, D., Tsodyks, M., Grinvald, A., & Arieli, A. (2003). Spontaneously emerging cortical representations of visual attributes. *Nature*, *425*(6961), 954–956.

Kurikawa, T., & Kaneko, K. (2012). Associative memory model with spontaneous neural activity. *EPL (Europhysics Letters)*, *98*(4), 48002. Retrieved from http://stacks.iop.org/0295-5075/98/i=4/a=48002

Luczak, A., Bartho, P., & Harris, K. D. (2009). Spontaneous Events Outline the Realm of Possible Sensory Responses in Neocortical Populations. *Neuron*, *62*(3), 413–425. Retrieved from

Parisi, G. (1986). A memory which forgets. *Journal of Physics A: Mathematical and General*, *19*(10), L617–L620. doi:10.1088/0305-4470/19/10/011

Smith, M. A., & Kohn, A. (2008). Spatial and Temporal Scales of Neuronal Correlation in Primary Visual Cortex. *J. Neurosci.*, *28*(48), 12591–12603.

Van Vreeswijk, C., & Sompolinsky, H. (1996). Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science (New York, N.Y.)*, *274*(5293), 1724–6.

Worgotter, F., Suder, K., & Zhao, Y. (1998). State-dependent receptive eld restructuring in the visual cortex. *Nature*, *396*(6707), 165–8. doi:10.1038/24157