

脳の視覚野における緩和計算について

科学技術振興事業団 岡田真人¹

アトラクタへの収束のダイナミクスを用いた情報処理は脳の情報処理メカニズムの有望な仮説の一つである。しかし残念ながら、この仮説を実証的に検証しようとする研究はそれほど多くない。我々はこれまで、計算論的な研究と心理学実験を用いて、運動知覚に関するアトラクタダイナミクスを用いた動的な情報処理の存在を議論してきた。ここではさらに連想記憶モデルの枠組の中で、物体認識に関係すると思われる生理学的な知見を説明するモデルを議論し、この知見もアトラクタダイナミクスの枠組の中で説明できることを述べる。

1 はじめに

Sugaseらは、表情を持った複数のヒトやサル顔および単純図形をサルに呈示し、Inferior Temporal cortex(IT野)等に存在する顔細胞(face-responsive neuron)について、ニューロン(神経細胞)のシングルユニットレコーディングを行ない、そのニューロンが持つ情報量の時間変化を議論した[1]。IT野は物体認識に深く関係している領野であると考えられている。最近の生理学実験は、情報理論的な手法を用いてニューロンの運ぶ情報量を定義し、その時間的な変化を定量的に議論できる段階まで達している。これらの手法に関しては色々なものが提案されているが、基本的には呈示する視覚刺激や運動課題等に関して分類や分割を導入して、それらについてニューロンがどの程度の情報量を持つかを議論している。Sugaseらは、ニューロンの発火の初期部分が、“顔対単純図形”や“ヒトの顔対その他”のような大分類に関する情報を持ち、それに続く後半部分が、ヒトの個体識別や表情の分類のような小分類に関する情報を持つことを発見した。彼女らはこの種のニューロンが大分類と詳細な分類の二つのコーディングを行なっていることから、これらをデュアル(コーディング)ニューロンと呼んだ。また、全てのニューロンがこのような二重のコーディングを行っているのではなく、大分類だけをコードするニューロン(ラフニューロン)や詳細な分類だけをコードするニューロン(ファインニューロン)も存在することを発見した。しかしこれらのニューロンに関しても、大分類および詳細な分類の情報が表現される時間関係はデュアルニューロンと同じであり、大分類に関する情報は発火の初期部分で運ばれ、詳細な分類はそれに続く部分で表現されている。

理論家やモデラーの立場からは、シングルニューロンの挙動だけではなく、ニューロン集団の巨視的な挙動を知りたいところである。しかし、現在の精密な実験技術を持ってしても、現実的にはこの願いを直接かなえることは非常に困難である。しかし、ニューロンに多数の入力を与える情報量解析の手法は、前述の理論家の願いをある程度かなえてくれる。多数の入力に対するシングルニューロンの挙動から、一つの入力に対するニューロン集団を挙動を推測することができるからである。Sugaseらの知見から、この

¹E-mail: masato@erato.atr.co.jp

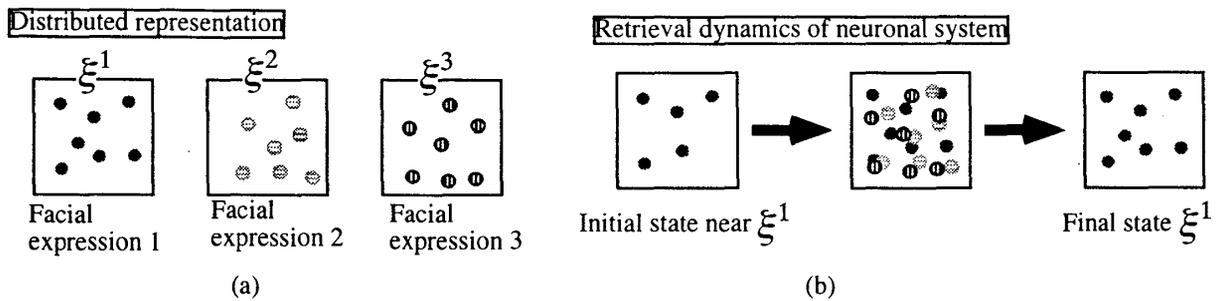


図 1: IT 野の顔に反応する細胞の発火ダイナミクス

ニューロン集団の挙動は図 1 のように推測される。まず各々の個体の顔や表情が分散コードされていると仮定する。このニューロン集団の発火パターン全体で個体の顔や表情が表現されていると考え、その発火パターンを平衡点型のアトラクタとして情報処理に用いるのである。例えば、視覚入力としてヒトの個体 1 が呈示された場合、そのニューロン集団の発火ダイナミクスは最終的に個体 1 のパターン (ξ^1) に収束する。Sugase らの知見から、この緩和過程の中間状態で、ニューロン集団がヒトの顔をコードするパターン全体の OR 的な混合状態を中間状態として経過していると推測される。

本報告の目的は、図 1 のようなダイナミクスを持つモデルを議論することである。Sugase らが用いた視覚刺激全体が持つ重要な性質の一つは、大分類や小分類を自然に導入することが出来るパターンの階層性である。そのような階層的な構造を持つパターンを憶えさせた連想記憶モデルの想起ダイナミクスにより、図 1 のダイナミクスを再現できることを示す。

2 モデル

モデルは N 個の素子から構成されている。 j 番目の素子から i 番目の素子への結合荷重は J_{ij} であり、素子の出力関数は符号関数 $\text{sgn}(\cdot)$ とする。モデルのダイナミクスを離散時間同期型とする、

$$x_i^{t+1} = \text{sgn}\left(\sum_{j \neq i}^N J_{ij} x_j^t\right). \quad (1)$$

ここで x_i^t は時刻 t での i 番目の素子の出力である。式 (1) は交換相互作用 J_{ij} をもつ Ising スピン系の Glauber ダイナミクスを同期型更新に代えたものに対応する。このスピン系の緩和のダイナミクスで §1 で述べたニューロンのダイナミクスのメカニズムを議論するのが本報告の目的である。

結合荷重 J_{ij} が、式 (1) のダイナミクスの固定点型のアトラクタを決める。このアトラクタの平衡パターンを次のように生成する。Sugase らの視覚刺激の特徴は階層性である。視覚刺激はまず顔と顔以外の画像に分類される。顔はヒトとサル顔に分類され、さらにそれぞれの個体や顔の表情に分類される。このような階層性を模倣するために超計量パターンを用いた [2]。簡単のために 2 段階の超計量パターンを用いた。より複雑で非一様な超計量パターンに議論を拡張することは容易であり、以下で得られる結果は定性的には変化しない。超計量パターンを生成する方法は数多く提案されているが、ここでは以下の方法でそのパターンを生成した。まず、そのパターンが属するグループの中心 (親パターン ξ^μ と呼び、 ± 1 の 2 値

で構成する)を生成する. 簡単のため親パターンは互いに無相関であるとし, その各成分は以下の確率で独立に生成する,

$$\text{Prob}[\xi_i^\mu = \pm 1] = \frac{1}{2}. \quad (2)$$

次に, その親パターン ξ^μ に対して一定の相関を持つように記憶パターン $\xi^{\mu\nu}$ を s 個生成する,

$$\text{Prob}[\xi_i^{\mu\nu} = \pm 1] = (1 \pm b\xi_i^\mu)/2. \quad (3)$$

同じ親パターン ξ^μ を持つ記憶パターン $\xi^{\mu\nu}$ 間の距離 (相関) は次の相関行列 \mathbf{B} で記述される,

$$(\mathbf{B})_{\nu\nu'} = E[\xi_i^{\mu\nu}\xi_i^{\mu\nu'}] = \delta_{\nu\nu'} + b^2(1 - \delta_{\nu\nu'}). \quad (4)$$

異なった親パターンを持つ記憶パターンは直交する. このような2段階の超計量パターンを平衡点にするために次の学習則で結合荷重 J_{ij} を決める,

$$J_{ij} = \frac{1}{N} \sum_{\mu=1}^{\alpha N} \sum_{\nu=1}^s \xi_i^{\mu\nu} \xi_j^{\mu\nu}. \quad (5)$$

クラスターの数 (親パターンの数) は系の素子数 N に比例するとし αN とした. α を記憶率 (loading rate) と呼ぶ. それぞれのクラスター (親パターン) はそれぞれ s 個の記憶パターンを持つ. つまり記憶パターンは全部で $\alpha s N$ 個である.

式 (2) と (3) で生成されるような階層構造を持つパターンを学習するより複雑で洗練された方法 ([3, 4] 等) も存在するが, それらには次のような二つの欠点が存在する. これらの学習法はパターンのサフィックス μ, ν に関して局所的ではない. 言い替えると, これらの方法では記憶パターンがどのクラスター (親パターン) に属するかの情報が必要になる. これは情報処理上の欠点である. もう一つの欠点は Sugase らの生理学実験の結果を取り扱ううえで本質的である. これらの学習法では記憶パターンの超計量的な構造は破壊されてしまうので, Sugase らが生理学的にえたニューロンダイナミクスを説明することはできないと予想される.

3 平衡状態の性質

このモデルでは, 記憶パターン $\xi^{\mu\nu}$ だけでなく記憶パターン ($\xi^{\mu\nu}$, $\nu = 1, \dots, s$) の対称混合状態 $\eta^\mu = \text{sgn}(\sum_{\nu=1}^s \xi^{\mu\nu})$ も平衡状態である. これら記憶パターン $\xi^{\mu\nu}$ と混合状態 η^μ の安定性を議論するために, この系の平衡状態のオーダパラメータ方程式を SCSNA[5] を用いて導出した. えられたオーダパラメータ方程式はレプリカ法 [6] の結果と一致する. $s = 3$ についてオーダパラメータ方程式を数値的に解いてえられた相図を図 2(a) に示す. 破線と一点鎖線はそれぞれ記憶パターンと混合状態の記憶容量を示す. それぞれのパターンは記憶容量より大きな記憶率では不安定になる. $b \leq b_C = 1/\sqrt{s-1} = 1/\sqrt{2}$ の時, 記憶パターンと混合状態は平衡状態として共存する. 我々のグループは, 図 2(a) の三角形で示される領域で, 2種類の対称混合状態が共存し双安定であることを発見した. 強磁性相に相当する記憶想起相で1次相転移が起こっている. 我々が知る限り, このような現象はこれまで報告されていない. この双安定性は任意の s に関して存在する. 我々は, 双安定性の存在を含めてここで述べた理論的な結果を, 計算機シミュレーションでも確認している.

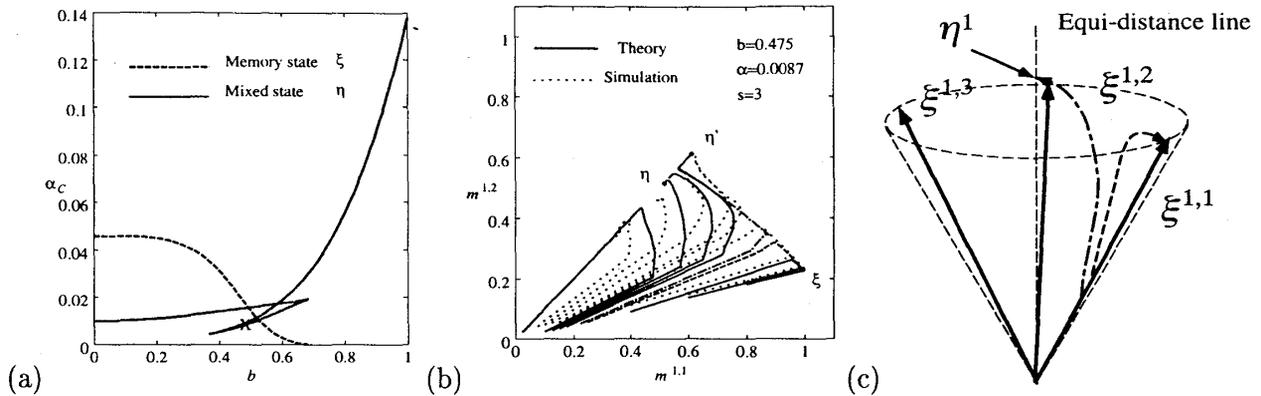


図 2: (a): 平衡状態の相図. (b): $s = 3$, $b = 0.475$, $\alpha = 0.0087$ での想起過程. (c): 図 (b) の破線と一点鎖線で示される系の軌跡の概念図.

4 想起過程の性質

次に想起過程を議論する. ここでは, 理論的に議論された想起過程の特徴と Sugase らの結果が定性的に一致することを説明する. 想起過程を理論的に議論するために統計神経力学のフルオード版を使う [7]. この理論は経路積分による厳密解 [8] と深い関係にあるが, 数値計算上の複雑さは経路積分法とは異なり適用可能である. この理論からえられる平衡状態や想起過程の動的な性質は, 計算機シミュレーションの結果を良く説明する. 図 2(b) に状態 x^t と想起パターン $\xi^{1,1}$ や $\xi^{1,2}$ とのオーバーラップ,

$$m_t^{1,\nu} = \frac{1}{N} \sum_{i=1}^N \xi_i^{1,\nu} x_i^t. \quad (6)$$

の軌跡を示す. 系のパラメータは $\alpha = 0.087$, $b = 0.475$, $s = 3$ であり, 相図上では混合状態の双安定相中に存在し, 図 2(a) の \times で示される. このパラメータでは記憶パターン (図 2(b) の ξ) と二つの混合状態 (η と η') が共存する. 初期状態 x^0 は以下の確率で各成分独立に決めた,

$$\text{Prob}[x_i = \pm 1] = \frac{1 \pm m_0^{1,1} \xi_i^{1,1}}{2}. \quad (7)$$

統計的にはこの初期状態を用いると $m_0^{1,2} = m_0^{1,3} = b^2 m_0^{1,1}$ となり, $m_t^{1,2} = m_t^{1,3}$ となるので, 図 2(b) には $m_t^{1,1}$ と $m_t^{1,2}$ だけを表示している. このようにして決めた初期状態は図 2(b) の中の $m^{1,2} = b^2 m^{1,1}$ 上に存在する. 点線は計算機シミュレーション ($N=30,000$) の結果を示す. 理論の結果は実線と破線および一点鎖線で示す. 記憶パターン $\xi^{1,1}$ が想起される場合, 全ての軌跡は図 2(b) の内側を通る. これは, 想起過程の最初の段階でモデルの状態 x^t が混合状態に一度近づき, そのあと混合状態から離れて, 最終的に記憶パターン $\xi^{1,1}$ に収束することを意味する. この軌跡を図 2(c) の破線で定性的にあらわす. 図 2(c) では記憶パターンと混合状態はそれぞれコーンの表面とそれらからの当距離線上に埋め込まれている. 図 2(c) の破線が図 2(b) の破線に対応する. この想起過程の特徴は, 図 1 で示される生理実験結果と定性的に一致することがわかる.

Sugase らはニューロンがコードする情報に従い, ニューロンをデュアルニューロン, ファインニューロンとラフニューロンの 3 種類に分類した. 一方, 副格子 (sub-lattice) 法に基づいてモデルを構成するニュー

ロンを 2^s 個のグループに分類することができる. $s = 3$ である場合, モデルのニューロンはそのコーディングに従い, $(\xi_i^{1,1}, \xi_i^{1,2}, \xi_i^{1,3}) = (\pm 1, \pm 1, \pm 1)$ の 8 グループに分類することができる. いま我々は発火率 50% の場合を議論しており, $m_i^{1,2} = m_i^{1,3}$ である場合を議論しているため, 統計的にはこれらの 8 グループの挙動は以下の 3 グループの挙動を考えれば十分である,

$$(\xi_i^{1,1}, \xi_i^{1,2}, \xi_i^{1,3}) = (+1, +1, +1), (+1, +1, -1), (+1, -1, -1). \quad (8)$$

この分類にしたがい, 例えば $(\xi_i^{1,1}, \xi_i^{1,2}, \xi_i^{1,3}) = (+1, +1, +1)$ とコードされたモデルのニューロンを $(+1, +1, +1)$ ニューロンとよぶ. $(+1, -1, -1)$ ニューロンは $\xi^{1,\nu}$, $\nu = 1, \dots, s$ の中から $\xi^{1,1}$ を区別することに関する情報を一番多く持っている. 一方 $(+1, +1, +1)$ ニューロンはそのような情報を持たない. 逆に, 同じクラスター内の記憶パターンには相関があるので, $(+1, +1, +1)$ ニューロンは記憶パターン $\xi^{1,\nu}$ が属するクラスター $\mu = 1$ に関する情報を多く持っている. まとめると, $(+1, -1, -1)$ ニューロンは同一クラスター内の記憶パターン $\xi^{1,\nu}$ を区別する詳細な分類に関する情報を担い, $(+1, +1, +1)$ ニューロンはクラスターの情報をコードしている. これらの考察から, Sugase らのファインニューロンがモデルの $(+1, -1, -1)$ ニューロンに相当し, ラフニューロンが $(+1, +1, +1)$ ニューロンに相当することがわかる. 図 2(b) の破線の軌跡であらわされる系に関して, ニューロンの出力を各ニューロングループごとに平均したものの時間変化を図 3(a) に示す. 図から $(+1, +1, +1)$ ニューロングループの緩和時間が最も速く, $(+1, -1, -1)$ が最も遅いことがわかる. この理論的な結果は Sugase らの生理実験的な結果に一致する. $t = 2$ 付近での $(+1, -1, -1)$ ニューロングループの平均出力の非単調な時間変化に注目してみよう. これは, ファインニューロンに相当する $(+1, -1, -1)$ ニューロングループで $t = 2$ 付近で状態 -1 をとるニューロンの割合が大きくなっていることを意味する. 実験で得られたデュアルニューロンは異なる時間領域で大分類と詳細な分類を二重にコードしている. このニューロングループが混合状態で取る状態は -1 である. この考察から, $(+1, -1, -1)$ ニューロンの一部がデュアルニューロンに対応することがわかる.

初期状態のオーバーラップ $m_0^{1,1}$ が小さい場合はノイジーな初期状態に対応する. この場合, 図 2(b) の一点鎖線に示すように, ネットワークの状態は記憶パターンではなく混合状態に収束する. これから想起の特性が定性的に変化する $m_0^{1,1}$ の臨界値が存在することがわかる. この場合の軌跡を図 2(c) の一点鎖線で定性的にあらわす. 図 2(c) の一点鎖線が図 2(b) の一点鎖線に対応する. 図 2(b) の一点鎖線の軌跡であらわされる系に関して, ニューロンの出力を各ニューロングループごとに平均したものの時間変化を図 3(b) に示す. $(+1, -1, -1)$ ニューロングループの平均出力をあらわす実線を図 3(a) と (b) に関して比較すると, 初期オーバーラップが臨界値より小さい場合 ($m_0^{1,1} < m_c^{1,1}$), $(+1, -1, -1)$ ニューロングループの平均出力が定性的に変化していることがわかる. もしこの変化が生理学実験的にえられれば大変興味深い. Shidara ら [9] や Amit ら [10] が行なったような, オリジナルな視覚刺激にノイズを加えた視覚刺激を使うことが, このような結果をえるための実験条件の一つである. そのような視覚刺激を Sugase らの実験に用いると, 画像に加えるノイズの量に臨界値が存在し, その値以上のノイズではデュアル (ファイン) ニューロンの時間変化が定性的に変わることが予想される.

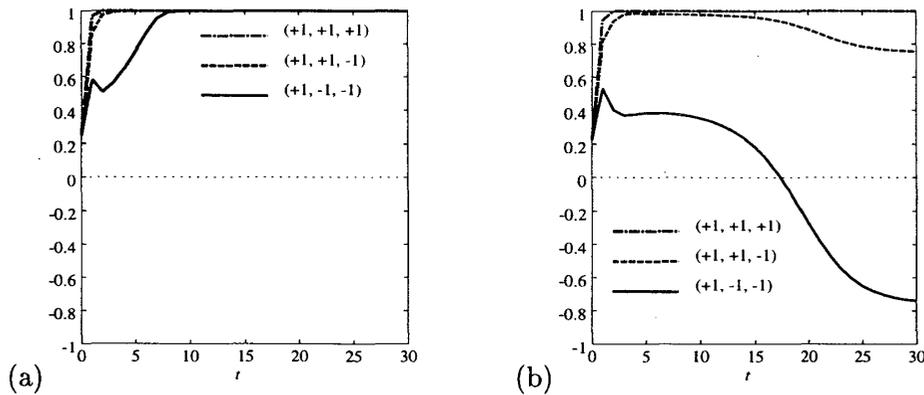


図 3: ニューロンの出力を各ニューロングループごとに平均したものの時間変化を示す. (a): 図 2(b) の破線で示される記憶パターンを想起する場合. (b): 図 2(b) の一点鎖線で示される混合状態を想起する場合.

5 まとめと議論

本報告では Sugase らによって報告された IT 野顔細胞のダイナミクスを, 典型的なアトラクタネットワークである連想記憶モデルの枠組の中で議論し, 彼女らの知見をアトラクタの想起過程のダイナミクスで説明した. 図 1 に示される生理学的な知見とここで議論した理論的な結果の食い違いの一つは, 中間状態として経由する混合状態の違いである. モデルでは同じクラスターに属する記憶パターンの各要素の多数決から決まる要素を持つ“多数決混合状態”を経由する. 一方, 生理学的には図 1 に示すように各要素の OR から決まる状態を経由する. これをここでは“OR 混合状態”と呼ぼう. 多数決混合状態と OR 混合状態は両方とも, 同一クラスターに属する記憶パターンからの等距離線上に存在する. それぞれの混合状態の違いはニューロンのしきい値が原因である. ところでここで議論している記憶パターンの発火率は 50% である. 発火率が小さい記憶パターンを用いることをスパースコーディングとよぶ. 脳内ではスパースコーディングが用いられていること示唆する生理学的な知見が存在する [11]. また理論的にもスパースコーディングは記憶容量が増大する等の利点を持つ [12, 13]. 我々は, 今回紹介したモデルにスパースコーディングを適用すると, ダイナミクスの中間状態で多数決混合状態ではなく OR 混合状態を経由することを発見した [14]. これはここで議論した 50% コーディングのモデルよりもよく Sugase らの知見を説明する. また生理学および解剖学的によりリアリスティックな興奮性 - 抑制性モジュールを用いたモデル [15] では, よりよく Sugase らの結果を説明できるがわかってきている. 我々は, このモジュールが IT 野の顔コラム [16] に対応すると考えている.

最近, Parga と Rolls は混合状態が座標変換に対して不変なパターン認識を行なうための重要な機構であるという仮説を提案している [17]. 彼らのモデルとここで議論したモデルには次のような“双対性”が存在する. ここでは構造を持った記憶パターンを構造を持たない単純な Hebb 則で学習している. 一方彼らのモデルでは, 記憶パターンは均一で構造を持たないが, 学習則によって構造が導入されている. つまり彼らのモデルとここで議論されているモデルは Hopfield モデルの典型的な二つのバリエーションであると考えられる. この観点から考えると, Miyashita アトラクタのモデルである相関アトラクタ [18] は Parga と

Rolls のモデルと同じものに属する。Parga と Rolls の研究 [17] やここで述べた結果から分かるように、これら二つのバリエーションは同じ性質を持つ部分もある。しかしながら、これら二つのバリエーションは、想起過程のような動的な性質は完全に異なっている。Parga と Rolls のモデルでは、ここで議論したような過渡的な現象は観測できないはずである。このような観点から、典型的なモデルを生理学的なデータと比較することにより、我々の外側にある複雑な超計量的な構造がどのように脳内にエンコードされているかを議論することが出来るかもしれない。

謝辞

この報告書の内容は、大阪大学大学院基礎工学研究科戸谷要氏、大分高専電気工学科木本智幸氏、奈良先端大情報科学研究科松本有央氏、科学技術振興事業団銅谷賢治博士との共同研究による。ここに感謝する。

参考文献

- [1] Y. Sugase, S. Yamane, S., S. Ueno, and K. Kawano, Global and fine information coded by single neurons in the temporal visual cortex. preprint.
- [2] R. Rammal, G. Toulous, and M. A. Virasoro, *Reviews of Modern Physics*, **58** (1986), 765.
- [3] H. Gutfreund, *Physical Review A*, **37** (1990), 570-577.
- [4] N. Brunel, *Journal de Physique I*, **3** (1993), 1693.
- [5] M. Shiino and T. Fukai, *Journal of Physics A: Mathematical and General*, **25** (1992), L375.
- [6] J. F. Fontanari, *Journal de Physique*, **51** (1990), 2421.
- [7] M. Okada, *Neural Networks*, **8** (1995), 833.
- [8] E. Gardner, B. Derrida, and P. Mottishaw, *Journal de Physique*, **48** (1987), 741-755.
- [9] M. Shidara, Z. Liu, and B. J. Richmond, Stimulus degradation has a similar effect on stimulus-related information in inferior temporal (IT) neurons and on the monkey's behavior. *Soc. Neurosci. Abstr.*, **22** (1996), 1615.
- [10] D. J. Amit, S. Fusi, and V. Yakovlev, *Neural computation*, **9** (1997), 1071.
- [11] Y. Miyashita, *Nature*, **335** (1988), 817.
- [12] M. V. Tsodyks and M. V. Feigel'man, *Europhysics Letters*, **6** (1988), 101.
- [13] M. Okada, *Neural Network*, **9** (1996), 1429.

- [14] 木本 智幸, 岡田 真人. IT 野の顔細胞のダイナミクスを説明する連想記憶モデル - 階層構造を持つスパースパターンを学習したモデル -, 日本神経回路学会第 9 回全国大会投稿中.
- [15] D. J. Amit and N. Brunel, *Network: Comput. Neural Syst.*, **8** (1997), 373.
- [16] G. Wang, K. Tanaka, K. and M. Tanifuji, *Science*, **272** (1996), 1665.
- [17] N. Parga and E. Rolls, *Neural Computation*, **10** (1998), 1507.
- [18] M. Griniasty, M. V. Tsodyks, and D. J. Amit, *Neural Computation*, **5** (1993), 1.