

大脳皮質視覚野の情報表現を眺める

科学技術振興事業団 岡田 真人*

(受付 2000 年 12 月 27 日 ; 改訂 2001 年 3 月 19 日)

要 旨

本解説では脳の視覚野の情報表現の解明に主成分分析 (PCA) や多次元尺度法 (MDS) がどのように使われているかを紹介する。大脳視覚野の解剖学および生理学的な知見を簡単に述べ、ニューロンの集団的な挙動を MDS で解析した例として Young と Yamane の研究を紹介する。彼らは AIT 野の顔ニューロンの集団を MDS で解析した。さらに顔の物理的な特徴ベクトルを MDS により解析し、それらは良く似ていることを発見した。Makioka らは、2 次元パターンに関する心理学的な空間とパターンの物理的な空間を MDS で比較し、それらが互いに似ていることを発見した。最近、我々は側頭葉の顔ニューロンの集団の動的な振舞を MDS を用いて解析した。その結果、ニューロン集団において、大分類と詳細な分類の情報表現に階層性があり、情報表現が大分類から詳細な分類へと時間的に変化していることが分かった。

キーワード：脳，視覚野，ニューロン，情報表現，PCA，MDS.

1. はじめに

日頃、脳科学や神経科学に接する機会が少ない読者でも、脳はニューロンと呼ばれる多数(ほぼ 140 億個)の神経細胞から構成されているのはご存じであろう。そして我々の知覚、運動、認知、記憶さらには思考などの高次機能が、ニューロンから構成される脳によって司られていると聞いても異論を挟まれる方は多くないだろう。脳が他の臓器と最も異なっているのは、情報処理を行なっている点である。脳が情報処理を行なっているのだから、脳の中には何らかの情報表現が存在して、その情報表現を使って計算をおこなっているはずである。脳内の情報表現を調べることは科学的な観点はもちろんのこと、脳型計算機等の工学的な観点からも重要なことは言うまでもない。しかし脳は非線形な素子であるニューロンが多数結合している系なので、脳内情報表現を探ると言ってもそれほどたやすいことではない。

脳内の情報表現をさぐる伝統的な手法の一つは電気生理学で、微小電極でニューロンの応答特性を測定する。最近は多重電極で複数個のニューロンを同時に測定することが可能になっている。それでは、多数のニューロンの応答特性に関する知見を積み重ねて行くだけで、脳内の情報表現が明らかにされるだろうか。この問いに Yes と答えるのはあまりにも楽観的すぎるだろう。得られるデータは多数のニューロンの高次元データであり、ニューロンのダイナミクスに

*現 理化学研究所 脳科学総合研究センター：〒351-0198 埼玉県和光市広沢 2-1。

Figures 1 and 3: Reprinted with permission from Young, P.M. and Yamane, S. (1992, Sparse population coding of faces in the inferotemporal cortex, *Science*, **256**, 1327-1331). Copyright (1992) American Association for the Advancement of Science.

より複雑な挙動を示すであろう。測定された高次元データをただ眺めているだけでは、何の知見も得られないことは容易に想像できる。その時まず試されるべき手法が主成分分析 (Principal Component Analysis, PCA) や多次元尺度法 (Multi-Dimensional Scaling, MDS) である。PCA や MDS は線形的手法であるので、一般には脳のような非線形な系の解析には向かないように思えるが、これから紹介するような工夫により、現在でもニューロンの集団的振舞の解析の標準的な手法となっている。本解説は、MDS や PCA を用いて脳の視覚野について情報表現についてどのような知見が得られているかを紹介する。

2. おばあさん細胞仮説と分散コード仮説

例えば自分のおばあさんを見て、そのおばあさんと認識した場合を考えてみよう。その時、ニューロンのどのような活動が認識に直接関与しているのだろうか。この問いに対して、大きくわけて二つの仮説が提案されている。一つ目の仮説は、おばあさん細胞仮説と呼ばれるものである。おばあさん細胞仮説では、一人のおばあさんに対応して一つまたは種類のニューロンが存在し、そのニューロンが応答することが最終的な認識に対応している。このような外界と脳の内部状態の単純な対応が、知覚や認識に直結していることに抵抗を持たれる読者も多いのではないだろうか。しかし次節で述べるような電気生理学的な知見より、脳内に自分のおばあさんに対応するニューロンが存在して、そのニューロンの応答がそのおばあさんを認識したことに対応すると考えるのもまったく根拠がないものではないとされている。おばあさん細胞仮説に対立する仮説としては分散コード仮説がある。分散コード仮説では、情報は 0, 1 のビット列に対応するようなニューロン集団の発火パターンでコードされる。おばあさん細胞仮説とは正反対に、分散コード仮説では、特定のニューロンが発火することが直接情報を表すのではなく、ニューロン集団の発火パターンに情報がコードされている。分散コード仮説の利点は、ニューロン集団の発火パターンで表されるベクトル間の距離で外界の事象の遠い近いを自然に表現できることである。以上説明した二つは両極端の仮説である。現状では、個々のニューロンはある程度の刺激選択性を持つが、物体や事象そのものを直接表現するほどの情報は持たず、ある程度のニューロンが集まって外界を表現すると考えられている (Fujita et al. (1992))。この立場では、一つのニューロンがどのような特徴選択性を持つかを調べることと、それらの個性を持ったニューロンが集団としてどう振舞い、その振舞が情報処理の観点からどのような機能を担っているかを調べるのが重要になる。そして、その振舞を解析する方法の一つが PCA や MDS である。

3. 視覚野の構造と受容野の構造

物体の認識を担当すると考えられている視覚野の部位の情報処理の流れは、

$$\text{LGN} \rightarrow \text{V1} \rightarrow \text{V2} \rightarrow \text{V4} \rightarrow \text{IT}$$

である。それぞれの視覚領野でどのような情報処理がなされているかを簡単に説明しよう。さらに詳しい内容については(甘利・外山(2000))等を参考にされたい。

網膜で捉えられた視覚入力はずまず LGN (Lateral Geniculate Nucleus, 外側膝状体) に送られる。LGN のニューロンは、視野中の限局した受容野と呼ばれる領域を持つ。LGN のニューロンの特性は同心円状のフィルタで説明できる。LGN のニューロンはオンセンターオフサラウンド型ニューロンと、その逆のオフセンターオンサラウンド型ニューロンに分類される。LGN のニューロンは入力画像に対して近似的に線形に反応することが知られているので、線形シス

テムを同定する手法を用いてフィルタの特性をくわしく調べることが出来る．LGN のフィルタは分散の違う二つのガウス関数の差で良く近似できるので、バンドパスフィルタになっている．入力画像をバンドパスフィルタでフィルタリングした結果が LGN に表現される．LGN を経由した視覚入力信号は第一次視覚野 (V1) に送られる．ここで、ノーベル医学生理学賞を受賞した Hubel と Wiesel の知見を紹介しよう (Hubel and Wiesel (1959))．Hubel と Wiesel は麻酔下のネコの視野に線分を呈示して、ネコの V1 のニューロンの応答を微小電極によって電気生理学的に調べた．V1 のニューロンは受容野に呈示された、特定の傾きを持つ線分に選択的に応答する．この性質を方位選択性と呼ぶ．V1 のニューロンのうち単純型細胞 (simple cell) と呼ばれるニューロンについても線形性が成り立つことが知られているので、LGN のニューロンと同様に線形システムを同定する手法を用いて、1 次の応答関数に対応するフィルタの特性をくわしく調べることが出来る．単純型細胞の特性はガウス関数と三角関数の積で表されるガボールフィルタ $G(x)$ でよく説明できることがわかっている(甘利・外山(2000)),

$$(3.1) \quad G(x) \propto \cos(k \cdot x - \varphi) \exp\left(-\frac{x \cdot x}{2\sigma^2}\right),$$

ここで x はフィルタリングする 2 次元空間の座標に対応する 2 次元ベクトルである．2 次元ベクトル k は三角関数の波数ベクトルであり、 φ は位相である． σ^2 はガウス分布の分散をあらわす．入力画像をガボールフィルタでフィルタリングした結果が V1 の単純型細胞で表現される．より正確には単純型細胞は、ガボールフィルタによるフィルタリング後に負の出力を 0 にする半波整流の非線形変換をほどこしたような応答特性を持つ．

ニューロンの応答が線形性を満たしていれば、線形システムの同定の手法を使ってニューロンの特性とその情報表現を系統的に調べることが出来る．しかし、線形性を用いたこの戦略は V1 ですでに破綻する．V1 には単純型細胞の他に複雑型細胞と呼ばれるニューロンが存在する．複雑型細胞は単純型細胞と同様に、受容野に呈示された特定の傾きを持つ線分に選択的に応答する．さらにその線分の呈示位置が線分の直交方向にずれた場合にも、複雑型細胞は応答する．この事実は複雑型細胞の特性が非線形であることを示唆する．なぜなら、もし複雑型細胞の特性が線形であれば、異なる位置にある同じ方向の線分を線形に加えた画像にも反応することになり、そのような線形フィルタでは方位選択性を説明することが出来ないからである．実際に複雑型細胞に線形システム同定の手法を適用しても、ガボールフィルタの特性は得られず、線形フィルタの範囲では方位選択性を説明することは出来ない．非線形な特性を取扱うためには、応答関数の次数を上げていく必要がある．複雑型細胞の数理的なモデルとしては、同じ方位選択性を持ち、少しずつ異なった位置に受容野を持つ単純型細胞の出力を加えたような特性を持つモデルが提案されている (Fukushima and Miyake (1982))．このモデルでは、線分の位置がずれても、それに対応する受容野をもつ単純型細胞が応答を出すので、複雑型細胞の出力は変わらない．このモデルは単純型細胞の半波整流特性のため応答特性は非線形であり、このモデルの 2 次の応答関数は実験的にえられた複雑型細胞の 2 次の応答関数を良く説明することが知られている (Sakai and Tanaka (2000))．

V1 の複雑型細胞が非線形特性を持つことから、V1 から投射を受ける V2 のニューロンが非線形な特性を持つことは明らかである．一般には非線形性が強くなればなるほど、複雑型細胞については 2 次で十分であった応答関数の次数を高くする必要があり、応答関数を測定する時間は組合せ論的に長くなってしまふ．そのため高次視覚野のニューロンの特性を応答関数で調べるのは現実的でなくなる．そこで登場するのが、おばあさん細胞仮説に影響を受けた“最適特徴”戦略である．V1 のニューロンについては、受容野内に最適方位に一致する線分が存在することと、ニューロンが応答することが 1 対 1 に対応していたわけである．V1 のニューロンは最適方位と言う最適特徴を持っていたわけである．最適方位の概念を拡張すれば、高次領野

も含めて視覚野のニューロンは最適な特徴を持つはずである。その考えに従い、個々のニューロンに対してその最適な特徴を見つけることで、ニューロンの特性を決めようという戦略を考えることができる。厳密な意味では、最適な特徴を見つけることと高次の応答関数を計算することはほとんど等価である。そのため、厳密な意味で最適な特徴を見つけることは組合せ論的に困難である。そこで普通に行なわれているのは、研究者の直観であらかじめ調べる視覚刺激を制限して、その中で最適な特徴を求める方法である。そのような戦略を用いて、V1 から投射を受ける V2 には、二本の線分を組み合わせた十字やコーナ、さらには渦巻に選択的に反応するニューロンが存在することが報告されている (van Essen (1999))。またさらに高次の視覚野である IT (Inferior Temporal) 野と呼ばれる領野では、より複雑な図形に反応するニューロンの存在が報告されている (Fujita et al. (1992))。

4. ニューロンのポピュレーションによる情報表現

V1 のニューロンの方位選択性は、“最適特徴” 戦略以外のもう一つの戦略を示唆している。V1 のニューロンの方位選択性は、ある特定の傾きの線分だけに反応するというのではなく、最適方位からのずれに応じて徐々に反応が小さくなっていくようなガウス関数状の特性を持つ。簡単のために、0 度から 180 度までの方位に対して、選択性を示すガウス関数の中心が 1 度ずつずれているような 180 個のニューロン集団を考えてみよう。この場合、入力された線分の方向 θ_0 は 180 次元のベクトル空間上の一点で表現される。この θ_0 を 0 度から 180 度まで連続的に変化させると、180 次元のベクトル空間上に連続的な閉じた曲線が描ける。この曲線は 180 次元空間の 1 次元の部分空間である。この部分空間を可視化するためには、PCA を用い 2 次元線形部分空間を抽出すればよいことは容易に推測できる。前に述べたことからわかるように、この方法を高次視覚野に適用する場合の問題点は、パラメタライズされた最適特徴をみつけることが組合せ的困難を内包していることである。そこで、前節の最適特徴戦略と同様に、研究者の直観であらかじめ調べる視覚刺激を制限して、その範囲内でニューロン集団の挙動を PCA や MDS を用いて低次元空間に射影して、情報表現をさぐる方法が考えられる。その際重要なのは、視覚刺激間の距離関係と脳内の表現の距離関係がどのような関係にあるかを調べることが重要である。次節では視覚刺激を適切に選ぶ研究者の直観と MDS による解析が有機的な形で組み合わせられている Young と Yamane の研究を紹介しよう (Young and Yamane (1992))。

5. 顔ニューロンのポピュレーションコーディング

IT には顔に反応するニューロン(顔ニューロン)が存在することが知られている。ここでは顔ニューロンのポピュレーション(集団)コーディングについて議論した Young と Yamane の研究を紹介しよう。線分は線分の方位で線分間の距離を定義することが出来る。しかし複雑な視覚刺激に対しては、何が重要なパラメータ(特徴)であるかが明らかではない。そこで Young と Yamane は図 1 の左の図のようなパラメータを顔の特徴量とし、その特徴量を元に顔画像間の距離を計算した結果を MDS で解析した。その結果を図 1 の右に示す。この 2 次元空間で顔画像間の距離の分散の 94% が説明できている。主観的にも似ている顔画像が近い場所に写像されているのがわかる。前節の議論にもとづけば、この 2 次元空間が直線の方位をあらわす 2 次元空間に対応し、MDS によって得られた物理的特徴に関する 2 次元空間とニューロンが張る空間に埋め込まれている低次元空間を比較することで、顔ニューロンによる情報表現を眺めることができる。そこで Young と Yamane は図 1 の顔を分類するタスクを行なっているサル (*Macaca fuscata*) の AIT と STP (Superior Temporal Polysensory area) からシングルユニット

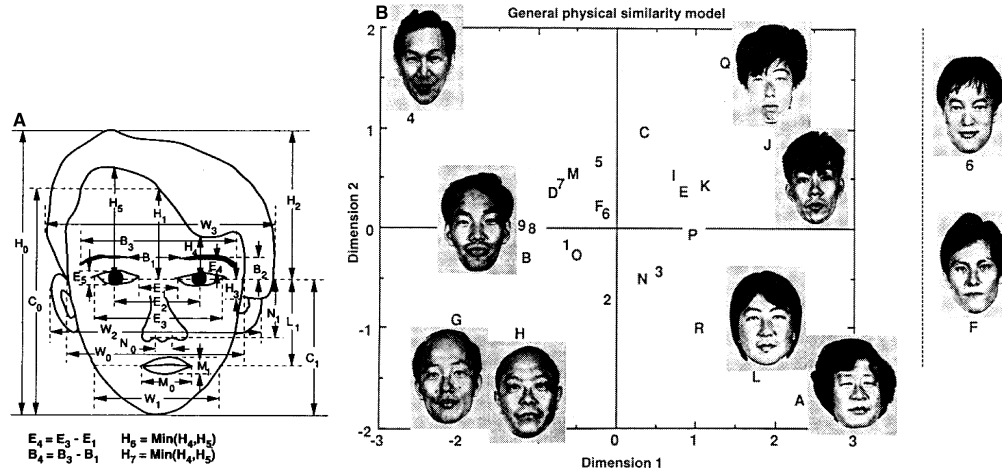


図 1. 顔の物理特性とその MDS の結果 . 文献 (Young and Yamane (1992)) より引用 .

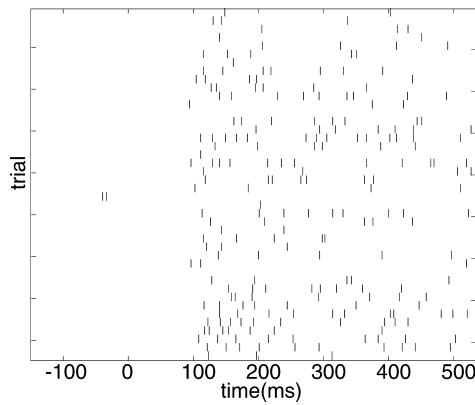


図 2. スパイククラスター .

レコーディングでニューロンの応答を測定した . 彼らはニューロンをそれらが属する領野にしたがった二つのグループにわけ , 以下のような手続きで求めたニューロンの応答から構成されるベクトルに MDS を適用した . ここでは AIT のニューロンの解析結果について紹介する . まずニューロンのスパイクをどのように処理するかを概説しよう . 普通 , ニューロンの活動は図 2 に示すようなスパイククラスターであらわされる . 時間を離散化し , その中に高々一個のスパイクしかないように時間幅を設定する . 図 2 では幅 1 msec で時間を離散化している . 視覚刺激呈示の一試行に対して , スパイクが発射された時刻に点を打ったものが図の横の一行に対応する . 図 2 は一枚の画像に対する全ての試行が記録されている . この例では試行数は 42 である . 縦軸が試行の番号に相当する . このようにして , 各時刻にスパイクが発射される確率を計算することができる . 実験では顔画像は 600 msec の間呈示されている . ニューロンは入力呈示されていない場合もスパイクを発射する . これは自発発火と呼ばれている . ニューロンの自発発火の影響を消去するために , 自発発火をベースラインとして図 2 の右の図の発火確率が

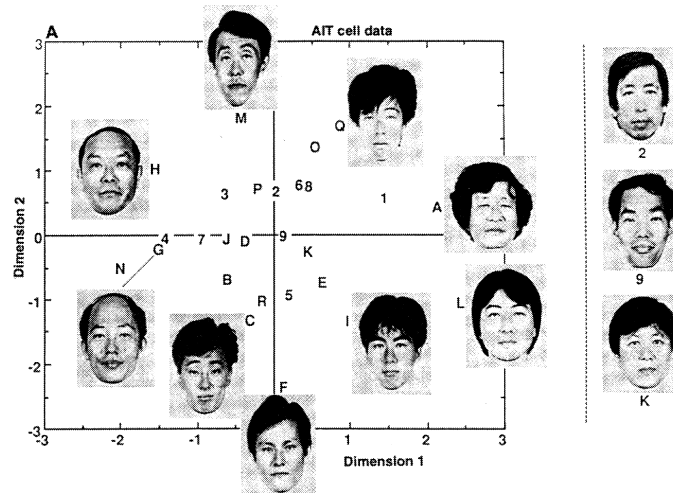


図 3. ニューロンの応答の MDS による解析の結果．文献 (Young and Yamane (1992)) より引用．

ら差し引く．顔画像呈示後の 100 msec の応答は無視し，残りの発火確率を平均して平均発火率を求める．彼らは AIT に関しては二頭のサルから合計 41 個のニューロンを測定した．つまり一つの顔画像 i に対して，41 個のニューロンの平均発火率を計算し，それらを要素とするベクトル v_i を作成する．顔画像は全部で 27 枚あるので，全ての顔画像に対応し 27 個のベクトルが存在する．これらのベクトル間の距離としてユークリッド距離 d を用いる，

$$d_{ij} = \|v_i - v_j\|.$$

Young と Yamane はこのようにして求めたユークリッド距離にもとづいて AIT のニューロン集団を MDS (ユークリッド距離を用いたので PCA と等価である) によって解析した．図 3 が AIT のニューロンに関する結果である．この 2 次元空間ではニューロンの発火率から構成されるベクトル空間での距離の分散の 70% が説明できている．図 3 の AIT のニューロンの平均発火率と図 1 の顔の物理的な特徴ベクトルに関する MDS の結果は類似性があることがわかる．彼らはより定量的に，最適な反転，回転，スケーリング変換を求めて，元の二つの空間を直接比較している．その結果，二つの空間は統計的に優位に関係していることがわかった ($r < 0.36$, $P < 0.05$)．ただし r は相関係数である．ここで重要であったのは，視覚刺激間の距離を研究者の直観により図 1 であらわせるような特徴ベクトルにもとづいて定義したことである．しかしながら，このような方法で特徴を定義しても必ずしも常にうまくいくとはかぎらない．何か系統的方法はないであろうか．次節ではそのヒントとなる Makioka らのパターン認識に関する心理学的知見とニューラルネットワークモデルに関する研究を紹介する (Makioka et al. (1996))．

6. 心理学的知見とニューラルネットワークモデル

前節では顔の物理的な特徴とサルの AIT の脳内表現の関係を議論した．ここでは 2 次元パターン認識についての心理実験とそれを説明するニューラルネットワークモデルを述べた Makioka らの研究 (Makioka et al. (1996)) を紹介しよう．心理学の分野ではパターン間の距離を，どのパターンとどのパターンを混同したかを表す混同行列 (confusion matrix) を用いて測



図 4. Rumelhart が用いたアルファベット文字 (Rumelhart (1971)). 文献 (Makioka et al. (1996)) より引用 .

表 1. MDS で求めた心理学的な 2 次元空間とニューラルネットワークで求めた 2 個の中間層による表現の間の相関 . 文献 (Makioka et al. (1996)) より引用 . 相関の計算では、反転と回転を用いて相関が最大値を取るように空間を変換した .

	第一次元	第二次元
Rumelhart のアルファベット	0.94	0.94
Blough のランダムドットパターン (前処理なし)	0.25	0.44
Blough のランダムドットパターン (複雑型細胞による前処理あり)	0.76	0.76

定し、それを用いてパターンの心理学的な空間を MDS で求める研究が数多く行なわれている . Rumelhart は図 4 に示すアルファベット 6 文字を用い、被験者にこれらの文字を短時間呈示した . 彼は被験者がどのパターンをどのパターンに間違えたかの確率を測定し、それらを成分とする混同行列を求めた (Rumelhart (1971)) . Makioka らはこの混同行列に MDS を適用し、似ているアルファベットが近い場所に写像されることを示した . Makioka らはこのような内部表現がどのように形成されるかを三層パーセプトロンと呼ばれるニューラルネットワークを用いて議論した . 図 4 のアルファベット 6 文字がビットパターンとして三層パーセプトロンの入力層に入力される . 入力層 中間層 出力層の順で変換される . それぞれの層をつなぐ結合は結合荷重と呼ばれ、各層で結合荷重で線形変換された後にシグモイド(ロジスティック)関数により非線形変換される . Makioka らはパーセプトロンにパターン認識と恒等写像の二つの課題を別々に学習させて、それぞれの中間層のニューロンがどのように応答するかを調べた . パターン認識課題では、それぞれのパターンに対して 6 個のニューロンが出力層に存在し、その中の一つのニューロンが発火することが、そのパターンを認識したことに対応する . 一方、恒等写像では出力層のニューロンの個数は入力層のニューロンの個数と同じである . 入力パターンがそのまま出力パターンになるようにパーセプトロンを学習する . それぞれの学習には出力誤差の自乗和を最小にするように、結合荷重を最急降下法(バックプロパゲーション)で求めた . Makioka らは学習終了後に 6 個の入力パターンをパーセプトロンに入力し、それらに対応する中間層のニューロンの応答(内部表現と呼ぶ)を求め、それらを MDS を用いて解析した . その結果、パターン認識を学習したパーセプトロンの内部表現は混同行列から MDS で抽出した心理学的な空間とほとんど相関を持たないことがわかった . 一方、彼らは恒等写像を学習したパーセプトロンの内部表現は心理学的な空間と非常に似ていることを発見した(表 1) . 実は三層パーセプトロンで抽出される内部表現は PCA で抽出される低次元部分空間とほぼおなじであることが知られている . つまりこの知見は、2 次元パターンが表現される心理学的空間の

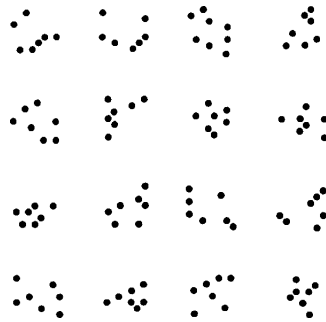


図 5. Blough が用いたランダムドットパターン (Blough (1985)). 文献 (Makioka et al. (1996)) より引用.

構造は、2次元パターンをビット表現するベクトル空間の構造とほぼ同じであるということを示唆する。

次に、彼らは図 5 の Blough のランダムドットパターン (Blough (1985)) に対して、先程と同様に心理学的にえられる空間と恒等写像を行なうパーセプトロンの内部表現を比較した。その結果先程の Rumelhart のパターンとは異なり、パーセプトロンの内部表現と心理学的に得られる空間は似ていないことがわかった(表 1)。そこで Makioka らは視覚野の構造にヒントを得て、図 5 のランダムドットパターンに対して、V1 の単純型細胞や複雑型細胞に相当するような変換をほどこした。単純型細胞はガボールフィルタの特性を持つが負の出力を 0 にする半波整流型の非線形特性を持つ。そのため、ランダムドットパターンは非線形に変換される。そのような特性から、単純型細胞は直線を検出する能力を持つ。図 5 のようなランダムドットパターンを見た場合にも、明らかに我々にはランダムドットパターンを直線で補間する能力があるので、このような非線形変換を施すのは自然な方法であろう。第 3 節で述べたように、複雑型細胞の数理的なモデルとして、同じ方位選択性を持ち、すこしずつ異なった位置に受容野を持つ単純型細胞の出力を加えたような特性を持つモデルが提案されている (Fukushima and Miyake (1982), Sakai and Tanaka (2000))。画素表示されている空間では、パターン全体が少しだけシフトしたようなパターン間の距離は一般には遠いが、複雑型細胞のようなフィルタで変換された空間では、それらは近い点に写像される。我々のパターン認識能力はシフトに対して頑健であるので、複雑型細胞のようなフィルタで変換された空間を用いるのも自然な方法である。Makioka らはそれらの変換されたパターンを用いて、恒等写像を行なうパーセプトロンを学習し、それらの内部表現を比較した。その結果、単純型細胞 複雑型細胞と入力画像の前処理が高度になるに従い、得られたパーセプトロンの内部表現と心理的な空間は似てくることがわかった(表 1)。

7. 視覚野の情報表現解明の一戦略とその工学的応用

Young-Yamane や Makioka らの結果は、視覚野の情報表現を系統的に探る一つの戦略を示唆している。1980 年代後半から 90 年代前半にかけて、三層パーセプトロンの中間層の情報表現から、視覚野のニューロンの特性を説明することが試みられていた。例えば、V4 のニューロンの色に対する応答が色の恒等写像を行なう三層パーセプトロンの中間層のニューロンの応答とよく似ていることが報告されている (Usui et al. (1991))。しかし最近では、視覚野の情報表現がこの戦略により解明されたという話は聞かない。Young-Yamane や Makioka らの結果

から、この理由は入力の実現が適切でなかったためだと推測できる。視覚野の研究を振り返ると、これまでは視覚入力の入口に相当する V1 と最終段階に相当する IT に関する知見が大変多く得られている。さらに最近では中間的な領野である V2 や V4 のニューロンの特性を調べた研究が多く見られるようになってきた。このように視覚野のすべての階層にわたって知見が得られはじめている今こそ、Young-Yamane や Makioka らの知見に基づき視覚野の情報表現を系統的に解明できる時期ではないかと個人的には思っている。第 3 節で紹介したように、V1 から投射を受ける V2 には二本の線分を組み合わせた十字やコーナ、さらには渦巻に選択的に反応するニューロンが存在することが報告されている (van Essen (1999))。例えば、V1 で抽出できるような線分から構成される比較的単純な図形に対する V2 の多数のニューロンの応答を系統的に測定し、Young-Yamane のように PCA を用いて解析する。次に、その図形を Makioka らのように V1 の複雑型細胞に相当するようなフィルタでフィルタリングした結果を PCA で解析し両者を比較する。それぞれの視覚刺激がどのように配置しているかで、V1 から V2 までにどのような情報処理が行なわれているかがわかる可能性がある。この手続きを繰り返し行なえば、V1 から IT まで首尾一貫した方法で情報表現を推測できるはずである。一部の生理学者は、低次領野の情報表現と高次領野の情報表現の関係を系統的に調べている (Kobatake and Tanaka (1994))。彼らのように低次領野と高次領野の情報表現の関係を調べる際に、ここで提案したような PCA を用いた手法を適用するのは大変興味深いテーマである。

Makioka らの結果は、高次視覚野での脳内表現は単に入力パターンである画像を PCA で次元圧縮したものではなく、低次視覚野のニューロンをカーネルとし画像を非線形変換した画像を PCA で次元圧縮したものに近いことを示唆している。図 5 のようなランダムドットパターンをパターン認識する場合、ヒトと同じように 間違える パターン認識装置を作るには、V1 の複雑型細胞のようなカーネルで入力パターンを非線形変換して得られるパターンを用いることであろう。工学的な目的として、ヒトと同じような能力を持つパターン認識装置を構築する場合、大脳視覚野の情報表現を参考にしながら、カーネルを設計するのは一つの方法であると考えられる。またニューラルネットワークの分野では、自己組織化の研究として、視覚野のニューロンのフィルタを入力画像だけから学習するアルゴリズムが数多く提案されている。これらのアルゴリズムもパターン認識装置の設計に役立つと思われる。

8. 動的な解析

ここまでの議論では、視覚野は単に入力である画像を非線形に変換(写像)することで、視覚情報処理を行なっているように感じられるかもしれない。しかし、視覚野にはその非線形変換を行なうフィルタに相当するフィードフォワード結合だけでなく、フィードバック結合も存在する。フィードフォワード結合だけでなくフィードバック結合をも用いて視覚情報処理を行なうのであれば、ニューロンの情報表現も時々刻々変化するはずである。その例として、Sugase らの研究を紹介する (Sugase et al. (1999))。Sugase らはヒトとサル顔および単純図形の 38 種類の画像をサルに呈示し、もっとも高次の視覚野であると考えられている側頭葉の単一ニューロン活動を記録した。第 5 節で述べたように、ニューロンの活動は図 2 に示すようなスパイククラスターであらわされる。このスパイククラスターの時間窓を 50 msec とし、Sugase らは画像 s と時間幅 50 msec の間のスパイク数 r の間の関係を求め、以下の相互情報量を計算した、

$$(8.1) \quad I(S; R) = \sum_s -p(s) \log p(s) - \left\langle \sum_s -p(s|r) \log p(s|r) \right\rangle_r.$$

ここで S は刺激 s のセットであり、スパイク数 r のセットを R とする。 $p(s|r)$ はスパイク数が r である条件で、刺激クラスが s である場合の条件付き確率である。 $p(s)$ は刺激 s の事前

確率で、ここでは一様としている。 $\langle \cdot \rangle_r$ はスパイク数 r に関する確率 $p(r)$ についての平均をあらわす。式 (8.1) の相互情報量の定性的な意味を以下の例で考えてみよう。2 種類の入力画像 A と B が呈示されたとしよう。今測定しているニューロンが A に対して応答し、B に対して応答しない場合、式 (8.1) の相互情報量は 1 ビットになる。これは、このニューロンの応答の状態を観測するだけで、A と B のうちどの画像が呈示されたかを決定できることを意味する。このニューロンは A に対して選択的に応答するので、このニューロンは A のおばあさん細胞に相当する。一方、発火確率が A でも B でも変わらないとすると、ニューロンの応答の状態を入力画像を分類することはできないので、相互情報量は最小値の 0 ビットをとる。実際は、試行毎にニューロンの応答が変化する場合もあるので、相互情報量は 0~1 ビットの間値を取る。ここでの説明からわかるように、相互情報量を用いたこの方法は、Hubel と Wiesel に代表されるニューロンの反応選択性の概念を情報理論的な立場で拡張したものと解釈することができる。

Sugase らはヒトとサルの顔および単純図形の 38 種類の画像に対し、“顔と単純図形” や “サルの顔とその他” のような分類を考え、それらを大分類と呼び、また “ヒトの個体分類” や “サルの表情分類” 等を詳細分類と呼んだ。図 2 に示すように、発火確率は時間的に変化するので、式 (8.1) の相互情報量も時間的に変化する。彼らはそれらの分類毎に決まる式 (8.1) の相互情報量の時間変化を求めた。その結果、大分類の相互情報量が最大値をとる時刻が詳細分類のそれより約 50 msec 先行していることがわかった。ニューロンの発火パターンは応答の開始時において大分類情報をコードし、それに遅れて詳細な分類情報をコードする。この結果は、ニューロンによる情報表現が動的に変化していることを意味する。

しかし単一ニューロンでの解析だけでは、ニューロンの集団としての情報表現やその経時の変化を明らかにすることはできない。そこで我々はニューロン集団の経時的な振舞を、MDS を用いて解析した(松本 他 (2000))。第 5 節で紹介したように、Young と Yamane は複数のヒトの顔をサルに呈示し、側頭葉の単一ニューロン活動を記録し、ニューロン集団を表す高次元ベクトルを MDS を用いて解析した (Young and Yamane (1992))。彼らはニューロン集団を表すベクトルの要素として、ニューロンの発火頻度の長時間平均を用いた。これから紹介する我々の解析は Young-Yamane を動的に拡張したものになっている。

8.1 ニューロン集団ベクトルの生成

ここで用いるニューロンのデータは同時記録されたものではない。そこで、個々のニューロンが統計的に独立に挙動すると仮定して解析を行なった。最近、側頭葉のニューロンに関して、この仮定が成り立っていることが報告されている (Rolls et al. (1999))。ニューロン集団を表すベクトルを以下のように生成した。Sugase らの生理学実験のデータから、全ての画像に対して応答が測定されている 45 個のニューロンを選び出した。一つの呈示画像 i に対して、図 2 で説明した方法で 45 個のニューロンの発火確率を計算する(図 2 参照)。50 msec の時間窓を用い、その時間窓中の平均発火率を計算する。それぞれの 50 msec の時間窓に対して、45 個のニューロンの応答が得られるので、画像は発火確率を各要素とする 45 次元のベクトルで表現される。時間窓は 0 ms から 300 ms まで 1 ms ずつスライドさせた。画像 i に対して、ある時間窓でのニューロンの平均発火率を成分とする 45 次元のニューロン集団のベクトルを v_i とする。各々の時間窓で、38 種類の呈示刺激に対応し 38 個のベクトルが存在する。これらのベクトル間の距離としてユークリッド距離 d を用いた、

$$d_{ij} = \|v_i - v_j\|.$$

8.2 解析結果

上で求めたベクトルを、各時間窓毎に Young-Householder (Young and Householder (1938)) の MDS 法を用いて 2 次元に次元圧縮した。そのため、時間窓毎に 2 次元線形部分空間は異なる。解析の結果、以下の事がわかった。[0 ms, 50 ms] の時間ウィンドウでは、サル、ヒト、単純図形の分布は重なっている。[90 ms, 140 ms] の時間ウィンドウではサルの分布全体とヒトの分布全体と図形の分布全体が分離されている。サルの表情の分布とヒトの個体の分布と図形の形の分布に関して詳細な構造は見られない。よってこの時間では大分類が行なわれているといえる。[140 ms, 190 ms] の時間ウィンドウではサルの表情の分布とヒトの個体の分布が最も分離される。つまりこの時間窓では詳細な分類が行なわれているといえる。[300 ms, 350 ms] の時間ウィンドウでは大分類と詳細な分類は減少するが、サルの分布全体とヒトの分布全体と図形の分布全体は重ならない。さらにサルの表情の分布どうしとヒトの個体の分布どうしは [0 ms, 50 ms] の時間ウィンドウに比べると分離している。MDS の寄与率については、[0 ms, 50 ms]、[90 ms, 140 ms]、[140 ms, 190 ms]、[300 ms, 350 ms] の時間ウィンドウではそれぞれ、元の空間での分散の 37%、68%、67%、57% が説明できている。これらの結果から、MDS で与えられた空間上で大分類と詳細な分類の情報表現が [140 ms, 190 ms] の時間ウィンドウで階層的な表現になることが分かった。また、[90 ms, 140 ms] の時間ウィンドウで大分類情報がニューロン集団に表現され、[140 ms, 190 ms] の時間ウィンドウで詳細な分類情報が表現されることも分かった。また、MDS による解析からえられるシステム全体の挙動は、Sugase らの単一ニューロンにおける情報量解析からえられる大分類情報と詳細な分類情報の経時的変化に矛盾しないことも分かった。

9. まとめ

本解説ではまず、ニューロンの集団を MDS で解析した例として、Young と Yamane の研究を紹介した (Young and Yamane (1992))。彼らは AIT の MDS の結果と顔の物理的な特徴ベクトルの MDS の結果は良く似ていることを示した。次に、Makioka らのパターン認識に関する心理学的知見とニューラルネットワークモデルに関する研究を紹介した。彼らの結果は、高次視覚野での脳内表現は単に入力パターンである画像を PCA で次元圧縮したものではなく、低次視覚野のニューロンをカーネルとし画像を非線形変換した画像を PCA で次元圧縮したものに近いことを示唆している。Young-Yamane や Makioka らの結果は、大脳視覚野の情報表現を系統的に探る一つの戦略を示唆している。本解説ではその戦略を紹介し、現在の生理学的な手法でも十分その戦略を推進できることを指摘した。また Makioka らの知見は、工学的な立場から最近パターン認識等に用いられている、カーネル PCA やサポートベクトルマシンとも関係が深い。工学的な目的として、ヒトと同じような能力を持つパターン認識装置をカーネルベースで構築する場合、大脳視覚野の情報表現を参考にしながら、カーネルを設計するのは一つの方法であると考えられる。またニューラルネットワークの分野で開発されている、視覚野のフィルタを入力画像だけから学習するアルゴリズムも、カーネルベースのパターン認識装置のカーネル設計に役立つと思われる。最後に、時間的に変化する情報表現を持つ例として、Sugase らの研究を紹介した (Sugase et al. (1999))。我々が行なった Sugase らのデータについての MDS の解析(松本 他(2000))を紹介し、ニューロン集団において大分類と詳細な分類の情報表現に階層性があり、情報表現が時間的に変化していることを紹介した。

PCA や MDS は線形的手法であるので、一般には脳のような非線形な系の解析には向かないように思えるのが、ここで紹介するような工夫により、現在でもニューロンの集団の振舞を解析する際に最初に試すべき標準的な手法となっている。本解説を読まれ、神経科学の分野に興

味を持たれる読者がおられるなら，それは筆者にとって望外の喜びである．

謝 辞

ここで紹介した内容の一部は，松本有央氏，銅谷賢治博士，菅生康子博士，山根茂博士，河野憲二博士との共同研究にもとづいている．特に貴重な刺激画像と生理学データを提供頂いた菅生康子博士，山根茂博士，河野憲二博士に深く感謝の意を表したい

参 考 文 献

- 甘利俊一，外山敬介 編 (2000). 『脳科学大辞典』，朝倉書店，東京．
- Blough, D. S. (1985). Discrimination of letters and random dot patterns by pigeons and humans, *Journal of Experimental Psychology: Animal Behavior Processes*, **11**, 261–280.
- Fujita, I., Tanaka, K., Ito, M. and Cheng, K. (1992). Columns for visual features of objects in monkey inferotemporal cortex, *Nature*, **360**, 343–346.
- Fukushima, K. and Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position, *Pattern Recognition*, **15**, 455–469.
- Hubel, D. and Wiesel, T. (1959). Receptive fields of single neurones in the cat's striate cortex, *Journal of Physiology*, **148**, 574–591.
- Kobatake, E. and Tanaka, K. (1994). Neuronal selectivities to complex object features in the ventral visual pathway of the macaque cerebral cortex, *Journal of Neurophysiology*, **71**, 856–867.
- Makioka, S., Inui, T. and Yamashita, H. (1996). Internal representation of two-dimensional shape, *Perception*, **25**, 949–966.
- 松本有央，岡田真人，木本智幸，銅谷賢治，菅生康子，山根茂，河野憲二 (2000). NC 研究会技術研究報告, NC99–159, 電子情報通信学会，東京．
- Rolls, E. T., Booth, M. C. A., Panzeri, S., Schlitz, S. R. and Treves, A. (1999). Information about objects and faces in the responses of simultaneously recorded inferior temporal cortex neurons, *Society for Neuroscience Abstract*, **25**, p. 528.
- Rumelhart, D. E. (1971). A multicomponent theory of confusions among briefly exposed alphabetical characters, Tech. Report, No. 22, Centre of Human Information Processing, La Jolla, California.
- Sakai, K. and Tanaka, S. (2000). Spatial pooling in the second-order spatial structure of cortical complex cells, *Vision Research*, **40**, 855–871.
- Sugase, Y., Yamane, S., Ueno, S. and Kawano, K. (1999). Global and fine information coded by single neurons in the temporal visual cortex, *Nature*, **400**, 869–872.
- Usui, S., Nakauchi, S. and Nakano, M. (1991). Internal color representation acquired by a five-layer neural network, *Artificial Neural Networks* (eds. T. Kohonen, O. Simula and J. Kangas), 867–872, North-Holland, Amsterdam.
- van Essen, D. C. (1999). Primate cerebral cortex: Structure, function, and development, Supplement to *Neuroscience Research*, **23**, S2.
- Young, G. and Householder, A. S. (1938). Discussion of a set of points in terms of their mutual distances, *Psychometrika*, **3**, 19–22.
- Young, P. M. and Yamane, S. (1992). Sparse population coding of faces in the inferotemporal cortex, *Science*, **256**, 1327–1331.

Visualizing Information Representation of Visual Area in the Cerebral Cortex

Masato Okada

(Kawato Dynamic Brain, Project, Japan Science and Technology Corporation)

In this review, I will explain how the Principal Component Analysis (PCA) and/or Multi-Dimensional Scaling (MDS) methods are utilized to explore information representation of visual areas in the brain. After reviewing the anatomical and physiological findings of the visual areas briefly, I introduce interesting works regarding comparison between physiological and physical spaces of face by Young and Yamane, and that between psychological and physical spaces of two-dimensional shape by Makioka et al. Young and Yamane analyzed the population of face-responsive neuron in the area AIT by means of the MDS. They compared this MDS result with that of some physical configurations of face, and found that they are similar to each other. Makioka et al. also found the similarity between the psychological and physical spaces. Recently, we analyzed the dynamical behavior of face-responsive neuron population in the monkey temporal cortex using the MDS. We found that the face-representation is embedded in the population of the face-responsive neurons with the hierarchical structure, and found that this internal representation dynamically changes: In the early phase (90 ms–140 ms), the population formed clusters corresponding to rough categories. In the later phase (140 ms–190 ms), the each cluster expanded to form sub-clusters corresponding to finer categories.