

脳活動から心を可視化する

堀川 友慈*・宮脇 陽一**・神谷 之康*,†

Visualizing Mind from Human Brain Activity

Tomoyasu HORIKAWA*, Yoichi MIYAWAKI** and Yukiyasu KAMITANI*,†

Whereas the human mind is a subject of many disciplines of biological and social sciences, it is often assumed that the externalization or visualization of states of the mind is in principle impossible because the mind is accessible only to the person. However, methods of neural decoding have rapidly developed over the last decade in neuroscience, allowing us to read out the contents of human perception and subjective mental states from measured brain activity patterns. Recent progress achieved the reconstruction of perceived visual images and the prediction of dream contents using machine learning-based analysis of functional magnetic resonance imaging (fMRI) of human brain activity. In this review, we first introduce basic procedures of neural decoding using fMRI data. Next, we overview studies on neural decoding of visual and subjective contents from human brain activity. Finally, we discuss challenges and prospects for future neural decoding studies.

Key words: visual perception, neural decoding, functional magnetic resonance imaging (fMRI), visual image reconstruction, dream, machine learning, multivariate pattern analysis (MVPA)

1. 脳活動から心を「見る」

科学技術の進歩は、ナノレベルでの測定を可能にする超高解像度顕微鏡や、毎秒数百万フレーム以上で動きを捉える超高速カメラなど、ヒトがもつ本来の能力では見ることができない対象を可視化する技術を多く開発してきた。これらの技術をもってしても、依然として見ることが難しい対象が、「心」である。古来より、心は、その人が発する言葉や行動などから推測するしかないものとされてきた。またそうであるがゆえ、心を読む技術への憧れが、SFや空想の世界で多く描かれてきた。しかしながら、近年の脳活動計測技術と解析技術の急速な進歩に伴い、心の状態を見ることが部分的に可能になりつつある。これを可能にする技術が「脳情報デコーディング」である。

脳情報デコーディングは、外界からの刺激や被験者の心的状態と脳活動の関係を探る手法として、神経科学の分野において広く利用されている。また、脳から読み出した情報を機械の操作に利用するブレイン-マシン-インター

フェイスの技術基盤としても、重要な役割を担っている。脳と機械との間での直接的な情報伝達が可能になるため、運動機能に障害をもつ人々の機能補綴を目指した医療応用としての期待も高い。このように、脳情報デコーディングは、基礎と応用の両方面において多岐にわたる発展性をもつ技術として広く注目を集めている。

脳情報デコーディング技術を用いた研究の中でも、機能的磁気共鳴画像法 (functional magnetic resonance imaging; fMRI) を用いたヒトの視覚研究は技術の進歩が著しい¹⁻⁵⁾。本稿では、「ヒトの視覚世界の可視化」をテーマとして、見ている対象を脳活動から読み出す技術という観点から、現在の脳情報デコーディング技術の実態について解説する。

2. 脳情報デコーディング

従来の fMRI 研究では、特定の感覚刺激・運動課題において、fMRI の信号強度に差が出る脳部位を同定するとい

* ATR 脳情報研究所 (〒619-0288 京都府相楽郡精華町光台 2-2-2) †E-mail: kmtn@atr.jp

**電気通信大学先端領域教育研究センター (〒182-8585 調布市調布ヶ丘 1-5-1)

う手法がとられていた。このように、fMRI 画像の個々の画素（ボクセル）ごとに刺激・運動課題への反応性の違いを調べることで脳部位と機能との対応づけを行うアプローチは、「脳機能マッピング」とよばれ、脳イメージング研究の主流をなしていた。一方、脳情報デコーディングでは、脳活動パターンに外的な刺激や自身の運動状態、心的内容などが符号化（コード）されていると見なし、そのコードを解読（デコード）することによって、刺激や運動の状態、心的内容の予測を行うというアプローチをとる³⁻⁵⁾。この方法では、複数のボクセルからなるパターンを解析することで、従来の脳機能マッピングでは調べることができなかった詳細な情報の脳内表現を探ることが可能となった。脳情報デコーディング技術は、神経科学研究における解析手法として、従来法からの大きなパラダイムシフトを引き起こしたといえる。

脳情報デコーディング研究では、機械学習アルゴリズムを用いた「教師あり学習」の手続きに従って解析が行われることが多い⁶⁾。具体的には、まず計測した個々の fMRI 画像に対して、被験者が知覚している刺激や被験者が行っている運動課題の条件に応じたラベルづけを行う。次に、解析対象とする脳部位（関心領域）を決定し、個々の脳画像における関心領域内のボクセルの信号値を並べたベクトルを作成する。そして、この信号値ベクトルから対応するラベルの情報を予測する機械学習モデル（デコーダー）の学習を行う。学習済みのデコーダーを使って、学習用データとは独立なデータセット（テストデータ）の信号値ベクトルからラベルの予測を行い、予測成績を評価する。以上の手続きにより、予測の正答率や誤差などの指標を使って、各関心領域に興味の対象となる情報が表現されているかどうかを定量的に評価することが可能となる。

3. 視覚像のデコーディング

脳情報デコーディングの特徴は、脳活動のパターンを利用することにある。脳活動のパターンに情報が表現されていることを示した先駆的な研究として、Haxby らの研究がある¹⁾。彼らは、人の顔や家などの数種類の物体カテゴリーの画像を被験者が観察した際の fMRI 計測を行い、同じカテゴリーに属する異なる画像を見せたときのボクセルパターンの相関が、異なるカテゴリーに属する画像を見せたときの相関よりも高い値を示すことを見いだした。また、見ている物体画像のカテゴリーを、腹側高次視覚野の脳活動パターンから高精度で予測できることを示した。同様の物体カテゴリー情報を機械学習の手法を使って読み出せることも、続く研究²⁾で示されている。

物体カテゴリーの情報は、数センチメートルのオーダーで空間的に広がりをもつ脳領域に表現されているため、ボクセルパターンの解析が有効であると考えられる¹⁾。しかし、線分の方位や運動方向などの視覚的特徴は、個々の神経細胞や「コラム」とよばれる幅数百 μm 程度の神経細胞のクラスターを単位として表現されており、3 mm 角程度の解像度である fMRI で調べることは困難であるように思える。しかし、同じ特徴に反応する神経細胞やコラムが空間的な広がりをもって繰り返し不規則に配置されているとすれば、ボクセルパターンにその分布の情報が反映されていても不思議ではない。この予測に基づき、Kamitani と Tong は、異なる方位をもつ縞模様を見ているときの被験者の脳活動パターンから、見ている画像の方位を予測することに成功した³⁾。予測成績が高かった初期視覚野の V1, V2 では、特定の方位に対して選択的に反応する神経細胞がコラム構造をなして並んでいることが知られているが、この研究でも標準的なサイズのボクセル（3 mm 角）を用いており、1つのボクセル内に異なる方位選択性をもつコラムが多数含まれてしまうことになる。彼らは、ボクセルよりも小さな空間構造に表現されている情報を読み出すことができる理由の説明として、個々のボクセルに含まれる特徴選択性細胞のばらつきが重要な情報源となっているという仮説を提唱した。この仮説を巡り数多くの議論がなされたが^{7,8)}、近年の研究によって、この仮説を強く支持する結果が得られている^{9,10)}。

これらの先行研究で用いられた方法は、脳活動から知覚内容を予測する上で有用であるが、あらかじめ学習に用いられた数種類の刺激のうちどれが被験者に提示されたものか、という予測しかできないという共通の制約がある。では、学習に使った画像に限定されず、どのような画像であっても、脳活動から予測することはできないだろうか。

われわれのグループは、近年、脳活動から見ている画像そのものを画像として再構成すること（「視覚像再構成」）に成功した⁴⁾（図 1, 図 2）。画像は、ピクセルの組み合わせによって膨大な数の例を作ることができる。従来の分類にもとづくデコーディング法で任意の画像を再構成するためには、天文学的な数の画像を刺激として提示して脳活動を同定する実験が必要となる。われわれが開発した視覚像再構成の方法で最も重要な点は、画像の局所的な小領域の状態を複数のデコーダーを用いて脳活動から並列に予測し、その予測値を組み合わせることで画像全体を表現することにある。個々のデコーダーは、画像の局所的な状態（われわれの実験では画像の局所コントラスト）を予測するだけであるが、それらを組み合わせることにより、膨大な数の画

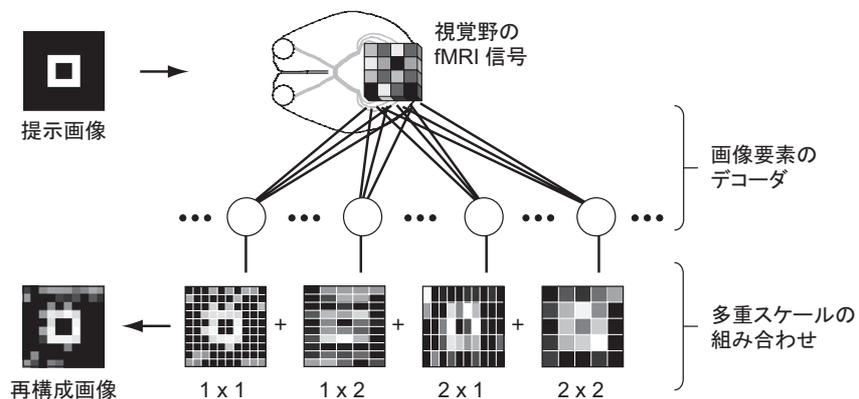


図1 視覚像再構成の方法. 視覚野の fMRI 信号と局所的な画像要素のコントラスト値の統計的関係性をデコーダが学習する. 局所的な画像要素は4つの空間スケールで冗長に定義される. これらの各スケールからの予測値を, 画像予測誤差が最小になるように組み合わせ, 再構成画像を生成する. 図中の提示画像と再構成画像のグレースケール値は, コントラスト値を表現している.

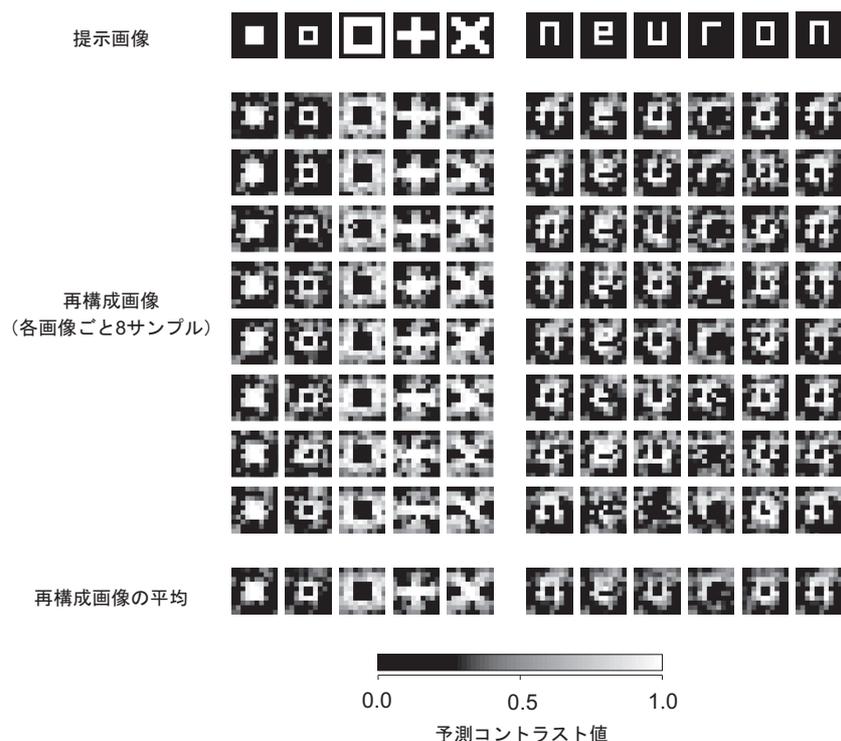


図2 視覚像再構成の結果. 最上段は提示画像, 中段は脳活動からの再構成画像, 最下段は再構成画像の平均画像を表す. 1つの画像種に対して, 8試行分実験しているので, 再構成画像は8枚分得られる. 再構成画像は, 精度の高い順に上から並べた. なお, アルファベット画像の最右列の“n”と最左列の“n”は同じ結果の複製である. 図1と同じく, 図中のグレースケール値はコントラスト値を表現している.

像のバリエーションに対応できるようになる. 要素の特徴を予測する個々のデコーダをモジュールとして考え, それらの組み合わせにより複雑な刺激や心の状態を予測する方法のことを「モジュラー・デコーディング」とよんでいる. 実験では, 10×10 のグリッドで定義されたランダムなコントラストパターン画像を合計 440 パターン観察した際の脳活動を fMRI で計測し, 初期視覚野の脳活動から画

像中の各小領域のコントラストを予測するデコーダを作成した. これらを組み合わせることにより, 1億通り以上のコントラストパターンを識別できる精度で視覚像再構成が可能であることを示した (図2).

高精度な画像再構成を可能にしたポイントは2つある. 第1に, 多重解像度表現による画像の冗長な予測である. ヒトの視覚系は, 複数の空間スケールで外界の視覚情報を

表現していることが知られている¹¹⁾。この生理学的知見に基づき、画像を4種類の異なるスケールで多重に分割し、それぞれのスケールで局所コントラストを冗長に予測している(図1)。これらの予測値は、画像内での位置に応じて予測誤差が最小となるように最適化して組み合わせられている。最適化の結果、中心視領域においては小さな空間スケールでの予測値が高い重みで使用され、周辺視領域においては大きな空間スケールでの予測値が高い重みで使用されていることがわかった。この結果は、中心視から周辺視にいたるにつれて、受容野サイズが大きくなっていくという生理学的知見とも符合する。

第2に、各小領域の状態の予測を行う上で重要なボクセルを自動的に選んで組み合わせるアルゴリズムを採用している点である。ヒトの初期視覚野には、視野の位置が相対的な位置関係を保ったまま表現されたレチノトピーという構造がある¹²⁾。これを同定するための標準的な実験手法は確立されており、どの脳部位がどの視野位置に対応するかを見いだすことはできる。しかし、これは繰り返し刺激を提示して得られたデータから推定される大まかな対応関係であり、単一の画像提示による脳活動から画像のコントラストパターンを予測することは難しい¹³⁾。視覚像再構成では、われわれが開発したスパースロジスティック回帰¹⁴⁾という機械学習アルゴリズムを用いることにより、初期視覚野の画素をどのように組み合わせるのが最適であるかを自動的に決定している。これにより、レチノトピーを同定する方法では捉えられない微細な画素のパターンを抽出して、高精度の予測を実現している。

4. 主観的な内容のデコーディング

脳情報デコーディング技術の適用範囲は、感覚刺激として提示される画像の情報を解読するのみにとどまらない。例えば、画像を実際に見ているときの脳活動を使ってデコーダーを学習し、被験者が心の中で画像をイメージしているときの脳活動に適用することで、被験者の主観的な内容の解読を行うことも可能となる。KamitaniとTongは、ヒトが視覚刺激のどの特徴に注意を向けているかを、脳活動から予測することに成功している³⁾。彼らは、被験者に45度と135度の方位をもつ2つの格子縞が重なりあった画像を提示し、いずれかの方位の格子縞にのみ注意を向けるよう教示を与え、fMRI計測を行った。この実験とは別に、45度と135度のそれぞれの方位をもつ格子画像を別々に見たときの脳活動データを使って、どちらの方位の画像を見ていたのかを予測するデコーダーを学習しておく。このデコーダーを、2つの方位をもつ格子縞が重なりあった

画像の一方の方位にのみ注意を向けているときの脳活動に適用した結果、初期視覚野の活動から注意を向けている方位を高精度に予測できることを示した。この結果は、被験者の心的内容を脳活動から解読する「マインド・リーディング」の実証例であり、視覚的意識(visual awareness)に関する初期視覚野の重要性を示唆している。

その後の研究から、画像刺激の提示を伴わないで、心の中で視覚的に想起したり、記憶として保持したりしている視覚情報に関しても、脳活動から予測できることが示された。Stokesらは、アルファベットのOあるいはXを実際に見ているときの脳活動から提示画像がどちらであったかを予測するデコーダーを学習し、そのデコーダーをOあるいはXを想起しているときの脳活動に適用して、どちらの文字を想起していたかを予測することに成功した¹⁵⁾。同様に、ある物体カテゴリーの画像を視覚的に想起した際の脳活動から、想起している物体カテゴリーを予測できるという結果も報告されている¹⁶⁾。また、HarrisonとTongは、2つの異なる方位をもつ格子画像を連続的に提示して、一方のみの記憶を保持している間の初期視覚野の脳活動から、どちらの方位を記憶しているかを解読することに成功した¹⁷⁾。

これらの先行研究は、脳の視覚野では、実際に画像を見ていない場合でも、心の中で画像を思い描いているときには、実際に画像を見ているときと類似した脳活動パターンが生じていることを示唆している。それでは、あたかも現実起こったことであるかのように錯覚するほど鮮明な視覚的経験を伴う「夢」の内容も、脳活動から解読することが可能であろうか。

われわれのグループでは、最近、実際に画像を見ているときの脳活動から、どの物体を見ているかを予測するデコーダーを構築し、それを睡眠中のヒトの脳活動データに適用することで、見ている夢の内容を解読することに成功した(図3、図4)⁵⁾。

夢見に関わる脳活動を研究する際の最大の困難は、大量のデータを収集することが難しいという点にある。一般に夢が頻繁に生じると考えられているレム睡眠は、入眠からおよそ90分経たないと現れず、この期間に見る夢を研究対象とすることは、データ収集の観点から効率が悪い。一方、最近の研究によって、入眠時にもレム睡眠中と類似した夢見体験が生じることが指摘されている¹⁸⁾。そこでわれわれは、入眠時に見る夢に注目し、fMRIを用いて入眠期の脳活動計測実験を行った(図3)。実験では、被験者にMRI装置内で脳波計測用の電極キャップを装着した状態で眠ってもらい、脳波から睡眠状態をリアルタイムに判定

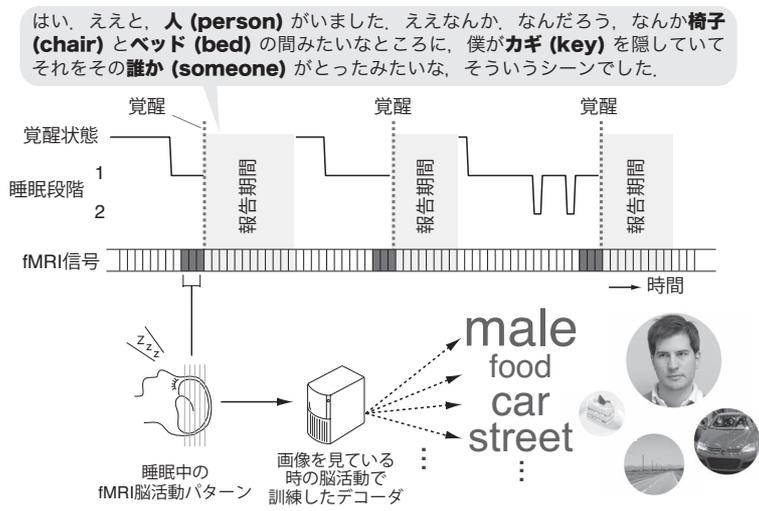


図3 夢のデコーディング研究の概要。吹き出しの中の夢の報告は実験において実際に得られたものであり、太字の部分が解析の対象となった物体や風景を表す単語である。

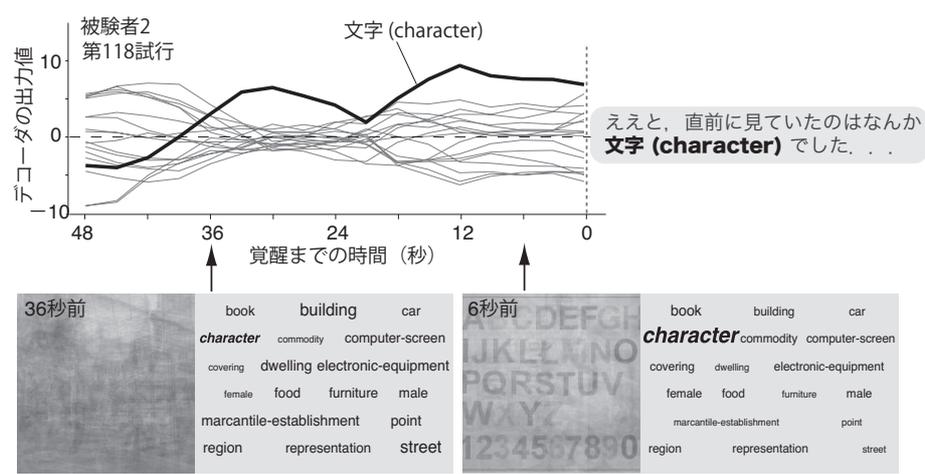


図4 デコーダの出力スコアの時間変化。下の図には覚醒の36秒前と6秒前のデコーダの出力スコアに応じてそれぞれの主要カテゴリーの画像を重ねあわせたものと、スコアに応じて文字サイズを調整したタグクラウドを示した。画像は夢で見た物体の形や色を再構成したものではなく、その夢にどのような種類の物体や風景が現れたかを示している。

しながら fMRI 計測を行った。夢見と強い関連があると知られている睡眠脳波が生じたタイミングで被験者を覚醒させ (睡眠開始から約2~3分後)、直前まで見ていた夢の内容を自由に報告させた (約30秒)。得られた報告を記録し、再び被験者に眠ってもらい、という手続きを何度も繰り返すことで、夢報告とそれに対応する脳活動データを大量に取得することに成功した (各被験者について約200回の夢報告)。

夢内容の報告を自由に行かせたため、実験で得られた夢報告の内容は非常に多岐にわたっていた。このような不定形な夢報告データから夢の内容を表すラベルを作成するために、われわれは言語データベースの「WordNet」を活用して、まず夢報告中に現れた物体や風景を表す単語を

WordNet で定義されるカテゴリーでまとめ、頻出した主要なカテゴリーを被験者ごとに特定した (各被験者につき約20個のカテゴリー)。そして、各夢報告に、これらの主要なカテゴリーが含まれているか否かを求め、各夢報告を主要カテゴリーの有無を表す二値のラベルベクトルで表現するという方略をとった。また、これらのカテゴリーに対応する画像を実際に見ているときの脳活動を使って、見ている物体を予測するデコーダを構築した。このデコーダを、各夢報告が得られた直前の睡眠中の脳活動に適用することで、見た夢の内容を脳活動から解読可能かどうかを定量的に評価した。

われわれはまず、画像を見ているときの高次視覚野の脳活動から、主要カテゴリーのペアごと (male vs. car など)

に二値判別器を構築し、覚醒させる直前(0~9秒前)の脳活動から物体の判別ができるかを検証した。その結果、夢に現れたほうのカテゴリーを高い精度で予測することに成功した。次に、各ペアについて構築した判別器の出力を組み合わせることで、各カテゴリーの有無を検出するデコーダーを構築した。このデコーダーは睡眠中の脳活動データが与えられると、各カテゴリーが存在する度合いを示す「スコア」を出力する。覚醒前の時間帯における出力スコアの時間変化を見てみると、特に覚醒直前において、報告に現れた物体や風景に対するスコアが高くなった(図4)。これは、夢報告の内容が覚醒直前の脳活動を反映していることを示している。また、報告に含まれるカテゴリーだけではなく、それと関連性が高いカテゴリー(characterに対して、bookやcomputer-screenなど)も高いスコアを示していたことから、報告はしなかったが、実際には夢に現れていた物体がデコーダーの出力に反映されている可能性が考えられる。視覚野を細かく分割して、それぞれの部位間でデコーディングの精度を比較すると、後頭葉から側頭葉にかけて広がる高次視覚野を用いた場合に、より高い精度が得られることがわかった。また、高次視覚野の中でも、人の顔や風景の画像に対して選択的に反応する脳領域の活動を用いたときには、それぞれ人や風景に関するカテゴリーに対して、高精度で予測が可能であった。以上の結果は、夢を見ているときにも、実際に画像を見ているときと類似する脳活動パターンが生じていることを示唆している。

本研究は、夢の内容が睡眠中の脳活動パターンに反映されていることを初めて実験的に証明したものである。今回の研究では、高次視覚野の脳活動から、夢に現れる物体や風景の情報を高い精度で解読できることはわかったが、夢の中に現れた色や形など、他の視覚特徴を解読できるかは明らかではない。われわれは、今後このアプローチを応用・発展させていくことで、より多様な夢内容の解読が可能かどうかを検証していく予定である。

脳情報デコーディングは、被験者の視覚世界の情報を脳活動から読み出す技術として日々進歩を遂げている。心的視覚内容の解読では、思い描いている像をありのまま高精度に再構成することはいまだ実現していないが¹³⁾、想像しているときや夢を見ているときに、初期視覚野に視覚内容の色や形などの特徴が表現されている可能性は大いに考えられる。これを検証することは、今後の研究の重要な課題であるだろう。

この技術の現在の課題点として、ヒトの脳活動計測の主

流となっているfMRIの時空間解像度の低さや計測の簡易性の低さが挙げられる。fMRIはヒトの脳活動計測機器の中では比較的高い空間解像度を有しているとはいえ、1つのボクセルの中には数100万個の神経細胞が含まれている。また、神経細胞の活動に伴う血流変化を計測しているため、時間解像度も低い。さらに、計測中、被験者が身動きをとれないという制約もある。したがって、より時空間解像度が高く、簡易性の高い計測方法の開発が、今後の技術の発展に必要となるであろう。この点に関して、近年、脳の皮質表面上に設置した電極から脳の電気信号を計測する皮質脳波の利用が進んでいる。皮質脳波計測装置は、時間解像度が高く携帯性も高いため、埋め込み型BMIシステムの脳計測基盤として、医療現場での実用化に大きな期待が寄せられている。

脳情報デコーディングは、本来ならば他人からはアクセスできない心的内容に迫ることを可能とする技術であり、適用方法次第では、倫理的に大きな問題へと発展する恐れがある。現時点の脳活動計測・解析技術では、解読可能な情報は限られたものであるが、今後の発展を見据えた上で、倫理的側面からの議論や、技術利用における法整備の検討が必要であろう。また、今後、解析技術のさらなる発展によって、すでに計測してある脳活動データから、現在想定されている以上の詳細な情報が解読可能になるかもしれない。そのため、脳計測データや個人情報の管理の徹底も同時に考えていく必要がある。

文 献

- 1) J. V. Haxby, M. I. Gobbini, M. L. Furey, A. Ishai, J. L. Schouten and P. Pietrini: "Distributed and overlapping representations of faces and objects in ventral temporal cortex," *Science*, **293** (2001) 2425-2430.
- 2) D. D. Cox and R. L. Savoy: "Functional magnetic resonance imaging (fMRI) "brain reading": Detecting and classifying distributed patterns of fMRI activity in human visual cortex," *NeuroImage*, **19** (2003) 261-270.
- 3) Y. Kamitani and F. Tong: "Decoding the visual and subjective contents of the human brain," *Nat. Neurosci.*, **8** (2005) 679-685.
- 4) Y. Miyawaki, H. Uchida, O. Yamashita, M. Sato, Y. Morito, H. C. Tanabe, N. Sadato and Y. Kamitani: "Visual image reconstruction from human brain activity using a combination of multi-scale local image decoders," *Neuron*, **60** (2008) 915-929.
- 5) T. Horikawa, M. Tamaki, Y. Miyawaki and Y. Kamitani: "Neural decoding of visual imagery during sleep," *Science*, **340** (2013) 639-642.
- 6) F. Pereira, T. Mitchell and M. Botvinick: "Machine learning classifiers and fMRI: A tutorial overview," *NeuroImage*, **45** (2009) 199-209.
- 7) H. P. Op de Beeck: "Against hyperacuity in brain reading: spatial smoothing does not hurt multivariate fMRI analyses?" *NeuroImage*, **49** (2010) 1943-1948.
- 8) Y. Kamitani and Y. Sawahata: "Spatial smoothing hurts localiza-

- tion but not information: Pitfalls for brain mappers,” *NeuroImage*, **49** (2010) 1949–1952.
- 9) D. Chaimow, E. Yacoub, K. Ugurbil and A. Shmuel: “Modeling and analysis of mechanisms underlying fMRI based decoding of information conveyed in cortical columns,” *NeuroImage*, **56** (2011) 627–642.
 - 10) A. Alink, A. Krugliak, A. Walther and N. Kriegeskorte: “fMRI orientation decoding in V1 does not require global maps or globally coherent orientation stimuli,” *Front. Psychol.*, **4** (2013) 493.
 - 11) F. W. Campbell and J. G. Robson: “Application of Fourier analysis to the visibility of gratings,” *J. Physiol.*, **197** (1968) 551–566.
 - 12) S. A. Engel, D. E. Rumelhart, B. A. Wandell, A. T. Lee, G. H. Glover, E. J. Chichilnisky and M. N. Shadlen: “fMRI of human visual cortex,” *Nature*, **369** (1994) 525.
 - 13) B. Thirion, E. Duchesnay, E. Hubbard, J. Dubois, J. B. Poline, D. Lebihan and S. Dehaene: “Inverse retinotopy: Inferring the visual content of images from brain activation patterns,” *NeuroImage*, **33** (2006) 1104–1116.
 - 14) O. Yamashita, M. A. Sato, T. Yoshioka, F. Tong and Y. Kamitani: “Sparse estimation automatically selects voxels relevant for the decoding of fMRI activity patterns,” *NeuroImage*, **42** (2008) 1414–1429.
 - 15) M. Stokes, R. Thompson, R. Cusack and J. Duncan: “Top-down activation of shape-specific population codes in visual cortex during mental imagery,” *J. Neurosci.*, **29** (2009) 1565–1572.
 - 16) L. Reddy, N. Tsuchiya and T. Serre: “Reading the mind’s eye: Decoding category information during mental imagery,” *NeuroImage*, **50** (2010) 818–825.
 - 17) S. A. Harrison and F. Tong: “Decoding reveals the contents of visual working memory in early visual areas,” *Nature*, **458** (2009) 632–635.
 - 18) Y. Nir and G. Tononi: “Dreaming and the brain: From phenomenology to neurophysiology,” *Trends Cogn. Sci.*, **14** (2010) 88–100.

(2013年9月17日受理)