

脳科学と情報科学：2つの接点

どれほど刺激的な音楽も、物理的には単なる空気の振動に過ぎない。私たちの脳はその単なる振動を処理して、豊かな経験をもたらしてくれる。その過程で私たちの脳はどのように活動しているのだろうか。また、その情報処理過程はどのように理解できるのだろうか。

この世に数多い動物の中で、ヒトが最も豊かな音楽文化を擁しているのは間違いないだろう。音楽と脳の間を研究するならば、本当はヒトの脳活動を詳細に調べるのが一番に違いない。しかしヒト神経細胞の直接記録は容易でなく、音楽との関係がよく分かっているとは言いがたい。ヒトで比較的信頼性が高く計測しやすいのは神経集団の活動と相関する血流量を観測する手法 (fMRI) で、その信号はミリメートルオーダーの空間解像度と秒オーダーの時間解像度のマクロな信号である。一方、動物を対象とすればミクロスケールの神経活動を計測できる。しかし、彼ら彼女たちにとっての音楽とはいったい何かという別の問題が発生してしまう。実際にはヒトや動物それぞれで研究が行われており、相互に補完している。

そのような脳研究において、情報科学はいまや欠かせない存在だ。その役割を分類すると、次の2種類に大きく分けられるだろう。

1. 脳活動の解析
2. 脳活動の理解 (情報処理モデル)

本稿ではこの2つの観点から、行列分解やスパースコーディング、深層学習 (Deep Learning) といった情報科学的手法が脳研究に近年使われている

例を解説する。音楽そのものに対するヒトの脳活動に加えて、音楽を構成する要素であるピッチがサルの脳でどのように形成されるのか、さらに聴覚に限らない広い脳の理解との関連にも触れる。

脳活動の解析と情報科学

私たちの神経細胞は、特にこれといった作業をしていなくても常に活動を続けている。そこで何かの音を聞くと、脳は普段とは異なる活動を示す。そして音楽を聴いた場合も、また違ったパターンが見られる。音楽に対する活動は、ほかの音に対する活動と何が違うのだろうか。

McDermott らは行列分解を利用して、特に強い仮説を用意せずにこの問いに答えた¹⁾。彼らはまず、165カテゴリーのさまざまな自然音刺激セットを用意した。被験者にはそれぞれの音刺激をMRI装置の中で聞いてもらい、その際の脳活動を記録しておく。脳活動は、聴覚関連領域を10,000個以上に分割した脳の3次元メッシュの1つ1つ(ボクセルと呼ぶ)に対して得られる。1つの音刺激に対して脳活動量を1つの値として表現すると、最終的には音の種類数×ボクセル数の行列が得られる(図-1左上)。

この活動行列をどのように使えばいいだろうか。彼らは独立成分分析に類似の非ガウス性制約付き行列分解を使い、6個の主なコンポーネントを同定した(図-1右上)。得られた基底ベクトルは、ボクセル軸で見ると脳活動の空間的パターンを示している。一方で音刺激の種類という次元で見れば、類似の脳活動を引き起こす音刺激の分布を表現している。

// 特集 // 音楽を軸に拡がる情報科学

脳活動の空間パターンを解析すると、それぞれのコンポーネントが比較的単純な音の要素（パワーやピッチ）に反応する領域、言語に反応する領域、そして音楽に反応する領域に対応することが分かった。側頭葉にある単純な音要素に反応する領域に対して、言語に反応する領域はその側方に、また音楽に反応する領域はその前方に位置していた（図-1下）。微細な構造を知ることは難しいものの、教師なしの解析で音楽に対する選択性を示す領域が明らかになった。

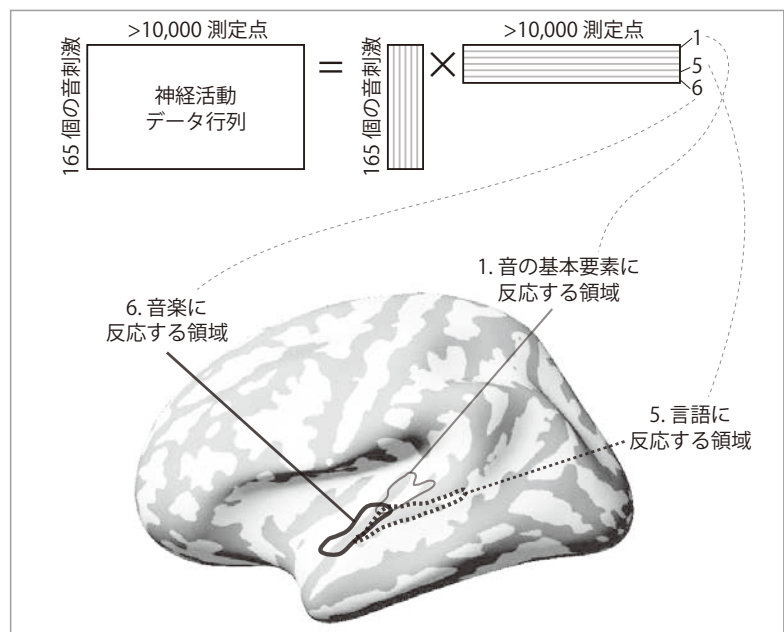


図-1 脳活動の解析と行列分解。さまざまな音刺激に対して得られる神経活動行列を分解することで、音楽や言語に反応する領域が明らかになった。文献1)をもとに筆者が作成。

脳活動の理解と情報科学

情報科学的手法は前章のようなデータ解析に威力を発揮するが、脳科学においては別の重要な役割も担っている。それは、脳の情報処理過程を理解するためのモデル化である。

脳研究の大きな目的の1つは、脳がどのような計算や学習を行っているのか理解したいというものである。多くの研究者は、脳をある種の情報処理装置として捉えている。そのような観点に立って理解を進めるということは、私たちが理解できる情報科学的手法——多くの場合、既存のよく知られている手法——との対応付けを行うことを意味する。

本章では、情報科学的手法がそのようなモデル化に利用されている例を2つ紹介する。音楽そのものの理解にはまだまだほど遠いが、音楽を構成する主要素の1つであるピッチなどの中程度に複雑な特徴に対応する神経表現がどのように獲得されているのか、ある程度の理解が進みつつある。

♪ 自然音とスパースコーディング

ある音のピッチ（音高）とは、その音を聞いたときの主観的な音の高さ [Hz] である。同じピッチ知覚はさまざまな音刺激によって引き起こされ得る。たとえば倍音で構成される音を聞くと、その基本周波

数がピッチとして知覚される。倍音ごとのパワーの組合せは色々な可能性があるが、音色は変わってもピッチ知覚は変わらない。ピッチがどのように計算されているのか、聴覚心理学では常に問題とされてきた。

ピッチに対応する計算を行っている場所の候補の1つが大腦皮質聴覚野である。近年になって、ピッチに対応する表現を示す神経細胞が聴覚野で報告されはじめた。特に興味深いのは、missing fundamental と呼ばれる非線形反応が見られることである。低次倍音を削って高次倍音だけを組み合わせた音は、基本周波数においてはパワーがないにもかかわらず、ピッチ知覚はしばしば基本周波数のままである。このような基本周波数の不変抽出と似た非線形な反応が、一部の聴覚野神経細胞で見られる。

Terashimaらは、このような基本周波数の抽出が倍音を多く含む自然音のスパースコーディングやトポグラフィック独立成分分析で説明できることを示した²⁾。倍音のさまざまな組合せを入力として受けて学習したネットワークは、まず倍音にピークを持つような基底ベクトルを学習した。さらに同じ基本周波数を持つ基底ベクトルが近くに集まり、それらが非線形に統合されることで自然と基本周波数を不変に抽出するような

「missing fundamental 細胞」が学習された。

この結果から、これらの手法が聴覚野の良いモデルだというのは言い過ぎだと思われるかもしれない。しかし、これらの手法はもともと視覚野のモデルとして提案されたものだった。視覚野と聴覚野は機能的には一見似ていないようにも見えるが、類似の解剖学的構造を持っている。自然画像を学習したモデルが視覚野に類似し、自然音を学習したモデルが聴覚野に類似する。両領域に共通する学習メカニズムを理解するには、これら情報科学的手法を共通言語として用いることが本質的である。

♪ 自然音と深層学習

ピッチの計算も重要だが、音楽の理解にはまだまだ距離がある。より複雑で高次な特徴の脳での表現は、どのようなモデルで理解できるだろうか。

視覚研究において、Yaminsらは深層学習がヒト視覚野の階層構造の良いモデルとなり得ることを示した³⁾。Kellらはこの結果をベースとして、自然音を学習する畳み込みニューラルネットワーク(CNN)を使って聴覚関連領域の階層性をモデル化しつつある⁴⁾。彼らは、ノイズを加えた音声のスペクトログラムを入力として与え、音声認識タスクでCNNをトレーニングした。

その結果、下位層には周波数などの低次特徴が、中間層にはピッチなどの中程度に複雑な特徴が、そして上位層にはより複雑な特徴が学習された。これら層の反応は、ヒトfMRIで得られる活動をよく説明することができ、さらに階層構造を見ることができる。すなわち、低次聴覚野はCNNの下位層を、高次聴覚野は上位層を使って最もよく説明することができる。前章と同様に視覚系と聴覚系が同様の情報科学的手法を用いて説明されており、両者を理解するためのモデルとして有望である。

係を明らかにしようとする近年の試みを解説した。情報科学的手法は脳活動の解析に役立つのみならず、脳の情報処理機構を理解しようとする際の枠組みをも提供する。行列分解、スパースコーディング、深層学習といった手法を活用して、ヒトのマクロな脳活動や動物のミクロな脳活動の理解が進んでいる。

音楽に関連する脳活動は、ほかの機能とは無関係なのだろうか。今回紹介した研究はいずれも、視覚研究と深く関係している。大脳新皮質は聴覚や視覚といったさまざまな機能を持つが、共通の解剖学的特徴は何らかの計算原理の存在を示唆している。したがって、視覚野の研究と聴覚野の研究はお互いに良い影響を与え得る。特に音楽の研究は、聴覚で重要とされるリズムと深くかかわっており、その原理が解明されればほかの感覚研究に与える影響は大きいだろう。

複数の感覚モダリティに共通の機能を記述しようとする際、各モダリティに固有の表現を使うことはできない。その意味で、普遍的な表現を与える情報科学的手法の重要性はますます高まっている。脳と音楽の関係はまだまだ分からないことも多いが、他領域の研究との接点から新展開がもたらされるだろう。

参考文献

- 1) Norman-Haignere, S., Kanwisher, N. G. and McDermott, J. H. : Distinct Cortical Pathways for Music and Speech Revealed by Hypothesis-free Voxel Decomposition. *Neuron*, Vol.88, pp.1281-1296 (2015).
- 2) Terashima, H. and Okada, M. : The Topographic Unsupervised Learning of Natural Sounds in the Auditory Cortex. *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, pp.2321-2329 (2012).
- 3) Yamins D. L. K., Hong, H., Cadieu, C., Solomon, E. A., Seibert D. and DiCarlo, J. J. : Performance-Optimized Hierarchical Models Predict Neural Responses in Higher Visual Cortex. *Proceedings of the National Academy of Sciences*, Vol.11, No.23, pp.8619-8624 (2014).
- 4) Kell, A., Yamins, D. L. K., Norman-Haignere, S. and McDermott, J. H. : Speech-trained Neural Networks behave Like Human Listeners and Reveal a Hierarchy in Auditory Cortex. *Cosyne Abstracts 2016*, pp.109-110, Salt Lake City, USA (2016).

(2016年3月1日受付)

寺島裕貴 terashima.hiroki@lab.ntt.co.jp

2009年、東京大学理学部情報科学科卒業。2011～14年、日本学術振興会特別研究員(DC1)。2014年、同大学院新領域創成科学研究科複雑理工学専攻博士課程修了。同年、NTTコミュニケーション科学基礎研究所入所。聴覚などの計算論的モデリングに従事。博士(科学)。

情報科学的手法を用いた 大脳皮質の理解に向けて

本稿では、情報科学的手法を用いて脳と音楽の関