

大脳皮質聴覚野ピッチ細胞は 視覚野複雑細胞と計算論的に相同か？

寺島 裕貴^{1,2,a)} 岡田 真人^{1,3,b)}

概要：大脳皮質聴覚野が担う複雑な情報処理の原理解明に向け、皮質の一様性を根拠に視覚系相同領野のモデルが応用されつつある。しかし、既存研究で議論されてきたのは視覚野神経細胞のうち線形応答を示す単純細胞のモデルに限られており、非線形な複雑細胞に相当する概念については未着手だった。本研究で我々は、視覚野複雑細胞の学習モデルであるトポグラフィック独立成分分析を自然画像の代わりに自然音に適応させた。学習結果である聴覚版「複雑細胞」の特性は、近年聴覚野で発見されたピッチ細胞と同様に“missing fundamental”類似の非線形性を示した。聴覚野ピッチ細胞が視覚野複雑細胞と同一のモデルによる自然刺激学習で獲得可能であるという結果は、両者の計算論的相同性を示唆する。

1. はじめに

大脳皮質一次聴覚野 (A1) は末梢聴覚系の出力を主に受け複雑な情報処理を行なっているが、その原理は不明で計算論的モデル研究が未発達である。一方、視覚系における相同領野である一次視覚野 (V1) では盛んにモデル研究が行われてきた。近年、両者の解剖学的類似性に着目して V1 単純細胞の学習モデル [1] を A1 神経細胞のモデル化に応用する試みが成功しつつある [2], [3], [4]。これらの結果は、A1 と V1 が学習則を共有しつつ、それぞれへの入力である自然音と自然画像に対して適応していることを示唆している。

A1 と V1 の相同性に関するこれまでの議論は、刺激に対し線形応答する V1 単純細胞のモデルを用いたものに限られてきた。しかし実際には、V1 にも A1 にも非線形性を示す神経細胞が多く存在し、これらの対応関係は依然として不明である。非線形応答する神経細胞として最も確立された概念は、V1 の複雑細胞である。複雑細胞は V1 研究の最初期 [5] から知られ、単純細胞と同様に特定方位の光線分選択的に反応する。しかし、単純細胞が刺激に対し線形に反応するのは対照的に、位置や位相のずれに対して応答が不変という非線形性を示す。この特性は、同一方位・異

位相に選択性を持つ複数の単純細胞の活動を統合することでモデル化され [6]、単純細胞の場合と同様に自然刺激の学習で獲得され得る [7]。このように複雑細胞は 50 年の時を経て視覚野研究の中では確立された概念だが、不思議なことに他の感覚野における対応概念については議論されてこなかった。もし聴覚野にも「複雑細胞」が存在するなら、刺激のどのような特性に対して不変性を示すのだろうか。また、その細胞の特性は V1 と同様に自然刺激学習の結果獲得され得るのだろうか。

異なるモダリティの実験結果を直接比較することは難しいが、モデル上であれば議論可能である。そこで本研究では過去の研究が示唆する学習則の共通性に基づき、V1 複雑細胞のモデルであるトポグラフィック独立成分分析の学習則をそのまま用い、自然画像の代わりに自然音を入力することで A1 版「複雑細胞」を学習させた。その特性を調べた結果、「複雑細胞」の一部が“missing fundamental”類似の非線形性を示し、近年 A1 で発見されたピッチ細胞 [8] に酷似していることを発見した。この結果は A1 ピッチ細胞と V1 複雑細胞の計算論的な相同性を示唆し、両領野の統一的な理解に寄与する。

2. モデル

本研究で扱うモデルはトポグラフィック独立成分分析 [7] の過完備拡張 [9] である (以後 TICA と呼ぶ)。これは 2 層ニューラルネットワークによる教師なし学習モデルで、V1 の単純細胞と複雑細胞、そしてそれらのトポグラフィ (地図) のモデルとして提案された。画像を入力として受け取り、第 1 層のユニットはこれに対し応答する単純細胞をモ

¹ 東京大学 新領域創成科学研究科 複雑理工学専攻
Dept of Complexity Sci and Eng, The University of Tokyo

² 日本学術振興会 特別研究員 DC1
Japan Society for the Promotion of Science

³ 理化学研究所 脳科学総合研究センター
RIKEN Brain Science Institute

a) teratti@teratti.jp

b) okada@k.u-tokyo.ac.jp

デル化している。ユニットのトポグラフィは両層間の結合を定める近傍関数により定義される。第2層のユニットは複雑細胞のモデルであり、その近傍にある第1層ユニットの活動を非線形に統合する。

白色化済みの入力 $\mathbf{I}(x) \in \mathbb{R}^d$ ($d < N$) をある N 個のベクトル $\mathbf{a}_i \in \mathbb{R}^d$ からなる基底で再構築することを考える。

$$\mathbf{I} = \sum_i s_i \mathbf{a}_i \quad (1)$$

ここで $s_i \in \mathbb{R}$ が第1層のユニットの活動にあたる。この第1層の活動を用いて、第2層の活動 $c_i \in \mathbb{R}$ を以下のように定義する。

$$c_i = \sum_j h(i, j) s_j^2 \quad (2)$$

ここで $h(i, j)$ はトポグラフィを定義する近傍関数で、 i と j が近傍であれば1を、さもなければ0を返す。本研究の2次元地図では、正方形型の 3×3 近傍窓を用いた。第1層の活動 s_i は以下のエネルギー関数最小化で定める。

$$E = -\log L(\mathbf{I}; \{\mathbf{a}_i\}, \{s_i\}) \quad (3)$$

$$= \left| \mathbf{I} - \sum_i s_i \mathbf{a}_i \right|^2 - \lambda \sum_i G(c_i) \quad (4)$$

第1項は再構築の誤差最小化を、第2項は第2層活動のスパース性を要請する。 λ はスパース性の相対的な寄与率を表し、 $G(c_i) = -\sqrt{\epsilon + c_i}$ である。最小化は以下の勾配を用いて最急降下法で行った。

$$\Delta s_i \propto \mathbf{a}_i^T \left(\mathbf{I} - \sum_j s_j \mathbf{a}_j \right) - \lambda s_i \left(\sum_j h(i, j) g(c_j) \right) \quad (5)$$

ここで $g(c_i)$ は $G(c_i)$ の微分で、 s_i の初期値は \mathbf{I} と \mathbf{a}_i の内積とした。基底 \mathbf{a}_i の学習は以下の対数尤度最大化とし、256個の入力ごとに以下の式で基底を更新した。

$$\log L(\mathbf{I}; \{\mathbf{a}_i\}) = \sum_i G(c_i) \quad (6)$$

$$\Delta \mathbf{a}_i = \eta \left\langle s_i \left(\mathbf{I} - \sum_j s_j \mathbf{a}_j \right) \right\rangle \quad (7)$$

演算子 $\langle \dots \rangle$ は繰り返し平均を表す。本研究で用いたパラメータ値は $\epsilon = 0.005$ (数値計算安定性のため)、 $\lambda = 0.91$, $\eta = 0.08$ である。

3. 結果

3.1 自然音を用いた学習

A1版「複雑細胞」をモデル上で検討するために、まずV1複雑細胞の学習モデルであるTICAを自然画像の代わりに自然音に対して適応させた。用いた自然音は、IPAハンド

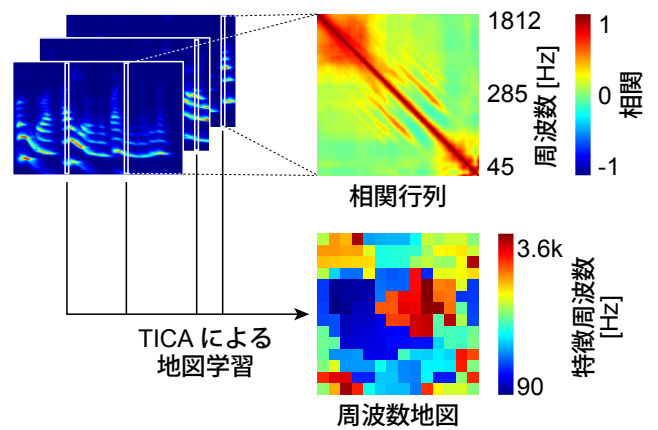


図1 遠相関を持つ自然音とその学習結果の周波数地図
Fig. 1 Natural sound statistics and the learned tonotopic map.

ブック [10] 収録の多言語ヒト音声データベースである。ヒト音声を用いたのは、ヒト音声への適応がA1神経細胞の特性をよく説明するという知見があるためである [2], [11]. 8 kHz にダウンサンプルした後、NSL toolbox [12] で末梢系の出力を模擬したスペクトログラムを作成した。例を図1左上に示す。スペクトログラムから128ピクセルの対数周波数スペクトル (24 pixel = 1 オクターブ) をランダム抽出した ($N = 100,000$)。

図1右上に、本研究で用いた自然音から生成した相関行列を示す (4 kHz にダウンサンプル)。近周波数間の相関を示す対角成分のみならず、これに平行な非対角成分が見られる。これはハーモニックな関係にある (単純な整数比で書ける) 離れた周波数間の相関を示している。このような遠相関は自然画像には見られない構造で、自然音に特徴的である [4]。

この自然音をPCAで100次元に圧縮し、TICAの入力に用いて学習させた。第1層ユニット (単純細胞) の周波数選択性地図 (トノトピー) を図1右下に示す。様々な周波数の純音入力に対する活動を計算し、最も強く活動する純音周波数をそのユニットの特徴周波数とした。トポグラフィは 14×14 の2次元トーラス ($2 \times$ 過完備) とした。既報通り、V1モデルとして自然画像を学習させた場合の視野再現地図よりも乱雑な周波数地図が獲得され [13]、この対照は近年の生理学的知見と符合する [14], [15]。

3.2 ピッチ選択的ユニットの出現

図1のように、自然音ではハーモニックな関係にある離れた周波数間に強い統計的依存性がある。聴覚心理学では、ハーモニーに関連した重要な知覚としてピッチ (音高) が知られている。ある音のピッチとは、その音を聴いた時に知覚される主観的な音の高さ (単位: Hz) のことで、倍音の組み合わせからなるハーモニックな聴覚刺激の場合には基本的にその基本周波数 f_0 [Hz] にあたる。興味深い現象として missing fundamental が知られており、 f_0 や $2f_0$

表 1 V1 複雑細胞と A1 ピッチ細胞の類比
 Table 1 An analogy between A1 pitch cells and V1 complex cells.

	反応	選択性	不変性	学習モデル
A1 ピッチ細胞	非線形	特定ピッチのハーモニー音	スペクトル	TICA + 自然音
V1 複雑細胞	非線形	特定方位の光線分	位相	TICA + 自然画像

といった低い倍音成分が欠けてもピッチ知覚は f_0 のまま不変である [16]. 例えば, 図 2A に示す多様なスペクトルはいずれも共通の f_0 というピッチ知覚を誘導する. このように多様なスペクトルから不変量を抽出するには非線形演算が必要だが, これを実現する神経メカニズムは未だ不明である.

近年, サルの A1 で missing fundamental 類似の非線形反応を示すピッチ細胞が報告された [8]. このニューロンは純音に対して f_0 選択的に反応するだけでなく, f_0 を基本周波数とするハーモニック刺激にも, また低次ハーモニックのパワーを欠く様々な missing fundamental 刺激に対しても反応するという非線形性を示した. また, このニューロンは A1 のトノトピー上で低周波領域にのみ見つかった.

この実験 [8] に倣って MF 刺激 (missing fundamental complex tones) を入力した時の活動度をモデル上で調べた. 入力の実験に倣って, 3 つの倍音成分から構成され, 図 2B に示すように最低ハーモニック次数が 1 から 10 までとした. 各ユニットに対する f_0 には, そのユニットの特徴周波数の前後 0.2 オクターブ (5 パターン) を試した. 活動度は, 与えた全 MF 刺激に対する最大の反応を 1 として正規化し, 最低次ハーモニックが f_0 から $4f_0$ の MF 刺激全てに対して強く反応 (> 0.4) したユニットをピッチ選択的ユニットとした.

その結果, 第 2 層の一部にピッチ選択的ユニットが見つかった [13]. 図 3A にそれらの MF 刺激に対する反応を示す ($n = 66$; シミュレーション 6 回; 縦棒は標準偏差). これらのユニットは, 最低ハーモニック次数が 6 前後までである程度反応し, それ以上では 0 付近に落ちるといった実験と同様の傾向を示す [8]. このようなユニットは第 1 層には見られなかった. また, 図 3B はピッチ選択的ユニットの位置を第 1 層のトノトピー (平滑化済) 上に重ねて示したものである. ピッチ選択的ユニットはトノトピー上で低周

波領域に集中しており, これもまた実験と一致する結果である. 第 2 層ユニットは元来 V1 複雑細胞のモデルであるから, この結果は, A1 で見られる非線形なピッチ選択的反応が V1 の複雑細胞と同様の学習アルゴリズムで獲得されている可能性を示唆する.

3.3 ピッチ選択性のメカニズム・V1 複雑細胞との類比

前節で見たピッチ選択的ユニットが示す missing fundamental 類似の非線形応答はどのようなメカニズムで実現されているのだろうか. モデル第 2 層の活動は, 近傍の第 1 層ユニットの活動を統合して計算される. 統合元の特性を調べるために, 同一ピッチ選択的ユニットの近傍内にある第 1 層ユニット同士でペアを作り, その特徴周波数差のヒストグラムを図 4A に示した. 図 4B はそれ以外の全ペアについてのヒストグラムである (図 3 と同じデータを用いて解析し, 3.0 オクターブ以上の差は省略した). 図 4A に見られるピークは, 以下のようにそれぞれ 2:3, 1:2, 1:3 というハーモニックな関係 (単純な整数比) に対応する.

- $0.59 = \log_2 1.5$
- $1.00 = \log_2 2.0$ (0 に次ぐ最大ピーク)
- $1.59 = \log_2 3.0$

これは, 実際の A1 ピッチ細胞が情報を受け取る近傍単純細胞の特徴周波数もハーモニックな関係に偏っていることを示唆する. 実験論文 [15] の図 5d を精査すると同様の傾向があるようにも見えるが, ピッチ細胞とそれ以外の区別が無く, さらに用いられた刺激が 0.25 オクターブ刻みでハーモニックな関係になりやすいバイアスもあり判断が難しい. この予測はより精密な実験で検証する必要がある.

以上のようにピッチ細胞は, ハーモニックな関係にある周波数に選択性を持つ単純細胞の活動を統合することで,

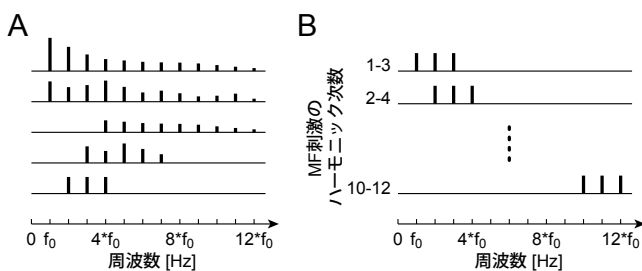


図 2 同じ高さのピッチ知覚を誘導する多様なスペクトル
 Fig. 2 Various spectra that induce a constant pitch.

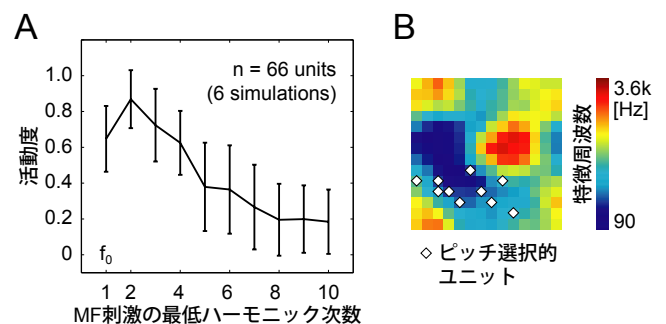


図 3 ピッチ細胞類似の非線形反応を示す A1 版「複雑細胞」
 Fig. 3 A1 “complex cells” resemble the nonlinear pitch cells.

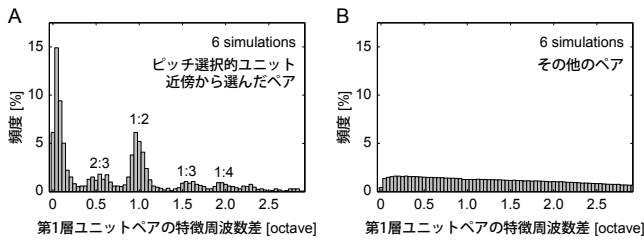


図 4 ピッチ選択的ユニットの近傍から抽出した第 1 層ユニットペアの特徴周波数はハーモニックな関係に偏る

Fig. 4 Harmonic relationships between characteristic frequencies of two first-layer units that are in the neighbourhood of a pitch selective unit.

図 2 のように多様なスペクトルが共有する基本周波数に対して頑健な選択性を獲得している。すなわち、あるピッチに対して選択性を示しつつ、多様なスペクトルに対して不変な非線形反応を示す。V1 複雑細胞との類比を表 1 にまとめた。V1 複雑細胞は特定方位の光線分に対して選択性を持ち、位相変化に対して不変という非線形性を示す。一方の A1 ピッチ細胞は、特定ピッチのハーモニックな音刺激に対して選択性を持ち、基本周波数 f_0 を共有するスペクトルの多様な変化に対して不変という非線形性を示す。

このような類比に加えて、本研究が示したように A1 ピッチ細胞の特性は V1 複雑細胞の学習モデルを通じて獲得可能であることから、両者の計算論的な相同性が示唆される。これは両領域の統一的理解の基盤となり、相対的に未発達な聴覚系モデルの今後の発展に寄与すると考えられる。例えば、視覚系で発達している多層モデルは大脳皮質の知見を基に単純細胞と複雑細胞を交互に重ねるアーキテクチャを共有しているが、聴覚系ではそもそも複雑細胞の相当概念が不明だったため多層モデルは未発達である。本研究の結果は聴覚版「複雑細胞」を理解するための初めての足掛かりであり、聴覚多層モデルの礎として有力な候補に成り得る。

4. まとめ

既存の計算論的研究は A1 と V1 による学習則の共有を示唆しているが、議論は V1 単純細胞相当の範囲に限られてきた。本研究は、視覚野でのみ確立された概念である複雑細胞の聴覚野における相同概念を検討すべく、複雑細胞学習モデルである TICA を自然音に適応させた。その結果、聴覚版「複雑細胞」の一部は“missing fundamental”類似の非線形を示し、A1 で近年発見されたピッチ細胞の特性と酷似した。A1 ピッチ細胞が同一ピッチの様々なスペクトルに対して示す不変性は、V1 複雑細胞が同一方位光線分の様々な位相に対して示す不変性と類比される。単一学習アルゴリズムで獲得可能であることと合わせ、両者は計算論的に相同であることが示唆される。

参考文献

- [1] Olshausen, B. A. and Field, D. J.: Emergence of simple-cell receptive field properties by learning a sparse code for natural images, *Nature*, Vol. 381, No. 6583, pp. 607–609 (1996).
- [2] Klein, D. J., Konig, P. and Kording, K. P.: Sparse spectrotemporal coding of sounds, *EURASIP Journal on Applied Signal Processing*, Vol. 2003, No. 7, pp. 659–667 (2003).
- [3] Terashima, H. and Hosoya, H.: Sparse codes of harmonic natural sounds and their modulatory interactions, *Network: Computation in Neural Systems*, Vol. 20, No. 4, pp. 253–267 (2009).
- [4] Terashima, H., Hosoya, H., Tani, T., Ichinohe, N. and Okada, M.: Sparse coding of harmonic vocalization in monkey auditory cortex, *Neurocomputing*, Vol. 103, pp. 14–21 (2013).
- [5] Hubel, D. H. and Wiesel, T. N.: Receptive fields, binocular interaction and functional architecture in the cat’s visual cortex, *The Journal of Physiology*, Vol. 160, No. 1, pp. 106–154 (1962).
- [6] Adelson, E. H. and Bergen, J. R.: Spatiotemporal energy models for the perception of motion, *Journal of the Optical Society of America A*, Vol. 2, No. 2, pp. 284–299 (1985).
- [7] Hyvärinen, A. and Hoyer, P. O.: A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images, *Vision Research*, Vol. 41, No. 18, pp. 2413–2423 (2001).
- [8] Bendor, D. and Wang, X.: The neuronal representation of pitch in primate auditory cortex, *Nature*, Vol. 436, No. 7054, pp. 1161–1165 (2005).
- [9] Ma, L. and Zhang, L.: A hierarchical generative model for overcomplete topographic representations in natural images, *International Joint Conference on Neural Networks (IJCNN2007)*, pp. 1198–1203 (2007).
- [10] International Phonetic Association: *Handbook of the International Phonetic Association: A Guide to the Use of the International Phonetic Alphabet*, Cambridge: Cambridge University Press (1999).
- [11] Saxe, A. M., Bhand, M., Mudur, R., Suresh, B. and Ng, A. Y.: Unsupervised learning models of primary cortical receptive fields and receptive field plasticity, *Advances in Neural Information Processing Systems 24 (NIPS2011)*, pp. 1971–1979 (2011).
- [12] Chi, T. and Shamma, S.: NSL Matlab Toolbox, <http://www.isr.umd.edu/Labs/NSL/Software.htm> (2003).
- [13] Terashima, H. and Okada, M.: The topographic unsupervised learning of natural sounds in the auditory cortex, *Advances in Neural Information Processing Systems 25 (NIPS2012)*, pp. 2321–2329 (2012).
- [14] Bandyopadhyay, S., Shamma, S. A. and Kanold, P. O.: Dichotomy of functional organization in the mouse auditory cortex, *Nature Neuroscience*, Vol. 13, No. 3, pp. 361–368 (2010).
- [15] Rothschild, G., Nelken, I. and Mizrahi, A.: Functional organization and population dynamics in the mouse primary auditory cortex, *Nature Neuroscience*, Vol. 13, No. 3, pp. 353–360 (2010).
- [16] Moore, B. C. J.: *An introduction to the psychology of hearing*, London: Emerald Group Publishing Ltd., 5th edition (2003).