

4章 脳の画像・音声処理戦略を解き明かすスパースモデリング

寺島 裕貴[†]

キーワード：スパースモデリング、スパースコーディング、大脳新皮質、自然刺激統計性、一次視覚野、一次聴覚野

1. まえがき

20XX年、地球に降り立った異星人は「テレビ」と呼ばれる奇妙な箱を発見し、興味を抱く。彼らはある部品を取り除くと映像が消えることに偶然気付き、映像処理ユニットを同定する。続けて音声処理ユニットも明らかになり、より深部の解明へと向かっていく——

できの悪いSFのあらすじのようだが、脳研究の実際にはこれと似たところがある。例えば戦争で損傷した脳部位と症状を対応付けていけば、この神経核は視覚系、あの神経核は聴覚系、といったことがわかる。しかし脳研究が真に難しいのは、われわれの興味が「学習」という動的な性質にあることだ。脳が高性能な学習器であることに異論はなく、もしその原理が理解できたなら波及効果は計り知れない。脳の学習という側面を人類が理解できる形で切出すためには、どのようなアプローチが適切なのだろうか。

本稿では、脳の学習様式を理解する試みの一例として大脳新皮質の画像・音声処理系におけるスパースモデリングの事例を紹介したい。興味深いことに、大脳新皮質の画像・音声表現には、人類が画像・音声処理技術として開発してきた高効率な（圧縮できる）表現との類似点が多い。どうやら脳は自然界が持つ統計的特徴を何らかの方法で上手に抽出していて、この「賢い」処理はヒトに至る進化の過程で獲得されたい。この学習戦略をわれわれが理解するためには、単に脳を構成する部品の性質を追うだけでは不十分で、明らかになった知見をうまく説明できる学習モデルを並行して探索する必要がある。その有望な候補がスパースモデリングだ。

桑谷らによる本特集第3章¹⁾では、観測データの生成過程が明らかな系においてはベイズ推論が有用であることを議論した。しかし、われわれがこの世に生を受けたまさにその時、自然界で触れることになる映像や音声といった刺激の生成メカニズムなど知る由もない。未知の構造を持つ

大量のデータにさらされるという状況は、殊に近年騒がれる「ビッグデータ」時代における人類の課題でもある。このような状況に生物がいかなる解決策を編み出してきたのか、その探求を無価値とは言えまい。学習の具体的なアルゴリズムについては庄野らによる本特集第2章²⁾を参照していただくこととして、ここでは神経科学と機械学習がどのように結びつくのか、その面白さを少しでも感じていただければ幸いである。

2. 大脳新皮質の謎：多様性と普遍性

もしも誰かが「あらゆる画像処理と音声処理を都合良く行える万能アルゴリズムが存在する」と学会で発言したなら一笑に付されるだろう。ノーフリーランチ定理を持ち出すまでもないし、画像処理も音声処理もそれぞれ固有の発展を遂げてきた大分野だ。それでは、こう言ったらどうだろうか？「自然界に出てくるほとんどすべての画像・音声を適切に処理できる単一アルゴリズムが存在する」——これでもまだほとんど妄想といったところだが、淡い希望を抱かせてくれる器官が実は私達の頭の中にある。それが大脳新皮質だ。

大脳新皮質は、ふつう「脳」と言われてほんやりと思い浮かべるだろう数々の表面に対応する薄いシート状の器官で、ヒトで特に大きく発達している。言語、推論、計算、社会性、…といった、ほとんどあらゆる「賢さ」に深く関わり、これらの機能は膨大な数の神経細胞が絡み合った複雑なネットワークによって実現されているというのが定説だ。これまでの脳研究により、大脳新皮質は多くの小領域（領野）に分かれ、各々が特有の機能を担っていることが知られている（図1左）。例えばこの領野は視覚野、あの領野は聴覚野、といった具合だ。その中でも視覚系であれば低次処理を担う順に一次視覚野（V1）、二次視覚野（V2）、…と細かく同定され、領野間の接続様式や階層構造が明らかになっている。

大脳新皮質の謎は、この多様な機能を呈す領野が、構造という観点からはかなり均質に見えることにある。どの領野も6層構造になっており、層間の結合様式も他器官との関係も非常に似ている（図1（右））。そのため、この多様な

[†] NTTコミュニケーション科学基礎研究所
"Sparse Modelling for Image and Sound Processing of the Neocortex" by
Hiroki Terashima (NTT Communication Science Laboratories, Atsugi)

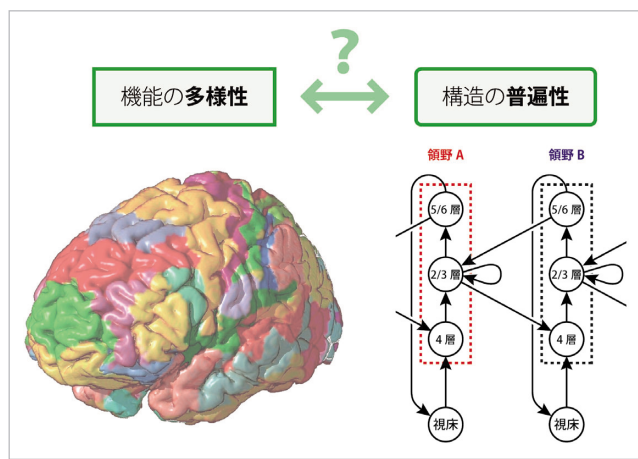


図1 大脳新皮質の多様性と普遍性

大脳新皮質は多くの小領域に分かれ、それぞれ多様な機能を担っている(左)。しかしこの多様な領域は、構造的にはどれも非常によく似ている(右)。何らかの普遍的な計算原理が存在して「賢さ」を支えているのかもしれない。

機能の裏には何らかの普遍的な計算原理——「万能」アルゴリズム——が存在するのではないかと、大脳新皮質研究の初期から漠然と考えられてきた。とはいえ現状ではわからないことが圧倒的に多く、夢物語の枠を出ていないというのが実際のところだ。この多様性と普遍性という一見するところの矛盾を解くのが大脳新皮質研究の大きな課題だが、実際の機能研究のほとんどは、それぞれの単一領域について別個に行われてきている。次節ではその具体例として、最も理解が進んでいる視覚野をまず紹介する。

3. 脳の画像処理とスパースモデリング

あなたがこの文章を読んでいる時、脳では何が起きているのだろう。まず目に光が入ってきて焦点を結び、網膜に文字が映し出される。ここで光が神経活動に変換され、視床と呼ばれる器官で中継された後、大脳新皮質の一次視覚野(V1)と呼ばれる領域に到達する(図2)。1958年、HubelとWieselはネコ視覚野の神経活動を計測中、特定の傾きを持った短い光線分をある視野領域に呈示すると激しい活動が現れることに気付いた³⁾(彼らは1981年にノーベル生理学・医学賞を受賞)。後の詳細な解析により、その反応パターンは2次元のGabor関数(Gaussianと三角関数の積)でよく近似できることがわかっている(図3)。個々のV1神経細胞は視野の特定小領域にある特定の傾きの光刺激に選択性を持っており、その多様な集合としてV1全体で原画像を符号化している。つまり、網膜表現がピクセル単位でのナイーブな画像表現だとすると、V1における表現は小さな線分を基本とする別基底(Gabor関数)を用いた表現になっている。

神経生理学者の精力的な解析によりV1神経細胞の刺激選択性が詳細に同定されていく一方で、生物が「なぜ」その刺激選択性を獲得するに至ったかという疑問に対して良い

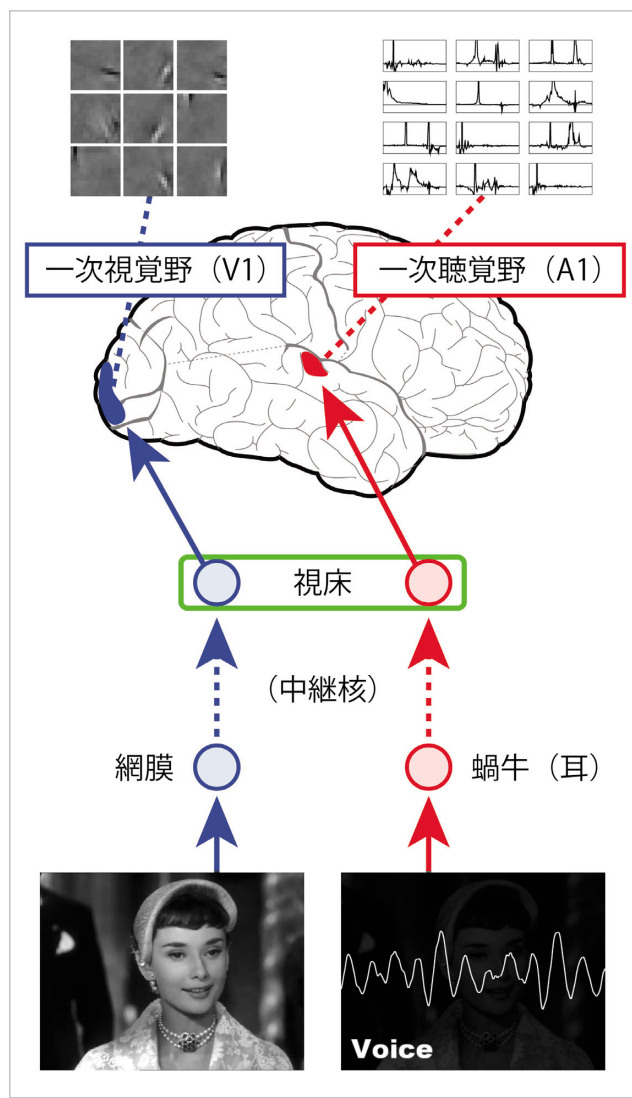


図2 一次視覚野(V1)と一次聴覚野(A1)

視覚・聴覚刺激が変換された神経活動はそれぞれ、V1とA1において初めて大脳新皮質に到達する。V1神経細胞は短い光線分に選択的に反応し、その特性はGabor関数で記述される(詳細は図3)。A1神経細胞は周波数等に複雑な選択性を示す。

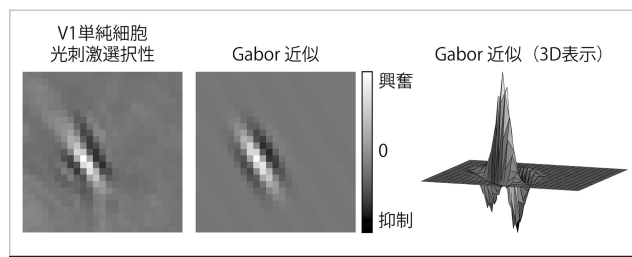


図3 V1神経細胞刺激選択性のGabor近似

個々のV1単純細胞の光刺激選択性は、図左のように描かれる。正方形が視野内の小領域に対応し、個々のピクセルでの光呈示がこの神経細胞を興奮させたり抑制したりする。V1単純細胞は特定の位置・特定の傾きを持つ光線分刺激に反応し、刺激選択性は2次元Gabor関数でよく近似される(中・右)。

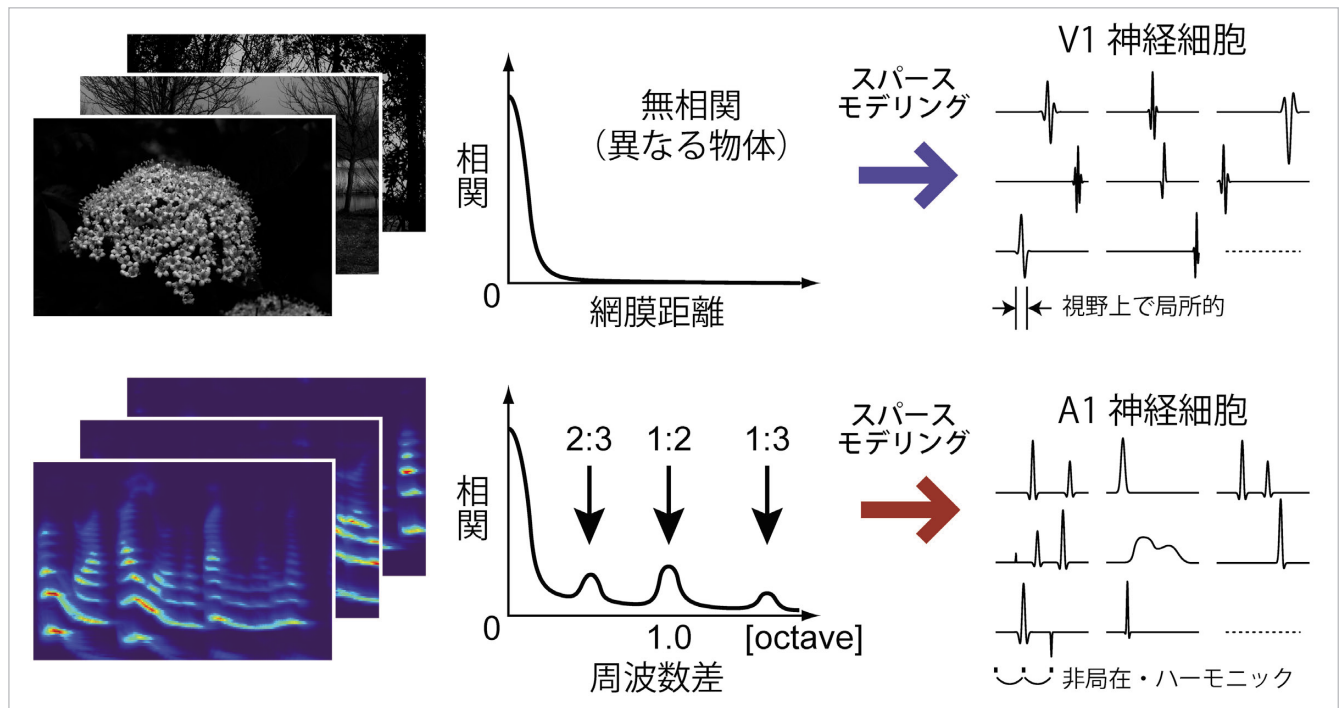


図4 スパースモデリングはV1とA1の特性を自然刺激への適応として説明する

スパースモデリングによれば、V1単純細胞の特性(右上)は自然画像(左上)に対する「最適」な基底である。また、A1の一見複雑な特性(右下)も同じアルゴリズムの自然音(左下)への適応として説明できる。V1とA1の違いは学習則に依るものではなく、自然刺激統計性の差異(中列：詳細は図5)の反映だと考えることができる。

解釈が出るまでにはやや時間を要した。テレビのトランジスタひとつの画像応答特性がわかっただけでは、画像処理ユニット全体の設計思想を理解できたとは言えないだろう。1996年、OlshausenとFieldはスパースコーディング仮説を発表した⁴⁾。このモデル研究によると、V1神経細胞の反応特性は自然画像に対して「最適化」された表現になっている。自然界に出てくる画像は各ピクセル値が完全にランダムに決まっているのではなく、物体の境界以外は概ね滑らかな変化を示す。もし少ないビット数で(神経活動量を抑えつつ)画像の情報を落とさずに表そうとするなら、最優先でエッジを表現すると効率が良い。彼らは自然画像を大量に用意し、それらの「最適」な(スパースな：少ない数の神経細胞だけが活動するような)表現を与える良い基底を探索する最適化問題を解いた*1。このシミュレーションで得られた基底は様々なエッジを表すフィルタになっており、V1神経細胞で知られるGabor基底と驚くほどよく似ていた(図2(左上))。ということは、V1で見られる刺激選択性は自然画像を入力とした最適化問題を解いた(自然画像に適応した)結果なのではないだろうか(図4(上))。

この“efficient coding”仮説は単体でも魅力的だが、その拡張もまた大きな成功を収めてきた。上記はV1神経細胞の中でも線形応答を示す単純細胞を対象にした理論だが、

位相に対する不変性を示す非線形な細胞である複雑細胞もまた、単純な拡張で再現できる⁵⁾。また、個々の神経細胞の反応特性だけでなく、それらが皮質表面でどのように配置されているかという地図構造についても、同様の理論で神経生理学的知見に解釈を与えることができる⁵⁾(モデルの詳細は本特集第2章²⁾を参照)。さらに、近年俄に注目を浴びている深層学習(deep learning)の基本構造は、大雑把に言ってしまうならこのスパースコーディングモデルを何層も重ねたものである。この階層化もまた大脳新皮質の構造にヒントを得たもので、この種の先駆的なモデルには福島によるネオコグニトロン⁶⁾が挙げられる。

ここまで見てきたように、大脳新皮質視覚野は大量の自然画像を入力として受け、その中からルールや規則性を抽出しているのだと解釈できる。この解釈を提供した数理手法が、スパースコーディングに代表されるスパースモデリングである。スパースモデリングは、少なくとも視覚野の理解については有用だと考えられる。しかし計算の普遍性を念頭に置いて見るなら、これだけではあくまで視覚系のモデルに過ぎない。さらに発展したスパースモデリングの成功例として、次節では聴覚系における近年の試みを紹介する。

4. 脳の音声処理とスパースモデリング

画像処理と音声処理が別の学問分野として発展したように、視覚系と聴覚系は別経路として発達してきた。一方で、

*1 ペイズ推定の言葉では、スパースな事前分布を与えることに対応する。自然画像に現れる物体の数は総ピクセル数に比べて非常に小さいという事前知識を利用していると解釈できる。

画像・音声処理のどちらも同じレイヤ（電子回路やソフトウェア）で実装可能なように、視覚系と聴覚系もまた神経細胞という同じ基本ユニットの組合せによって実装されている。

聴覚系で音波が神経活動に変換される場所は、耳の奥にある蝸牛と呼ばれる器官である（図2）。ここで音波は周波数帯ごとに分解され、例えば200 Hzに反応する細胞や400 Hzに反応する細胞が見つかる（フーリエ変換を思い浮かべて欲しい）。その後いくつかの中継核を経て、大脳新皮質の一次聴覚野（A1）に神経活動が到達する。聴覚系でまずスパースモデリングが発達したのは、大脳新皮質というよりは末梢側の神経表現だった。Lewickiらは自然音の波形をスパースに表現する基底を探し、結果が聴神経で知られるフィルタの形状とよく似ていることを指摘した⁷⁾⁸⁾。この結果は、スパースモデリングが単に視覚系に留まらない普遍性を持つことを示唆していた。

聴覚末梢においてスパースモデリングが有用なのであれば、中枢の大脳新皮質ではどうだろうか。V1神経生理学がその機能理解を促進させたのとは対照的に、大脳新皮質聴覚系の入口であるA1の機能理解は充分に進んでいないというのが実情である。神経活動記録の難易度が異なることも理由の一つだが、加えて、刺激選択性の種類が幅広いことの影響が大きい。200 Hzのように単一周波数に反応するものがあるかと思えば、他の神経細胞は200 Hzと400 Hzの両方のみに選択性を示したり、あるいはほとんどすべての周波数に弱い選択性を示したりする。この多様な選択性を目の当たりにした神経生理学者は、V1神経細胞におけるGabor関数のような簡潔な記述は難しいのではないかという印象を受け、なぜこうなっているのかという解釈はなおのこと難しいように思われた。

しかし近年の研究によれば、この一見理解が難しそうなA1の特性もまた、実はV1と同様のスパースモデリングで解釈可能である^{9)~11)}。V1におけるスパースモデリングの図式を思い出すと、V1神経細胞の刺激選択性は、自然画像が持つ統計性への適応として定式化できた。同様に、A1神経細胞の刺激選択性は、自然音を持つ統計性への適応として解釈することができるのだ（図4（下））。ではなぜ、A1の選択性はV1と違って複雑に見えるのだろうか。このモデルによれば、A1神経細胞がすっきりと記述できない特徴を持っているのは、そもそも自然音統計性があるような性質を持っているからである¹¹⁾¹²⁾。すなわち、学習則は共通なのだが、入力異なるために学習結果である神経生理学的特徴が別物に見えてしまうのだ。

自然音と自然画像の統計的な特徴はどのように異なるのだろうか？ まず網膜に映った自然画像について考えてみると、近隣の2点には概ね同じ物体が映っているからその光量は強く相関する。しかし離れた2点には異なる物体が映ることが多く、その明るさは無相関になるだろう。つまり、

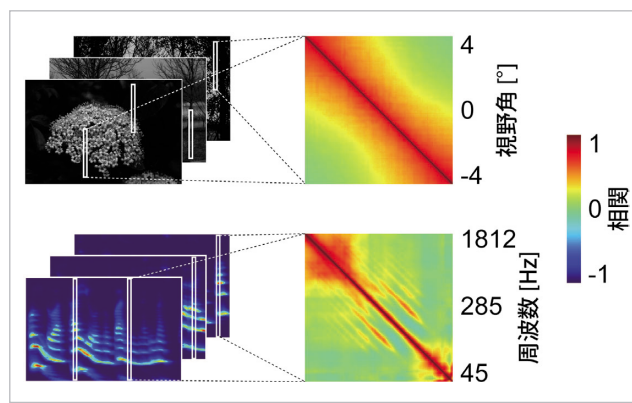


図5 局所的な自然画像統計性と非局所的な自然音統計性

自然画像と自然音それぞれから作った相関行列を示す。離れた網膜上の2点は基本的に異なる物体を映すので、明るさに相関があるのは視野上での近距離（狭い視野角）に限られる。一方自然音には、音声に含まれる倍音構造のように、周波数軸上で必ずしも局在しない構造が多く含まれる。図4（中列）はこの関係をまとめたものである。

自然画像は局所的な特徴を多く含んでいる（図5（上））。一方の自然音を周波数軸上で見てみると、まず自然画像と同様に、近い周波数は共起しやすいので局所相関が存在する。自然音の場合にはこれに加え、離れた周波数が必ずしも無関係でない、という非局所的な特徴が多く含まれる。われわれの音声を例にとって考えると、倍音構造が強く現れるために、200 Hzと400 Hzが強く相関する（そして300 Hzとは無相関）ということが充分に起こりうる。すなわち、自然音は非局所的な統計性を持ち、局所的な自然画像とは対照的である（図5（下））。このような相異を含む入力に対してV1やA1がそれぞれ適応しているのだから、V1が視野上で局在したコンパクトな基底表現を持ち、またA1が対照的に周波数軸上で分散した基底表現を持っているのは自然なことと言ってもいいだろう¹²⁾（図4）。

より解釈が難しそうなA1の地図構造についてもまた、V1と同様にモデル化できる。さらには、これまで類似性がまったく指摘されていなかったV1とA1における別概念までもが実は相同なものだという議論がスパースモデリングを通して始まっている¹²⁾。神経生理学的な特性記述を見るだけでは気付かなかった類似性が、スパースモデリングを通すことで初めて見えてきたという一例である。実は、V1とA1の学習戦略が同一ではないかという発想は突飛なものではなく、これを示唆する実験がある。視覚入力を強制的に聴覚経路に流し込むと、その後学習が進んだA1神経細胞の反応特性は、通常のV1神経細胞と似たものとなることが報告されている¹³⁾。すなわち、V1やA1に見られる特性は入力の統計性にかなりの程度依存するものらしい。

これらの結果はいずれも、スパースモデリングが単なるV1の理論に留まらず、何らかの普遍性を有する可能性を示唆している。これが最初に触れた大脳新皮質の「計算原理」とどう関係してくるのか、はっきりしたことはまだ何

も言えない。しかし原理解明のためには、モデル仮説をまず提唱し、複数領域のデータとすりあわせて修正するというループを繰り返す営みが必要だ。V1で発展した理論がA1に適用されつつあるという現状は、まさにその第一歩に相当するだろう（なお、言語学における普遍文法を巡る議論との類比が指摘されている¹⁴⁾）。

5. むすび

厳しい自然の中で生き抜くため、生物は目や耳から入ってくる大量の自然刺激からルールや規則性を抽出できるよう脳を進化させてきた。本稿では、大脳新皮質における画像・音声処理様式の理解がスパースモデリングによって進みつつある現状を概観した。

まず、局所光線分に選択的に反応するV1神経細胞が、自然画像をスパースに符号化しようとする最適化の結果であると示唆するモデル研究を紹介した。一見複雑で理解し難いと思われていたA1神経細胞についても、同様の戦略を用いた自然音への適応であることが指摘されつつある。そして大脳新皮質の均質性は、この種のモデリングが画像・音声処理のみならずより普遍的な計算原理の理解にも有用である可能性を示唆している。

近年注目を集める深層学習の例に見るように、脳の学習機構に学ぶべきことはまだまだ多い。折良くオープンな勉強会^{*2}が立ち上がる等、企業の方々の関心も高まってきている。本稿も科学を発展させる異分野交流のささやかな一助となったなら、この上ない喜びである。

(2014年8月30日受付)

〔文 献〕

- 1) 桑谷, 永田, 岡田: “マルコフランダムフィールドモデルによる計測画像データからの潜在構造抽出”, 映情学誌, 68, 12, pp.892-896 (2014)

*2 全脳アーキテクチャ勉強会, <http://www.sig-agi.org/wba>

- 2) 庄野, 佐々木: “スパースモデリングとTopographic ICA”, 映情学誌, 68, 12, pp.888-891 (2014)
- 3) D.H. Hubel and T.N. Wiesel: "Receptive fields, binocular interaction and functional architecture in the cat's visual cortex", The Journal of Physiology, 160, 1, pp.106-154 (1962)
- 4) B.A. Olshausen and D.J. Field: "Emergence of simple-cell receptive field properties by learning a sparse code for natural images", Nature, 381, 6583, pp.607-609 (1996)
- 5) A. Hyvärinen and P.O. Hoyer: "A two-layer sparse coding model learns simple and complex cell receptive fields and topography from natural images", Vision Research, 41, 18, pp.2413-2423 (2001)
- 6) 福島邦彦: “位置ずれに影響されないパターン認識機構の神経回路のモデル - ネオコグニトロン -”, 信学論A, J62-A, 10, pp.658-665 (1979)
- 7) M.S. Lewicki: "Efficient coding of natural sounds", Nature Neuroscience, 5, 4, pp.356-363 (2002)
- 8) E.C. Smith and M.S. Lewicki: "Efficient auditory coding", Nature, 439, 7079, pp.978-982 (2006)
- 9) D.J. Klein, P. König and K.P. Körding: "Sparse spectrotemporal coding of sounds", EURASIP Journal on Applied Signal Processing, 2003, 7, pp.659-667 (2003)
- 10) H. Terashima and H. Hosoya: "Sparse codes of harmonic natural sounds and their modulatory interactions", Network: Computation in Neural Systems, 20, 4, pp.253-267 (2009)
- 11) H. Terashima, H. Hosoya, T. Tani, N. Ichinohe and M. Okada: "Sparse coding of harmonic vocalization in monkey auditory cortex", Neurocomputing, 103, pp.14-21 (2013)
- 12) H. Terashima and M. Okada: "The topographic unsupervised learning of natural sounds in the auditory cortex", In Advances in Neural Information Processing Systems 25, pp.2321-2329 (2012)
- 13) M. Sur, P. Garraghty and A.W. Roe: "Experimentally induced visual projections into auditory thalamus and cortex", Science, 242, 4884, pp.1437-1441 (1988)
- 14) H. Terashima: "Computational Model for Auditory Cortex: An Analogy to Visual Cortex", PhD thesis, the University of Tokyo (2014)



寺島 裕貴 2009年、東京大学理学部情報科学科卒業。2011～2014年、日本学術振興会特別研究員(DC1)。2014年、同大学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。大脳新皮質の計算論的モデリングに従事。博士(科学)。