

視覚野・聴覚野地図の同一適応アルゴリズムによる解釈

An integrated interpretation of adaptive maps in visual and auditory cortices

寺島裕貴*1*2

Hiroki TERASHIMA

岡田真人*1*3

Masato OKADA

*1 東京大学 新領域創成科学研究科 複雑理工学専攻

Department of Complexity Science and Engineering, The University of Tokyo

*2 日本学術振興会 特別研究員 DC1

JSPS Research Fellow DC1

*3 理化学研究所 脳科学総合研究センター

RIKEN Brain Science Institute

While the anatomically uniform cerebral cortex is expected to use some universal intelligent algorithms, recent physiological studies has revealed difference of map structures between sensory cortices: unlike the smooth retinotopic map of the primary visual cortex (V1), the primary auditory cortex (A1) has a disordered tonotopic map. We hypothesize that the difference results from adapting, using a same algorithm, to different statistics of natural stimuli in vision and audition. The present study shows that a V1 map model, topographic independent component analysis, can also generate A1-like untidy maps, adapting to artificial and natural stimuli of high “auditoriness”.

高度な知能を実現するアルゴリズムの探求にあたり、脳はその謎を解く鍵を握っている。その中でも大脳皮質は、高等哺乳類において社会性等の高次機能獲得と共に急速に発達した脳部位であり、特に重要だと考えられている。大脳皮質は様々な機能を担う領野に分かれるにも関わらず、解剖学的には均質であるという大きな特徴を持つ。そのため、異なる領野間に共通する何らかの情報処理機構、それも知性の根幹に関わる情報処理系を有していると考えられ、その解明は脳科学の一大トピックとなっている。

一方で近年における測定技術の進歩は、上の考えと一見矛盾するかのような領野間の多様性もまた明らかにしつつある。最も盛んに研究されている領野である一次視覚野 (V1) の個々の神経細胞は、特定の視覚刺激パターン (受容野) に対して反応する。皮質表面で近い位置にある V1 神経細胞の受容野は視野内で似た位置にあり、V1 の表面全体が視野全体に対応する 2 次元状の地図構造を形成している (レチノトピー) [Tootell 88]。この地図は非常に滑らか、すなわち近傍神経細胞の受容野の類似度は非常に高い [Ohki 05, Ohki 06]。それに対して、音情報の処理を担う一次聴覚野 (A1) の地図 (トノトピー: 最も強く反応する周波数の並び) は乱雑、すなわち近接神経細胞の受容野が大きく異なることが分かってきた [Rothschild 10, Bandyopadhyay 10]。この地図構造の違いは、領野間に共通の情報処理機構が存在するという考えと矛盾するものだろうか。

我々は、この V1 と A1 の一見異なる地図構造が実は同一の適応アルゴリズムから生成されたものであり、自然界における刺激を持つ統計性の違いが反映されたものだという仮説を提唱する。V1 は A1 とは違って理論的側面からも盛んに研究されており、生理学的知見を上手く説明する理論がいくつか存在する [Olshausen 96, Bell 97]。その示唆するところによると、V1 神経細胞の受容野構造は自然視覚刺激に対して最適化されていると考えられる。すなわち、視覚刺激はランダムではなく、局所的な線分の組み合わせとして符号化すると都合が良い統計的特徴を持っているので、V1 神経細胞の受容野はそれに

対応する構造になっている。同様の考え方を用いて、レチノトピー等の地図構造をシミュレーションで再現することもできる [Hyvärinen 01a]。一方、A1 のモデル研究は非常に少なく、地図構造まで考慮したモデルは存在しなかった。我々は先行研究で、個々の神経細胞受容野に関しては A1 も V1 と同様のモデルで説明可能であることを示した [Terashima 09, Terashima 10]。自然聴覚刺激は離れた周波数間の相関という点で特徴的であり、対照的に視野空間で局所相関を示す視覚刺激との差が受容野構造の差として現れていると考えられる。本研究ではこの考えを地図構造にも拡張し、一見異なる V1 と A1 の地図構造もまた、同一のモデルを用いて解釈できることを示す。具体的には、過去に V1 の地図モデルとして提案されたトポグラフィック独立成分分析 (TICA) [Hyvärinen 01a, Hyvärinen 01b] を「聴覚らしい」人工刺激・自然刺激に対して適用した時に、A1 で見られるような乱雑な地図が現れることを示す。

本稿の構成は以下の通りである。まず第 1 章で、過去に V1 のモデルとして提案された TICA の概要を示し、定量的評価のための乱雑度 D の定義など本研究で用いた手法の詳細を述べる。そして第 2 章に人工刺激を用いた実験の結果、第 3 章に実データを用いた実験の結果をそれぞれ示す。第 4 章は本稿のまとめである。

1. 手法

1.1 トポグラフィック独立成分分析

本研究で用いるモデルであるトポグラフィック独立成分分析 (TICA: topographic independent component analysis) [Hyvärinen 01a, Hyvärinen 01b] は、独立成分分析 (ICA) [Bell 97] にトポグラフィーを加えた拡張である。TICA は以下のように近傍概念を導入して空間構造を定義し、自然画像の統計性から V1 のレチノトピーや方向選択性の滑らかな地図を再現した。ここでは [Hyvärinen 01a] の記法に従い、モデルと学習則を概説する。

TICA は 2 層の神経回路モデルであり、1 層目の活動 s_i は入力 I と結合荷重 (フィルタ) w_i との内積で決まる。1 層目と 2 層目の結合は近傍関数 $h(i, j)$ で定義され、近傍の 1 層目の活動を集めた “local energy” c_{it} を 2 層目の活動とし、これ

を用いて尤度を書く．なお $h(i, j)$ には、近傍で 1、それ以外では 0 をとるものを考える．($h(i, j) = \delta_{ij}$ のとき ICA と等価)

$$c_{it} = \sum_j h(i, j) \langle \mathbf{w}_i, I_t \rangle^2 \quad (1)$$

$$\log L(\{I_t\}; \{\mathbf{w}_i\}) = \sum_i \sum_t G(c_{it}) \quad (2)$$

ここで $G(c_{it}) = -\sqrt{\epsilon + c_{it}} = \log p_i(c_i)$ で、2 層目の活動にスパースな事前分布を与える．また、 \mathbf{w}_i の学習は尤度を勾配法で最大化することで行い、入力 I は白色化して \mathbf{z} とする．

$$\Delta \mathbf{w}_i = E \{ \mathbf{z} \langle \mathbf{w}_i, \mathbf{z} \rangle r_i \} \quad (3)$$

$$r_i = \sum_k h(i, k) g \left(\sum_j h(k, j) \langle \mathbf{w}_j, \mathbf{z} \rangle^2 \right) \quad (4)$$

ここで g は上記 G の微分である．なおこの学習則は、[Hyvärinen 01b] の付録にて導出されており、いわゆる Hebb 則に“トップダウン”の項 r_i が加わった変形になっている．

本研究のシミュレーションは、TICA の開発者らが公開している Matlab ライブラリ `imageica` に拡張を加えて行なった．パラメタの値は $\epsilon = 0.005$ である．

1.2 人工データ：「視覚的」な入力と「聴覚的」な入力

本研究では離れた座標間の相関に注目するため、まず人工データを対象に実験を行う．人工データ $I(x)$ は 1 次元 $x = 0, 1, \dots, 15$ という座標を考え、境界の影響を排除するために両端が接続された 1 次元トラスとする．

まず「視覚的」な信号を作成すべく、視覚刺激の特徴である局所相関性を以下のように模擬する．ランダムな座標 x を中心とし、これを平均とする $\sigma = 2$ の正規分布から生成した k 個の整数座標に 2 を加える (k は 3, 4, 5, 6 の一様分布からサンプル毎に生成)．このようにして、局所的な相関を持つ入力ができる．

次に「聴覚的」な入力を、「聴覚らしさ」を表すパラメタ p_a を用いて定義する．上で整数座標に値を加える際に、確率 p_a で座標 $x_{dist} = x + 5$ にも値 2 を加えることにする．これは離れた座標で相関を持つという聴覚刺激の特性を模擬したもので、 p_a の値が大きいくほど入力は「聴覚的」となる．

1.3 実データ：画像入力と音声入力

人工刺激だけではなく、以下のような画像と音声の実データを用いたシミュレーションも行い、地図の乱雑度を比較した．

まず画像については、TICA の作者らによる Matlab デモプログラム `imageica` に付属の自然画像をそのまま用いた．モデルへの入力は 1 辺 25 pixel の正方形をランダムな位置から選び、 $N = 20,000$ 個作成した．また主成分分析を用い、625 次元のうち寄与の大きい方から 400 個の成分だけを採用することで、実質的なローパスフィルタリングを行った．

音声データには、ピアノ演奏の録音を自然界の聴覚刺激の代表として用いた．図 4A のスペクトログラムが示すようにピアノ演奏は離れた周波数間での相関が顕著で聴覚刺激の特徴をよく表しており、また A1 受容野の解釈に適切な統計性を持っていることが既存研究により示唆されている [Terashima 09, Terashima 10]．録音データはまず 4 kHz にダウンサンプリングし、そのスペクトログラムを細かく区切って入力とした．スペクトログラムの作成には長さ 64 の hann 窓を $\frac{1}{4}$ ずつ重ねて用い、結果として周波数方向は 2 kHz まで等間隔に 32 点で表現した．個々の入力は、時間方向に 18 の幅 (約 0.2 秒) をランダムに抽出して $N = 20,000$ 個作成し、画像データと同様に 400 個の主成分を採用した．

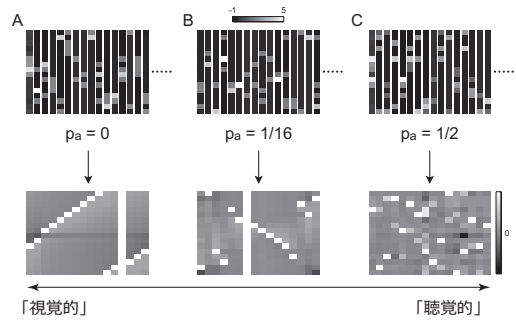


図 1: 人工入力の「聴覚らしさ」 p_a と地図の乱れ．下段地図の各列が \mathbf{w}_i で、絶対値最大の点が白 (= 1)、灰色が 0 となるように各々をスケールした．

1.4 乱雑度 D ：地図の乱雑さの指標

本研究では、特微量地図の乱雑さの度合いを比較する．しかし特微量が異なる次元を持つとそのままでは比較できないので、正規化した乱雑度 D を各座標に対して以下のように定義する．

まず、地図の座標が 1 次元整数の場合に、ある座標 x_c における乱雑度を定める．注目する特微量 $f(x)$ の値の範囲が $[0, L]$ のとき、その値が $[0, 1]$ に収まるように $\tilde{f}(x) = \frac{f(x)}{L}$ としてスケールする．その上で、近傍座標の集合 $NB(x; l) = (x - l, x - l + 1, \dots, x + l)$ を考え (元の数 $N_{NB} = 2l + 1$)、その範囲で座標 x を説明変数として \tilde{f} を最小二乗法で線形回帰して得られる残差を $r(x)$ と書くことにする．このとき、座標 x_c における乱雑度 $D(x_c)$ を、以下のように定める．

$$D(x_c) = \sqrt{\frac{\sum_{x \in NB(x_c)} r(x)^2}{N_{NB}}} \quad (5)$$

すなわち、乱雑度 D は、 l で定義される近傍 NB における残差平方和の平均値の平方根である．なお、本研究で用いる近傍は全て $l = 2$ である (1 次元なら $N_{NB} = 5$) ．

地図の座標が 2 次元 x, y の場合も、説明変数が 2 つとなることと、近傍が 2 次元で定義されることを除けば同様に定義できる．本研究で用いる 2 次元の近傍は 5×5 の正方形である．

ただし、人工データの座標空間は 1 次元トラスとして定義したので、両端の扱いを考える必要がある．本研究では近似的な手法として、近傍中で $[0, \frac{1}{2})$ の範囲にある正規化された特微量 \tilde{f} に 1 を加えた場合の D も同時に計算し、より良く線形回帰が当てはまる方の値 (小さな D) を採用した．本研究では近傍の大きさが L に比べて十分小さいので ($N_{NB} = 5 < 16$)、副作用はあまり大きくならない．

2. 結果：人工データへの適用

2.1 TICA による地図の作成

自然界の聴覚刺激の特徴である遠い周波数間での相関が A1 で見られる地図構造の乱れの原因であるという仮説を検証するため、まず人工データを用いてシミュレーションを行った．

実験毎に、まず用いる刺激の「聴覚らしさ」 p_a の値を決め、1.2 節の方法に従って入力を生成した．そして TICA モデル (1.1 節) を用いて、16 個のフィルタ \mathbf{w}_i からなる 1 次元地図を生成する．

異なる p_a の値を用いた際の地図の違いを図 1 に示す．局所的な相関が強い視覚的な入力を用いた場合 ($p_a \sim 0$) には滑

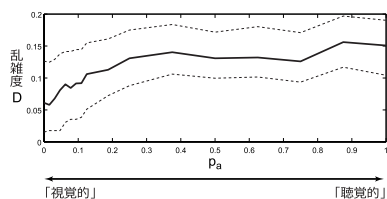


図 2: 人工入力 of 「聴覚らしさ」と地図の乱雑度は相関する。

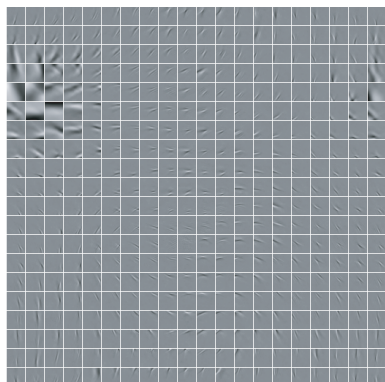


図 3: TICA による画像入力からの地図生成. 1 マスが 1 つのフィルタ w_i にあたる (白が最大値, 0 がグレー).

らかな地図が得られたのに対し, 聴覚らしさが増すに従って地図が乱れていく。

2.2 入力の「聴覚らしさ」と地図の乱雑さ

次に, p_a の値による地図の乱雑さの変化を定量的に調べるため, まず特徴量 f を, 得られた各フィルタの中で絶対値最大の値を持つ (入力空間での) 座標とする ($f = 0, 1, \dots, 15$). その上で 1.4 節の手法を用い, 入力空間と地図座標のトーラス構造を考慮しつつ, 特徴量 f の乱雑度 D を地図の各座標に対し計算していく。

p_a の値毎に, その値を用いたシミュレーション全てで得られた D の値を考え, p_a の変化に伴う D の推移を図 2 に示す. 乱雑度 D は「視覚的」な入力 ($p_a \sim 0$) に対しては 0.05 付近の非常に低い値を持つが, 入力の「聴覚らしさ」が増すに従って単調に増加しており, 入力の「聴覚らしさ」が TICA で得られた地図の乱雑さと正の相関を持つ。

3. 結果: 実データへの適用

3.1 TICA による画像入力からの地図生成

前章の 1 次元人工データに続き, 2 次元の実データを用いたときの振る舞いについて調べる。

まず, 2 次元画像を入力として TICA で 2 次元地図を作成する. これは既に [Hyvärinen 01a, Hyvärinen 01b] 等で示されているものと同様であり, あくまで聴覚刺激から作成する地図との比較用である. 座標空間は 20×20 のトーラス, 用いた近傍の大きさは 5×5 である。

用いた入力は Matlab デモプログラム `imageica` に付属の自然画像から生成した正方形パッチである (1.3 節). この入力を対象に十分学習した地図を図 3 に示す. 個々のフィルタは局在し, 方向選択性を持ち, バンドパスであり, 2 次元ガボールフィルタでよくフィットできる. すなわち V1 単純細胞の受容野と酷似している [Jones 87]. そして座標や傾きといったパラメタについて, 地図座標上で滑らかに変化することが分かる。

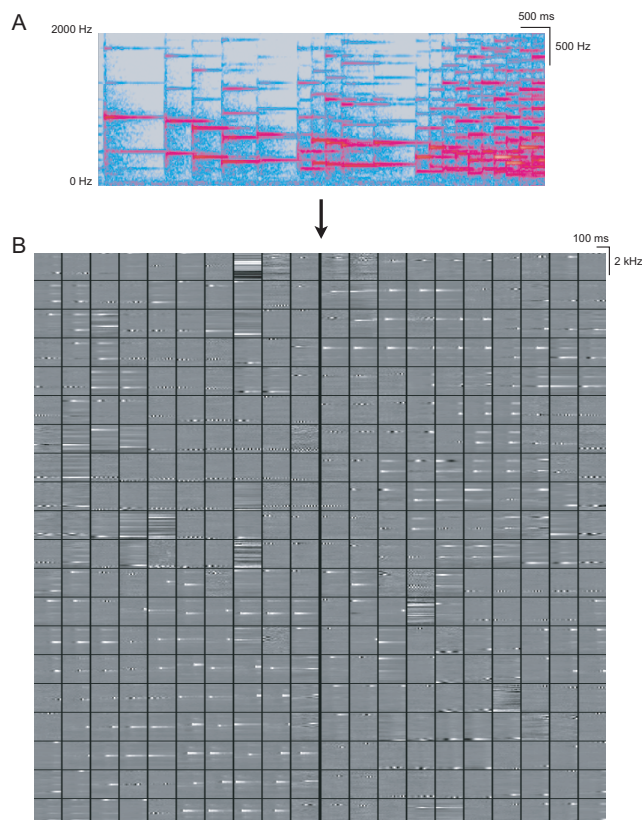


図 4: TICA による音声入力からの地図生成.

3.2 TICA による音声入力からの地図生成

さらに, 音声入力を用いても同様に地図を作成した. 離れた周波数間での相関が明瞭な例として, 入力にはピアノ演奏の録音を用いた (1.3 節). ハーモニックな周波数間の相関構造は図 4A のスペクトログラム上でもはっきり確認することができる。

この入力を対象に十分学習した結果を図 4B に示す. 個々のフィルタは時間周波数空間 (特に周波数空間) で局在し, A1 の受容野の特徴と類似する [Klein 03] と類似のフィルタが得られた。

さらに地図座標上での変化を見ると, 滑らかな部分も一部に存在するものの, 大きな変化が頻繁に見受けられる地図となった. 例えば地図の右下では, 高周波に選択性を持つフィルタと低周波のフィルタが隣接していたり, 地図右中央にはハーモニーに関連するフィルタが集まっているように見えるが, その境界では特徴が激しく変化する. このような不連続性は, 画像入力を用いた場合の非常に滑らかな地図とは対照的である。

3.3 視覚地図と聴覚地図の乱雑さ

以上で得られた視覚地図 (図 3) と聴覚地図 (図 4B) の乱雑さを定量的に比較するため, 各座標で正規化済み乱雑度 D を求めた。

まず, 特徴量 f を各々の地図の各座標で定義する. 視覚地図については 2 次元ガボール関数をフィットし, その中心座標 μ_x, μ_y それぞれについて図 5A のように特徴量地図を作成した. 中心座標は概ね滑らかな変化を示していることが分かる。

一方の聴覚地図については, 特徴周波数に対応する指標として, フィルタの中で絶対値最大の値を持つピークの周波数を特徴量として選んだ. これは図 4B の表示においては縦軸の座標

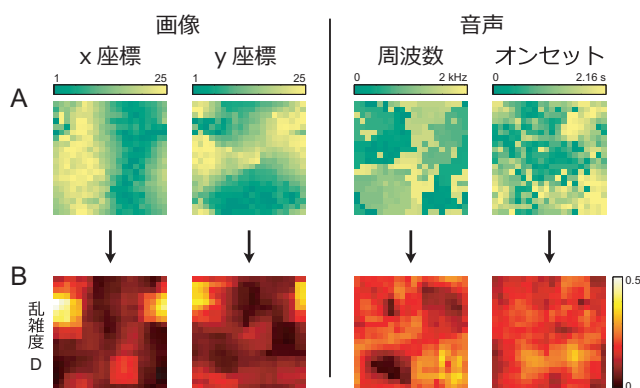


図 5: 視覚刺激と聴覚刺激から生成された特徴量地図と、その乱雑度地図。音声地図は視覚地図に比べ乱雑である。

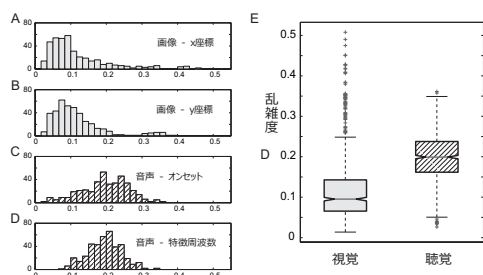


図 6: 聴覚地図の乱雑度は、視覚地図に比べ有意に高い。Eの「視覚」はA, Bを、「聴覚」はC, Dをまとめたもの。

にあたる。また同時に、横軸の座標も別の特徴量としてとる。この値は、生物学的には刺激提示後何秒後に反応が起きるかというオンセット時間(遅れ)として解釈できる。フィルタ地図を見て受ける印象と同様に、滑らかに変化する部分も存在するものの、大きな変化もしばしば見受けられる。

以上4つの特徴量地図について、各座標における正規化済みの乱雑度 D を 5×5 の近傍を用いて求めた結果を図5Bに示す。画像から作った地図の乱雑度 D は概して低い値をとる。例外的に乱雑度が高い領域が左上にあるが、これは元々のフィルタの周波数が低く部分的にしか見えていないため、ガボール関数のフィッティング誤差が大きいためと考えられる。

画像特徴量の滑らかな地図に対し、音声から作った地図の乱雑度 D は全体的に比較的大きな値をとる。特徴量地図を直接見て得られる直感と同様に、音声からTICAで作ったフィルタの地図は相対的に乱雑な構造を示している。

それぞれの特徴量について、乱雑度 D の分布を図6に示す。画像特徴量2つを合わせた分布と音声特徴量2つを合わせた分布を比較すると、中央値は有意に異なる ($p < 10^{-7}$; Wilcoxon rank test)。画像特徴量乱雑度の中央値が0.09なのに対し、音声特徴量の乱雑度中央値は0.20だった。

4. まとめ

本研究では、V1地図のモデルであるTICAを人工刺激と音声刺激に対して適応させ、「聴覚らしい」刺激に対してはA1のように乱れた地図が生成されることを示した。この結果は、近年明らかになったV1とA1の地図構造の違いが、刺激が持つ統計性の反映として解釈され得ることを示唆する。すなわち、網膜空間で局所的な相関を持つ視覚刺激はV1地図の滑らかさ

を、また周波数空間で離れた相関を持つ聴覚刺激はA1地図の乱雑さを導いている。またこの結果は、V1とA1の違いが本質的な情報処理機構の相違に基づくのではなく、同一アルゴリズムによる自然界の構造への適応の結果であるという先行研究の仮説 [Terashima 09] を改めて支持するものである。

参考文献

- [Bandyopadhyay 10] 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)
- [Bell 97] Bell, A. J. and Sejnowski, T. J.: The “independent components” of natural scenes are edge filters, *Vision Research*, Vol. 37, No. 23, pp. 3327–3338 (1997)
- [Hyvärinen 01a] 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)
- [Hyvärinen 01b] Hyvärinen, A., Hoyer, P. O., and Inki, M.: Topographic independent component analysis, *Neural Computation*, Vol. 13, No. 7, pp. 1527–1558 (2001)
- [Jones 87] Jones, J. P. and Palmer, L. A.: An evaluation of the two-dimensional Gabor filter model of simple receptive fields in cat striate cortex, *Journal of Neurophysiology*, Vol. 58, No. 6, pp. 1233–1258 (1987)
- [Klein 03] 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)
- [Ohki 05] Ohki, K., Chung, S., Ch’ng, Y. H., Kara, P., and Reid, R. C.: Functional imaging with cellular resolution reveals precise micro-architecture in visual cortex, *Nature*, Vol. 433, No. 7026, pp. 597–603 (2005)
- [Ohki 06] Ohki, K., Chung, S., Kara, P., Hübener, M., Bonhoeffer, T., and Reid, R. C.: Highly ordered arrangement of single neurons in orientation pinwheels, *Nature*, Vol. 442, No. 7105, pp. 925–928 (2006)
- [Olshausen 96] 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)
- [Rothschild 10] 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)
- [Terashima 09] 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)
- [Terashima 10] Terashima, H. and Hosoya, H.: Sparse codes of harmonic sound and their interaction explain harmony-related response of auditory cortex, *BMC Neuroscience*, Vol. 11, No. Suppl 1, p. O19 (2010)
- [Tootell 88] Tootell, R. B., Switkes, E., Silverman, M. S., and Hamilton, S. L.: Functional anatomy of macaque striate cortex. II. Retinotopic organization, *Journal of Neuroscience*, Vol. 8, No. 5, pp. 1531–1568 (1988)