

## ハーモニーのスパースコーディングとその相互作用

## Sparse codes of harmonic sounds and their modulatory interactions

寺島裕貴 (PY)<sup>†</sup>, 細谷晴夫<sup>‡</sup>

Hiroki Terashima(PY), and Haruo Hosoya

<sup>†</sup> 東京大学新領域創成科学研究科複雑理工学専攻 <sup>‡</sup> 東京大学理学部情報科学科

hterashima@mns.k.u-tokyo.ac.jp

**Abstract**— Sparse coding has been successful to explain peripheral auditory neuronal response properties. We continue this line of research in a higher stage of auditory system, focusing on harmonic structure of behaviourally important natural sounds. We show frequency-domain sparse codes of harmonic sounds have harmonic relations in response properties, which are qualitatively consistent with physiological report.

**Keywords**— auditory system, sparse coding, auditory cortex, computational neuroscience, harmony

## 1 はじめに：スパースコーディングとハーモニー

スパースコーディングとその関連手法は、初期感覚情報処理の特性を見事に説明してきた。その最も顕著な例は [1] である。自然画像をスパース (疎) に表現できる基底の組をシミュレーションで学習させると、得られた基底は一次視覚野の単純細胞が示す反応選択性に似た性質を示した。また、聴覚系においても、末梢の聴覚神経の特性が独立成分分析 (ICA) などを使って再現されている。生物学的には神経発火の回数、すなわちエネルギー消費量を抑えられるという利点があり、生体の感覚情報処理における原理として有望である。解釈としては、環境からの刺激を学習、あるいはそれに適応し、ある種の特徴抽出を行っていると考えられる。一方で、高位の情報処理段階においては一般に神経細胞の刺激に対する反応は複雑であり、うまく説明された例は少なく、聴覚系もその例外ではない。しかし環境への適応という観点からすると、刺激の持つ複雑な構造を何らかの形で捉えていることが期待できる。

我々は、自然音に含まれる複雑な特徴としてハーモニーに注目した。ここでハーモニー (harmonic な音) とは、単純な整数比で表される周波数成分が多く含まれる音を指す。一般に自然界の音 (動物の発声, 楽器音, etc.) はハーモニーを豊富に含んでおり、人間の知覚上でもハーモニーは心地良く感じられることから、何らかの形でハーモニーの検出が行われていると思われる。実際、サル的一次聴覚野 (A1) において、単周波数刺激と 2 周波数刺激を用いたときの反応特性に harmonic な関係が見られる [2]。我々はこの関係はハーモニーを豊富に含む音のスパースコーディングの結果だと考えた。そこで、[1] のモデルとアルゴリズムを用いて周波数領域でのスパースな表現を与える基底を学習させ、これをモデルニューロンとみなした。シミュレーションで [2] の方

法に倣ってその反応特性を調べると、同様の harmonic な関係を見出すことができた。この関係は、ニューロン (基底) 間の相互作用によって生じると考えられる。

## 2 モデルとアルゴリズム

本研究では [1] のアルゴリズムをほぼそのまま用いた。入力  $A(x)$  を基底  $\phi_i(x)$  の線形結合で表し、基底  $\phi_i(x)$  をモデルニューロン、係数  $a_i$  をその活動とみなす。

$$A(x) \simeq \sum_i a_i \phi_i(x)$$

ここで 2 つの問題を解く必要がある。ひとつは入力  $A(x)$  と基底の集合  $\phi_i(x)$  が与えられたとき、スパースな係数  $a_i$  を決定する問題。もうひとつは、多くの入力に対してスパースな係数  $a_i$  を与えるような基底の集合  $\phi_i(x)$  を求めるという問題、つまり基底を入力列に対して適応させるという問題である。この 2 つの問題は共にコスト関数  $E$  の最小化問題として定式化できる。

$$E = \frac{1}{2} \sum_x \left[ A(x) - \sum_i a_i \phi_i(x) \right]^2 + \lambda \sum_i |a_i|$$

右辺の第 1 項が二乗誤差、第 2 項が係数のスパース性を表す。 $\lambda$  はスパース性の重みを表す定数である。この表現を使うと、2 つの問題はそれぞれ次のようにかける。

$$\Delta a_i = \alpha \left[ \sum_x \phi_i(x) A(x) - \sum_j a_j \sum_x \phi_i(x) \phi_j(x) - \lambda \cdot \text{sign}(a_i) \right] \quad (1)$$

$$\Delta \phi_i(x) = \eta \left\langle a_i \left[ A(x) - \sum_i a_i \phi_i(x) \right] \right\rangle$$

$\alpha, \eta$  はそれぞれ更新率と学習係数を、 $\langle \dots \rangle$  は繰り返しに対する平均を表す。なお基底の更新においては、むやみな発散を避けるため、全基底の値の 2 乗平均が定数  $v_{\text{const}}$  になるようにスケールした。

## 3 シミュレーション

まず、3 種類の音を入力として用いた基底を学習した。音断片 (256 or 320 sample) を高速フーリエ変換 (FFT) した振幅スペクトルを入力とし、上記アルゴリズムで基底 (80 or 128 個) を学習した。ハーモニーの多い音としてピアノ演奏と人間の声<sup>1</sup>を、harmonic でない自然音と

<sup>1</sup> 国際音声学会ハンドブックの American-English narratives

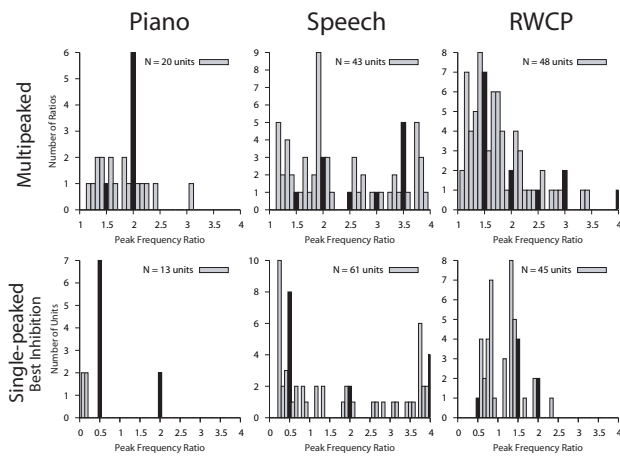


図 1: Multipeaked と single-peaked 遠位抑制ピーク周波数比. 特にピアノで harmonic な比に集中する. (黒塗:  $\frac{1}{2}$  の倍数)

して RWCP 実環境音声・音響データベース<sup>2</sup>を利用した. ピアノ演奏と人間の声はローパスフィルタをかけた後 4 KHz にダウンサンプリングしたものを用いた.

次に, 得られた基底をモデルニューロンとみなし, [2] に倣って単周波数刺激に対する応答で基底を single-peaked と multipeaked の 2 種類に分類した. 絶対値最大の係数を与える周波数を CF とし<sup>3</sup>, (CF での係数)  $\times \rho_{\text{peak}}$  以上の係数を与える周波数が他にあれば multipeaked とする. 1 つの基底が持つピークの中で周波数が最低のものに対する周波数比のヒストグラムを作った (図 1 上段). 入力として harmonic な音を使ったときは harmonic な比に集中しているのに対し, 入力が harmonic でないときはそのような傾向は見られない.

Single-peaked の基底に対しては CF ともう 1 つの周波数を混合した 2 周波数刺激を与え, その周波数を変化させて活動の変化を調べた. [2] では, CF 両側抑制のさらに外側 (遠位) にある促進あるいは抑制のピークを扱っている. その手法に倣い活動変化率  $\delta = (R_2 - R_1)/R_1$  を導入し, 側抑制の外側で閾値  $\delta_+$ ,  $\delta_-$  を超えるピークを検出した.  $R_1, R_2$  はそれぞれ, 単周波数刺激と 2 周波数刺激に対する係数を表す. このようなピーク周波数の CF 比の分布を調べた (抑制の結果は図 1 下段). ここでも, 入力に harmonic な音を使った場合は結果も harmonic な比に集中し, 入力に harmonic でないときはそういった傾向が失われる.

#### 4 生理学的知見との対応と計算論的解釈

サル A1 ニューロンにおいては (1)multipeaked のピーク周波数比は harmonic な関係にある, (2)single-peaked の遠位抑制ピーク周波数比は harmonic な関係にある, という傾向が認められる [2]. 我々の得た分布は, harmonic な音, 特にピアノ演奏を入力として使ったときに類似の性質を示した (図 1 左). 0.5 と 2.0 での強いピークは [2] の結果と一致する. 一方で, harmonic でない RWCP を入力として使ったときには harmonic な偏りは見られなかった (図 1 右). 人間の声を使った

<sup>2</sup>非音声音のドライソース (楽器と電子音を除く)

<sup>3</sup>特徴周波数. 係数が負の場合は基底と係数の符号を反転させた.

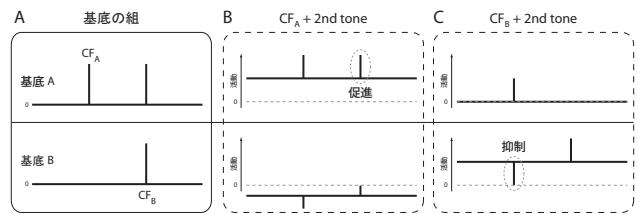


図 2: 遠位促進および抑制の 2 基底による単純なモデル

ときに得られた中間的な傾向と合わせて考えると, A1 で見られる harmonic な反応特性は入力音の構造を反映したものだと考えられる.<sup>4</sup>

CF に対して harmonic な第 2 周波数を与えたときに single-peaked な細胞で起こる反応変化は, 単純なモデルを用いてニューロン間の相互作用を考えることで説明できる. 図 2A のような基底 2 つで考えると, スパース性最大化によって基底 A, B は  $CF_A, CF_B$  で最も強く反応する.  $CF_A$  と他の周波数の 2 周波数刺激に対する反応を考えると, 基底 A の活動は  $CF_B$  で促進される (図 2B). 同様に, 基底 B に対して  $CF_B$  と他周波数の刺激を与えると,  $CF_A$  で活動が抑制される (図 2C). ここで, 学習対象として harmonic な成分が豊富に含まれる音を使うと得られる基底も harmonic な構造を持つので, 結果として harmonic な関係にある周波数において促進や抑制が現れることになる.

このような相互作用は式 1 に見られる項  $\phi_i(x)\phi_j(x)$  によっても, 似た形を持つ基底間, つまり類似の性質を持つニューロン間に起こる一種の「側抑制」として説明できるだろう. 生理学的にも A1 内水平結合は等方的でなく, 同じ CF を持つ方向への結合が多く存在することが知られており, 似た反応特性を持つニューロン間の相互作用の重要性を示唆している.

最後に, 我々の仮説を確かめるための生理学実験を提案する. 1 つ目に, harmonic な音を全く呈示されない極端な環境下で動物を飼育すると, A1 で harmonic な反応特性は見受けられないと予想される. 一方通常環境下で飼育された動物では上で述べたようにニューロン間相互作用が起きていると考えられるため, 1 つのニューロンだけに注目しても分からない. そこで 2 つ目として, 互いに harmonic な関係にある CF を持つニューロン A, B の活動を同時に計測することを提案し, 以下の結果を予想する. もし B の CF で A の促進が起きて同時に B の活動が無くなるなら, B の抑制は A の CF で A の活動上昇と同時に起きる. また, 逆も成り立つ.

#### 参考文献

- [1] B. A. Olshausen and D. J. Field (1996) “Emergence of simple-cell receptive field properties by learning a sparse code for natural images.” *Nature*, **381**, 607–609.
- [2] S. C. Kadia and X. Wang (2003) “Spectral Integration in A1 of Awake Primates: Neurons With Single- and Multipeaked Tuning Characteristics.” *Journal of Neurophysiology*, **89**, 1603–1622.

<sup>4</sup>なお [2] では (3)single-peaked の遠位促進ピーク周波数は harmonic でない比に比較的ばらけるという傾向も指摘されているが, 本研究では遠位促進でも harmonic な関係が見受けられた.