

スパースな局在興奮パターンによる自己組織マップを用いた 楽器音の分離

和田 浩司[†], 倉田 耕治[‡]

メキシカンハット型の層内結合を持つ神経場にランダム性を導入すると、平衡状態にまばらな局在興奮をもつようになる。我々は、このまばらな局在興奮を用いた情報表現と、その自己組織化のモデルを提案、考察した。提案モデルでは情報は局在興奮の位置と、同じ位置での局在興奮の細かな発火パターンの異りとして表現される。提案モデルに楽器音を学習させることで、楽器音の非線形第1種成分が局在興奮の位置で、第2種成分が発火パターンで表現されることを確認した。

Sound separation using self-organizing map with sparse and local excitation

Koji Wada[†], Koji Kurata[‡]

A nervefield with random Mexican hat type intralayer connection pattern has stable local excitation patterns in which active units are sparsely and randomly distributed. A new self-organization model based on such nervefield is proposed and studied. In this model information is coded not only by the position of excitation but by detailed difference among random sympatric excitation patterns. It is shown by simulation that the first principal component of input data is coded by the position of excitation pattern and the second by detailed difference of it.

1 はじめに

1.1 自己組織マップと非線形主成分分析

Kohonen の SOM(Self-Organizing Map) などのトポロジカルマップ形成のモデル [1][2] を用いると、線形主成分分析では抽出不可能な非線形の主成分も抽出することができる。しかし従来のマップ形成のモデルでは、第2以降の主成分についてはこれを単独で抽出することができない。それは、第1主成分と第2主成分を持つような信号空間のマップを、1枚の1次元の神経場上に自己組織的に形成させると、SOM に代表される従来のトポロジカルマップ形成の方法では、第2主成分を全く無視した第1主成分についてのマップができるか、あるいは全体としては第1主成分をマップしながら局所的には第2主成分方向に波打つようなマップが形成されるからである [1]。

このようなマップでは第1、第2主成分の特定の組合せに対して選択的に発火する素子は形成できるが、第2主成分を単離抽出することはできない。白倉らは2枚の神経場よりなる連続多重マップモデルを使って第1、第2主成分を単離抽出することに成功している [3]。

1.2 スパースな局在興奮による自己組織化

今回われわれは、ランダムかつスパースな局在興奮パターンを持つような神経場を用いたトポロジカルマップ形成のモデルを提案する。本論文では計算量を減らすため、

図 1(a) に示すような 1 次元の神経場を用いた。このモデルでは発火する領域の位置と、発火している素子群のランダムな違いとの 2 つを用いて情報を表現することができる。本論文では図 1(a) の f_1 と f_3 のような興奮パターンの違いを位置方向の違い、 f_1 と f_2 のような違いをランダム方向の違いといい表すことにする。本研究のテーマは、このような興奮パターンを持つ神経場に SOM のような学習則を組合せると、どのようなマップができるかということである。

一般に、信号空間から神経場への自己組織マップがトポロジーを持つことは、近傍学習の効果による。本モデルでも発火パターンの広がりによって近傍学習をとり入れており、これによってトポロジカルなマップが形成される。発火パターン相互の遠近は、発火パターンをベクトルとみなしたとき、その方向余弦で計測できる。スパースな発火パターンであっても位置のずれとともに方向余弦は小さくなるので、位置方向へトポロジカルなマップを形成することができる。

しかし、同じ領域でのランダムな発火パターン相互にも、その方向余弦は大きい場合と小さい場合がある。このため本モデルでは、ランダム方向についても、入力信号の遠近(入力信号相互の方向余弦)を興奮パターン間の遠近にある程度反映させることができるはずである。

図 1 にその模式図を示す。 v_1, v_2, v_3 は入力信号を表し、 f_1, f_2, f_3 はそれらに対応すると予想される神経場の発火パターンを表している。ここで、神経場の発火している素子は黒く塗りつぶされている。 v_1 と v_2 は第2主成分のみが異なり、 v_1 と v_3 は第1主成分のみが異なる。入力信号

[†] 埼玉大学大学院理工学研究科
Graduate School of Science and Engineering, Saitama University
[‡] 琉球大学工学部
Faculty of Engineering, Ryukyu University

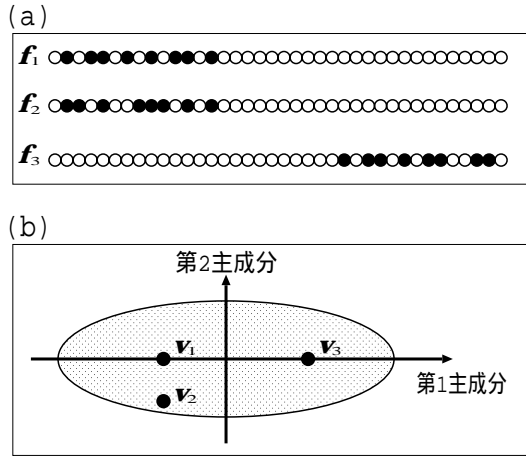


図 1: (a) スパースな局在興奮. (b) 入力信号.

入力信号 v_1, v_2, v_3 のそれぞれに、神経場の発火パターン f_1, f_2, f_3 が対応すれば、入力の第 1 主成分は発火の位置によってコードされ、第 2 主成分は発火パターンの細かな違いによってコードされていることになる。 v_1, v_2, v_3 のそれぞれに、 f_1, f_3, f_2 が対応するなら、入力の第 2 主成分が発火の位置によってコードされ、第 2 主成分は発火パターンの細かな違いによってコードされていることになる。

v_1, v_2, v_3 のそれぞれに、神経場の発火パターン f_1, f_2, f_3 が対応すれば、入力の第 1 主成分は発火の位置によってコードされ、第 2 主成分は発火パターンの細かな違いによってコードされていることになる。また、もし v_1, v_2, v_3 のそれぞれに、 f_1, f_3, f_2 が対応するなら、入力の第 2 主成分が発火の位置によってコードされ、第 2 主成分は発火パターンの細かな違いによってコードされていることになる。ただしこのとき、 f_2 と f_3 が同じ発火パターンを平行移動したもになっていることまでは、本モデルでは期待できない。

ところで、このモデルで第 1 主成分が位置方向、ランダム方向どちらの方向に抽出されるかはそれら 2 つの方向の相対的な長さに依存し、より長い方に第 1 主成分が、より短い方に第 2 主成分がそれぞれ抽出されるはずである。

したがって、もし発火パターンのなす集合の位置方向の長さがランダム方向の長さに比べて充分長ければ、位置方向に第 1 主成分、ランダム方向に第 2 主成分のトポロジカルマップが形成され、また逆に、ランダム方向の長さが位置方向の長さに比べて充分長ければ、ランダム方向に第 1 主成分、位置方向に第 2 主成分のトポロジカルマップが形成されることになる。

このように、第 2 主成分が神経場上の位置方向にマップされれば、学習完了後に素子のしきい値を下げるなどして発火パターンを密なものにすることにより、連続多重マップモデル [3] と同じように第 2 主成分を分離抽出することができる。

2 モデル

2.1 層内結合

本モデルでは神経場内の層内結合にランダム性を導入することで、ランダムでスパースな安定局在興奮パターンをもつ神経場を実現している。その様式は、距離が近い素子との結合荷重は、その値が正負等確率で生成され、距離が遠い素子との結合は、高い確率で抑制性である、という確率的性質を持つ対称結合である。

具体的には、神経場上の素子 x と x' との距離を d としたとき、神経場内の素子 x と素子 x' との結合荷重 $J_{xx'}$ は、 $d > B$ のとき、平均 $-\frac{L}{B}d$ 、標準偏差 $M(1 - \frac{d}{B})$ の正規分布よりランダムに決定し、 $d \geq L$ のときは $J_{xx'} = -L$ で一定とする。ここで L, B, M は定数であり、実際の実験で用いた具体的な数値は表 1 に掲げる。本論文では 1 次元の神経場を離散的に取り扱う。したがって x は整数であり、隣合う素子間の距離は 1 としている。

2.2 膜電位のダイナミクス

ある入力信号ベクトル v が与えられた時の神経場の平衡状態は、以下の手順で求める。ここで、神経場上の位置 x に存在する素子と入力信号との結合荷重ベクトルを $w(x)$ とする。

手順 1. 神経場内の全素子に対して膜電位のダイナミクスの初期値 $u(x) = w(x) \cdot v$ を求める。

手順 2. 神経場における膜電位の分布の平均 $\hat{\mu}$ を $\hat{\mu} = \frac{1}{N} \sum_x u(x)$ 、標準偏差 $\hat{\sigma}$ を $\hat{\sigma} = \sqrt{\frac{1}{N} \sum_x (u(x) - \hat{\mu})^2}$ と推定し、膜電位の値が $\hat{\mu} + c\hat{\sigma}$ よりも大きな素子について、その出力 $a(x)$ を 1 とする (発火)。それ以外の素子の出力は 0 とする (非発火)。ここで、 N は素子数、 c は定数であり、実際の実験で用いた具体的な数値は表 1 に掲げる。

手順 3. 層内結合にしたがって、同期的に次の状態の膜電位を決定する。つまり、素子 x の膜電位を $u(x) := \sum_{x'} J_{xx'} a(x')$ とする。

手順 4. 手順 2, 3 を、神経場が平衡状態に達するまで繰り返す。

手順 2 は発火率を一定に保つためのソーティングを用わない簡便な方法である。この方法を用いることにより、計算量を大幅に軽減することができる。このダイナミクスでは、入力の役割は直接には膜電位の初期値にしか使われていない。これは、外部入力のスナプスより神経場の内部結合のスナプスの方がはるかに強いということを意味する。そのため、内部結合によって安定発火パターンの集合は決定される。しかし外部入力も、多くの場合は安定発火パターンのなかから、ひとつの発火パターンを指定するという重要な役割を果たす。これは自己組織モデルの多くに共通した性質である。

今回のようなダイナミクスでは、神経場の発火状態は平衡状態に収束する場合と、2 パターンの発火状態が交替する周期解に収束する場合とが存在した。

2.3 学習

平衡状態に収束した際に発火している素子について、以下のように Kohonen の内積型 SOM[1] と同じ正規化つき Hebb 学習を行なう。

$$w(x, t+1) = \frac{w(x, t) + \varepsilon v(t)}{\|w(x, t) + \varepsilon v(t)\|} \quad (1)$$

なお、二つの発火パターンが交替する場合には、両方の発火パターンで共通に発火している素子について式 1 をそのまま適用し、発火と静止とを繰り返している素子については ε を半分にして式 1 を適用した。いずれの場合も発火しない素子の結合荷重ベクトルは変更しない。

3 計算機実験

計算機実験によって、入力を提示し、平衡状態を求め学習することをくりかえし、学習神経場の位置方向とランダム方向とにマップが広がることを確認した。

3.1 入力信号

PCM 音源からパイプオルガン、ピアノ、フルートの 3 つの楽器音の C, C#, D, D#, E の 5 つ、合計 15 個のデータを用いた。それぞれの音を 2 秒間発音させ、これを高速フーリエ変換 (FFT) したものを、2048 次元のベクトルとし、それを正規化したものを入力ベクトルとして用いた。

図 2 に各入力の類似関係の例を示す。この図より、これら 15 個の入力ベクトルは、同じ音色で異なる音程の入力間の類似度 (近さ) の方が、同じ音程で異なる音色の入力間の類似度よりも大きいことがわかる。よってこれらの入力信号について、音程の成分が第 1 主成分、音色の成分が第 2 主成分だということができる。またこの図より、“C, C#, D, D#, E” は、隣り合う音程同士がベクトルとして最も近く、音程が離れると類似度も小さくなることも分かる。

以上のことより、神経場の位置方向とランダム方向とのより長い方に音程のマップが、短い方に音色に関するマップが形成されるはずである。

3.2 結果

神経場は 1 次元で、300 個の素子からなるものを用いた。実験に用いた各種パラメータは表 1 のとおりである。マップができあがるまでの学習回数は、素子の数や層内結合に依存するが、今回用いたパラメータでは約 20 万から 30 万回のバッチ学習で完成した。

学習終了後の神経場の、各入力に対する局在興奮の重心位置を図 3 に示す。これより、入力の音程の成分が、局在興奮の重心として神経場上にトポロジカルにマップされて

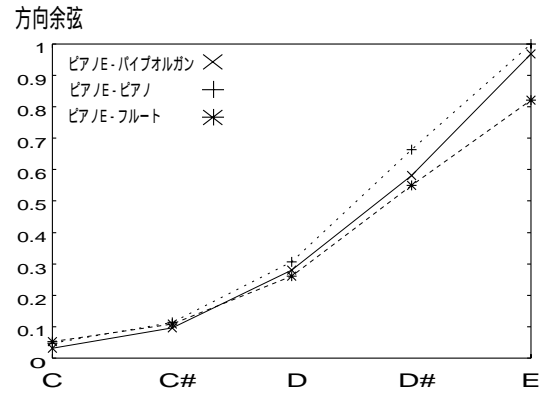


図 2: ピアノの E と各楽器各音程の音の違い。

ピアノの E と、各楽器の各音程との近さを、方向余弦で表した。同じ音程で異なる音色の入力の方が、同じ音色で異なる音程の入力よりもより遠い。また、各音程は隣り合う音程同士が最も近く、音程が離れると類似度も小さくなる。

表 1: 実際の計算機実験で用いたパラメータ

B, L, M は 2.1 節で述べた層内結合生成に関するパラメータ、 ε は学習係数、 c は 2.2 節の計算で使う、発火する素子の多さに関するパラメータである。

素子数 N	300
B	600
L	10
M	10
ε	$0 \leq t < 500$ 0.01 $t \geq 500$ $\frac{0.01}{t-499}$
c	$0 \leq t < 280$ $-2 + 0.01t$ $t \geq 280$ 0.8

いること、また、入力の音色の成分は局在興奮の重心位置ではほとんどコードされていないことがわかる。

次に、局在興奮の重心位置ではほとんどコードされていない入力信号の音色の成分がランダムな発火パターンで表現されていることを確かめた。図 4 に、ピアノの D に対する発火パターンと、全入力に対する発火パターンとの類似関係を示す。類似度は、発火パターン同士の方向余弦で表現した。同じ音程で音色の異なる入力同士に対する局在興奮の類似度が異なることから、それら局在興奮の発火パターンが異なったものであることがわかる。ピアノ以外の各楽器についても同様に、同じ音程の入力でも音色が異なればそれに対応する発火パターンも異なるという結果が得られている。

これらのことより、入力の音色の成分が、ランダムな発火パターンで表現されていることがわかる。以上の結果より、入力信号の音程の成分が神経場の位置方向に、音色の成分がランダム方向にマップされていることが確認できる。

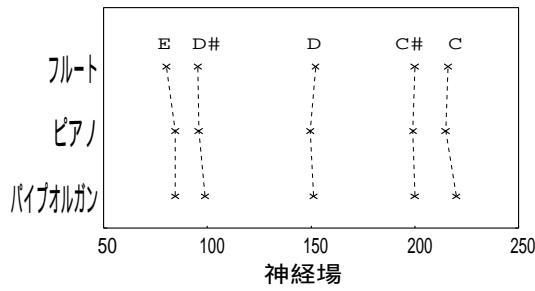


図 3: 音程, 音色と神経場の局在興奮の重心.

入力信号の音程と音色と, そのときの局在興奮の重心をプロットし, 同じ音程を持つものを破線で結んだ. この図より, 入力信号の音程の成分が, 局在興奮の重心として神経場上にトポロジカルにマップされていることと, 音色の成分は局在興奮の重心位置ではほとんどコードされていないことがわかる.

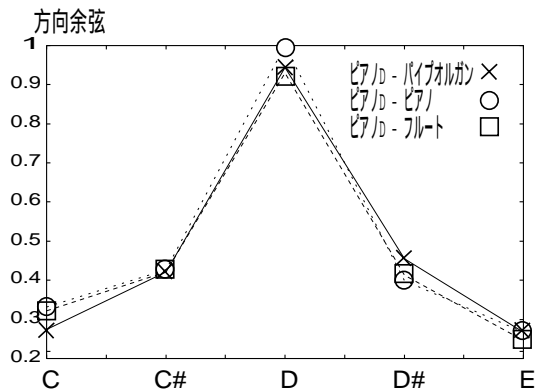


図 4: ピアノの D に対応する局在興奮と, その他の全ての入力に対応する局在興奮との重なり.

同じ音程で音色の異なる入力同士に対する局在興奮の距離が異なることから, それら局在興奮の発火パターンが異なったものであることがわかる. ピアノ以外の各楽器についても同様に, 同じ音程の入力でも, 音色が異なるとそれに対応する発火パターンは異なるとう結果が得られている. これらのことより, 入力の音色の成分が, ランダムな発火パターンで表現されていることがわかる.

4 考察

4.1 音色の成分を位置方向にマップすることについて

表 1 のパラメータを用いた今回の実験では, 入力信号空間の第 1 主成分である音程が神経場の位置方向に, 第 2 主成分である音色がランダム方向にマップされた.

ところで, 本モデルのマップのトポロジーは局在興奮パターンの重なりによって作られるので, 神経場の位置方向とランダム方向とのどちらに入力信号の第 1 主成分がマップされるかは, 局在興奮の集合の神経場の位置方向への長さ, ランダム方向への長さとの大小関係に依存する. 今回の入力データの場合, より長い方向に音程が, 短い方向に音色が抽出されるはずである.

このことより, 局在興奮をよりスパースにしてランダム

方向の距離を長くし, かつ局在興奮の幅をより広くして位置方向への距離を短くすれば, 音程がランダム方向にマップされ, 音色が局在興奮の位置にマップされることが予想される. これらの性質は神経場の層内結合及び発火する素子の個数により決定される.

しかしながらランダム方向の距離の延長には, 多数の素子のよりスパースな発火が必要であり, そのためには神経場全体がさらに多数でなければならない. そして, 層内結合の生成にもランダム性があるため, 毎回の試行で安定した結果を得るためには, やはり多くの素子が必要である. また本モデルでは層内結合に基づく発火のダイナミクスの計算が毎回必要であるため, 計算機の計算速度の問題から, 数値実験で試すことのできる神経場の素子数は比較的少数に限られる. このため今回は, 音色が局在興奮の位置として安定にマップされるような実際のモデルは作成できなかった.

しかしそれでも, 十分にスパースな局在興奮をもつ, 十分に大きな神経場を用いれば, 本モデルの神経場で, 位置方向に音色の成分がマップされる可能性は充分にある. このことは本モデルと良く似たモデルを用いて, 人工的なデータを用いた計算機実験で確認している.

4.2 従来のモデルとの関係

本モデルでは通常, スパースな局在興奮パターンを得るために層内結合生成のパラメータ M, B や発火する素子の数に関するパラメータ c を, 発火パターンがスパースになるように調節する. しかし逆に, これらのパラメータを調節することで, 密な安定発火パターンをもつ神経場を作成することもできる. そのとき本モデルは, 従来の自己組織マップ形成のモデルとほとんど同じものとなり, よって本モデルは従来の自己組織マップ形成のモデルを拡張したものであるということが出来る.

参考文献

- [1] T.Kohonen, Springer Sries in Information Sciences: Self-Organizing Maps, Springer, Berlin, 1995.
- [2] D.J. Willshaw and C. von der Malsburg (1976), How patterned neural connections can be set up by self-organization, *Proc. R. Soc., Lond. B.*, 194, pp.431-445.
- [3] “連続多重マップによる主成分の自己組織的抽出”, 信学技報, Vol.98, No.623, NC97-124(1998-03), 1998.

謝辞 本研究を行うにあたりお忙しい中多大なご支援を頂いた庄野逸博士および入力データを提供して頂いた紀田馨さんに心から感謝致します.