

F-20

エネルギースペクトルによる音楽情報の自己組織化分析  
Energy Spectrum-based Analysis of Musical Works using Self-organizing Map馬杉 正男\*  
Masao MASUGI\*

## 1. はじめに

音楽情報の特徴量に基づいて、楽曲パターンを分析・評価する取り組みは、音楽からの感情値の抽出や音楽検索の応用等の検討を進める上でも重要である。しかしながら、音楽の印象は、音高、音量、リズム等の多様な要因の組み合わせに依存するため、特徴量の選択問題を含めてさらなる評価手法の検討が課題となっている [1]-[4]。

本報告は、自己組織化マップ [5], [6] を用いた音楽情報の評価方法に関するものである。ここでは、複数の周波数帯域別にエネルギースペクトルを導出して、自己組織化学習する方法の概要について述べるとともに、音楽作品のクラスタリング評価を通して、適用手法の有効性を検証する。

## 2. 自己組織化マップによる音楽情報の評価法

## (1) 自己組織化マップの処理概要

本報告では、文献 [5] の方法を用いて入力データを2次元平面上に自己組織化マッピングする方法を想定する。まず、多次元入力ベクトルと2次元出力層上の参照ユニット  $i$  との結合関係を図1に示す。同図において、入力ベクトル  $\mathbf{x}$  と出力層（競合層）上のユニット  $i$  が参照ベクトル  $\mathbf{m}_i$  により結合した様子を示しており、離散ステップを  $t$  として、次式による処理に基づいて学習を行う。

$$\mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h(t) \{ \mathbf{x}(t) - \mathbf{m}_i(t) \} \quad (1)$$

ここで、 $h(t)$  は学習係数を考慮した近傍関数であり、離散ステップ  $t$  の増加とともに単調減少する。学習過程では、距離  $|\mathbf{x}(t) - \mathbf{m}_i(t)|$  を最小にするノード  $i$  を探索し、類似度に対応するように入力ベクトルが2次元出力層上にマッピングされる。

## (2) 音楽情報の特徴量の抽出

人間が受ける音楽や音刺激の印象は、音高、音量、リズム等の多様な要素の組み合わせに依存し、時間変動及び周波数特性の両面から、その特徴量を抽出・評価することが求められる。本報告では、音楽情報中の周波数軸上に分布するエネルギー量に注目して、複数の周波数帯域別に導出したエネルギースペクトルを元に音楽作品を自己組織化マッピング評価する方法について検討する。

いまここで、周波数帯域 ( $\omega_1 \sim \omega_2$ ) のエネルギー

スペクトル  $E$  は次式で定義できる。

$$E(\omega_1, \omega_2) = \int_{\omega_2}^{\omega_1} |P(f)|^2 df \quad (2)$$

但し、上式において、 $P$  は周波数スペクトルであり、時間-周波数の変換に際してはハニング関数を窓関数として用いる。

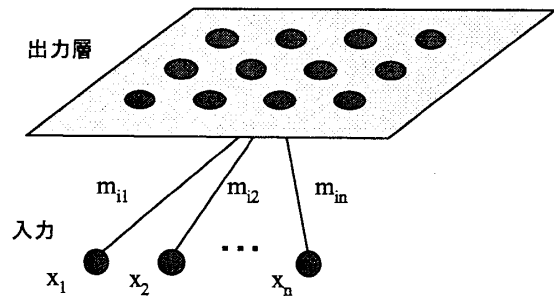


図1. 自己組織化マップイメージ

## 3. 評価例

本例では、6つの音楽作品 (A: ピアノソロ [北米]、B: トラッド系民族音楽 [ギリシャ]、C: クラシック・弦楽四重奏 [英国]、D: クラシック・ロック融合 [英国]、E: ラテン系ポップス [スペイン]、F: ハードロック [英国]) の各開始1分長を評価対象とし、各々のエネルギースペクトルを導出する。またエネルギースペクトルの導出に際しては、標準音高 (440Hz) を考慮して、DC~110Hz [F1]、110~220Hz [F2]、220~440Hz [F3]、440~880Hz [F4]、880~1760Hz [F5]、1760~22 kHz [F6] の6つの周波数帯域に分割する。なお、音源は市販の音楽CDとし、PC上にてWaveファイルを作成した後に時間-周波数変換を行う。

まず、全エネルギー量について、各作品間の相対的な関係を比較した結果を図2に示す。同図では、作品Aの値を1として規格化しており、作品Aに対して他作品のエネルギー量が20~600倍程度であり、人間に与える音刺激として大きな較差があることが確認できる。次に、各作品中に占める帯域別エネルギーの分布レベルを図3に示す。同図より、各作品における各帯域別エネルギーが様々な分布パターンを示しており、周波数成分の比率やエネルギー特性に違いがあることがわかる。

続いて、2次元平面上への各作品のマッピング例を図4に示す。同図の学習では、5×5の正方形マップに対して、各作品の6帯域分のエネルギースペクトル

\*: 日本電信電話株式会社 NTTネットワークサービスシステム研究所, NTT Network Service Systems Laboratories, NTT Corporation

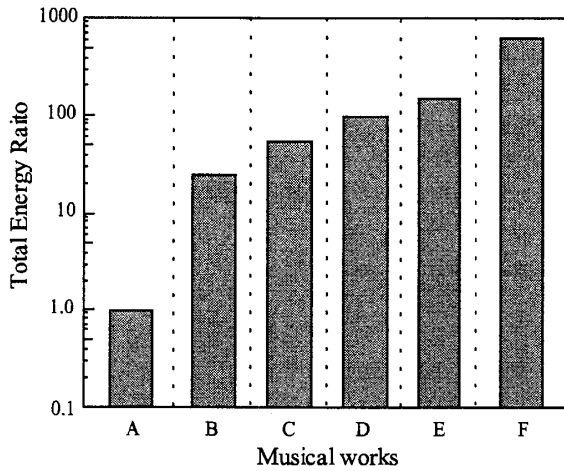


図2. 作品間の全エネルギー相対比

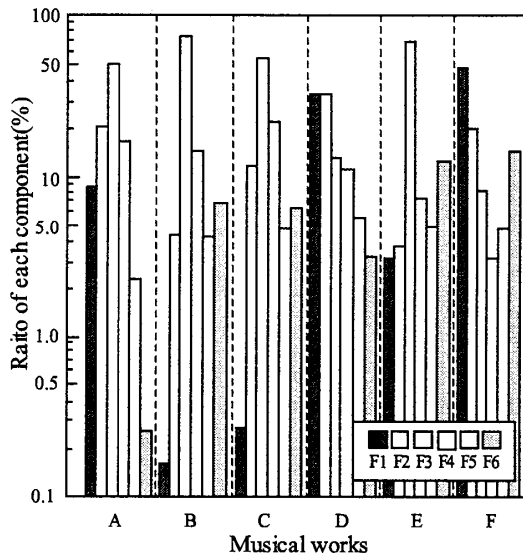


図3. 各作品に占める帯域別エネルギー比

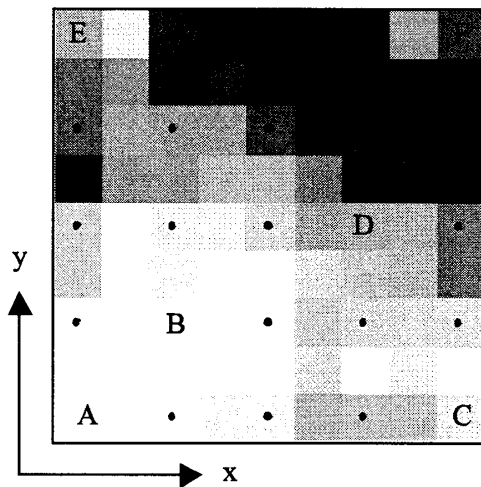


図4. マッピング例

ルを入力変数として与えている。また学習回数は5000回とし、ガウス型関数を近傍関数として採用している。さらに図4では、マップユニット間のクラス特性の差が大きい程、より濃い色が割り当てられている[6]。

図4において、2次元平面上の左下箇所が全周波数帯域にわたってエネルギーレベルが比較的低い領域に対応し、作品Aが最も左下点に配置されていることが確認できる。また上方部は、エネルギーレベルが高い領域に相当し、全エネルギー合計値で最大レベルを有する作品Fが右上点に配置されていることもわかる。一方、全エネルギー合計値で2番目、周波数帯域220~440Hz [F3]について最大エネルギーを有する作品Eは左上点に配置されていることも図4から確認できる。

本評価例は、開始1分長を対象としたものであり、処理時間長、対象周波数領域、周波数分割数等についてさらなる検証が必要であるが、音楽作品の特徴に応じて2次元平面上にマッピングされることが確認され、音楽特性のパターン分析への適用性を示唆するものであると考えられる。

#### 4. まとめ

本報告では、自己組織化マップを用いた音楽情報の処理方法の概要と評価例について述べた。自己組織化マップへの入力データとしては、複数の周波数帯域別に導出したエネルギーベクトルを用いて、音楽作品開始1分長のデータを元に学習処理を行った。その結果、各音楽作品のエネルギー分布特性に応じて2次元平面上にマッピングされ、音楽分析への適用の可能性を示すものとなった。

今後は、音楽信号の処理時間長、周波数分割領域数、自己組織化マップへの入力パラメータ選択、等の評価を行っていく予定である。

#### (参考文献)

- [1] 片寄, 今井, 井口, “音楽における感性情報抽出の試み”, 人工知能学誌, Vol. 3, No. 6, pp. 748-754, 1988-11.
- [2] 坂本, 梶川, 野村, “音楽感性空間における非線形判別分析を用いた曲印象グループの分割”, 情処学論, Vol. 40, No. 4, pp. 1901-1909, 1999-8.
- [3] 後藤, “音響信号を対象としたメロディーとベースの音高推定”, 信学論D-II, Vol. J84-D-II, No. 1, pp. 12-22, 2001-1.
- [4] 堀野, 佐藤, 黒木, 北上, “旋律の音高変化に基づく楽曲特徴量についての検討”, 情報処理学会・研究報告MUS40-5, pp. 29-34, 2001-5.
- [5] T. Kohonen, “Self-Organizing maps,” Springer, Berlin, Heidelberg (Second extended edition), Chapter 3, 1997.
- [6] A. Ultsch & H. P. Siemon, “Kohonen’s Self-Organizing Feature Maps for Exploratory Data Analysis,” in Proc. of INNC’90, Int. Neural Network Conference, Netherlands, pp. 305-308, 1990.