

エネルギー変化の線形予測符号化に基づく リズム特徴量を用いた音楽印象識別

三好 真人^{1,a)} 柘植 覚^{2,b)} 福見 稔^{1,c)}

受付日 2012年7月1日, 採録日 2013年1月11日

概要: 本論文では音楽の印象識別を高精度で行う特徴量として, 線形予測符号化に基づくリズム特徴量 (Rhythm feature based on Linear Predictive Coding: RLPC) を提案する. RLPC は, 音響信号のエネルギー変化に対して線形予測符号化を適用することにより求められるケプストラムであり, 音楽におけるリズムの周期性をとらえることが可能である. 7つの印象に対する音楽印象識別実験により, ジャンル分類ならびに印象分類における5種類の従来のリズム特徴量との比較を行った. 実験結果より, RLPCを用いた場合の平均識別率は83.7%であり, 従来のリズム特徴量を用いた場合より1.3ポイント高い識別率が得られた. さらに, 音量・音色・和音特徴量にRLPCを併用した場合の平均識別率は89.5%であり, 音量・音色・和音特徴量のみを用いる場合と比較して2.0ポイント, 従来のリズム特徴量を併用した場合よりも0.6ポイント高い識別率が得られた. また, 各印象においてRLPCと従来のリズム特徴量で仮説検定を行った結果, 4種類の従来のリズム特徴量に対して, RLPCを用いた場合の識別精度が有意であった.

キーワード: 音楽印象識別, リズム特徴量, 線形予測符号化

Rhythm Features Based on Linear Predictive Coding of Energy Variations for Musical Mood Classification

MASATO MIYOSHI^{1,a)} SATORU TSUGE^{2,b)} MINORU FUKUMI^{1,c)}

Received: July 1, 2012, Accepted: January 11, 2013

Abstract: In this paper, we propose a novel rhythm feature, which we call Rhythm feature based on Linear Predictive Coding (RLPC), to improve mood classification performance. The proposed feature is extracted with Linear Predictive Coding (LPC) on energy variations of an audio signal and is able to represent periodicity of rhythm in musical audio signals. To evaluate the proposed feature in comparison with 5 conventional rhythm features, mood classification experiments were conducted for 7 moods. From these experimental results, average accuracy of the proposed feature was 83.7% and 1.1 point higher than that of the conventional features. In addition, in case of combining base features, which indicate intensity, timbre, and harmony features, with RLPC, average accuracy was 89.5%. The accuracy was 2.0 point higher than that of base features and 0.6 point higher than that of combining base features with the conventional features. From results of hypothesis test on each mood, accuracies of the proposed feature were significant against that of the 4 conventional features.

Keywords: musical mood classification, rhythm feature, linear predictive coding

¹ 徳島大学
The University of Tokushima, Tokushima 770–8506, Japan
² 大同大学
Daido University, Nagoya, Aichi 457–8530, Japan
a) mirasilh@gmail.com
b) tsuge@daido-it.ac.jp
c) fukumi@is.tokushima-u.ac.jp

1. はじめに

近年, インターネットによる音楽配信や大容量の携帯型オーディオプレイヤー・PCなどの出現によって, 膨大な音楽データを個人で保有することが可能となり, 容易に利用できるようになってきている. このような膨大な音楽データ

を効率的に扱うために、有効な音楽検索システムが要求されている。音楽検索システムにおいて利用される検索キーとしては、アーティスト名や音楽ジャンル名、音楽印象などがあげられる。音楽印象による音楽検索システムは、「明るい」や「穏やかな」といった印象を検索キーとして使用し、指定された印象と一致する音楽を検索する。このようなシステムでは、アーティスト名や音楽ジャンル名といった音楽に関する情報について、ユーザが未知であっても検索をすることが可能であるという利点がある。また、直感的にユーザの嗜好に応じた音楽検索を行うことが可能である。印象による音楽検索システムを実現するためには、検索に使用される全音楽データに対してあらかじめ適切な印象情報を付与しておく必要がある。しかし、データ数が膨大である場合、このような作業を人手で行うことは困難である。そこで、印象付与を自動化するために自動音楽印象識別手法が必要となる。ユーザが音楽を聴くことで受ける印象を印象識別手法により識別し、識別結果を印象情報として付与することで自動印象付与が可能である。

現在までに音楽の印象識別手法に関しては多くの研究がなされてきた [1], [2], [3], [4], [5], [6], [7], [8], [9]。これらの手法では音楽内容を示す音量・音色・リズム・和音などに関連する音響特徴量を多数併用し、サポートベクタマシン (Support Vector Machine : SVM) などの機械学習手法によって印象識別することが一般的である [2], [3], [6], [7]。しかし、文献 [2], [3], [6], [7] で報告されている平均印象識別精度は 80.0% から 85.0% であり、識別率が 80.0% 以下となる印象も存在することから、識別精度は十分とはいえず、さらなる改善が必要である。従来より、印象識別手法においてはフレーム内エネルギーやスペクトラルセントロイド、メル周波数ケプストラム係数などの音量・音色特徴量は抽出が容易であるため多用される。しかし、リズム・旋律・和音特徴量は音量・音色特徴量には含まれない音楽の構造に関する情報 (リズムパターンや和音遷移パターンなど) を含んでおり音楽印象識別に有効であると考えられるが、正確な抽出が困難であるため、あまり使用されていない。そこで、さらなる音楽印象識別の改善のためにはリズム・旋律・和音特徴量などの音楽の構造に関する特徴量の研究が必要である。自動音楽ジャンル分類において、音楽の構造に関する特徴量は有効であることが報告されている [10], [11], [12], [13], [14], [15]。ジャンル分類と印象分類は分類するクラスは異なるが、音楽のクラス分類問題としては類似していると考えられる。両分類とも音楽の構造 (リズム, 和音など) をとらえた特徴を用いてクラスを分類する点は共通であるため、ジャンル分類で有効性が確認されている特徴は、印象分類においても有効性が高いことが期待される。しかし、音楽印象識別についてこれら特徴量に着目した研究は少数であり [16], [17], 十分に有効性が検討されているとはいえない。そこで、本論文では、

これらの音響特徴量の中からリズム特徴量に着目する。リズムは旋律や和音と並んで音楽における 3 大要素の 1 つであり、音楽を構成するうえで重要な要素である。また、バスドラムやスネアドラムなどの打楽器音によって周期的なリズムが演奏されることが多いため、周期的なリズム特徴をとらえることは印象識別に有効であると考えられる。

従来のリズム特徴量として、発音検出に基づくリズム特徴量である Inter-Onset Interval Histogram (IOIH) 特徴量 [11], [18] や周期性に基づくリズム特徴量である Beat Histogram (BH) 特徴量 [19], Periodicity Histogram (PH) 特徴量 [11], [12] が存在する。IOIH は、検出されるすべての発音間で発音間隔 (Inter-Onset Interval : IOI) を計算し、各発音間隔において IOI 数を累積することで求められるリズム特徴量であり、音楽音響信号における発音数および優勢な発音間隔を表現する。周期性に基づくリズム特徴量である BH は、一定時間長の音楽音響信号から計算される自己相関関数のピーク数を累積することで求められるリズム特徴量であり、優勢なリズムの周期を表現する。また、PH はコムフィルタによって出力されるエネルギーを累積することで求められるリズム特徴量であり、優勢なリズムの周期を表現する。

本論文では、音楽音響信号において周期的に演奏されるリズムを示す特徴量として線形予測符号化に基づくリズム特徴量 (Rhythm feature based on Linear Predictive Coding : RLPC) を提案する。RLPC は、スペクトルのエネルギー変化を表現するスペクトルエネルギー信号に対して、線形予測符号化によるスペクトル包絡をケプストラム係数へ変換することで抽出される。スペクトルエネルギー信号が周期的に変化する場合はスペクトル包絡においてリズム周期に対応する周波数でピークが発生し、周期的に変化しない場合はピークが発生しない。そのため、RLPC は音楽における周期的に演奏されるリズムをとらえることが可能であり、印象識別に有効であると考えられる。

本論文は以下のように構成される。2 章で提案リズム特徴量である RLPC について述べる。3 章で RLPC の有効性比較のための従来のリズム特徴量、および補助特徴量としての有効性を検証するために使用する音量・音色・和音特徴量について述べる。4 章で RLPC の有効性を検証するための音楽印象識別実験について述べる。5 章で本論文のまとめと今後の課題について述べる。

2. 線形予測符号化に基づくリズム特徴量

本章では、音楽音響信号におけるバスドラムやスネアドラムなどの打楽器が演奏する周期的なリズムを表現することが可能な線形予測符号化に基づくリズム特徴量 (Rhythm feature based on Linear Predictive Coding : RLPC) について述べる。RLPC の抽出手順を図 1 に示し、以下で処理の流れに従い各処理について詳細に述べる。

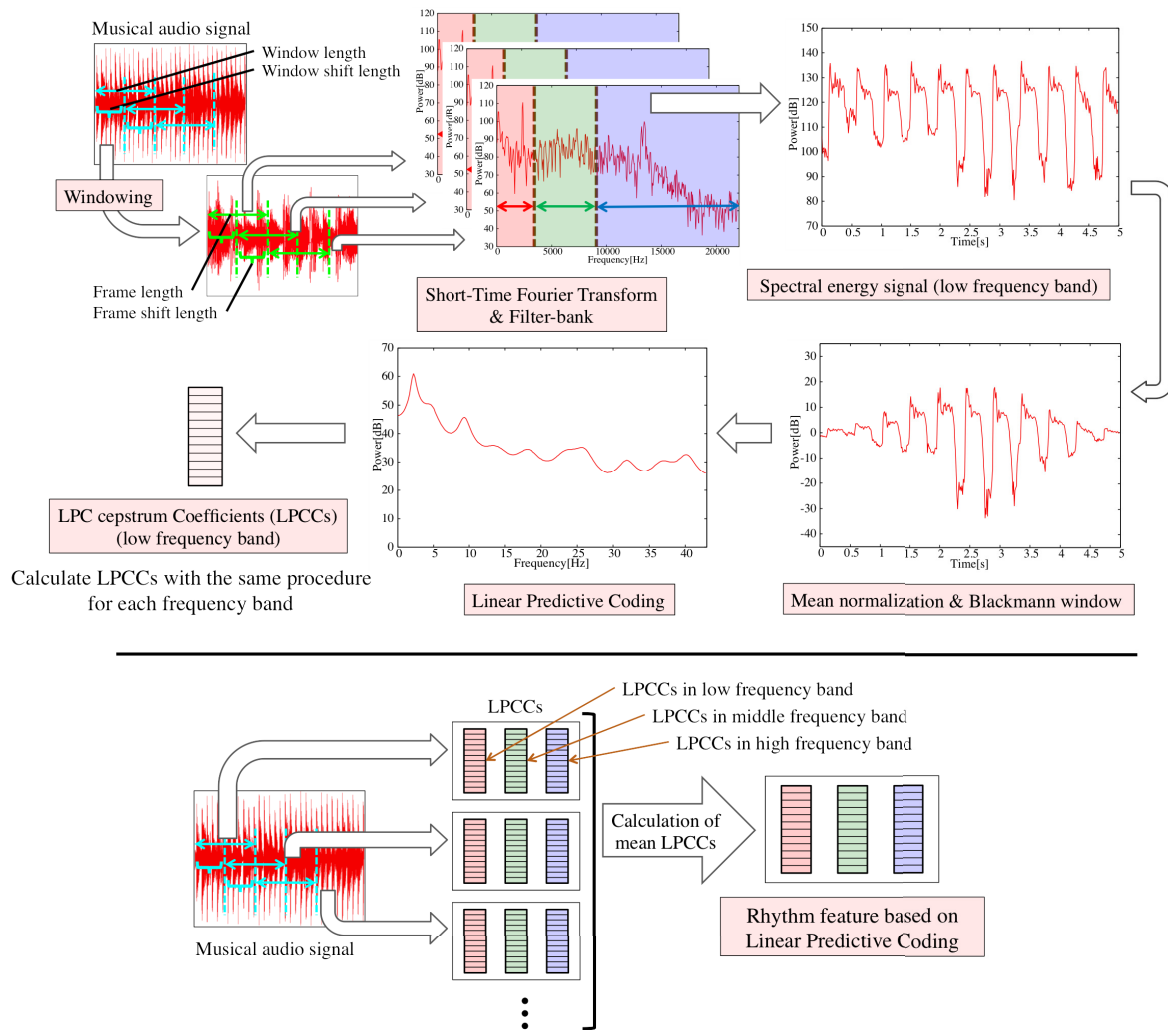


図 1 RLPC の抽出手順
Fig. 1 Extraction procedure of RLPC.

2.1 音響信号の分割・フィルタバンク処理

音楽において一定時間はリズムの変化がないと仮定し、入力音響信号から一定時間長の音響信号を切り出し、その区間ごとに RLPC を抽出する。本論文では、音響信号を切り出す処理を音響信号のウィンドウ化 (Windowing) と呼び、ウィンドウの長さをウィンドウ長、ウィンドウを時間シフトさせる時間をウィンドウシフト長と呼ぶ。ウィンドウ化処理は、入力音響信号 ($x(n)$) に対し以下の処理を行うことである。

$$x_w(n) = x(n + (w - 1) \cdot M_w) \quad (n = 1, \dots, N_w, w = 1, \dots, W) \quad (1)$$

ここで、 $x_w(n)$ はウィンドウ w 内の音響信号、 M_w はウィンドウシフト長、 N_w はウィンドウ長、 W はウィンドウ総数を示す。4 章で述べる音楽印象識別実験において、ウィンドウ長およびウィンドウシフト長は 5.0s および 0.5s に設定した。

切り出されたウィンドウ単位の音響信号におけるリズム周期は、各ウィンドウに存在する楽器音の発音を検出し、そ

の周期性をとらえることにより求めることができる。楽器音の発音は、スペクトルエネルギーが急峻に変化した部分に出現しやすいことが知られている [20]。そこで、ウィンドウ化処理により切り出された音響信号の局所的なスペクトルの変化を求めるために、式 (2) によりフレーム分割および窓掛け処理を行い、各フレームに対して短時間フーリエ変換 (Short-Time Fourier Transform : STFT) を行う。

$$x_{w,k}(n) = x_w(n + (k - 1) \cdot M_u) \cdot W(n) \quad (n = 1, \dots, N_u, k = 1, \dots, K) \quad (2)$$

ここで、 $x_{w,k}(n)$ はウィンドウ w の k 番目のフレーム内の音響信号、 M_u 、 N_u はフレームシフト長およびフレーム長、 K はフレーム総数、 $W(n)$ は窓関数を示し、この信号から求められる f Hz におけるフーリエスペクトルを $X_{w,k}(f)$ とする。

音楽データの多くは複数の楽器の演奏により構成され、各楽器の周波数帯域は異なる。各楽器ごとのリズムパターンをとらえるため、フーリエスペクトルに対しフィルタバンク処理を行う。この処理はビートトラッキングや発音検

出において有効性が示されている [20], [21]. 本論文では, 周波数帯域分割に関する予備実験において, 最も精度が高かった以下の3つの周波数帯域分割を用いる.

● 低周波数帯域 (0~300 Hz)

ベースラインやバスドラム音が出現する周波数帯域であり, これらを演奏する楽器のリズムパターンをとらえる.

● 中周波数帯域 (300~3,000 Hz)

最も多くの楽器音 (弦楽器や打楽器, 管楽器など) が含まれる周波数帯域であり, これらを演奏する楽器のリズムパターンをとらえる.

● 高周波数帯域 (3,000~22,050 Hz)

スネアドラム音など打楽器音が顕著に出現する周波数帯域であり, これらを演奏する楽器のリズムパターンをとらえる.

2.2 スペクトルエネルギー信号

スペクトルの変化を表現するため, 前節で求めたフィルタバンク処理されたフーリエスペクトルに対し, 各ウィンドウごとに各周波数帯域におけるパワーの総和を求め, パワーを総和することでスペクトルエネルギーの急峻な変化を強調する. さらに, 対数変換を行うことで人間の聴覚特性を反映する. 本論文では, スペクトルエネルギーの変化を表現する信号をスペクトルエネルギー信号と呼ぶ. ウィンドウ w の k 番目のフレームにおけるフィルタバンク処理された各周波数帯域 b ($b = 1, 2, 3$ (1は低周波数帯域, 2は中周波数帯域, 3は高周波数帯域を示す)) におけるスペクトルエネルギー信号 ($SE_{w,b}(k)$) は次式で計算される.

$$SE_{w,b}(k) = 10 \log_{10} \left\{ \sum_{f=f_{b,S}}^{f_{b,E}} |X_{w,k}(f)|^2 \right\} \quad (3)$$

ここで, $f_{b,S}$, $f_{b,E}$ は周波数帯域 b の開始および終了周波数を示す. 求められたスペクトルエネルギー信号は後の処理のため, 式 (4) に示すとおり平均 0 に正規化を行い, Blackmann 窓 ($W_{blk}(k)$) により窓掛け処理を行う.

$$\tilde{SE}_{w,b}(k) = (SE_{w,b}(k) - E[SE_{w,b}(k)]) \cdot W_{blk}(k) \quad (4)$$

ここで, $E[\cdot]$ は期待値を示す.

2.3 線形予測符号化

音楽音響信号において, 周期的にリズムが演奏される場合はスペクトルエネルギー信号のスペクトルにおけるリズム周期に対応する周波数でピークが発生し, そうでなければピークが発生しない. したがって, スペクトル包絡形状を特徴量とすることで周期的に演奏されるリズムを示すリズム特徴量が得られる. スペクトル包絡情報を得るために, スペクトルエネルギー信号に対して線形予測符号

化 (Linear Predictive Coding : LPC) を適用する. 線形予測符号化は現時刻の信号サンプルを過去の信号サンプルの線形結合により近似可能であると仮定したモデルであり, スペクトル包絡を推定するために有効なモデルである [22]. 平均値正規化を行ったスペクトルエネルギー信号 ($\tilde{SE}_{w,b}(k)$) に対し, 線形予測符号化を適用することで線形予測係数 ($a_{w,b}(i)$ ($i = 1, \dots, p$)) を計算する. 線形予測係数推定手法として Levinson-Durbin 法 [22] および Burg 法 [23] を用いて 4 章で述べる実験を行った結果, 大きな差が発生しなかったため, 本論文では Levinson 法を用いることとした.

求められた線形予測係数からスペクトル包絡形状を直接的に示す特徴量である LPC ケプストラム係数 (LPC cepstrum Coefficients : LPCCs) に変換する. LPCC は対数化された LPC スペクトル包絡の逆フーリエ変換として定義され, 線形予測係数よりも頑健で有効な特徴量であることが示されている [22]. ウィンドウ w における周波数帯域 b の LPCC ($c_{w,b}(o)$ ($o = 0, \dots, O$)) は式 (5)~(7) に示す再帰式により線形予測係数 ($a_{w,b}(i)$) から計算される.

$$c_{w,b}(0) = \log(\sigma_{w,b}^2) \quad (o = 0) \quad (5)$$

$$c_{w,b}(o) = -a_{w,b}(o) - \sum_{k=1}^{o-1} \frac{k}{o} \cdot c_{w,b}(k) \cdot a_{w,b}(o-k) \quad (1 \leq o \leq p) \quad (6)$$

$$c_{w,b}(o) = -\sum_{k=1}^p \frac{k}{o} \cdot c_{w,b}(k) \cdot a_{w,b}(o-k) \quad (o > p) \quad (7)$$

ここで, $\sigma_{w,b}$ は線形予測符号化におけるゲイン項を示す.

2.4 RLPC 抽出

RLPC ($R_b(o)$) は, 各周波数帯域において計算される LPCC の全ウィンドウに対する平均値として式 (8) で計算される.

$$R_b(o) = \frac{1}{W} \sum_{w=1}^W c_{w,b}(o) \quad (8)$$

周波数帯域を3つに分割し, 各周波数帯域における RLPC の次元数が $(O+1)$ であった場合, 音楽印象識別を行う各楽曲片からは $3 \times (O+1)$ 次元 ((周波数帯域) \times (次元数)) の RLPC が抽出される.

例として, 周期的なリズムを持つポピュラー音楽データの周波数帯域におけるスペクトルエネルギー信号と RLPC による対数スペクトルを図 2 および図 3 に示す. また, リズムが周期的でないクラシック音楽データの周波数帯域におけるスペクトルエネルギー信号と RLPC による対数スペクトルを図 4 および図 5 に示す. 図 2 のスペクトルエネルギー信号は周期的な変化を示すため, 図 3 の対数スペクトル包絡は周期に対応する周波数でピークを持ち, さらにその高調波でピークを持つ周期的で緩やかな

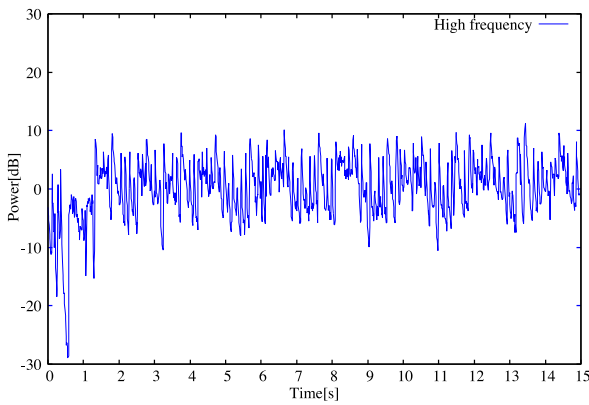


図 2 ポピュラー音楽データから求められた高周波数帯域におけるスペクトルエネルギー信号

Fig. 2 Spectral energy signal of a popular musical piece in the high frequency band.

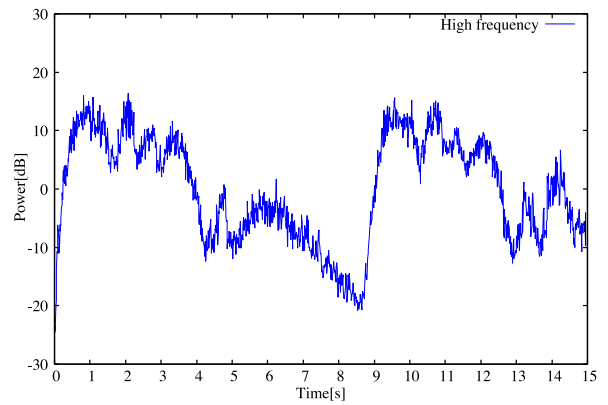


図 4 クラシック音楽データから求められた高周波数帯域におけるスペクトルエネルギー信号

Fig. 4 Spectral energy signal of a classical musical piece in the high frequency band.

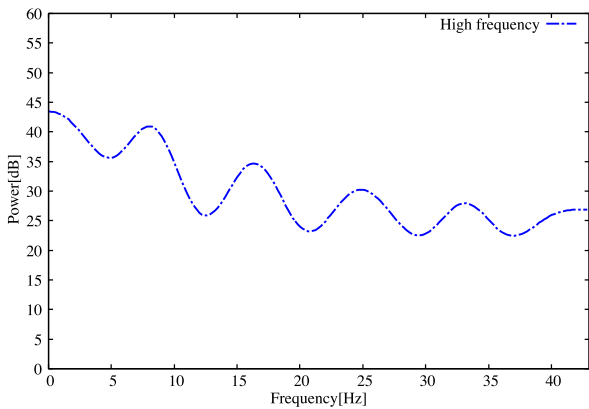


図 3 ポピュラー音楽データにおける高周波数帯域の対数スペクトル包絡

Fig. 3 Logarithm spectral envelope of a popular musical piece in the high frequency band.

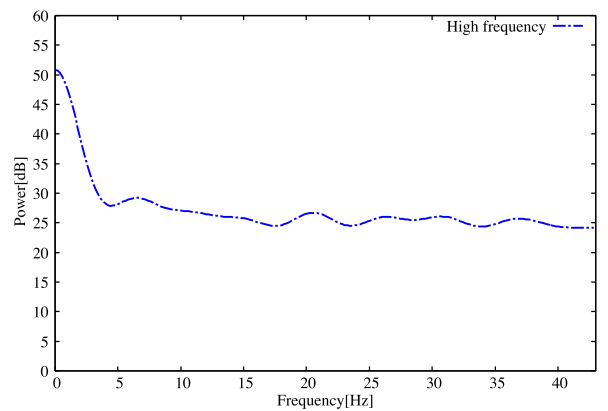


図 5 クラシック音楽データにおける高周波数帯域の対数スペクトル包絡

Fig. 5 Logarithm spectral envelope of a classical musical piece in the high frequency band.

スペクトル包絡形状となる。一方、図 4 のスペクトルエネルギー信号は緩やかな変化を示している。これより、図 5 の対数スペクトル包絡では主に低周波数成分にスペクトルエネルギーが集中するため、低周波数で急峻なスペクトル包絡形状となる。これらの図より、RLPC が音楽音響信号における周期的なリズムをとらえることが分かる。

3. 自動音楽分類のための音響特徴量

本章では、従来のリズム特徴量について述べる。リズム特徴量は、従来の音量・音色・和音特徴量などと併用することにより性能を改善することが可能である [16], [17]。そこで、従来の音量・音色・和音特徴量についても本章で述べる。

3.1 リズム特徴量

3.1.1 Inter-Onset Interval Histogram 特徴量

Inter-Onset Interval Histogram (IOIH) は、文献 [18] において音楽音響信号のリズム構造解析のために提案され、

音楽の自動ジャンル分類手法におけるリズム特徴量としても使用された [10], [11]。IOIH は、検出されるすべての発音間で発音間隔 (Inter-Onset Interval : IOI) を計算し、各発音間隔において IOI 数を累積することにより、音楽音響信号における発音数および優勢な発音間隔を表現する。本論文では、移動平均フィルタにより平滑化した各周波数帯域のスペクトルエネルギー信号に対して、一次回帰係数を求め、閾値処理により発音時間を検出する。各周波数帯域で検出された発音時間を 1 つにまとめ IOI を計算する*1。IOI 数を楽曲片全体で累積し、ガウス窓を畳み込むことで IOIH を得る。IOI 数を楽曲片全体で累積して得られたヒストグラムを $IO\tilde{I}H(\tau)$ 、ガウス窓を $W_g(l)$ とすれば IOIH ($IOIH(\tau)$) は次式で計算される

$$IOIH(\tau) = \sum_{l=1}^L W_g(l) \cdot IO\tilde{I}H(\tau - l) \quad (9)$$

$$W_g(l) = \exp\left(-\left(\frac{l}{L}\right)^2\right) \quad (10)$$

*1 ここで重複する発音時間は 1 つの発音時間として扱う。

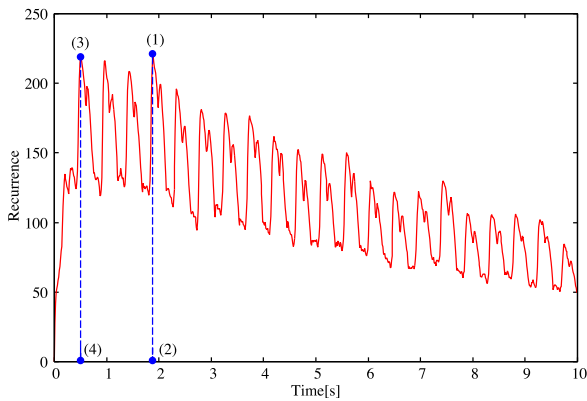


図 6 Inter-Onset Interval Histogram
Fig. 6 Inter-Onset Interval Histogram.

ここで L はガウス窓の窓幅を示す。図 6 に求められた IOIH の例を示す。IOIH から文献 [11] で使用された特徴量を含む以下の 11 次元特徴量を抽出する。また、図 6 内に以下の特徴量 (1) から (4) の対応を示す。

- (1) 1 番目に大きなピーク値,
- (2) 1 番目に大きなピークが示す周期,
- (3) 2 番目に大きなピーク値,
- (4) 2 番目に大きなピークが示す周期,
- (5) 1 番目に大きなピークが示す周期と
2 番目に大きなピーク値が示す周期の比,
- (6) 算術平均 [11], (7) 幾何平均 [11],
- (8) 総エネルギー [11]*2, (9) 平坦性 [11],
- (10) 歪度 [11], (11) 尖度 [11]

3.1.2 Power Spectrum Peaks 特徴量

音楽印象推定に関する文献 [4], [9] において、一定時間長の音響信号から検出される楽音の発音数がリズム特徴量として使用されている。そこで、類似特徴量として、スペクトルエネルギー信号において検出される楽音の発音数とその変化度合いをリズム特徴量として使用する。本論文では、このようなリズム特徴量を Power Spectrum Peaks (PSPs) 特徴量と呼ぶ。3.1.1 項で述べた発音検出により検出される発音数の全ウィンドウにおける平均値および標準偏差の 6 次元特徴量 ((2 次元特徴量) × (3 つの周波数帯域)) を PSPs 特徴量として使用する。

3.1.3 Beat Histogram 特徴量

Beat Histogram (BH) は、文献 [19] において音楽の自動ジャンル分類のためのリズム特徴量として提案され、印象識別におけるリズム特徴量としても使用された [3], [14]。BH は、一定時間長の信号から計算される自己相関関数のピーク数を累積することにより、音楽音響信号における優勢なリズムの周期を表現する。本論文では、移動平均フィルタによって平滑化した各周波数帯域のスペクトルエネルギー信号に対して、標準化を適用し、自己相関関数を計算する。自己相関関数の一次回帰係数を計算し、閾値処理に

*2 本論文では dB 値に変換して使用。

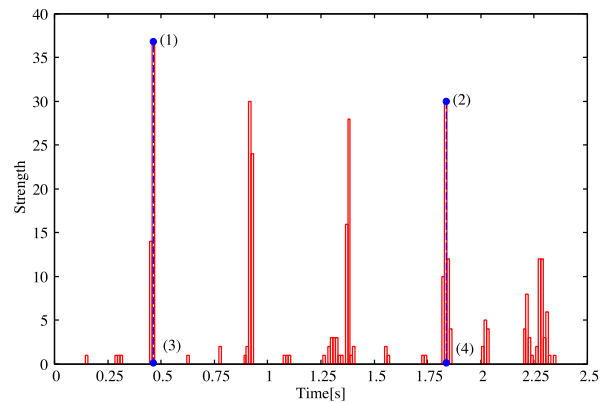


図 7 Beat Histogram
Fig. 7 Beat Histogram.

よりピークを検出する。検出された各周波数帯域における自己相関関数のピーク数を曲全体で累積し、BH を得る。図 7 に求められた BH の例を示す。BH から文献 [19] で使用された以下の 5 次元特徴量を抽出する。また、図 7 内に以下の特徴量 (1) から (4) の対応を示す。

- (1) 1 番目に大きなピーク値 [19],
- (2) 1 番目に大きなピークが示す周期 [19],
- (3) 2 番目に大きなピーク値 [19],
- (4) 2 番目に大きなピークが示す周期 [19],
- (5) ヒストグラム値の総和 [19]

3.1.4 Auto-Correlation Function Peaks 特徴量

3.1.3 項で述べた BH 特徴量だけでなく、リズムの周期性を検出するために自己相関関数は広く使用されている [10], [21]。そこで、自己相関関数を用いたリズム特徴量として、一定時間長のスペクトルエネルギー信号における自己相関関数のピーク数とその変化度合いをリズム特徴量として使用する。本論文では、このようなリズム特徴量を Auto-Correlation Function Peaks (ACFPs) 特徴量と呼ぶ。3.1.3 項と同様に、各ウィンドウの各周波数帯域における自己相関関数を計算し、ピークを検出する。検出された各周波数帯域におけるピーク数の全ウィンドウにおける平均値および標準偏差の 6 次元特徴量 ((2 次元特徴量) × (3 つの周波数帯域)) を ACFPs 特徴量として使用する。

3.1.5 Periodicity Histogram 特徴量

Periodicity Histogram (PH) は、文献 [12] において音楽間の類似性尺度として提案され、音楽の自動ジャンル分類手法におけるリズム特徴量としても使用された [11]。PH は、音楽音響信号を人間の聴覚特性を考慮した信号表現へ変換し、コムフィルタバンクによって出力されるエネルギーを累積することにより、音楽音響信号における優勢なリズム周期を表現する。PH は文献 [12] で述べられた方法で抽出する。図 8 に求められた PH の例を示す。PH から文献 [11] で使用された特徴量を含む以下の 12 次元特徴量を抽出する。また、図 8 内に以下の特徴量 (1) から (4) の対応を示す。

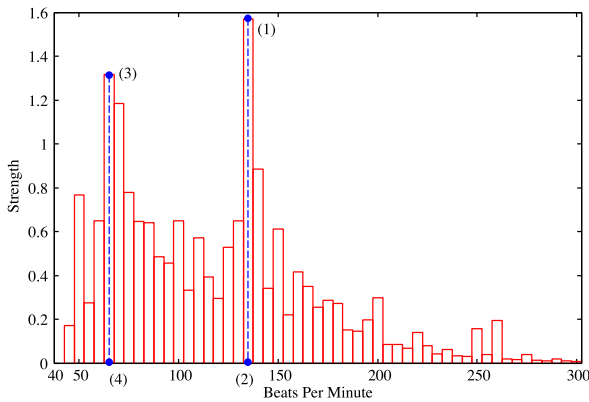


図 8 Periodicity Histogram

Fig. 8 Periodicity Histogram.

- (1) 1 番目に大きなピーク値,
- (2) 1 番目に大きなピークが示す周期 [11],
- (3) 2 番目に大きなピーク値,
- (4) 2 番目に大きなピークが示す周期,
- (5) 1 番目に大きなピークが示す周期と
2 番目に大きなピーク値が示す周期の比 [11],
- (6) PH のセントロイド [11], (7) 算術平均 [11],
- (8) 幾何平均, (9) ヒストグラム値の総和 [11],
- (10) 平坦性, (11) 歪度, (12) 尖度

3.1.6 従来のリズム特徴量と提案リズム特徴量の比較

発音検出に基づくリズム特徴量である IOIH 特徴量と PSPs 特徴量は、ヒストグラム値や平均発音数によって音楽音響信号に含まれる発音数や優勢な発音間隔を表現する。たとえば、音楽音響信号に含まれる発音数が多い場合は、多くの IOI 数を累積することになるため、ヒストグラム値が大きくなる。しかし、周期的に演奏されていても発音数が少なければ、ヒストグラム値が小さくなる。このように、これらの特徴量は音楽音響信号における発音数に基づくため、音楽音響信号における周期的なりズム特徴をとらえていない。周期性に基づくリズム特徴量である BH 特徴量、ACFPs 特徴量、PH 特徴量は、ヒストグラム値や平均ピーク数によって優勢なりズム周期を表現する。音楽音響信号においてテンポの速い周期的なりズムが演奏されていれば、ヒストグラム値や平均ピーク数は大きくなる。しかし、テンポの遅い周期的なりズム、もしくは周期的でないリズムが演奏されていれば、ヒストグラム値や平均ピーク数は小さくなるため、周期的なりズムの存在を明確に示すことができない。一方、提案特徴量においては、2.4 節で述べたように、音楽音響信号で周期的なりズムが演奏されていれば、対数スペクトル包絡においてピークが発生し、そうでなければピークが発生しない。そのため、従来の特徴量と比較して周期的なりズム特徴を明確に示すことができると考えられる。

3.2 音量・音色・和音特徴量

音量・音色・和音特徴量は、音楽印象識別において一般的に使用される音響特徴量であり、リズム特徴量と併用することにより識別精度を改善することが可能である [16], [17]. 本論文では、以下に示す音量・音色・和音特徴量を使用する。また、音量・音色・和音特徴量において、フレーム単位で抽出される特徴量は全フレームの平均値および標準偏差を音響特徴量として使用する。(・) は各音響特徴量の次元数を示す。

- 音量特徴量 (7 次元)
 - フレーム内エネルギー (2) [9],
 - 対数周波数パワー (2) [24],
 - Δ 対数周波数パワー (2) [24],
 - 低エネルギーフレーム数 (1) [19]
- 音色特徴量 (60 次元)
 - スペクトラルセントロイド (2) [19],
 - 周波数帯域幅 (2) [9],
 - スペクトラルロールオフ (2) [19],
 - スペクトラルフラックス (2) [19],
 - スペクトラルコサイン類似度 (2) [24],
 - 零交差点数 (2) [19],
 - メル周波数ケプストラム係数 (24) [19],
 - Δ メル周波数ケプストラム係数 (24) [19]
- 和音特徴量 (4 次元)
 - クロマベクトルフラックス (2) [24],
 - メジャーおよびマイナーコード成分 (2) [24]

4. 音楽印象識別実験

本章では、RLPC の有効性を検証するために行った以下に示す 2 つの音楽印象識別実験について述べる。

- (1) リズム特徴量による音楽印象識別実験
 - 2 章および 3.1 節で述べた各リズム特徴量を単独で使用したときの音楽印象識別精度比較
- (2) リズム特徴量と音量・音色・和音特徴量併用による音楽印象識別実験
 - 3.2 節で述べた音量・音色・和音特徴量 (以下では Base 特徴量と呼ぶ) と各リズム特徴量を併用したときの音楽印象識別精度比較

次節以降で使用する音楽データセットおよび実験条件、実験結果、考察について述べる。

4.1 音楽データセット

本実験では RWC 研究用音楽データベース [25] の「クラシック音楽データベース」、「ジャズ音楽データベース」、「ポピュラー音楽データベース」、「著作権切れ音楽データベース」の音楽、合計 226 曲を用いた。これらの音楽から一定の印象を与えられられる箇所を 15 秒ごと切り出し、合計 406 曲片の音楽データを作成し、実験に用いた。

表 1 各印象ラベルにおける音楽データ数

Table 1 The number of music data for each mood label.

| 印象ラベル | Yes データ数 | No データ数 | 総数 |
|---------|----------|---------|-----|
| 明るい | 171 | 76 | 247 |
| 賑やかな | 174 | 105 | 279 |
| アップテンポな | 105 | 150 | 255 |
| 軽快な | 170 | 72 | 242 |
| 穏やかな | 161 | 94 | 255 |
| 力強い | 71 | 117 | 188 |
| さわやかな | 70 | 73 | 143 |

音楽データのフォーマットは、ステレオ、サンプリング周波数 44.1 kHz、16 bit 量子化であり音響特徴量抽出時にはステレオからモノラルへ変換した。印象ラベルには、SD 法による音楽の印象測定に関する先行研究 [26] において、印象測定に用いられている形容詞の中から、「明るい」、「賑やかな」、「アップテンポな」、「軽快な」、「穏やかな」、「力強い」、「さわやかな」の 7 つを選択した。各印象ラベルにおける音楽データセットを構築するために全音楽データを用いてアノテーション実験を行った。アノテーション実験の被験者は 3 名である。各被験者は音楽データを試聴して 7 つの印象ラベルが示す印象を受けるかどうか判定した。本実験には全被験者の判定結果が一致した音楽データのみを使用した。各印象ラベルに対する音楽データ数を表 1 に示す。表 1 において、「Yes データ数」は印象ラベルが示す印象を受けると判定された音楽データの数、「No データ数」は印象ラベルが示す印象を受けないと判定された音楽データの数をそれぞれ示す。

4.2 実験条件

音響分析時のフレーム長およびフレームシフト長はリズム・音量・音色特徴量抽出時には 23.2 ms および 11.6 ms、和音特徴量抽出時には 185.8 ms および 80.0 ms、窓関数には Blackmann 窓を用いた。IOIH 特徴量を除くリズム特徴量におけるウィンドウ長およびウィンドウシフト長は 5.0 s および 0.5 s とした。RLPC 抽出のための線形予測次数は 15 次、各周波数帯域におけるケプストラム次数は 0 次から 17 次まで (18 次元) を使用した。したがって、RLPC の次元数は 54 次元 ((18 次元) × (3 つの周波数帯域)) である。IOIH 特徴量と PSPs 特徴量における移動平均フィルタ次数は 6 次と 3 次、発音検出のための閾値は 0.8 と 1.0 に各々設定した。BH 特徴量と ACFPs 特徴量における移動平均フィルタ次数は 5 次、ピーク検出のための閾値は 0.01 に設定した。低エネルギーフレーム数における閾値は全音楽データから抽出されたフレーム内エネルギーの平均値とした。スペクトラルロールオフにおける閾値は 0.8 に設定した。メル周波数ケプストラム係数と Δ メル周波数ケプストラム係数抽出に使用するメルフィルタバンクのチャンネル数は 60 とし、パワー項を除く低次項から 12 次元を使用

した。 Δ 対数周波数パワーと Δ メル周波数ケプストラム係数を計算するためのフレーム数は 3 フレームとした。これらの値は予備実験の結果、最適であった結果を用いた。

本実験における識別器には SVM を使用し、LIBSVM [27] により実装した。SVM の学習に用いる特徴量は各特徴量軸における最大値および最小値を用いて -1~1 の範囲に正規化した。SVM のカーネルには線形カーネルを使用し、ソフトマージンのパラメータは 0.1 とした。実験は 5-fold Cross-Validation により行った。各印象ラベルにおいて音楽データを 5 つに分割し、学習に 4 つの分割データ、評価に 1 つの分割データを使用した。これを 5 試行繰り返した。

4.3 実験結果

4.3.1 リズム特徴量の比較

各リズム特徴量における印象ラベルに対する平均識別率を表 2 に示す。表 2 における各印象ラベルの (・) 内は各試行で得られた識別率の標準偏差を示す。表 2 より、各リズム特徴量の全印象ラベルにおける平均識別率は、IOIH 特徴量では 81.5%、PSPs 特徴量では 82.3%、BH 特徴量では 80.0%、ACFPs 特徴量では 82.4%、PH 特徴量では 78.6%、RLPC では 83.7% であり、RLPC が最も高い平均識別率であることが分かる。この結果より、提案する RLPC は音楽印象識別に有効なリズム特徴量であるといえる。

次に、各印象ラベルの識別率を比較すると、「明るい」、「アップテンポな」、「穏やかな」、「さわやかな」において RLPC は最も高い識別率であることが分かる。また、印象ラベル「力強い」においては、発音検出に基づくリズム特徴量である IOIH 特徴量および PSPs 特徴量と比較すると、RLPC の方がそれぞれ約 8.0 ポイント、6.0 ポイント高い識別率が得られた。RLPC と従来のリズム特徴量の各印象における識別率において、Dunnett の方法を用いて有意水準 5% で片側仮説検定を行った結果、RLPC は「アップテンポな」と「軽快な」において PH 特徴量、「力強い」において IOIH 特徴量と有意差が確認できた。これら実験結果より、RLPC は従来のリズム特徴量と比較して印象識別に有効な特徴量であるといえる。

表 2 において、リズム特徴量の各印象に対する識別率を比較すると、「穏やかな」、「力強い」、「さわやかな」の 3 つの印象に対する識別率はその他の印象と比較すると低い識別率であることが分かる。この原因を調べるため、各印象において印象の有無*3 と各特徴量との相関係数を計算した。計算結果より、識別精度の低い 3 つの印象における相関係数 (穏やかな: 0.316, 力強い: 0.238, さわやかな: 0.243) はその他の印象における相関係数 (明るい: 0.381, 賑やかな: 0.403, アップテンポな: 0.417, 軽快な: 0.419) と比較して低いことが分かった。このため、これら 3 つの

*3 印象を受ける、もしくは受けないとアノテーションされることを示す。

表 2 実験結果 (各リズム特徴量の比較)

Table 2 A comparison among rhythm features.

| 特徴量 | 明るい | 賑やかな | アップテンポな | 軽快な | 穏やかな | 力強い | さわやかな | 平均 |
|-------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|--------------|
| RLPC | 87.0% (6.3%) | 86.0% (4.9%) | 89.4% (4.7%) | 91.3% (2.9%) | 78.8% (4.3%) | 71.3% (6.2%) | 74.8% (7.0%) | 83.7% |
| IOIH | 83.0% (3.9%) | 85.7% (2.3%) | 89.0% (1.1%) | 91.7% (4.0%) | 76.5% (5.1%) | 63.8% (3.6%) | 72.0% (4.7%) | 81.5% |
| PSPs | 84.2% (5.9%) | 86.4% (2.0%) | 89.0% (1.8%) | 92.6% (3.6%) | 77.6% (4.3%) | 65.4% (3.5%) | 72.0% (5.2%) | 82.3% |
| BH | 85.0% (6.4%) | 82.8% (4.6%) | 85.9% (2.1%) | 86.8% (3.2%) | 76.9% (3.3%) | 66.0% (6.0%) | 67.8% (11.2%) | 80.0% |
| ACFPs | 86.2% (4.0%) | 85.3% (3.5%) | 88.2% (2.4%) | 90.9% (1.2%) | 78.0% (3.7%) | 69.1% (4.0%) | 70.6% (5.3%) | 82.4% |
| PH | 83.0% (6.2%) | 84.2% (4.9%) | 76.5% (7.6%) | 84.3% (4.2%) | 74.9% (7.3%) | 72.9% (5.1%) | 68.5% (9.7%) | 78.6% |

表 3 実験結果 (すべてのリズム特徴量併用時)

Table 3 Identification results of all rhythm features.

| 特徴量 | 明るい | 賑やかな | アップテンポな | 軽快な | 穏やかな | 力強い | さわやかな | 平均 |
|------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------|
| All Rhythm | 89.9% (5.6%) | 89.2% (2.3%) | 92.9% (3.3%) | 92.6% (1.7%) | 83.1% (3.1%) | 75.5% (7.9%) | 78.3% (6.5%) | 86.9% |

表 4 実験結果 (Base 特徴量併用時における各リズム特徴量の比較)

Table 4 A comparison among rhythm features with base features.

| 特徴量 | 明るい | 賑やかな | アップテンポな | 軽快な | 穏やかな | 力強い | さわやかな | 平均 |
|--------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|------------------------|--------------|
| +RLPC | 89.5% (5.5%) | 94.6% (3.1%) | 94.1% (3.1%) | 95.0% (1.1%) | 84.7% (4.0%) | 82.4% (7.2%) | 79.7% (3.5%) | 89.5% |
| Base | 88.7% (5.9%) | 93.9% (2.4%) | 90.2% (2.4%) | 93.8% (1.5%) | 83.5% (5.4%) | 81.9% (11.0%) | 72.0% (7.3%) | 87.5% |
| +IOIH | 87.9% (7.4%) | 95.7% (2.4%) | 94.1% (4.6%) | 95.5% (1.7%) | 85.5% (5.3%) | 80.3% (8.5%) | 74.8% (2.1%) | 88.9% |
| +PSPs | 88.7% (6.7%) | 94.6% (1.7%) | 94.1% (2.8%) | 93.4% (3.2%) | 85.5% (6.0%) | 80.3% (9.5%) | 75.5% (3.2%) | 88.6% |
| +BH | 89.1% (5.4%) | 93.9% (1.6%) | 88.6% (2.6%) | 94.2% (1.7%) | 83.1% (4.3%) | 84.6% (8.8%) | 72.7% (1.8%) | 87.7% |
| +ACFPs | 89.1% (5.3%) | 94.3% (1.9%) | 93.3% (2.2%) | 93.4% (0.8%) | 83.5% (4.7%) | 81.9% (8.8%) | 69.2% (4.7%) | 87.8% |
| +PH | 87.4% (5.6%) | 93.9% (1.6%) | 90.2% (3.7%) | 91.3% (3.4%) | 83.9% (6.3%) | 82.4% (10.5%) | 69.2% (6.9%) | 86.8% |

印象に対する識別精度がその他の印象に対する識別精度よりも低くなったといえる。

さらに、すべてのリズム特徴量を併用した場合の実験結果を表 3 に示す。表 3 より、すべてのリズム特徴量を使用することで単独でリズム特徴量を使用するよりも高い識別率が得られた。また、これはすべてのリズム特徴量の組合せの中で最も高い識別率であった。

4.3.2 Base 特徴量とリズム特徴量併用時の比較

Base 特徴量を併用した場合の各リズム特徴量における印象ラベルに対する平均識別率を表 4 に示す。Base 特徴量と RLPC 併用時の識別率の比較により、RLPC の全印象ラベルにおける平均識別率は 89.5%、Base 特徴量の平均

識別率は 87.5%であり、RLPC が 2.0 ポイント高い平均識別率 (誤り削減率で 15.9%) であることが分かる。各印象ラベルの識別率を比較すると、すべてのラベルにおいて提案した RLPC を併用することにより、Base 特徴量の識別率が向上することが分かる。これは、Base 特徴量には含まれない音楽音響信号のリズム特徴を RLPC がとらえているためである。

次に、従来のリズム特徴量併用時と RLPC 併用時の識別率を比較する。表 4 より、各リズム特徴量併用時の全印象ラベルにおける平均識別率は、IOIH 特徴量において 88.9%、PSPs 特徴量において 88.6%、BH 特徴量において 87.7%、ACFPs 特徴量において 87.8%、PH 特徴量におい

表 5 実験結果 (Base 特徴量とすべてのリズム特徴量併用時および最適なリズム特徴量併用時)

Table 5 Identification results of all rhythm features and ideal combination of rhythm features.

| 特徴量 | 明るい | 賑やかな | アップテンポな | 軽快な | 穏やかな | 力強い | さわやかな | 平均 |
|--------------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-----------------|-------|
| +All Rhythm | 89.9% (4.6%) | 93.5% (2.7%) | 94.5% (2.6%) | 95.5% (1.7%) | 84.3% (2.8%) | 81.4% (8.5%) | 79.0% (3.2%) | 89.2% |
| +(RLPC+PSPs+ACFPs) | 91.1% (4.7%) | 95.3% (2.9%) | 94.5% (2.1%) | 94.2% (1.6%) | 85.9% (4.4%) | 85.1% (5.9%) | 79.0% (4.1%) | 90.2% |

て 86.8%であり, RLPC が最も高い平均識別率であることが分かる. また, 各印象ラベルの識別率を比較すると, 印象ラベル「賑やかな」, 「軽快な」, 「穏やかな」において IOIH 特徴量, 「アップテンポな」において PSP 特徴量, 「力強い」において BH 特徴量, 「明るい」, 「さわやかな」において RLPC が最も高い識別率である. しかし, RLPC 併用時以外の上記リズム特徴量併用時では, Base 特徴量単独使用時と比較して識別率が低下した印象ラベルが存在するが, RLPC 併用時ではすべての印象ラベルで識別率が向上している. また, RLPC 併用時と Base 特徴量および従来のリズム特徴量併用時の各印象における識別率において, Dunnett の方法を用いて有意水準 5%で片側仮説検定を行った結果, RLPC は「アップテンポな」において BH 特徴量, 「軽快な」において PH 特徴量, 「さわやかな」において Base 特徴量, ACFPs 特徴量, PH 特徴量と有意差が確認できた. これら実験結果より RLPC は従来のリズム特徴量よりも Base 特徴量に対して有効な補助特徴量であるといえる.

さらに, Base 特徴量に対してすべてのリズム特徴量を併用した場合, およびリズム特徴量の組合せの中で最も高い識別率であった実験結果を表 5 に示す. 表 5 より, すべてのリズム特徴量を併用した場合は RLPC を併用する場合と比較して平均識別率が 0.3 ポイント低下することが分かった. また, Base 特徴量に対して, RLPC, PSPs 特徴量, ACFPs 特徴量を併用した場合が最も識別率が高く, 平均識別率が 90.2%であった.

4.4 考察

本節では, 提案した RLPC が従来のリズム特徴量と比較して高い識別率であった理由について考察する. 前節の実験結果を詳細に分析した結果, 印象ラベル「アップテンポな」において BH 特徴量は印象を受けると判定された音楽に対して印象を受けないと識別する誤りが多い傾向があった. この原因を分析するため, BH 特徴量において誤識別し, RLPC において正識別したジャズ音楽データから抽出した BH を図 9, RLPC による各周波数帯域における対数スペクトル包絡を図 10 に示す. また, 学習データに用いた, 印象を受けると判定されたポピュラー音楽データから抽出した BH を図 11, RLPC による各周波数帯域における対数スペクトル包絡を図 12 に示す. 図 9, 図 11 に示

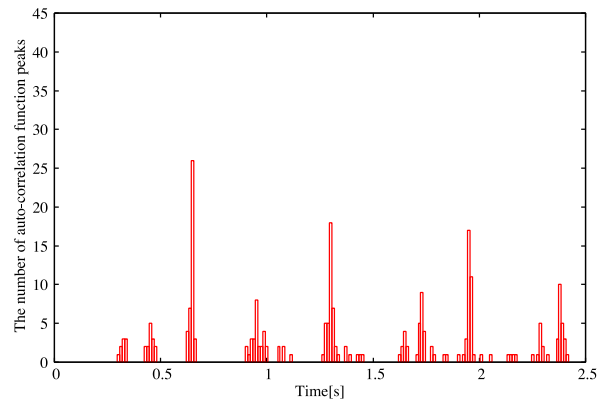


図 9 ジャズ音楽データから抽出された Beat Histogram

Fig. 9 Beat Histogram extracted from a jazz musical piece.

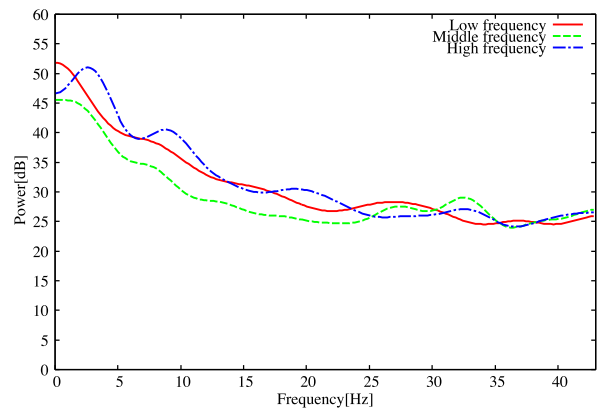


図 10 ジャズ音楽データにおける対数スペクトル包絡

Fig. 10 Logarithm spectral envelope of a jazz musical piece.

した BH は一定の時間間隔でピークが発生していることが分かる. しかし, 図 9 は図 11 と比較してヒストグラム値の総和が相対的に小さいことが分かる. このため, 図 9 に示した BH を持つジャズ音楽データは誤識別されている. 一方, 図 10 と図 12 に示した RLPC の対数スペクトル包絡では, ともに高周波数帯域の対数スペクトル包絡においてピークが発生していることが分かる. RLPC では, ピークが発生するスペクトル包絡を持つ音楽データは印象を受ける音楽データ, そうでなければ印象を受けない音楽データへ識別するため, ジャズ音楽データは正識別される. ACFPs 特徴量である全ウィンドウにおけるピーク数の平均値においても, RLPC と比較した場合に同様の傾向が見られた.

また, 表 2 により, RLPC は発音検出に基づくリズム特

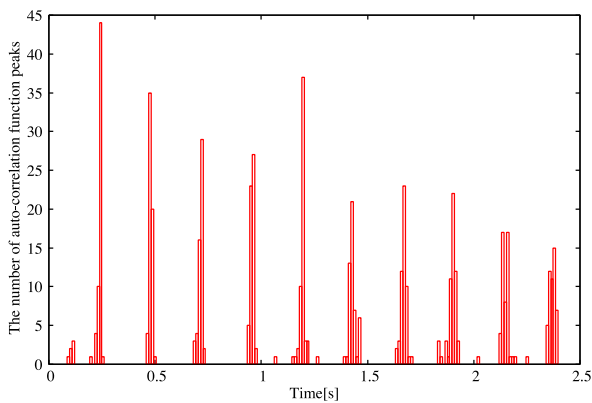


図 11 ポピュラー音楽データから抽出された Beat Histogram
Fig. 11 Beat Histogram extracted from a popular musical piece.

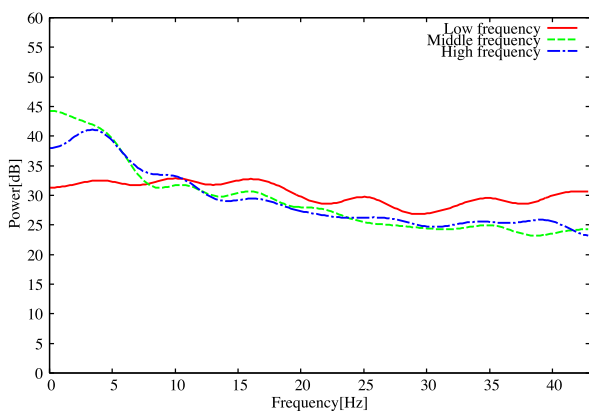


図 12 ポピュラー音楽データにおける対数スペクトル包絡
Fig. 12 Logarithm spectral envelope of a popular musical piece.

微量である IOIH 特徴量, PSPs 特徴量と比較して, 印象ラベル「アップテンポな」においては同程度の識別率, 「力強い」においては高い識別率を示すことが分かる. この結果を分析するため, 「アップテンポな」と「力強い」における, IOIH 特徴量と印象を受ける場合と受けない場合の関係を調査した. 図 13 と図 14 に「アップテンポな」と「力強い」における IOIH 特徴量を示す. これらの図より, 「アップテンポな」(図 13) では, 印象を受ける場合の特徴量と受けない場合の特徴量間に重なりが少ないが, 「力強い」(図 14) においては, 特徴量が重なっていることが分かる. このため, 識別器である SVM の識別境界が正しく学習できず識別性能が低くなったと考えられる. この傾向は, PSPs 特徴量における各周波数帯域の平均発音数においても同様であり, 識別性能が低い原因と考えられる.

さらに, 印象ラベル「力強い」において, RLPC を用いた場合に正識別し, IOIH 特徴量および PSPs 特徴量を用いた場合に誤識別した音楽データを調べた結果, ポピュラー音楽が多いことが分かった. これらのポピュラー音楽データを試聴により確認したところ, アクセントを持つ周期的なリズムで演奏されている傾向が強いことが分かっ

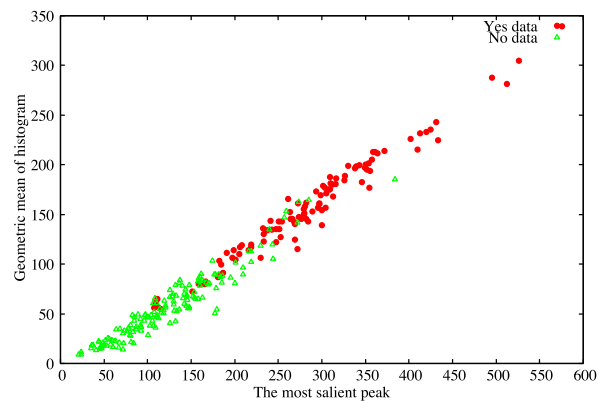


図 13 「アップテンポな」における IOIH 特徴量分布 (横軸: 1 番目に大きなピーク値, 縦軸: 幾何平均)
Fig. 13 Distribution of IOIH features on mood label “Up-tempo” (horizontal axis: the most salient peak, vertical axis: geometric mean of IOIH).

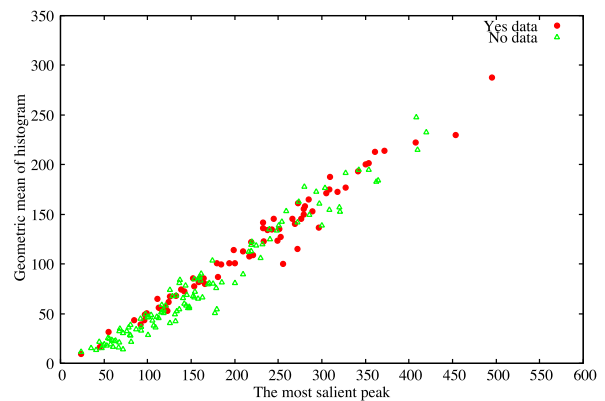


図 14 「力強い」における IOIH 特徴量分布 (横軸: 1 番目に大きなピーク値, 縦軸: 幾何平均)
Fig. 14 Distribution of IOIH features on mood label “Heavy” (horizontal axis: the most salient peak, vertical axis: geometric mean of IOIH).

た. このような周期的なリズムで演奏される音楽データは, アノテーションにおいて印象を受けるとラベル付けされる*4傾向が強い. 2.4 節および RLPC と BH 特徴量との比較においても述べたとおり, 周期的なリズムを持つ音楽データは対数スペクトル包絡においてピークが発生するため, RLPC を用いた場合には正しく識別される傾向がある. しかし, これら音楽データの中には周期的なリズムを持つが, 発音数が大きい音楽データと発音数が小さい音楽データの両方が存在することが試聴により分かった. 発音数に着目する特徴量 (IOIH 特徴量, PSPs 特徴量) では, リズムの周期性をとらえることができず, 「力強い」と印象ラベルが付与された音楽データにおいても正確な識別ができずに識別率が低下したと考えられる.

5. おわりに

本論文では音楽音響信号において, 周期的に演奏される

*4 本実験における正解のラベルとなる.

リズムを示す特徴量として、線形予測符号化に基づくリズム特徴量 (Rhythm feature based on Linear Predictive Coding) を提案した。RLPC では、従来のリズム特徴量では明確にとらえることができなかった周期的なリズム特徴を、LPC スペクトル包絡を用いることによりとらえることができる。

本論文では、RLPC の有効性を比較するため7つの印象ラベルに対して、(1) リズム特徴量による音楽印象識別実験、(2) リズム特徴量と Base 特徴量併用による音楽印象識別実験、をそれぞれ行った。リズム特徴量による印象識別実験結果より、RLPC における全印象ラベルの平均識別率は 83.7% であり、これは従来のリズム特徴量よりも高い平均識別率であった。また、リズム特徴量併用による印象識別実験結果より、RLPC 併用時の平均識別率は 89.5% であり、Base 特徴量のみの平均識別率である 87.5% より 2.0 ポイント高い平均識別率 (誤り削減率は 15.9%) が得られた。これは従来のリズム特徴量併用時の平均識別率よりも高い識別率であった。以上の実験結果より、提案する RLPC は印象識別において、単独使用時および補助特徴量としても有効なリズム特徴量であることを示した。

RLPC は線形予測符号化を適用する中間信号表現であるスペクトルエネルギー信号に依存する。文献 [16] において、リズム特徴量を正確に抽出するために、音源分離手法を用いて音響信号における打楽器音成分の強調を行っている。そこで、今後の課題としては、スペクトルエネルギー信号に対して音源分離手法を適用することで改善を行い、RLPC の有効性を向上させる予定である。また、RLPC を用いた類似音楽検索手法や自動ジャンル分類手法を提案する予定である。さらに、本論文で提案した印象識別精度に対するユーザ満足度を調査し、実用面から十分な精度であるか調査する予定である。

謝辞 本研究の一部は科学研究費基盤研究 B 21300060 の援助を受けて行った。

参考文献

- [1] Kim, Y.E., Schmidt, E.M., Migneco, R., Morton, B.G., Richardson, P., Scott, J., Speck, J.A. and Turnbull, D.: Music Emotion Recognition: A State Of The Art Review, *ISMIR2010*, pp.255-266 (2010).
- [2] Schmidt, E.M., Turnbull, D. and Kim, Y.E.: Feature Selection for Content-Based, Time-Varying Musical Emotion Regression, *MIR2010*, pp.267-273 (2010).
- [3] Trohidis, K., Tsoumakas, G., Kalliris, G. and Vlahavas, I.: Multi-Label Classification of Music Into Emotions, *ISMIR2008*, pp.325-330 (2008).
- [4] Mion, L. and Poli, G.D.: Score-Independent Audio Features for Description of Music Expression, *IEEE Trans. Audio, Speech, and Lang. Process.*, Vol.16, No.2, pp.458-466 (2008).
- [5] Skowronek, J., Mckinney, M. and van de Par, S.: A Demonstrator for Automatic Music Mood Estimation, *ISMIR2007* (2007).
- [6] Li, T. and Ogihara, M.: Content-Based Music Similarity Search and Emotion Detection, *ICASSP2004*, pp.V705-V708 (2004).
- [7] Li, T. and Ogihara, M.: Detecting Emotion in Music, *ISMIR2003*, pp.239-240 (2003).
- [8] Liu, D., Lu, L. and Zhang, H.-J.: Automatic Mood Detection from Acoustic Music Data, *ISMIR2003*, pp.81-87 (2003).
- [9] 平江 遼, 西 隆司: 感性に基づくクラシック音楽の分類, *日本音響学会誌*, Vol.64, No.10, pp.607-615 (2008).
- [10] Dixon, S., Pampalk, E. and Widmer, G.: Classification of Dance Music by Periodicity Patterns, *ISMIR2003*, pp.159-165 (2003).
- [11] Gouyon, F., Dixon, S., Pampalk, E. and Widmer, G.: Evaluating Rhythm Descriptors for Musical Genre Classification, *AES 25th Intl. Conf.*, pp.196-204 (2004).
- [12] Pampalk, E., Dixon, S. and Widmer, G.: Exploring Music Collections by Browsing Different Views, *ISMIR2003*, pp.201-208 (2003).
- [13] Rauber, A., Pampalk, E. and Merkl, D.: Using Psycho-Acoustic Models and Self-Organizing Maps to Create a Hierarchical Structuring of Music by Sound Similarity, *ISMIR2002*, pp.71-79 (2002).
- [14] Tsuchihashi, Y., Kitahara, T. and Katayose, H.: Using Bass-Line Features for Content-Based MIR, *ISMIR2008*, pp.620-625 (2008).
- [15] 角尾衣未留, Tzanetakis, G., 小野順貴, 嵯峨山茂樹: 音楽音響信号の低音旋律パターンのクラスタリングと自動ジャンル認識への応用, *情報処理学会 第 81 回音楽情報科学研究会* (2009).
- [16] Tsunoo, E., Akase, T., Ono, N. and Sagayama, S.: Music Mood Classification by Rhythm and Bass-line Unit Pattern Analysis, *ICASSP2010*, pp.265-268 (2010).
- [17] Cheng, H.-T., Yang, Y.-H., Lin, Y.-C., Liao, I.-B. and Chen, H.H.: Automatic Chord Recognition for Music Classification and Retrieval, *ICME2008*, pp.1505-1508 (2008).
- [18] Gouyon, F., Herrera, P. and Cano, P.: Pulse-Dependent Analysis of Percussive Music, *AES 22nd Intl. Conf.*, pp.396-401 (2002).
- [19] Tzanetakis, G. and Cook, P.: Musical Genre Classification of Audio Signals, *IEEE Trans. Speech and Audio Process.*, Vol.10, No.5, pp.293-302 (2002).
- [20] Bello, J.P., Daubet, L., Abdallah, S., Duxbury, C., Davies, M. and Sandler, M.B.: A Tutorial on Onset Detection in Music Signals, *IEEE Trans. Audio, Speech, and Lang. Process.*, Vol.13, No.5, pp.1035-1047 (2005).
- [21] Gainza, M. and Coyle, E.: Tempo Detection Using a Hybrid Multiband Approach, *IEEE Trans. Audio, Speech, and Lang. Process.*, Vol.19, No.1, pp.57-68 (2010).
- [22] Rabiner, L. and Juang, B.-H. (著), 古井貞熙 (監訳): 音声認識の基礎上・下, NTT アドバンステクノロジー (1995).
- [23] 日野幹雄: スペクトル解析, 朝倉書店 (1977).
- [24] Miyoshi, M., Tsuge, S., Oyama, T., Ito, M. and Fukumi, M.: Feature Selection Method for Music Mood Score Detection, *ICMSAO2011*, pp.713-718 (2011).
- [25] 後藤真孝, 橋口博樹, 西村拓一, 岡 隆一: RWC 研究用音楽データベース: 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, *情報処理学会論文誌*, Vol.45, No.3, pp.728-738 (2004).
- [26] 杉原太郎, 森本一成, 黒川隆夫: SD 法を通してみた音楽に対する感性の基本特性, *電子情報通信学会信学技報*, pp.57-63 (2001).
- [27] Hsu, C.-W., Chang, C.-C. and Lin, C.-J.: A Practical Guide to Support Vector Classification (online), avail-

able from <http://www.csie.ntu.edu.tw/~cjlin/libsvm>
(2010).



三好 真人

2010年徳島大学工学部知能情報工学科卒業。2012年同大学大学院先端技術科学教育部システム創生工学専攻博士前期課程修了。音楽印象識別の研究に従事。



柘植 覚 (正会員)

2001年3月徳島大学大学院博士後期課程修了。1997年ATR音声翻訳通信研究所研修研究員，2000年徳島大学工学部助手，2006年徳島大学大学院ソシオテクノサイエンス研究部講師，2010年大同大学情報学部准教授。博士(工学)。音声認識，話者認識，情報検索の研究に従事。日本音響学会，電気学会各会員。



福見 稔 (正会員)

1987年3月徳島大学大学院工学研究科修士課程修了。同年同大学助手。1996年徳島大学助教授，現在，同大学院ソシオテクノサイエンス研究部教授。博士(工学)。ヒューマンセンシング，生物型情報処理，デジタル信号処理等の研究に従事。計測自動制御学会・信号処理学会の論文賞等。電気学会，計測自動制御学会，電子情報通信学会，IEEE，日本顔学会等の各会員。