Vol.2017-MUS-114 No.8
2017/2/27

大域・局所リズムパタンテンプレートマッチングに基づく楽曲の伴奏スタイル識別

篠井 暖[1,a)]　前澤 陽[1]

概要：本稿では，楽曲が持つ伴奏の特性（伴奏スタイル）を分析するための手法について述べる．楽曲の伴奏スタイルを決定づける要素としてリズムパタンが重要と考え，リズムパタンに基づく伴奏スタイル識別問題に取り組む．ここで，リズムパタンの特徴量による識別を行う際に，(1) 楽曲の音響信号からリズムパタンの特徴量を教師情報なしで精度よく抽出する手法が確立されていない (2) 伴奏スタイルのアノテーションが付与されたデータセットが存在しない という問題がある．そこで，本稿では伴奏スタイル種別のアノテーションを付与した伴奏パタンのテンプレートデータベースを事前に構築しておき，(1) 楽器音およびリズムパタンの教師情報利用によるリズムパタン特徴抽出の精度向上 と (2) 楽曲をクエリとした伴奏スタイルテンプレートの類似検索問題としての定式化 による伴奏スタイル識別手法を提案する．

Music Style Identification with Template Matching Using Global and Local Rhythmic Pattern Features

Dan Sasai[1,a)]　Akira Maezawa[1]

1. はじめに

音楽情報処理による音楽の理解を実現する上で，楽曲の伴奏が持つ特性を把握することは重要である．なぜならば，90 年代のロック楽曲にはギターやドラムといった伴奏部分に特有の演奏パタンが存在し，ジャズ楽曲の伴奏にはまた別の演奏パタンが存在するといったように，音楽の印象は伴奏によって大きく特徴づけられるためである．また楽曲の伴奏が持つ特性は楽曲推薦やジャンル識別などの有用な特徴量となると考えられるため，これらの応用における基礎技術となる．そこで本稿では「90 年代ロック」や「70 年代ディスコ・ミュージック」といったそれぞれの音楽カテゴリが特有に持つ伴奏パタンを「伴奏スタイル」と定義し，楽曲の音響信号から伴奏スタイルを識別する問題を取り扱う．

伴奏スタイルの識別を行うためには，音楽がもつ伴奏スタイルを決定づける要素を考える必要がある．ここでは以下 3 つの要素が重要と考える．

楽器編成　音楽のジャンルごとに使用される楽器は変わってくる．たとえば，ロック楽曲であればギター，ベース，ドラムという編成が一般的であるし，ジャズ楽曲であればピアノ，ベース，ドラム，金管楽器といった編成になる．このように，ジャンルごとの音楽スタイルを特徴づける要素として楽器編成は重要な要素として考えられる．

リズムパタン　同一ジャンル内での細かな音楽スタイルの違いを特徴づけるのがリズムパタンと考えられる．たとえば，ロックで考えると 8 ビートを刻んでいるストレートなロックなのか，変則的なシャッフルリズムを刻んでいるのか，またはツーバスを多用しているヘヴィ・メタルなのかといったように，リズムパタンにより細かな音楽スタイルの違いが生み出されていると捉えられる．細かな音楽スタイルの違いを特徴づける要素としてリズムパタンは特に重要である．

テンポ（BPM）　ジャンルごとに用いられる典型的なテンポ（BPM:Beats Per Minute）値が存在する．たと

[1] ヤマハ株式会社
Yamaha Corporation, Iwata, Shizuoka 438-0192, Japan
[a)] dan.sasai@music.yamaha.com

えばポップスであれば BPM120 前後，ダンスであれば BPM140 前後などである．ジャンルごとの違いを特徴づける要素として BPM は有効と考えられる．

これら 3 つの要素に着目した伴奏スタイル識別手法の確立を本研究の目的とする．そのためには，(1) 楽曲から上記 3 要素に対応した特徴量の抽出，(2) 抽出した特徴量を用いての伴奏スタイル識別 の 2 つの課題を解決する必要がある．まず（1）に関して，楽器編成とリズムパタンの特徴抽出には，非負値行列因子分解（NMF）[8] に代表される楽器音分離手法により音響信号を楽器音基底とアクティベーションに分解することで楽器編成とリズムパタンを同時に推定するアプローチが考えられる．しかし，楽曲のような混合音に対して教師情報なしで精度よく楽器音の分離を行う手法が確立されていないという問題がある．また，（2）に関して，伴奏スタイル識別はジャンル識別と問題の枠組み自体は似ているが，ジャンル識別が対象とする「ロック」「ジャズ」といった分類よりも細分化された粒度の識別を必要とするため，既存のジャンル識別用に構築されたデータセットによる学習では対応できないという問題がある．

そこで，本稿では，伴奏スタイル種別のラベルを付与した伴奏パタンのテンプレートのデータベースを構築し，(1) 楽器音およびリズムパタンの教師情報利用によるリズムパタン特徴抽出の精度向上 および（2）楽曲をクエリとした伴奏パタンテンプレートの類似検索問題としての定式化 により伴奏スタイルの識別を実現する手法を提案する．

2. 関連研究

ここでは関連分野の研究としてジャンル識別の既存研究を挙げ，本研究が対象とする伴奏スタイル識別問題へ適用する際の問題点を説明する．

音楽音響信号のジャンル識別は音響特徴抽出と教師あり学習によるものが主流である [3]．予めジャンルの正解ラベル付きの楽曲データベースを用意しておき，学習用の楽曲から音響特徴量を抽出し，それらの特徴量から識別器を学習し識別を行う．特に MFCC（Mel Frequency Cepstral Coefficients）やビート特徴量などの音響特徴量を用いて SVM（Support Vector Machine）で識別を行うアプローチが広く行われている [4]．近年では特徴抽出に非負値行列因子分解を用いた手法 [5][6] や特徴抽出に Deep Neural Network を用いる手法 [7] も存在する．

これらの手法を伴奏スタイル識別に適用する際の問題点としては（1）ジャンルよりも細かい粒度を持つ伴奏スタイルの識別を行うための学習データセットが存在しない（2）特徴量がリズムパタンを詳細に捉えるものになっていない点が挙げられる．

（1）に関して，本研究で対象とする粒度の伴奏スタイルのアノテーションが付与された既存のデータセットは存在しないので，リズムパタンのテンプレート音響信号と伴奏スタイルのアノテーションが組になったデータベースを新たに構築する必要がある．

（2）に関して，従来のジャンル識別手法でも MFCC の時間変化やビート特徴量などのリズムパタンに対応する特徴量は用いられていたが，これらの特徴量では先に述べたような 8 ビートとシャッフルリズムといった細かなリズムの違いを判別する能力はないと考えられ，より細かな差異を判別できる特徴量が必要になる．

3. 伴奏スタイル識別手法

本研究で提案する伴奏スタイル識別手法の全体構成を図 1 に示す．

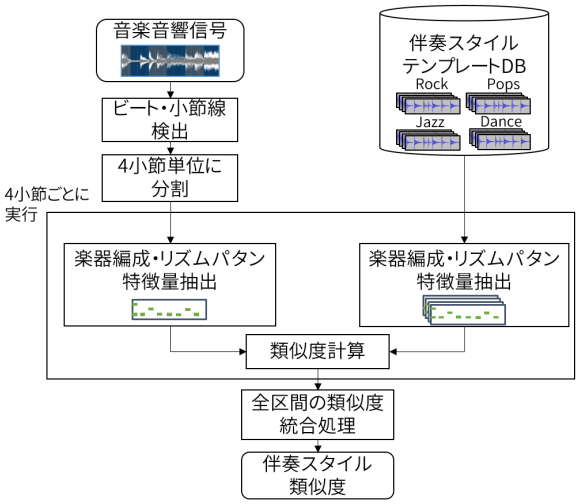

図 1 伴奏スタイル識別手法の全体構成

事前に伴奏スタイルのテンプレートデータベースを構築しておき，楽曲の音響特徴量と伴奏スタイルの音響特徴量とのテンプレートマッチングにより両者の類似度を算出する．ここで楽曲の時間長と伴奏スタイルの時間長は異なるが，シンプルなテンプレートマッチングによる類似度計算を可能にするために楽曲の音響信号を伴奏スタイルの時間長単位で分割する．今回用いている伴奏スタイルのテンプレート（詳細は後述）は 4 小節分の時間長を持っているので，楽曲の音響信号に対して事前にビート・小節線検出を行っておき，4 小節単位での分割を行う．続いて分割した楽曲の音響信号と伴奏スタイルそれぞれに対し楽器編成，リズムパタン，BPM の特徴量を抽出し，それらの特徴量の類似度を総当りで計算する．この処理を 4 小節ごとに行い，最後に 1 曲全体にわたって類似度を統合する処理を行う．

以下に提案法を構成する各部の詳細を述べる．

3.1 伴奏スタイルテンプレートデータベース

伴奏スタイルのテンプレートデータベースとして，ヤマハ株式会社の電子鍵盤楽器製品が持つ「スタイル機能」[1]

と呼ばれる自動伴奏機能で用いられている伴奏パタンのデータを用いる．これは各ジャンルにおける伴奏部分（ドラム、ベース，ギター，キーボードなど）の典型的な演奏パタン（スタイル）の MIDI データを内蔵しておき，鍵盤楽器奏者がテンポとコード進行を指定することで伴奏部分のテンポとコード進行を制御し，奏者の実現したい演奏を実現する自動伴奏機能である．各スタイルには「60's VintageRock」「80's PowerRock」「EuroTrance」といった名前が付与されており，この名前が伴奏スタイルの種別として利用できると考え，下記の情報をもつデータベースを「伴奏スタイルテンプレートデータベース」として構築した．

(1) 打楽器パートの音響信号
(2) 非打楽器パートの音響信号
(3) 伴奏スタイルの名前
(4) 伴奏スタイルの所属ジャンル
(5) 伴奏スタイルの標準 BPM

（1）と（2）の音響信号はスタイル機能を搭載する電子鍵盤楽器製品である Tyros4[2] を用いコード：C Major，テンポ：標準 BPM の条件で打楽器パートと打楽器以外のパートをそれぞれ 4 小節分再生したスタイルを録音することで作成した．Tyros4 には各ジャンルごとの伴奏スタイルが計 2000 種類存在し，そのうち打楽器パートが存在する 1892 種類の伴奏スタイルからなるデータベースを作成した．ジャンルごとの伴奏スタイル数の内訳を表 1 に示す．

表 1 ジャンルごとの伴奏スタイル数

ジャンル	伴奏スタイル数
Pop	264
Rock	128
Ballad	204
Dance / Electronic	92
R&B/Soul	196
Hip Hop / Rap	20
Country	120
Jazz	156
Latin	220
Gospel & Worship	64
Easy Listening	96
Traditional & Folk	200
Soundtrack	92
Holidays & Events	24
Blues	16

3.2 楽器編成・リズムパタン特徴抽出の概要

音響信号から局所的なリズムの違いも捉えられる特徴量を抽出するために，（1）楽器ごとの大まかなリズムパタンの変動を捉える特徴量（大域リズムパタン特徴量）と（2）局所的なリズムの差異を捉える特徴量（局所リズムパタン特徴量）の 2 通りの特徴量を抽出する．（1）の特徴量は各楽器が鳴っているタイミングを大まかに捉えることを目的とし，（2）の特徴量は規則正しくリズムを刻んでいるのか，それともシャッフルリズムのようにビート位置から微妙に逸脱したリズムを刻んでいるのか，といった違いを捉えることを目的とする．以下でそれぞれの詳細を説明する．

3.3 大域リズムパタン特徴量: 教師あり GaP-NMF アクティベーション

大域リズムパタン特徴量の抽出では楽器ごとの発音タイミング変動を捉えることが目的となる．そこで，本手法では伴奏スタイルテンプレートを教師データとした教師あり非負値行列因子分解（NMF）による大域リズムパタン特徴量の抽出を行う．具体的には，伴奏スタイルテンプレート音響信号からジャンルごとの楽器音基底を学習しておき，それらの基底を教師基底として固定した上で楽曲のアクティベーションを教師あり NMF により推定する．

3.3.1 GaP-NMF による伴奏スタイルテンプレートの基底学習

基底学習フェーズにおいては、ガンマ過程 NMF（GaP-NMF）[9] により伴奏スタイルテンプレートの音響信号を少数の基底スペクトルとアクティベーションに分解する．本手法では，音響信号はまず定 Q 変換（CQT）により CQT スペクトログラムに変換され，さらにビート情報を利用して時間軸をビート単位に変換した CQT スペクトログラム $\mathbf{X}$ を得る．そして $\mathbf{X}$ を基底行列 $\mathbf{W} \in \mathbb{R}^{F\times K}$ とアクティベーション行列 $\mathbf{H} \in \mathbb{R}^{K\times T}$，および非負値ベクトル $\boldsymbol{\theta} \in \mathbb{R}^{K}$ に分解する．ここで，F は X の周波数ビン数，T は X のビート数，K は GaP-NMF の基底数である．以下に GaP-NMF のモデルを示す：

$$X_{ft} \approx \sum_{k=1}^{K} \theta_k W_{fk} H_{kt} \tag{1}$$

ここで，θ_k は k 番目の基底の重み，W_{fk} は k 番目の基底の周波数 f におけるパワー，H_{kt} は k 番目の基底の時刻 t のアクティベーションを示す．$\mathbf{W}$ の各列は k 番目の基底スペクトルを，$\mathbf{H}$ の各行は k 番目の基底の時間変化パターンを示している．

上記の GaP-NMF による基底スペクトルの学習を伴奏スタイルのジャンルごとに行い，さらに打楽器パート p_d の音響信号と非打楽器パート p_n の音響信号それぞれに対し NMF を実行し基底学習を行う．基底学習の全体像を図 2 に示す．まず，伴奏スタイルテンプレートデータベースよりジャンル $g \in \{g_1, ..., g_G\}$ に属し，スタイル番号 $s \in \{s_1, ..., s_N\}$ をもつ伴奏スタイル $S_s^{(g)}$ の集合 $S^{(g)} = \{S_{s_1}^{(g)}, ..., S_{s_N}^{(g)}\}$ を抽出する．ここで，G はジャンルの総数，N はジャンル g に属する伴奏スタイルの数，s_i はジャンル g に属する伴奏スタイルのスタイル番号を示す．次に，$S_s^{(g)}$ のパート

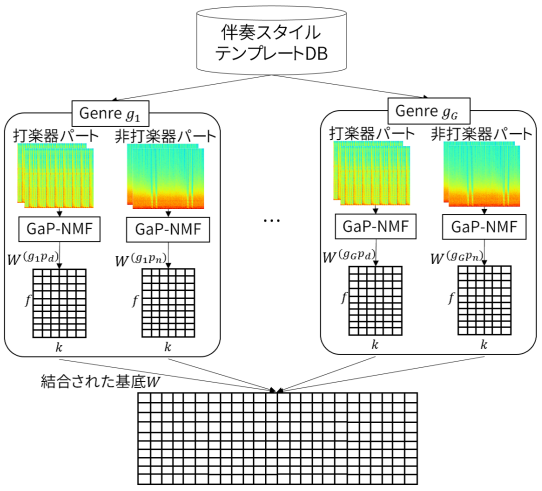

図 2 GaP-NMF による伴奏スタイルテンプレートの基底学習

$p \in \{p_d, p_n\}$ の CQT スペクトログラムを $\mathbf{X}_s^{(gp)}$ とし，ジャンル g に属する全ての $\mathbf{X}_s^{(gp)}$ を時間方向に結合した 1 つの巨大な CQT スペクトログラム $\mathbf{X}^{(gp)}$ を作成する．

$$\mathbf{X}^{(gp)} = \left[\begin{array}{cccc} \mathbf{X}_{s_1}^{(gp)} & \mathbf{X}_{s_2}^{(gp)} & \ldots & \mathbf{X}_{s_N}^{(gp)} \end{array} \right] \tag{2}$$

この $\mathbf{X}^{(gp)}$ に対して GaP-NMF を実行する．

$$X_{ft}^{(gp)} \approx \sum_{k=1}^{K} \theta_k^{(gp)} W_{fk}^{(gp)} H_{kt}^{(gp)} \tag{3}$$

ジャンル g の基底 $W^{(g)}$ は打楽器パートの基底 $W^{(gp_d)}$ と非打楽器パートの基底 $W^{(gp_n)}$ を基底次元方向に結合することで得られる．

$$W_{fk}^{(g)} = \left[\begin{array}{cc} W_{fk}^{(gp_d)} & W_{fk}^{(gp_n)} \end{array} \right] \tag{4}$$

さらに，ジャンル $g = \{g_1, g_2, ..., g_G\}$ それぞれの基底 $W_{fk}^{(g)}$ を基底次元方向に結合することで伴奏スタイルテンプレート全体の基底 W_{fk} が得られる．

$$W_{fk} = \left[\begin{array}{cccc} W_{fk}^{(g_1)} & W_{fk}^{(g_2)} & \ldots & W_{fk}^{(g_G)} \end{array} \right] \tag{5}$$

3.3.2 教師あり GaP-NMF による楽曲のアクティベーション推定

推定フェーズにおいては，まず楽曲のビート・小節線検出を実行した後に，伴奏スタイルと同様の特徴量を抽出する．すなわち，時間軸をビートに変換した CQT スペクトログラム X_{ft} を得てから，X_{ft} と基底学習フェーズにおいて得られた伴奏スタイル基底 W_{fk} を用いて教師あり NMF の枠組みにより NMF アクティベーション H_{kt} を推定する．これを大域リズムパタン特徴量として用いる．

3.3.3 GaP-NMF の推論アルゴリズム

GaP-NMF の推論は Hoffman ら [9] と同様に変分ベイズ法による推論を行う．以下に推論アルゴリズムの概要を示す．詳細は Hoffman らの論文を参照されたい．まず，$\mathbf{W} \in \mathbb{R}^{F \times K}$, $\mathbf{H} \in \mathbb{R}^{K \times T}$, $\boldsymbol{\theta} \in \mathbb{R}^{K}$ の値がそれぞれある生成過程に従い確率的に生成されたと仮定する．ここでは，それぞれ以下の Gamma 分布を事前分布とした．

$$W_{fk} \approx Gamma(a^{(W)}, b^{(W)}) \tag{6}$$
$$H_{kt} \approx Gamma(a^{(H)}, b^{(H)}) \tag{7}$$
$$\theta_k \approx Gamma(\frac{\alpha}{K}, \alpha c) \tag{8}$$

K は観測スペクトログラムの楽器音の数より十分に大きい数であり，c は $\mathbf{X}$ の平均値の逆数である．つまり，$c = \frac{1}{FT} \sum_f \sum_t X_{ft}$ となる．

また，変分事後分布には一般化逆ガウス分布（GIG）を仮定する．

$$q(W_{fk}) = GIG(\gamma_{fk}^{(W)}, \rho_{fk}^{(W)}, \tau_{fk}^{(W)}) \tag{9}$$
$$q(H_{kt}) = GIG(\gamma_{kt}^{(H)}, \rho_{kt}^{(H)}, \tau_{kt}^{(H)}) \tag{10}$$
$$q(\theta_k) = GIG(\gamma_{k}^{(\theta)}, \rho_{k}^{(\theta)}, \tau_{k}^{(\theta)}) \tag{11}$$

GIG 分布は Gamma 事前分布を一般化した確率分布になっており，モデルの表現力を向上させるとともに解析的な更新式を導出することが可能である．

GIG 分布のパラメータ推論のために，まず変分下限の更新を行う．具体的には，以下の式によりパラメータ ϕ と ω の更新を行う．

$$\phi_{fkt} \propto \mathbb{E}_q \left[\frac{1}{\theta_k W_{fk} H_{kt}} \right]^{-1} \tag{12}$$
$$\omega_{ft} = \sum_k \mathbb{E}_q \left[\theta_k W_{fk} H_{kt} \right] \tag{13}$$

次に，更新された ϕ と ω を用いて変分事後分布のパラメータを更新する．

$$\gamma_{fk}^{(W)} = a^{(W)} \tag{14}$$
$$\rho_{fk}^{(W)} = b^{(W)} + \mathbb{E}_q[\theta_k] \sum_t \frac{\mathbb{E}_q[H_{kt}]}{\omega_{ft}} \tag{15}$$
$$\tau_{fk}^{(W)} = \mathbb{E}_q \left[\frac{1}{\theta_k} \right] \sum_t X_{ft} \phi_{fkt}^2 \mathbb{E}_q \left[\frac{1}{H_{kt}} \right] \tag{16}$$

$$\gamma_{kt}^{(H)} = a^{(H)} \tag{17}$$
$$\rho_{kt}^{(H)} = b^{(H)} + \mathbb{E}_q[\theta_k] \sum_f \frac{\mathbb{E}_q[W_{fk}]}{\omega_{ft}} \tag{18}$$
$$\tau_{kt}^{(H)} = \mathbb{E}_q \left[\frac{1}{\theta_k} \right] \sum_f X_{ft} \phi_{fkt}^2 \mathbb{E}_q \left[\frac{1}{W_{ft}} \right] \tag{19}$$

$$\gamma_{k}^{(\theta)} = \frac{\alpha}{K} \tag{20}$$
$$\rho_{k}^{(\theta)} = \alpha c + \sum_f \sum_t \frac{\mathbb{E}_q[W_{fk} H_{kt}]}{\omega_{ft}} \tag{21}$$
$$\tau_{k}^{(\theta)} = \sum_f \sum_t X_{ft} \phi_{fkt}^2 \mathbb{E}_q \left[\frac{1}{W_{fk} H_{kt}} \right] \tag{22}$$

3.4 局所リズムパタン特徴量: ビートスペクトル

NMF による大域リズムパタン特徴量に加えて，リズム構造の細かな違いを捉えることが可能な特徴量を導入する．伴奏スタイル識別のためには，(1) ビート位置で規則正しく刻んでいるリズム と (2) ビート位置からの微妙なずれが存在するリズム（例：シャッフルリズム）のようなリズム構造の違いを判別できることが重要となる．これらの違いはオンセット特徴量が周期的なピークを持っているのか，それとも周期から外れたところにピークが立っているのか，を見ることで判別可能と考えられる．本手法ではオンセット特徴量の周期性を捉える特徴量としてビートスペクトル [10][11] を用いる．

ビートスペクトルは対数スペクトルや MFCC といったスペクトル特徴量の周期性に基づくテンポ特徴量として定義され，いくつかの計算方法が提案されている [10][11]．ここでは Kurth らの手法 [11] を採用する．以下にビートスペクトルの計算手順を示す．

まず入力のスペクトログラム $\mathbf{X}$ に対し，スペクトルの時間差分 $N[x](t)$ を計算する．

$$N[x](t) = \sum_{f=0}^{F/2-1} max(|X(t+1,f)| - |X(t,f)|, 0) \quad (23)$$

続いて，以下のコムフィルタを適用する．

$$y_p(t) = (1-\alpha)N[x](t) + \alpha y_p(t-p) \quad (24)$$

ここで，p は共振周期を示すパラメータで，サンプリング周波数を固定した場合に BPM と対応している値であることが知られている．α は共振因子と呼ばれる定数で，$\alpha = 0.5$ が通常用いられる．

次に，以下の平滑化を行いビートスペクトル $\mathbf{B}$ が得られる．

$$B(t,p) = \sum_{\tau=-r}^{r} |y_p(t+\tau)|^2 \quad (25)$$

時刻 t から $2r+1$ サンプル分の $y_p(t)$ の値を加算する式になっており，$r = 2300$ が通常用いられる．これは時刻 t を中心とし，20 秒分の区間において $y_p(t)$ と時刻 $y_p(t+\tau)$ の共振度合いを評価していることに相当する．

3.4.1 BPM によるビートスペクトル正規化

本項ではビートスペクトルをリズムパタン特徴量として利用するための正規化処理について説明する．

ビートスペクトルは通常 BPM の分析などに用いられる．ビートスペクトルの例を図 3 に示す．

図 3 はジャンル：ロックに属し BPM の異なる 2 つの伴奏スタイル（Hard Rock と 70's Rock）のビートスペクトルを重ねて表示している．両者はリズムパタンとしては似た 8 ビートのリズムを刻んでいるが，BPM が異なるためにピーク間隔が異なっている．ビートスペクトルをリズムパタン特徴量として用いる際は，BPM が多少異なっていてもリズムパタンが似ているものは似たものとして扱いたいので，共振周期 p から BPM の影響を排除するために，時間軸を BPM により正規化する．具体的には，BPM 情報を利用してビートスペクトルの計算に入力するスペクトログラム X の時間軸をフレーム単位からビート単位に変換しておくことで，共振周期 p をビート単位に変換する．これにより BPM に依存しないビートスペクトルが得られる．BPM による正規化を行ったビートスペクトルを図 4 に示す．両者でピークの位置が一致し，単純な比較でリズム構造の差異を評価できるようになったことがわかる．

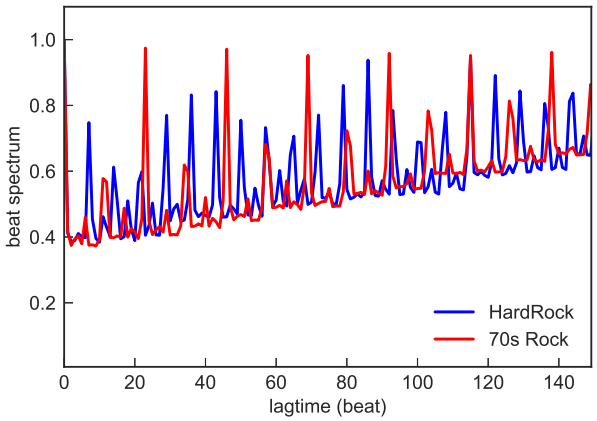

図 3 BPM 正規化前のビートスペクトル

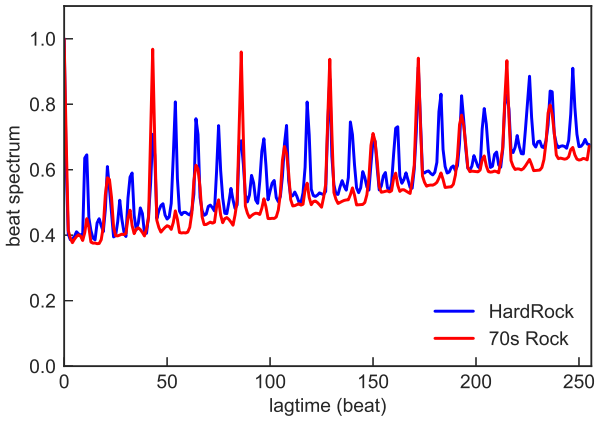

図 4 BPM 正規化後のビートスペクトル

3.5 リズムパタン特徴のマッチング

本節では楽曲と伴奏スタイルの類似度計算方法について説明する．大域リズムパタン特徴量（NMF アクティベーション）と局所リズムパタン特徴量（ビートスペクトル）それぞれに対し類似度を計算し，それらを重み付けした上で加算する．さらに楽曲と BPM が大きく離れている伴奏スタイルは類似度を下げる目的で BPM に基づく罰則項を設ける．類似度の計算式を以下に示す．

$$Sim(\mathbf{x}_i, \mathbf{x}_j) = w_t Sim_t(\mathbf{H}_i, \mathbf{H}_j) + w_r Sim_r(\mathbf{H}_i, \mathbf{H}_j) + w_b Sim_b(\mathbf{B}_i, \mathbf{B}_j) - P(b_i, b_j) \quad (26)$$

ここで，$\mathbf{H}_i$ は音響信号 $\mathbf{x}_i$ の NMF アクティベーション特徴量，$\mathbf{B}_i$ は $\mathbf{x}_i$ のビートスペクトル特徴量，b_i は $\mathbf{x}_i$ の BPM を示す．$Sim_t(\mathbf{H}_i, \mathbf{H}_j)$ は NMF アクティベーション特徴量の音色に関する類似度，$Sim_r(\mathbf{H}_i, \mathbf{H}_j)$ は NMF アクティベーション特徴量のリズムに関する類似度，$Sim_b(\mathbf{B}_i, \mathbf{B}_j)$ はビートスペクトルの類似度を示す．$P(b_i, b_j)$ は BPM による罰則項を示す．以下にそれぞれの類似度および罰則項の詳細を説明する．

3.5.1 NMF アクティベーションの類似度

NMF アクティベーションの類似尺度には以下で定義される相関係数を用いる．

$$c(\mathbf{x}, \mathbf{y}) = \frac{\sum_{i=1}^{n}(x_i - \overline{\mathbf{x}})(y_i - \overline{\mathbf{y}})}{\sqrt{\sum_{i=1}^{n}(x_i - \overline{\mathbf{x}})^2}\sqrt{\sum_{i=1}^{n}(y_i - \overline{\mathbf{y}})^2}} \quad (27)$$

ここで，NMF アクティベーションの類似度を測る際には（1）楽器編成（=音色）の類似度 と（2）リズムパタンの類似度 の両方を評価したい．楽器編成の類似度を測るには，NMF アクティベーションの基底次元方向を 1 つのベクトルと見て相関係数の計算を行い，時間次元方向に平均すればよい．

$$Sim_t(\mathbf{H}_i, \mathbf{H}_j) = \frac{1}{T}\sum_{t=1}^{T} c(\mathbf{H}_i(t, 1:K), \mathbf{H}_j(t, 1:K)) \quad (28)$$

リズムパタンの類似度を測るには，時間次元方向を 1 つのベクトルと見て相関係数の計算を行い，基底次元方向に平均すればよい．

$$Sim_r(\mathbf{H}_i, \mathbf{H}_j) = \frac{1}{K}\sum_{k=1}^{K} c(\mathbf{H}_i(1:T, k), \mathbf{H}_j(1:T, k)) \quad (29)$$

3.5.2 ビートスペクトルの類似度

ビートスペクトルの類似尺度にも NMF アクティベーションと同様に相関係数を用いる．

$$Sim_b(\mathbf{B}_i, \mathbf{B}_j) = \frac{1}{T}\sum_{t=1}^{T} c(\mathbf{B}_i(t), \mathbf{B}_j(t)) \quad (30)$$

3.5.3 各類似度の重み付け

$Sim_t(\mathbf{H}_i, \mathbf{H}_j)$，$Sim_r(\mathbf{H}_i, \mathbf{H}_j)$，$Sim_b(\mathbf{B}_i, \mathbf{B}_j)$ のスケールを合わせるための重み付けを行う．この 3 種類の類似尺度は全て相関係数を使用しているが，入力の取りうる値によりスケールが変わってくるので「それぞれの類似尺度が現実的に取りうる値の平均」の逆数を重みとすることでスケールを一致させる．具体的には，重み w_t，w_r，w_b は伴奏スタイルテンプレートデータベースを用いて，伴奏スタイル間で総当りの類似度計算を行い，その平均の逆数をとることで計算される．さらに，音色類似度の重み w_t とリズム類似度の重み w_r，w_b の寄与率が同等となるように，w_t の値は 2 倍しておく．

$$w_t = 2\left(\frac{1}{D^2}\sum_{i=1}^{D}\sum_{j=1}^{D} Sim_t(\mathbf{H}_i, \mathbf{H}_j)\right)^{-1} \quad (31)$$

$$w_r = \left(\frac{1}{D^2}\sum_{i=1}^{D}\sum_{j=1}^{D} Sim_r(\mathbf{H}_i, \mathbf{H}_j)\right)^{-1} \quad (32)$$

$$w_b = \left(\frac{1}{D^2}\sum_{i=1}^{D}\sum_{j=1}^{D} Sim_b(\mathbf{B}_i, \mathbf{B}_j)\right)^{-1} \quad (33)$$

$$(34)$$

ここで，D はデータベースに存在する伴奏スタイルの数である．最後に，$w_t + w_r + w_b = 1$ となるように正規化を行う．

3.5.4 BPM による罰則項

BPM による罰則項 $P(b_i, b_j)$ を以下の式で定義する．

$$P(b_i, b_j) = \begin{cases} 0 & (1-\beta < \frac{b_j}{b_i} < 1+\beta) \\ \gamma & (otherwise) \end{cases} \quad (35)$$

ここで，β は罰則を課す BPM の範囲を決定するためのパラメータ，γ は罰則の値を決定するためのパラメータである．楽曲と伴奏スタイルの BPM の比が $1-\beta$ から $1+\beta$ の範囲内であれば罰則を課さず，この範囲から外れた時に罰則を課す．パラメータの値は本手法では $\beta = 0.25$，$\gamma = 0.5$ を用いた．

3.6 楽曲全体と伴奏スタイルの類似度計算

楽曲全体と伴奏スタイルの類似度は以下の式で計算される．

$$Sim_{all}(\mathbf{x}_i, \mathbf{x}_j) = \frac{1}{M}\sum_{m=1}^{M} Sim(\mathbf{x}_i^{(m)}, \mathbf{x}_j^{(m)}) \quad (36)$$

M は楽曲 $\mathbf{x}_i$ を 4 小節単位に分割した際の区間の総数，m はその区間のインデックスを示す．$\mathbf{x}_i^{(m)}$ は $\mathbf{x}_i$ の区間 m の音響信号を示す．4 小節ごとに楽曲と伴奏スタイルの類似度が計算され，全ての区間で類似度を平均した値を楽曲全体と伴奏スタイルの類似度とする．

4. 評価

本章では提案手法の有効性を確認するための評価について述べる．以下の 2 種類の評価を実施した．

(1) 提案法の特徴量と類似尺度の有効性評価
(2) 楽曲に対する伴奏スタイル識別性能評価

4.1 実験条件

本実験では，音響信号は楽曲，伴奏スタイルともに

44.1kHz，16bit，ステレオの信号を用いた．なおチャネルについては特徴抽出時にモノラルに変換している．実験で用いた分析パラメータを表 2 に示す．このパラメータを

表 2 分析パラメータ

	パラメータ	値
CQT	フレームサイズ	4096
	ホップサイズ	1024
	時間軸のビート単位	1/8 拍
	最小周波数	65.40639Hz
	最大周波数	8372.018Hz
	CQT ビン数	168
GaP-NMF	$a^{(W)}$	1.0
	$b^{(W)}$	1.0
	$a^{(H)}$	1.0
	$b^{(H)}$	1.0
	α	1.0
	K	100
ビートスペクトル	フレームサイズ	1024
	ホップサイズ	512
	時間軸のビート単位	1/8 拍

用いて 3.1 節で説明した伴奏スタイルテンプレートデータベースに対して基底学習および特徴量抽出を行った．

4.2 特徴量の伴奏スタイル識別性能評価

まず，提案法の特徴量と類似尺度の伴奏スタイル識別性能について評価した．評価は多次元尺度構成法（MDS）[12] を用いた特徴量の可視化により行った．伴奏スタイルテンプレートデータベース内の全ての伴奏スタイルに対し大域リズムパタン特徴量および局所リズムパタン特徴量を抽出し，3.5 節で述べた類似尺度により総当りの類似度行列を作成した上で，類似度行列に対し MDS を適用し各伴奏スタイルの特徴量を 2 次元空間に可視化した．なお，リズムパタン特徴量自身の判別能力を評価するため，3.5.4 項で述べた BPM 罰則項はここでは類似尺度に含めていない．可視化の結果を図 5 に示す．

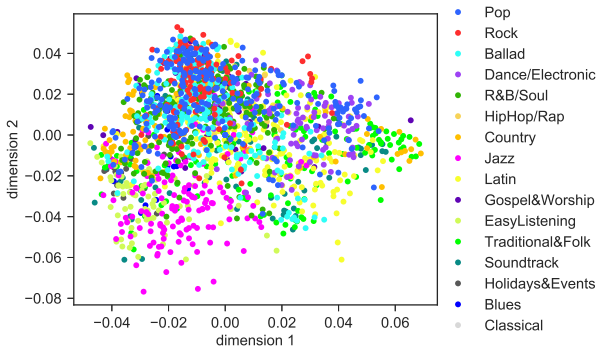

図 5 MDS による伴奏スタイル特徴量の可視化

MDS の可視化結果を観察すると，左下部に Jazz のクラスタがあり，Rock は左上部を中心に分布しており，Dance は右上部に分布しているものが多い，といったようにある程度ジャンルごとにクラスタが形成されていることが見て取れる．ただし，Pop，Ballad，Latin は広く分布しており他のジャンルと重なっている部分が多い．ここで（1）左上部の Pop，Rock，Ballad が重なっている領域 と（2）右上部の Pop と Dance が重なっている領域に着目する．まず（1）の領域にある伴奏スタイルはエレキギターの入ったバンド編成のものが固まっており，Pop の中でも比較的 Rock 寄りのものがこの領域に現れている．また，Ballad が重なっているのはリズムパタン特徴量を BPM で正規化しているため，BPM の差に依存しない特徴量になっていることによるものと考えられる．Ballad は BPM 以外は Pop や Rock と楽器編成やリズムパタンが似ているものが多く，今回用いた特徴量では Pop や Rock と近いところに分布するのは音響的には自然なことと考えられる．（2）の領域は Dance と Pop が重なっているが，出現している Pop のスタイルは Pop の中でもダンスビート主体のものである．こちらも音響的には似たものが出てきている．

以上より，今回用いた特徴量はジャンル識別能力はあまり高くないものの，「ジャンルは異なるがリズムパタンが似ているものを特定する」ことには長けていると言える．

4.3 楽曲の伴奏スタイル識別性能評価

次に，楽曲に対する伴奏スタイル識別性能を評価した．まず評価用データセットとして，洋楽 51 曲の音響信号に対し，最も類似している伴奏スタイルのアノテーションを人手で付与したデータセットを構築した．これを楽曲に対する正解の伴奏スタイルとする．表 3 に楽曲と正解伴奏スタイルの例を示す．

表 3 楽曲と正解伴奏スタイルの例

アーティスト&曲名	正解スタイル名
Bon Jovi - Livin' On A Prayer	80's PowerRock
Queen - Radio Ga Ga	80's Pop
Carpenters - Yesterday Once More	70's PopDuo1
Brian Adams - Please forgive me	SoftRock
Michael Jackson - Earth Song	VocalPopBallad
Eric Clapton - Tears in Heaven	Acoustic8BeatBallad
Jamiroquai - Canned Heat	90's Disco
Madonna - Borderline	SynthPop
Aretha Franklin - Think	FranklySoul
Sting - Brand New Day	ModernShuffle

提案法により楽曲の音響信号から伴奏スタイルの類似度計算を行い，正解スタイルが何位に出現するかの評価を行った．評価結果を図 6 に示す．

横軸は正解スタイルが出現した順位，縦軸は評価楽曲全体のうち，その順位までに正解スタイルが出現した楽曲の割合を示している．約 60%の楽曲で 20 位以内に正解が出

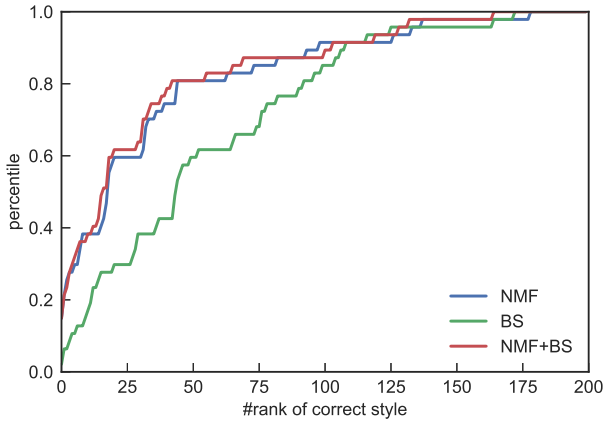

図 6 正解スタイル順位のパーセンタイル

現し，約 80%の楽曲で 40 位以内に正解が出現している．提案法は正解を 1 位で当てるほどの精度はないが，8 割の楽曲で上位 3%以内には正解が出現する程度の精度を備えている．提案法は楽曲に合う伴奏スタイルを選択するタスクにおいて上位の候補をユーザに提示してその中から選択してもらう，あるいは楽曲推薦において上位に出現した伴奏スタイルの組合せを楽曲の特徴量として推薦に用いる，といった応用には十分利用可能な性能となっている．また，特徴量を（1）NMF アクティベーションのみ（2）ビートスペクトルのみ（3）NMF アクティベーションとビートスペクトル に変更した際の精度の変化を見てみると，識別性能に対する寄与としては NMF アクティベーションが支配的であるものの，ビートスペクトルの追加により正解率が向上していることが確認できる．

5. まとめ

本稿では，楽曲の伴奏スタイルを識別するための新しい手法を提案した．伴奏スタイルテンプレートデータベースを構築し， さらに楽器編成と細かなリズムパタンの違いを捉えられる特徴量を考案することによりジャンルよりも細かい区分の伴奏スタイルを識別することが可能となった．今後の課題としては，NMF を用いたリズムパタン特徴量抽出における楽曲と伴奏スタイルの音色やピッチの差異を考慮した手法への拡張や，提案法の前提になっているビート・小節線検出において誤認識が発生した際に精度を落とさない枠組みの導入などが考えられる．

参考文献

[1] ヤマハ株式会社: スタイルの楽しみ方，入手先 ⟨http://jp.yamaha.com/products/musical-instruments/keyboards/fun/style/⟩
[2] ヤマハ株式会社：*Tyros 4*，入手先 ⟨http://usa.yamaha.com/products/musical-instruments/keyboards/arranger_workstations/tyros4/⟩
[3] Tzanetakis, G., Essl, G., Cook, P.: *Musical genre classification of audio signals*, IEEE Transactions on Speech and Audio Processing vol. 10, No. 5, pp. 293-302, Jul. 2002.
[4] Xu, C. , Maddage, N.C., Shao, X., Cao, F., Tian, Q.: *Musical Genre Classification Using Support Vector Machines*, In Proc. of IEEE ICASSP 2003, pp.429-432
[5] Holzapfel, A., Stylianou, Y.: *Musical Genre Classification Using Nonnegative Matrix Factorization-Based Features*, IEEE Transactions on Audio, Speech, And Language Processing, vol. 16, No. 2, pp. 424-434, Feb. 2008.
[6] Markov, K., Matsui, T.: *Nonnegative Matrix Factorization Based Self-Taught Learning With Application To Music Genre Classification*, IEEE International Workshop on Machine Learning For Signal Processing, Sept. 2326, 2012.
[7] Hamel, P., Eck, D.: *Learning Features From Music Audio With Deep Belief Networks*, 11th International Society for Music Information Retrieval Conference (ISMIR 2010).
[8] Smaragdis, P., Brown, C.B.: *Non-Negative Matrix Factorization for Polyphonic Music Transcription*, IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA), pp. 177-180 (2003)
[9] Hoffman, M.D., Blei, D.M., Cook, P.R.: *Bayesian Nonparametric Matrix Factorization for Recorded Music*, Proceedings of the 27th International Conference on Machine Learning (ICML), pp. 439-446 (2010).
[10] Foote, J., Uchihashi, S.: *The Beat Spectrum: A New Approach To Rhythm Analysis*, Proceedings of IEEE International Conference on Multimedia and Expo (ICME), pp. 881-884 (2001).
[11] Kurth, F., Gehrmann, T. , Muller, M.: *The Cyclic Beat Spectrum: Tempo-Related Audio Features for Time-Scale Invariant Audio Identification*, In ISMIR, Victoria Canada, 2006.
[12] Young, F.W., Hamer, R.M.: *Multidimensional Scaling: History, Theory and Applications*, Erlbaum, New York (1987).