

リズムマップ：音楽音響信号からの 単位リズムパターンの抽出と楽曲構造の解析

角尾 衣未留[†] 小野 順 貴[†] 嵯峨山 茂樹[†]

概要：本報告では、音楽音響信号に含まれる小節単位の打楽器パターンを複数種類抽出し、それらの演奏箇所も同時に推定することにより、ビートパターンの地図を作る方法を提案する。本方法が目指す応用事例は、音楽自動ジャンル分類や、音楽情報検索、さらに打楽器パターンを置換するなどの音楽の加工等である。One-pass DP 法を用いた小節への最適なセグメンテーションと k -means クラスタリング法によるそれら打楽器パターンのクラスタリングを反復的に行い情報量規準により推定された最適な数の打楽器パターンを抽出し、その結果それらによる楽曲構造を示すマップを得るアルゴリズムを紹介する。

Rhythm Map: Extraction of Unit Rhythmic Patterns and Analysis of Rhythmic Structure from Music Acoustic Signals

EMIRU TSUNOO,[†] NOBUTAKA ONO[†] and SHIGEKI SAGAYAMA[†]

This paper discusses an approach to extract constituent percussive bar-long unit patterns in a music piece given as acoustic signals and to analyze the music structure with a map of constituent rhythmic patterns. Possible applications include music genre classification, music information retrieval (MIR) and music modification such as replacing rhythmic patterns with others. We propose a mathematical method based on One-pass DP algorithm and k -means clustering to extract the optimum number estimated using information criteria of unit percussive rhythmic patterns. As the result of identifying and localization the unit patterns in the entire piece we obtained a music structure in the form of a map of rhythmic patterns.

1. はじめに

音楽情報検索に関する研究の中でも、特に音楽ジャンル分類のタスクにおいて、リズムに関する特徴量は非常に有力である。例えば、サンバやタンゴなどの楽曲ではそれらの典型的な小節単位の打楽器パターンには大きな特徴が見られる。

リズムは音楽を形成する最も基本的で重要な要素の一つであることは明らかである。ミクロな視点で見れば、単位リズムパターンはビートから構成される小節となることがほとんどである。また、マクロな視点で見れば、楽曲全体に含まれる複数の単位リズムパターンによって楽曲の構造を形成している事が多い。それら複数のリズムパターンがそれぞれ正確に抽出でき、単位リズムパターンから形成される楽曲構造が解析できたならば、リズムパターンに関する音楽解析や音楽ジャンル分類、音楽情報検索に非常に役立つであろうと考えられる。

従来の、リズム解析における最も基本的な研究としては、ビートトラッキングが挙げられる¹⁾。このシステムでは、調波音のオンセット、コードの遷移、打楽器のパターンによって実時間で推定を行っている。ここでの打楽器パターンはスペクトルの低周波域にエネルギーが局在しているバスドラムと広い周波数域にエネルギーが広がっているスネアドラムの組合せであり、音楽的に一般的なパターンに基づく事前知識によってビートの位置を推定している。また、音楽ジャンル分類の研究²⁾では音色特徴量、ピッチ情報特徴量、リズム情報特徴量がそれぞれ抽出される。特にリズム情報特徴量としてのビートヒストグラムは音楽音響信号の時間包絡と自己相関から抽出されたテンポの統計量として計算される³⁾。その他のリズムパターンに関する関連研究としては、リズムパターン間距離の測定が挙げられる⁴⁾。この研究ではスペクトルセントロイドやMFCC(Mel-Frequency Cepstrum Coefficient)等がリズムパターン特徴量として抽出され、この特徴量パターンを Dynamic Time Warping 法 (DTW) を用いて比較する。打楽器のみの音源では良く似たパターンほど距離が近くなる事が確認されたが、実音楽への応用までは至っていなかった。実音楽からリズムパター

[†] 東京大学大学院情報理工学系研究科
Graduate School of Information Science and Technology,
The University of Tokyo

ンを抽出する研究は、音響信号のパワーの周期的なパターンをヒューリスティックな方法により抽出したもの⁵⁾ やスペクトルの周期性に基づく特徴量を抽出したもの⁶⁾ が挙げられ、サンバ、タンゴ等、リズムが特徴的な楽曲の区別が出来る程度の成果が挙げられている。

本報告では、入力音響信号から小節単位の打楽器パターンを抽出し、それら構成打楽器パターンのマップとしての楽曲構造を推定する方法について議論する。次節では小節単位の打楽器パターンを音響信号から抽出するアルゴリズムについて述べ、3節では提案アルゴリズムを実音楽に対し適用し、評価を行う。4節でまとめと今後の展望について述べる。

2. 打楽器パターンの最適なセグメンテーションとクラスタリングアルゴリズム

2.1 リズム解析における問題

打楽器パターンのように周期的に同じパターンが繰り返している中からその単位パターンを抽出したい。仮に一つのパターンの繰り返しであれば比較的簡単な問題になるが、実際の楽曲では複数の単位打楽器パターンがそれぞれのジャンルに適した演奏法に従って演奏される。このような問題は鶏と卵問題であり、単位パターンへのセグメンテーションが与えられていれば複数の単位パターン自体を決定することは難しくなく、逆も然り、複数の単位パターン自体が与えられていれば入力打楽器パターン列のセグメンテーションは容易になると考えられる。その他の問題点としては、テンポ変動の可能性が挙げられる。楽曲中の単位打楽器パターンが伸縮した場合にも正確にセグメンテーションを行う必要がある。これらは打楽器のみの楽曲から単位打楽器パターンを抽出する場合においても問題となる事だが、実際の楽曲、特に現代のポピュラー音楽やジャズ音楽から抽出するには、打楽器のみでなくメロディーや和音が含まれ、それらによって打楽器パターンが部分的に隠されるため更に難しい問題となる。

そのため、打楽器音のテンプレートのような事前知識なしに入力音響信号から小節単位打楽器リズムパターンを抽出する際の問題点は以下の4点にまとめられる。

- (i) 入力音響信号は打楽器音のみでなく、メロディーや和音を含む
 - (ii) 演奏者によってテンポや打楽器パターン自体が変動する
 - (iii) 単位パターンへのセグメンテーションが未知である
 - (iv) 複数の単位打楽器パターン自体が未知である
- この節ではこれらの問題を解決する手法について議論する。

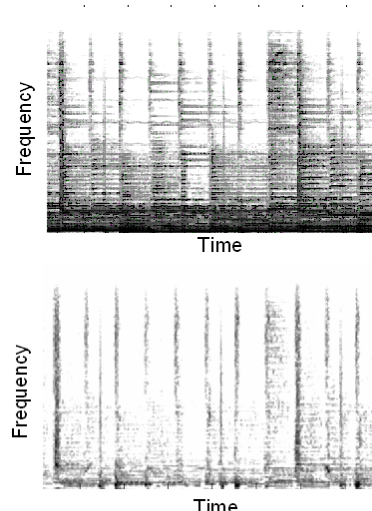


図1 ポピュラー音楽 (RWC-MDB-G-2001 No.6) の元のスペクトログラム (上) と打楽器を強調したスペクトログラム (下)

2.2 打楽器音の強調

上で挙げた問題 (i) はつまり、入力音響信号から得られるスペクトログラムに調波音と打楽器音が混在することである。このようなスペクトログラムから打楽器音を事前知識無しに分離強調する方法として宮本らの手法⁷⁾ が挙げられる。調波音は音高を持つため、そのスペクトログラムはある特定の周波数に偏ってエネルギーが大きくなり、時間的には一定時間演奏される。逆に打楽器音は広い周波数域にエネルギーが広がり、時間的には瞬間しかエネルギーが存在しない。このようなスペクトログラム上の音色の特徴の違いを利用し、時間軸方向に滑らかな調波音と周波数軸方向に滑らかな打楽器音に分離するマスクを MAP 推定の枠組で推定する。このような手法によって図1の上図のスペクトログラムから調波音と打楽器音を分離するマスクが推定され、打楽器音が強調されたスペクトログラムを得る事ができる (図1の下図)。

2.3 リズムによる楽曲構造と単位打楽器パターンの反復的な推定

仮に正解の小節単位打楽器パターンがテンプレートとして与えられていれば上述の問題 (iii) は連続音声認識に類似した問題と考える事が出来るため、One-pass DP 法⁸⁾ を用いて発声された各言葉の箇所を推定するように各单位打楽器パターンの演奏箇所としての楽曲構造の推定を行う事ができる。また、このアルゴリズムは時間の伸縮を許すため、上述の問題 (ii) も同時に解決される。

しかし、実際には問題 (iv) にも挙げられるように正解の単位パターンは与えられない鶏と卵問題であり、当初はセグメンテーションと正解単位パターンの両方が未知であるため、これらを同時に推定する必要があ

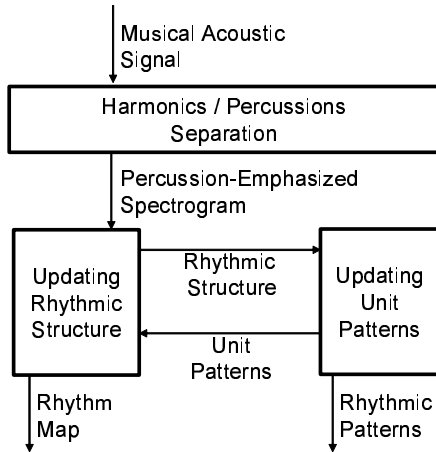


図2 提案アルゴリズムのフロー図

る。つまり教師なし学習問題と捉える事が出来る。そのため、One-pass DP 法と組み合わせる事が考えられる。つまり、クラスタ要素の平均化による単位パターンの更新と楽曲構造の推定を反復的に行うという事である。図2はこのアルゴリズムのフロー図である。各テンプレートパターンからある確率分布に従って入力楽曲の打楽器パターンが演奏されるとすると One-pass DP 法で逐次的に対数尤度を計算し、その総和を最大化する問題として定式化する事が出来る。

2.3.1 One-pass DP 法を用いたリズムによる楽曲構造解析

One-pass DP 法を用いる事で、連続音声認識のように与えられた各单位パターンのテンプレートから最適なセグメンテーションとそれぞれのテンプレートに対応するアラインメントを得ることができる。打楽器パートを含むほとんどの楽曲ではテンポは大きく変化しないため、時間変動の許容幅を小さくするため動的計画法に用いる局所的なパスを図3のように設計できる。

各分析フレーム間のスペクトルの形状はその時刻で演奏される音色を表す。例えばバスドラムのスペクトルは低周波域に大きなエネルギーが存在するのに対し、ハイハットは高周波域に大きなエネルギーが存在する。このようなスペクトルの形状の違いを表したい。入力楽曲の各時間フレームでの各周波数帯域でのエネルギーがテンプレート打楽器パターンの対応する時間フレームの周波数帯域のエネルギーを期待値とする確率分布を成して観測されると仮定すると、確率的なモデルが設計できる。そのため、このような出現確率の尤度が大きい程テンプレートに近い音色であると考えられる事が出来る。

打楽器スペクトログラムの時間周波数成分を $P(t, f)$ とする。ただし t は時刻、 f は対数周波数を表す。実際

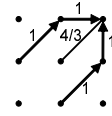


図3 One-pass DP 法での局所経路と経路コスト

には観測値 t, f は離散的なので $P(t_x, f_y) = P_{x,y}$ とし、周波数 bin 数を Y として、さらにある時刻 t_x での周波数成分 $P_{x,1}, \dots, P_{x,Y}$ を要素にもつベクトルを $r_x = (P_{x,1}, \dots, P_{x,Y})^t$ と表す事が出来る。時間周波数成分 $M_m(t_i, f_y) = M_{m,i,y}$ からなる One-pass DP 法でのテンプレートスペクトログラム ($m = 1, \dots, M$) を用意し、これも同様に $\mu_{m,i} = (M_{m,i,1}, \dots, M_{m,i,Y})^t$ と表せる。

このとき時刻 x のスペクトル r_x がテンプレート m の時間フレーム i から出現する確率は $e_{m,i,x} = (\mu_{m,i} - r_x)$ とおいて、

$$p_{m,i}(r_x) = \frac{1}{(2\pi)^{\frac{Y}{2}} |\Sigma_{m,i}|^{\frac{1}{2}}} \exp\left(-\frac{1}{2} e_{m,i,x}^t \Sigma_{m,i}^{-1} e_{m,i,x}\right) \quad (1)$$

と表す事が出来る。ただし、 $\Sigma_{m,i}$ はテンプレート m の時間フレーム i におけるスペクトル周波数 bin 各要素の対角分散行列である。

これを用いて One-pass DP 法アルゴリズムにより、図3の局所経路に従い、各フレームでの対数尤度 $\ln(p_{m,i}(r_x))$ に経路コスト w を掛けたものを逐次的に加えていく。経路コストは図3のように設計するといずれの経路を通ったとしても同じ点にたどり着くものは等価に扱う事が出来る。結果的に、最も尤もらしい経路を捜す事で、各テンプレートパターンに基づいたアラインメントが得られ、用意されたテンプレート打楽器パターンによる楽曲構造が推定される。

対数尤度の総和が最大になる経路として、各時刻フレーム $x(a)$ のスペクトル $r_{x(a)}$ に対してテンプレート $m = m(a)$ の時刻フレーム番号 $i = i(a)$ のスペクトル $\mu_{m(a),i(a)}$ との対応が得られた時、最終的な対数尤度の総和 D_A は $e_{m,i,x}(a) = (\mu_{m(a),i(a)} - r_{x(a)})$ として

$$D_A = -\frac{1}{2} \left(\sum_{a=1}^A (Y \log(2\pi) + \log |\Sigma_{m(a),i(a)}|) \cdot w(a) + \sum_{a=1}^A e_{m,i,x}(a)^t \Sigma_{m(a),i(a)}^{-1} e_{m,i,x}(a) \cdot w(a) \right) \quad (2)$$

と表せる。

2.3.2 k-means クラスタリング法を用いたテンプレート単位打楽器パターンの更新

One-pass DP 法によって入力打楽器パターンが最適なセグメンテーションを行い、それらのセグメントがクラスタリングされる。次に、単位打楽器パターンを推定するため、k-means クラスタリング法のように

各クラスタの中心となるパターンが計算され、新しいテンプレートパターンとして更新される。それぞれの中心パターンは One-pass DP 法によって与えられたアラインメントに基づいて、同じクラスタとしてラベル付けされたセグメントを平均化することで計算できる。実際には、経路コストによる重み付け平均を行うことにより、上述の尤度を最大化することができる。このことにより反復的に One-pass DP 法によるセグメンテーションと各クラスタ中心パターンの計算を行う事で、One-pass DP 法の対数尤度の総和が増加し続け、収束性が保証される。

具体的な更新は最尤法により求まるパラメータ更新 $\hat{\theta} = (\hat{\mu}_{m,1}, \dots, \hat{\mu}_{m,I_m}, \hat{\Sigma}_{m,1}, \dots, \hat{\Sigma}_{m,I_m})$ に従えばよい。

$$\frac{\partial D_A}{\partial \mu_{m,i}} = -\frac{1}{2} \sum_{a \in A_{m,i}} e_{m,i,x}(a)^t \Sigma_{m(a),i(a)}^{-1} e_{m,i,x}(a) w(a)$$
 (3)

よりこれを 0 として、

$$\hat{\mu}_{m,i} = \frac{\sum_{a \in A_{m,i}} r_{x(a)} \cdot w(a)}{\sum_{a \in A_{m,i}} w(a)} \quad (4)$$

ただし、 $A_{m,i} = \{a | m(a) = m, i(a) = i\}$ とした。同様に

$$\begin{aligned} \frac{\partial D_A}{\partial \Sigma_{m,i}} &= -\frac{1}{2} \sum_{a \in A_{m,i}} \Sigma_{m,i} \cdot w(a) \\ &\quad - \frac{1}{2} \sum_{a \in A_{m,i}} e_{m,i,x}(a) e_{m,i,x}(a)^t \cdot w(a) \end{aligned} \quad (5)$$

よりこれを 0 として、

$$\hat{\Sigma}_{m,i} = \frac{\sum_{a \in A_{m,i}} e_{m,i,x}(a) e_{m,i,x}(a)^t \cdot w(a)}{\sum_{a \in A_{m,i}} w(a)} \quad (6)$$

となり、この更新により再度 One-pass DP 法で計算される D'_A はそれ以前に計算された D_A に対して、

$$D'_A \geq \hat{D}_A = \max_{\theta} D_A \geq D_A \quad (7)$$

となり反復更新は目的関数を少なくとも減少させないため、収束性が保証される。

2.4 情報量規準を用いた最適なクラスタ数の決定

この提案アルゴリズムでは、クラスタ数の決定が問題になることが考えられる。つまり、上述の k -means クラスタリング法の k によって結果が大きく変わってしまう可能性がある。当然、クラスタ数が増えれば楽曲中の打楽器パターンをよりうまく表現する事が出来るため、尤度の総和である目的関数 (式 (2)) は大きくなり、楽曲中の小節数分のクラスタを用意すれば最大になると考えられる。しかし、それが最適であるとは考えにくい為、最適なクラスタ数を決定する方法として情報量規準を利用する。尤度が大きくなる程情報量規準が小さくなるが、冗長な情報が多くなる程、

情報量規準は大きくなるため、これを最小化するクラスタ数として決定できる。情報量規準の一つとしてベイズ情報量規準 (BIC) を用いることができる。

上述のアルゴリズムによって観測ベクトル列 r_1, \dots, r_X が小節にセグメント分割され、 R_1, \dots, R_N の N 個のスペクトログラムパターンに分けられるとする。ここで $R_n = (r_{x(n-1)+1}, \dots, r_{x(n)})$ とし、 $x(0) = 0, a(N) = A$ を満たすとする。楽曲の各小節で演奏される打楽器パターン R_1, \dots, R_N が出現する確率はそれぞれの参照パターンを中心とするガウス分布を成すと仮定できる。そこで各クラスタ $m = 1, \dots, M$ にモデル：

$$p_m(\mathbf{R}; \theta) = \left(\prod_{n=1}^N \frac{1}{(2\pi)^{\frac{Y}{2}} |\Sigma_{m,i}|^{\frac{1}{2}}} \right) \exp \left(-\frac{1}{2} \sum_{n=1}^N \left(\sum_{a=a(n-1)+1}^{a(n)} e_{m,i,x}(a)^t \Sigma_{m(a),i(a)}^{-1} e_{m,i,x}(a) w(a) \right) \right) \quad (8)$$

を当てはめることができる。

これも最尤法により式 (8) を最尤とするパラメータは式 (4),(6) となる。このパラメータを用いた最大対数尤度 $l(\hat{\theta}_m; p_m) = \log(p_m(\mathbf{R}; \hat{\theta}))$ を用いてベイズ情報量規準は、

$$\begin{aligned} \text{BIC} &= -2 \sum_{m=1}^M l(\hat{\theta}_m; p_m) + \log(N) \dim \hat{\theta} \\ &= -2 \sum_{m=1}^M l(\hat{\theta}_m; p_m) + M \cdot Y \cdot \log(N) \end{aligned} \quad (9)$$

と表す事ができ、これを最小化するクラスタ数 M が最適なクラスタ数であるといえる。

クラスタを作成する方法としては、まずは一つのテンプレート単位打楽器パターンを用意し、そこから LBG アルゴリズム⁹⁾ のような方法でクラスタ数を増やす事が考えられる。最も分散の大きいクラスタを二分割する事によって一つずつクラスタを増やす。クラスタを増やした後に上述のアルゴリズムによって、最適な単位打楽器パターン推定と楽曲構造解析を行い、情報量規準 (式 (9)) を計算し、これを繰り返し行い情報量規準が増加した時点で打ち切ることで、最適なクラスタ数が決定できる。

2.5 提案アルゴリズムの手順

上で議論したアルゴリズムは以下の手順にまとめられる：

- (1) 打楽器パターンの強調：調波音・打楽器音分離により打楽器音成分が分離、強調されたスペクトログラムを得る。
- (2) 最適なセグメンテーションを行う：式 (2) を最大化する経路を One-pass DP 法によって計算する。
- (3) テンプレートパターンを更新する： k -means クラスタリング法のようにクラスタ中心を式 (4),(6) として計算し、新たにテンプレートパターンと

- して更新する。
- (4) 反復：ステップ 2 と 3 を目的関数 (式 (2)) が収束するまで繰り返し行う。
 - (5) 情報量規準計算：以前の情報量規準の値 (式 (9)) より増加した場合終了。
 - (6) クラスタ分割：LBG アルゴリズムのように最も大きい分散のクラスタを二分割する。
 - (7) ステップ 2 へ戻る。

3. 実験評価

上述の小節単位打楽器パターンを抽出するアルゴリズムが実際の打楽器音のみでなく調波音も含む楽曲に対して適用出来るかを確認するため、評価実験を行った。

3.1 データセット

RWC 音楽データベース¹¹⁾ の WAV ファイルを 22.05kHz、1ch 信号にダウンサンプリングしたものをを用いた。1024 点の長さのハミング窓を半分ずつオーバーラップしながら短時間フーリエ変換 (STFT) を施したスペクトログラムに対し調波音・打楽器音分離を適用した。この時、スペクトル形状の違いを認識する事と計算量削減の両方を実現するため、スペクトログラムの対数周波数を 8 つの帯域に等分割しそれぞれ平均値を求めた。

3.2 結果

上のデータセットのダンス音楽:RWC-MDB-G-2001 No.16 を用いた。テンプレートの単位打楽器パターンを反復的に更新し、収束した後の情報量規準が最小となった時のアラインメントが図 4 となり、学習された個々の基本小節単位打楽器パターンは図 5 となった。

図 4 は演奏開始から一定時間はパターン 1 が演奏される中パターン 2 が演奏されていることを示している。実際の楽曲を聴くと、パターン 1 が繰り返し演奏される中、4 小節に一度パターン 2 が演奏されるのが確認できる。このような基本リズムに続いて、間奏の打楽器パターンが演奏され (パターン 3)、クライマックス部の打楽器パターンが続く (パターン 4)。このように、図 4 は打楽器パターンの地図としての楽曲構造を示していると捉える事ができ、「リズムマップ」と呼ぶことにする。

他の「リズムマップ」の例はポピュラー音楽:RWC-MDB-G-2001 No.6 から推定された、図 6 に示されるような打楽器パターンによる楽曲構造のようになり、図 7 が対応する単位打楽器パターンであった。

3.3 評価

提案したアルゴリズムが人間が行うような打楽器パターン分類を実現できるかを評価する実験を行った。4 つの楽曲:RWC-MDB-G-2001 No. 6, No. 16, No. 19 と No. 26 について同じアルゴリズムを適用した。

まず、これらの楽曲に対して人手による打楽器パ

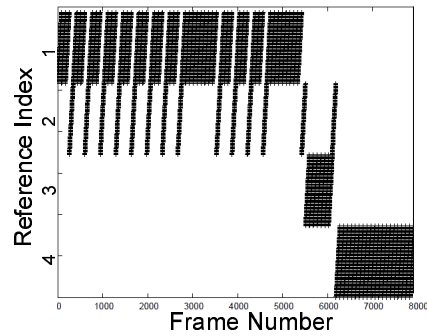


図 4 ダンス音楽 (No. 16) の最適に推定されたアラインメント: 「リズムマップ」

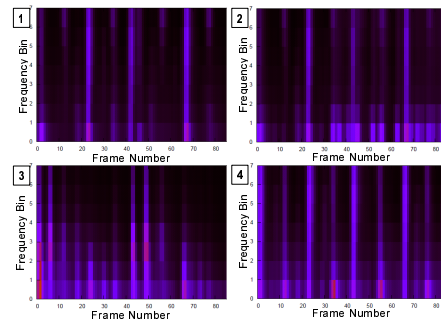


図 5 ダンス音楽 (No. 16) から抽出された 4 つの単位打楽器パターンとそのインデックス番号

ターンのラベル付けを行った。正解のラベルを与えるために、楽曲中の打楽器パターンを分類するためのインターフェースを開発し、数人に正解ラベル付けを行ってもらった。そして得られた正解ラベルと提案アルゴリズムによって推定されたセグメンテーションと分類とを比較した。評価基準として正しく分類されているフレーム数の楽曲全体の割合を計算した。

表 1 に結果をまとめた。少数のデータであるが基本的に提案アルゴリズムは人間が行うように分類できていると考えられる。特にダンス音楽などの打楽器音のエネルギーの大きく正確に抽出しやすい楽曲に対しては効果的な結果が得られた。

4. まとめ

音楽音響信号から小節単位の打楽器パターンを抽出し、それらの地図として楽曲構造を示し解析する手法について議論した。調波音・打楽器音分離手法を用いて音響信号から打楽器を強調したスペクトログラムを得る事ができ、それをを用いて単位打楽器パターンと「リズムマップ」を One-pass DP 法と k -means クラスタリング法を組み合わせさせた方法により反復的に学習するアルゴリズムを提案した。また、最適なクラスタ数を情報量規準を用いて決定する手法を紹介した。実験的に様々なジャンルの楽曲から妥当な単位打楽器パ

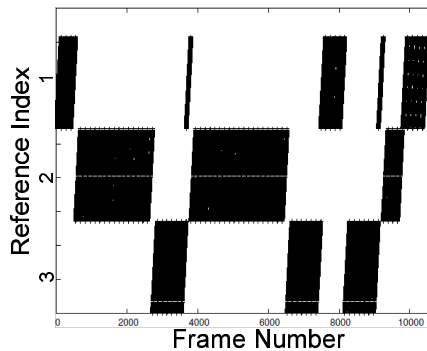


図 6 ポピュラー音楽 (No. 6) の「リズムマップ」

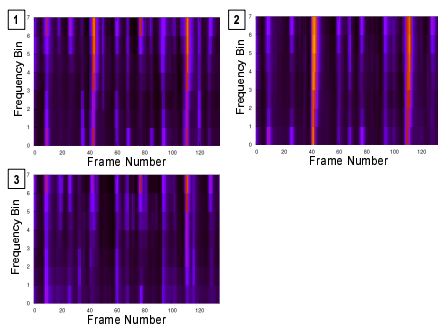


図 7 ポピュラー音楽 (No. 6) から抽出された 3 つの単位打楽器パターン

表 1 正確に分類された区間の割合

Music	Correct Frames (%)
No.6 (Pop)	83.32
No.16 (House)	97.87
No.19 (Techno)	87.25
No.26 (Soul)	69.73

ターンの抽出とそれらの演奏される箇所としてのリズムマップ推定がなされることが確認された。

今後の展望としては、典型的な打楽器パターンによる事前知識の利用や、調波音を用いたコードの遷移情報の利用によりセグメンテーションの精度向上が考えられる。応用としては、このアルゴリズムで抽出される個々の単位打楽器パターンにより、ジャンル分類に適用する事が考えられる。また、楽曲構造を示すリズムマップを利用して、楽曲検索の性能を向上させる事も考えられるであろう。

参 考 文 献

- 1) Goto, M., "An audio-based real-time beat tracking system for music with or without drum-sounds," *Journal of New Music Research*, Vol. 30, No. 2, pp. 159–171, June 2001.
- 2) Tzanetakis, G. and Cook, P., "Musical genre classification of audio signals," *IEEE Transaction on Speech and Audio Processing*, Vol. 10, No. 5, pp. 293–302, 2002.
- 3) Tzanetakis, G., Essl, G. and Cook, P., "Audio analysis using the discrete wavelet transform," *Proc. of*

WSES Int. Conf. on Acoustics and Music: Theory and Applications, 2001.

- 4) Paulus, J. and Klapuri, A., "Measuring the similarity of rhythmic patterns," *Proc. of the 3rd Int. Conf. on Music Information Retrieval (ISMIR 2002)*, pp. 150–156, IRCAM Centre Pompidou, 2002.
- 5) Dixon, S., Guyon, F. and Widmer G., "Towards characterization of music via rhythmic patterns," *Proc. of the 5th Int. Conf. on Music Information Retrieval (ISMIR2004)*, Barcelona, Spain, 2004.
- 6) Peeters, G., "Rhythm classification using spectral rhythm patterns," in *Proc. of the 6th Int. Conf. on Music Information Retrieval (ISMIR2005)*, pp. 644–647, London, UK, September 2005.
- 7) Miyamoto, K., Kameoka, H., Ono, N. and Sagayama, S., "Separation of harmonic and non-harmonic sounds based on anisotropy in spectrogram," *Proc. of ASJ Spring Meeting*, pp. 903–904, March, 2008.
- 8) Ney, H., "The use of a one-stage dynamic programming algorithm for connected word recognition," *Int. Conf. on Acoust., Speech, Signal Processing (ICASSP)*, pp. 263–271, 1984.
- 9) Linde, Y., Buzo, A. and Gray, R., "An algorithm for vector quantizer design," *IEEE Trans. Commun.*, vol. 28, pp. 84–95, 1980.
- 10) Rabiner, L. and Juang, B. H., "Fundamentals of Speech Recognition," chap. 4, pp. 200–238, Prentice Hall, 1993.
- 11) Goto, M., Hashiguchi, H., Nishimura, T. and Oka, R., "RWC music database: music genre database and musical instrument sound database," *Proc. of the 4th Int. Conf. on Music Information Retrieval (ISMIR 2003)*, pp. 229–230, October 2003.