

調波音・打楽器音分離手法を用いた 音楽音響信号からのリズム特徴量の抽出*

角尾衣未留、宮本賢一、小野順貴、嵯峨山茂樹 (東大・情報理工)

1 はじめに

近年、膨大な音楽の中から所望の楽曲ファイルを効率よく探し出すことを目的に、コンテンツベースの楽曲検索や分類に関する研究が盛んに行われているが、音色やメロディー、ピッチ、テンポ等の特徴量を別々に抽出するものが多い。

その中の一部であるリズム特徴量について、従来では、時間情報を含むものとして、音響信号の振幅値から自己相関を計算したテンポ情報、いわゆるビートと呼ばれるものが用いられている [4]。打楽器の音色に関しては、スペクトルの形状、例えば、スペクトル重心やフラットネス、MFCC 等の特徴量があり、打楽器認識に用いられる [1]。これら音色情報に時間情報を加えた打楽器パターンについては、パターン間距離の測定に関する研究はなされている [3] が、実際の楽曲は打楽器音のみでなく調波音も含んでいるため、スペクトル形状のみからの打楽器パターン抽出は容易でない。

今回我々は、調波音・打楽器音分離手法を用いて打楽器音に着目して楽曲特有の打楽器パターンを抽出する手法を検討したので報告する。

2 本研究のアプローチ

2.1 打楽器音の強調

一般の楽曲は調波音と打楽器音の両方を含むことが多いが、そこから打楽器音を分離強調したい。この目的には宮本らの手法 [6] が利用できる。調波音のスペクトルは時間軸方向には滑らかであるが、周波数はある特定の周波数付近に偏って音高を成すことが多い。逆に、打楽器音のスペクトルは周波数軸方向には急激に変化しないが、時間的には瞬間的にしか存在しないことが多い。そのため、時間軸方向と周波数軸方向の滑らかさをコストとして目的関数に加え EM アルゴリズムでマスクを推定することで、打楽器音を分離強調できる。

2.2 打楽器パターンの抽出

ポピュラー音楽等、打楽器を含む音楽では、あるリズム構造の打楽器パターンが繰り返し用いられることが多いので、それを構成する最小パターンを抽出し、それをリズム特徴量とすることを考える。複数のパターンが存在する場合も多いが、それぞれの打楽器パターンを、上述の打楽器音のスペクトル系列から抽出したい。

ここでの問題は、複数種類のパターンの区間は当初は未知であること、どれが同じパターンであるか

を推定すること、それらのパターンは演奏やアレンジ等により多少の変動を受けていることである。

これらの問題に対処するためには、連続音声認識で用いられる One-pass DP 法 [2] の利用が考えられる。個数既知の複数の適切な初期参照パターンがあれば、観測入力をこれらの参照パターンとマッチングして、最適なパターン分類とセグメント分割を行える。DP マッチング経路を適切に設定すれば、パターンの変動を吸収することができる。パターン分類とセグメント分割に基づいて参照パターンを更新して、この手順を繰り返せば、 k -means アルゴリズムの原理により、局所最適に更新された参照パターンとセグメント分割とセグメントごとのパターン分類が行える。このような学習アルゴリズムにより、楽曲の音響信号から打楽器パターン抽出法を抽出することができる。

3 One-pass DP 法を用いた打楽器パターン抽出法の定式化

One-pass DP で計算される距離は参照パターンと入力パターンが近くなるほど小さくなるため、One-pass DP の距離を極小化する問題として定式化できる。

入力パターン列を $X_1, \dots, X_m, \dots, X_M$ 、参照パターン R_v に対するパターン列を $Y_{v1}, \dots, Y_{vn}, \dots, Y_{vN_v}$ とする。今回用いたノード間の距離関数は

$$d(X_m, Y_{vn}) = (X_m - Y_{vn})^t \Sigma^{-1} (X_m - Y_{vn}) \quad (1)$$

とした。 Σ は入力パターンに対する対角分散行列である。

One-pass DP により、参照パターンの各フレームそれぞれに対して入力パターンのフレームとの対応関係が得られる。参照パターンのフレーム Y_{vn} に対応する入力パターンのフレームを X_{vn} とすると、One-pass DP で計算される距離 D は

$$\begin{aligned} D &= \min_{1 \leq v \leq V} [D_A(M, N_v, v)] \\ &= \min_{1 \leq v \leq V} \left[\sum_{Y_{vn}} d(X_{vn}, Y_{vn}) \right] \end{aligned} \quad (2)$$

と書く事ができるので、 $\sum_{Y_{vn}} d(X_{vn}, Y_{vn})$ を最小化するには Y_{vn} を対応する X_{vn} の平均とすればよく、更

* "Rhythmic Features Extraction from Music Acoustic Signals using Harmonic/Non-Harmonic Sound Separation" by Emiru TSUNOO, Ken-ichi MIYAMOTO, Nobutaka ONO, and Shigeki SAGAYAMA, The University of Tokyo.

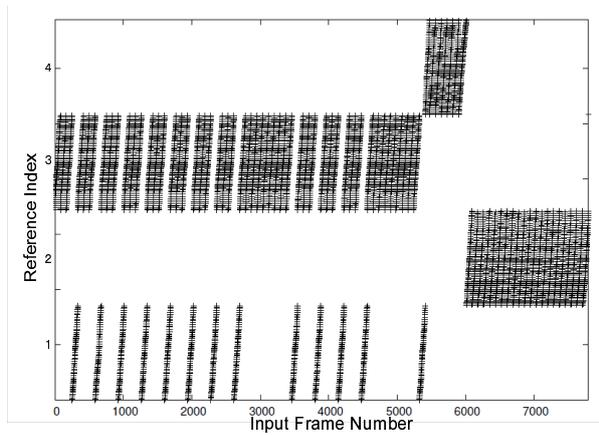


Fig. 1 学習された打楽器パターンに対して One-pass DP 法によって求めたアラインメントの結果

新後の距離 D' は

$$\begin{aligned}
 D' &= \min_{1 \leq v' \leq V'} [D_A(M, N_{v'}, v')] \\
 &\leq \sum_{X_{vn}} d(X_{vn}, \bar{X}_{vn}) \\
 &\leq \sum_{Y_{vn}} d(X_{vn}, Y_{vn}) \\
 &= D
 \end{aligned} \quad (3)$$

となる。距離の総和 D は正であり、各更新により減少するか少なくとも非増大であるので、収束することが保証される。

4 実験

上で述べたアルゴリズムを実際の楽曲に適用し、打楽器パターンを抽出する実験を行った。RWC 音楽データベース [5] 中の WAV ファイルを 1 チャンネル、22.05kHz にダウンサンプリングし、調波音・打楽器音分離手法により得られた打楽器強調音を、長さ 1024 点、シフト 512 点のフレームで短時間フーリエ変換し、得られたパワースペクトルを入力パターンとした。パターン次元を減らすため、パワースペクトルは周波数方向に対数的に 8 つの bin に平均化した。初期参照パターンとしては BPM 120 の典型的な打楽器パターンと思われるスペクトルパターン 2 秒間分 (1 小節に相当) の打楽器パターン数分の種類を与えた。参照パターンのフレーム数は固定した。

上のデータのうち 4 種類の打楽器パターンを持つダンス音楽、RWC-MDB-G-2001 No. 16 に対して One-pass DP 法を用いたクラスタリングを行い、収束した結果を図 1 に示す。横軸は入力パターンのフレーム数、縦軸は参照パターンを縦に並べたものである。また、図 2 は学習されたそれぞれの参照パターンのスペクトルである。低周波のみの音、高周波のみの音、周波数全域に広がって鳴る音、という具合に見ると、これらが打楽器の音色パターンを表している事が確認できる。

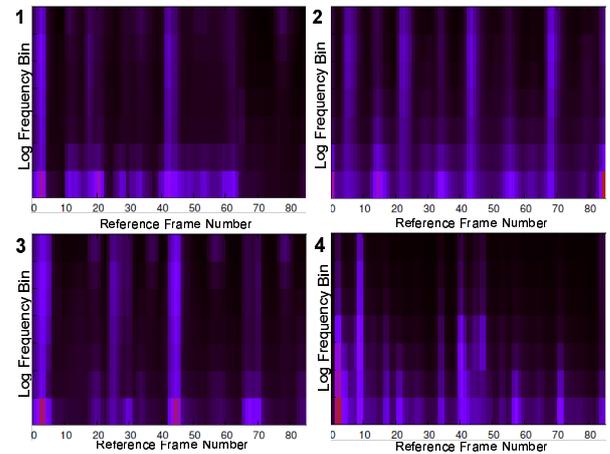


Fig. 2 学習された 4 種類の打楽器パターンのスペクトルとそれぞれのインデックス番号

5 おわりに

調波音と打楽器音の分離手法と One-pass DP 法を用いた学習法により、楽曲の打楽器パターンを抽出する手法を検討した。打楽器パターン数を既知として 1 楽曲内で局所解に収束させる事で複数の打楽器パターンが抽出できる事が実験により確認できた。今後の応用としては、大量の楽曲で打楽器パターンを抽出し、それらのリズム特徴量を辞書として持つ事で検索の性能向上や、音楽のクラス分類に利用できる事が考えられる。課題としては、打楽器パターン数が未知である場合にも対応する必要がある。

参考文献

- [1] Herrera, P. *et al.*, “Automatic labeling of unpitched percussion sounds,” *Presented at the 114th Convention of the Audio Engineering Society*, Amsterdam, Netherlands, 2003.
- [2] Ney, H. and Ortmanns S., “Dynamic programming search for continuous speech recognition,” *IEEE Signal Processing Magazine*, Vol. 16, No. 5, pp. 64–83, Sept. 1999.
- [3] Paulus, J. and Klapuri, A., “Measuring the similarity of rhythmic patterns,” *In Proceedings of the 3rd International Conference on Musical Information Retrieval*, pp. 150–156, IRCAM Centre Pompidou, 2002.
- [4] Tzanetakis, G. and Cook, P., “Musical genre classification of audio signals,” *IEEE Transactions on Speech and Audio Processing*, 10(5), pp. 293–302, 2002.
- [5] 後藤 他, “RWC 研究用音楽データベース:音楽ジャンルデータベースと楽器音データベース,” 情報処理学会 音楽情報科学研究会 研究報告, 2002-MUS-45-4, Vol. 2002, No. 40, pp. 19–26, May 2002.
- [6] 宮本 他, “スペクトログラムの滑らかさの異方性に基づく調波音・打楽器音の分離,” 日本音響学会 春季研究発表会講演論文集, 2-5-8, Mar. 2008.