動的計画法に基づく音楽構造解析とその音楽信号符号化への応用*

☆ 藏 悠子 (東大・工)，鎌本 優 (NTT・CS 研)，
小野 順貴 (NII)，嵯峨山 茂樹 (東大院・情報理工)

1　はじめに

近年、MP3 や AAC のような高圧縮かつ高音質を実現する非可逆圧縮技術の発展により、音楽デジタルデータの普及が目覚ましい。ユーザは携帯オーディオプレーヤーや携帯電話などの端末を通じて音楽データを大量に取得・保有することが可能となっている。しかしその一方で、修正離散コサイン変換と聴覚心理モデルを基盤とした現在の圧縮技術は飽和状態にあるとも言える。

我々は、楽曲を曲毎にダウンロードして端末に保存した後に受聴する応用に目的を限定し、フレームごとの逐次符号化処理を諦める代わりに、音楽の繰り返し構造や大域的冗長性を利用することにより圧縮率を向上させる新しい音楽信号符号化方式を検討している。本研究ではその第一歩として、音楽構造解析手法について論じる。

2　音楽構造を利用した音楽信号符号化

音楽構造を利用した符号化の基本的なアイディアは、楽曲中で似ている箇所を音響信号のみから自動検出し、検出された音楽構造とともに、2 回目以降の類似箇所は 1 回目との差分のみを符号化することで情報量を削減することを目指すものである (Figure 1 参照)。音楽構造解析自体は、楽曲のサムネイル作成、ビデオと音楽の同期、音楽情報検索などの応用からも研究されている (例えば [1]) が、本研究における構造解析の目的は、作曲者の意図した楽曲構造や楽曲を特徴付ける構造を推定することではなく、楽曲中の類似箇所とそれらの間の対応関係を自動的に求めることである。本研究では、最適経路問題としてこの問題を定式化することを検討した。

3　動的計画法に基づく音楽構造解析

3.1　構造解析の特徴量

楽曲中の類似箇所であっても波形が似ているとは限らないが、スペクトログラム同士は似ていることが多い (Figure 2 参照)。我々が別途研究をすすめているスペクトログラムの振幅位相符号化 [2, 3] の適用も想定し、本研究では構造解析の特徴量として対数振幅スペクトログラムを用いた。M を周波数ビンの数、A_{im} を i 番目のフレームにおける m 番目の周波数ビンの振幅スペクトルとすると、状態 $\mathbf{x}_i$ は

$$\mathbf{x}_i = \begin{pmatrix} \log A_{i1} \\ \log A_{i2} \\ \vdots \\ \log A_{iM} \end{pmatrix} \tag{1}$$

と表される。

3.2　フレーム間の類似確率

一般に、特徴量 $\mathbf{x}_i$ と $\mathbf{x}_j$ が近いほど、i 番目と j 番目のフレームが繰り返し構造として対応する確率は高いと考えられる。ここでは最も単純に、この確率を

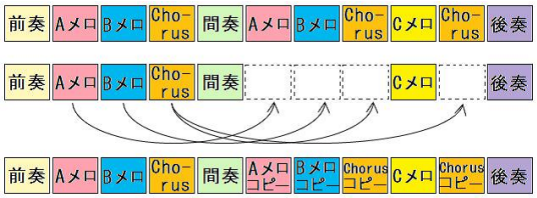

Fig. 1　音楽構造を利用した符号化の概念図。繰り返しを含む楽曲 (上) の構造を用い、類似箇所の 2 回目以降は 1 回目のコピーに基づき差分のみを符号化する

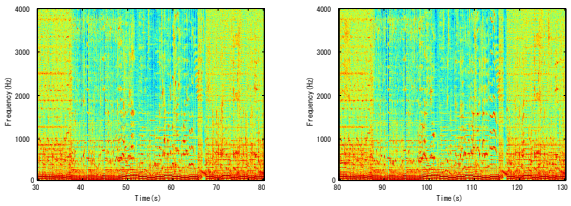

Fig. 2　ポップスの楽曲 (RWC-MDB-P-2001 No.1[5]) の 1 番 (左) と 2 番 (右) 部分の振幅スペクトログラム

多次元正規分布を用いて以下のようにモデル化する。

$$P(i,j) = \left(\frac{1}{\sqrt{2\pi\sigma^2}}\right)^M \exp\left(-\frac{|\mathbf{x}_i - \mathbf{x}_j|^2}{2\sigma^2}\right) \tag{2}$$

Figure 3 は式 (2) の対数値を可視化したものであり、いわゆる類似度行列に相当する。値が小さいほど濃い赤色で表されている (分析条件は後述)。ここでの問題は、この行列上で、類似している箇所を結ぶ経路を求める問題となる。ただ似ている対応箇所をつなげただけでは、たくさんのジャンプを含む、符号化に適さない経路が求まってしまうため、経路にも音楽進行中での起こりやすさを反映した確率を導入する必要がある。

3.3　音楽進行の確率モデル

楽曲の繰り返し構造を状態遷移によりモデル化すると、以下の典型的な進行があり得ると考えられる。(括弧内は類似度行列中での表現)

1. 新しい状態へ進行する、もしくは対応箇所と同じテンポで進行する (対角線、もしくは斜め 45 度の線分)
2. 対応箇所と少し異なるテンポで進行する (斜め 45 度の線分から少しずれる経路)
3. 離れた状態へジャンプする

それぞれの確率を p_{const}, p_{deviate}, p_{jump} とする。例えば全フレーム数 N に対し、音楽構造ブロック間のジャンプが n_{jump} 回起こるのであれば、

$$p_{\mathrm{jump}} = \frac{n_{\mathrm{jump}}}{N} \tag{3}$$

---

*Structural Analysis of Musical Signal by Dynamic Programming and Its Application to Audio Coding by Yuko ZOU (The University of Tokyo), Yutaka KAMAMOTO (NTT), Nobutaka ONO (NII) and Shigeki SAGAYAMA (The University of Tokyo)

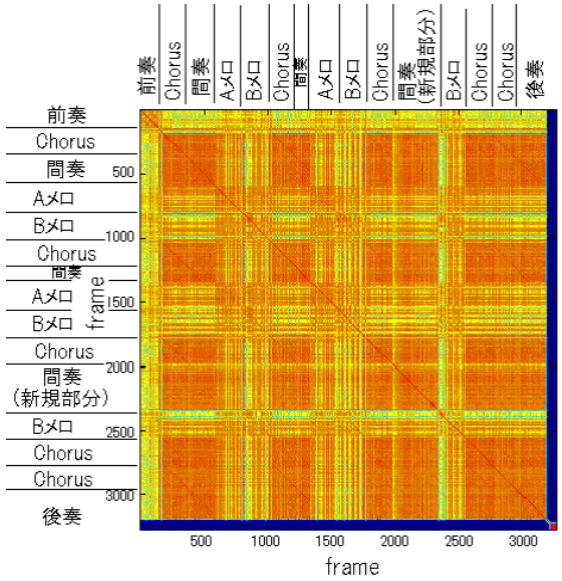

Fig. 3　フレーム間での対数振幅スペクトルの類似確率 (フレーム同士が似ているほど濃い赤色)

のように表されるが、実際には n_{jump} は未知であるので、適当な値を予め定めておくこととする。

以上のモデルに基づき、最も確率が高くなる経路を選ぶ問題は、最適経路探索問題となる。各経路の尤度は類似確率と遷移確率の積で表されるが、対数尤度を考えることにより動的計画法が適用可能となり、少ない計算量で大域的最適解を求めることができる。

4　実験

4.1　実験条件

前節で述べた提案法の有効性を検証する実験を行った。本来は符号化した符号長により圧縮率を評価するべきであるが、ここではその前段階として、提案法により得られた最適経路と手動でラベル付けした音楽構造を比較することによりその有効性を検証した。なお、すでに複数の楽曲に対して本手法を試しているが、ここでは紙面の制約から、RWC 音楽データベースのポップスの楽曲 (RWC-MDB-P-2001 No.1[5]) に対して適用した例のみを示す。

本実験ではまず、楽曲を 8 kHz にダウンサンプリングし、ハミング窓を用いた短時間フーリエ変換によりスペクトログラムを求めた。フレーム長は 1024 点、フレームシフトは 512 点、とした。

本手法の確率モデルには、いくつかのパラメータが含まれている。式 (2) の σ^2 は、予め RWC 音楽データベースの 10 曲分のデータに対し実験を行い、最も良い結果になったものとして $\sigma^2 = 2500$ を採用した。また、状態遷移確率も実験的に $p_{\mathrm{deviate}} = 500/3000$, $p_{\mathrm{jump}} = 5/3000$, $p_{\mathrm{const}} = 1 - p_{\mathrm{deviate}} - p_{\mathrm{jump}}$ と定めた。これらの値は、本来は学習により求めることが望ましく、今後の課題の 1 つである。

また本実験では、参照フレームを楽曲の 1 番に相当する部分に制約して最適経路を求めた。

4.2　実験結果

Figure 3 は前述の通り、式 (2) の対数値を可視化したものである。手動でラベリングした結果と比べると、音楽構造が反映されていることが確認できる。

Figure 4 は、参照フレームを曲の 1 番が終わる 1222 フレーム目で制限した最適経路である。ラベルと照らし合わせると、新しく現れた部分である間奏部分以外は、繰り返し構造の対応関係が検出されていることが確認できる。

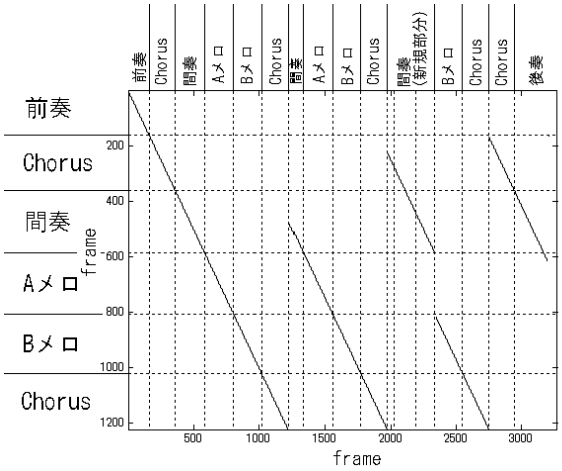

Fig. 4　動的計画法により求まった最適最適経路

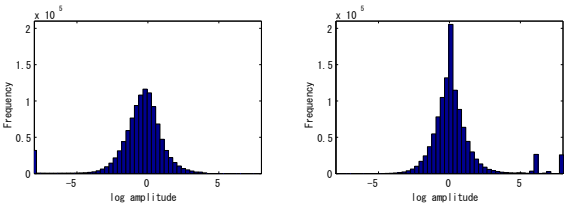

Fig. 5　対数振幅ヒストグラム (左) と残差ヒストグラム (右) の比較

Figure 5 は、1223 フレーム以降の、対数振幅スペクトルそのもののヒストグラムと、最適経路に基づき対応する類似箇所との差分のヒストグラムを比較したものである。対応する類似箇所を検出して差分のみを求めることにより、ヒストグラムがより急峻になっていることが確認でき、情報量の削減が期待できる。

5　おわりに

本稿では音楽信号符号化を目的に、楽曲の音楽構造を自動的に求める手法について検討した。信号の振幅スペクトルの類似確率と遷移確率を使った動的計画法により、音楽構造が検出できることを確認した。今後はこれに基づき、符号長や音質の評価を行っていきたい。

謝辞　本研究は、日本学術振興会科学研究費補助金 (挑戦的萌芽研究：23650083) の助成を受けたものである。

参考文献

[1] Levy *et al.*, *IEEE TASLP*, Vol. 16, No. 2, pp. 318–326, 2008.

[2] 佐藤 他, 音講論 (春), pp. 337–338, 2011.

[3] 佐藤 他, 音講論 (秋), pp. 229–230, 2011.

[4] 松井 他, 情処研報, 2002(14), 33-38, 2002.

[5] 後藤 他, 音講論 (春), pp.705–706, 2002.