

## コード進行モデルを用いた楽曲のコード認識の検討

平賀 弘平<sup>†</sup> 山田 武志<sup>‡</sup> 北脇 信彦<sup>‡</sup>

筑波大学第三学群情報学類<sup>†</sup> 筑波大学システム情報工学研究所<sup>‡</sup>

### 1. はじめに

楽曲の和音構成を記号で表したものをコードといい、その並びをコード進行という。一般に音楽は、メロディー、リズム、和音、の三要素で構成されていると考えられている。音楽音響信号からコードを認識する技術は、その楽曲の要約を得ることに似ており、楽曲の分析・演奏・作曲・編曲活動をするのに非常に有用である。また得られたコード進行を用いると、楽曲検索、自動伴奏に応用できる重要な技術である。

楽曲からコード認識を行うには、複雑な混合音と多様な音域中から最も適したコードを同定する必要がある。従来、入力音響信号と学習データを各次元が 12 音名に対応する 12 次元のベクトルに変換し、k-NN 法を用いてコードを同定する手法が提案されている[1]。さらに N-gram コード進行モデルを使い、同定精度を向上させる手法も提案されている[2]。

これらの研究により、一般の楽曲に対する和音同定手法の枠組みが構築された。しかし、歌声や打楽器音を含む、より複雑な楽曲に対しては、なかなか十分な性能が得られていない。その一つの理由は、N-gram モデルを構築する為の、十分な学習データを用意することが困難なことが挙げられる。また、コードが推移する時間位置を含めたコード認識は未解決となっている。これは楽曲に自動伴奏させる場合に重要である。

そこで本研究では、以上の問題に対処し、より複雑な楽曲からコード進行を同定する手法を検討する。

### 2. 提案手法

#### 2.1. 概要

入力音響信号のパワースペクトルを、フレームごとに 12 次元の特徴ベクトル(クロマベクトル[3])に変換する。特徴ベクトルの各次元は、12 音名に対応する。次に、ビタービアルゴリズムを用いて、特徴ベクトルからコスト最小にな

るコード進行を認識する。この時、ビタービアルゴリズムのコスト関数に、和音の特徴ベクトル(テンプレート)と楽曲から抽出した特徴ベクトルの余弦と、後述する N-gram コード進行モデルを用いることにする。また楽曲のテンポ情報を与えられる場合は、コード長に制約を与えて、より正確なコード認識を行えるようにする。

具体的な手順を以下に述べる。

#### 2.2. 特徴ベクトル抽出

入力音響信号から特徴ベクトルを抽出する方法を説明する。

入力音響信号を標準化周波数 16kHz、量子化ビット数 16bit のモノラル音響信号に変換し、窓幅 256msec のハニング窓、シフト幅 128msec で FFT を行い、パワースペクトルを得る。フレームごとに、パワースペクトルを対数スケール周波数から平均律スケールに変換しつつ、各次元が 12 音名に対応する 12 次元の特徴ベクトルに変換する。現在は、オクターブ位置 3 未満のパワースペクトルは特徴ベクトルに変換していない。

詳細については文献[3]を参照されたい。

#### 2.3. N-gram コード進行モデルの学習

コード認識の精度向上のために、N-gram コード進行モデルを学習する。

コードの生起は直前のコードに依存するという仮定に基づき、次に来るコードの確率を N-gram で求める。本研究では、Bigram(N=2)について検討した。学習は、WWW からポップスのコード進行データ約 800 曲分を収集し、例えば「コード $\alpha$ からコード $\beta$ への推移数 / コード $\alpha$ から全コードへの推移数」で、推移確率を求める。これをすべてのコードについて行う。

また、学習する楽曲の調性に偏りがあると、出現するコード進行も偏ってしまう問題がある。そこで、楽曲を移調したコード進行も学習することで、この問題を解決する。楽曲中に出てくるコードのルート音名を半音ずつシフトしていき、12 調性すべてのコード進行を学習させた。

#### 2.4. 認識

和音特徴ベクトル(テンプレート)と N-gram コード進行モデルを用いて、入力特徴ベクトルからコード認識をする方法を説明する。

一般に和音は、二分音符や全音符の長さでの

Chord Recognition Using N-gram Chord Sequence Model: Kohei Hiraga, Takeshi Yamada, and Nobuhiko Kitawaki (Univ. of Tsukuba)

音で構成されていることが多い。単一フレームからの特徴ベクトルでは区間長が短いので、アルペジオ等の場合、和音の構成音が一部出現しないという問題がある。そこで、隣り合った複数フレームから、特徴ベクトルの平均ベクトルを求め、その区間の特徴ベクトルとする。

$V = \{v_1, v_2, v_3, \dots, v_L\}$  を、長さ  $L$  の入力特徴ベクトル列、 $S_k = \{s_1, s_2, s_3, \dots, s_M\}$  を特徴ベクトル区間列と定義する。ここで  $|S_k|$  は区間数、 $s_i$  は区間長とする。例えば、 $s_1 = \{v_1, v_2\}$  であり、区切り位置は下記のビタービアルゴリズムにより最適なものが設定される。

$C_k = \{c_1, c_2, c_3, \dots, c_M\}$  をコード列とする。 $c_i$  は  $\{s_i\}$  のコード認識候補を表す。認識によって求めたい出力区間列  $\hat{S}$  と、出力コード列  $\hat{C}$  は、式(1)となる。

$$(\hat{S}, \hat{C}) = \underset{S, C}{\operatorname{argmin}} \sum_{i=1}^{|S|} \{ \operatorname{cost}(s_i, c_i) \times (1 + \operatorname{con}(c_{i-1}, c_i) \times w) \times |s_i| \} \quad \dots(1)$$

( $w$  は定数、現在は 0.5 を設定している。 $c_0$  は接頭コードという特別なコードとする)

$\operatorname{cost}(s_i, c_i)$  は、特徴ベクトル区間  $s_i$  と和音  $c_i$  の特徴ベクトルの類似度によるコスト関数である。類似度は、 $s_i$  と、基本周波数成分のみで構成される  $c_i$  の特徴ベクトル(テンプレート)を、それぞれ長さが 1 になるよう正規化した後、内積を求めて、「1-内積の値」をコストとした。

$\operatorname{con}(c_{i-1}, c_i)$  は、コード  $c_{i-1}$  と  $c_i$  の接続コスト関数である。この関数は N-gram コード進行モデルに基づき、「1 -  $c_{i-1}$  から  $c_i$  への推移確率(0~1)」をコストとした。

式(1)を満たす  $S$  と  $C$  を求めるには、ビタービアルゴリズムを用いた。本問題に適用する際、特徴ベクトル区間長  $|s_i|$  は、 $4 \leq |s_i| \leq 15$  の範囲に限定した。これは、512msec より短い長さでコードが推移することは稀で、1920msec 以上の長い区間も必要になることは稀であるという考えから決定した。

楽曲のテンポ情報が与えられる場合、特徴ベクトル区間長  $|s_i|$  に二つの制約を加える事で、コードが推移する時間位置を考慮したコード認

識ができる。一つは、特徴ベクトル区間長  $|s_i|$  は、二分音符または全音符の長さとする。もう一つは、二分音符の区間長は必ず 2 つ続くこととする。

### 3. 認識実験

本手法の有効性を検証するために認識実験を行った。RWC 研究用音楽データベースのポピュラー音楽 CD より、曲番号 13 の A メロ部分 8 小節(図 1(a))を N-gram コード進行モデルなし(b)、Bigram 適用(c)、Bigram+テンポ情報適用(d)でそれぞれ認識した結果、図中赤丸の部分で(b)より(c)に改善が確認できた。また、(d)ではコードが推移する時間位置が小節と一致することが確認できた。以上より、本手法の有効性を確認できた。

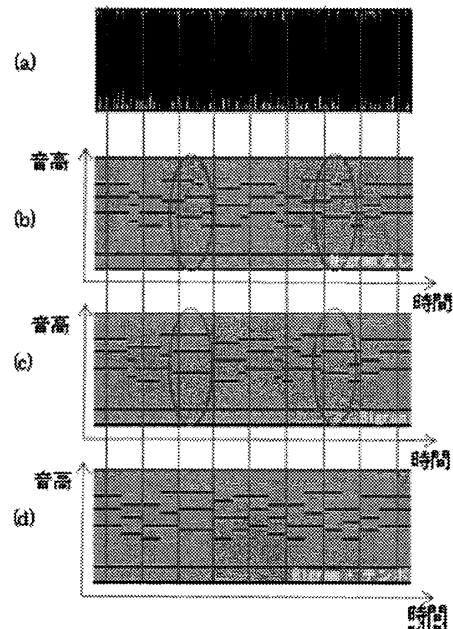


図 1: コード認識結果の一例

### 4. まとめ

本研究は、複雑な音楽音響信号からコード認識を行うという問題に対して、N-gram コード進行モデルとビタービアルゴリズムを用いた手法について検討した。

今後はさまざまな楽曲にコード進行データをラベリングし、本手法の性能評価を定量的に行いたい。また、Trigram コード進行モデルについても検討し、よりロバストなコード認識手法に拡張して行きたいと考えている。

#### 参考文献

- [1]山田 他, “音楽音響信号を対象とした和音名同定手法,” 日本音響学会講演論文集, pp. 641-642, Sep. 2002.
- [2]菅野 他, “ガウス混合モデルを用いた楽曲のコード認識,” 信学技報, CAS2003-26, Jun. 2003.
- [3]後藤 他, “リアルタイム音楽情景記述システム:サビ区間検出手法,” 情報処理学会研究報告, 2002-MUS-47-6, Oct. 2002.