和音認識のためのスペクトル特徴量の検討*

☆上田 雄, 内山 裕貴, 小野 順貴, 嵯峨山 茂樹 (東大·情報理工)

1 はじめに

本稿では音楽音響信号からの自動和音認識について扱う。西洋音楽などの調性音楽において、和声は音楽おける重要な要素の一つであり、実演奏からの自動和音認識はコード譜の作成だけでなく自動採譜や、カバー曲同定、音楽データベースの自動タグ付けなどの音楽情報検索 (MIR) の分野への応用の手掛かりとなるであろう。

音響入力を対象とした和音認識においては Pitch Class Profile (PCP) またはクロマベクトル [1] が特徴量として広く用いられており、我々も既にクロマベクトル特徴量を改善する手法を提案してきた [2, 3]。クロマベクトルは音楽的知見に基づく有効な特徴量である一方、倍音などの影響により成分ごとに相関を持つため、学習に必要なパラメータ数が少なくない。また、オクターブの相異を圧縮するため和音の転回形を認識できないことや、音域の情報を無視してしまうため、例えば和音の根音が低音域で演奏されやすい、といった和音認識において有意な情報が失われてしまう問題点がある。そこで我々は今回、情報量の損失の少ない学習パラメータの削減手法という観点から 1) クロマの成分を無相関化する手法、2) スペクトログラムの主成分分析 (PCA) に基づく手法を提案し、それらを実験的に検証する。

2 和音特徴量の抽出

2.1 クロマベクトル特徴量

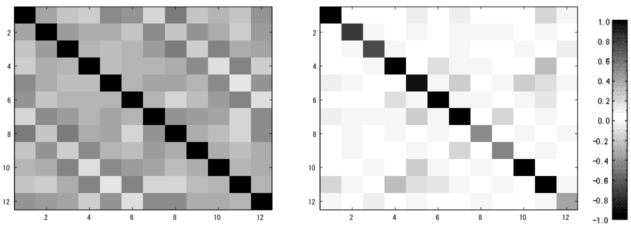
和音は、さまざまなオクターブに渡って演奏されたり、いくつかの転回形や開離形、 密集形など様々な音高配置で演奏されたりする。このような和音の音高配置によらない特徴量として、クロマベクトル [1] がある。クロマベクトルの $c(k,t)$ は、

$$c(k,t) = \log\left(\sum_{i=0}^{I-1} H(12i+k,t)\right), k = 0\cdots 11 \quad (1)$$

のようにパワースペクトルを半音ごとに複数オクターブ間で足し合わせることで得られる。ただし、$H(i,t)$ はスペクトログラムの周波数 bin i、時刻フレーム t でのパワー、I は取得するオクターブ数を表す。

2.2 フーリエ変換によるクロマの無相関化

クロマベクトルの各成分は倍音の影響や、複数のピッチが同時に演奏されることにより相関を持つ。そのため、クロマベクトルの共分散行列 Σ は Fig. 1 (a) のように対角要素以外にも 0 でない値を持つ。ここで、すべての音は同じ調波構造を持ち、同じ音程の組み合わせ (C と G, D と A など) が同じ頻度で出現するという仮定を置くと、Σ は巡回行列となる。この仮定は一曲のデータでは一般に成り立たないが、多くの曲のデータを用いることで平均的に成り立つと考えられる。実際に Σ は Fig. 1 (a) に示すようにほぼ巡回行列となることが分かる。巡回行列はその値によらず離散フーリエ変換行列により対角化されることが数学的に示されている [4]。従って、離散フーリエ変換行列を F とすると、クロマベクトル c の離散フーリエ変換 Fc を特徴量とすることにより、その共分散行列 $F\Sigma F^{-1}$ は Fig. 1 (b) のように確かにほぼ対角行列となる。この変換で得られた共分散行列を対角行列で近似することにより、共分散行列に必要な学習パラメータ数を 66 から 12 に削減することができる。



(a) chroma　　(b) FT of chroma

Fig. 1　ポップスの楽曲 207 曲における共分散行列: (a) クロマベクトル (b) クロマベクトルのフーリエ変換

2.3 スペクトログラムの主成分分析による特徴量

クロマベクトルには、オクターブの相異を圧縮するため和音の転回形などを認識できないことや、音域の情報が失われてしまうという問題点がある。スペクトルのオクターブ情報をなるべく圧縮せずに用いることで情報量を維持することができると考えられるが、高次元の特徴量となるため学習を十分に行うことができず、データを過学習してしまい認識率の低下を招く可能性がある。そこで、クロマのようにオクターブの圧縮を行うのではなく、主成分分析 (PCA) による次元圧縮を考える。PCA はデータを最小二乗誤差に基づき直交基底に分解する手法であり、データの共分散行列の固有値分解として定式化できる。得られた基底ベクトルを第一主成分から順に $\{a_1, a_2, \cdots, a_n\}$ とすると、そのうち $\{a_1, a_2, \cdots, a_m\}(m < n)$ にデータを射影することで、二乗誤差が最小となる m 次元の特徴量に圧縮することができる。

3 隠れマルコフモデルによる和音認識

和音進行を N-gram の確率過程として近似できると仮定し、特徴ベクトルは隠れ状態である和音から確率的に出力され、和音間は $N-1$ つ前までの和音に依存して確率的に遷移するという仮定を置くと、和音進行と特徴ベクトルの関係は隠れマルコフモデル (HMM) により表現できる [5, 6]。これにより、和音認識問題は特徴ベクトルから背後の和音を推定するという確率的逆問題として定式化できる。ここでは和音の移り変わりは直前の和音のみに依存するとして 2-gram でモデル化し、和音のエルゴディックな HMM を用いて定式化する。観測された特徴ベクトル系列を x、求める和音進行系列を c とおくと、最尤の和音

*Investigation of spectral features for chord recognition. by UEDA Yushi, UCHIYAMA Yuki, ONO Nobutaka, SAGAYAMA Shigeki (The University of Tokyo)

Table 1　クロマベクトルとクロマベクトルの FT 特徴量の対角共分散/全共分散における和音認識率

Features	chroma		FT of chroma	
covariance	diagonal	full	diagonal	full
	68.3%	77.4%	77.1%	77.4%

Table 2　クロマベクトルとスペクトログラムの PCA 特徴量における和音認識率

Chroma Vector	77.4%
Spectrogram(60 basis)	77.1%
PCA of Spec. 12 basis	73.2%
PCA of Spec. 20 basis	78.0%
PCA of Spec. 30 basis	77.9%
PCA of Spec. 40 basis	77.6%
PCA of Spec. 50 basis	77.4%

系列 c はベイズの定理により

$$\underset{c}{\operatorname{argmax}}\, p(c|x) = \underset{c}{\operatorname{argmax}}\, p(x|c)p(c) \tag{2}$$

$$\simeq \underset{c}{\operatorname{argmax}}\, p(x_0|c_0)p(c_0)\prod_{t=1}^{T} p(x_t|c_t)p(c_t|c_{t-1})$$

となり、和音毎の特徴ベクトルの出力確率 $p(x_t|c_t)$ と、和音間の遷移確率 $p(c_t|c_{t-1})$ をあらかじめ取得していれば Viterbi アルゴリズムにより効率的に計算することができる。また、和音進行系列の正解を持っていれば、Baum-Welch アルゴリズムにより、$p(x_t|c_t)$ と $p(c_t|c_{t-1})$ の局所最適値を学習することができる。連続的な観測の場合、出力確率の確率密度関数として混合ガウス分布を仮定することが多い。また、各和音に複数状態を割り当てることもできる。本研究では HMM の混合数・状態数は各和音につき 1 混合、1 状態とした。

4　和音認識実験

4.1　実験条件

提案するクロマベクトルのフーリエ変換特徴量、スペクトログラムの PCA 特徴量とクロマベクトルとの和音認識率の性能比較を行った。データセットとして The Beatles の楽曲 180 曲, Carole King の 7 曲, Queen の 20 曲の計 207 曲を用い、楽曲データはすべて 11 kHz サンプリング、モノラルの Wave ファイルであった。スペクトログラム及びクロマベクトルの生成には 55.0 Hz–1760 Hz の 5 オクターブを用いた。すべての特徴量には前処理として [3] と同様に調波音強調、チューニング補正を行った。PCA の基底数は 12、20、30、40、50、60(スペクトログラムそのもの) とした。学習・正解ラベルは [7] で提供されているものを用いた。和音の種類は、長三和音と短三和音の 24 種類としたが、このラベルは、長三和音、短三和音以外の和音も含むため、根音と第三音から、長三和音または短三和音とみなした。学習・認識はデータセットを 3 分し、2/3 を学習データ、残り 1/3 を認識データとして、3 分割交差検定により認識率を得た。認識率は 100 ms 毎にサンプリングされたフレームの (正解フレーム数)/(全フレーム数) により求めた。

4.2　実験結果

クロマベクトル、クロマベクトルのフーリエ変換を特徴量とした和音認識率として Table 1 を得た。全共分散の場合、クロマのフーリエ変換特徴量とクロマベクトルの認識率は厳密に一致した。これは、離散フーリエ変換は線形な可逆変換であるため、パラメータ数が同じ場合は情報量が等価であるためである。対角共分散の場合、全共分散の場合と比べて、クロマベクトルでは認識率が大きく低下したが、クロマベクトルをフーリエ変換することにより同等の認識率を得た。これは、クロマのフーリエ変換特徴量において、対角共分散による近似がほぼ成り立ち、情報量の損失が少なかったためと考えられる。

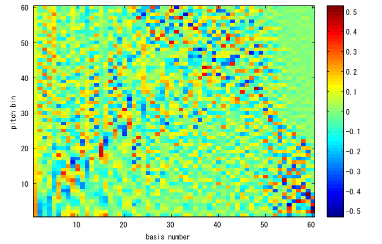


Fig. 2　主成分分析により得られた基底：左から第 1 主成分…第 60 主成分

クロマベクトル、スペクトログラムの PCA を特徴量とした和音認識率として Table 2 を得た。また、基底として Fig. 2 を得た。クロマベクトルの認識率と比較して、スペクトログラムの PCA 特徴量が 12 次元、60 次元とした場合を除いて認識率が上回っていることが分かる。また、60 次元のスペクトログラムと比較して、PCA による次元を削減により基底数が 12 の場合を除いて同程度の性能が得られ、基底数が 20 の場合認識率が最も上回ることが確認できた。これは、基底数を削減することにより学習に必要なパラメータ数が減り、データへの過学習を防ぐことが出来た一方、基底数を一定数以上削減すると和音認識に必要な情報が失われ、認識率が低下したと考えられる。

5　おわりに

本稿では和音特徴量としてクロマベクトルのフーリエ変換特徴量とスペクトログラムの PCA に基づく特徴量を実験的に検討した。クロマのフーリエ変換特徴量により共分散行列をほぼ対角行列として近似でき、全共分散の場合と同等の認識性能を得られること、スペクトログラムの PCA 特徴量で適切な基底数に圧縮すればクロマベクトルを上回る認識性能を得ることを確認した。

今後は異なるデータセットや和音の種類を増やした実験、HMM の混合数・状態数の検証を行う予定である。

謝辞　本研究の一部は、文部科学省科学研究費補助金基盤研究 (A) (課題番号 00303321) と科学技術振興機構 CrestMuse プロジェクトの支援を受けて行われた。

参考文献

[1] Fujishima, in *Proc. ICMC*, pp. 464–467, 1999.
[2] 内山他, 音講論 (春), pp. 901–902, 2008.
[3] 上田他, 情報処理学会研究報告, MUS-81, 2009.
[4] G. Golub and C. Van Loan, *Matrix Computations*, Johns Hopkins University Press, 1996.
[5] 川上他, 電気関係学会北陸支部大会講演論文集, F-61, p. 361, 1999.
[6] Sheh and Ellis, in *Proc. ISMIR*, pp. 183–189, 2003.
[7] http://isophonics.net/