Vol.2013-MUS-99 No.46
2013/5/12

圧縮やダウンサンプリングがクロマベクトルと和音認識に与える影響について

植村あい子[†1]　　石倉和将[†1]　甲藤二郎[†1]

本研究は，音楽情報処理で使用される特徴量クロマベクトルとそれを用いた和音認識について，楽曲の周波数特性の変化の影響を調査する．ダウンサンプリングや圧縮後のデータは周波数特性が変化していることから，音楽要素の認識やコンテンツベースな情報検索の結果に影響を及ぼすことが懸念される．そこで本稿では，ビットレートやサンプリングレートの異なる楽曲に関し，クロマベクトルの解析を行い，その品質について和音認識用データセットを用いて評価を行った．

Effects of Audio Compression and Downsampling for Chroma Vector and Chord Recognition

AIKO UEMURA[†1]　KAZUMASA ISHIKURA[†2]
JIRO KATTO[†3]

Feature analysis of music encoded with different bit rates and sampling rates is necessary to achieve high accuracy in musical content recognition and content-based music information retrieval (MIR). Bit rate and sampling frequency differences are expected to adversely affect musical content analysis and content-based MIR results because the frequency response might be changed by the encoding. In this paper, we specifically examine its effect on the chroma vector, which is a commonly used feature vector for music signal processing. We analyze chroma vectors extracted from encoded music files with different bit rates and compare them with chroma features of original songs obtained using datasets for chord recognition.

1. はじめに

音楽のコンテンツ配信携帯型音楽プレーヤやインターネット上でのストリーミング配信など，音楽に触れる機会の増加に伴い，MP3（MPEG Audio MPEG Audio Layer-3）や AAC （Advanced Audio Coding）などの様々な音声ファイルフォーマットが提供されるようになった．音楽情報処理では，これらのファイルを対象とするため，圧縮やダウンサンプリング後のデータは周波数特性が変形していることが，音楽要素の認識やコンテンツベースの情報検索の結果に影響を及ぼすことが懸念される．

筆者らは以前に異なるビットレートの楽曲から抽出された MFCC による楽曲検索への影響を調査し，非圧縮楽曲のみからなるデータセットとビットレートの異なる圧縮楽曲から構成されるデータセットを用いた場合では楽曲検索の結果が大きく異なることを示した[1].

同様に，音楽情報処理で用いられる特徴量クロマベクトルを用いた認識や検索結果も圧縮の影響を受ける可能性がある．ここで，図 1　異なるビットレートの楽曲から抽出されたクロマベクトル（The Beatles ”Let It Be”の冒頭 50 フレームに異なるビットレートの楽曲から算出したクロマベクトルを示す．これらはすべて同一楽曲であるがクロマベクトルのパワーの分布が異なっており，圧縮によるクロマベクトルへの影響は無視できない．

クロマベクトル[2]とは, 12 音名の各音名の周波数のパワーを複数のオクターブに渡って加算した 12 次元のベクトルで，和音認識に一般的に用いられる特徴量である．

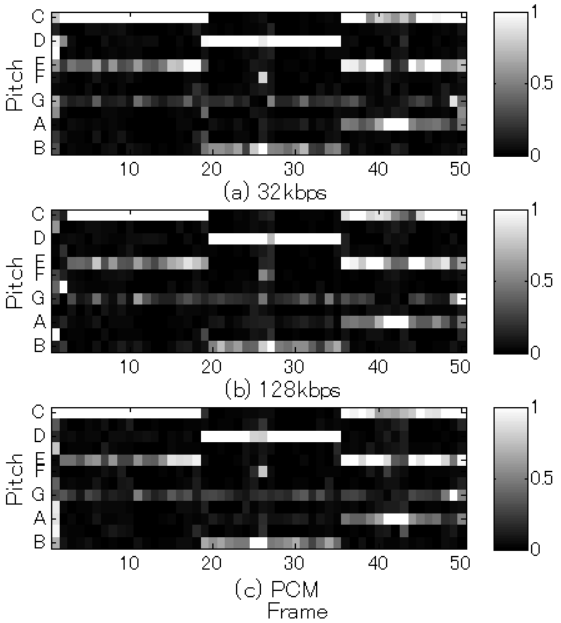

図 1　異なるビットレートの楽曲から抽出されたクロマベクトル（The Beatles ”Let It Be”の冒頭 50 フレーム）
Figure 1 Chroma vectors from different bit rate MP3 files: extracted from first 50 frames of “Let It Be” by The Beatles.

†1 早稲田大学大学院基幹理工学研究科
Graduate School of Fundamental Science and Engineering, Waseda University.

ⓒ 2013 Information Processing Society of Japan

クロマベクトルの算出方法は数多く提案されている[3-5]. Harte らはチューニングを行って定 Q スペクトルのパワーがほかの bin にまたぐのを防ぐ方法を提案した[3]．また，Ellis らは和音がビートに合わせて演奏されることに基づき，ビート同期を取ることで改善を行った[4].

表 1　各コーデックのエンコード条件

Table 1　Encoding settings.

コーデック	MP3	AAC	Ogg Vorbis
拡張子	mp3	m4a	ogg
エンコーダ	LAME [7]	NeroAACEnc [8]	oggenc2 [9]
ビットレート (kbps)	32 - 320	12 - 320	32 - 320

他にも，文献[5]では周波数スペクトルのフレームは理想的な音パターンの線形結合で表せるという仮定をし，NNLS（Non-Negative Least Square）問題を解くことで得られる NNLS chroma を提案している．さらに Müller らは，対数スケールと Discrete Cosine Transform（DCT）を用いた Chroma DCT-Reduced log Pitch（CRP） chroma を提案した [6].

本稿では，ビットレートとサンプリング周波数を変化させた楽曲からクロマベクトルを算出し，圧縮率とダウンサンプリングの影響について和音認識用データセットを用いてクロマベクトルの解析および和音認識性能の評価を行う．クロマベクトルの解析では，Peak signal-to-noise ratio（PSNR）を用いて PCM ファイルと圧縮後のファイルのクロマベクトルの劣化度を調査し，和音認識性能評価により周波数特性の変化が和音認識に及ぼす影響を明らかにする．

2. クロマベクトル解析

2.1 楽曲の圧縮とダウンサンプリング

本稿では表 1 のように MP3 （MPEG1 Audio Layer-3），AAC（Advanced Audio Coding），Ogg Vorbis の 3 種に圧縮を行う. ここで扱う PCM ファイルはサンプリング周波数 44.1 kHz，量子化ビット数 16 bit，2 ch の WAV フォーマットとし，すべて CBR（Constant Bit Rate）モードで圧縮を行う．MP3 ファイルへの圧縮には，LAME (ver. 3.99.5) [7] を，32 kbps から 320 kbps までの 14 種類のファイルを得た．AAC については，NeroAACEnc (ver. 1.5.1.0) [8] を用いて圧縮を行い，12 kbps から 320 kbps まで 16 種のファイルを抽出した．Ogg Vorbis では，Oggenc2.87 [9]を用いて 32 kbps から 320kbps までの 14 種類のファイルを抽出した. ただし, Ogg Vorbis では VBR（Variable Bit Rate）圧縮を基本としており，ビットレートは一定にならないため，マネジメントモードを利用し，最大・最小・平均ビットレートを同様にすることで CBR を実現した.

ダウンサンプリングにおいては，サンプリング周波数 44.1 kHz の PCM ファイルを，それぞれ 5.5 kHz から 32 kHz までダウンサンプリングを行い，結果として 7 種類のファイルを得た.

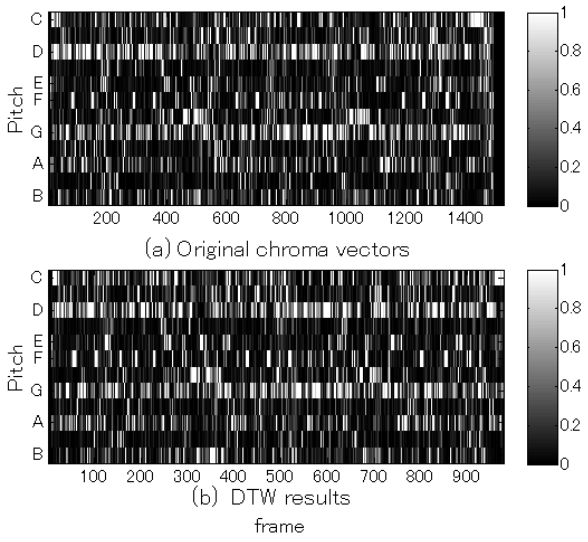

図 2　The Beatles ”Hard Days a Night” （MP3 ファイル 32 kbps）のクロマベクトルの DTW 結果

Figure 2 DTW results of 32 kbps MP3 files from “A Hard Day’s Night” by The Beatles.

2.2 クロマベクトル抽出

本稿のクロマベクトルは Ellis による手法[4]に基づき 12 次元のクロマベクトルを算出する．はじめに，ISP toolbox[10]を用いて細かいフレーム単位でクロマベクトルを算出する．このときクロマは，ウインドウ幅 93 ms，オーバーラップサイズ 75% で離散フーリエ変換を用いて算出される．最終的なクロマとして，文献[4]では推定されたビートごとに平均をとり，ビート同期を行ったクロマベクトルをするが, 本研究ではタイムスケールを合わせるため, 100 ms のウインドウでの平均をとることでクロマベクトルを算出した.

2.3 DTW

エンコードファイルは楽曲圧縮アルゴリズムの影響で，同一楽曲であってもサンプル数が変化する場合がある．その差は 0.1 秒程度であるが，算出されるクロマベクトルのフレーム数（時間方向）に違いが生じてしまう．そこで，ビットレートの異なるクロマベクトルの差異を評価するにあたり，我々は DTW（Dynamic Time Warping）を利用して 2 つのクロマベクトルのフレーム長を合わせるようにした．図 2 に DTW の結果の例を示す．DTW は，時間の異なる 2 つの信号シーケンスを柔軟に変化させて距離を算出するアルゴリズムである.

2.4 PSNR による評価

PSNR (Peak Signal-to-Noise Ratio)は主に画像信号の量子化歪みの評価に用いられるが，本稿では，PCM ファイルと圧縮ファイルのクロマベクトルの差異を評価するために

ⓒ 2013 Information Processing Society of Japan

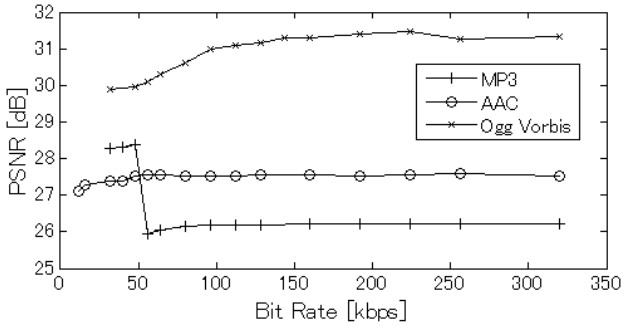

図 3　各ビットレートにおけるクロマベクトルの PSNR 評価結果

Figure 3 PSNR evaluation results of chroma vectors for 307 songs encoded at different bit rates.

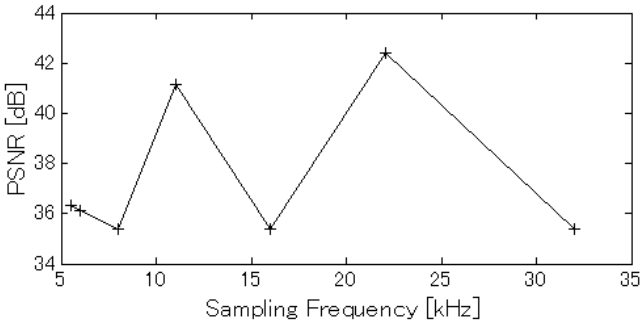

図 4　各サンプリング周波数におけるクロマベクトルの PSNR 評価結果

Figure 4 PSNR evaluation results of chroma vectors for 307 songs sampled at different frequencies.

PSNR を利用する．PSNR は次式のように定義される．

$$PSNR = 10\log_{10}\left(\frac{MAX^2}{MSE}\right) = 20\log_{10}\left(\frac{MAX}{\sqrt{MSE}}\right) \quad (1)$$

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left(C_{pcm}(i,j) - C_{encode}(i,j)\right)^2 \quad (2)$$

$C_{pcm}(i,j)$，$C_{encode}(i,j)$はそれぞれ PCM ファイルと圧縮ファイルから抽出したクロマベクトルであり，i，j は bin 数とフレーム数を表す．式(2)における MAX は，画像の場合は画素が取り得る 最大値であり，8 ビット画像の場合は 255 になるが，本稿ではクロマベクトルを扱うため，全楽曲のクロマパワーの最大値とした．

PSNR は値が高ければ高いほど，圧縮ファイルから抽出されたクロマベクトルは PCM ファイルから抽出されたクロマベクトルに類似しているということを表す．なお，画質評価において PSNR の値は，40 dB 程度は参照画像との差がわからず，30 dB 程度では多少のノイズがあり，20 dB 程度ではノイズが非常に多く，見るに堪えない画質とされている．

2.5 和音認識への適用

和音認識に対する周波数特性変化の影響を調査するため，ビットレートとダウンサンプリングした楽曲からそれぞれ抽出されたクロマベクトルを用いて和音認識を行う．ここでは認識には HMM を使用して認識を行う．モデルでは，major と minor の全 24 種の和音を扱い，学習データの数を増やすために根音の異なる同じ種類の和音においてクロマベクトルをシフトさせて学習を行う．ここでは，はじめに抽出されたクロマベクトルをすべて根音に基づき Cmajor と Cminor にシフトを行い，2 つの和音に対してモデルを算出する．その次に，Cmajor と Cminor のモデルをそれぞれ遷移させて 24 種類のモデルを得る．和音のクラス分類には，単一正規分布を用いて出力確率を算出し，forward-backward アルゴリズムで得られる各和音の出力確率の最大値を取り和音の識別を行う．

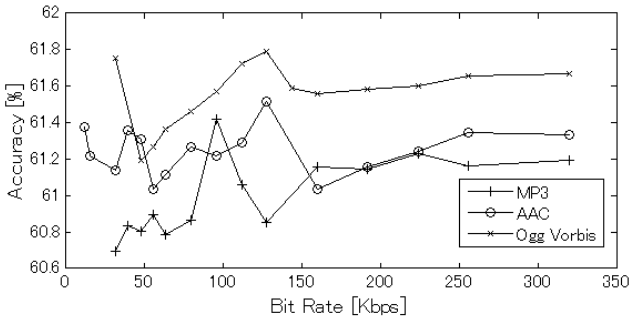

図 5　各コーデックにおけるビットレートごとの和音認識結果

Figure 5 Recognition-rate accuracy each chroma codec with different bit rate.

3. 評価実験

評価･学習データには The Beatles のアルバム曲（180 曲），Queen のアルバム曲（20 曲），C. King のアルバム曲（7 曲）と RWC 研究用音楽データベースよりポピュラー音楽（100 曲）の全 307 曲を用いた．正解ラベルには[]で提供されている正解データを使用する．本研究で扱う和音は major と minor の 24 種であるが，和音データには major，minor 以外の和音も含まれているため，人手によって根音と第 3 音により major と minor に分けた．例えば，Csus4 や Caug は Cmaj に，Cmin7 や Cdim は Cmin に分類した．

3.1 PSNR による評価

各コーデックにおけるビットレートごとの PSNR 平均を図 3 に示す．図 3 を見ると，コーデックごとに PSNR の値に差が出たものの，ビットレートでの変化はほとんど見られなかった．一方，サンプリング周波数が異なる場合の PSNR を図 4 に示す，ここでは，8 の倍数ごとに PSNR が下

ⓒ 2013 Information Processing Society of Japan

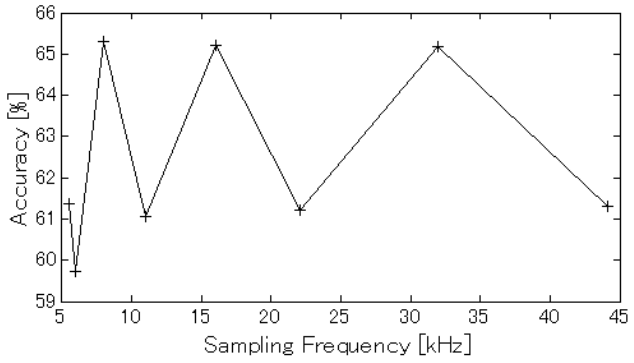

図 6　各サンプリング周波数における和音認識結果
Figure 6 Recognition-rate accuracy at different sampling frequencies.

がる傾向がみられた．

3.2 和音認識による性能評価

認識率はフレームごとの正解率とし，次式のように定めた．

$$\text{正解率} = \frac{\text{正しく出力されたフレーム数}}{\text{全フレーム数}} \qquad (3)$$

10 分割交差検定によるビットレートごとの和音認識結果を図 5 に示す．なお，このとき PCM ファイルによる和音認識結果は 61.5%であった．ここではすべてのビットレートとコーデックを含め最大でも 2%程度の差しか出なかった．一方で，サンプリング周波数を変化させた場合の和音認識結果を図 6 に示す．ビットレートが異なる場合では認識率への影響が小さかったことに対し，サンプリング周波数の変化では最大約 5%の差が生じた．また，8 の倍数ごとに認識率が上がる傾向がみられた．

4. 考察

圧縮において認識率が大きく変化しなかったのは，圧縮処理自体が有意な周波数を残す処理であることと，クロマベクトルがオクターブを足し合わせた抽象度の高い特徴量であることが原因であると考えられる．圧縮後であっても基本周波数が残っていれば，クロマにおける必要なパワーは維持されるといえよう．

また，ダウンサンプリングしたファイルに比べて圧縮ファイルのクロマベクトルの PSNR が全体的に低かったのは，圧縮によって有意な周波数以外が削られて，クロマにおいて点在していた細かなノイズが消えてしまったためと考えられる．

5. おわりに

本稿では，ビットレートとサンプリング周波数を変化させた楽曲からクロマベクトルを算出し，圧縮率とダウンサンプリングの影響について和音認識用データセット 307 曲を用いて評価を行った．圧縮では PCM ファイルからのクロマベクトルの劣化は見られたものの，和音認識率に関してはビットレートごとに大きな影響は見られなかった．また，サンプリング周波数の変化による分析では，8 の倍数において特徴がみられた．

今後はクロマベクトル算出方法による違いやフィルタ処理による周波数特性の変化がどのように影響を及ぼすのか調査していく予定である．

参考文献

1) S. Hamawaki et al.，“Feature Analysis and Normalization Approach for Robust Content-Based Music Retrieval to En-coded Audio with Different Bit Rates,” 15th Intl. Multimedia Modeling Conference，Jan. 2009.
2) T. Fujishima，“Real-time Chord Recognition of Musical Sound: a System using Common Lisp Music,” Proc. ICMC，pp. 464-467，Oct. 1999.
3) C. Harte and M. Sandler，“Automatic Chord Identification using a Quantised Chromagram,” Proc. Audio Eng. Soc.，Spain，May 2005.
4) D. P. W. Ellis and G. Poliner，“Identifying Cover Songs with Chroma Features and Dynamic Programming Beat Tracking,” Proc. ICASSP，pages IV 1429-1432，2007.
5) M. Mauch and S. Dixon，“Approximate Note Transcription for the Improved Identification of Difficult Chords,” Proc. ISMIR，Aug. 2010.
6) M. Müller and S. Ewert，“Towards Timbre-invariant Audio Features for Harmony-based Music,” IEEE Trans. on Audio，Speech，and Language Processing，vol. 18，no. 3，pp. 649-662，2010.
7) LAME MP3 Encoder
http://lame.sourceforge.net
8) Nero AAC Codec
http://www.nero.com/enu/company/about-nero/nero-aac-codec.php
9) RAREWARES – oggenc2
http://www.rarewares.org/ogg-oggenc.php
10) Intelligent Sound Processing
http://kom.aau.dk/project/isound/
11) Chroma Toolbox: Pitch，Chroma，CENS，CRP
http://www.mpi-inf.mpg.de/resources/MIR/chromatoolbox/
12) Isophonics
http://isophonics.net/
13) M. Goto，“AIST Annotation for the RWC Music Database,” Proc. the International Conference on Music Information Retrieval (ISMIR 2006), pp.359-360，Oct. 2006.

ⓒ 2013 Information Processing Society of Japan