

異なるビットレートの楽曲を用いたクロマベクトルの解析

植村 あい子[†] 石倉 和将[†] 甲藤 二郎[†]早稲田大学大学院 基幹理工学研究科[†]

1 はじめに

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、楽曲の圧縮の影響を調査する。圧縮後のデータは周波数特性が変形していることから、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に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。

そこで本稿では、ビットレートを変化させた楽曲のクロマベクトルを算出し、圧縮率の影響について和音認識用データセット 207 曲を用いて評価を行った。

2 楽曲のクロマベクトル解析

2.1 クロマベクトル抽出

2.1.1 楽曲圧縮

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 44.1kHz, 量子化ビット数 16bit, 2ch の WAV フォーマットとし、すべて CBR (Constant Bit Rate) モードを用いて圧縮する。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、平均として CBR を実現した。

2.1.2 クロマベクトル

クロマベクトルは、パワースペクトルをオクターブごとに吸収したもので、和音認識で一般的に用いられている特徴量である。本稿のクロマベクトルは Ellis による手法[2]に基づきビート同期した 12bin クロマベクトルを算出する。クロマベクトルの算出には ISP toolbox[3]を用いて、推定されたビートに基づき同期を行う。

2.2 クロマベクトル解析

2.2.1 DTW (Dynamic Time Warping)

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。さらに、クロマベクトルはそれぞれ推定されたビートに基づき同期処理を行っているため、推定値によって数フレームの差が生じることがある。そこで、ビットレートの異

表 1 各コーデックのエンコード条件

コーデック	MP3	AAC	Ogg Vorbis
拡張子	mp3	m4a	ogg
エンコーダ	LAME	NeroAACEnc	oggenc2
ビットレート (kbps)	32~320	12~320	32~320
種類	14	16	14

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。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。

2.2.2 PSNR (Peak Signal-to-Noise Ratio)

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、圧縮によるクロマベクトルの劣化度を調べるため 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 PSNR を利用する。PSNR は次式のように定義される。

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

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

$C_{\text{pcm}}(i, j)$, $C_{\text{encode}}(i, j)$ 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、実験的に 3.2 と定めた。

2.2.3 和音認識

和音認識に対する楽曲圧縮の影響を調査するため、クロマベクトルのみを使用し和音認識を行う。認識には HMM を使用し、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った。また認識時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。

評価・学習データには The Beatles(180 曲), Queen(20 曲), C. King(7 曲)の 207 曲と[4]で提供されている正解ラベルを使用する。なお、本稿では major と minor の 24 和音を扱うこととし、提供されているラベルは人手で 24 和音に分類した。

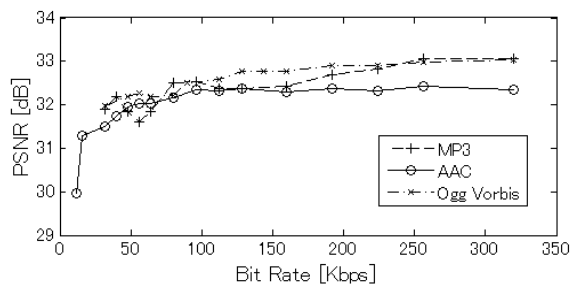


図1 各ビットレートのPSNRの平均値

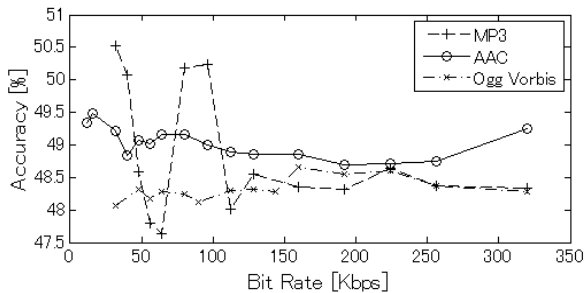


図2 各ビットレートの和音認識結果

2.2.4 正解周波数からのずれ

算出されたクロマが正解周波数からずれている場合、認識率は低くなると推測される。そこで、和音認識用テンプレート[5]を参考に、構成音の6倍音まで考慮したものを正解と仮定し、正解と算出されたクロマの比較を行う。ここでは正解ラベルがわかっていることから、比較対象の楽曲のビート推定値に対して、そのフレームに相当する正解ラベルに基づき生成する。例えば Cmajor や Cminor は図のようなパワーを持つことになる。なお、ここでは正解ラベルに含まれる和音全種(major, minor, dominant seventh, diminished, augmented, etc.)のテンプレートを扱った。周波数のずれはテンプレートと楽曲の各フレームについてユークリッド距離を求め、1曲ごとに全フレームの平均値を算出する。

3 評価実験

3.1 PSNR

各コーデックにおけるビットレートごとの PSNR 平均を図1に示す。図1を見ると、Ogg Vorbis は 112kbps 以上が推奨標準であることから、100k を超えたあたりから PCM ファイルのクロマ成分に近づいていくことは妥当であるといえる。AAC と MP3 は 100kbps 以下において劣化がみられ、低レート圧縮では認識精度に影響を及ぼす可能性があることがわかる。

3.1 和音認識結果

9 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2に示す。なお、PCM ファイルによる和音認識結果は 48.5%であった。図2においてすべてのビットレートとコーデックを含め最大でも 3%程度の差しか出なかった。また、PCM ファイルのクロマが圧縮ファイルのクロマと比べて認識率が低いのは、もともとの信号に高周波ノイズが多く含まれていることが原因として考えられる。

3.2 正解周波数からのずれ

各ビットレートにおけるユークリッド距離 207 曲の平均を図3に示す。なお、PCM ファイルの距離は 0.49 であった。図2,3において算出されたクロマが正解周波数

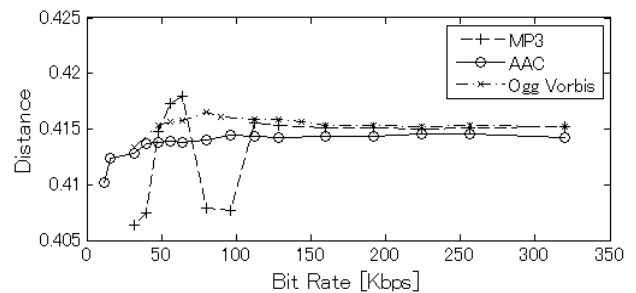


図3 各ビットレートの正解周波数からのずれ

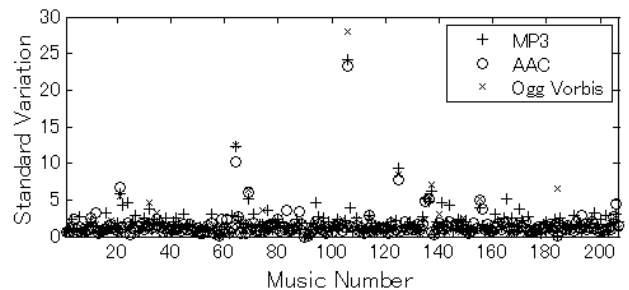


図4 楽曲ごとの認識率の標準偏差

から離れている場合、認識率が下がることが確認できる。AAC は低ビットレートにおいて PCM データから劣化しているものの、正解周波数はレートごとの大きな変化はない。このことから AAC は幅広いビットレートに対して圧縮の影響は小さく、有意な音が得られているといえる。

4 考察

図2において和音認識結果にビットレートの強い影響は見られなかったが、楽曲によっては圧縮の影響を受ける可能性がある。図4ではいくつかの楽曲では標準偏差が大きく、圧縮により認識率が大きく変化している様子がわかる。一方、多くの楽曲は和音認識に対して圧縮の影響は小さいことが確認できる。これはオーディオ圧縮が本質的には有意な周波数を残す作業であり、クロマの算出処理に類似していることから、かなりビットレートを下げても影響が小さいためだと考えられる。また、クロマベクトル抽出法は数多く提案されており、手順の違いがクロマベクトルの差異を生じさせる可能性があるため、手法による違いの調査も必要である。

5 おわりに

本研究は、ビットレートを変化させた楽曲を用いて、クロマベクトルの PSNR 評価、和音認識と正解周波数からのずれについて解析を行った。今後はクロマベクトル抽出法による違いや周波数スペクトルベースでの比較について検討を行っていきたい。

参考文献

- [1] S. Hamawaki et al.: "Feature Analysis and Normalization Approach for Robust Content-Based Music Retrieval to Encoded Audio with Different Bit Rates," 15th International Multimedia Modeling Conference, Jan.2009..
- [2] 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.
- [3] Intelligent Sound Processing: <http://kom.aau.dk/project/isound/>.
- [4] isophonics: <http://isophonics.net/>
- [5] L. Oudre, et al.: "Template-Based Chord Recognition : Influence of the Chord Types," Proc. ISMIR, pp. 153-158, Oct.2009.