1P－7

ビットレートの異なるTwinVQオーディオデータの類似曲検索のための特徴量

墳崎 英明　小早川 倫広　大西 建輔　星 守　大森匡
電気通信大学 大学院 情報システム学研究科*

1　研究の背景と目的

近年，音楽配信が注目されており，楽音圧縮技術を用いた配信システムが普及しつつある．そのような音楽データベースでは，楽音圧縮技術と検索技術が重要である．本来，圧縮技術と検索技術とは別々に研究されているが，圧縮データが多用されている現在，この2つの技術を組み合わせて考えることが必要である．そこで，本研究ではMPEG-4[1]に採用されているTwinVQ[2]圧縮技術を用いた楽音データを用い，曲の一部を検索キーとした類似曲検索システムを考える．

このようなシステムを構築するには，効率的なデータ管理，検索を行うために，圧縮データから直接索引を生成することが望ましい．また，TwinVQデータはビットレートを変えることで，品質，圧縮率，通信量など使用目的に合わせて利用するため，ビットレートが異なっていても検索可能であることも重要である．本稿では，以上の2点を実現可能な検索のための特徴量を提案する．

2　TwinVQの概要

TwinVQ圧縮データはフレーム単位でビットストリームを形成し，各種補助情報(LSP，pitch，Bark scale envelope，power)と平坦化されたMDCT係数(flattened MDCT coefficient)の5つから構成される(図1)．線形

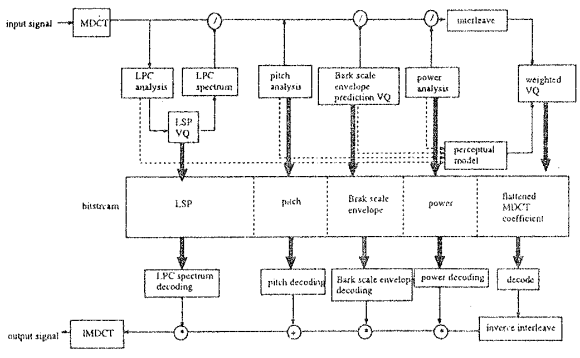

図1: MPEG-4で使用されるTwinVQのブロック図

予測分析部(LPC analysis)ではMDCT係数を入力とし，LPCスペクトルを求めている(図2)．LPCスペクトルは，MDCT係数の概形を表現し，MDCT係数を平坦化する目的で計算される．つまり，LPCスペクトルは曲の特徴を簡単に表現しているといえる．本稿では，線形予測分析部に着目し，曲検索の索引となる特徴量を検討する．

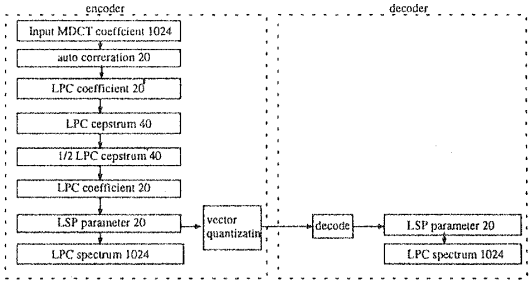

図2: 線形予測分析部のエンコードとデコード処理

3　特徴量の検討と算出法

3.1　類似曲検索のための特徴量

先に述べた曲検索システムを実現するために，線形予測分析部より抽出可能な特徴量を考える．元信号に対しMDCTをかけて得られるMDCT係数は，低周波成分に比べ高周波成分は非常に小さく，ビットレートに応じて高周波成分を取り除き，データ量を削減している．すなわち，ビットレートはMDCT係数の低周波成分の使用数に対応する．使用する係数列に対し，0を挿入することで$f(=1024)$次元にし，MDCT係数を平均的に引き伸ばす．この係数列を入力として，以下の式により自己相関係数r_iを求める．

$$r_i = \sum_{k=0}^{f} (MDCT\ 係数_k)^2 cos(i\theta) \quad (i = 1, \cdots, p)$$

ここで，r_iは自己相関係数，fは次元数とする．線形予測分析部の入力が高周波成分を削除し(ビットレートによる)，0を挿入していることから，上式で変換して求めた，ビットレートの異なる$p(=20)$次元の自己相関係数列をビットレート比で伸縮するとほぼ一致することがわかる．すなわち，ビットレートB_1のときの自己相関係数を$r = (r_1, \cdots, r_p)$，ビットレートB_2のときの自己相関係数を$r' = (r'_1, \cdots, r'_p)$とし，$B_1 < B_2$ならば，$j = \left[i\frac{B_2}{B_1}\right] \leq p$を満たす最大の$i$を$k$とすると，

$$r_i \sim r'_j \quad (i = 1, \cdots, k) \qquad (1)$$

*Feature for similarity retrieval of TwinVQ audio data with different bitrate , H.Tsukazaki M.Kobayakawa K.Onishi M.Hoshi T.Ohmori (U.Electro-Comm.)

という関係が成り立つ．

3.2 特徴量抽出のための予備実験

1フレームの自己相関係数 ($B_1 = 16kbit/s/ch$, $B_2 = 32kbit/s/ch$) をプロットしたものを図3に示す．図3を

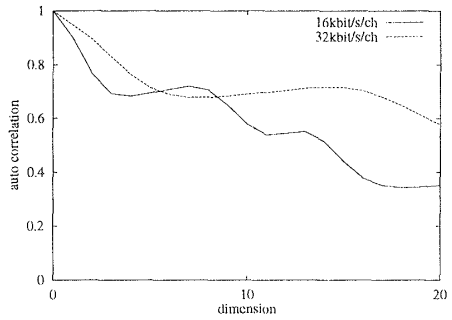

図 3: 1フレームにおける自己相関係数

みると，3.1節で示した関係 (1) が成り立っていることがわかる．そこで，楽曲全体に対しても同様に成立するか確認するため [3] に示された時系列に対する手法を適用し，実験を行った．サンプリング周波数44.1kHz，量子化ビット数16bit，チャンネル数2channelを入力としたときの，ビットレート $B_1 = 16kbit/s/ch$ と $B_2 = 32kbit/s/ch$ の自己相関係数を計算する．このとき，$i = 1, 2, \cdots, 10$, $j = 2, 4, \cdots, 20$ である．実験条件としてはDCTをかける窓幅を512フレームとして1フレームずつずらして行う．図4は，曲約1分 (約5000フレーム) に対する1次と2次のDCT係数の特徴空間上の系列を表示した．図

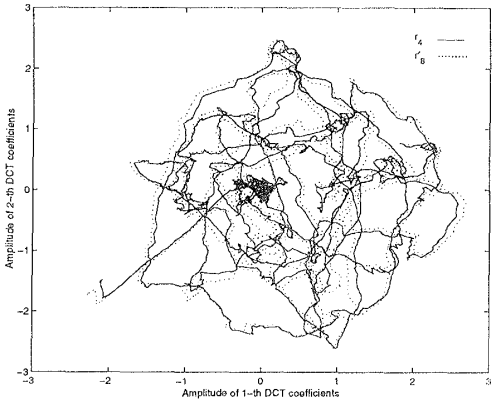

図 4: 曲1分に対する r_4, r'_8 の1次と2次のDCT係数

4をみると類似した軌跡を描いていることがわかる．このことから，自己相関係数の時系列に対しても関係 (1) が成り立っているといえる．

3.3 自己相関係数の算出方法

ここで，自己相関係数の算出法を考える．自己相関係数は線形予測分析部のエンコードの際に抽出することができる (図2)．また，逆にTwinVQ圧縮データからも計算可能である．圧縮データよりLSPパラメータを抽出し，これを順次図2に示したエンコードの逆変換を行うことで自己相関係数を計算する．LSPパラメータと線形予測パラメータには以下の関係が成立している [4]．

$$\begin{aligned} F(z) &= 1 + \alpha_1 z^{-1} + \ldots + \alpha_p z^{-p} \\ &= \frac{1}{2}(P(z) + Q(z)) \end{aligned}$$

ここで，

$$P(z) = (1 - z^{-1}) \prod_{i=2,4,\ldots,p} (1 - 2z^{-1}\cos\omega_i + z^{-2}),$$

$$Q(z) = (1 - z^{-1}) \prod_{i=1,3,\ldots,p-1} (1 - 2z^{-1}\cos\omega_i + z^{-2})$$

ここで α_i は線形予測係数，ω_i はLSPパラメータである．

求めた線形予測係数をLPCケプストラム係数に変換し，その係数列を2倍して線形予測係数に変換する．さらに以下の式より自己相関係数を計算する．ここで α_i は線形予測係数，r_i は自己相関係数，p は次元数を表す．ここで $\alpha_0 = r_0 = 1.0$, $R = (r_1, \cdots, r_p)^T$, $A = (\alpha_1, \cdots, \alpha_p)^T$ とする．

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{p-1} \\ r_1 & r_0 & & r_{p-2} \\ \vdots & & \ddots & \vdots \\ r_{p-1} & r_{p-2} & \cdots & r_0 \end{pmatrix} A = -R$$

これを変形すると，次式が得られる．

$$\begin{pmatrix} \alpha_0 + \alpha_2 & \alpha_3 & \cdots & \alpha_p & 0 \\ \alpha_1 + \alpha_3 & \alpha_0 + \alpha_4 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \alpha_{p-2} + \alpha_p & \alpha_{p-3} & & \alpha_0 & 0 \\ \alpha_{p-1} & \alpha_{p-2} & \cdots & & \alpha_0 \end{pmatrix} R = -A$$

この連立方程式を解いて自己相関係数 r_i を求める．

4 まとめ

本稿では，本研究の目的とする検索システムを構築するために，自己相関係数を使用することが有効であることを示した．圧縮データから自己相関係数を算出し，索引を生成したときの検索精度，大規模データに対する有効性については今後報告したい．

参考文献

[1] ISO/IEC JTC 1/SC 29/WG11 N2203. "Working Draft of ISO/IEC CD 14496-3". 5 1998.

[2] Naoki Iwakami, Takehiro Moriya, and Satoshi Miki. "High-quality audio-coding at less than 64kbits/s by using transform-domain weighted interleave vector quantization". In *Proc.IEEE ICASSP '95*, pages 3095–3098, May 1995.

[3] Christos Faloutsos, M.Ranganathan, and Yannis Manolopoulous. "Fast Subsequence Matching in Time-Series Databases". In *ACM SIGMOD 94*, pages 419–429, 5 1994.

[4] 守谷 健弘. "音声符号化". 電子情報通信学会, 1998.