楽曲類似検索における特徴量抽出の高速化

青木　圭子　　神田　龍一　　帆足　啓一郎　　柳原　広昌
KDDI研究所

1．はじめに

一般に楽曲の類似検索においては，楽曲ファイルの音響的特徴量を量子化し，コサイン距離等でその類似度を測定する手法が用いられている．中でも音声認識のモデル学習で知られるMFCC(Mel-Frequency Cepstrum Coefficient)が特徴量として広く用いられている[1]．

一方，昨今の音響圧縮技術の発達により，楽曲データは，MP3(Mpeg-1 Audio Layer-3)やAAC(Advanced Audio Coding)等の圧縮形式で保存することが多くなった．MFCCを求める場合には，これらの圧縮ファイルを一旦非圧縮のPCM形式にデコードした後，特徴量の計算を行う必要がある．そのため，楽曲の類似検索においては，特徴量抽出過程がシステム構築にかかる時間の大半を占めている．

そこで本稿では，携帯端末を対象とした音楽配信サービスにおいて主流となりつつある，HE-AAC(High-Efficiency Advanced Audio Coding)のファイルから直接MFCCに相当する特徴量（AACCEP）抽出を行う手法を提案し，他の手法との比較も含め，実装・評価を行った．

2．MP3CEP

MP3データから特徴量抽出を行う手法として，MP3CEP[2]がある．MP3CEPはMP3データをフィルタバンク出力部分までデコードし，その各サブバンドデータに離散コサイン変換(DCT)を行うことで特徴量算出を行う手法である．MP3符号化では一旦時間領域のフィルタで32サブバンドに分割した後にMDCTを行うのに対し，HE-AACでは入力サンプルに直接MDCTが行われるため，本方法は適用できない．

3．提案手法

MFCCに近い特徴量を抽出するため，まずMFCC抽出におけるFFTスペクトルの代わりに，AACのデコード過程（図1）で得られる1024の帯域における MDCT係数を，メル周波数軸上における12のメルフィルタバンクに写像する（図2）．

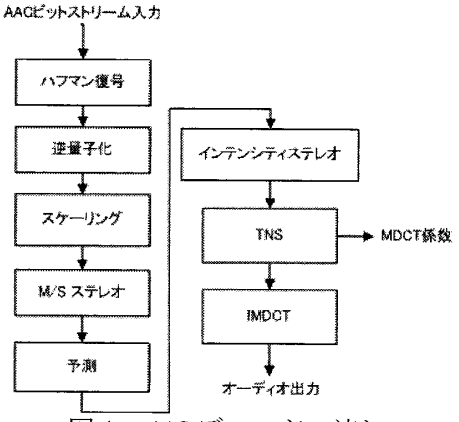

図1　AACデコードの流れ

図2　MDCT係数の写像の例

次に帯域フィルタバンク出力として，各メルフィルタバンク内に含まれるMDCT係数にMFCCと同様の窓関数を掛けてメルフィルタバンク毎に加算する．加算されたMDCT係数に対して対数コサイン変換を行った結果を特徴量（AACCEP）とする．

尚，本方式ではMDCT係数を使用するため，HE-AACのSBR部分については考慮していない．

4．実験及び考察

特徴量抽出速度及びその特徴量を用いた検索精度を検証するため，MP3形式，HE-AAC形式の楽曲ファイルを用意し，MFCC，MP3CEP，AACCEPそれぞれについて，測定を行った．実験に使用した環境は表1の通りである．

表1　実行環境

マシン	Endeavor MT7800
CPU	Core2 Duo E6700 2.66GHz
OS	Vine Linux 4.1
メモリ	3GB

A Fast Method of feature extraction for Music Retrieval,
Keiko Aoki, Ryuichi Kanda, Keiichiro Hoashi, Hiromasa Yanagihara,
KDDI R&D Laboratories Inc.

4－1．特徴量抽出時間

ランダムに抽出した 100 曲の楽曲データについて，提案方式での特徴量抽出時間の測定を行った．比較データとして(1)HE-AAC 形式から PCM 形式に変換した後，MFCC を抽出する時間 (2)HE-AAC 形式から直接 AACCEP を抽出する時間を測定した．参考までに (3)MP3 形式から直接 MP3CEP を抽出する時間についても測定した（図 3）．

その結果，AACCEP は MFCC に対して約 3.3%の処理時間で実行できることが分かった．AACCEP では IMDCT の処理が省けるのに対し，MP3CEP では IMDCT 処理が必要となるため，MP3CEP では処理時間に大きな改善は得られなかった．今回の測定ではファイル I/O の時間を考慮していないため，実際にはさらに実行時間が短縮される．

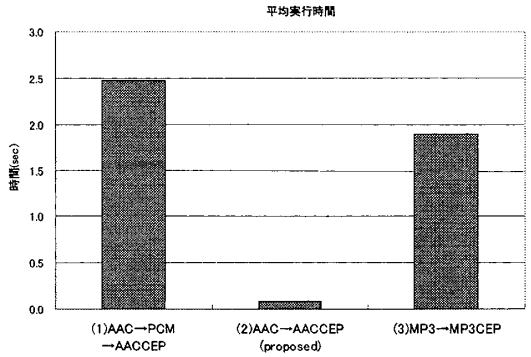

図 3　平均実行時間

4－2．検索精度

検索精度を測定するため，各特徴量に対してツリーベクトル量子化処理(TreeQ)を行い（図 4），特徴空間を作成した．

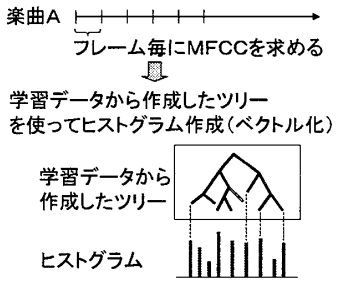

図 4　ベクトル量子化の流れ

検索では，[1]と同様に本特徴空間におけるコサイン距離によって類似度を測定した．ランダムに抽出した 100 曲のデータについて，1 曲のクエリーに対する MFCC, MP3CEP, AACCEP のそれぞれを用いた特徴空間における距離を 5 回測定してその平均値を算出し，距離の近い順にグラフ上にプロットした（図 5）．検索においてはコサイン距離の順序関係が重要であるため，MFCC を正解データと仮定して，MFCC と同様に単調増加となることが理想である．そこで MP3CEP, AACCEP と MFCC との相関係数を測定した（表 2）．

その結果，AACCEP の方が MFCC に近い相関の得られることが分かった．MP3CEP では等間隔に分割されたサブバンドデータを利用するのに対し，AACCEP ではメルフィルタバンクを基準として，窓関数を用いた抽出を行っていることに効果があったと思われる．

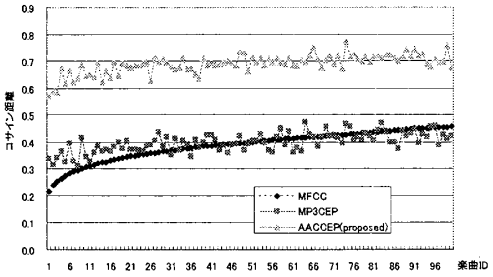

図 5　クエリー楽曲との距離

表 2　MFCC との相関係数

方式	相関係数
MP3CEP	0.68
AACCEP	0.76

5．おわりに

本稿では，楽曲データの類似検索における特徴量抽出処理を高速化するため，HE-AAC の MDCT 係数を MFCC に写像し，圧縮データから直接特徴量を求める手法を提案した．従来の PCM 形式を経由した特徴量抽出手法に対し，処理時間を約 3.3%に高速化し，検索精度も MFCC に近いものが得られた．

一部実施した主観評価では，まだ検索精度で MFCC に及ばないケースもあるため，今後は高域強調処理を加えた改善方式や HE-AAC 形式の SBR 部分を考慮した処理を検討予定である．

参考文献

[1] Keiichiro Hoashi et al., “FEATURE SPACE MODIFICATION FOR CONTENT-BASED MUSIC RETRIEVAL BASED ON USER PREFERENCES”, pp. 517-520, ICASSP 2006.

[2] David Pye, “Content-Based Methods for the Management of Digital Music”, pp. 2437-2440 vol. 4, ICASSP 2000.