修士論文要旨 (2013 年度)

冗長ウェーブレット変換を用いた楽曲特徴量
Music Features using Undecimated Wavelet Transform

11N5100023I 鈴木雄亮

電気電子情報通信工学専攻　久保田研究室

1. 研究背景と目的

近年, 音楽を取り巻く環境の変化により, 大量の楽曲を容易に入手, 保持できるようになったが, ユーザが所望の楽曲を見つけ出す負担もまた大きくなった. こうした背景よりコンピュータが音楽を自動的に理解し , ユーザの嗜好に合わせて楽曲を推薦するシステムが注目されている. 実装例では Pandora[1] などがある. 本稿では音楽理解のために冗長ウェーブレット変換を用いた楽曲特徴量の抽出手法について述べる.

2. 関連技術

本研究では, 楽曲信号に多重解像度解析を行う. 多重解像度解析は, 複数のバンドパスフィルタから構成されるフィルタバンクによって, 入力信号を個々の周波数成分に分解する手法である. ここでは代表的な多重解像度解析法である離散ウェーブレット変換とその亜種である冗長ウェーブレット変換について説明する.

2.1 離散ウェーブレット変換

離散ウェーブレット変換（DWT）は, ウェーブレットと呼ばれるさざ波の形をした基底信号を拡大縮小と平行移動によって適用することで, 入力信号を表現しようとする時間-周波数解析法である. FFT 同様に変換前の信号におけるデータサイズは保持され, また逆変換を行うことで元信号を復元できる. 図 1 に入力信号 $X[n]$ に対する解析手順を示す.

直交ミラーフィルタであるハイパスフィルタ h とローパスフィルタ g による畳み込み演算後, 解像度を 1/2 に落とすダウンサンプリングを行う. ハイパスフィルタ h はウェーブレットの母関数でもあり, 基底信号の波形を決定する. またダウンサンプリング後のローパス信号を

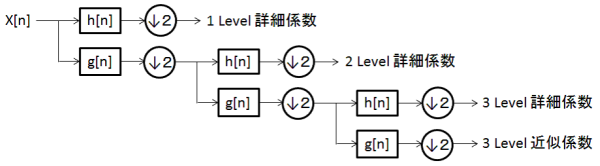

図 1: 離散ウェーブレット変換

近似係数, ハイパス成分を詳細係数呼び, この近似係数について直交ミラーフィルタリングとダウンサンプリングを繰り返す中で現れる詳細係数を個々のバンド信号として分解することができる. この際, ダウンサンプリングによってバンド信号においての基底の周期は, 実質的には分解レベルごとに 2 倍ずつ大きくなり, オクターブの周波数分解能を持つことになる. これは人間の聴覚特性とも合致する.

2.2 冗長ウェーブレット変換

DWT はオクターブ分解能を持つ時間-周波数解析法であるが, ダウンサンプリングにより, 各バンドごとに時間分解能が異なる性質を持つ. この性質はバンド信号間の比較計算で不便になるので, 冗長ウェーブレット変換（UWT）を導入する. UWT は DWT の時間分解能を元信号の時間分解能に揃えることを目的をする. 図 2 に解析手順を示す.

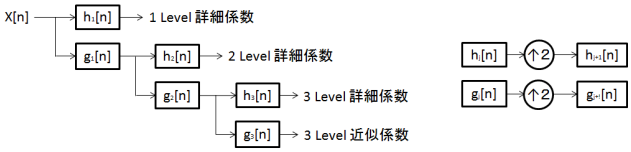

図 2: 冗長ウェーブレット変換

DWT では信号をダウンサンプリングさせることでフィルタのサイズを変えずにウェーブレットの拡大を表現し

たのに対して, UWT ではフィルタをアップサンプリングすることで, ウェーブレットの拡大を表現していることである. これにより変換後のデータサイズは変換前の分解レベル倍に大きくなり, 冗長性を持つと同時に各バンドごとに時間分解能を揃えることができる. また UWT の各バンド係数をダウンサンプリングさせることで,DWT の各係数に変換することもできるため逆変換も可能になる.

3. 提案手法

本研究では音楽信号を冗長ウェーブレット変換を用いた多重解像度解析を行い, 複数のバンド信号成分から楽曲特徴量を抽出する. 各バンド信号を絶対値処理後, 短時間フレームごとに始端から終端までシフトさせながら, フレーム数個の局所特徴量を導出し, さらに全フレームの局所特徴量から最終的な特徴量を求める. 図 3 に全体の流れを示す.

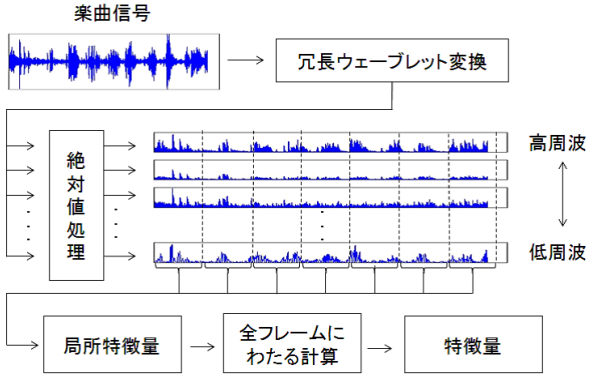

図 3: 提案手法のフロー

局所特徴量には各バンド信号成分の大きさや変化の激しさなどの各バンド信号の固有の特徴量と, 各バンド信号間における相関などのバンド信号どうしの比較による特徴量の 2 パターンを用意した.

3.1 前処理

ここから楽曲特徴量の抽出法について説明する. まず信号全体に微小のホワイトノイズを付加した. これは入力信号の大きさを除算に用いることがあり, その際に起こりうる 0 の除算を防ぐために導入した. 続いて楽曲信号の正規化を行う. 楽曲信号を信号成分の大きさで除算することで, 楽曲ごとの音量のバラつきを抑えることができる. N 次元の入力信号ベクトル $\vec{x}$ について, 要素の絶対値平均で各要素を除算する.

$$DC = \frac{1}{N}\sum_{i=1}^{N}|x(i)| \tag{1}$$

$$\vec{x_{reg}} = \frac{1}{DC}\vec{x} \tag{2}$$

N は入力信号 $\vec{x}$ のサンプル数,DC を音量に関する特徴量とする. この $\vec{x_{reg}}$ に対して UWT を行う.

$$\boldsymbol{U} = \begin{pmatrix} \vec{U_1} \\ \vdots \\ \vec{U_L} \end{pmatrix} = \begin{pmatrix} U_1(1) & \dots & U_1(1) \\ \vdots & \ddots & \vdots \\ U_L(1) & \dots & U_L(N) \end{pmatrix} \tag{3}$$

L は分解バンド数,$\vec{U}$ は各バンド信号,U はウェーブレット係数である. 本研究ではウェーブレット係数をすべて絶対値処理した値を使用する.

3.2 局所特徴量

多重解像度信号 $\boldsymbol{U}$ から短時間フレームごとに局所特徴量を求める. フレーム数を M としてフレーム番号 j, バンド番号 i のバンド信号を $\vec{U_i^j}$ のように切り出す.

3.2.1 平均 $\boldsymbol{AV}$

平均の局所特徴量を配列 $\boldsymbol{AV}$ に格納する.

$$\boldsymbol{AV} = \begin{pmatrix} A\vec{V}_1 \\ \vdots \\ A\vec{V}_L \end{pmatrix} = \begin{pmatrix} AV(1,1) & \dots & AV(1,M) \\ \vdots & \ddots & \vdots \\ AV(L,1) & \dots & AV(L,M) \end{pmatrix} \tag{4}$$

$$AV(i,j) = \frac{1}{F}\sum_{k=1}^{F} U_i^j(k) \tag{5}$$

$$(F = NM^{-1} \quad i \le L, j \le M \quad i,j \in \mathbb{N}_+)$$

$\mathbb{N}_+$ は 0 を含まない自然数である. 絶対値処理された係数の平均は, そのバンド信号の大きさを意味する.

3.2.2 変動係数 $\boldsymbol{CV}$

変動係数の局所特徴量を配列 $\boldsymbol{CV}$ に格納する.

$$\boldsymbol{CV} = \begin{pmatrix} C\vec{V}_1 \\ \vdots \\ C\vec{V}_L \end{pmatrix} = \begin{pmatrix} CV(1,1) & \dots & CV(1,M) \\ \vdots & \ddots & \vdots \\ CV(L,1) & \dots & CV(L,M) \end{pmatrix} \quad (6)$$

$$CV(i,j) = \frac{1}{F\dot{A}V(i,j)} \sqrt{\sum_{k=1}^{F} \left(U_i^j(k) - AV(i,j) \right)^2} \quad (7)$$
$$(F = NM^{-1} \quad i \leq L, j \leq M \quad i,j \in \mathbb{N}_+)$$

変動係数は標準偏差に対して平均を除算したものである. 成分のバラつき具合である標準偏差に対して, 成分の大きさである平均を除算することで AV と相関のない成分を抽出できる.

3.2.3 相関係数 $\boldsymbol{CO}$

相関係数の局所特徴量を配列 $\boldsymbol{CO}$ に格納する.

$$\boldsymbol{CO} = \begin{pmatrix} C\vec{O}_1 \\ \vdots \\ C\vec{O}_B \end{pmatrix} = \left(C\vec{O}^1 \dots C\vec{O}^M \right) \quad (8)$$

式 (8) に示すように行ベクトルと列ベクトルで $\boldsymbol{CO}$ を表現する.B は 2 つのバンド信号の組み合わせであり,$B = {}_LC_2$ で示される. 相関係数は一般式 (9) で示される.

$$corr(x,y) = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (9)$$

二つの共分散をそれぞれの標準偏差で正規化したものである. 二つの異なるバンド信号 $\vec{U_i^j}$,$\vec{U_k^j}$ について

$$C\vec{O}^j \xleftarrow{\text{Trans Vector}} corr(\vec{U_i^j}, \vec{U_k^j}) \quad (10)$$
$$(i < k \leq L \quad j \leq M \quad i,j,k \in \mathbb{N}_+)$$

バンド信号間に相関が存在することは, 楽曲信号に単一のウェーブレット波形が異なるスケールにおいて観測されることを意味する. なお変数 i,k の取り方による $C\vec{O}^j$ 内でのベクトル要素の並び順は一意的に定まり, 各楽曲において変化しないものとする.

3.3 全フレームにわたる計算

最後に全フレームにおける局所特徴量の計算から, 最終的な特徴量を求める. 時間フレームごとの局所特徴量ベクトル $A\vec{V}$,$C\vec{V}$ については平均と変動件数, $C\vec{O}$ については平均と分散を用いる. また音量正規化で用いた DC も付け加える. 図 4 に本研究で用いた特徴量を示す.

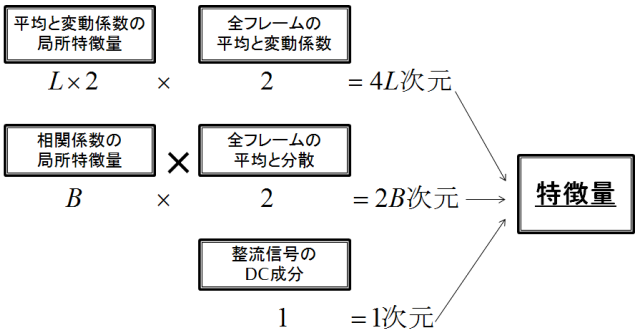

図 4: 特徴量と次元

4. 実験

特徴量の評価として GTZAN データセット によるジャンル判別を行った.10 ジャンル (Blues, Classical, Country, Disco, Hiphop, Jazz, Metal, Pop, Reggae, Rock) で各ジャンル 100 曲, 合計 1000 曲の楽曲ファイルで 30 秒,22.05kHz サンプリングのモノラルデータである. フレーム幅は 4096 サンプル点で実時間で約 0.18 秒と設定した. 今回は分解レベルにおける高周波成分より上位 14 レベルの信号を使う. ウェーブレットは Symlets20 ステップを用いた. 特徴量は平均, 変動係数の局所特徴量から 56 次元, 相関係数の局所特徴量から 182 次元と, 音量正規化に用いた DC 成分から 1 次元より, 合計 239 次元とした.

表 1: Confusion Matrix

BL	CL	CO	DI	HI	JA	ME	PO	RE	RO	⇐ classification as
84	0	1	5	0	2	2	2	3	1	blues
1	97	0	0	0	2	0	0	0	0	classical
3	2	83	1	0	3	3	1	0	4	country
1	1	3	78	5	1	1	1	7	2	disco
2	0	1	5	78	1	1	4	6	2	hiphop
4	3	3	0	1	84	3	0	0	2	jazz
0	0	1	1	0	0	95	0	0	3	metal
2	1	7	1	6	2	1	77	1	2	pop
5	1	4	4	7	2	1	3	71	2	reggae
4	0	5	5	0	2	9	4	1	70	rock

評価方法としては交差検証法を用いた.1 曲を未知データ, 残りを学習データとして未知のデータがどのジャンルに属するかを未知データを入れ替えながら全曲繰り返し判別を行う. 表 1 は本実験の判別結果であり, 横列にある各ジャンルの未知データと置いた曲が, 縦列のどこのジャンルに分類されたかを示す. 対角線に位置している値が各ジャンルにおける正答率であり, 全ジャンルにおいては 81.7 % となった. 詳細を見ると classical や jazz,metal のような音色やリズムに際立った特徴が存在する曲は高い判別精度を有するが,pop や rock などのジャンルの境界が曖昧なものは誤認が多かった. 特に rock に関しては他ジャンルが rock に誤認する場合も多く, データセットを通して平均的な特徴量を持った曲が多いと考えられる. また reggae に関してはリズムパターンに大きな特徴があるのにも関わらず判別制度が低い. これは UWT がオクターブ分解能なので, 曲のテンポやリズムなどに対して周波数分解能が足りず, このような複雑な特徴を捉えきれないと思われる. なお分類器にはデータマイニングツール WEKA より多項式カーネルを用いたサポートベクターマシン (SMO) を用いた.

4.1 考察とまとめ

今回は局所特徴量から特徴量を求める際には, 平均や分散, 変動係数など基本的な統計量を用いたが, これだけでは詳細なリズム成分や楽曲構造を捉えきれない. 今後は詳細な局所特徴量の動きについても調査したい. また基底のウェーブレット波形を変更することによって, 正答率に大きく差が生じたため, 今回用いる事がなかったウェーブレットを試すと同時に, バンド数や時間フレーム幅なども, バリエーションを増やして調査してみたい.

参考文献

[1] http://www.pandora.com/

[2] 大塚玲朗, 梶川嘉延, 野村康雄, “PCM データに対応した感性語による音楽データベース検索システムに関する研究”, 第 14 回データ工学ワークショップ（DEWS2002）,8-P-5 (2003-03).

[3] Zhouyu Fu, Guojun Lu, Kai Ming Ting, and Dengsheng Zhang “A Survey of Audio-Based Music Classifcation and Annotation” IEEE TRANSACTIONS ON MULTIMEDIA, VOL. 13, NO. 2, APRIL 2011

[4] Juan P. Bello ”Measuring Structural Similarity in Music” IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, VOL. 19, NO. 7, SEPTEMBER 2011