ケプストラム分析による心理聴覚実験 に基づいた音楽ジャンル分類

○伊藤直樹(神奈川工科大院・情報工学), 前田功弥, 浜田健史,
西口磯春, 徳弘一路(神奈川工科大・情報メディア), 高澤嘉光(電通大・電気通信)

Classification of Music Genres Using by Cepstrum Analysis Based on
Psychological Listening Experiments

ITOH Naoki(Grad. Kanagawa Institute of Technology), MAEDA Isami, HAMADA
Takeshi, NISHIGUCHI Isoharu, TOKUHIRO Ichiro(Kanagawa Institute of Technology),
TAKASAWA Yoshimitsu(The University of Electro-Communications)

1. はじめに

近年普及したコンテンツ検索手法にあいまい検索が挙げられる。しかし音楽検索では、依然としてタイトルなどの文字情報による一意検索が主流であり、音楽におけるあいまい検索(例えば類似楽曲検索)の手法が望まれている。

我々は、楽曲に含まれる特徴量を利用して、このようなあいまい検索も含めた、広く楽曲の“管理”を行えるソフトウェアの開発のための研究を行っており、その一手法にジャンル分類¹⁾を用いている。しかし、現在楽曲の特徴量としてゆらぎ情報²⁾を主に用いているが、より高いジャンル分類精度実現のための特徴量抽出が課題となっている。

今回、楽曲中の倍音情報を新たな楽曲特徴量として提案し、その抽出法を考えた。また、若者へのジャンルに関する心理聴覚実験の結果を用いて、正準判別分析によるジャンル分類実験を行い、倍音情報の有効性を検証したので報告する。

2. 楽曲に含まれる倍音情報

倍音とは、楽器音に含まれる基音の整数倍の周波数成分の音のことである。倍音はピッチのあるあらゆる音に含まれている。

楽曲は様々な楽器で演奏され、そのスペクトルは複雑であり、波形から楽器ごとの音に分解することは容易ではない。しかし、メロディ音のパワーが大きい瞬間ではスペクトルにメロディ音の倍音成分が出現する。その例を図1に示す。

図1. メロディ音のパワーの大きい瞬間のスペクトル

楽曲中に現れるこのような特徴の度合いやその時間変化は楽曲によって異なる。したがって、これらの情報を倍音情報とすればその楽曲の特徴量となると考えられる。そこで、ケプストラム分析を用いて楽曲の波形から倍音情報を定量的に抽出する方法を考えた。

ケプストラムとは対数振幅スペクトルを逆フーリエ変換したもので、その系列はケプレ

ンという時間と同次元のものとなる。スペクトルに倍音成分が存在すると基本周波数の逆数のケプレンスにピークが表れるという性質がある。

3. 倍音情報抽出法

倍音情報抽出の具体的な方法を以下に示す。抽出対象のデータは、22.05kHzサンプリング、16bit量子化のwave形式とした。

まず、wave時系列波形を1024点(約46ms)ごとにフレーム分割し、各フレームをFFT処理してスペクトルを計算し、更に対数振幅スペクトルをIFFT処理してケプストラムを計算する。次に、フレーム毎にケプストラムのケプレンス0.5~10ms(基本周波数100~2000Hzに対応)の部分で以下の5項目について求める。

- (1) 平均 (以下ceps_avg)
- (2) 分散 (以下ceps_var)
- (3) 標準偏差 (以下ceps_std)
- (4) 最大値 (以下ceps_max)
- (5) 正規化した最大値=(最大値-平均)/標準偏差 (以下ceps_nmax)

これらはそれぞれ、フレーム区間においてどれだけ倍音構造の特徴が現れているかを表す尺度として考えたものである。これらを全フレームで求めることにより、倍音構造の出現の度合いの時間変化を表す時系列が計算される。例として、あるインストルメンタル楽曲のceps_varの時間変化を図2に示す。図中の網掛部はバイオリンによるメロディ音が存在する区間である。

図2. ある楽曲の ceps_var の時系列

次に、これら5つの時系列の2048点(約95秒)についてそれぞれ次のものを計算する。

- (1) ゆらぎ(傾き・y切片)
パワースペクトルの対数周波数系列での最小二乗法による回帰直線で表す。
- (2) 平均
楽曲全体に含まれる倍音の量を表す。
- (3) 分散
倍音の量の時間変化の大小の開きを表す。
- (4) 標準偏差
倍音の量の時間変化の大小の開きを表す。
- (5) 最大値

倍音の量の瞬間最大量を表す。

以上で楽曲の倍音情報が5×6=30種の特徴量として抽出される。これらの特徴量を倍音特徴量と呼ぶこととする。なお、分散と標準偏差は統計量としては類似だが、特徴量としては違いがある可能性があるため、個別に扱った。

4. ジャンル度計測のための心理聴感実験

通常、音楽ジャンルの判定にはテンポなどの絶対的基準の他に主観を伴い、また一つの基準だけで決まるとは限らない。そこで13ジャンル計1300曲(1ジャンル100曲)について、0~7の8段階評価でジャンルらしさ(ジャンル度)を計測する心理聴感実験を行った。13ジャンルを以下に示す。

ボーカル有り

(1) Hard Rock, (2) Pops, (3) Ballade,

(4) Enka, (5) Oldies (含 Rock'n'Roll, R&B, Blues),

(6) Dance (含 Euro Beat, Disco), (7) Hiphop (含 Rap),

(8) Techno (含 House, Trance), (9) Reggae, (10) Acapella

ボーカル無し

(11) Jazz (含 Fusion), (12) Classic (含 Sound Track),

(13) Mood

被験者は、大学・大学院生13名である。各楽曲とも曲頭約3秒後~103秒後までの約100秒間を試聴した後の印象で評価させた。また、この実験は被験者への負担が高いため、各被験者のペースに応じて、複数回に分けて行うことを認めた。

5. 倍音特徴量のジャンルごとの分布

13ジャンル1300曲から倍音特徴量を抽出し、ジャンルごとの倍音特徴量の分布を調べた。ここでは ceps_var の各特徴量の分布グラフを図3に示す。尚、グラフの曲目は左から100曲ずつ、4章で示したジャンル毎に並んでいる。

図3. ceps_var の各特徴量のジャンル分布

(上:ゆらぎ(傾き/y切片),中:平均/分散,下:標準偏差/最大値)

これらの分布のグラフから,ceps_var の各特徴量の値それぞれについてジャンルごとの傾向が現れていることが分かる。

同様に, ceps_var 以外の特徴量も含めた全ての特徴量にジャンルごとの傾向が現れた。

6. 倍音特徴量を使用したジャンル分類実験

13ジャンル1300曲のうち、複数ジャンルの混

合がないと判断された5ジャンル426曲を用い、正準判別分析によるジャンル分類実験を行った。ジャンル内訳はRock80曲, Enka97曲, Oldies75曲, Jazz83曲, Mood91曲である。このうち各ジャンルからランダムに10曲ずつ計50曲をジャンル未知楽曲とし、残りの376曲を解析用楽曲とした。

最初に、解析用楽曲の倍音特徴量30種を説明変数、ジャンル5種を目的変数として正準判別分析を行い、各ジャンルで正準得点のジャンル内平均を求めた。次に、解析用楽曲376曲についてジャンル推定を行った。推定方法は、楽曲の正準得点と正準得点のジャンル内平均とのユークリッド距離が最も近くなるジャンルをその楽曲の推定ジャンルとする方法である。結果を表1に示す。的中率はJazzについてはやや低くなっているが、全体的には高くなっており、倍音特徴量による分類ができていているといえる。

表1. 解析用楽曲376曲の Confusion Matrix

	Rock	Enka	Oldies	Jazz	Mood	計	的中率(%)
Rock	67	1	0	2	0	70	95.7
Enka	0	83	0	1	3	87	95.4
Oldies	3	2	57	3	0	65	87.7
Jazz	5	1	11	55	1	73	75.3
Mood	0	4	0	4	73	81	90.1
曲数	75	91	68	65	77	376	89.1

(縦の項目:実際のジャンル 横の項目:推定のジャンル)

続いて、解析用楽曲から得られた正準得点のジャンル内平均を使用し、ジャンル未知楽曲50曲についても同様にジャンル推定を行った。結果は表2のようになり、ジャンル未知楽曲についても高い的中率となった。

表2.ジャンル未知楽曲50曲の Confusion Matrix

	Rock	Enka	Oldies	Jazz	Mood	計	的中率(%)
Rock	10	0	0	0	0	10	100
Enka	0	9	1	0	0	10	90.0
Oldies	0	1	8	1	0	10	80.0
Jazz	0	1	1	8	0	10	80.0
Mood	0	1	0	0	9	10	90.0
曲数	10	12	10	9	9	50	88.0

(縦の項目:実際のジャンル 横の項目:推定のジャンル)

7. 結論

倍音特徴量を用いたジャンル分類実験で高い的中率となったことから、倍音情報がジャンル分類に有効であることがわかった。今後の課題は、倍音情報抽出法を改善し、抽出される特徴量の冗長性を排除することである。

参考文献

- [1] Ito Naoki, Tokuhiro Ichiro, Nishiguchi Isoharu, Takasawa Yoshimitsu, "AMuGCS - Automatic Music Genre Classification System - with Multivariate Analysis", ISMA, (2004)
- [2] 田中実, 伊藤直樹, 飯島秀亮, 山川佑介, 横関圭, 吉澤友孝, 渡邊誠之, 徳弘一路, 西口磯春, 高澤嘉光, "楽曲のゆらぎ情報およびスペクトル情報の抽出", 日本音響学会, 音楽音響研究会, (2001).

伊藤直樹

(現在, 北陸先端科学技術大学院大学 知識科学研究科)
連絡先: 923-1292 石川県能美市旭台1-1 北陸先端科学技術大学院大学, 知識科学研究科 西本研究室
n-itou@jaist.ac.jp