音楽情報からの特徴抽出

岩崎　真哉・島田　英之*・塩野　　充*・宮垣　嘉也*

岡山理科大学大学院工学研究科修士課程情報工学専攻

*岡山理科大学工学部情報工学科

（1997年10月6日　受理）

1．まえがき

1980年代初期からのカラオケの急速な普及や，CMのタイアップによる音楽の販売戦略等で古くからあった音楽[1]は一層身近なものになり，今では様々な音楽が溢れている。

一口に音楽といっても何世紀も前からある伝統的なものから電子楽器やコンピュータを使用したものまで様々であり，その分類というものは非常に困難である。

近年の楽曲に関する研究として，楽曲の自動採譜[2]や曲中の楽器をスペクトルから検出[3]する方法等，幅広く行われている。これらは楽曲における個々の音の性質や楽器の役割を厳密に捉えようとする研究であるが，曲全体を大局的に捉えたアプローチによる研究はあまり行われていない。

音を聞くことのできる人間は，音楽を聞いてある程度の識別が可能である。例えば，演歌を耳にした時に，「これは明らかに演歌である」と感じるような事である。それは，その音楽に「演歌」と認識させる何か性質があるためだと考えられる。

本実験では，様々な楽曲にそれぞれ特有の性質があることを前提とし，その特徴を抽出，分類することを目的とする。

2．サンプルの選択及び収集

2．1　サンプルの選択

本実験では対象をユーロビート（eurobeat），ジャズ（jazz），クラシック（classicism），演歌，童謡（日本童謡）とした。これらは全て明確に定義されたジャンルであるといい難い。しかし，これらは明らかに人間が試聴して分類可能なカテゴリであると考え，サンプルとしては十分流用可能であるとした。以後特に明記しない場合，ジャンルをカテゴリと記し，それぞれの楽曲データをサンプルと記すことにする。

2．2　収集方法

市販の音楽CDよりサンプリング周波数8kHzでサンプリングを行い，楽曲データとした。選曲はCDアルバムのジャケットにそれぞれのカテゴリを明記してあったものを条件とし，その中から無作為に30サンプルを用いた。したがって，以後5カテゴリ各30サンプルを使用して実験を行っていく。

2.3　使用機器

楽曲データのサンプリングに使用した機器は次のものである。

SONY NWS-5000UI オーディオインターフェース

CD の標本化（サンプリング）周波数は44.1kHzであるので，本来その周波数でサンプリングするのが理想である。上記の機器も性能的にその周波数によるサンプリングが可能である。サンプリング周波数を大きくすればそれだけ分解能が上がり，より詳しい波形を得ることができる利点がある。しかし，他の処理の際に多大な演算時間を要するなどの欠点がある。

今回は欠点の方が利点を上回るので，本機器最低限のサンプリング周波数を用いてデータ数を極力少なくした。ジャンルを判別するにはこの程度のサンプリング周波数で十分と判断した。

3. 特徴抽出

特徴量として，(1)演奏時間，(2)パワースペクトルの2つの点を取り上げる。

3.1　演奏時間による傾向

各曲の演奏時間をカテゴリ別に統計を取る。それにより演奏時間にどのような傾向が見られるかを類推する。全てのサンプルの歌詞カードに演奏時間が記載されている訳では無い。したがって，今回は実際にサンプリングしたデータとサンプリング周波数を用いて次のように演奏時間を求めた。

$$\text{演奏時間} = \frac{\text{サンプルのデータ数}}{\text{サンプリング周波数}}[\text{sec}] \tag{1}$$

しかし，楽曲データをサンプリングする際，1秒～2秒程度の無音部分もデータとして含まれているものもあるため，上式で求められる値が完全に正確であるというわけではない。

3.2　パワースペクトルを用いた特徴量

3.2.1　前　処　理

まず，各サンプルの波形データを秒単位に分割する（図1）。そして，分割した秒を t [sec]，分割した結果できた波形データの個数を N とおき，各サンプルの分割数を便宜的に次の様に表す。

$$N_t\,[\text{個}] \tag{2}$$

例えば，$N_t = 30_{15}$ とすると，「あるサンプルの波形データにおいて，15秒単位では30個に分割された波形データを作成できた」という意味である。本実験では $t = 1, 15, 30$ について分割を行った。

次に，これらの分割した波形データより，それぞれのパワースペクトルを FFT により求める（図2）。したがって，パワースペクトルの波形もサンプルの分割数 N だけ生成される。

また FFT のアルゴリズムにおいて，要素数（データの個数）n の値については

$$n = 2^G \qquad (G\text{:正整数}) \tag{3}$$

が成立する場合のものを用いている[4]。ここで分割した波形データが式(3)を満たさない場合に不都合が生じてくる。そこで分割した波形データに以下のような処理を行った。

(1) 分割した波形データのデータ個数に一番近い 2 のべき乗の個数を用いる。

(2) 増えたデータ個数分のデータ値は 0 を使用する。

増加分のデータ値を 0 として扱ったのは，エリアシングを避けるためである。表 1 に

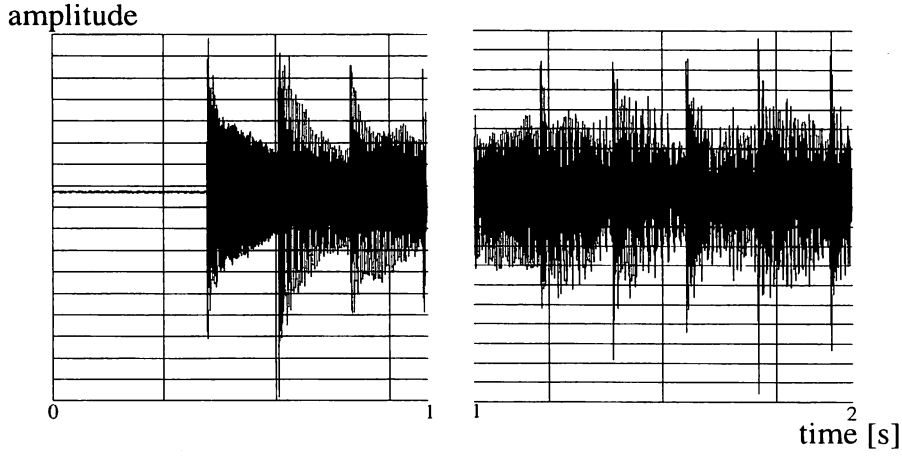

図 1　サウンドデータの分割

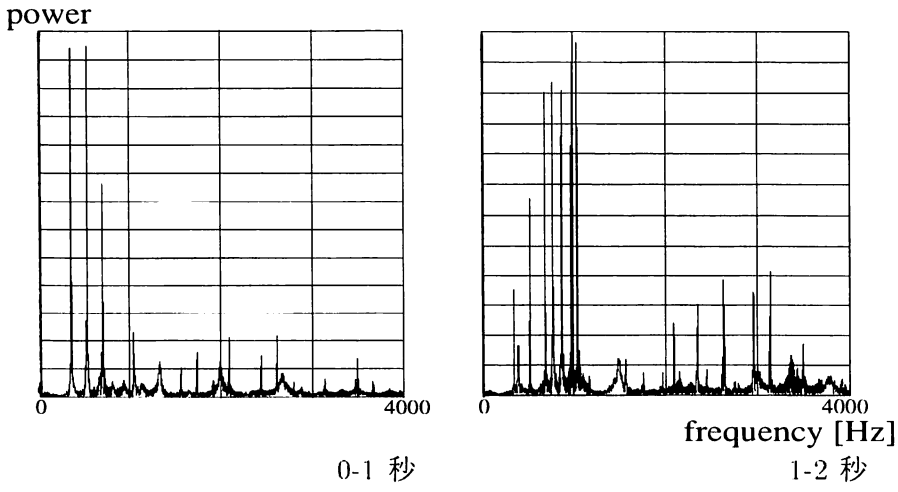

図 2　FFT による各パワースペクトルの算出

分割時間と処理前及び処理後の音楽波形のデータ数を示す。

FFT の際に，今回は窓関数を特に考慮せずに波形データを分割した。したがって，切り出した波形データは矩形窓を用いたことになる。

各サンプルの一曲の長さが統一されていないのはすでに述べた通りである。これはつまり，同じ t の値でもサンプルによって N の値が異なることを示している。したがって，それから求められるパワースペクトルの波形の個数 P もサンプルによって異なることがわかる。

そこで，周波数のパワーの度合いを各サンプルの特徴とすること，特徴の評価のための基準を設けることという 2 つの目的から，今回は前処理で得られた周波数パワースペクトルをそれぞれの周波数ごとに一曲分を加算し，その平均を取ることにより各サンプルを比較できるようにした。その結果得られたパワースペクトルを特徴量とする（図 3）。そして得られた特徴量から更にサンプリング定理と直流成分を考慮すると，それぞれの分割時間における特徴次元数は表 2 のようになり，データの個数を n とすると以下のように算出できる。

$$\text{特徴次元数} = \frac{n}{2} - 1 \tag{4}$$

表 1　分割時間と音楽波形データ数

分割時間［sec］	波形データ数	処理後
1	8,000	8,192
15	120,000	131,072
30	240,000	262,144

表 2　分割時間と特徴次元数

分割時間［sec］	次元数［次元］
1	4,191
15	65,535
30	131.071

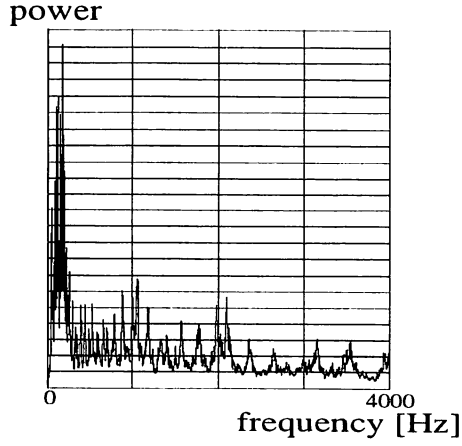

図 3　各パワースペクトルから求まる平均スペクトル

3.2.2 ジャンルの判別

今回得られた特徴の評価方法として，パターンマッチング法の一つである単純類似度法[5]による認識実験を行った。各カテゴリ15サンプルを学習サンプルとし，残り15サンプルを未知サンプルとした。標準パターンは学習サンプルから作成され，これは学習サンプルの特徴量の平均をとったものである。

単純類似度法は，パターンをベクトル表現した場合その2つ（標準パターンと入力パターン）のベクトルがなす角度を求め，どれくらい類似しているかを決める方法である。lをカテゴリ，標準パターンを$x^{(l)}$，入力パターンをxとすると，

$$s_s^{(l)}(x) = \cos\theta^{(l)} = \frac{(x, x^{(l)})}{\|x\|\|x^{(l)}\|} \tag{5}$$

x, x^lが具体的なパターンであることを考えると，通常この式は0以上1以下の値をとり，2つのパターンが似ているほどそれらを表す2つのベクトル間の角度は狭くなるから，1に近い値となる。

そこで，上式に基づいた識別方法は，

$$s_s^{(l_0)}(x) = max_l s_s^{(l)}(x) \tag{6}$$

ならば

xのカテゴリ名はl_0 (7)

となる。

4. 実験結果と考察

4.1 演奏時間の傾向

表3はサンプルの演奏時間を分単位で統計を取ったものである。ユーロビート，演歌，童謡に関してはそれぞれ3分台，4分台，1分台にそれぞれ多く集まっていることがわかる。また，ジャズとクラシックの演奏時間はさまざまであることがわかる。

しかし，クラシックに関してはある題目の中の章ごとについて別々に調べているので，全体を通してみれば一番演奏時間が長いのはクラシックということがいえる。

表3 サンプルの演奏時間の統計

category/time [sec]	0—1	1—2	2—3	3—4	4—5	5—6	6—7	7—
ユーロビート	0	1	0	23	6	0	0	0
ジャズ	0	0	9	6	5	3	4	3
クラシック	0	6	7	5	3	2	1	6
演歌	0	0	0	9	21	0	0	0
童謡	4	21	3	2	0	0	0	0

4．2　パワースペクトルを用いた特徴量

表4は実験(2)において行った認識結果である。

全体的に学習サンプルに対しての認識率が高く，未知サンプルに対しての認識率はカテゴリに分かれる。分割時間を増やすことによる傾向として，学習サンプルに対しての認識率は上がるのだが，未知サンプルに対しての認識率は上がるカテゴリもあれば下がるカテゴリもあることから，最適な分割時間の存在も考えられる。

4．2．1　各カテゴリの傾向

ユーロビートは全ての認識率において100%であった。これは，どのサンプルの特徴量もピーク周波数が近い値で存在し，また，一曲分のスペクトルの外形をとってみてもほぼ同じような波形を示す。

また，童謡の認識率が全体的に高く，分割時間を長くしていくほど認識率が上がっている。このことから更に分割時間を長くすれば認識率の向上が望まれる可能性が高い。

演歌においても分割時間を多くすれば認識率が上がることから童謡と同じことがいえると考えられる。

ジャズとクラシックについて，学習サンプルに対する認識率はほぼどの分割時間を用いてもあまり変化しなかった。しかし，未知サンプルに対しては徐々に悪化する傾向が見られ，しかも良い認識率が出なかった。

以上をまとめると，ユーロビートに関しては認識率が常に100%であることから，本手法により抽出された特徴量が有効であることを示している。また，認識率の高かった演歌と童謡が単調な繰り返しから成る楽曲であり，認識率の低かったジャズやクラシックは非常にさまざまな形態（使用されている楽器やリズム等）が取られており，また，演奏者によっても変化することから，ある程度単調な楽器に対しては本手法による特徴量の抽出が有効ではないかと考えられる。

5．ま　と　め

本研究の音楽の特徴を抽出するという目的のうち，今回は5つのジャンルに絞り次の2

表4　認識率

	1 [sec]		15 [sec]		30 [sec]	
カテゴリ	学習	未知	学習	未知	学習	未知
ユーロビート	100	100	100	100	100	100
ジャズ	73	60	80	33	80	40
クラシック	100	40	100	33	100	33
童謡	93	86	100	93	100	93
演歌	100	60	93	60	100	73

点に着目して実験を行った。

- 演奏時間の長さ
- パワースペクトルの分布

それにより，どちらも特徴として有効なジャンルとそうでないジャンルに分かれた。パワースペクトルを用いた方法では，構成や曲調等似通っている楽曲には有効であることがわかった。本実験における課題として次のものが挙げられる。

- サンプル数の追加
- カテゴリの分類方法の再考
- 分割時間をさらに変化させた場合の調査

カテゴリは今回一般的に用いられている「ジャンル」に分けて実験を行ったが，2．1節でも述べたようにこれらに厳密な定義があるわけではない。カテゴリの分類方法として，被験者を用意してのその楽曲のイメージにより分類することなどが考えられる。また，全体の結果が出るまでの処理に時間がかかるので，波形データを有効に利用した実験の検討が必要である。

以上のことから本研究の課題として，

- カテゴリの構成の検討
- 他の手法による特徴抽出

が挙げられる。

参考文献

1）小川博司：“メディア時代の音楽と社会”，音楽之友社，1993.
2）大照，橋本：“仮想音楽空間”，オーム社，1995.
3）後藤，村岡：“打楽器音を対象にした音源分離システム”，信学論（D-II），Vol. J77-D-II，pp. 901-911，1996.
4）佐川，貴家：“高速フーリエ変換とその応用”，昭晃堂，1993.
5）舟久保登：“視覚パターンの処理と認識”，啓学出版株式会社，1990.

An Experiment to Extract a Feature from Music

Shinya IWASAKI, Hideyuki SHIMADA*, Mitsuru SHIONO*
and Yoshiya MIYAGAKI*

Graduate School of Engineering,
**Department of Information & Computer Engineering,*
Faculty of Engineering,
Okayama University of Science,
Ridai-cho 1-1, Okayama, 700-0005 Japan

(Received October 6, 1997)

There are many reserches concerned with music or sounds in various field: to separate sound sources by melody, making a musical score automatically, and so on.

Our purpose in this paper is to extract the feature that show what we can identify the music as itself. Each category contains 30 samples.

Now, we separated music janls into 5 categories (*eurobeat, jazz, classicism, ENKA, DOUYOU* (is a japanese children's song)) and examined about two points.

1. The perfoming time of each categories

2. The distribution of the power spectrum

In second experimentation, we experimented on pattern recognition, and got 100% for recognition rate of *eurobeat*, on the other hand we got only 33% for recognition rate of *jazz* or *classicism*.