

独立成分分析を用いた楽曲ジャンル分類 Musical genre taxonomy using Independent Component Analysis

松田 拓也[†] 六井 淳[†]
Takuya Matsuda Jun Rokui

1. はじめに

PC やデジタルオーディオプレーヤーの HDD 容量が大きくなり、大量の楽曲を所有、利用することが可能となった。これに伴い、PC やデジタルオーディオプレーヤーを使用している人が増えている。楽曲の数が膨大になれば、アルバムを探すこともプレイリストを作成するための操作も難しくなる。膨大に存在する楽曲の中から、ユーザの聞きたい楽曲をわずらわしい操作なしに選曲する研究が行われてきた。

楽曲のフレーズの特徴から似ているフレーズを手がかりに多くの異なる楽曲を連続的に楽しむことのできる新たな音楽システムが提案されている[1]。音楽再生の様子やシステム動作を効果的に視覚化して、ユーザが大規模な音楽ライブラリを楽しむための作業負荷の軽減を実現している。この研究では、楽曲を時間的に統合されたフレーズの集合として扱い、類似度の計算には、曲単位に平均化された特徴ではなく、各フレーズのもつ特徴量を用いる。

また、特徴量を用いる際に、楽曲の調波信号と打楽器信号とを分離した後の特徴量抽出を行う研究がある[2]。楽曲は旋律楽器のみでなく、打楽器も同時に演奏される場合が多い。そのため、低音旋律パターンを推定する際にこれらを分離し調波音を強調することによってより正確なパターン抽出を行うことができる。小節単位の打楽器信号パターンはジャンルごとに特徴的であるので、それらはジャンル分類を行う上で非常に有利に分類できるとしている。

本研究は、各楽曲をジャンルごとに分類・識別するための手法を提案する。本研究では、これまでの研究において、音響的特徴量が様々な音源から成っている楽曲に対して使われていることに着目した。楽曲の特徴量を抽出する際、様々な音源から成る楽曲にそのまま特徴量抽出を行ってはいくつか特徴を捉えることができない。また、ジャンルが異なる楽曲を比較する際、十分な結果が得られない。そこで、音源分離を行ってから特徴量抽出を行う手法を提案する。楽曲は複数の独立した音源から成り立っているため、音源分離することによって、ひとつひとつを単音源として捉えることで、単音源ごとに特徴量抽出を行う。複数の音源から成った楽曲にそのまま分析を行うより、音源分離を行って単音源ごとに分析を行う方が、より詳しい分析が可能である。本研究では、音源分離を行う手法として、独立成分分析（ICA、Independent Component Analysis）を挙げる。

ICA は信号の統計学的な性質を利用して多次元信号の分離を行う手法である。音源ごとにフレーム分割を行い、各フレームごとに特徴量を算出し、複数個の特徴量を求める。フレーム分割を行うことで、楽曲の平均化された特徴ではなく、フレームごとに特徴を算出するため、個々のフレームの詳しい特徴を得ることができる。

本研究では音響分析手法として、線形予測符号化（LPC、Linear Predictive Coding）から得られる LPC ケプストラム係数を特徴量として用いた。検証としてジャンルのクラス識別をサポートベクターマシン（SVM、Support Vector Machine）を用いて行った。

2. 提案法

従来の音響分析手法では、複数の音源から成る楽曲に対して、音響的特徴を用いている。本研究では、複数の音源のまま音響的特徴を算出すると十分な結果が得られないと考え、音源ひとつひとつを音源分離する手法を提案する。また、楽曲では常に同じ音源が使われていないため、フレーム分割を行うことで、より詳しい分析が可能であると考える。提案法の手順は図 2.1 である。

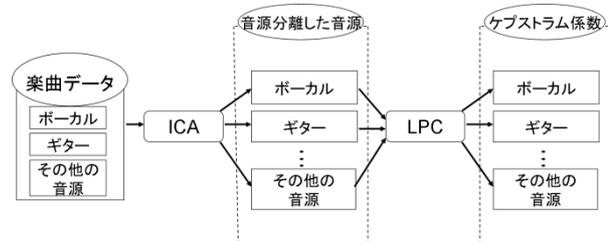


図 2.1 提案法の手順

具体的な手順は次のようになる。

- STEP 1 : 分析楽曲のダンプを行う。
- STEP 2 : 楽曲をフレーム分割する。
- STEP 3 : フレームごとに ICA を行う。
- STEP 4 : 音源分離された音源ひとつひとつに対して LPC を行う。
- STEP 5 : SVM を行う。

まず、wave 形式の楽曲をダンプする。本研究で用いる楽曲のサンプリングレートは 11,025 Hz としている。STEP 2 としてダンプした楽曲データをフレーム分割する。混合信号の変数の数が独立している信号の変数の数より大きい場合、混合信号は線形従属となるため、フレーム分割を行う前に楽曲データの無相関化を行った。無相関化を行

[†] 島根大学総合理工研究科数理情報システム専攻
Interdisciplinary Faculty of Science and Engineering
Department of Mathematics and Computer Science Shimane
University

う手法として、主成分分析を用いた。STEP 3として各フレームごとにICAを行う。楽曲の信号源の数が不明なので、減次を用いることによるICAを行う。また、独立成分を逐次的に求める方法としてFast-ICAと呼ばれるアルゴリズムを用い、独立成分の導出には不動点法を用いた[3]。観測信号としてのデータは、フレームをさらに分割することで得た。各フレームから得た単音源ひとつひとつに対して、LPCを行う。LPCケプストラム係数の数は8として算出した。最後に、SVMを用いてジャンル分類を行う。本研究では、各ジャンルのLPCケプストラム係数を用いて、1対多クラスの2クラスとして分析を行った。また、n-foldクロスバリデーションを行わず、分類を行った。

3. 楽曲の音源分離検証

実際に、ICAで音源分離を行うことで、楽曲の音源をどの程度分離できるかを確認し、音源の数を固定して、どの程度誤差が出るのかを調べた。

実験に使用した楽曲はMIDIで作成し、音源の数を10個とした。楽曲の構成は表3.1のようになっている。ジャンルはロックである。音源はギター、ドラム、ベース、管楽器を複数使用した。元楽曲の信号は図3.1のようになる。グラフの横軸は時間、縦軸は振幅となっている。

表 3.1 楽曲データの構成

| | |
|-----------|----------|
| データファイル形式 | WAVEファイル |
| サンプリングレート | 11,025Hz |
| チャンネル数 | 1 (モノラル) |

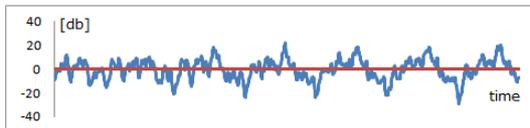


図 3.1 元楽曲の音源の信号

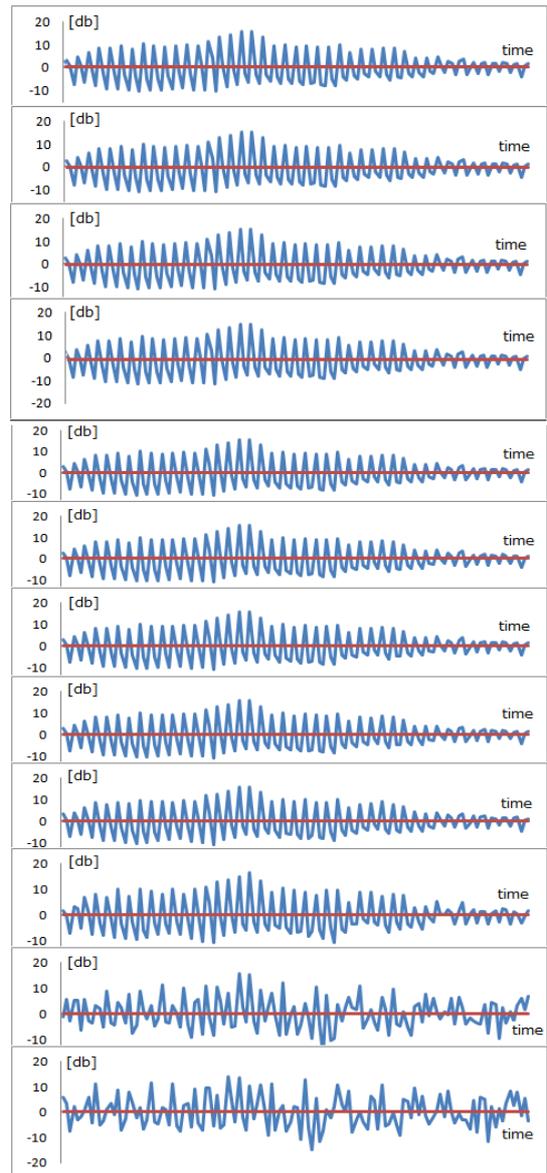
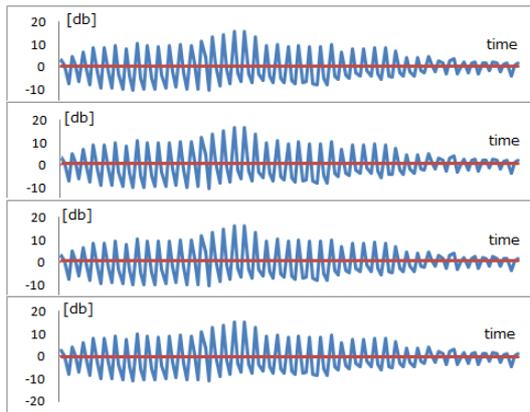


図 3.2 ICAによって分離した単音源

図 3.2 に示すように、楽曲にICAを用いると、単音源ごとに異なった信号を抽出することができる。この楽曲での実験では、実際の音源の数は10個であったのに対し、ICAを行った結果、出力された音源の数は16個となった。このことからICAでは、音源の数が分かっている場合でも、実際の音源の数通りに正確に分離することはできない。

4. ジャンル分類実験

本研究では、検証としてSVMを用いて、正しくジャンル分類が可能であるかどうかについて実験を行った。多クラス分析は識別したいジャンルとそれ以外のジャンルといったような1対多で行っている。SVMに用いるLPCケプストラム係数の次元数は8次元とし、分類には楽曲のフレームごとに求めたLPCケプストラム係数の全てを用いて分

類を行う。ここでのLPCとは楽曲のダンプファイルに対してICAを行わず、LPCを行い、LPCケプストラム係数を算出したものを用いている。

表 5.1 ジャンル識別率

| | 提案法 | LPC |
|------------|-----------|-----------|
| 識別率（クラシック） | 81.3706 % | 82.5324 % |
| 識別率（演歌） | 81.3103 % | 79.96 % |
| 識別率（ジャズ） | 81.3806 % | 78.5027 % |
| 識別率（J-ポップ） | 81.3505 % | 77.1858 % |
| 識別率（ロック） | 74.8794 % | 81.8659 % |

提案法は、ジャンルがロック以外のジャンルの楽曲はすべて80%を超える識別率であった。ジャンルがクラシックとロックである識別は、源信号から得たLPCケプストラム係数を算出した識別の方が識別率が良い結果となった。この結果からクラシックとロックには何かしらの相関があると考えられる。

5. LPC ケプストラム係数値の検証

先ほどのジャンル分類の実験で、各ジャンルの識別率を得ることができたが、本研究における提案手法と、単にLPCを行う手法でのLPCケプストラム係数の違いについて検証を行った。分析を行う楽曲データの詳細は表4.1、のとおりである。使用したジャンルはクラシック、演歌、ジャズ、J-ポップ、ロックで、各6曲ずつ用いた。

表 6.1 楽曲データの構成

| | |
|-----------|----------|
| 楽曲データ数 | 36曲 |
| データファイル形式 | WAVEファイル |
| サンプリングレート | 11,025Hz |
| チャンネル数 | 1（モノラル） |

特徴量のLPCケプストラム係数の次元数は8次元とした。また、LPCケプストラム係数の最大値を1として、正規化を行った。

5.1 実験結果

各ジャンルの楽曲のLPCケプストラム係数をグラフにしたものが図6.1～図6.10である。なお、図の横軸はLPCケプストラム係数の次数、縦軸はLPCケプストラム係数の係数値とする。ここでのLPCとは、楽曲のdumpファイルにICAを行わず、LPCを行ってLPCケプストラム係数を算出したものを用いている。

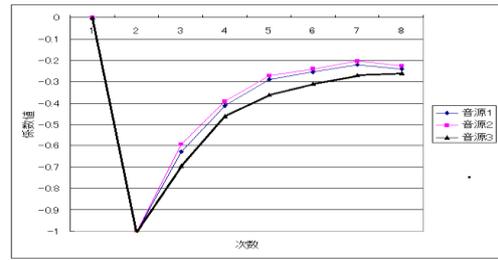


図 6.1 クラシック：提案法

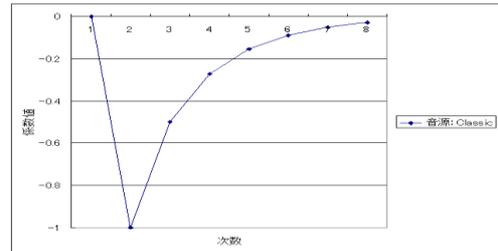


図 6.2 クラシック：LPC

図6.1は、ジャンルがクラシックである楽曲データにICAを行ってから、LPCケプストラム係数を図示したものである。図6.2は、ジャンルがクラシックである楽曲データのLPCケプストラム係数を図示したものである。

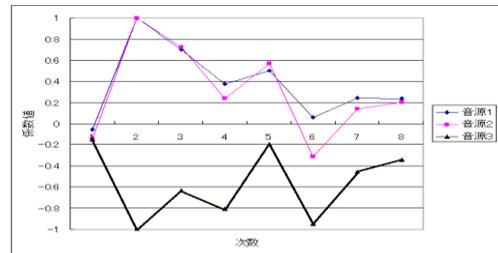


図 6.3 演歌：提案法

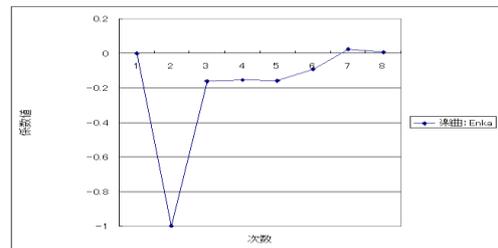


図 6.4 演歌：LPC

図6.3は、ジャンルが演歌である楽曲データにICAを行ってから、LPCケプストラム係数を図示したものである。図6.4は、ジャンルが演歌である楽曲データのLPCケプストラム係数を図示したものである。

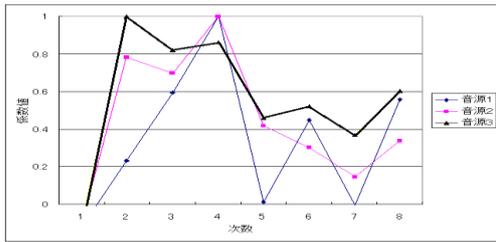


図 6.5 ジャズ：提案法

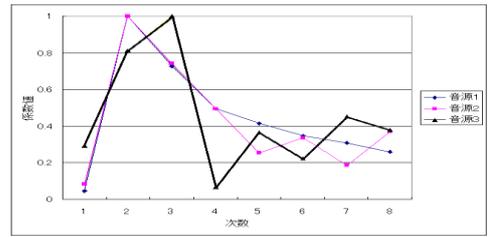


図 6.9 ロック：提案法

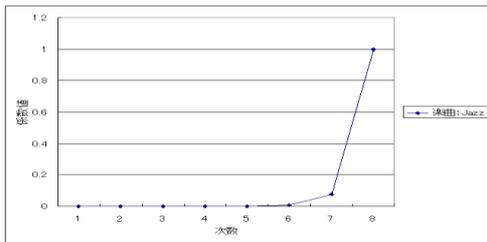


図 6.6 ジャズ：LPC

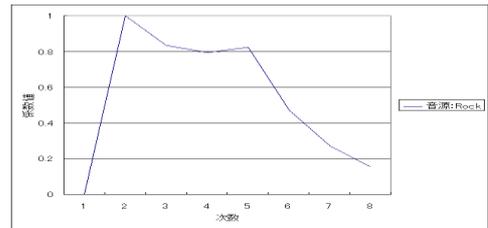


図 6.10 ロック：LPC

図 6.5 は、ジャンルがジャズである楽曲データに ICA を行ってから、LPC ケプストラム係数を図示したものである。図 6.6 は、ジャンルがジャズである楽曲データの LPC ケプストラム係数を図示したものである。

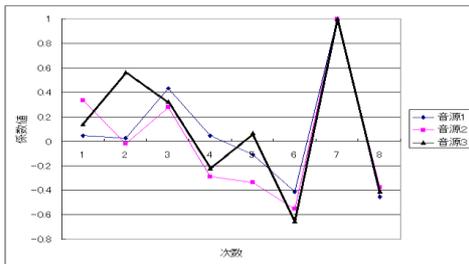


図 6.7 J-ポップ：提案法

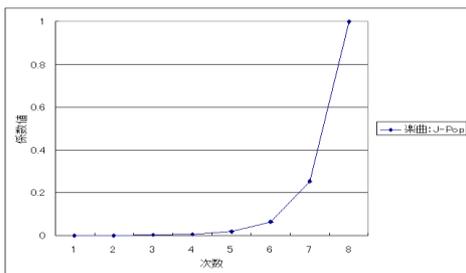


図 6.8 J-ポップ：LPC

図 6.7 は、ジャンルが J-ポップである楽曲データに ICA を行ってから、LPC ケプストラム係数を図示したものである。図 6.8 は、ジャンルが J-ポップである楽曲データの LPC ケプストラム係数を図示したものである。

図 6.9 は、ジャンルがロックである楽曲データに ICA を行ってから、LPC ケプストラム係数を図示したものである。この図は、楽曲データのある時系列において、音源別に値がなるべく異なるケプストラム係数をいくつか図示した。図 6.10 は、ジャンルがロックである楽曲データのケプストラム係数を図示したものである。この図は、ICA を行ってから、LPC ケプストラム係数と同じ時系列のものを用いた。本研究の提案手法では、音源によって、異なる値を算出している。クラシックとロックの LPC ケプストラム係数におけるグラフの形状が似ている。このことから、クラシックとロックについては ICA の分離効果が小さく、そのような楽曲については、音源分離効果が得られないと考えられる。

6. まとめ

本研究では、音響分析手法を行う前に、音源分離を行う手法を提案した。ICA を用いて、音源分離を行うことで、単音源ごとの分析が可能となり、より詳しい分析を行うことができる。今後の研究課題として、より良い特徴量や識別手法を探し、識別率をより向上させることが必要と考えられる。

参考文献

- [1] 堀内直明・藺田俊行・田中浩司・田中淳一・長沢秀哉・萩山真一：Song Surfing:類似フレーズで音楽ライブラリを散策する音楽再生システム, PIONEER R&D(Vol.17, No.2/2007)
- [2] 角尾衣未留・George Tzanetakis・小野順貴・嵯峨山茂樹：音楽音響信号の低音旋律パターンのクラスタリングと自動ジャンル認識への応用, 情報処理学会, 2009
- [3] 村田昇：入門独立成分分析/電機大出版局, 2004