くし型フィルタと Multi-class SVMによる混合音からの演奏楽器推定

くし型フィルタと Multi-class SVM による混合音からの演奏楽器推定

山森　一人 a)・青島　大河 b)・相川　勝 c)

Musical Instrument Identification in Polyphonic Music using Comb Filters and Multi-class SVM

Kunihito YAMAMORI, Taiga AOSHIMA, Masaru AIKAWA

Abstract

Automatic transcribing is one of the attractive application for computer, and many software have been developed. However, most of these software can not address polyphonic sound. The process of transcription for polyphonic sound is divided into two processes; sound source separation and musical instrument identification. In this paper, we propose a method to combine comb filters for sound source separation and multi-class SVM for musical instrument identification. Experimental results showed that our method could identify musical instrument with 57.6% accuracy.

Keywords: comb filter, multi-class SVM, MFCC, musical instrument identification

1. はじめに

近年、コンピュータを用いて音響信号から自動で楽譜を作成する「自動採譜」と呼ばれる研究が盛んに行われている[1][2]。自動採譜を目的としたソフトウェアも数多く開発され、音楽知識や採譜経験なしに誰でも簡単に採譜が行えるようになった。しかし、ほとんどの採譜ソフトウェアには、複数の楽器音が混在する音響信号を正しく採譜できないという問題がある。

自動採譜の工程の中に、対象となる音源を構成している楽器が何であるかを特定する工程があり、これを楽器推定と呼ぶ。複数楽器からなる混合音に対する楽器推定の場合は、前処理として混合音をそれぞれの単楽器音へと分離した後、それぞれの音源について楽器推定を行う。混合音をそれぞれの単楽器音へと分離する工程には三輪ら[1]のくし型フィルタを用いた手法などがあり、分離された音源について楽器推定をする工程には北原ら[2]のSVMを用いた手法などがあるが、分離精度、楽器推定精度とも不十分である。

本研究では、複数の楽器音が混在する音響信号をそれぞれの単楽器音へ分離し、その単楽器音が何の楽器の音であるかを識別することを目的とする。そこで、北原らの手法をベースとし、音源分離にくし型フィルタを用いる手法を提案する。

a) 情報システム工学科教授

b) 情報システム工学科

c) 宮崎大学工学部教育研究支援技術センター技術職員

2. フーリエ解析

フーリエ解析とは、時間領域の信号波形を周波数領域へ移すフーリエ変換を行って信号を解析することである。周波数領域へ移されたデータはスペクトラムと呼ばれる。信号波形$f(t)$をスペクトラム$F(\omega)$に変換するには式(1)を用いる。このとき、eは自然対数の底、tは時間、ωは周波数、iは虚数単位である。

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-i\omega t}dt. \quad (1)$$

スペクトラムを対数変換し、再度フーリエ変換を行うことにより、ケプストラムを得ることができる。ケプストラムの低次成分には楽器の音響特性、高次成分には音高情報が現れる。スペクトラム$F(\omega)$にフーリエ変換を行い、ケプストラム$C(q)$に変換する式を式(2)に示す。このとき、qはケフレンシと呼ばれる。

$$\mathrm{C(q)} = \int_{-\infty}^{\infty} \log_{10} \mathrm{F(\omega)e^{-iq\omega}d\omega}. \quad (2)$$

2.1. メル周波数ケプストラム係数

ヒトは音高の低い音ほど音高の差異に敏感で、音高の高い音ほど音高の差異を認識できない。この音高差に関する感度の尺度をメル尺度（mel-scale）[3]と呼ぶ。周波数fをメル尺度f_{mel}に変換する式を式(3)に示す。

図 1:MFCC 算出の流れ。

$$f_{mel} = \alpha \times \log_{10}\left(1 + \frac{f}{700}\right). \qquad (3)$$

メル周波数ケプストラム係数（MFCC）は、ケプストラムの低次成分に対し、メル尺度を重み付けすることにより得ることができる。MFCC 算出の流れを図 1 に示し、以下で説明する。

STEP1：短時間フーリエ変換

音響信号に対し重み付けと短時間フーリエ変換を行い、スペクトラムを求める。短時間フーリエ変換の窓幅は 4,096 点がよく用いられるので[2]、本研究でも 4,096 点の窓幅を採用する。重み付けに用いる窓関数には式(4)に示すハミング窓を用いる。

$$w(x) = 0.54 - 0.46\cos 2\pi x \qquad (0 \le x \le 1). \qquad (4)$$

STEP2：メルフィルタバンク

メルフィルタバンクは、バンドパスフィルタをオーバーラップさせながら並べて構成されている。バンドパスフィルタの数をチャネル数と呼ぶ。チャネル数には経験的に 20 が用いられる[3]。20 個あるバンドパスフィルタ毎にスペクトラムの和をとることで、スペクトラムを 20 次元のベクトルで表す。さらにベクトルの各要素の対数をとり、メル周波数スペクトラム（MFS）へ変換する。

STEP3：離散コサイン変換

離散コサイン変換は短時間フーリエ変換と同様に、離散波形信号を周波数領域へ変換する。メル周波数スペクトラムに対し離散コサイン変換を行うことにより、メル周波数スペクトラムをメル周波数ケプストラム（MFC）へ変換する。

STEP4：低次成分の抽出

メル周波数ケプストラムの低次成分を取り出す。この低次成分の次元数は、本研究では 12 を用いる。メル周波数ケプストラムの低次成分を取り出したものが MFCC である。MFCC は、3 章で説明する SVM への入力特徴量として用いる。

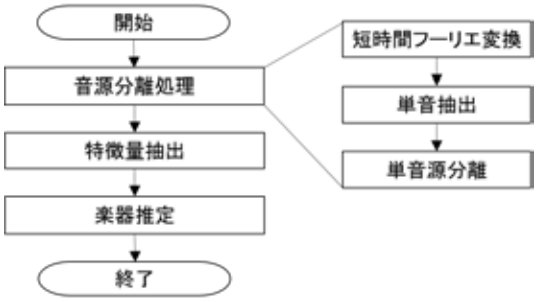

図 2:提案手法の流れ。

3. Multi-class SVM

3.1. Support Vector Machine

サポートベクタマシン（SVM : Support Vector Machine）は、入力されたパターンを 2 クラスに分類する識別器を構成する手法である。SVM は、サポートベクトルと呼ばれる、クラスの境界近傍に位置する学習ベクトルからクラス境界までの距離を最大化するように分離超平面を構築し、クラス分類を行う。このとき、サポートベクトルと分離超平面との距離をマージンと呼ぶ。

SVM はパターン認識の分野でよく用いられている階層型ニューラルネットワークに比べ汎化能力に優れるほか、2 次の最適化問題として定式化されるため、学習により最適解が得られることが保証されている。

3.2. Multi-class SVM

SVM を多クラスの識別問題を対象にできるよう拡張したものが、多クラス SVM（Multi-class SVM）である[4]。

多クラスの識別問題を 2 クラスの分類器で扱うためには、複数の 2 クラス識別モデルを組み合わせることになる。この組み合わせ方として、一対一分類法（One-versus-One）と呼ばれる方法がある。これは、あるクラスとそれ以外のクラスへの分類を全クラスに対し行い、得られた複数の境界で分離超平面を構築する方法である。

本研究では、楽器推定を行う識別器を Multi-class SVM を用いて構成する。

4. 提案手法による楽器推定

提案手法の流れを図 2 に示し、以下で説明する。

STEP1：前処理

入力された音響信号に対して短時間フーリエ変換を行いスペクトラムを求める。短時間フーリエ変換には 4,096 点の窓幅を採用し、窓関数はハミング窓を使用する。また、次ステップの音源分離のため、楽曲中に含まれる各単音の発音時刻、消音時刻、音高を抽出する。なお、本研究では、これらはそれぞれ所与のものとして扱う。

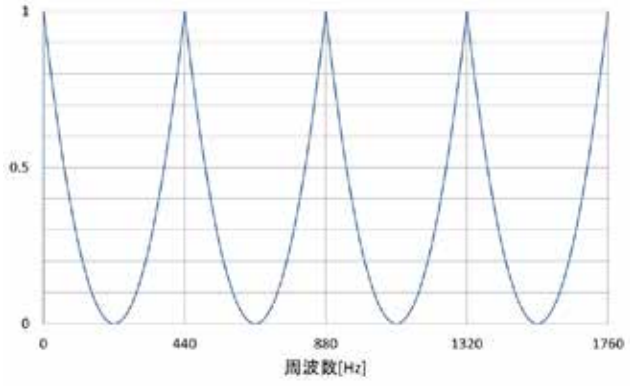

図 3: A4(=440Hz) に対するくし型フィルタ。

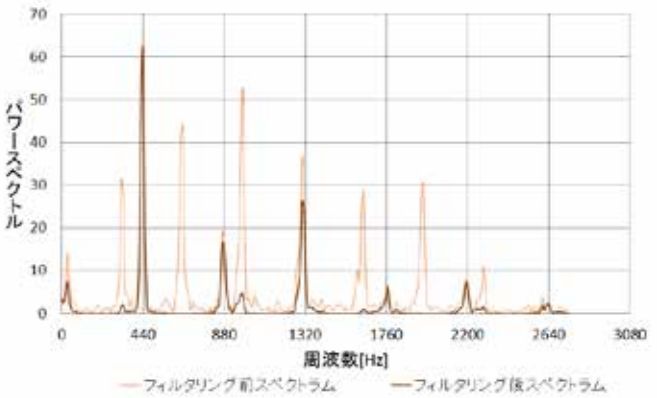

図 4: フィルタリングによるスペクトラムの変化。

STEP2：音源分離処理

単音の音高の周波数とその整数倍の周波数にピークを持つくし型フィルタを用い音源分離を行う。例として、ヴァイオリンを音高 A4（440Hz）で 1 秒間、トランペットを音高 E4（約 330Hz）で 1 秒間同時に演奏した楽器音のスペクトラムに対し、くし型フィルタによりヴァイオリンの楽器音のスペクトラムのみを抽出することを考える。ヴァイオリンの音を抽出する場合、図 3 に示す音高 A4 に対応するくし型フィルタにより、ヴァイオリンの楽器音スペクトラムのみを抽出する。2 楽器を同時演奏した楽器音のスペクトラムに対し、図 3 のくし型フィルタを作用させる前と後のスペクトラムの比較を図 4 に示す。図 4 に示すように、ヴァイオリンの基本周波数ピークとその倍音のピークのみを抽出できていることが分かる。

STEP3：特徴量抽出

音源分離処理により分離した全単音に対し、特徴量として 12 次元 MFCC を抽出する。

STEP4：楽器推定

楽器推定には、単音源の学習データをあらかじめ学習させた Multi-class SVM を用いる。本研究で対象とする楽器とその音域、音長を表 1 に示す。楽器音は MML で記述し MIDI 出力したものをシンセサイザーを用いて wav 形式に出力したものを使用する。これらの音源の 12 次元 MFCC を Multi-class SVM に学習させることにより、楽器識別モデルを作成する。

表 1: 特徴量に使用する単音源。

楽器	Clarinet（Cl）、Classic guitar（Cg）、Electric bass（Eb）、Piano forte（Pf）、Trumpet（Tr）、Violin（Vn）
音域	3 オクターブ 36 音階 C3（約 130.8[Hz]）～B5（約 2,093.0[Hz]）
音長	1 秒

図 5: 使用する楽曲の楽譜。

5. 評価実験

5.1. 評価実験

表 1 に示されている音源すべてについて、それぞれ 12 次元 MFCC を算出し、Multi-class SVM の学習データに用いる特徴量とする。テスト用音源は 2 種類の楽器を同時に発音させた楽曲に対し音源分離処理を行い、1 種類の楽器音からなる単音源に分離したものを用いる。具体的には、2 種類の楽器を同時に発音させた楽曲に短時間フーリエ変換を行いスペクトラムに変換する。次に、各単音の開始時刻から終了時刻までに対応するスペクトラムに対し、単音の音高とその倍音にピークを持つくし型フィルタを作用させ、単音の成分のみを抽出する。最後に、くし型フィルタを通過したスペクトラムに対し逆フーリエ変換を行い、1 種類の楽器音からなる単音源へと変換する。

使用する楽曲の楽譜を図 5 に示す。この楽譜を、表 1 に示されている楽器 6 種類のうち 2 種類を用いて演奏し、実験に使用する楽曲とする。

5.2. 実験結果

楽器推定の精度は、式(5)で示す正解率で評価する。

$$\text{正解率} = \frac{\text{正解率}}{\text{テストデータ総数}} \times 100. \qquad (5)$$

各楽器ごとの正解率を図 6 に示す。図 6 より、トランペット、クラリネット、ヴァイオリンの 3 楽器については 50%以上の正解率が得られている。一方、クラシックギター、エレクトリックベース、ピアノの 3 楽器については正解率が 50%に達していない。

図 6 の結果のうち、最も正解率の高いヴァイオリンについて、推定された楽器別の内訳を図 7 に示す。図 7 より、ヴァイオリンはピアノ、クラリネットとして認識されなかったことがわかる。また、誤って認識された回数

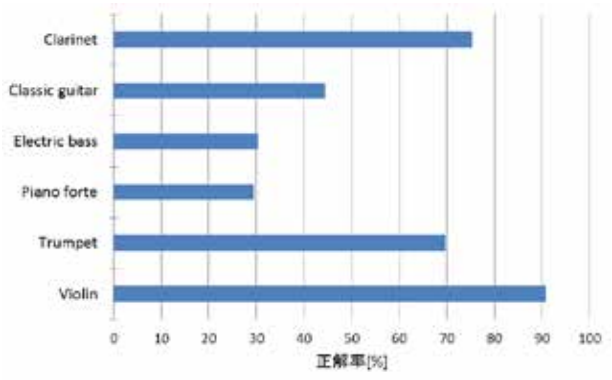

図 6: 各楽器ごとの正解率。

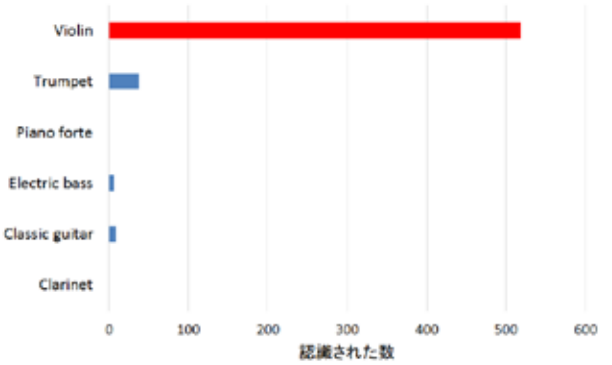

図 7: ヴァイオリンの推定結果の内訳。

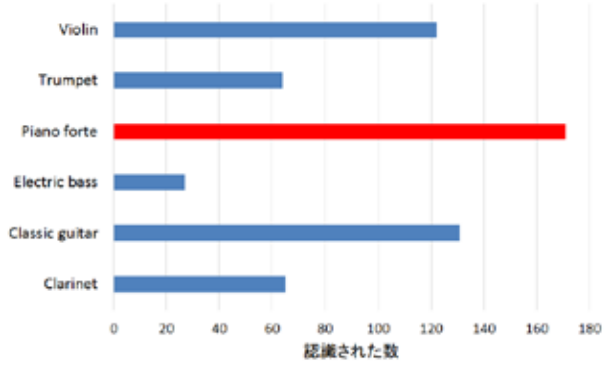

図 8: ピアノの推定結果の内訳。

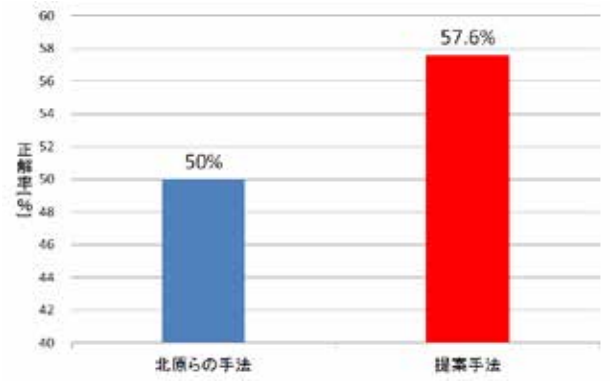

図 9: 総合的な正解率の比較。

が最も多いトランペットについても、誤認識率は 7%以下にとどまっていることがわかる。一方で、最も正解率の低いピアノがどの楽器音と推定されたかの内訳を示した図 8 より、ピアノはヴァイオリンやクラシックギターと誤認識されることが多く、エレクトリックベースと誤認識されることが少ないことがわかる。

また、図 6 に示した各楽器ごとの正解率を平均した総合的な楽器推定の正解率と、本研究でベースとした北原らの手法による正解率との比較を図 9 に示す。図 9 より、提案手法は北原らの手法に比べ、正解率が 7.6%向上している。

5.3. 考察

図 6 より、各楽器ごとの正解率に大きな差異があることが分かる。比較的正解率の高い楽器であるトランペット、クラリネット、ヴァイオリンの 3 楽器と、比較的正解率の低い楽器であるクラシックギター、エレクトリックベース、ピアノの 3 楽器の相違点を考えてみると、音の持続性の観点から、前者は持続楽器に、後者は減衰楽器に分類されることに気付く。また図 7 より、持続楽器のヴァイオリンは持続楽器、減衰楽器を問わず、他の楽器と誤認識されることが少ないことが分かる。一方、図 8 より減衰楽器のピアノは持続楽器のヴァイオリン、トランペット、クラリネットと誤認識されることが多いことが分かる。このことから、持続楽器の正解率が高く減衰楽器の正解率が低いのは、持続楽器と減衰楽器の混合音の場合では、減衰楽器の成分を抽出した場合であっても、減衰楽器の音は時間経過とともに減衰してしまい、持続楽器の除去しきれなかった成分が多く残留することが原因と考えられる。

6. おわりに

本研究では、複数の楽器音が混在する音響信号をそれぞれの単楽器音へ分離し、その単楽器音が何の楽器の音であるかを識別することを目的とした。そのため、北原らが提案した MFCC を特徴量とした SVM による楽器推定法に、くし型フィルタを用いた音源分離手法を組み合わせる手法を提案した。

提案手法について、北原らの手法による楽器推定の正解率と提案手法による楽器推定の正解率との比較や、各楽器ごとの正解率の比較を行なった。実験の結果、提案手法は北原らの手法より 7.6%高い楽器推定の正解率を得ることができた。また、各楽器での正解率の比較から、楽器音の持続性が識別率に大きな影響を与えているという示唆を得た。

今後の課題としては、楽器音の持続性を考慮して正解率に偏りが発生しないような改良を施す必要がある。

参考文献

1) 三輪多恵子, 田所嘉昭, 斎藤努, “くし形フィルタを用いた採譜のための音源分離”, 電子情報通信学会

技術研究報告. SST, スペクトル拡散, Vol. 97, No. 169, pp. 61‐66 (1997).

2) 北原聡志, 甲藤二郎, “楽器の階層的分類を考慮した SVM による音源同定”, 電子情報通信学会総合大会講演論文集, Vol. 2006, No. 1, p.152 (2006).

3) S. Davis, I. Signal Technology, C. Santa Barbara and P. Mermelstein, “Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences”, Acoustics, Speech and Signal Processing, IEEE Transactions on, Vol.4, pp.357–366 (2003).

4) 栗田哲平, 近山隆, “多クラス Support Vector Machine を用いた一般物体認識での複数候補提示下における分類性能の傾向”, 電子情報通信学会技術研究報告. MVE, マルチメディア・仮想環境基礎, Vol. 108, No. 328, pp. 251‐258 (2008).