

# サポートベクターマシンを用いた 演奏楽器推定に関する研究

函館工業高等専門学校 情報工学科

花田 慧悟, 東海林 智也

# 目次

1 章 はじめに .....	1
1-1. 目的 .....	1
1-2. 本論文の構成 .....	2
2 章 音について .....	3
2-1. オクターブと音名について .....	3
2-2. 音の3要素と倍音特性 .....	3
3 章 Resonator型くし形フィルタ .....	4
3-1. Resonator型くし形フィルタとは .....	4
3-2. 利得の導出 .....	5
4 章 MFCCによる特徴抽出 .....	6
5章 SVMによるクラス分類 .....	7
5-1. SVMについて .....	7
5-2. 線形SVMと非線形SVM .....	7
6 章 演奏楽器推定システムの構成 .....	9
7 章 実験1 (単音での演奏楽器推定) .....	10
7-1. 実験概要 .....	10
7-2. 結果 .....	10
7-3. 考察 .....	10
8 章 実験2 (2和音での演奏楽器推定) .....	11
8-1. 実験概要 .....	11
8-2. 結果 .....	11
8-3. 考察 .....	11
9 章 まとめ .....	13
参考文献 .....	14

# Estimation of Musical Instruments Using SVM

Hakodate National College of Technology

Keigo Hanada, Tomoya Tokairin

**Abstract :** This paper describes a system to estimate musical instruments. First, this system builds SVMs (Support Vector Machines) from MFCCs (Mel-Frequency Cepstrum Coefficients) of each musical instruments that is obtained from the resonator comb filter's output. Secondly, the system calculates MFCCs of a music that are also obtained from the resonator comb filter's output. Finally, the system estimates musical instruments that are included in the music by using the SVMs. In order to verify the effectiveness of the system, we have conducted several computer simulations.

**Key words :** resonator comb filter, MFCC, Support Vector Machines

# 1章 はじめに

## 1-1. 目的

現在，楽曲検索，楽曲推薦，自動作曲等の音楽情報処理技術の精度向上が望まれている[1][2][3]．そのためには演奏されている楽曲から採譜をおこなって楽曲情報をデータベース化して管理する必要があるが，世の中に存在する楽曲数は膨大であるため，これを人の手で行うには音楽に関する知識と多くの時間を要する．そこで自動採譜システムを用いて自動的に楽曲のデータベース化を行うことが好ましいと思われる．自動採譜システムの構築においては，音高，音長，拍子等の推定が必要であるが，それと同時に演奏楽器の推定もおこなう必要がある[4][5]．

従来の研究では，Resonator 型くし形フィルタ[6][7]を用いて楽曲の実演奏音声から音源を分離した後で MFCC[8]を求め，マハラノビスの距離[9][10]によるテンプレートマッチングを行って演奏楽器を推定する手法が提案されているが，満足のいく結果が得られていない[7]．そこで本研究では，マハラノビスの距離の代わりにサポートベクターマシン(以下 SVM : Support Vector Machines)[11]を用いて演奏楽器を推定するシステムを開発する．開発するシステムの処理の流れを図1に示す．

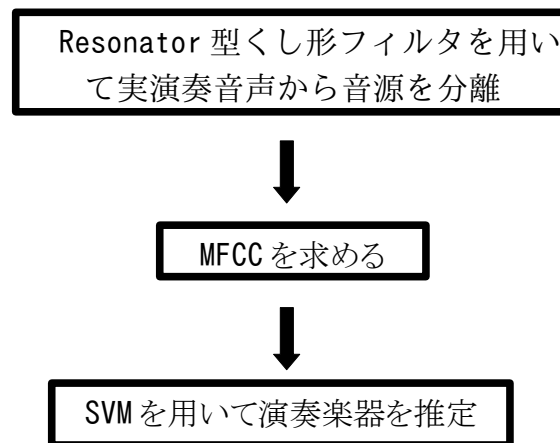


図1. 演奏楽器推定の流れ

更に，提案した手法の有用性を調べるために RWC 研究用音楽データベース[12]に収録されている実演奏音を用いてシミュレーション実験を行う．RWC 研究用音楽データベースは，研究者が研究目的に利用するために用意された音楽情報処理研究用 DB である．

## 1-2. 本論文の構成

1章では本論文の背景，目的，全体の内容の説明を行っている。

2章では音についての基礎的知識の説明を行っている。

3章では，Resonator型くし形フィルタについての説明を行っている。

4章ではMFCCによる特徴抽出の説明をおこなう。

5章ではSVMについての説明をおこなう。

6章では具体的な演奏楽器推定方法について説明をおこなう。

7章では単音での演奏楽器推定の条件・結果・考察を書いている。

8章では2和音での演奏楽器推定の条件・結果・考察を書いている。

9章では本論文のまとめをおこなっている。

## 2章 音について

### 2-1. オクターブと音名について

音名は日本語の五十音や英語のアルファベットと同じように、1つ1つの音につけられた名前である。本論文では英語表記を使って音名を表す(表1)。

またCからC, AからA, のように一回り上または下の同じ音までの間を1オクターブという[13]。ピアノの1番左(低い音)のオクターブはオクターブ1となりピアノの中央のオクターブはオクターブ4となる。

表1. 音名の表

イタリア表記	ド	ド#	レ	レ#	ミ	ファ	ファ#	ソ	ソ#	ラ	ラ#	シ
英語表記	C	C#	D	D#	E	F	F#	G	G#	A	A#	B

### 2-2. 音の3要素と倍音特性

音は音色, 音量, 時間的特性の3つの要素を持つ。音色は基本周波数とその倍音成分による倍音構造で決定されている[14]。倍音とは基本周波数の整数倍の音のことである(図2)。図2を見てわかるように倍音成分は基本周波数の整数倍の周波数であり楽器によりそのスペクトル構造が異なる。音量は音波の振幅の大きさによって決まる。時間的特性は, 楽器が演奏されてからの音色や音量などの時間的変化のことをいう。なお音高は基本周波数によって決まる[15]。また, 基本周波数の周波数が2倍になるとオクターブが1つ上がる。

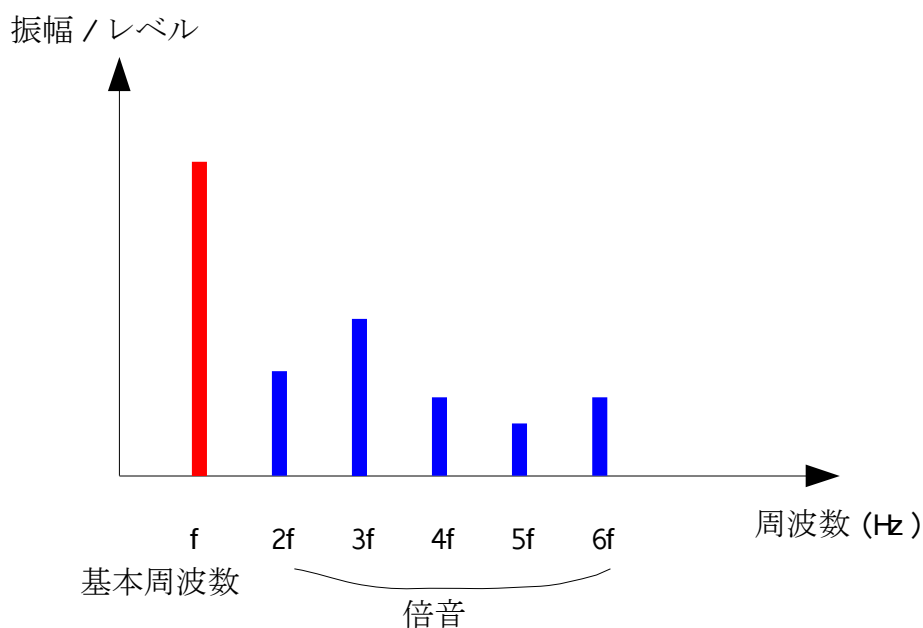


図2. 倍音構造の例

### 3章 Resonator型くし形フィルタ

#### 3-1. Resonator型くし形フィルタとは

オクターブを整数  $i$ ，音高番号を整数  $p$  とすると，任意の音は  $i$  と  $p$  の組み合わせ  $(i, p)$  で表される．また， $(i, p)$  音の基本周波数を  $f_{i,p}$  (Hz) とし，サンプリング周波数を  $f_s$  (Hz) とする．このとき  $(i, p)$  音に対応する正整数の定数として  $N_{i,p}$  を  $[f_s/f_{i,p}]$  ( $[\ ]$  : 整数への端数処理) と定義すると  $(i, p)$  音に対応する Resonator 型くし形フィルタは次の伝達関数で表される [6][7]．

$$H(i, p) = \frac{1-a}{1-a \cdot z^{-N_{i,p}}} \quad (1)$$

このフィルタは時刻  $n$  における出力音声  $y_{i,p}(n)$  を  $N_{i,p}$  時刻だけ遅延させて入力音声に加算する単純な無限インパルス応答フィルタであり，その利得特性は図3に示される．図3から，Resonator 型くし形フィルタは  $(i, p)$  音に含まれる基本周波数とその倍音成分以外の周波数成分を減衰させる特性を持つことが分かる．なお，フィルタ係数  $a$  ( $0 \leq a < 1$ ) は任意に設定できるパラメータであり， $a$  が 1 に近づくほど基本周波数とその倍音成分周辺の利得が急になり 0 に近づくほど平坦になる性質を持つ．

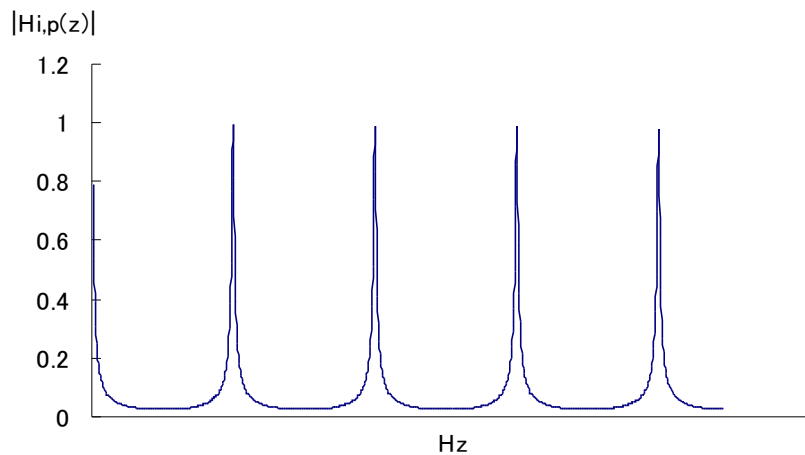


図3. Resonator 型くし形フィルタの利得特性の例 ( $a = 0.9$ )

### 3-2. 利得の導出

伝達関数(1)式より図3のような利得が得られることを証明する。(1)式の  $z$  を  $e^{-j\omega\tau}$  に置き換えてフーリエ変換を求める。

$$H(\omega t) = \frac{1-a}{1-a \cdot e^{-j\omega\tau N_{i,p}}} \quad (2)$$

(2)式の両辺の絶対値を取って利得を得る。

$$|H(\omega t)| = \left| \frac{1-a}{1-a \cdot e^{-j\omega\tau N_{i,p}}} \right| \quad (3)$$

ここでオイラーの公式より

$$|H(\omega t)| = \frac{1-a}{\sqrt{1-2a\cos\omega\tau N_{i,p}+a^2}} \quad (4)$$

となるため図3のような利得が得られる。



## 4 章 MFCCによる特徴抽出

MFCC( Mel-Frequency Cepstrum Coefficients : メル周波数ケプストラム係数 )とは音声認識に一般に用いられる特徴量で, 計算が簡素で処理速度が速いという利点がある. 普通の音声データではデータ数が膨大で計算処理をするのが難しいが, MFCCを求めることによりデータ数を大幅に減らすことができ, 計算処理が容易に行えるようになる. MFCCの計算手順は以下の通りである[8].

1. 周波数軸上にL個の三角窓を配置し, フィルタバンク分析によりスペクトル分解を行う(図4). 図中の $l$  は フィルタバンクの番号を示す
2.  $L$ 個の帯域におけるパワーに対してDCT( Discrete Cosine Transform : 離散コサイン変換)を行う

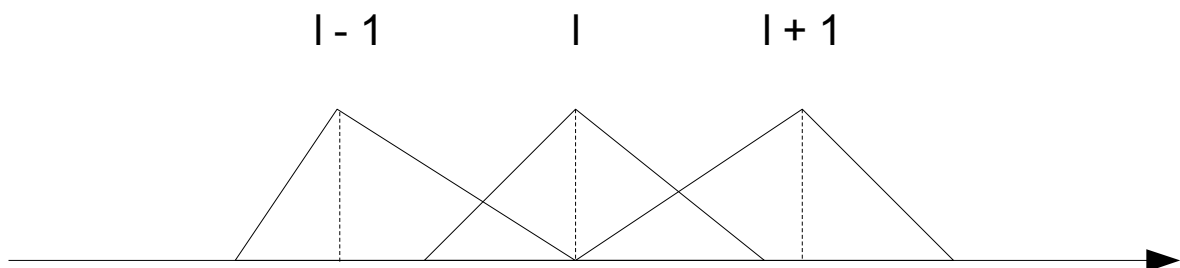


図4. フィルタバンク分析の例

## 5章 SVMによるクラス分類

### 5-1. SVMについて

SVM ( Support Vector Machines : サポートベクターマシン ) は, パターン認識の学習モデルの1つである. SVMは2クラスの識別器を構成する手法であり, 未学習データに対して高い識別性能を持っている [11]. SVMは与えられた学習用サンプルからマージン最大化という方法で判別曲線を学習する(図5). マージンとは判別曲線とデータとの距離である. 学習して得られた判別曲線を用いて, 未学習データのクラス分類をおこなう.

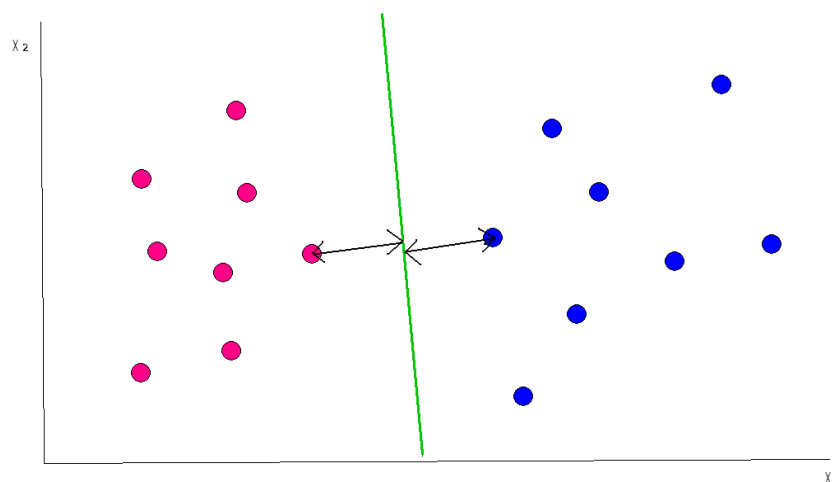


図5. SVMによる2次元データの分類の例

### 5-2. 線形SVMと非線形SVM

線形SVMの識別関数は次のように定義される[11].

$$f(\mathbf{x}) = \langle \mathbf{w} \cdot \mathbf{x} \rangle + b \quad (5)$$

$\mathbf{x}$  は入力ベクトル, ベクトル  $\mathbf{w}$  とスカラー  $b$  は関数を決定するパラメータであり, ラグランジュ乗数法によってこの  $\mathbf{w}$  と  $b$  を求める.

しかし, 線形SVMでは線形分離不可能なデータに対する識別能力に限界がある. そこで, カーネル関数を用いた非線形SVMを導入することで線形分離不可能なデータに対しても高い精度での識別が可能になる. 非線形SVMの識別関数は次のように定義される [11].

$$f(\mathbf{x}) = \sum_i \alpha_i y_i K(\mathbf{x}_i, \mathbf{x}) + b \quad (6)$$

スカラー  $\alpha_i, b$  は関数を決定するパラメータである。  $y_i$  はクラスを識別する既知のパラメータであり 1 及び -1 の値を取る。 また  $K(\mathbf{x}_i, \mathbf{x})$  をカーネル関数といい、ベクトルの内積を一般化したものである。カーネル関数にはガウス関数が用いられる事が多い。ガウス関数は次のように定義される。この  $\sigma^2$  は分散を表す。

$$K(\mathbf{x}, \mathbf{z}) = \exp(-\|\mathbf{x} - \mathbf{z}\|^2 / \sigma^2) \quad (7)$$

## 6 章 演奏楽器推定システムの構成

本研究で提案する楽器推定手法の手順は以下のとおりである。

1. ある時間間隔ごとに，並列接続された全てのフィルタ出力  $y_{i,p}(n)$  を求める(図6)．時刻  $n$  における入力音声を  $x(n)$  としたとき，全ての  $(i, p)$  音に対応した Resonator 型くし形フィルタ  $H_{i,p}(z)$  を並列接続することで，入力音声に含まれる  $(i, p)$  音が  $y_{i,p}(n)$  として分離されて出てくる．
2. 各  $y_{i,p}(n)$  から MFCC を求めて  $(i, p)$  音別にクラス分類を行う．今回はクラス分類手法としてフィルタ出力の MFCC とテンプレート音声の各楽器音から求めた MFCC を用いた SVM によるクラス分類をおこなう．分類された楽器を現在演奏中の楽器としてその楽器番号を記録する．
3. 楽曲内で短時間の間に頻繁に楽器が切り替わることは通常考えられないため，フレーズ単位で楽器推定を行う．フレーズ終了時に記録した楽器番号の頻度を計算し，高い頻度順に楽器名を挙げる．

なお，入力された演奏音声は Resonator 型くし形フィルタの利得の影響を受けるため，通常の楽器演奏音声からそのまま MFCC を求めてテンプレートを作成して用いることは出来ない．その代わりに，各楽器の  $(i, p)$  音を Resonator 型くし形フィルタ  $H_{i,p}(z)$  に通して出てきた出力音声から求めた MFCC をその楽器の  $(i, p)$  音に対するテンプレートとして用いる．

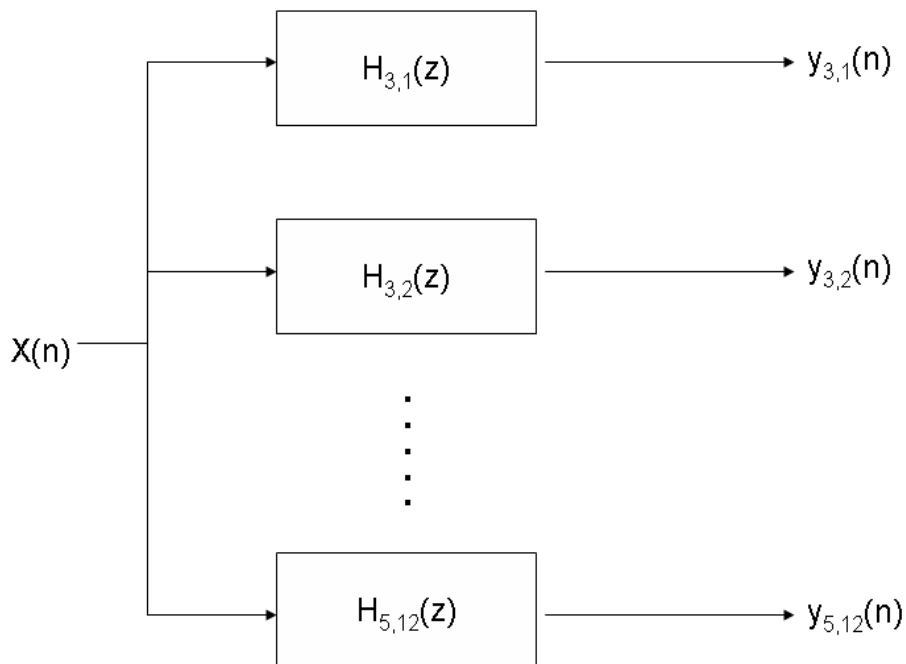


図6. Resonator 型くし形フィルタの並列接続例

## 7 章 実験1 (単音での演奏楽器推定)

### 7-1. 実験概要

始めに単音での楽器推定実験を行なった。実験対象の楽器はアルトサクソ(AS), コントラバス(CB), クラリネット(CL), フルート(FL), ホルン(HR), ピアノ(PF), トランペット(TR), チェロ(VC), ビオラ(VL), バイオリン(VN)の10楽器とした。実験に使用した音声はRWC研究用音楽データベース[12]に収録されている通常奏法, 音の強さ”強”の実演奏楽器音声(44.1Hz, 16bit, モノラル)であり, テンプレートに用いた音声と同じものである。使用音はオクターブ4のCからBまでである。なお, RWC研究用音楽データベースに含まれる楽器音は最低音から最高音までつながっているため, 各音高ごとにWAVEファイルを分割して用いた。

### 7-2. 結果

結果を表2に示す。表中の数字は推定頻度(全推定回数からその楽器であると推定された回数の割合)を表す。表の行は推定された楽器, 列は推定対象の楽器である。表中の灰色の項目は正しく推定された楽器を示している。また, 「未分類」はSVMによる分類が失敗したものである。

表2. 単音での推定結果(%)

	AS	CB	CL	FL	HR	PF	TR	VC	VL	VN	未分類
AS	100	0	0	0	0	0	0	0	0	0	0
CB	0	97.8	0	0.9	0	0	0	0	0	0.9	0.4
CL	0	0	99.2	0	0	0	0	0.8	0	0	0
FL	0	0	0	99.7	0	0	0.3	0	0	0	0
HR	0	0	0	0	99.7	0	0.3	0	0	0	0
PF	0	0.7	0	1.1	0	95.7	0	0	1.8	0	0.7
TR	0	0	0	0	1.1	0	97.7	0	0	1.1	0
VC	0	0.5	0	0	0	0	0	98.4	0	1.1	0
VL	0	0	0	0.3	0.3	0	0	0	99	0	0.3
VN	0.5	0	0	0	0	0.5	0	0	0	98.4	0.5

### 7-3. 考察

全ての楽器で高い精度で楽器推定が成功した。

## 8 章 実験2 (2和音での演奏楽器推定)

### 8-1. 実験概要

次に2和音での楽器推定実験を行なった。実験対象の楽器はアルトサククス(AS), コントラバス(CB), クラリネット(CL), フルート(FL), ホルン(HR), ピアノ(PF), トランペット(TR), チェロ(VC), ビオラ(VL), バイオリン(VN)の10楽器とした。実験に使用した音声はRWC研究用音楽データベース[12]に収録されている通常奏法, 音の強さ”強”の実演奏楽器音声(44.1Hz, 16bit, モノラル)であり, テンプレートに用いた音声と同じものである。その内オクターブ4のCからBまでの楽器音を用いて2和音のwaveファイルを作成した。更に楽器の特性により音域により音量が変化するため音量をノーマライズして実験を行なった。

### 8-2. 結果

結果を表3, 表4に示す。表3の各行は対象楽器の組み合わせを示し, 各列は推定対象の楽器名, 表中の数字は推定頻度(全推定回数からその楽器であると推定された回数の割合)を示している。「未分類」の列は, SVMによる分類が失敗したものである。ここで, 表3中の灰色の項目は正しく推定された楽器を示し, 黒枠で囲まれた項目は誤って推定された楽器を示している。表4は楽器個別の推定頻度と全体での平均推定頻度を示している。なお表4の推定頻度は「未分類」を除いて求めた。

### 8-3. 考察

実験の結果, 全楽器での正解率として88.89%と高い推定頻度が得られた。しかし, ホルン(HR), ピアノ(PF), トランペット(TR)に関しては推定頻度が50%以下になるものもあり, 高い推定頻度が得られなかった。また, SVMによるクラス分類の失敗の割合も全体的に高くなった。これらの原因としては, Resonator型くし形フィルタによる和音の分離が上手くいっていないため, 推定が失敗したことが考えられる。

表3. 2和音での推定結果

	AS	CB	CL	FL	HR	PF	TR	VC	VL	VN	未分類
AS-CB	37.64	34.07	0.66	0.1	0	0	0	5.03	0.31	0.69	21.49
AS-CL	58.38	0.43	25.01	0.54	0	0	0.08	3.56	0.35	2.75	8.9
AS-FL	53.38	0.64	1.28	25.84	0	0	0	0.86	0.86	1.97	15.17
AS-HR	67.02	1.71	0.25	0.87	16.89	0	0.36	0.29	0.69	2.77	9.14
AS-PF	87.7	1.67	0.65	1.02	0	0	0	0.14	0.79	0.93	7.1
AS-TR	82.31	0	0	0.04	0	0	15.97	0.04	0	0.55	1.09
AS-VC	56.9	0.73	1.1	0.73	0.16	0	0	19.7	0.16	1.1	19.44
AS-VL	57.84	0.29	0	0.03	0	0	0	0.06	26.73	4.78	10.26
AS-VN	71.21	0.09	0.14	0.05	0	0	0.05	0.51	0	25.36	2.6
CB-CL	0.28	51.95	25.66	1.46	0.05	0	0	2.87	0.05	0	17.68
CB-FL	0.23	30.29	1.16	50.23	0	0.04	0.01	4.09	0.64	0.11	13.2
CB-HR	0.39	60.78	2.25	6.87	6.96	0.6	0.15	1.42	2.64	0.04	17.9
CB-PF	0.36	80.83	1.21	3.8	0.03	1.64	0	0.82	3.9	0.16	7.24
CB-TR	1.38	31.45	0.37	4.38	0	0.1	7.4	4.92	10.06	3.72	36.22
CB-VC	0.04	69.1	2.77	1.24	0	0	0	14.12	0	0.07	12.66
CB-VL	2.08	30.05	0.85	5.21	0.12	0	0	1.17	27.95	4.44	28.12
CB-VN	0.75	47.64	1.13	2.44	0	0	0	6.1	3.32	16.14	22.49
CL-FL	0.09	1.4	17.75	64.46	0	0	0.01	3.42	0.31	0.12	12.44
CL-HR	0.28	2.6	23.49	17.67	13.32	0.15	0.13	4.07	2.69	0.24	35.35
CL-PF	0.9	4.37	59.77	4.27	0.6	15.01	0.13	3.87	2.47	0	8.61
CL-TR	8.23	0.16	17.66	3.66	0.36	0.18	14.72	2.23	6.88	2.43	43.49
CL-VC	0.59	2.59	64.57	2.07	0.07	0	0.11	21.24	0.48	0.18	8.09
CL-VL	1.84	0.83	10.4	4.91	0.12	0	0.02	1.36	43.93	1.2	35.38
CL-VN	4.23	2.15	33.11	0.9	0.03	0	0.12	3.17	1.56	26.04	28.69
FL-HR	0.01	2.13	0.36	77.25	1.54	0.3	0.35	0.09	1.72	0.28	15.96
FL-PF	0.23	1.91	1.34	86.39	0.02	0.42	0.08	1.12	2.01	0.37	6.1
FL-TR	0.31	0.08	0.07	52.62	0.01	0.07	4.9	0	7.04	3.88	31.03
FL-VC	0.26	1.85	3.21	66.85	0	0	0.06	12.36	0.15	0.5	14.75
FL-VL	0.07	0.27	0.2	56.61	0.03	0	0.12	0.14	30.21	0.68	11.68
FL-VN	0.61	0.36	1.15	68.03	0	0	0.11	0.63	2.7	16.25	10.17
HR-PF	0.06	0.81	0.03	1.55	72.97	16.77	0.77	0	2.29	0	4.74
HR-TR	0.02	0	0	3.56	4.97	0.04	64.02	0	13.8	0.07	13.53
HR-VC	0.61	9.14	9.75	9.71	28.41	1.11	0.73	17.78	0.84	0.11	21.8
HR-VL	0.08	0	0	1.3	7.71	0.15	0.44	0	82.04	0.69	7.58
HR-VN	4.11	1.16	0.28	8.43	15.53	0.67	1.47	0	12.16	30.79	25.4
PF-TR	0.22	0.64	0.03	1.79	1.32	6.89	79.42	0	3.61	0.03	6.05
PF-VC	1.16	17.74	10.61	11.11	0.25	0	0	22.97	3.81	0.25	32.09
PF-VL	0.22	0.32	0	0.35	0.38	1.67	0.19	0	92.56	0.4	3.91
PF-VN	5.19	3.42	0.22	7.34	0	0.83	0.11	0.17	14.18	50.28	18.27
TR-VC	4.02	0.88	2.61	4.23	0.04	0.04	29.28	18.36	4.3	0.28	35.98
TR-VL	0.09	0	0	0.41	0	0	18.85	0	66.7	2.79	11.16
TR-VN	8.4	0	0.03	2.46	0	0.06	22.95	0	3.26	30.21	32.64
VC-VL	4.47	1.09	2.57	3.44	0.06	0	0	10.21	46.04	2.27	29.85
VC-VN	2.06	0.31	1.62	0.87	0	0	0	28.74	2.62	40.15	23.63
VL-VN	2.5	0	0.05	0.24	0.35	0	0	0	70.24	18.97	7.65

表4. 楽器個別の推定精度

AS	CB	CL	FL	HR	PF	TR	VC	VL	VN	全体での正解率
100	100	100	100	66.67	44.44	77.78	100	100	100	88.89

## 9章 まとめ

本研究では、Resonator型くし形フィルタにより実演奏音声から音源を分離した後でMFCCを求め、SVMによるクラス分類を行って演奏楽器を推定する手法を提案した。また、10楽器音を用いて楽器推定実験を行った。実験の結果、単音、2和音両方の場合に関して高い推定精度が得られた。

今後の課題としては、音源分離の精度を向上させることが挙げられる。また、実際の楽曲に対する演奏楽器推定実験を行うことが挙げられる。



## 参考文献

- [1] 帆足啓一郎, 上月勝博, 菅谷史昭 : 楽曲配信サービスを支える音楽情報検索技術, 電子情報通信学会誌, Vol. 49, No. 7, pp.529-534, 2005.
- [2] 藤原弘将, 後藤真孝 : VocalFinder : 声質の類似度に基づく楽曲検索システム, 情報処理学会研究報告, MUS-71, pp.27-32, 2007.
- [3] 吉井和佳, 後藤真孝, 駒谷和憲, 尾形哲也, 奥乃博 : 楽曲推薦システムの効率性とスケーラビリティの改善のための確率的推薦モデルのインクリメンタル学習法, 情報処理学会研究報告, MUS-71, pp.19-26, 2007.
- [4] 東海林智也, 北見伸一郎 : くし形フィルタと自己相関関数による演奏楽器推定, 情報処理学会第70回全国大会講演論文集 ( 2 ), pp.429-430, 2008.
- [5] 馬場貴之, 山田武志, 北脇信彦 : HMMとMFCCを用いた楽器音の音源同定の検討, 情報処理学会研究報告, MUS-61, pp.79-84, 2005.
- [6] 藤原道, 山口満, 斎藤努, 田所嘉昭 : Resonator型くし形フィルタによる打楽器音を含む楽器音の音高推定法の検討, 電気情報通信学会技術研究報告, EA2003-80, pp.19-23, 2003.
- [7] 室谷良平, 東海林智也 : Resonator 型くし形フィルタを用いた演奏楽器推定手法, 第79回音楽情報科学研究会, MUS-79, pp.43-46, 2009.
- [8] 鹿野清宏, 伊藤克亘, 河原達也, 武田一哉, 山本幹雄 編著 : 音声認識システム, オーム社出版, 2001.
- [9] 武藤真介 : 初等多変量解析, 朝倉書店, 1999.
- [10] 長谷川勝也 : ゼロからはじめてよくわかる多変量解析, 技術評論社, 2004.
- [11] Nello Cristianini, John Shawe-Taylor 著, 大北剛 訳: サポートベクターマシン入門, 共立出版, 2005.
- [12] 後藤真孝, 橋口博樹, 西村拓一, 岡隆一 : RWC研究用音楽データベース : 研究目的で利用可能な著作権処理済み楽曲・楽器音データベース, 情報処理学会論文誌, Vol. 45, No. 3, pp.728-738, 2004.  
<http://staff.aist.go.jp/m.goto/RWC-MDB/index-j.html>

[13] 水谷友香, キッシー岸田 : CD で覚えるやさしい楽譜の読み方, 成美堂出版, 2005.

[14] 宮本賢一, 亀岡弘和, 西本卓也, 小野順貴, 嵯峨山茂樹 : 調波構造・時間包絡・音色の統合的クラスタリングによる楽音分析, 情報処理学会研究報告, MUS-71, pp.155-160, 2007.

[15] 春日正男, 船田哲男, 林伸二, 武田一哉 : 音声情報処理, コロナ社, 2001.