

卒業論文

Resonator型くし形フィルタを用いた 演奏楽器推定手法

2009年 3月 5日

東海林研究室

5年情報工学科28番 室谷良平

目次

1 章 はじめに	2
2 章 音について	4
2-1 音の要素	4
2-2 楽器音と倍音特性	4
2-3 RWC研究用音楽データベース	4
3 章 Resonator型くし形フィルタ	5
3-1 原理と性質	5
3-2 利得の導出	6
4 章 MFCCによる特徴抽出	8
5 章 マハラノビスの距離によるテンプレートマッチング	9
6 章 演奏楽器推定システムの構成	10
7 章 テンプレートと実験データの作成	11
8 章 実験1 (単音での演奏楽器推定)	12
8-1 実験概要	12
8-2 結果	12
8-3 考察	12
9 章 実験2 (2和音での演奏楽器推定 1)	13
9-1 実験概要	13
9-2 結果	13
9-3 考察	13
10 章 実験3 (2和音での演奏楽器推定 2)	15
10-1 実験概要	15
10-2 結果	15
10-3 考察	15
11 章 まとめ	17
参考文献	18
付録	20

1 章 はじめに

現在、携帯音楽プレーヤーや携帯電話の音楽再生機能の普及により音楽が今まで以上に身近なものとなった。いまや音楽はCDやレコードを買って楽しむのではなく、インターネットを介して楽曲をデータとして入手し楽しむ時代になっている。2002年から配信が開始された着うたは現在では年間2億曲を越える楽曲数がダウンロードされるようになり、場合によっては一曲の着うたが600万を越えるダウンロード数を誇るなど、90年代に見られたCDシングル・アルバムのミリオンセラー連発の現象が、現在では音楽配信の業界で起こっている。

また近年のテクノ・ハウス・エレクトロ等といった音楽ジャンルの過熱や、初音ミクといったボーカロイドの人気はDTM[1]の進歩が背景にある。DTMを用いれば楽器を演奏出来ない人でも比較的簡単に作曲して楽しむことが出来るため、SNSサイトなどで同じ音楽の趣味を持つ者同士が集まって自作の曲をアップロードするなど、音楽を多種多様に楽しむ時代になって来ている。

このように、いまや楽曲数はインターネット上に膨大な数が存在するため、それにつれて楽曲検索[2][3]、楽曲推薦[4]、自動作曲等の音楽情報処理技術のクオリティ向上が望まれている。そのためには採譜をおこなって楽曲情報を知識データベース化して管理する必要があるが、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。さらに世の中に存在する楽曲数は膨大であるため、自動採譜システムを用いて自動的にデータベース化を行うことが好ましい。自動的に採譜を行なうためには音高、拍子、音長推定の他にも演奏されている楽器の推定を行うことが必要である[5][6]。

そこで本研究では、Resonator 型くし形フィルタ[7]を用いて演奏音声から音源を分離した後でMFCC[6][8][20]を求め、マハラノビスの距離[9]を用いて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を提案する。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[14]に収録されている実演奏音を用いてコンピュータ実験を行う。

従来、楽器推定に用いられている手法としてNotch 型くし形フィルタ[10]やハーモニッククラスタリング[11][12]、調波・非調波統合モデル[13]による推定手法等が提案されている。ここで、本研究で提案する手法とNotch 型くし形フィルタを用いる手法との違いについてここで述べる。Notch 型は特定の調波成分を取り除く性質を持っているのに対して、Resonator 型は特定の調波成分を分離して取り出す性質を持っている。従って、Notch 型で特定の調波成分を抽出するためにはフィルタを多段縦続接続する必要があり低音域のノイズが発生する。一方、Resonator 型で特定の調波成分を抽出するためにはフィルタを並列接続するだけで済むためにノイズの発生を抑えることができる。さらに、Resonator 型は計算処理が単純であり、計算速度が速くなるという利点もある。

本論文の構成は以下のとおりである。

- 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と実験に用いた音源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おこなっている。原理と性質の説明，利得の導出過程を説明している。
- 4 章ではMFCCによる特徴抽出の説明を行う。
- 5 章ではマハラノビス距離によるテンプレートマッチングの説明をおこなう。
- 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 章ではテンプレートの作り方の説明をおこなう。
- 8 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。
- 9 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。
- 10 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。
- 11 章では本論文のまとめをおこなっている。

Musical Instruments Estimation System by Resonator Comb Filters and MFCC

This paper describes a system to estimate musical instruments. First, the system calculates MFCC of several musical instruments as template data. Next, the system also calculates MFCC of an input music through resonator comb filters. Then, the system calculates Mahalanobis distance between MFCC of the input music and MFCC of template data, and carries out template matching. As a result, the system estimates the musical instruments which is used in the music.

Key words : A Musical Instrument Estimation, Comb Filters, MFCC, Mahalanobis distance

2 章 音について

2-1 音の要素について

音は音色，音量，時間的特性の3つの要素を持つ。音色は楽器のもつ基本周波数と倍音構造により決まる。倍音とは基本周波数の整数倍の音のことであり，この違いにより楽器を聞き分けることが出来る。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。

2-2 楽器音と倍音特性

音高は基本周波数によって決まる。例えば440Hzはオクターブ4のAである。また，音の基本周波数が2倍になるとオクターブが1つ上がるため880Hzはオクターブ5のAとなる。しかし，音高が同じでも楽器により音色が異なるのは倍音構造のスペクトルが異なるからである[15]。倍音構造の例を図1に示す。

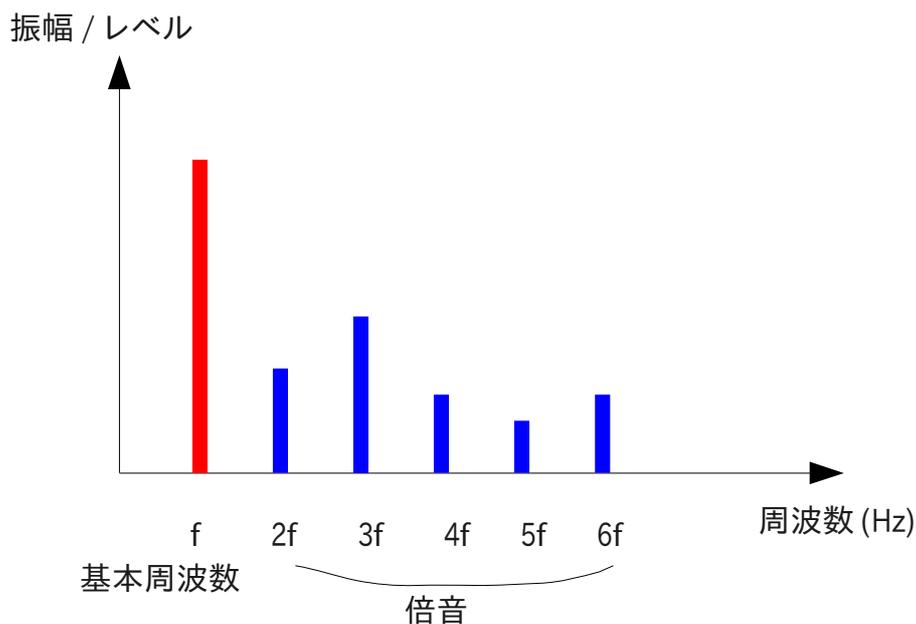


図 1. 倍音構造の例

2-3 RWC研究用音楽データベース

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上で，共通利用の自由，学術利用の自由が確保された音楽情報処理研究用DBである[14]。技術研究組合新情報処理開発機構RWC音楽DBサブワーキンググループにより構築され，産業技術総合研究所RWC音楽データベース管理責任者によって研究者へ実費配布されている。

3 章 Resonator型くし形フィルタについて

3-1 Resonatorくし形フィルタの原理と性質

オクターブを整数 i ，音高番号を整数 p ($p = 1, 2, \dots, 12$, $p = 1$ がC, 12 がBに対応) 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．例えば $(4, 3)$ はオクターブ4のD, $(3, 12)$ はオクターブ3のB音に対応する．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(1)で表される．

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

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2に示される．図2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 は $0 \leq a < 1$ の範囲の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．従って，時刻 n における入力音声 $x(n)$ 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) 音が分離されて出てくる．

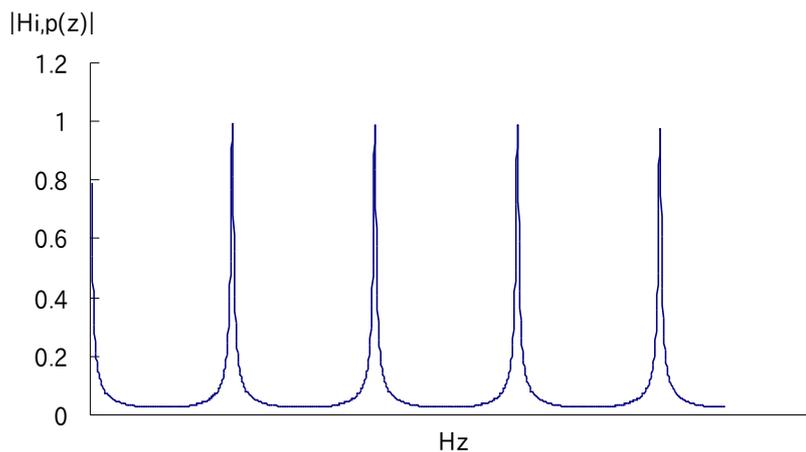


図2. Resonator 型くし形フィルタの利得特性の例 ($a = 0.9$)

2-2 Resonatorくし形フィルタの利得の導出

ここで(1)式より図2のような利得が得られることを説明する。始めに、フーリエ変換とラプラス変換，そしてz変換について説明する。フーリエ変換とは時間領域の関数 $f(t)$ を周波数領域の関数 $F(\omega)$ に変換するもので，定義は

$$F(\omega) = \int_0^{\infty} f(t) \cdot e^{-j\omega t} dt \quad (2)$$

となる。

しかし， $t \rightarrow \infty$ のとき $f(t)$ が収束せずにフーリエ変換の値が ∞ になるときがある。そこで $e^{-\sigma t}$ を掛けて $f(t) \cdot e^{-\sigma t}$ を収束させる。

すなわち

$$\lim_{t \rightarrow \infty} |f(t) \cdot e^{-\sigma t}| = \cdot \quad (3)$$

とする。この式をさらにフーリエ変換すると

$$\int_0^{\infty} f(t) \cdot e^{\sigma t} e^{-j\omega t} dt \quad (4)$$

となり，ここで

$$s = \sigma + j\omega \quad (5)$$

とおくと

$$F(s) = \int_0^{\infty} f(t) \cdot e^{-st} dt \quad (6)$$

という式が得られる。これがラプラス変換の定義である。

Z変換は離散時間信号 $f(t)$ をラプラス変換 $e^{s\tau}$ し，を z と置き換えたものである。 τ はサンプリング間隔である。ここで伝達関数(1)式はくし形フィルタのインパルス応答 $h[t]$ の z 変換であるので，(1)式の z を $e^{s\tau}$ に置き換えてラプラス変換(7)を求める。

$$H(s) = \frac{1-a}{1-a \cdot e^{-s\tau N_{i,p}}} , \quad (s = \sigma + j\omega) \quad (7)$$

次に、 $\sigma = 0$ を代入することで(7)式をフーリエ変換(8)にする。

$$H(\omega t) = \frac{1-a}{1-a \cdot e^{-j\omega\tau N_{i,p}}} \quad (8)$$

(8)式の両辺の絶対値を取って利得を得る。

$$|H(\omega t)| = \left| \frac{1-a}{1-a \cdot e^{-j\omega\tau N_{i,p}}} \right| \quad (9)$$

ここでオイラーの公式より

$$|H(\omega t)| = \frac{1-a}{\sqrt{1-2a \cos \omega \tau N_{i,p} + a^2}} \quad (10)$$

となるため図2 のような利得が得られる。

4 章 MFCCについて

MFCC(メル周波数ケプストラム係数 : Mel-Frequency Cepstrum Coefficient)とは音声認識に一般に用いられる特徴量[16]で, 計算が簡素で処理速度が速いという利点がある.

MFCC を求める手順は以下の図 3 に示される.

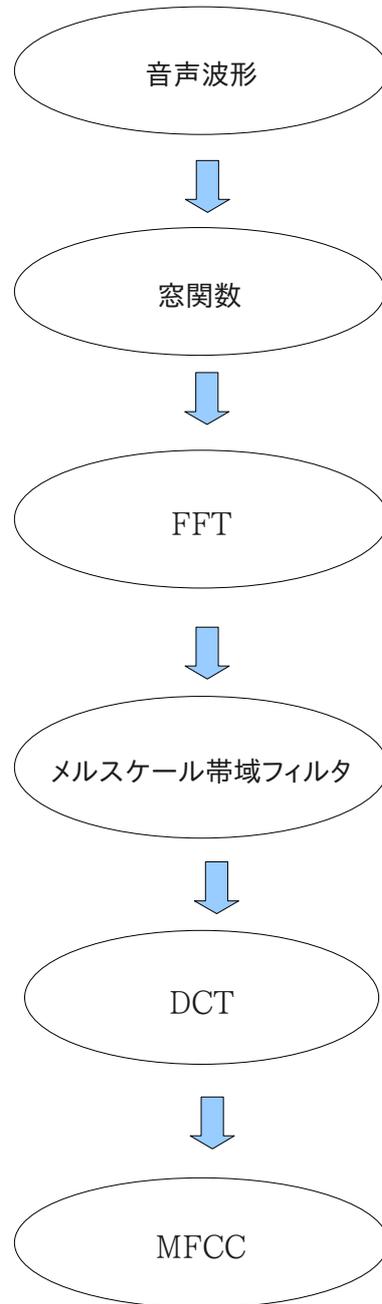


図 3. MFCC の計算手順

5 章 マハラノビスの距離によるテンプレートマッチングについて

マハラノビス(mahalanobis)の距離とは座標間の距離であるが、ユークリッド距離とは異なり重み付けがされた距離である。マハラノビス距離の例を図4に表す。ピアノ、バイオリン音声から求めた特徴量が特徴空間内にプロットされており、各群の平均値を群の重心とする。ここで入力音声の特徴量が特徴空間内の中央の座標にプロットされる時、ユークリッド距離では入力音声はバイオリンの重心よりもピアノの重心に近い位置にある。つまり、ユークリッド距離では「入力音声はピアノである」と推定される。一方、マハラノビスの距離の場合は重み付けされた距離を考慮するため「入力音声はバイオリンである」と推定される。判別曲線もマハラノビスの距離を用いると曲線になり、ユークリッド距離を用いると直線になるという違いもある[17]。

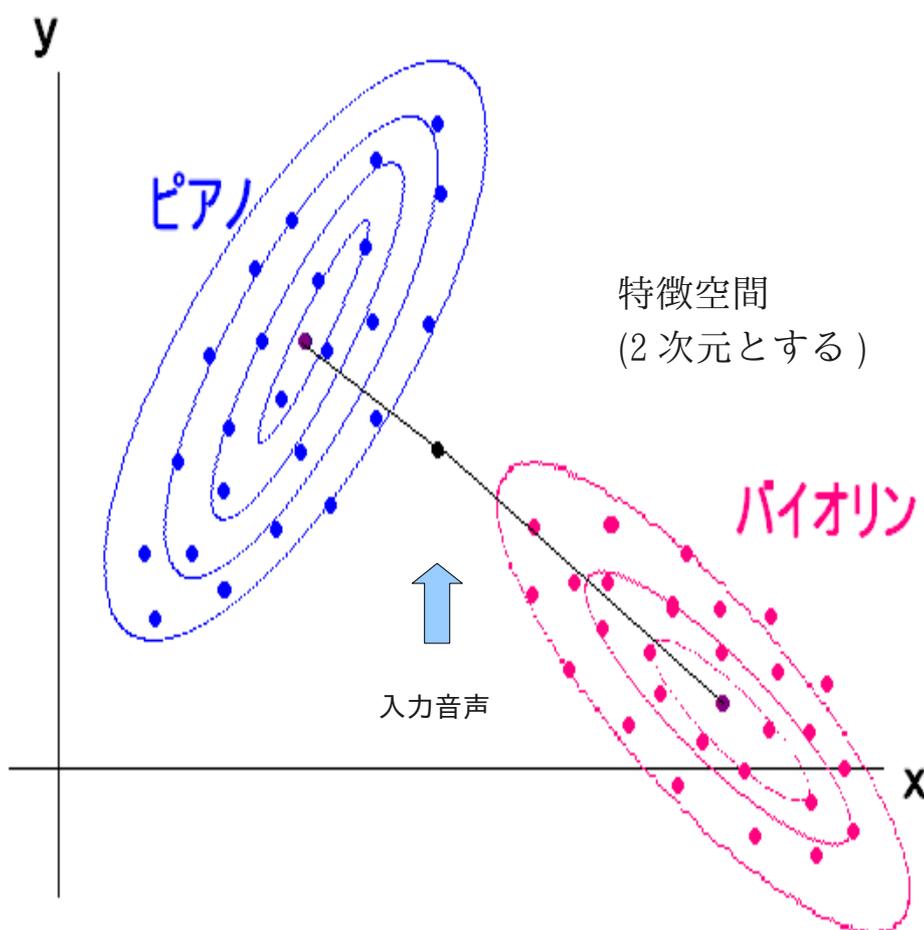


図4. マハラノビスの距離の例

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 5)。
- (2) 各 $y_{i,p}(n)$ から MFCC を求めてテンプレートマッチングを行う．今回はテンプレートマッチング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 との間マハラノビス距離を使用する．最も距離の短い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．

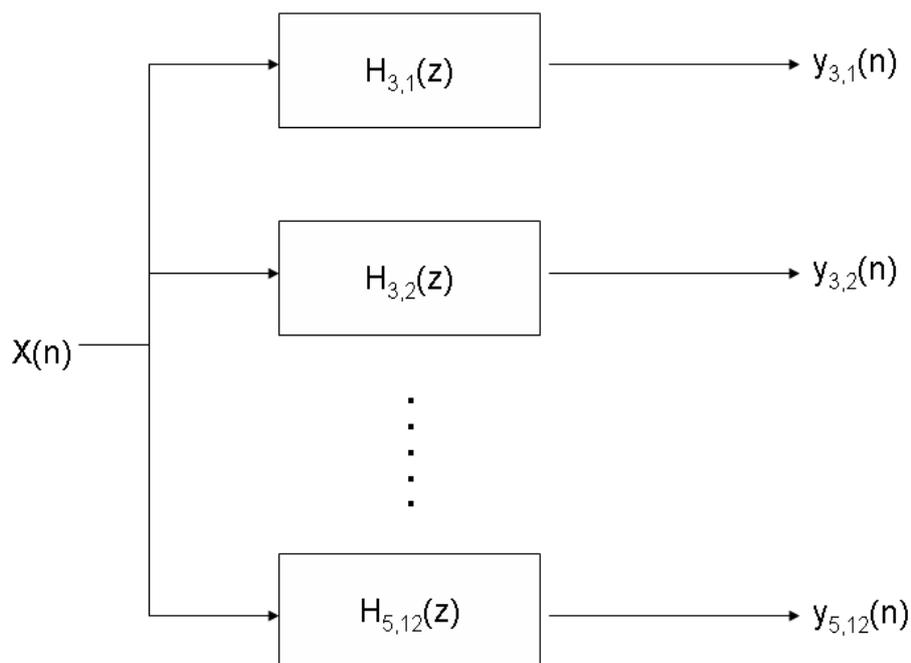


図 5. Resonator 型くし形フィルタの並列接続例

7 章 テンプレートと実験データの作成

入力された演奏音声はResonator型くし形フィルタの特性を受けるため、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、各楽器の(i,p)音をResonator型くし形フィルタ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。なお、RWC研究用音楽データベースの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ファイルを分割した後、それぞれのWAVEファイルのMFCCを計算してテンプレートを作成した。2和音実験用のデータは、各音高・各楽器全ての組み合わせのWAVEファイルを合成して作成した(図6)。使用した音声はRWC研究用音楽データベース[14]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)である。

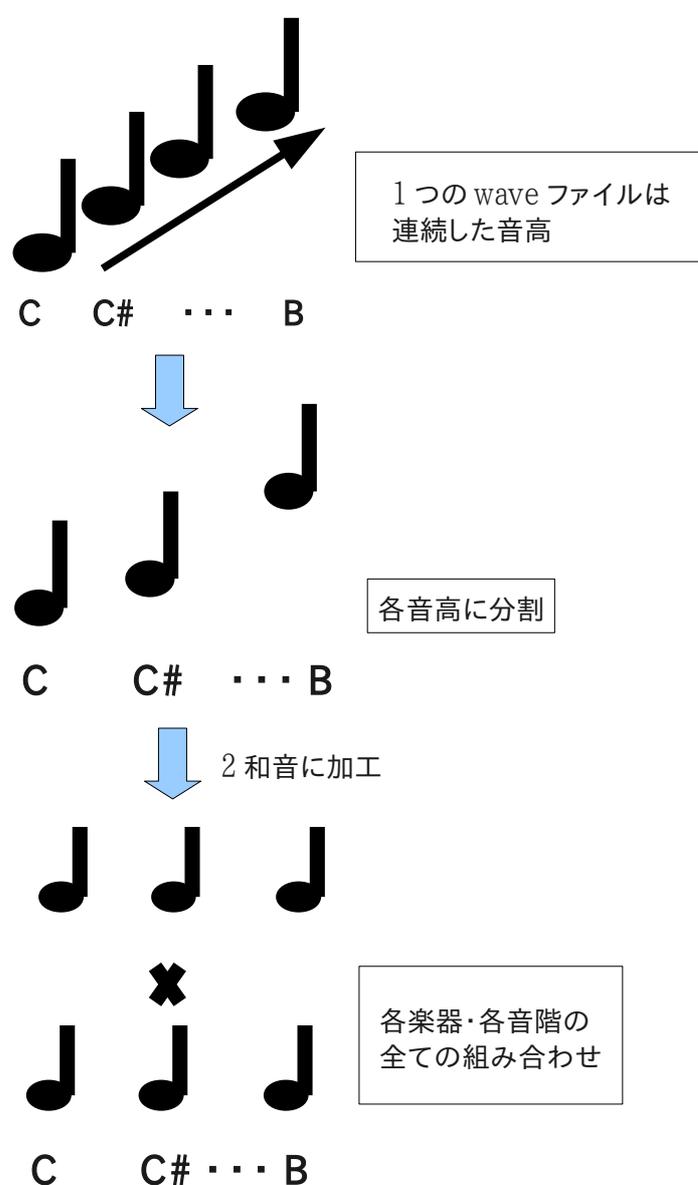


図6. 音声の分割

8 章 実験1（単音での演奏楽器推定）

8-1 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[9]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。

8-2 結果

結果を表1に示す。表中の数字は推定精度を表す。表の行は推定された楽器, 列は推定対象の楽器である。

表1. 単音での推定結果

	AS	CB	CL	FL	HR	PF	TR	VC	VL	VN
AS	94.7	0	0	0	0	0	0	0.8	0.5	4
CB	0	96.3	0	0	2.1	3.3	0.4	0	2.8	8.6
CL	0.4	1.2	53.1	0.1	3.2	3.1	0	35.3	1.6	2.1
FL	0	0	2.5	95.5	0	0	7.8	0	2.7	1.1
HR	0	1.9	13	1.3	88	3.6	0	0	2.1	0.5
PF	0.4	1.4	3	0.4	6.2	65.9	0	19.9	0.8	0
TR	1.1	3.5	0	2	4.1	3.7	96.5	0	4.6	0
VC	0	1.9	4	2.3	2.4	15.3	1.1	16.5	44.5	0
VL	0	0	7.5	0	5.5	0.2	0.6	1	88.5	3.7
VN	0	2.3	0	0	5.5	8.9	1.7	2.2	6.9	42.6

8-3 考察

単音に関しては全体的にうまく推定された。しかし, チェロ(VC)に関してはビオラに推定結果が引きずられてしまい正しく推定されなかった。これはチェロとビオラの音が非常に似ていることが理由であると考えられる。

9 章 実験2 (2和音での演奏楽器推定 1)

9-1 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), ビオラ(VL), バイオリン(VN)の9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[14]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の各音階の楽器音を用いて8章の説明した方法で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。

9-2 結果

結果を表2, 表3に示す。

9-3 考察

実験の結果, 全楽器での推定正解率として81.9%と高い推定精度が得られた。個別の楽器については, 特にアルトサククス(AS), コントラバス(CB), クラリネット(CL), フルート(FL), ビオラ(VL), バイオリン(VN)に関して非常に高い推定精度が得られた。しかし, ホルン(HR), トランペット(TR)に関しては推定正解率が50%以下になるものもあり, 高い推定精度が得られなかった。

この原因としては, 本手法ではMFCCを用いたマハラノビスの距離によるテンプレートマッチングを行ったため, アタック, サステイン等のエンベロープ(ADSR)[18][19](付録3)を全く考慮していなく, さらにオクターブ4の範囲でのみ推定を行ったため各楽器の中心的な音域による特徴的な音色を掴めなかつことが考えられる(付録1, 2)。

この改善点として, 各楽器特性のエンベロープを考慮することや, 各楽器の中心的な音域を用いてテンプレートを作成すること等が考えられる。

表 2. 2 和音での推定結果（入力音声の音の強さ”強”）

	AS	CB	CL	FL	HR	PF	TR	VL	VN
AS-CB	55.6	36.3	0.1	0.0	0.2	0.7	0.0	1.9	5.3
AS-CL	56.2	0.7	34.2	0.1	0.1	0.6	0.0	1.9	6.1
AS-FL	60.0	1.8	0.4	24.9	0.2	1.1	0.0	5.0	6.6
AS-HR	61.4	0.2	0.0	0.1	17.2	16.3	0.1	1.0	3.7
AS-PF	62.3	0.4	0.0	0.0	0.1	31.1	0.0	1.8	4.3
AS-TR	55.2	0.1	0.1	0.1	0.5	5.8	25.7	1.5	11.2
AS-VL	54.8	0.0	0.0	0.1	0.1	1.5	0.0	34.0	9.5
AS-VN	56.4	0.0	0.0	0.0	0.0	0.8	0.0	1.7	40.9
CB-CL	4.4	52.2	21.1	3.3	2.3	1.8	0.0	8.7	6.3
CB-FL	2.6	58.7	3.2	10.9	1.2	7.3	0.0	11.5	4.6
CB-HR	1.3	52.0	2.4	5.7	15.1	19.6	0.0	2.5	1.4
CB-PF	4.2	71.2	5.3	1.0	3.9	5.6	0.0	4.5	4.3
CB-TR	4.7	21.4	1.3	4.0	4.2	11.2	7.9	27.9	17.3
CB-VL	4.0	24.9	5.3	0.8	2.3	5.9	0.0	48.5	8.2
CB-VN	3.3	50.8	2.7	2.4	3.0	2.8	0.0	14.3	20.6
CL-FL	2.3	9.3	46.3	15.1	1.4	7.4	0.0	11.5	6.7
CL-HR	1.4	7.1	28.4	8.0	19.4	24.5	0.0	8.5	2.8
CL-PF	5.7	9.7	62.7	0.9	1.4	11.3	0.1	3.8	4.3
CL-TR	5.3	3.1	7.8	2.5	4.0	19.7	8.9	23.7	25.0
CL-VL	4.0	8.8	19.0	1.5	1.0	3.7	0.0	55.0	6.8
CL-VN	3.1	11.0	36.9	2.2	0.3	0.6	0.0	17.5	28.3
FL-HR	0.6	3.1	1.0	61.2	7.8	18.7	0.0	5.3	2.3
FL-PF	1.4	2.8	0.8	69.2	0.5	8.7	0.0	12.5	4.1
FL-TR	2.4	2.0	0.5	27.4	3.0	15.2	4.1	20.1	25.2
FL-VL	1.6	2.5	0.9	25.7	0.5	8.2	0.0	52.3	8.3
FL-VN	0.8	4.9	0.7	39.1	0.3	3.9	0.0	13.6	36.7
HR-PF	0.5	0.6	0.0	0.1	50.0	48.1	0.1	0.3	0.2
HR-TR	9.4	0.4	1.8	0.2	10.3	41.6	12.2	17.9	6.1
HR-VL	2.1	0.2	0.1	0.6	14.9	18.8	0.2	59.1	4.1
HR-VN	0.8	2.7	0.2	4.3	22.9	15.6	0.1	25.1	28.2
PF-TR	5.3	0.4	0.0	0.1	2.5	32.1	48.1	3.7	7.7
PF-VL	6.0	1.3	6.1	0.5	2.3	14.7	0.1	54.4	14.5
PF-VN	2.1	2.3	7.9	0.5	4.8	18.7	0.1	12.5	50.9
TR-VL	5.8	0.4	0.0	0.1	2.4	9.2	24.2	44.3	13.6
TR-VN	3.9	0.3	0.1	0.8	1.2	4.8	35.3	9.1	44.4
VL-VN	2.1	2.2	2.9	0.4	0.4	7.6	0.0	63.1	21.2

表 3. 楽器個別の推定精度

AS	CB	CL	FL	HR	PF	TR	VL	VN	全体での正解率
100.0%	100.0%	87.5%	87.5%	25.0%	87.5%	50.0%	100.0%	100.0%	81.9%

10 章 実験3 (2和音での演奏楽器推定 2)

10-1 実験概要

次に、テンプレートとは異なるフォーマットの音声を用いて2和音での楽器推定実験を行った。実験対象の楽器はアルトサククス(AS), 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), ビオラ(VL)の7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[14]に収録されている通常奏法, 音の強さ”中”の実演奏楽器音声(44.1Hz, 16bit, モノラル)である。

10-2 結果

結果を表4, 表5に示す。

10-3 考察

実験の結果, 全楽器での推定正解率として64.3%の推定精度が得られた。個別の楽器についてはホルン(HR), トランペット(TR)といった金管楽器に関しては推定正解率がおよそ30%以下になるものもあり, 高い推定精度が得られなかった。

表 4. 2 和音での推定結果（入力音声の音の強さ”中”）

	AS	CL	FL	HR	PF	TR	VL
AS-CL	36.4	25.2	7.6	1.1	2.7	0	27
AS-FL	34.5	5.2	13.7	0.9	6.9	0.1	38.7
AS-HR	37.1	1.1	2.4	12	26.8	0	20.6
AS-PF	38.1	1.8	1.7	3.1	33.5	0.1	21.8
AS-TR	35.7	1.1	1.8	7.1	22.2	2.1	30
AS-VL	31.7	0.8	2.1	0.8	1.7	0.1	62.9
CL-FL	2.1	32.6	26.9	2.2	5.1	0.2	30.9
CL-HR	4.9	23.5	15.1	17	17.9	0.1	21.5
CL-PF	5.7	35.3	13	6.7	20.1	0.1	19.2
CL-TR	7.8	17.1	13.4	11.1	17.3	1.4	31.9
CL-VL	8	21.4	9	3.4	5.7	0.1	52.4
FL-HR	0.9	5.1	36.8	13.7	25.6	0	17.9
FL-PF	5.9	15	28.6	5.2	14.2	0.2	30.9
FL-TR	5	1.4	25.2	7.7	20.5	2	38.2
FL-VL	7.1	8.3	11.5	2.7	13.5	0	56.9
HR-PF	1.4	0.4	5.1	34.1	58.2	0	0.8
HR-TR	13.4	3.2	3.5	25.7	41.7	3.5	8.9
HR-VL	7.7	0.8	5.2	16.5	27.8	0.1	41.9
PF-TR	10.8	0.7	1.9	16.5	58.4	4.2	7.4
PF-VL	13.7	10.7	3.9	6.5	20.6	0.1	44.5
TR-VL	12.4	0.8	3.3	9.4	27.3	2.9	43.9

表 5. 楽器個別の推定精度

AS	CL	FL	HR	PF	TR	VL	全体での正解率
100.0%	66.7%	66.7%	33.3%	83.3%	0.0%	100.0%	64.3%

11章 まとめ

本研究では，Resonator 型くし形フィルタにより演奏音声から音源分離した後で，MFCC を求め，マハラノビスの距離を用いて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を提案した．オクターブ4の範囲で実演奏楽器推定の実験を行った結果，単音に関しては全体として高い推定精度が得られた．2和音に関しては，推定対象となる入力音声とテンプレート音声と同じ音の強さ”強”の場合は81.9%の推定精度が得られたが，入力音声とテンプレート音声と異なる音の強さ”中”の場合は64.3%の推定精度となった．

今後の課題として，各楽器特性のエンベロープや各楽器の中心的な音域を用いたテンプレートマッチングを考案することや，最適なフィルタ係数 a を求めることが挙げられる，また，Notch 型くし形フィルタを用いた場合との比較を行う必要がある．更に，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] 関和則, DTMのための全知識: リットーミュージック社, 1997.
- [2] 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,
- [3] 藤原弘将, 後藤真考: 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.
- [4] 吉井和佳, 後藤真考, 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.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 大石康智, 後藤真考, 伊藤克亘, 武田一哉: 局所的・大局的な特徴を利用した歌声と朗読音声の識別, 情報処理学会研究報告, MUS-61, pp.1-6, 2005.
- [9] 武藤真介, 初等多変量解析: 朝倉書店, 1999.
- [10] 坂内秀幸, 夏井雅典, 田所嘉昭: くし形フィルタに基づく自動採譜システムの実現, 情報処理学会研究報告, 2007-MUS-71, pp.13-18, 2007.
- [11] 亀岡弘和, 西本卓也, 嵯峨山茂樹: ハーモニッククラスタリングと情報量基準による音楽の音高/音源数の推定, 情報処理学会研究報告, MUS-62, pp.27-32, 2005.
- [12] 宮坂広純, 松島俊明, 小田弘良: 多重音の基本周波数評価尺度の最小値選択による実時間音高・音源数推定の試み, 第7回情報科学技術フォーラム講演論文集(2), pp.241-243, 2008.
- [13] 糸山克寿, 後藤真考, 駒谷和憲, 尾形哲也, 奥乃博: 複数楽器個体による事前分布を用いた調波・非調和統合モデルのパラメータ推定, 情報処理学会第70回全国大会講演論文集(2), pp.435-436, 2008.

- [14] 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>
- [15] 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.
- [16] 石井健一郎, 上田修功, 前田英作, 村瀬洋 : わかりやすいパターン認識, オーム社, 1999.
- [17] 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.
- [18] 松前公高, シンセサイザー入門 : リットーミュージック社, 2007.
- [19] 松武秀樹, 別冊大人の科学マガジン : 学研, 2008.
- [20] 佐藤政次, 音声認識システム : オーム社, 2002.

付録

付録1. 実験に使用した各楽器の音域

表6. 各楽器の音域

アルトサククス(AS)	D \flat 3 - A \flat 5
ホルン(HR)F管	B \flat 2 - A5
ピアノ(PF)	A0 - C8
トランペット(TR)	E3 - B \flat 5
ビオラ(VL)	A3 - E6
バイオリン(VN)	G3 - E \flat 7
コントラバス(CB)	1E - B \flat 4
クラリネット(CL)E \flat 管	G3 - C6
フルート(FL)	C4 - D7

付録2. 実験に使用した各楽器の音域図

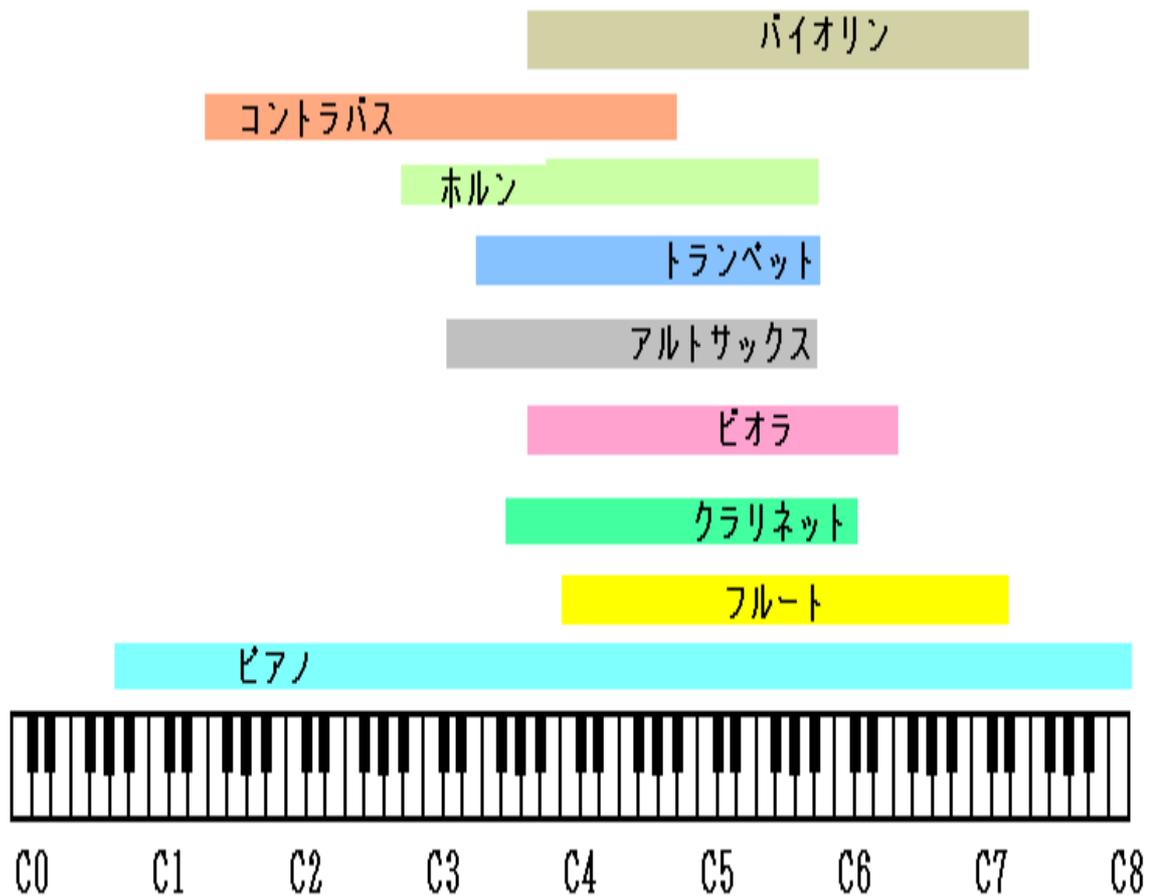


図 7. 音域図

付録3. エンベロープ特性(ADSR)

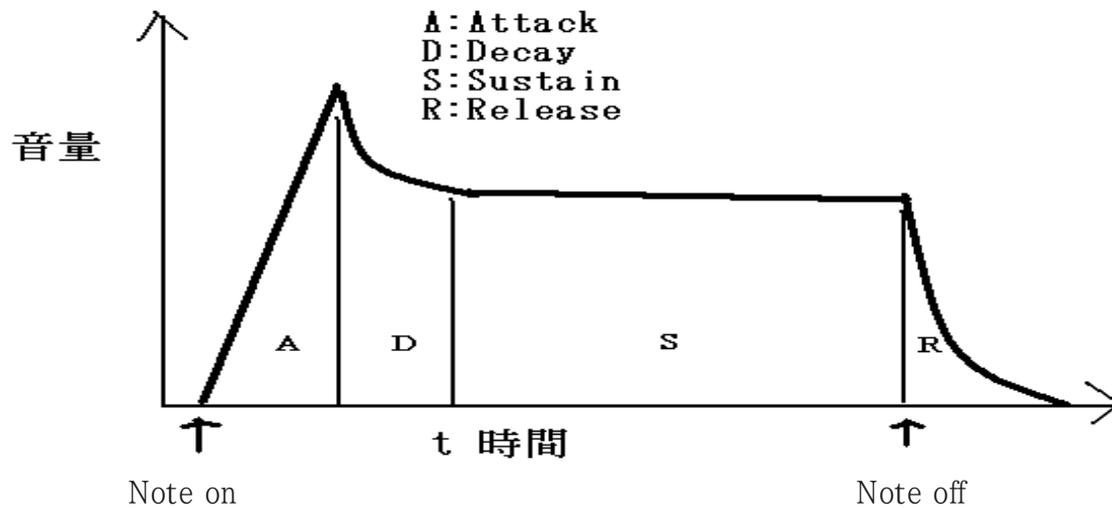


図8. エンベロープ特性

- Attack : 音の立ち上がりから最大音量に達するまでの時間
- Decay : 立ち上がりの最大音量から音量が持続していくまでの時間
- Sustain : 定常音量が持続する時間
- Release : 鍵盤を離してから音が消えていくまでの時間

なお、弦楽器や木琴・鉄琴などの場合はReleaseのことをSustainという。