

Resonator 型くし形フィルタを用いた演奏楽器推定手法

室谷良平† 東海林智也†

†函館工業高等専門学校情報工学科

本稿では, Resonator型くし形フィルタを用いて分離した楽音の調波成分から演奏楽器を推定する手法について検討する. 始めに, 全ての楽器と音階の組み合わせに対する調波成分から MFCC を求めてテンプレートを作成する. 次に推定対象の楽音を, 並列接続した Resonator 型くし形フィルタ群に通してから各音階別に MFCC を求め, 各音階別にマハラノビス距離を用いてテンプレートマッチングを行う. 実際の楽器音に対して行った実験の結果, 81.9%の精度で楽器推定を行えることを確認した.

Instrument Estimation by Resonator-Type Comb Filters

Ryohei Muroya† and Tomoya Tokairin†

†Hakodate National College of Technology

This paper describes a system to estimate musical instruments. This system divides polyphonic musical sound into harmonic components through parallel connected resonator comb filters. Then, this system estimates musical instruments by using Mahalanobis distance between MFCC of the separated harmonic components and MFCC of template data. To verify the efficiency of the proposed system, we conducted a computer simulation. Experimental result showed that 81.9% accuracy was obtained.

1. はじめに

現在, 携帯音楽プレーヤーや着うた等の携帯電話の音楽再生機能の普及により音楽が身近なものとなっている. また近年の DTM 人気により, 楽器を演奏出来ない人でも自由に音楽を作成して楽しむ時代になって来ている. それにつれて楽曲検索, 楽曲推薦, 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が, そのためには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. 楽曲情報を抽出するためには採譜が必要であるが, 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[1][2].

そこで本稿では, Resonator 型くし形フィルタ[3]

を用いて演奏音声から音源を分離した後で MFCC を求め, マハラノビスの距離[4]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.

これまで, 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[5]やハーモニッククラスタリング[6][7], 調波・非調波統合モデル[8]による推定手法等が提案されているが, 特に本稿で提案する手法と 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 型くし形フィルタ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 型くし形フィルタ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 型くし形フィルタ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してしまう. 一方, Resonator 型くし形フィルタ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ノイズの発生を抑えることができる. さらに, Resonator 型くし形フィルタ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.

また、我々は提案した手法の有用性を調べるためにコンピュータを用いた実験を行った。我々はこれまで電子音のみを用いて楽器推定の実験を行ってきたが、今回はRWC研究用音楽データベース[9]に収録されている実演奏音を用いて実験を行った。

2. Resonator 型くし形フィルタ

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 $N_{i,p} = [f_s / f_{i,p}]$ ($[]$: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[1][3]。

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

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1に示される。図1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 $H_{i,p}(z)$ を図2のように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。

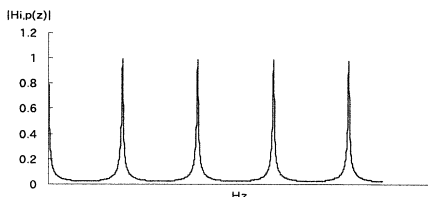


図1. Resonator 型くし形フィルタの利得特性 ($a = 0.9$)

3. 提案手法

本稿で提案する楽器推定手法の手順は以下のとおりである。

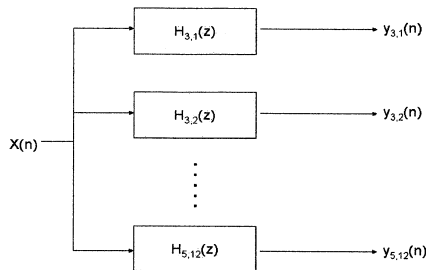


図2. Resonator 型くし形フィルタの並列接続例

- (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。
- (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 との間のマハラノビス距離[4]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。

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。

4. 実験

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ックス(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[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, 16 bit, モノラル)である。その内オクターブ4の各音階の楽器音を切り出して、計 108 (C~Bまでの12音階×9楽器)個のテンプレート用 wave ファイルを作成した。さらに推定で用いる wave ファイルとして各楽器・各音階の全ての組み合わせの音声を合成し2和音の wave ファイルを

作成した。ただし、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。また、ユニゾンに関しては音源分離が困難なため考慮しないものとした。なお、フィルタ係数を $a = 0.90$ 、テンプレートマッチングを行う時間間隔を 0.01 秒とした。

実験結果を表 1 及び表 2 に示す。表 1 の各行は対象楽器の組み合わせ、各列は推定対象の楽器名、表中の数字は推定精度(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。また、表中の灰色の項目は正しく推定された楽器を示し、太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 2 は楽器個別の推定精度と全体での平均推定精度を示している。

5. 考察

実験の結果、全体での平均推定精度として 81.9% と高い推定精度が得られた。個別の楽器については、特にアルトサクソ(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特性(ADSR)を全く考慮してなく、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ったことが考えられる。

この改善点として、楽器のエンベロープ特性も考慮することが考えられる。特にホルンやトランペット等の管楽器はエンベロープ特性が特徴的な楽器であるためエンベロープ特性を考慮することで高い推定精度を得られることが期待できる。更に、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も考えられる。

6. まとめ

本稿では、Resonator 型くし形フィルタにより楽音から調波成分を分離し、分離した調波成分の MFCC を求め、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て演奏楽器を推定する手法を提案した。また、9 楽器・2 和音での楽器推定実験を行い、結果として 81.9% の推定精度が得られた。

今後の課題として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て楽音から調波成分を分離してテンプレートマッチングを行った場合との比較を行う必要がある。

参考文献

- [1] 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.
- [2] 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.
- [3] 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.
- [4] 武藤眞介, 初等多変量解析, 朝倉書店, 1999.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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-i.html>

	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1. 推定結果

AS	CB	CL	FL	HR	PF	TR	VL	VN	平均推定精度
100.0%	100.0%	87.5%	87.5%	50.0%	66.7%	62.5%	100.0%	100.0%	81.9%

表2. 楽器別の楽定結果