

くし形フィルタと自己相関関数に基づく調波成分が重複する 混合音の音高推定

山口 満[†] 田所嘉昭[‡]

[†]豊橋創造大学・情報ビジネス学部・キャリアデザイン学科
[‡]豊橋技術科学大学・工学部・情報工学系

あらまし 本稿では、一方の調波成分が他方の調波成分と重複している混合音（基本周波数比が整数）に対する音高推定法として、くし形フィルタと自己相関関数による一手法を提案する。提案法では、重複する調波成分間には若干の周波数差が存在するものと仮定し、くし形フィルタにより一方の音が除去された際に現れるフィルタ出力の自己相関関数の変化を観測することによって、単音か混合音か、また混合音の場合はその重複関係を判別する。実験の結果、信号の立上りから 400[ms] 経過後のデータに対して約 85% の正答率で推定が可能であることを明らかにした。

Pitch Estimation for Mixed Tones Containing Overlapping Harmonic Components Based on Comb Filters and Auto-correlation Functions

Michiru YAMAGUCHI[†], Yoshiaki TADOKORO[‡]

[†]Dept. of Career Design, Faculty of Business and Information, Toyohashi Sozo University

[‡]Dept. of Information and Computer Sciences, Faculty of Engineering, Toyohashi University of Technology

Abstract A pitch estimation method for mixed tones which containing overlapping harmonic components was proposed. This method uses comb filters which can eliminate corresponding harmonic components, and auto-correlation functions which reveals the periodicity of a signal. Experimental results show that proposed method can estimate a pitch relation with the correct estimation rate of 85%.

1 はじめに

自動採譜の実現において、楽音の音高推定は不可欠な処理の一つであり、これまで様々な研究が行われている [1]~[9]。筆者らもまた、音高推定を実現するための手法として、くし形フィルタの零出力検出に基づいた時間領域における処理手法を検討している [5]~[8]。

音楽の世界で一般的に用いられている平均律音階では、オクターブ 4 の A の基本周波数 ($A_4:440[\text{Hz}]$) を基準として、1 オクターブ中に含まれる 12 音 ($C, C\#, \dots, B$) の基本周波数是对数的に等間隔に割り当てられている。このため、隣り合う音高間の周波数比は $2^{1/12}$ であり、各基本周波数の間隔は不等間隔となっている。このような平均律音階においては、特定の音同士による混合音においては、一方の調波成分が他方の調波成分と重複する場合が生じる。例えば、ある音高 p と 1 オクターブ上の (高い方に 12 音の隔たりをもつ) $p+12$ 音の基本周波数比は $2^{12/12} = 2$ 、

p と $p+19$ 音では $2^{19/12} \cong 2.996614 \cong 3$ のように、その比が整数または整数に近い関係となる。このため、上述の関係にある調波成分が重複する混合音の音高推定は非常に困難な問題とされている [10]。筆者らが以前に提案している手法 (くし形フィルタの零出力検出による方法) においても、そのような 2 音に対しては低い音高 p の 1 音しか推定することができない。また、既存の研究としては、混合音のスペクトルに関していくつかの仮定を設けた統計的分析に基づく推定法 [3] や音楽的知識を利用した推定法 [4] が提案されているが、必ずしも仮定が成立しないことや事前知識を要するなどの点において、決定的な手法であるとは言い難い。

これに対し、筆者らは、対象音の最低音高が既知でかつ重複する調波成分間には若干の周波数差が存在するという条件の下、信号処理的アプローチにより、信号が単音であるか、あるいは 2 音 ($1:r, r=2,3$) であるかの判別法を検討している [9]。本

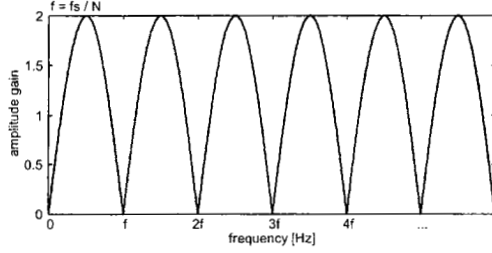


図1 くし形フィルタ $H_N(z) = 1 - z^{-N}$ の振幅特性

稿では、新たな手法の一つとして、くし形フィルタと自己相関関数に基づいた方法を提案する。本手法では、数個のくし形フィルタを用いて入力音をフィルタリングし、その出力における自己相関関数の変化を観測することによって、上記の判別を実現する。

なお、対象とする楽音の音域は $C_3 \sim B_5$ の範囲とする。また、ユニゾン（基本周波数比 = 1 : 1 の混合音）については、今回は対象として扱っていない。

2 従来の音高推定法および問題点

本節では、本研究で使用するくし形フィルタとそれに基づく音高推定の概要、および、問題点について簡単に説明する。

2.1 くし形フィルタによる音高推定

遅延数 N のくし形フィルタの伝達関数を式 (1) に、また、差分方程式を式 (2) のように表現する。

$$H_N(z) = 1 - z^{-N} \quad (1)$$

$$y_N(n) = x(n) - x(n - N) \quad (2)$$

いま、サンプリング周波数を f_s [Hz]（本研究では 44.1 [kHz]）とすると、式 (1) の $H_N(z)$ は、周波数領域において図 1 に示すように f_s/N [Hz] 間隔に零点をもつ。このため、 $H_N(z)$ は $x(n)$ に含まれる f_s/N [Hz] の整数倍の周波数成分を除去することが可能である。すなわち、ある音高 p ($= 1, 2, \dots, 36$, $C_3 \sim B_5$ の 36 音高を表現) の調波成分を除去する遅延数 N_p をもつくし形フィルタを用いて、その出力信号が零出力か否かを判別することによって入力音高を推定することができる（零出力検出による音高推定）。なお、零出力検出は、出力信号の振幅レベルとしきい値との比較により行われる。

これまで、筆者らは、 $H_N(z)$ の縦続接続構成 [5], [6]（図 2）、あるいは、並列接続構成による音高推定法をいくつか提案している [7], [8]。

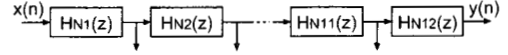


図2 $H_N(z)$ の縦続接続構成による音高推定

2.2 問題点

前述のように、基本周波数比が 1 : 2（例えば p と $p + 12$ ）、あるいはおよそ 1 : 3（例えば p と $p + 19$ ）の関係にある 2 音については、見かけ上、その周波数配置が低い音高 p が 1 音である場合と等しくなる。このため、 p 音を除去するくし形フィルタの出力はおよそ零となり、低い音の音高 p のみが推定され、もう一方の音を検出することは困難であった。このことから、 p 音を推定した後の処理として、単音/混合音の判別を実現する処理が必要とされていた。

3 くし形フィルタ出力の自己相関関数の特徴

本研究では、調波成分が重複する混合音に対する音高推定法として、くし形フィルタによる調波成分の除去、および、その出力の自己相関関数の変化検出に基づいた手法を提案する。はじめに、自己相関関数の計算手順について述べる。次に、単音および混合音を入力とした際の、くし形フィルタ出力で得られる自己相関関数の特徴について述べる。

3.1 自己相関関数とそのピークインデックス検出

$H_N(z)$ の出力 $y_N(n)$ に対して、式 (3) のように正規化された自己相関関数を計算する。

$$R_N(k) = \frac{\sum_{m=0}^{M-1} y_N(m+N)y_N(m+N+k)}{\sum_{m=0}^{M-1} y_N(m+N)y_N(m+N)} \quad (3)$$

信号が周期信号であるならば、計算される自己相関関数も周期関数であり、その周期（1 周期のサンプル数）に対応するラグ k でピーク値をとる。このため、信号周期を判別することが可能である。本研究では、以下のように相関関数のピーク値を求める。

元信号の既知の音高 p ($p = 1, 2, \dots, 36$) の調波成分を除去するくし形フィルタ遅延数 N_p を基に、式 (4) のように $R_N(N_p)$ の近傍において極大値を示す $R_N(k)$ を求め、これを $R_N(k_{\max})$ とする。

$$R_N(k_{\max}) = \max_{N_p - k_1 \leq k \leq N_p + k_2} R_N(k) \quad (4)$$

ここで、 $N_p - k_1$ および $N_p + k_2$ は、既知の音高 p の ± 50 [cent] の周波数に対応するサンプル数（遅延数）である。本研究では、1 : r ($r = 2, 3$) の関係

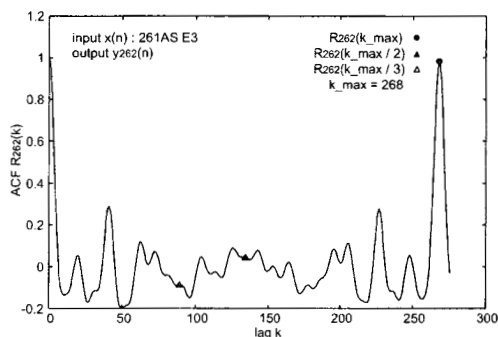


図3 入力音: alto-sax E_3 , $y_{262}(n)$ の自己相関関数 $R_{262}(k)$ (● は $R_{262}(k_{\max})$, ▲ は $R_{262}(k_{\max}/2)$, △ は $R_{262}(k_{\max}/3)$ を示す)

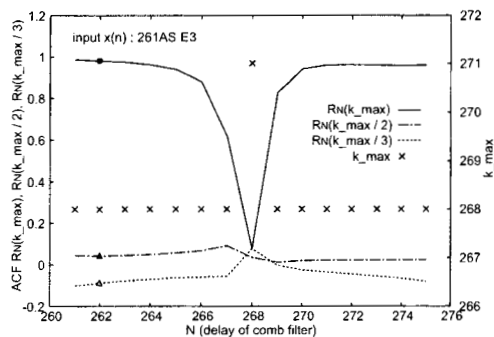


図4 入力音: alto-sax E_3 , 各遅延数 N における $y_N(n)$ から検出した $R_N(k_{\max}/j)$ ($j = 1, 2, 3$) (左軸) および k_{\max} (右軸)

にある2音の判別を目的とするため、さらに元の音高 p の r 倍の周期に相当する相関関数 $R_N(k_{\max}/r)$ の値に注目することとする。

例として、入力信号 $x(n)$ を alto-sax E_3 音 ($p = 5$, $f_s = 44.1$ [kHz], $N_p = 268$) として $N = 262$ のくし形フィルタ $H_{262}(z)$ に入力した際の、出力 $y_{262}(n)$ に対する自己相関関数 $R_{262}(k)$ の計算結果を図3に示す。同図中には、あわせて $R_{262}(k_{\max})$, $R_{262}(k_{\max}/2)$, および $R_{262}(k_{\max}/3)$ を示している。図3の例では、 $H_{262}(z)$ が $x(n)$ に含まれる E_3 音成分をすべて除去できないため、 $k = 268$ でピーク ($\cong R_{262}(0)$) が生じ、 $y_{262}(n)$ に周期信号成分が残っていると判断できる。このように、 $H_N(z)$ が $x(n)$ の調波成分をすべてを除去できなかった場合には、自己相関関数において信号周期に相当する k でピーク値が観測されることがわかる。

3.2 単音入力時に検出される自己相関関数 (調波成分を十分に除去/抑圧できる場合)

入力信号 $x(n)$ (alto-sax E_3 音, $p = 5$, $N_p = 268$) に対し、フィルタ遅延数 N を N_p の近傍で変化させた場合について、検出される $R_N(k_{\max}/j)$ ($j = 1, 2, 3$) をプロットしたものを図4に示す。 $N = 268$ において、 $R_{268}(k_{\max})$ が大きく減少していることが確認できる。これは、 $x(n)$ の成分がほぼ除去され、 $y_{268}(n)$ にはほぼ周期信号成分が残っていないことを意味している。すなわち、 $x(n)$ が単音 (信号周期はサンプル数で268) であることが容易に推定される。なお、図4から、 $R_N(k_{\max}/2)$ および $R_N(k_{\max}/3)$ もまた、 $N = 268$ において0に近づいていることが確認できる。

3.3 単音入力時に検出される自己相関関数 (調波成分の除去が不十分な場合)

$x(n)$ を alto-sax B_3 音 ($p = 12$, $N_p = 179$) として、 N を変化させたときに得られる $R_N(k_{\max}/j)$ (正規化後) を図5に示す。図5を見ると、前例 (図4) では明確であった周期性の消失、すなわち、 $R_N(k_{\max})$ の減少について、極めて判断が困難であることがわかる。これは、フィルタ遅延数の制約 (整数値であること) から $H_N(z)$ による調波成分の除去が不十分であり、結局 $R_N(0) \cong R_N(k_{\max})$ となっていることを意味している。

以上から、仮に入力 $x(n)$ が単音であり、適切なくし形フィルタを選択して除去を試みたとしても、その出力の正規化自己相関関数 $R_N(k_{\max})$ は必ずしも小さくはないといえる。このため、 $x(n)$ が単音か否かを判別する際には、上記のような $R_N(k_{\max})$ の振る舞いに注意する必要がある。

3.4 調波成分が重複する2音を入力した場合に検出される自己相関関数

$x(n)$ を理想信号 $C_3 + C_4$ 音 ($p = 1$, $N_p = 337$) として、基本周波数比が1:2の関係にある2音の混合音に対し、同様に $R_N(k_{\max}/j)$ を求めた。なお、理想信号は鋸歯状波とし、 C_4 の基本周波数 f_{13} は +10cent のずれを持つものとした。図6に、 N の変化に対する $R_N(k_{\max}/j)$ の値を示す。同図より、 $N = 335$ の場合に $R_N(k_{\max}/2)$ が大きく減少していることが確認できる。これは、 $H_{335}(z)$ が C_3 成分以上に C_4 成分を大きく減衰させ、 $y_{335}(n)$ がほぼ C_3 単音とみなせる周波数成分で構成されていることを意味する。一方、 $N = 337$ においては、 $R_N(k_{\max})$ と $R_N(k_{\max}/2)$ が非常に近い値となる (近接している) ことが確認できる。これは、 $H_{337}(z)$ が C_4

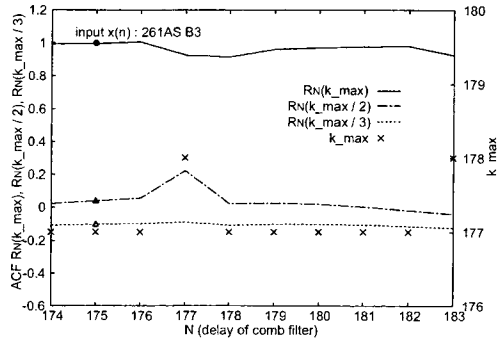


図5 alto-sax B_3 に対する N と $R_N(k_{\max}/j)$ の関係 ($j = 1, 2, 3$, 左軸) および検出された k_{\max} (右軸)

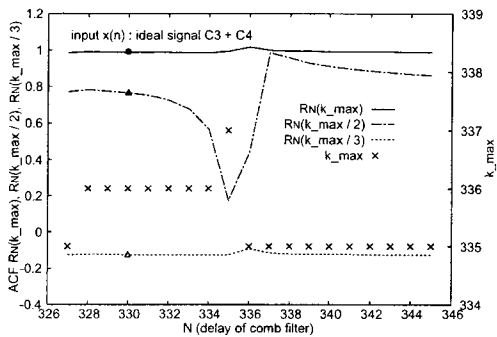


図6 理想信号 $C_3 + C_4$ に対する N と $R_N(k_{\max}/j)$ の関係 ($j = 1, 2, 3$, 左軸) および k_{\max} (右軸)

成分以上に C_3 成分を大きく減衰させ、 $R_{337}(0) \cong R_{337}(k_{\max}) \cong R_{337}(k_{\max}/2)$ となっていることを意味している。すなわち、 $y_{337}(n)$ はほぼ C_4 単音とみなせる周波数成分で構成されていると判断されるため、 C_4 音の存在を推定することができる。一方、 $R_N(k_{\max}/3)$ については、 N の変化に対してほとんど変動が見られないことが確認できる。

これらの特徴は、基本周波数比が $1:3$ の場合も同様に生じ、その場合は $R_N(k_{\max}/3)$ に同じ特徴が現れる。以上から、零出力検出では判別困難であった2音についても、基本周波数に若干の周波数差が存在すれば、自己相関関数の変化から2音の存在を示す特徴を見出すことが可能であるといえる。

3.5 まとめ

以上より、くし形フィルタ出力に対する自己相関関数の特徴は、以下のようまとめることができる。

- $x(n)$ が単音の場合、 $H_N(z)$ が信号成分をほぼ完全に除去しているならば、非常に小さな値の $R_N(k_{\max})$ が観測される (周期性が弱くなるた

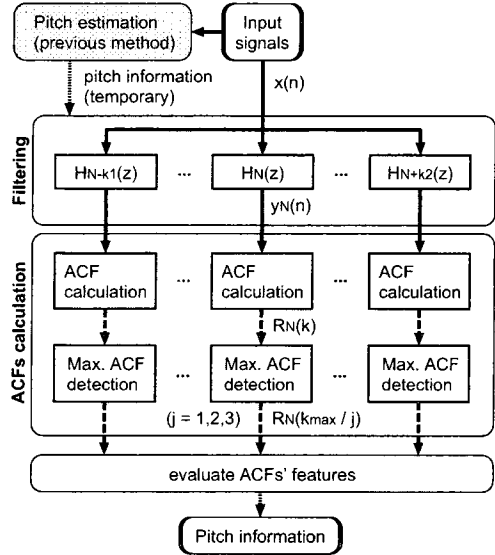


図7 提案する単音/混合音種別推定の流れ

め)。ただし、 N の制限 (整数値) による影響から、 $H_N(z)$ による調波成分の除去が不十分な場合もあり、信号が単音であったとしても必ずしも $R_N(k_{\max})$ が小さな値になるとは限らない。その場合、 $R_N(k_{\max}/j)$ ($j = 1, 2, 3$) には大きな変動は見られない。

- 信号が2音でその基本周波数関係が $1:r$ ($r = 2, 3$) の場合、2音のうち一方を十分に除去/抑圧させる $H_N(z)$ でフィルタリングしたとき、 $R_N(k_{\max}/r)$ に特徴が現れる。それは、
 - $R_N(k_{\max}/r)$ に変動が見られる
 - $R_N(k_{\max})$ と近接することである。

4 単音/混合音種別の推定法 (音高推定法)

4.1 推定手順

前節の結果を基に、図7の流れにしたがい、単音あるいは基本周波数比 $1:r$ の混合音 (2音) の判別を行う。以下に手順を説明する。

1. 入力 $x(n)$ に対し、既知の音高 p ($= 1, 2, \dots, 36$) を基に、 N_p の前後の遅延数で $H_N(z)$ によるフィルタリングを行う。今回は、

$$N_p - k_1 \leq N \leq N_p + k_2 \quad (5)$$

とした。 k_1, k_2 は、それぞれ音高 p の $\pm 50[\text{cent}]$ 範囲に相当するように設定している。

- 各出力 $y_N(n)$ に対して、それぞれ自己相関関数 $R_N(k)$ を求める。
- $R_N(k)$ において、 N_p の近傍でピーク値をとる $R_N(k_{\max})$ を検出する。また、その k_{\max} を基に $R_N(k_{\max}/2)$ 、 $R_N(k_{\max}/3)$ を求める。
- $R_N(k_{\max}/j)$ ($j = 1, 2, 3$) について、それらの最小値がしきい値 Th_1 以下であるか否かを評価する（単音らしさの判定）。すべて Th_1 以下であれば、 $x(n)$ は単音 p として処理を終了する。そうでないならば、5. へ。
- $R_N(k_{\max}/j)$ の変化の幅 $V(j)$ が、しきい値 Th_2 以下であるか否かを評価する（単音らしさの判定）。なお、 $V(j)$ は式 (6) で計算される。すべての $V(j)$ が Th_2 以下であれば、入力信号は単音 p であるとして処理を終了する。そうでないならば、6. へ。

$$V(j) = \max_N R_N \left(\frac{k_{\max}}{j} \right) - \min_N R_N \left(\frac{k_{\max}}{j} \right) \quad (6)$$

- $R_N(k_{\max}/2)$ および $R_N(k_{\max}/3)$ について、より $R_N(k_{\max})$ に近接する方を求め、これを $R_N(k_{\max}/r')$ とする。その後、式 (7) にしたがって $R_N(k_{\max})$ との差 d_N を計算し、その最小値がしきい値 Th_3 以下であるか否かを評価する（混合音らしさの判定）。

$$d_N = R_N(k_{\max}) - R_N \left(\frac{k_{\max}}{r'} \right) \quad (7)$$

d_N の最小値が Th_3 以下であるならば、 $x(n)$ は混合音で2音の関係は $1:r'$ であると推定する。そうでなければ、混合音らしさを検出できなかったと判断され、 $x(n)$ は単音 p として処理をする。

4.2 推定に要するデータ長

本手法で音高推定を実現するために必要な信号サンプル数は、 $H_N(z)$ によるフィルタリング、および、 $R_N(k)$ の計算に使用するデータ点数から、

$$\begin{aligned} \text{サンプル数} &= (N_p + k_2) + (M + N_p + k_2) \\ &= 2(N_p + k_2) + M \end{aligned}$$

となる。このため、 $M = N_p$ (M は $R_N(k)$ の算出に使用する点数) とすると、最も多くサンプルを観測しなければならない場合（音高が $p = 1$ の場合）においても、そのサンプル数は1027点程度 ($\cong 23[\text{ms}]$) である。したがって、非常に短い観測時間で音高推定が実現可能な手法であるといえる。

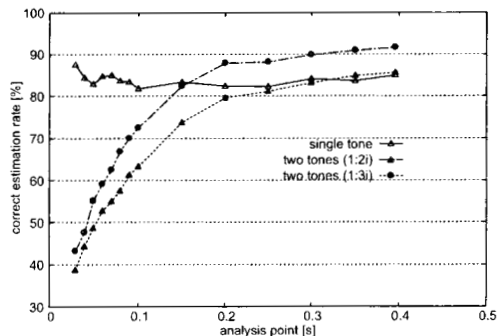


図8 推定開始時刻に対する音高推定正答率の変化

5 実験

5.1 実験条件

RWC 研究用音楽データベース [11] を用いて、提案する音高推定法による実験を行った。使用した楽器音は、10種類の楽器（ピアノ、バイオリン、ビオラ、トランペット、ホルン、アルトサクソ、テナーサクソ、オーボエ、クラリネット、フルート）により演奏された音で、オクターブ3～5の音高範囲のものである（全1032音）。また、混合音は単音の線形和により作成した（ $1:2i$ の混合音は28814種類、 $1:3i$ の混合音は14911種類、ただし、 $i = 1, 2$ ）。

楽音の音高 p は、従来の音高推定法により既知であるとすると、また、実験で使用する各パラメータは、

$$\begin{cases} M = N_p \\ Th_1 = 0.2 \\ Th_2 = 0.3 \\ Th_3 = 0.3 \end{cases}$$

と設定した。

5.2 実験結果

音高推定処理を行う時刻を30[ms]～400[ms]まで変化させたときの、単音、 $1:2i$ の混合音、 $1:3i$ の混合音のそれぞれについて推定正答率を求めた結果を図8に示す。図8より、立上り直後の不安定な信号データを使用した場合の正答率が低いことがわかる。これは、本手法で行っているくし形フィルタによるフィルタリングにおいて、一方の音の除去が不十分であり、推定に使用するための特徴が得られていないことに起因している。結果として、信号の立上りから約200[ms]経過した後に、約80%の正答率で推定可能となることが確認できる。

ここで、処理に使用するデータの開始点が約400[ms]の場合における実験結果の詳細を表1に示す。 $1:2i$ の混合音、および、 $1:3i$ の混合音につ

表 1 信号立上りから 400[ms] 付近における音高推定結果

入力	対象音数	推定結果	割合 [%]	
単音	1032	単音	877	85.0
		1:2i	95	9.2
		1:3i	60	5.8
1:2i	28814	単音	3137	10.8
		1:2i	24668	85.6
		1:3i	1009	3.5
1:3i	14911	単音	1211	8.1
		1:2i	45	0.3
		1:3i	13655	91.6

表 2 基本周波数比 1:2i・1:3i の混合音に対する実験結果の詳細

入力の種類	総数	正答数	割合 [%]
オクターブ 3 と 4 (1:2)	9471	8198	86.6
オクターブ 4 と 5 (1:2)	11862	9855	83.1
オクターブ 3 と 5 (1:4)	7481	6615	88.4
低音がオクターブ 3 (1:3)	8914	8203	92.0
低音がオクターブ 4 (1:3)	4023	3584	89.1
低音がオクターブ 3 (1:6)	1974	1868	94.6

いては、単音と推定する誤りが多いことがわかる。これは、2音の重複する調波成分部分において、くし形フィルタで一方の音を個別に除去（抑圧）できる程度の周波数差が存在しないことが原因である。すなわち、およそ単音とみなせる周波数配置となっていることによるものであり、重複する周波数成分において周波数差を仮定した本手法の限界であると考えられる。さらに、1:2iの混合音と1:3iの混合音について、入力の種類により分類して整理した結果を表2に示す。表2より、オクターブ4と5(1:2)の混合音、および、低音がオクターブ4(1:3)の混合音に対する正当率が多少低いものの、混合音の種類によって極端に正答率が異なる傾向は見られなかった。

6 まとめ

本稿では、調波成分が重複する混合音に対する音高推定を実現するための一手法として、くし形フィルタと自己相関関数に基づく手法について提案した。提案法は、数回のくし形フィルタによるフィルタリングおよび自己相関関数計算により、短時間(23[ms]以内)のデータで音高判別を行える点に特徴がある。実験の結果、信号の立上り付近での推定正答率に課題が残るものの、約400[ms]経過後の信号データに対しては、単音で約85%、基本周波数

比1:2iの混合音で約85%、1:3iの混合音で約91%の正答率で音高推定が行えることを確認した。

今後は、(1)判別法の見直しによる推定精度の改善、特に信号立上り部分における精度の向上、(2)周波数差と推定性能の関係の解明、(3)ユニゾンに対する処理法の確立、などを検討する必要がある。

謝辞

本研究は、文部科学省科学研究費補助金基盤研究(C)「楽音の解析とそのIT社会への応用に関する研究」の援助によって行われた。

参考文献

- [1] 長嶋洋一, 橋本周司, 平賀 譲, 平田圭二, コンピュータと音楽の世界—基礎からフロンティアまで, bit 別冊, 共立出版, 1998.
- [2] A. Sterian and G.H. Wakefield, "Music transcription systems: From sound to symbol," Proc. AAAI-2000 Workshop on AI and Music, 2000.
- [3] 白土 保, "混合音からの基本周波数分離抽出," システム制御情報学会論文誌, vol.12, no.1, pp.56-62, 1999.
- [4] 柏野邦夫, 木下智義, 中臺一博, 田中英彦, "音楽情景分析の処理モデルOPTIMAにおける和音の認識," 信学論 (D-II), vol.J79-D-II, no.11, pp.1762-1770, 1996.
- [5] 三輪多恵子, 田所嘉昭, 斎藤 努, "くし形フィルタを利用した採譜のための異楽器音中のピッチ推定," 信学論 (D-II), vol.J81-D-II, no.9, pp.1965-1974, Sept. 1998.
- [6] 三輪多恵子, "くし形フィルタによる自動採譜システムに関する研究," 博士論文, 豊橋技術科学大学大学院工学研究科, 2001.
- [7] 山口 満, 三輪多恵子, 田所嘉昭, "並列構成くし形フィルタと特異値分解による多重唱の音高推定," 信学論 (D-II), vol.J87-D-II, no.4, pp.1020-1029, April 2004.
- [8] 森田健夫, 山口 満, 田所嘉昭, "並列構成くし形フィルタの出力値に注目した採譜のための音高推定法," 信学論 (D-II), vol.J87-D-II, no.12, pp.2271-2279, Dec. 2004.
- [9] 瀬戸康裕, 夏井雅典, 田所嘉昭, "くし形フィルタと相関関数による音高推定困難和音の音高推定法," 情処学音楽情報科学研報, 2006-MUS-68, pp.1-6, Dec. 2006.
- [10] 柏野邦夫, "音楽を聞き分けるコンピュータ," システム制御情報学会誌「音楽情報処理研究への誘い特集号」, vol.46, no.11, pp.659-664, 2002.
- [11] M. Goto, "Development of the RWC Music Database," Proc. of ICA 2004, pp.I-553-556, April 2004.