

縦続接続くし形フィルタ構成による多重唱の音高推定の検討

山口 満 三輪 多恵子 田所 嘉昭

豊橋技術科学大学 情報工学専攻

あらまし 筆者らは、自動採譜システムのための音高推定の一手法として、くし形フィルタに基づく方式を提案した。各音に対応したくし形フィルタは、その音の基本周波数成分と倍音成分を除去することから、縦続接続したくし形フィルタの零出力検出により音高推定が可能である。しかし、楽器音と比較して、歌唱音は周波数成分にずれやゆらぎを持つことから、縦続接続構成では出力に十分な減衰を得ることができず、音高推定が困難となる問題が生じた。本稿では、従来の手法を基に、二重くし形フィルタを併用した新たな縦続接続構成による音高推定法を提案する。さらに、提案法を用いて実際の歌唱音に対する音高推定を行い、本方式の有効性を確認する。

Pitch Estimation of Sung Songs

Using Comb Filters Connected in Cascade

Michiru Yamaguchi Taeko Miwa Yoshiaki Tadokoro

Department of Information and Computer Sciences, Toyohashi University of Technology

Abstract We have proposed a method of the pitch estimation of musical instrument sounds using comb filters for transcription. But, owing to the fluctuation of song pitches, it was difficult to estimate the pitch of sung songs by the previous method which is based on comb filters connected in cascade. This paper proposes a new cascade structure with double comb filters to adapt the fluctuation of song pitches. We have carried out some computer simulations by the proposed algorithm. As a result, it is clear that the proposed method is useful to estimate the pitch of sung songs.

1 はじめに

楽器音や歌声から楽譜を作成する作業を採譜という。採譜には多くの専門的知識や経験が必要であり、このような能力を兼ねそろえた人でも短時間で正確な採譜は困難であると言われている。そこで近年、電子計算機を用いて採譜を行う自動採譜システムの検討が盛んに行われている [1]-[11]。

現在まで、音楽信号の音高を推定する手段として、高速フーリエ変換 (FFT) [1]-[4] や適応フィルタ [5],[6]、拡張ノッチフーリエ変換 (E-NFT) [7] を用いた方法が提案されている。しかし、平均律音階では隣り合う音高間の周波数比が $2^{1/12}$ 倍と不等間隔であるため、FFT による周波数分析では周波数補正等の処理が必要である。また、楽音には基本周波数成分の他に多数の高調波 (倍音) 成分が含まれているため、音高推定には多周波信号の解析が必要となる。このため、適応フィルタや E-NFT による周波数分析では、計算量の増大や処理系の複雑化の問題が生じた。

上述の問題に対して、我々は $C, C\#, \dots, B$ の 12 音に対応したくし形フィルタ $H_{q,p}(z_p)$ ($p = 1, 2, \dots, 12, q$: フィルタ次数, オクターブに關係) の縦続接続による楽器音解析の一手法を提案し、電

子楽器による異種複数楽器音に対する解析を行った [8]-[10]。 $H_{q,p}(z_p)$ は、 $0[\text{Hz}] \sim fs_p$ (sampling frequency) [Hz] を q 等分する箇所に零点を持ち、問題とされていた倍音成分を除去する性質があるため、その影響を回避することが可能である。また、時間領域において、 $H_{q,p}(z_p)$ はサンプル点の差分により実現されるため、演算量の削減および処理系の単純化が図れる。

しかし、解析対象を人間の歌声 (歌唱音) としたとき、その非定常性から、従来法におけるアルゴリズムでは音高推定が困難となることが明らかとなった。この問題を解決する方法として、我々は二重くし形フィルタ $H_{q,p}^d(z)$ を提案し、この並列接続による音高推定システムを用いてシミュレーションを行うことでその有効性を確認した [11]。ただし、このシステムへの入力は独唱および二重唱に限定しており、多重唱に対応したシステムへ拡張するためには、非常に多数のくし形フィルタを必要とすることから非効率的であるという問題が生じた。以上から、本稿では、再び縦続接続による歌唱音の音高推定法を検討する。

本稿では、はじめに本研究で用いるくし形フィルタの原理を述べる。次に、これまでのくし形フィルタを用いた音高推定法の問題点を説明する。そ

の後、従来の縦続接続構成による音高推定システムを基に、二重くし形フィルタを組み合わせた新たなシステムを提案し、これによる音高推定アルゴリズムを説明する。最後に、実際の歌唱音を用いて計算機シミュレーションを行い、提案方式の有効性を明らかにする。なお、本稿では、オクターブ3, 4の音域における独唱、二重唱を採譜対象とする。また、採譜対象音域をオクターブ3, 4としたため、6次バターワース低域フィルタ(LPF:カットオフ周波数500[Hz])を用いて、入力音の帯域を制限する。

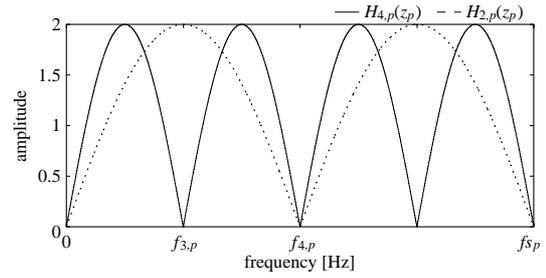


図1 くし形フィルタの周波数特性

2 くし形フィルタによる音高推定の原理

本章では、くし形フィルタによる音高推定原理を説明する。これは、電子楽器音に対して開発したアルゴリズムである[8]-[10]。

2.1 くし形フィルタ

本方式で使用するくし形フィルタの伝達関数は式(1)で表現され、式(2)に示す簡単な減算で実現される。ここで、 $p = 1, 2, \dots, 12$, $q = 2$ (オクターブ4), 4 (オクターブ3)である。

$$H_{q,p}(z_p) = 1 - z_p^{-q} \quad (1)$$

$$y_p(n) = x_p(n) - x_p(n - q) \quad (2)$$

この $H_{q,p}(z_p)$ は、図1に示すように、周波数領域において $0[\text{Hz}] \sim f_{s_p}[\text{Hz}]$ を q 等分する箇所に零点を持つ。なお、 $f_{i,p}$ はオクターブ番号 i の p 音の基本周波数であり、 $f_{s_p} = 2 \cdot f_{4,p}$ である。

図1から、 $H_{4,p}(z_p)$ は p 音の全ての周波数成分を除去できるため、 $y_p(n)_{q=4} = 0$ の検出により入力音 p が判別される。一方、 $y_p(n)_{q=4} \neq 0$ の場合は、 p 音以外の音が入力されたと判断する。

さらに、 $H_{q,p}(z_p)$ の次数と零点の関係を利用して、オクターブの判別を行う。図1に示すように、 $H_{4,p}(z_p)$ がオクターブ3, 4の p 音の全周波数成分を除去可能であるのに対し、 $H_{2,p}(z_p)$ ではオクターブ3の奇数倍音成分が除去不可能である。この性質より、 $H_{4,p}(z_p)$ の零出力を検出(音名を判別)した後、フィルタ次数を4次 \rightarrow 2次に変化させることで、以下のようにオクターブ判別が可能である。

$$\text{output of } H_{2,p}(z_p) \begin{cases} = 0 & \rightarrow \text{octave 4} \\ \neq 0 & \rightarrow \text{octave 3} \end{cases}$$

ただし、このくし形フィルタを縦続接続するため、実際には一定の高いサンプリング周波数 f_s による単位遅延 z^{-1} を用いて各 z_p^{-1} を近似する。こ

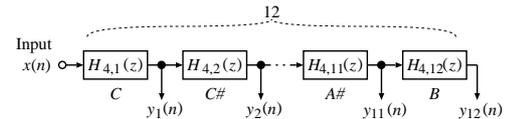


図2 $H_{4,p}(z)$ の縦続接続によるシステム構成

れにより、式(3)に示すように各 $H_{q,p}(z_p)$ を近似的に実現している(オーバーサンプリング法[9])。なお、本方式では f_s として44.1[kHz]を用いる。

$$H_{q,p}(z) = 1 - z^{-q_p} \quad (3)$$

$$q_p = \left\lceil \frac{q \cdot f_s}{f_{s_p}} \right\rceil \quad [\text{整数化}] \quad (4)$$

2.2 音高推定システム

各音に対応した $H_{4,p}(z)$ を図2のように縦続接続することにより音高推定を行う。このシステムは、入力 $x(n)$ に対して音名 C に対応した $H_{4,1}(z)$ から音名 B に対応した $H_{4,12}(z)$ までを1音ずつ順に除去し、 $x(n)$ に含まれる周波数成分をすべて除去し終えた段階で初めて零値を出力する。 $y_p(n)$ で零出力が得られた場合、入力は少なくとも音名 P の音を含んでいることがわかり、零出力検出の前段の出力を用いてオクターブ判別を行い、音高が推定される。次に、 $H_{4,P}(z)$ を縦続の先頭に移動させ、 $x(n)$ を入力する。 $x(n)$ が単音であれば、先頭に移動した $H_{4,P}(z)$ で再度零出力となるため、処理を終了する。一方、 $x(n)$ が和音の場合、 $H_{4,P}(z)$ の出力は非零値となるため、このときの出力 $y'_p(n)$ を新たな入力 $x'(n)$ として、同様に繰り返す。上述の処理により、 $x(n)$ を構成する各音を一音ずつ検出(除去)可能である。

なお、実際的には、上記の処理において完全な零出力を得ることはできない。したがって、まず各 $y_p(n)$ の最大振幅値 a_p から入出力振幅比 $|a_p/a_{p-1}|$ を計算する。求められた12個の入出力振幅比から最小値を求め、これが適当なしきい値以下のとき、その値を示したくし形フィルタにおいて零値が出力されたと判断する。

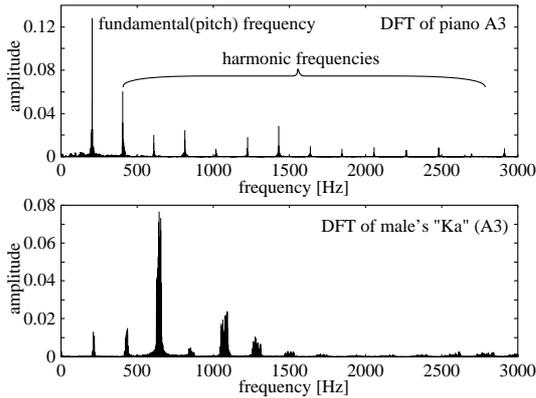


図 3 電子楽器音と歌唱音のスペクトル比較

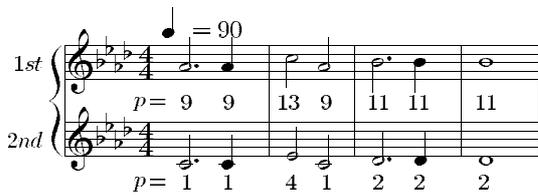


図 4 採譜対象旋律

3 歌唱音への適用と問題点

3.1 歌唱音の特徴

図 3 に、音高 A_3 (オクターブ 3 の A 音、基本周波数 220[Hz]) のピアノ音と男声歌唱“か”の振幅スペクトルを示す。図 3 から明らかなように、楽器音、歌唱音ともに、音高を決定する基本周波数成分の他に多数の高調波 (倍音) 成分が含まれていることがわかる。ここで、楽器音では音高 (周波数) が比較的安定していることから線スペクトルを示しているのに対し、歌唱音はスペクトルに幅を持つことが確認できる。これは、人間が厳密に決められた音高で歌唱を行うことが困難であり、周波数が変動するためである。

3.2 従来システムによる音高推定結果

図 2 の従来法を用いて歌唱音の音高推定を行った結果を示す。ここで、オーバーサンプリング周波数 f_s は 44.1[kHz] とし、音高推定の対象となる歌唱は図 4 による男声二重唱とした。また、前述の LPF により入力の帯域を制限したものを使用した。

採譜対象範囲をオクターブ 3, 4 としたため、音高推定結果は、各時刻 (50[msec] 毎) で推定された音高を $C_3 : p = 1, C\#_3 : p = 2, \dots, B_4 : p = 24$ として出力した。なお、入出力比計算は、その時刻から 400 点のサンプル (採譜対象範囲の下限: $f_{3,1} = 130.81$ [Hz] の 1 周期分以上) を観測して各サンプル値の絶対値を求め、その中で最大のものを振幅 a_p

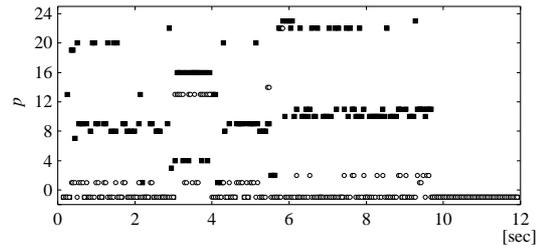


図 5 従来法 (図 2) による音高推定結果

として行った。そして、求められた最小の入出力比がしきい値 (今回は 0.5) 以上となった場合には、信号が十分に減衰していないと判断して、音名判別不能の $p = -1$ を出力している。

図 5 に音高推定結果を示す。図より、音高の誤判別が多く、また、それ以外は判別不能となっている。2 章で述べた $H_{4,p}(z)$ の性質より、理想的な入力信号に対しては $y_p(n)$ は零値を出力するため、これを検出することで音名判別が可能である。しかし、歌唱音は基本周波数にずれやゆらぎを持ち、 $H_{4,p}(z)$ の零点付近の振幅利得が比較的急峻な傾きを持つことから、 $H_{4,p}(z)$ が入力音に対応する場合でも $y_p(n)$ に十分な減衰が得られない。このため、図 5 のような結果が生じたものと考えられる。

3.3 二重くし形フィルタ

前節の問題を解決するために、筆者らは、 $H_{q,p}(z)$ の遅延数を $\pm \Delta q_p$ (整数) だけ変化させた 2 種類のくし形フィルタの縦続接続である、式 (6) の二重くし形フィルタ $H_{q,p}^d(z)$ を提案した [11]。ここで、従来の $H_{q,p}(z)$ における最大利得 2 に対応させるため、式 (6) では係数 1/2 を乗じている。

$$H_{q,p}^{\pm}(z) = 1 - z^{-(q_p \pm \Delta q_p)} \quad (5)$$

$$H_{q,p}^d(z) = \frac{1}{2} \{ H_{q,p}^+(z) \times H_{q,p}^-(z) \} \quad (6)$$

例として、 F_3 に対応する $H_{4,6}^d(z)$ ($\Delta q_6 = 3$ の場合) の周波数特性を図 6 に示す。この図より、 $H_{q,p}^d(z)$ が $H_{q,p}(z)$ と比較して零点付近に広い減衰域を持つため、歌唱音の持つ周波数変動に対応可能である。

実際に隣り合う $F_3(174.61$ [Hz]), $E_3(164.81$ [Hz]) 音の間隔は 9.8[Hz] であり、入力がこの間隔の $1/3 \cong 3.3$ [Hz] 程度の周波数変動を持つ場合、基本周波数成分に生じる $H_{4,6}(z)$ の利得は最大で 10.8% となる。一方、同じ周波数変動に対して $H_{4,6}^d(z)$ ($\Delta q_6 = 3$) の利得は最大で 0.3% 程度であり、基本周波数成分に対する出力誤差の振幅は $H_{4,6}(z)$ を使用した場合の 1/36 程度に減少する。

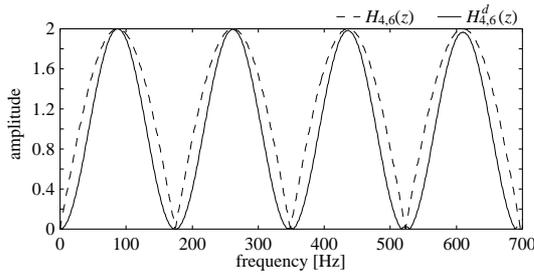


図 6 一重と二重くし形フィルタの周波数特性の比較

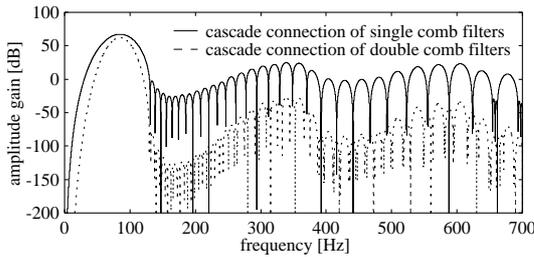


図 7 くし形フィルタの 12 個縦続接続時の周波数特性

3.4 縦続接続システムの周波数特性

一重くし形フィルタ $H_{4,p}(z)$ および二重くし形フィルタ $H_{4,p}^d(z)$ の 12 個縦続接続時の周波数特性を図 7 に示す。これより、くし形フィルタの縦続接続は低周波領域の振幅利得を非常に大きくし、逆に信号領域の振幅利得を極端に低下させてしまうことがわかる。この差は、 $H_{4,p}^d(z)$ を用いた場合にはより顕著なものとなる。このため、雑音成分と信号成分の振幅利得差が広がり、音高推定を困難にさせる結果となった。以上から、 $H_{4,p}^d(z)$ は縦続接続には適さないことがわかる。また、 $H_{4,p}(z)$ においても、縦続接続数を削減して振幅利得差を減少させることが望ましいといえる。

3.5 並列接続システムによる二重唱の音高推定

$H_{4,p}^d(z)$ の縦続接続による音高推定が困難であることから、筆者らは、入力音を独唱または二重唱に限定した図 8 に示すシステムを提案した [11]。図 8 の並列接続構成により、独唱および二重唱の音高推定が行えることを確認した。ただし、図 8 の構成は $H_{4,p}^d(z)$ を 78 個必要とする。三重唱へ対応させるためには 298 個となり、多重唱に拡張するためには非常に多数のくし形フィルタを必要とすることから、非効率であるという問題が生じた。

以上から、従来のくし形フィルタ $H_{4,p}(z)$ の縦続接続による手法を再度検討する必要性が生じた。

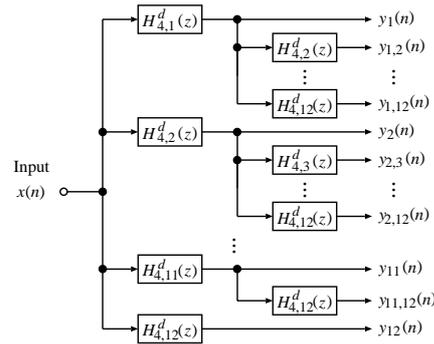


図 8 $H_{4,p}^d(z)$ の並列接続によるシステム

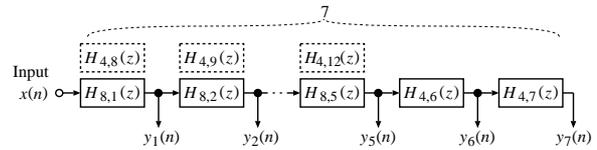


図 9 7 個縦続接続による音高推定システム

4 縦続数の削減および二重くし形フィルタを用いたシステム

本章では、くし形フィルタの縦続接続において縦続数を削減する手法を示し、これを基に、二重くし形フィルタを利用する新たな音高推定システムについて説明する。

4.1 縦続数を削減したシステム

オクターブ i の p 音 ($p = 1 \sim 12$ 各音名に対応) の基本周波数を $f_{i,p}$ で表した場合、平均律音階における隣り合う音高の周波数比は $2^{1/12}$ 倍であるため、以下の関係が成り立つ。

$$\frac{f_{i,p+7}}{f_{i,p}} = 2^{\frac{7}{12}} \cong 1.498307 \cong 1.5$$

したがって、 $H_{4,p}(z)$ の零点を 2 分割する点に新たな零点を設けることで、すなわち、 $H_{4,p}(z)$ の代わりに $H_{8,p}(z)$ を使用することで、 p 音および $p+7$ 音の両方を除去することが可能である。筆者らは、この関係を利用して、くし形フィルタの縦続数を図 9 に示すように 7 個に削減する手法を提案した [10]。ただし、このシステムは、2.2 節で説明した処理の他に、重複音判別の処理を必要とする (4.3 節で説明)。

このシステムにおけるくし形フィルタ 7 個縦続接続時の周波数特性を図 10 に示す。図 10 から、縦続数を削減することにより、低域振幅利得の増加の抑制、および、信号領域における振幅レベルの低下を回避する効果が期待できる。

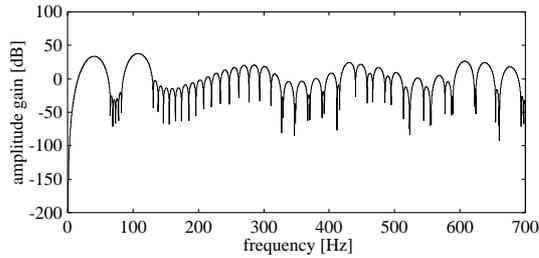


図 10 くし形フィルタの 7 個縦続接続時の周波数特性

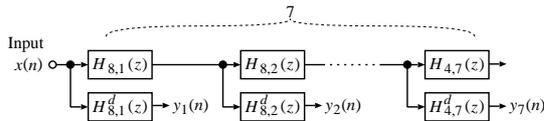


図 11 二重くし形フィルタを併用したシステム

4.2 二重くし形フィルタの利用

図 9 のシステムは、従来のくし形フィルタを使用しているために、歌唱音の周波数変動には十分に対応できない。そこで、図 9 に対し、二重くし形フィルタを組み合わせた図 11 のシステムを新たに提案する。

このシステムは、入出力比計算のために観測する $y_p(n)$ を、二重くし形フィルタを用いて求める。これにより、入力音が周波数変動を持つ場合でもフィルタに対応する音を除去することが可能となる。つまり、零・非零出力が明確となり、誤検出を減少させることができる。また、零出力検出後、縦続接続の先頭に配置するフィルタも二重くし形フィルタとすることで、より音高推定精度が向上するものと考えられる。

図 11 により、信号領域における振幅レベル低下の回避、かつ、歌唱音の非正常性への対応という両者の要求を満たすことが可能となる。

4.3 提案システムによる音高推定アルゴリズム

図 11 を用いた音高推定アルゴリズムを説明する。

まず、2.2 節で述べた基本的な音高推定アルゴリズムと同様に、図 11 において零出力検出を行い、零値を出力する $H_{q,p}^d(z)$ を明らかにする。ここで、 $H_{q,p}^d(z)$ が他音と重複関係を持たないもの ($H_{4,6}^d(z), H_{4,7}^d(z)$) である場合、前述と同様に p 音のオクターブ判別を行い、 $H_{q,p}^d(z)$ を縦続の先頭に移動させる。

一方、 $H_{q,p}^d(z)$ が他音と重複関係を持つ場合には、零出力が得られた前段の出力を新たな入力と考え、個別に重複音の判別を行う。すなわち、 $H_{4,p}^d(z)$ および $H_{4,p+7}^d(z)$ に入力して零出力を検出すること

で、音名 p 、もしくは $p+7$ を推定することが可能である。非零値であった場合には、両者の縦続接続 $H_{4,p}^d(z) \times H_{4,p+7}^d(z)$ に入力し、零出力検出により p と $p+7$ の和音と推定することができる。この後、同様にオクターブ判別を行い、音高を検出する。

上記の処理を繰り返すことにより、入力 $x(n)$ の音高を順次推定可能である。

5 シミュレーション

5.1 シミュレーション条件

3.2 節でのシミュレーション条件と同じく、 $f_s = 44.1$ [kHz] であり、入力 LPF により帯域を制限した図 4 に示す旋律の男声二重唱とした。音高推定結果も同様に、 $p = 1 \sim 24$: $C_3 \sim B_4$ に対応しており、入出力振幅比がしきい値 (0.5) 以上となった場合には、音名判別不能の $p = -1$ を出力する。

5.2 シミュレーション結果

縦続数を 7 個に削減したシステム (図 9)、および、今回新たに提案する二重くし形フィルタを組み合わせたシステム (図 11) のそれぞれを用いて音高推定を行った結果を図 13, 14 に示す。なお、評価基準として、二重くし形フィルタ ($\Delta q_p = 3$) の並列接続方式 (図 8) を用いた場合における図 12 の結果を使用する。図 8 は、二重くし形フィルタを用いて、かつ最小の縦続数であるため、最も精度よく音高を推定するものとして評価基準とした。

図 12 を基準 (100%) として、各結果を比較したときの正答率を表 1 に示す。なお、括弧内の値は、オクターブ情報を除いた音名推定結果のみを比較したときの正答率である。表より、縦続数を削減することで、音高推定精度の改善はかれることを確認できる。また、二重くし形フィルタを利用する提案法を用いた場合では、さらに約 10% 程度、精度が改善される。ソフトウェア的な後処理を施すことにより、これらの音高推定結果上に短時間 (連続せずに) 存在する誤判別を整形することで、音価 (音の継続時間) の検出も実現可能である。

以上の結果から、提案方式が、二重唱の音高推定法として有効であることを確認した。

6 むすび

本稿では、歌唱音の音高推定の一手法として、くし形フィルタの縦続接続による方式を検討した。従来の $H_{4,p}(z)$ の 12 個縦続接続方式、縦続数を削減した方式、および、今回新たに提案した縦続数の削減かつ二重くし形フィルタを組み合わせる手法

表 1 音高推定正答率の比較:単位 [%]

| method (result) | upper tone | lower tone |
|-----------------|----------------------|----------------------|
| Fig.2 (Fig.5) | 41.35 (51.48) | 45.99 (51.90) |
| Fig.9 (Fig.13) | 62.87 (68.78) | 59.49 (64.98) |
| Fig.11 (Fig.14) | 73.00 (82.28) | 69.20 (78.48) |

() : only pitch name

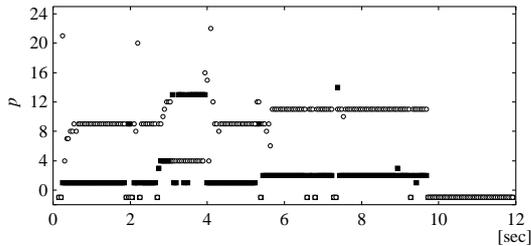


図 12 図 8 による音高推定結果

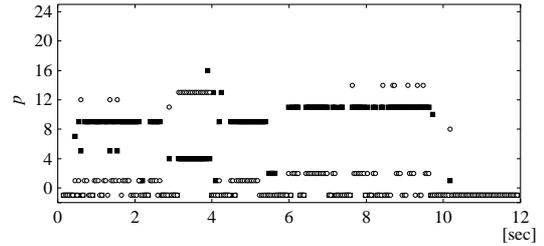


図 13 図 9 による音高推定結果

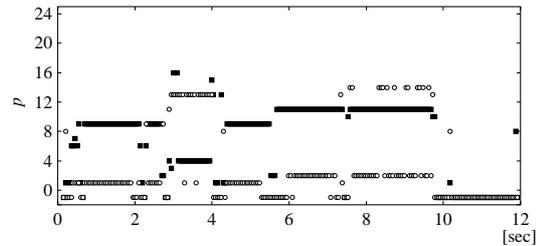


図 14 図 11 による音高推定結果

のそれぞれを用いて、実際の二重唱に対する音高推定を行った。この結果、縦続数を削減することで音高推定精度が向上することが明らかとなった。さらに、 $H_{q,p}^d(z)$ の併用により良好な結果を得ることが可能であり、提案方式が有効であることを確認した。

今後の課題としては、以下が挙げられる。

(1) 三重唱以上の多重唱に対する検討

本稿では、二重唱を入力として扱い、提案方式の有効性を検証した。しかし、二重唱に対して音高推定が可能であることを示したに過ぎず、今後は、提案方式による三重唱以上の音高推定が可能であるかを検討する必要がある。

(2) 採譜対象音域の拡大

現在は、採譜対象音域をオクターブ 3~4 と定めて検討を行っているが、これは実際の歌唱における音域を網羅しているとは言いがたい。対象音域を現実的な範囲に拡大させる検討が必要である。

以上の課題について、今後引き続き検討を行い、最終的に実用的な歌唱の自動採譜システムの構築を目指したい。

参考文献

[1] 井口征士, “音楽情報の処理-電子計算機を用いた自動採譜,” 計測自動制御学会論文集, Vol.19, No.3, pp.314-319, 1980.

[2] 原裕一郎, 井口征士, “複素スペクトルを用いた周波数同定,” 計測自動制御学会論文集, Vol.19, No.9, pp.718-723, 1983.

[3] 新原高水, 今井正和, 井口征士, “歌唱の自動採譜,” 計測自動制御学会論文集, Vol.20, No.10,

pp.940-945, 1984.

[4] 高沢嘉光, “計算機による採譜,” 情報処理学会誌, Vol.29, No.6, pp.593-598, 1988.

[5] 梅本敏孝, 青島伸治, “適応技術を用いた採譜に有効な周波数分析法,” 計測自動制御学会論文集, Vol.28, No.5, pp.619-625, 1992.

[6] 梅本敏孝, 青島伸治, “適応フィルタによる周波数分析におけるステップサイズパラメータの選定,” 計測自動制御学会論文集, Vol.28, No.10, pp.1257-1262, 1992.

[7] 門藤博, 田所嘉昭, “拡張ノッチフーリエ変換による採譜のための音高抽出,” 計測自動制御学会論文集, Vol.31, No.12, pp.2005-2010, 1995.

[8] 三輪多恵子, 田所嘉昭, 斎藤努, “零出力に注目したくし形フィルタによる音階検出,” 電気学会論文誌 C, Vol.118-C, No.1, pp.57-64, 1998.

[9] 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,” 電子情報通信学会論文誌 D-II, Vol.J81-D-II, No.9, pp.1965-1974, 1998.

[10] 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.

[11] Yoshiaki Tadokoro and Michiru Yamaguchi, “Pitch Detection of Duet Song using Double Comb Filters,” European Conference on Circuit Theory and Design, 2001(ECCTD'01), Espoo, Finland, Aug.2001.