

くし形フィルタによる 実楽器音を対象とした演奏楽器推定

三輪 多恵子[†] 田所 嘉昭[†] 斎藤 努^{††}

[†]豊橋技術科学大学・工学部・情報工学系
^{††}豊田高専・電気工学科

[†]〒 441-8580 愛知県豊橋市天伯町字雲雀ヶ丘 1-1

^{††}〒 471-8525 愛知県豊田市栄生町 2-1

著者らは、周波数軸上で等間隔に零点を持ち、サンプリング周波数に対応した音高中の全ての周波数成分を除去できるくし形フィルタを利用して、音名判別、和音分離（スペクトル抽出）、スペクトル構造を利用した楽器推定法を提案した。しかし、実際の楽器音にはピッチのずれや変動が含まれ、固定零点のくし形フィルタによる手法では、その精度に問題が生じた。この問題を解決するために、くし形フィルタの零点を適応的に移動させ、入力音のピッチのずれに対応する方式を提案する。適応的くし形フィルタにより、厳密な入力ピッチの箇所に零点を設けることで精度良く和音分離が可能であり、これにより楽器の推定精度の向上が図れる。本稿では、提案手法を説明し、実楽器音を用いた計算機シミュレーションにより、その有効性を確認する。

採譜、ピッチ推定、演奏楽器推定、適応的くし形フィルタ

The musical instruments estimation of a real musical sound using comb filters for transcription

Taeko MIWA[†] Yoshiaki TADOKORO[†] Tsutomu SAITO^{††}

[†]Toyohashi University of Technology
^{††}Toyota National College of Technology

[†]Hibarigaoka 1-1, Tempaku-cho, Toyohashi, 441-8580 Japan

^{††}2-1, Eisei-cho, Toyots, 471-8525 Japan

In this paper, we propose a method of the pitch frequency estimation and the estimation of playing musical instruments for a real musical sound using adaptive comb filters. The frequency components of the musical tone can be eliminated by comb filters, if zero points are corresponding to the pitch frequency. The real musical sound which is not synthesized has a difference of the pitch frequency from a standard temperament scales. In the proposed method, the positions of the zero points of an adaptive comb filter are changed as decreasing the amplitude of its output, and strict pitch frequencies can be estimated. We have done some computer simulations on transcribing the mixed sounds of the real musical tone. As a result, it is clear that the proposed method is useful to estimate pitch frequencies and musical instruments.

transcription, pitch estimation, musical instruments estimation, adaptive comb filters

1 まえがき

採譜とは楽音から楽譜を作成する作業であり、専門的な知識や経験が必要とされるため、計算機を用いた自動採譜の研究が盛んに行なわれている [1]-[11]。また、近年では異楽器和音の音高及び演奏楽器の推定へ、採譜対象が拡張される傾向にあるが、以下の理由から、その解析は大変難しいとされている [5]-[8],[10],[11]。

- 音高の推定—平均律音階では、隣接する音の周波数間隔が不等間隔 ($2^{1/12}$ 倍) であるため、FFT では周波数補正等の特別な処理が必要 [2]。
- 楽器の判別—各楽器の倍音構造を利用したテンプレートマッチング法が一般的であるが、異楽器により発せられた各音の倍音成分が複雑に重畳し、各音の個別認識が非常に困難 [8]。

筆者らは、上記の問題点を考慮し、 $C, C\#, \dots, B$ の 12 音に対応したくし形フィルタ $H_{q,p}(z_p) = 1 - z_p^{-q}$ ($p = 1, 2, \dots, 12$, q : フィルタ次数) による楽音解析法を提案した [9]-[11]。 $H_{q,p}(z_p)$ は、 $0 \sim f_s p$ (sampling frequency) [Hz] を q 等分する箇所に零点を持つため、楽音に含まれる倍音成分を全て除去可能であり、その影響を回避できる。

この性質を利用し、縦続接続した $H_{q,p}(z_p)$ の零出力による音高検出法 [9]、及び $H_{q,p}(z_p)$ による異楽器和音の分離方法 [10]、スペクトル構造に注目した楽器推定法 [11] を提案した。また、MIDI(KORG:05R/W) 音源を用いた計算機シミュレーションを行ない、本方式による異楽器音解析の可能性を示した。

しかし、実楽器音は一般に基本周波数のずれや変動を持ち、その解析は容易ではない。本稿では、基本周波数のずれに対して $H_{q,p}(z_p)$ の遅延数 (零点の位置) を適応的に変化させる方式を提案する。また、本方式の性能を明らかにする目的で、調律の合っていない piano、素人の演奏した管楽器 (clarinet, alto-sax, cornet, horn) に対して音高推定、和音分離、スペクトル構造に注目した演奏楽器推定を行なった結果について報告する。

なお、対象音域は #3 (オクターブ 3) ~ #5 の 36 音とし、予め録音した各楽器の単音を計算機上で合成することで異楽器和音を作成した。また、複数の楽器により同一音高が演奏されないことを入力条件とする。

本稿では、1. で実楽器音の性質を述べる。2. で、固定零点のくし形フィルタ $H_{q,p}(z_p)$ の原理を述べ、 $H_{q,p}(z_p)$ を利用した音高推定法、および和音分離法を説明する。また、ここでは、基本周波数のずれに対する $H_{q,p}(z_p)$ の特性と適応的くし形フィルタ $\hat{H}_{q,p}(z_p)$ の原理について、併せて述べる。4. で、スペクトル構造に基づいた楽器推定法の原理を説明する。最後に、5. では計算機シミュレーションにより、本方式の性能を明らかにする。

2 実楽器音の性質

本章では、本方式の検討課題である、実楽器音の基本周波数のずれの条件とその誤差範囲について述べる。

2.1 音階全体の基本周波数がずれを持つ場合

平均律音階の基本周波数は、基準ピッチ A_4 (#4 の A 音の基本周波数) を 440.0 [Hz] とし、隣接音の周波数比が $2^{1/12}$ 倍となるように定められている。しかし、近年では、「音に張りが出る」「ホールの隅々まで音が響く」等の理由から、基準ピッチを高めにチューニングするケースが多い [12]。さらに、楽器は温度や湿度等の条件によりピッチが変動するため、時と場所に応じて基準ピッチを変化させることは、ごく一般的に行なわれる。

基準ピッチの変動は、オーケストラに依存するもの、ホールの音響条件によるもの等があるが、一般に 440.0 ± 5.0 [Hz] 程度とされている [12]。この場合、 $A_4 = 440.0$ [Hz] に基づく平均律音階からのずれ幅は、半音の周波数間隔の $5.0 / \{A\#_4 - A_4 = 26.16\} = 19.11\%$ 程度となる。

2.2 特定音の基本周波数がずれを持つ場合

複数音源による演奏では、和声理論に基づき、和音に含まれる特定音の基本周波数のみをずらす場合がある [13]。ここでは、 C, E, G の長三和音を例にとり説明する。

オクターブ番号 o の p 音の基本周波数を $f_{o,p}$ で表すと、平均律音階における各音の周波数関係は式 (1) となる。

$$\begin{aligned} f_{o,E} &= 2^{\frac{4}{12}} \cdot f_{o,C} \cong 1.260 \cdot f_{o,C} \\ f_{o,G} &= 2^{\frac{7}{12}} \cdot f_{o,C} \cong 1.498 \cdot f_{o,C} \end{aligned} \quad (1)$$

ここで、 $f_{o,E}$ を 4 倍、 $f_{o,G}$ を 2 倍すると、それぞれ C の 5 倍音、3 倍音と非常に近い値であることが分かる。

$$\begin{aligned} 4 \cdot f_{o,E} &= 4 \times 2^{\frac{4}{12}} \cdot f_{o,C} \cong 5.039 \cdot f_{o,C} \\ 2 \cdot f_{o,G} &= 2 \times 2^{\frac{7}{12}} \cdot f_{o,C} \cong 2.997 \cdot f_{o,C} \end{aligned} \quad (2)$$

この性質から、 E の 4 倍音と G の 2 倍音が C の対応した成分 ($5 \cdot f_{o,C}$, $3 \cdot f_{o,C}$) と一致するように、故意に基本周波数をずらして演奏される。新たな基本周波数 $f'_{o,p}$ は式 (3) となり、半音の周波数間隔の E : 13.45%, G : 1.67% 程度のずれが生じる。

$$\begin{aligned} f'_{o,E} &= 5.0 \cdot f_{o,C} \div 4.0 \cong 0.992 \cdot f_{o,E} \\ f'_{o,G} &= 3.0 \cdot f_{o,C} \div 2.0 \cong 1.001 \cdot f_{o,G} \end{aligned} \quad (3)$$

上記以外に、ピアノ等の調律のずれや演奏者の技量に依存するピッチのずれ等があり、それらのずれ幅を予め予測することは困難である。後述のように、本方式においてピッチ推定結果が保証されるずれ幅は、半音の周波数間隔の 1/3 程度であるため、それ以上のずれが生じる場合には誤判別の可能性がある。

3 くし形フィルタを用いた音高検出と和音分離

従来の固定零点のくし形フィルタ $H_{q,p}(z_p)$ のピッチのずれに対する特性、および、周波数軸上での零点の移動による適応的くし形フィルタ $\hat{H}_{q,p}(z_p)$ の特性を説明する。なお、説明を容易にするために、従来の固定零点の $H_{q,p}(z_p)$ の原理、 $H_{q,p}(z_p)$ による音高検出、および和音分離法を簡単に説明する [9]-[11]。

3.1 固定零点のくし形フィルタの原理

固定零点のくし形フィルタ $H_{q,p}(z_p)$ の伝達関数は式 (4) であり、図 1 に示すような周波数特性を持つ。また、式 (4) は式 (5) に示す減算で実現される。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、フィルタ次数 q は各オクターブの周波数間隔に応じて #3 で $q=8$ 、#4 で $q=4$ 、#5 で $q=2$ の値をとる。

$$H_{q,p}(z_p) = 1 - z_p^{-q}, \quad p = 1, 2, \dots, 12 \quad (4)$$

$$y_{q,p}(n) = x_p(n) - x_p(n-q) \quad (5)$$

ここで、 $H_{8,p}(z_p)$ は #3~#5 の p 音の全ての周波数成分を除去できるため、 $y_{8,p}(n) = 0$ より入力音 p が検出される。逆に、 $y_{8,p}(n) \neq 0$ の場合は、それ以外の音が入力されたと判断する。

$$y_{8,p} \begin{cases} = 0 & \rightarrow \text{tone } p \\ \neq 0 & \rightarrow \text{otherwise} \end{cases}$$

さらに、オクターブ検出は各 $H_{q,p}(z_p) : q = 2, 4, 8$ の零点の位置関係から、 $H_{q,p}(z_p)$ の次数を $8 \rightarrow 4$ 次、 $4 \rightarrow 2$ 次に変化させ、非零出力を検出することで順次判別可能である。

$$y_{8,p} = 0 \ \& \ y_{4,p} \neq 0 \quad \Rightarrow \quad \#3$$

$$y_{8,p}, y_{4,p} = 0 \ \& \ y_{2,p} \neq 0 \quad \Rightarrow \quad \#4$$

$$y_{8,p}, y_{4,p}, y_{2,p} = 0 \quad \Rightarrow \quad \#5$$

なお、各 $H_{q,p}(z_p)$ は $f_{s,p}$ が異なるため、over-sampling 法により、その縦続接続を実現する。over-sampling 周波数 f_s による $x(n)$ の単位遅延 z^{-1} ($1/f_s$ [s]) を使用し、

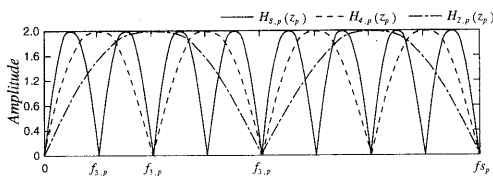
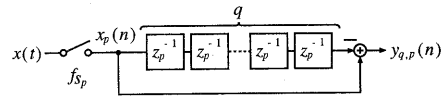
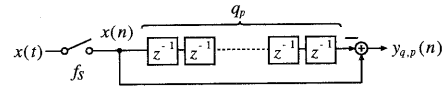


図 1 くし形フィルタの周波数特性



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



(b) $H_{q,p}(z_p) \cong 1 - z_p^{-q_p}$

図 2 Over-sampling 法による $H_{q,p}(z_p)$ の近似

$x_p(n)$ の単位遅延 z_p^{-1} ($1/f_{s,p}$ [s]) を近似することで、各 $H_{q,p}(z_p)$ を式 (6) のように近似的に実現する (図 2)。

$$H_{8,p}(z_p) \cong 1 - z^{-q_p} \quad (6)$$

ここで、

$$q_p \cong q \times \frac{f_s}{f_{s,p}} \quad [\text{四捨五入で整数化}] \quad (7)$$

なお、over-sampling 法により、 $H_{q,p}(z_p)$ の零点周波数には最大で 0.13% 程度のずれが生じる [9]。

3.2 音高検出

本方式では、C, C#, ..., B に対応した 12 種類のくし形フィルタ $H_{8,p}(z_p)$ を、図 3 に示すように縦続接続することで、音高検出システムを構成する。

このシステムでは、入力 $x(n) = \sum \text{tone } p_i$ ($1 \leq i \leq m$) に対して、最も高い音 p_m に対応した $H_{8,p_m}(z_{p_m})$ で最初にゼロ出力が検出される。この結果から、 $x(n)$ が少なくとも p_m 音を含むことが判別できる。

step 1. $y_{8,p_m} = 0 \rightarrow$ detecting tone p_m

次に、 $H_{8,p_m}(z_{p_m})$ を縦続の先頭に移動する。

step 2. $H_{8,p_m}(z_{p_m})$ is placed in front of $H_{8,1}(z_1)$

$x(n)$ が単音であれば、先頭に移動した $H_{8,p_m}(z_{p_m})$ で再びゼロ出力が検出されるため、処理を終了する。一方、

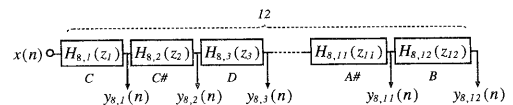


図 3 $H_{8,p}(z_p) : p = 1, 2, \dots, 12$ の縦続接続によるシステム構成

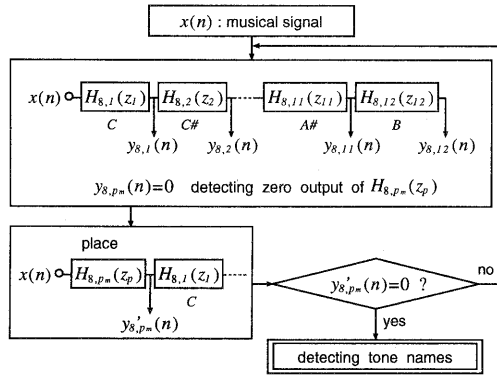


図4 音名判別アルゴリズム

$x(n)$ が和音の場合、移動した $H_{8,p_m}(z_{p_m})$ では非ゼロ出力が検出される。

step 3. $y'_{8,p_m} \begin{cases} = 0 & \rightarrow \text{monophony } p_m. \\ \neq 0 & \rightarrow \text{polyphony} \end{cases}$

step 3. で $y'_{8,p_m} \neq 0$ である場合、再び $H_{8,1}(z_1)$ から順に同様な処理を行なう。なお、移動させた $H_{8,p_m}(z_{p_m})$ により、次の $H_{8,p_1}(z_{p_1})$ に対する入力 $x'(n)$ からは p_m 音成分が除去され、 $x(n)$ 中の音の数は $m-1$ となる。

step 4. $x'(n) = x(n) - [p_m] \Rightarrow$ step 1.

上述の処理を繰り返すことで、 $x(n)$ を構成する各音が1音づつ検出(除去)される(図4)。さらに、オクターブ推定を行ない音高(基本周波数)を検出する。

なお、実際には、非調和周波数成分、ピッチのずれ等の影響を考慮し、 $|y_{8,p_i}(n)|/|y_{8,p_{i-1}}(n)|$ の比から、最も大きく減衰した段を入力音と判別する。次節では、ピッチのずれ幅に対する $H_{q,p}(z_p)$ の特性から、この音高判別法における周波数のずれ幅の許容範囲を明らかにする。

3.3 ピッチのずれに対する $H_{q,p}(z_p)$ の特性

入力音のピッチがずれを持つ場合、前節の音高判別で最も厳しい条件は隣接音の検出である。ここで、オクターブ番号 o の p_1 音と、隣接する p_2 音の和音を仮定し、対応した $H_{q,p_1}(z_{p_1})$ 、 $H_{q,p_2}(z_{p_2})$ の出力特性を考察する。なお、 p_1 音の基本周波数を f_{o,p_1} とし、 p_2 音の基本周波数は規定される f_{o,p_2} から $-\Delta f_{p_2}$ のずれを持つと仮定する。

ここで、 $H_{q,p_2}(z_{p_2})$ の各零点 $k \cdot f_{o,p_2}$ と p_2 音の倍音周波数のずれは、式(8)より、 $k \cdot \Delta f_{p_2}$ となる。

$$\underbrace{k \text{ 倍音の周波数}}_{k \cdot (f_{o,p_2} - \Delta f_{p_2})} - k \cdot f_{o,p_2} = -k \cdot \Delta f_{p_2} \quad (8)$$

k 倍音の振幅を b_k とし、各倍音に生じる利得を M_k とするならば、 $H_{q,p_2}(z_{p_2})$ の出力振幅 Y_{p_2} は式(9)となる。こ

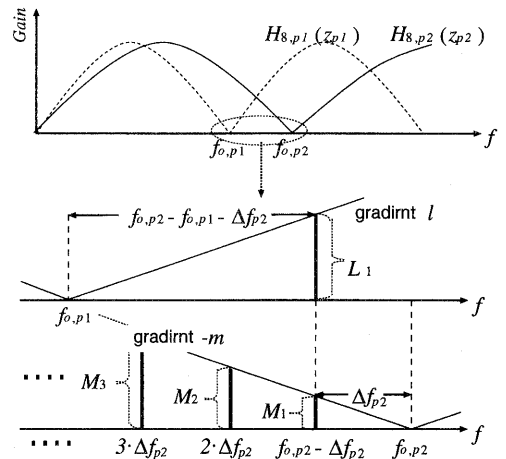


図5 周波数のずれ Δf_2 に対する $H_{q,p_i}(z_{p_i})$ の特性

ここで、図5に示すように、零点付近の $|H_{q,p_2}(z_{p_2})|$ の傾きを m の直線で近似すると、 M_k は式(10)で得られる。

$$Y_{p_2} = \sum_{k=1} M_k \cdot b_k \quad (9)$$

$$M_k = m \cdot k \cdot \Delta f_{p_2} \quad (10)$$

一方、 $|H_{q,p_1}(z_{p_1})|$ の $f_{o,p_2} - \Delta f_{p_2}$ 成分に生じる利得を L_k とすると、その出力振幅 Y_{p_1} は式(11)となる。また M_k と同様に、零点付近の傾きを l で直線近似すると L_k は式(12)で得られる。

$$Y_{p_1} = \sum_{k=1} L_k \cdot a_k \quad (11)$$

$$L_k = l \cdot k \cdot (f_{o,p_2} - f_{o,p_1} - \Delta f_{p_2}) \quad (12)$$

ここで、 p_1 音の k 倍音の振幅を a_k とおくと、 $p_1 + p_2$ の和音に対する各 $H_{q,p}(z_p)$ の出力は次のように変化する。

$$\begin{aligned} \text{input polyphony} & \quad \sum a_k + \sum b_k \\ & \quad \downarrow \\ \text{output of } H_{q,p_1}(z_{p_1}) & \quad \sum 0 \cdot a_k + \sum L_k \cdot b_k \\ & \quad \downarrow \\ \text{output of } H_{q,p_2}(z_{p_2}) & \quad \sum M_k \cdot L_k \cdot b_k \end{aligned}$$

前述のように、 p_2 音の検出には、 $H_{q,p_2}(z_{p_2})$ を通過する際に前段 $H_{q,p_1}(z_{p_1})$ の通過時以上の減衰が必要となる。従って、 p_2 音検出のための各 $H_{q,p_i}(z_{p_i})$ の利得 L_k 、 M_k の条件は式(13)となる。

$$\frac{\sum L_k \cdot b_k}{\sum b_k + \sum a_k} > \frac{\sum L_k \cdot M_k \cdot b_k}{\sum L_k \cdot b_k} \quad (13)$$

ここで、 $H_{q,p1}(z_{p1})$ と $H_{q,p2}(z_{p2})$ の傾きの絶対値 $|m| \cong |l|$ がほぼ等しく、両音の振幅に大きな違いが無い $\sum a_k \cong \sum b_k$ とするならば、式 (10), (12) より、式 (13) は次式となる。

$$\frac{\sum L_k}{2 \sum 1} > \frac{\sum M_k}{\sum 1} \text{ より,}$$

$$f_{o,p2} - f_{o,p1} - \Delta f_{p2} > 2 \Delta f_{p2}$$

$$\Delta f_{p2} < \frac{f_{o,p2} - f_{o,p1}}{3} \quad (14)$$

従って、ピッチのずれが半音の周波数間隔の 1/3 程度ならば、各 $H_{q,p}(z_p)$ の出力の減衰により音名判別が可能である。しかし、振幅 b_k が小さい場合や、 $f_{o,p2} - \Delta f_{p2}$ を極端に減衰させる $H_{q,p}(z_p)$ を通過する場合、許される変動幅は式 (14) より狭くなる。

3.4 和音分離とスペクトル抽出

入力された異楽器和音 $x(n) = \sum \text{tone } p_i (1 \leq i \leq m)$ に対して、式 (15) に示す合成フィルタ $G_{8,p_r}(z_{p_r})$ を使用することで、 p_r 音の抽出が可能である。

$$G_{8,p_r}(z_{p_r}) = \prod_{i=1, \neq r}^m H_{8,p_i}(z_{p_i}) \quad (15)$$

図 6 に示すように、 $x(n)$ に含まれる全ての音高に対して $G_{8,p_r}(z_{p_r})$ を構成することで、 $x(n)$ が単音成分に分離される。但し、合成フィルタの周波数特性は式 (16) となるため、抽出音の各周波数成分の振幅は、 $|G(e^{j\theta})|$ で計算される利得が乗じられた値に変化する。

$$|G(e^{j\theta})| = \prod_{i=1, \neq r}^m |H_{8,p_i}(e^{j\theta})|$$

$$= \prod_{i=1, \neq r}^m \left| 1 - \exp\left(-j \frac{8 \cdot 2\pi f}{f_{8p_i}}\right) \right|$$

$$= \prod_{i=1, \neq r}^m \sqrt{2 - 2 \cos\left(\frac{8 \cdot 2\pi f}{f_{8p_i}}\right)} \quad (16)$$

さらに、DFT を用いて抽出音の周波数解析を行なう。DFT では、入力 $\{x(nT) : n = 0, 1, \dots, N-1\}$ に対する

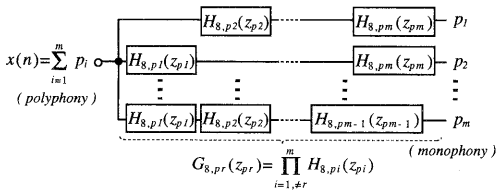


図 6 和音分離システム

周波数スペクトル $\{X(e^{j\omega_k T})\}$ が式 (17) で与えられる。なお、 $\omega_k = 2\pi k/N T = k \cdot \omega_1$ であり、DFT により解析可能な周波数は ω_1 の整数倍に制限されている。

$$X(e^{j\omega_k T}) = \sum_{n=0}^N x(nT) \exp\left(-j \frac{2\pi k}{N} n\right) \quad (17)$$

抽出された p_r 音の周波数成分は f_{3,p_r} の整数倍に限られるため、式 (18) を満たす N_{p_r} 点の DFT により周波数解析が可能となる。ここで、 p_r 音のサンプル値系列 $x_{p_r}(n)$ の周波数スペクトルは式 (19) となる。

$$N_{p_r} = \frac{2\pi f_s}{\omega_1} = \frac{f_s}{f_{3,p_r}} \text{ [四捨五入で整数化]} \quad (18)$$

$$X_{p_r,k} = X_{p_r}(e^{j\omega_k})$$

$$= \sum_{n=0}^{N_{p_r}} x_{p_r}(n) \exp\left(-j \frac{2\pi k}{N_{p_r}} n\right) \quad (19)$$

なお、前述のように、 $G_{8,p_r}(z_{p_r})$ の利得により抽出音のスペクトルは元信号とは異なる振幅を持つ。従って、DFT で得られた結果を、式 (16) より得られる各利得で除算することで元のスペクトルを復元する。

3.5 適応的くし形フィルタの原理

入力音の基本周波数にずれがある場合、前述の固定零点のくし形フィルタ $H_{q,p}(z_p)$ では除去しきれない成分が残され、和音分離（スペクトル抽出）の精度に問題が生じる。本方式では、式 (20) に示す適応的くし形フィルタ $\hat{H}_{q,p}(z_p)$ を構成することで、この問題を解決する。

$$\hat{H}_{q,p}(z_p) = 1 - z^{-\{q_r + \alpha\}}, \quad \alpha: \text{可変遅延数} \quad (20)$$

オクターブ番号 o の p 音の基本周波数が $f_{o,p}$ から Δf_p のずれを持つ場合、 $H_{q,p}(z_p)$ の零点 $k \cdot f_{o,p}$ とその倍音周波数のずれは、式 (8) の関係より、 $k \cdot \Delta f_p$ となる。ここで、 k 倍音の振幅を a_k とし、3.3 で述べたように、零点付近の $|H_{q,p}(z_p)|$ の傾きを m の直線で近似するならば、各倍音に生じる利得 M_k およびは $H_{q,p}(z_p)$ の出力振幅 Y_p は式 (21), (22) で得られる。

$$Y_p = \sum_{k=1} M_k \cdot a_k \quad (21)$$

$$M_k = m \cdot k \cdot \Delta f_p \quad (22)$$

一方、 $\hat{H}_{q,p}(z_p)$ の各零点周波数 $k \cdot \hat{f}_{o,p}$ は式 (23) であり、前述の零点と倍音周波数のずれ $k \cdot \Delta \hat{f}_p$ 、その出力振幅 \hat{Y}_k 、利得 \hat{M}_k は、それぞれ、式 (24), (25), (26) となる。なお、 \hat{m} を $|\hat{H}_{q,p}(z_p)|$ の零点付近の傾きとし、 δ を $f_{o,p} - \hat{f}_{o,p}$ とする。

$$k \cdot \hat{f}_{o,p} = k \cdot \frac{f_s}{q_p + \alpha} \quad (23)$$

$$\begin{aligned} k \cdot \Delta \hat{f}_p &= k \cdot (f_{o,p} + \Delta f_p - \hat{f}_{o,p}) \\ &= k \cdot (\Delta f_p + \hat{\delta}) \end{aligned} \quad (24)$$

$$\hat{Y}_p = \sum_{k=1} \hat{M}_k \cdot a_k \quad (25)$$

$$\hat{M}_k = \hat{m} \cdot k \cdot (\Delta f_p + \hat{\delta}) \quad (26)$$

ここで、 $\alpha = \pm 1$ による \hat{Y}_p の変化から Δf_p を推定する。

$\alpha = \pm 1$ の各 $\hat{f}_{o,p}$ と $f_{o,p}$ の差を δ^+ 、 δ^- とし、 $|\hat{H}_{q,p}(z_p)|$ の零点付近の傾きを m^+ 、 m^- とすると、式 (25)、(26) より、出力振幅 \hat{Y}_p^+ 、 \hat{Y}_p^- は式 (27) となる。

$$\hat{Y}_p^\pm = \sum_{k=1} \hat{m}^\pm \cdot k \cdot (\Delta f_p + \hat{\delta}^\pm) \cdot a_k \quad (27)$$

この場合、 α が微小であるため $\delta^+ \cong -\delta^-$ (δ と表記する)、 $\hat{m}^+ \cong \hat{m}^- \cong m$ とみなすと、式 (28) が得られる。

$$\hat{Y}_p^\pm = \sum_{k=1} m \cdot k \cdot (\Delta f_p \pm \delta) \cdot a_k \quad (28)$$

ここで、 $\sum m \cdot k \cdot a_k = A$ とおくと、式 (28) は

$$\hat{Y}_p^\pm = A \cdot (\Delta f_p \pm \delta)$$

となり、

$$A = \frac{\hat{Y}_p^+ + \hat{Y}_p^-}{2 \cdot \delta}$$

である。また、式 (21)、(22) より、 $Y_p = A \cdot \Delta f_p$ となるため、基本周波数のずれ Δf_p は式 (29) で得られる。

$$\begin{aligned} \Delta f_p &= \frac{Y_p}{A} \\ &= \frac{2 \cdot \delta \cdot Y_p}{\hat{Y}_p^+ + \hat{Y}_p^-} \end{aligned} \quad (29)$$

但し、入力音に含まれる複数の音が周波数のずれを持つ場合、 Y_p 、 \hat{Y}_p^\pm には他音のずれ成分が影響を与える。このため、和音に含まれる各音に対応した $\hat{H}_{q,p}(z_p)$ の α を繰り返し更新し、全ての出力が収束した段階で各音の厳密な周波数を決定する。

4 スペクトル構造を利用した楽器の推定

楽器は、その発音機構および共鳴機構の違いにより、固有の倍音構造を持つことが知られている。本方式では、この性質を利用して、楽器 v : piano, violin, trumpet, ... etc. の周波数スペクトル $\{T_{v,k} : v$ の単音の k 次倍音 $\}$ をテンプレートとし、前述の抽出音 p_r のスペクトル $\{X_{p_r,k}\}$ と比較することで演奏楽器の推定を行なう。ここで、各 $\{T_{v,k}\}$ を作成する際に、各楽器の単音に対して STFT を行ない (図 7)、その平均をとることで、楽器

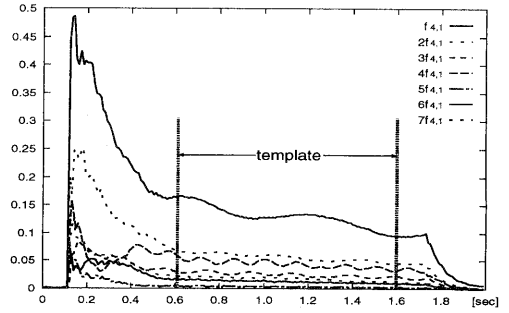


図 7 ピアノ音 C_4 の STFT 結果の例

音に含まれる非調和周波数成分や、ピブラート等による周波数変動の影響を回避する。

図 7 より、波形の立上りから 0.3[sec] 程度の間は、各倍音の振幅変化が安定しておらず、また、各波形がほぼ安定した後 (時間軸上の 0.6[sec] 以後) でも、各倍音成分の振幅に振動が生じていることが分かる。このため、STFT 結果の定常状態の部分 (図 7 の例では 0.6~1.6[sec] 程度) を切り出し、各倍音成分ごとに平均をとったものを、その楽器のテンプレート $\{T_{v,k}\}$ とする。また、上述の理由から、抽出音 p_r のスペクトル構造 $\{X_{p_r,k}\}$ についても同様な手法で計算する。なお、処理の簡単化のために、 $\{T_{v,k}\}$ 、 $\{X_{p_r,k}\}$ は基本周波数成分の振幅が 1.0 となるように、あらかじめ規格化する。

同一楽器音であれば、そのスペクトル構造は類似していると考えられるため、式 (30) に示す差分 S_v が最小となる v を検出することで、楽器推定が可能である。

$$S_v = \sum_{i=1}^{max} |X_{p_r,i} - T_{v,i}|, \quad max : \text{倍音数} \quad (30)$$

これまでの研究により、MIDI(KORG:05R/W) 音を用いた場合には、近接する音のスペクトル構造が類似している性質を利用し、採譜対象の 36 音に対して各楽器 9 種類程度のテンプレートにより楽器推定が可能であることを確認した [11]。しかし、実際の楽器音のスペクトルは音域、音量等の条件により変化するため、1 種類の楽器に対してでも、状況に応じた多数のテンプレートを用意する必要がある。

5 シミュレーション結果

5.1 シミュレーション条件

本方式の性能を明らかにする目的で、長年調律されていない piano 音、素人の演奏した管楽器 (clarinet, alto-sax, cornet, horn) 音を用いて計算機シミュレーションを

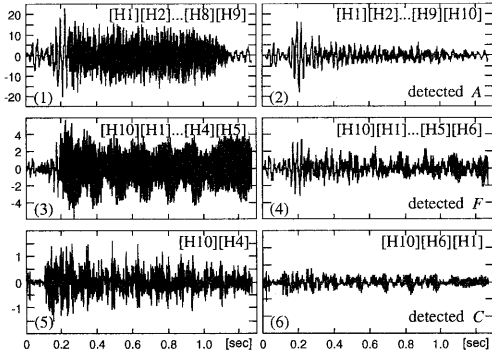


図8 異楽器和音 piano C_4 +clarinet F_4 +alto-sax A_4 の音名判別結果

行なった。ここで、使用したA/D変換器の特性上、オーバーサンプリング周波数 $f_s = 54054.054$ [Hz] とした。

なお、各くし形フィルタ $H_{q,p}(z_p)$ は、基準ピッチ $A_4 = 440.0$ [Hz] の平均律音階に基づいて構成した。また、あらかじめ各楽器の単音を録音し、必要な楽器音を計算機上で合成することで、所望の異楽器和音を作成した。

5.2 音名判別結果

piano C_4 , clarinet F_4 , alto-sax A_4 から成る和音に対して、縦続接続したくし形フィルタ $H_{q,p}(z_p)$ の出力振幅により、音名判別を行なった結果を図8に示す。ここで、図中の $[H_p]$ は固定零点のくし形フィルタ $H_{s,p}(z_p)$ を意味している。なお、この図では、(1) から (2), (3) から (4), (5) から (6) で順に、音名判別された前後段の $H_{q,p}(z_p)$ の出力波形を示し、各波形の右上部に、その出力を得るための $H_{s,p}(z_p)$ の並びを示した。

この結果より、実楽器音の持つ基本周波数のずれの影響から、各 $H_{q,p}(z_p)$ の出力は厳密には零にならないが、前段の出力と比較して十分な減衰が得られており、音名の判別が可能であることが確認できる。

5.3 厳密な基本周波数の推定結果

和音1: piano C_4 , alto-sax F_4 , clarinet A_4

和音2: cornet C_4 , horn E_4 , piano G_4

に対して、適応的くし形フィルタ $\hat{H}_{q,p}(z_p)$ により、ずれ周波数の推定を行なった結果を表1, 2に示す。ここで、 K は $\hat{H}_{q,p}(z_p)$ の可変遅延数 $\alpha = 0 \pm 1$ での周波数推定回数を1とした場合の、 α の総更新回数である。

この結果より、各楽器の単音をDFTした場合とほぼ等しい結果が得られており、適応的くし形フィルタ $\hat{H}_{q,p}(z_p)$ を用いて、基本周波数のずれを推定できることが確認で

表1 $\hat{H}_{q,p}(z_p)$ による和音1の周波数推定結果

音名	$f_{4,i}$	$q_p + \alpha$	K	$f'_{4,i}$	DFT 結果
C_4	261.62	413-1	2	262.40	262.92
F_4	349.23	310-6	3	355.62	356.54
A_4	440.0	246-1	2	441.26	440.89

表2 $\hat{H}_{q,p}(z_p)$ による和音2の周波数推定結果

音名	$f_{4,i}$	$q_p + \alpha$	K	$f'_{4,i}$	DFT 結果
C_4	261.62	413-3	3	263.68	263.45
E_4	329.63	328-1	2	330.61	330.95
G_4	392.00	276-2	2	394.56	394.63

きた。なお、推定される各周波数は、式(23)において $k=1$ とした場合に得られる離散周波数に依存する。

5.4 楽器推定結果

前述の和音1, 2に対して楽器推定を行なった結果を以下に示す。なお、実際の楽器音を用いる場合には、テンプレート作成が重要な課題となるが、今回のシミュレーションでは、録音された各楽器の単音からの合成音を解析対象としているため、それらの音を利用して各楽器音について1対1対応のテンプレートを用意した。

ここで、和音分離およびスペクトル抽出の例として、和音1. において抽出されたpiano音、およびpiano C_4 のテンプレートを図9に示す。得られた2つのスペクトル構造が非常に類似していることから、分離音 C_4 の雑

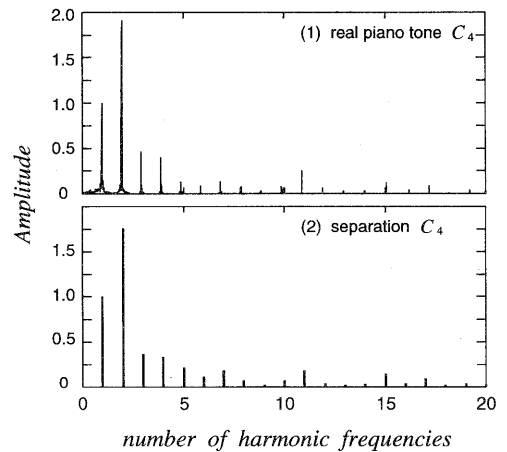


図9 piano C_4 と和音1. から抽出した C_4 のスペクトル構造

表3 和音1 piano C_4 ,clarinet E_4 ,A-sax G_4 のスペクトルの差分値

	piano	clarinet	alto-sax
C_4	0.001	0.241	0.183
F_4	0.314	0.028	0.188
A_4	0.236	0.119	0.005

表4 和音2 cornet C_4 ,horn E_4 ,piano G_4 のスペクトルの差分値

	cornet	horn	piano
C_4	0.002	0.100	0.221
E_4	0.147	0.024	0.129
G_4	0.246	0.189	0.003

音成分が少ないことが分かり、本方式により、精度の良い和音分離が可能であることがいえる。

また、和音1, 2に含まれる他の音に対しても同様にスペクトル抽出を行ない、各スペクトルの差分により楽器推定を行なった結果を表3, 4に示す。なお、ここでは、各スペクトルの10倍音までについて、式(30)に従い差分の平均値を計算した。この結果より、抽出音に対応した楽器との差分値(太字で記入)が最も小さくなっており、楽器の判別が可能であることが分かる。

6 むすび

本論文では、実楽器音を持つピッチのずれに対応させた適応的くし形フィルタ $\hat{H}_{q,p}(z_p)$ を用いた和音分離、楽器推定を行なうことで、以下の点を確認した。

(1) 入力楽器音がピッチのずれを持つ場合でも、そのずれ範囲が隣接する音との周波数間隔の1/3以内であれば、固定零点のくし形フィルタの出力の減衰を見ることで入力音名の判別が可能である。ただし、隣り合う音が音量に著しい違いを持つ場合は、誤判別の可能性がある。

(2) 入力和音に含まれる音の種類を判別した後に、可変遅延を持つ $\hat{H}_{q,p}(z_p)$ の出力振幅の変化をみることで、ずれピッチの検出が可能である。

(3) 厳密なピッチ推定を行なった後、入力周波数に対応したくし形フィルタを用いることで、精度良く和音を分離することが可能である。これにより、元音に忠実なスペクトル抽出が可能となり、ピッチにずれを持つpiano, clarinet, alto-sax, cornet, hornを用いた計算機シミュレーションにより、テンプレートとの差分を用いた楽器推定が可能であることを確認した。

今後の課題としては、以下の問題が挙げられる。

(1) 現段階では、計算機シミュレーションで使用した楽器や入力和音の数が少なく、十分に性能評価できたとは言えない。より多くの楽器音を用い、様々な状況を想定

して検証を行ない、本方式の性能を明らかにする必要がある。

(2) 今回の計算機シミュレーションでは、使用した異楽器和音が計算機上で合成されたものであるため、入力音と1対1対応のテンプレートを用いて楽器推定を行なうことができた。しかし、実際は、演奏によりスペクトル構造が変化するため、楽器推定のためのテンプレート作成法を検討する必要がある。

参考文献

- [1] M. Piszczalski and B. A. Galler, "Automatic music transcription," Computer Music Journal, vol.1, no.4, pp.24-31, (1977)
- [2] 原, 井口, "複素スペクトルを用いた周波数同定," 計測制御, Vol.19, No.9, pp.718-723 (1983)
- [3] 高沢, "計算機による採譜," 情処学論, Vol.29, No.6, pp.593-598 (1988)
- [4] W. J. Pielemeier and G. H. Wakefield, "A high-resolution time-frequency representation of musical instrument signals," J. Acoust. Soc. Am. vol.99, no.4, Pt.1, pp.2382-2396, 1996
- [5] 永田, 安部, 城戸, "多数センサによる音源波形の推定," 日本音響学会論文誌, Vol.47, No.4, pp.268-273 (1995)
- [6] 柏野, 木下, 中臺, 田中, "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, Vol.J79-D-II, No.11, pp.1762-1770 (1996)
- [7] 植田, 橋本, "音源分離のためのブラインドデコンポジションアルゴリズム," 情処学論, Vol.38, No.1, pp.146-156 (1997)
- [8] 小野, 斎藤, 小沢, "自動採譜のためのGAを用いた混合音推定," 計測制御, Vol.33, No.5, pp.417-423 (1997)
- [9] 三輪, 田所, 斎藤, "零出力に注目したくし形フィルタによる音階検出," 電学論, Vol.118-C, No.1, pp.57-64 (1998)
- [10] 三輪, 田所, 斎藤, "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, Vol.J81-D-II, No.9, pp.1965-1974 (1998)
- [11] 三輪, 田所, 斎藤, "くし形フィルタによる異楽器音中のピッチ推定と演奏楽器推定," 情処研報, Vol.98, No.96, pp.15-20 (1998)
- [12] 高沢, 西川, "基準ピッチ $A_4 = 440.0[\text{Hz}]$ をめぐる," 日本音響学会誌, Vol.52, No.5, pp.368-374 (1996)
- [13] J.R. ピアース 著, 村上陽一郎 訳, "音楽の科学," 日経サイエンス社