

音階分離フィルタを用いた楽音信号のピッチ検出法

鈴木 佑[†] 大竹孝平[‡]

[†], [‡] 法政大学工学部

〒184-8584 東京都小金井市梶野町 3-7-2

E-mail: [†] suzuki@otk.k.hosei.ac.jp, [‡] otake@k.hosei.ac.jp

あらまし 従来のくし形フィルタなどフィルタ形のピッチ検出法では、検出する音階の基本周波数および倍音周波数で零出力とする遮断形のくし形フィルタを用いているため、音階検出手順が煩雑で、検出精度や処理量の点で問題があると考えられる。本論文では、各音階の基本周波数および倍音周波数でシャープなバンドパス特性を持つ通過形のくし形フィルタである音階分離フィルタバンクを用いたピッチ検出法を提案している。音階分離フィルタバンクと音階検出特性を向上させる音階レベル検出器の構成法を明らかにし、また、計算機シミュレーションにより音階検出特性を示している。

キーワード 音階分離フィルタ, くし形フィルタ, ピッチ検出, 倍音

Pitch Detection of Musical Tones for Transcription Using Scale Separation Filters

Yuu SUZUKI[†] Kohei OTAKE[‡]

[†], [‡] Faculty of Engineering, Hosei University

3-7-2 Kajino-cho, Koganei-shi, Tokyo, 184-8584 Japan

E-mail: [†] suzuki@otk.k.hosei.ac.jp, [‡] otake@k.hosei.ac.jp

Abstract A filter type pitch detection method for transcription so far reported, uses band-stop type comb filters at the fundamental and harmonic frequencies of each musical scale. Therefore the pitch detection process becomes very complicated, resulting in the low detection accuracy and high computation cost. In order to solve the problem, in this paper we propose a pitch detection method using band-pass type comb filters, called scale separation filters, which have sharp band-pass characteristics only at the fundamental and harmonic frequencies. We show the construction methods of the scale separation filter bank and the scale level detector to enhance the pitch detection performance. We also show the pitch detection characteristics by computer simulation.

Keyword scale separation filter, comb filter, pitch detection, harmonic tone

1. まえがき

作曲や楽音を解析する上で自動採譜は極めて有用であり、その研究が盛んに行われている。しかし、自動採譜技術の中で、楽音の音階を検出するピッチ検出は特に重要であるが、技術的に困難な問題も多く、未だ完成された技術が確立されているとはいえない。

ピッチ検出法としてこれまでに、高速フーリエ変換(FFT)[1]、適応フィルタ[2]、拡張ノッチフーリエ変換(E-NFT)[4]や、くし形フィルタ[5]-[7]を用いた方法が提案されている。音階の周波数間隔が不均一なため、FFTを用いる方法よりもフィルタ形の方が有望と考えられる。しかし、従来のくし形フィルタなどフィルタ形のピッチ検出法では、検出する音階の基本周波数及び倍音周波数で零出力とする遮断形のくし形フィルタを用いているため、音階検出手順が煩雑になり、ピッチ検出精度や処理量の点で問題があると考えられる。

本論文では、この問題に対処するため、各音階の基本周波数及び倍音周波数でシャープなバンドパス特性を持たせた通過形のくし形フィルタを用いたピッチ検出法について述べている。このようなフィルタを音階分離フィルタと呼ぶ。平均率音階では、ある音階の倍音周波数が他の音階の基本周波数や倍音周波数に極めて近い値をとる場合がある。従って、音階分離フィルタの特性をいかにシャープにしても、それだけで音階を分離検出することは困難になる。このため本論文では、音階分離フィルタ出力とその基本周波数成分のレベルを用いた音階出力レベル検出器を設けて音階分離特性を向上させる構成法を提案している。本構成法について計算機シミュレーションを行い、その有効性を明らかにしている。

2. 単音のモデル化と音階分離フィルタ

楽音を構成する各単音は平均律音階で調律されているとする。今、音階 C, C#, D, D#, E, F, F#, G, G#, A, A#, B の 12 音階を、それぞれ、 $p=1, 2, \dots, 12$ で表す。また、オクターブを $q=0, 1, 2, \dots$ で表すと、各単音は (p, q) の 2 重添字で表示できる。第 (p, q) 単音の基本周波数は、次式で与えられ、その単音のピッチという。

$$f_{p,q} = 2^{(p-10)/12+q-4} f_{10,4} \quad (1)$$

ここで、 $f_{10,4} = A_4 = 440[\text{Hz}]$ である。

単音 (p, q) の信号波形 $x_{p,q}(t)$ は、楽器固有のインパルス応答 $\varphi_{p,q}(t)$ を持つ線形フィルタに、ピッチ周期が $T_{p,q} (=1/f_{p,q})$ で、時定数が $\tau_{p,q}$ ($T_{p,q}$ で規格化) のインパルス列を印加して生成されると仮定して次式でモデル化する。

$$x_{p,q}(t) = \varphi_{p,q}(t) * \sum_{n=0}^{\infty} e^{-nT_{p,q}/\tau_{p,q}} \delta(t - nT_{p,q}) \quad (2)$$

上式をフーリエ変換し、変形すると次式のような単音

(p, q) のスペクトルが得られる。

$$X_{p,q}(f) = \sum_{k=-\infty}^{\infty} a_k \delta(f - kf_{p,q}) \quad (3)$$

ただし、

$$a_k = \frac{\tau_{p,q} \Phi_{p,q}(kf_{p,q})}{1 + j2\pi k \tau_{p,q}} \quad (a_0 = 0) \quad (4)$$

$\Phi_{p,q}(f) : \varphi_{p,q}(t)$ のフーリエ変換

上式から、単音 (p, q) のスペクトルは、基本周波数 $f_{p,q}$ とその整数倍の倍音周波数 $nf_{p,q}$ ($n=2, 3, \dots$) 上に現れる複素振幅 a_n の線スペクトルの形で表わされることが分かる。 n に対する a_n の特性は、楽器の種類に依存し、その単音の倍音構造を規定する。

ここで、楽音は、一般に単音の線形結合で構成されているとしてモデル化して考え、楽音から単音 (p, q) の信号成分を抽出するフィルタを単音 (p, q) の音階分離フィルタと呼ぶ。それは上記の線スペクトルの周波数位置でシャープな帯域通過特性を有するフィルタにより実現できる。

楽音のピッチ検出は、検出対象オクターブの周波数範囲で、存在し得る各単音の音階分離フィルタを並べたフィルタバンクを設置し、一定の閾値レベルを越えた出力を生じるフィルタに対応した単音が存在すると判定することにより実現できる。

次に、音階分離フィルタの具体的構成法と問題点、並びに改善法について述べる。

3. 音階分離フィルタの構成

音階分離フィルタの構成法は種々考えられるが、ここでは、最も単純な 1 次音階分離フィルタと、2 次音階分離フィルタについて述べる。2 次フィルタについては、縦続構成と並列構成がある。また、音階分離フィルタ出力のレベルを検出するレベル検出器について述べる。

3.1 1 次音階分離フィルタ

図 1 に示すフィルタで、入力楽音 $x(n)$ の標本化周波数を $f_s (=1/T_s)$ とすると、伝達関数は

$$F_{p,q}(z) = \frac{1 - e^{-1/\tau}}{1 - e^{-1/\tau} z^{-L_{p,q}}} \quad (5)$$

振幅特性は次式で与えられる。

$$|F_{p,q}(e^{j\omega})| = 1 / \sqrt{1 + \gamma^2 \sin^2(\omega L_{p,q} / 2f_s)} \quad (6)$$

ただし、 $\gamma^2 = 4e^{-1/\tau} / (1 - e^{-1/\tau})^2$ である。

このフィルタは中心周波数が $f_k = kf_s / L_{p,q}$ ($k=0, 1, \dots$) で、帯域幅が $B = (2f_s / \pi L_{p,q}) \sin^{-1}(1/\gamma)$ なるくし形フィルタである。ここで、 $L_{p,q} \equiv f_s / f_{p,q}$ に選択すると近似的に単音 (p, q) を通過させるフィルタが実現できる。

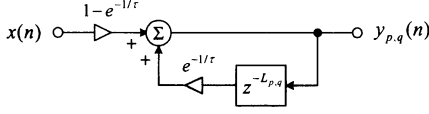


図 1 1次音階分離フィルタ

中心周波数を $kf_{p,q}$ により厳密に一致させる改良形の構成を図3に示す。このフィルタの伝達関数は次式で表され、

$$F_{p,q}(z) = \frac{1 - e^{-1/\tau}}{1 - e^{-1/\tau}[(1-\lambda)z^{-L_{p,q}} + \lambda z^{-(L_{p,q}+1)}]} \quad (7)$$

中心周波数は極めてよい精度で次式で与えられる。

$$f_k \equiv k[(1-\lambda)/L_{p,q} + \lambda/(L_{p,q}+1)]f_s \quad (8)$$

従って、パラメータ $\lambda \equiv (L_{p,q}+1)(1-L_{p,q}f_{p,q}/f_s)$ に選択すればよい。ただし、 $L_{p,q} = \lfloor f_s/f_{p,q} \rfloor$: ガウス記号。

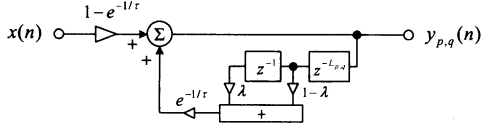


図 2 1次音階分離フィルタ（改良形）

3.2 2次音階分離フィルタ

角周波数 ω_0 を null 点とする2次のノッチフィルタは、零点が $e^{\pm j\omega_0}$ 、極が $\rho e^{\pm j\omega_0}$ ($\rho \equiv 0.99 \sim 0.999$) である次式の伝達関数で与えられる[8]。ここで、 ρ は1に近い程、急峻な特性になる。

$$H(z) = \frac{1 - \alpha z^{-1} + z^{-2}}{1 - \rho \alpha z^{-1} + \rho^2 z^{-2}} \quad (9)$$

ただし、 $\alpha = 2\cos(\omega_0)$ である。このノッチフィルタを用いて、音階分離フィルタは、以下の縦続形と並列形の2種の構成法が考えられる。

(A) 縦続形構成

単音 (p, q) を遮断するくし形フィルタは、その単音の N 倍音までを考慮すると、遮断周波数が、それぞれ $kf_{p,q}$ ($k=1, \dots, N$) のノッチフィルタの縦続接続で構成でき伝達関数は次式となる。

$$H_{p,q}(z) = \prod_{k=1}^N H_{k,p,q}(z) \quad (10)$$

ただし、

$$H_{k,p,q}(z) = \frac{1 - \alpha_{k,p,q} z^{-1} + z^{-2}}{1 - \rho \alpha_{k,p,q} z^{-1} + \rho^2 z^{-2}}$$

$$\alpha_{k,p,q} = 2\cos(2\pi kf_{p,q}/f_s)$$

このフィルタの周波数応答 $H_{p,q}(e^{j\omega})$ は、周波数 $kf_{p,q}$ ($k=1, \dots, N$) で0になり、 ρ が1に十分近いと、そのごく近傍でのみ0に近く、それ以外の周波数では1に近いので、単音 (p, q) の音階分離フィルタは

$$F_{p,q}(z) = 1 - H_{p,q}(z) \quad (11)$$

により実現できる。図3にその構成図を示す。

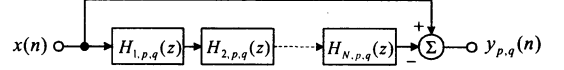


図 3 2次音階分離フィルタ（縦続形）

(B) 並列形構成

本構成法では、通過周波数が $kf_{p,q}$ の狭帯域フィルタを次式のようにして作る。

$$F_{k,p,q}(z) = 1 - H_{N,p,q}(z) = \frac{(1-\rho)(\alpha_{k,p,q} - (1+\rho)z^{-1})z^{-1}}{1 - \rho\alpha_{k,p,q}z^{-1} + \rho^2z^{-2}} \quad (12)$$

この $F_{k,p,q}(z)$ を $k=1, \dots, N$ まで N 個並列接続して

$$F_{p,q}(z) = \sum_{k=1}^N F_{k,p,q}(z) \quad (13)$$

なる単音 (p, q) の音階分離フィルタを実現する。図4にその構成図を示す。

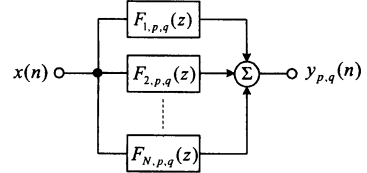


図 4 2次音階分離フィルタ（並列形）

3.3 レベル検出器

本提案の楽音のピッチ検出法では、以下のようにして検出を行う。

対象とするオクターブ範囲での音階分離フィルタ $F_{p,q}(z)$ ($p=1, \dots, 12$) から成るフィルタバンクを構成する。被検出楽音を入力 $x(n)$ に加え、各フィルタ出力 $y_{p,q}(n)$ のレベルを監視し、それがある閾値を越えた場合、そのフィルタに対応するオクターブの音階が存在すると判定する。

従って、この検出を行うには、 $y_{p,q}(n)$ を信号レベルに変換するレベル検出器が必要になる。ここでは、 $y_{p,q}(n)$ のレベル $w_{p,q}(n)$ を、処理量の少ない次式により検出を行う。

$$w_{p,q}(n) = (1-\lambda)|y_{p,q}(n)| + \lambda w_{p,q}(n-1) \quad (14)$$

この検出器は $y_{p,q}(n)$ の絶対値を低域フィルタに通してレベル $w_{p,q}(n)$ を検出するもので、パラメータ λ ($0 < \lambda < 1$) はその低域遮断周波数を決める。以下の実験ではレベルの脈動を少なくするため $\lambda = 0.999$ を用いた。

4. 音階分離フィルタバンクを用いたピッチ検出

4.1 フィルタバンクの特性

音階分離フィルタの周波数特性は、作成した各構成法で殆ど同様の特性であった。特に、並列形構成は、後述する基本周波数レベルを用いた処理が容易なために、以下では本構成による特性につき述べる。なお、楽音データの標本周波数は全て $f_s = 22.050$ [kHz] である。

まず、並列形で構成した第3オクターブ ($q=3$) の各音階分離フィルタの振幅特性 $|F_{p,3}(e^{j\omega})|$ を図5に示す。パラメータは $\rho = 0.9995$, $N=5$ とした。図には基本周波数範囲でのみ示しているが、実際には各音階フィルタとも第5倍音までの5つのピーク特性が生じる。

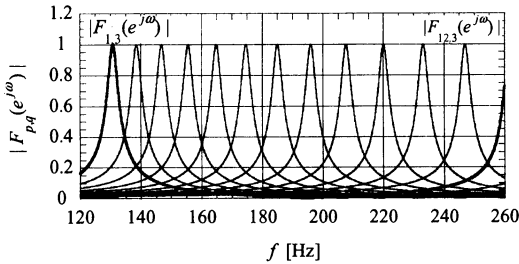
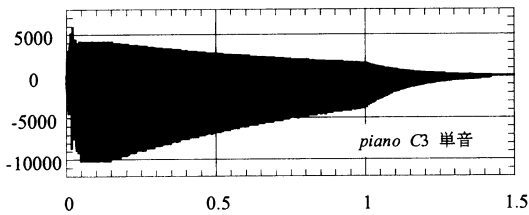


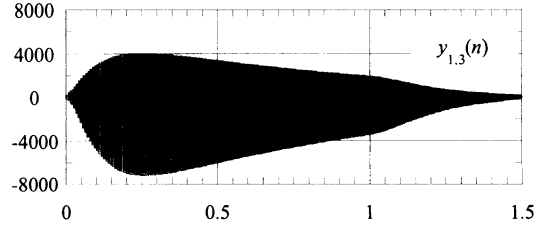
図5 音階分離フィルタの振幅特性 $|F_{p,3}(e^{j\omega})|$

次に、上と同じパラメータ条件で、入力 $x(n)$ として piano C3 ($p=1, q=3$) の単音を加えたとき(図6(a))、音階フィルタ $F_{1,3}(z)$ の出力応答波形 $y_{1,3}(n)$ を図6(b)に、また、このときの第3オクターブの各音階分離フィルタ出力レベル $w_{p,3}(n)$ ($p=1, \dots, 12$) を図6(c)に示す。

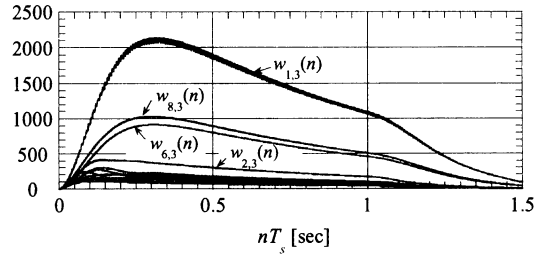
ρ を1に極めて近づけて急峻な帯域通過特性を持たせているので、図6(b)より $y_{1,3}(n)$ は鈍った応答波形になっていることが分かる。応答特性を改善するには、 ρ を小さく設定すればよいが、音階分離特性は劣化する。



(a) 入力信号 $x(n)$ piano C3 単音



(b) フィルタ $F_{1,3}(z)$ の出力応答波形



(c) 出力レベル検出特性

図6 音階分離フィルタの特性

また、図6(c)で、C3単音を入力しているので出力 $w_{1,3}(n)$ のレベルが最も高いのはよいが、それ以外の $w_{6,3}(n)$, $w_{8,3}(n)$ にもかなり大きな出力レベルが生じている。ピッチ検出では、ある閾値を越えるレベルが検出されたとき、その音階があると判定するため、この場合には誤判定となり大きな問題になる。

周知のように、平均率音階では、同一オクターブ内でも、2つの異なる音階のそれぞれのある倍音の周波数が極めて接近する場合がある[5]。これが誤判定を生む理由である。例えば、今の場合、音階(1,3)の3倍音と4倍音の周波数は、それぞれ、音階(8,3)の2倍音および音階(6,3)の3倍音の周波数と、約0.5 [Hz]程度の差しかない。このため、 $w_{6,3}(n)$, $w_{8,3}(n)$ にもかなりのレベルが検出されることになる。

次節では、この問題を解決する改良形音階レベル検出法について述べる。

4.2 改良形音階レベル検出法

異なる音階の倍音周波数での重なりを、音階分離フィルタの帯域通過特性の急峻化により分離識別しようとするのは応答特性の一層の劣化を招き不可能に近い。この対策として、音階(p, q)のフィルタ出力 $y_{p,q}(n)$ のレベル $w_{p,q}(n)$ だけではなく、 $y_{p,q}(n)$ に含まれる基本周波数 $f_{p,q}$ の成分のレベル $w_{1,p,q}(n)$ にも着目する。両レベルを検出し、それらがともにある閾値を越えていれば、音階(p, q)があると判定すればこの問題は解決できる。

これを実現するには、基本周波数成分のレベルを検出し易い図4の並列形構成が有利である。これを用いた音階分離フィルタと改良形レベル検出器の構成図を図7に示す。図で、フィルタ $F_{1,p,q}(z)$ の出力 $y_{1,p,q}(n)$ に対して式(14)と同様のレベル検出器を設け、その出力レベルを $w_{1,p,q}(n)$ とする。このレベル $w_{1,p,q}(n)$ と $y_{p,q}(n)$ の積 $y'_{p,q}(n)$ を求め、 $y'_{p,q}(n)$ に対してレベル検出を行い、最終的なレベル出力 $w'_{p,q}(n)$ を得る。このようにして得られたレベル出力は、 $y_{p,q}(n)$ に基本周波数成分が含まれなければ零に近くなり、所望のレベル出力を与えることになる。

本方法は、同一オクターブ内だけでなく、ある音階の倍音周波数が、上位のオクターブの他の音階の倍音周波数と重なる場合にも有効である。しかし、上位のオクターブのある音階の基本周波数と重なるときは検出が難しくなる。

例えば、C3とC4との複音を検出する場合である。この場合、C3の有無の判定は、明らかに上述の $w'_{1,3}(n)$ を用いて正しく行える。このとき、C3が存在しない場合には、C4の判定は $w'_{1,4}(n)$ により正しく行える。しかし、C3が存在する場合には、その倍音によりC4が存在しなくても、 $w'_{1,4}(n)$ にあるレベルの出力が現れるために、C4の存在の有無の正しい判定は難しい。この場合には、一般に、基本周波数成分は最も大きく、倍音成分は小さいとして、 $w'_{1,4}(n)$ が一定の閾値を越えればC4が存在するとして判定することになる。

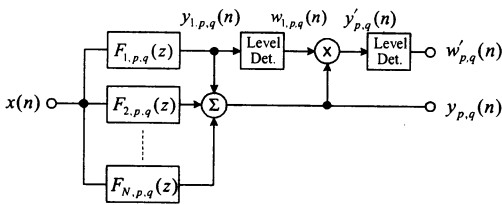


図7 音階分離フィルタと改良形レベル検出器

4.3 改良形レベル検出法の特性

上記の改善したレベル検出法を用いた音階分離フィルタの出力レベル特性につき検討する。パラメータ f_s, p, N は前と同じ値とし、入力も同じC3のpiano単音としたときの出力レベル $w'_{p,3}(n)$ ($p=1, \dots, 12$)の特性を図8に示す。図より明らかに、 $w'_{1,3}(n)$ 以外はレベルが殆ど検出されず、正しく音階(1,3)のみの存在を判定できることが分かる。

次に、入力をも同じ第3オクターブ内のC3, D3 piano複音としたときの出力レベル $w'_{p,3}(n)$ ($p=1, \dots, 12$)の特性を図9に示す。存在する音階(1,3), (3,3)に対応した出力レベル $w'_{1,3}(n)$, $w'_{3,3}(n)$ が大きく検出できていることが

分かる。 $w'_{3,3}(n)$ の方がレベルが高くなっているが、複音入力のD3音のレベルが元々高いためである。なお、各音階に隣接する音階の出力レベル $w'_{2,3}(n)$, $w'_{4,3}(n)$ も若干検出されている。これは音階分離フィルタの遮断域での減衰量が不足に漏れ込んだためである。しかし、この程度であれば正しい音階判定が得られると考えられる。

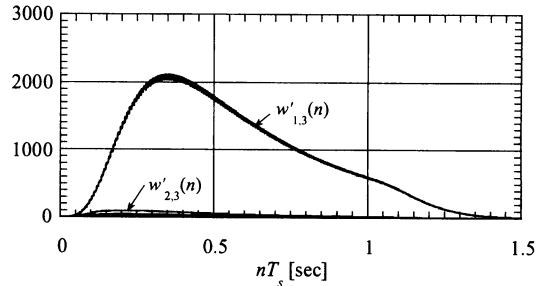


図8 C3 piano単音入力時の出力レベル検出特性

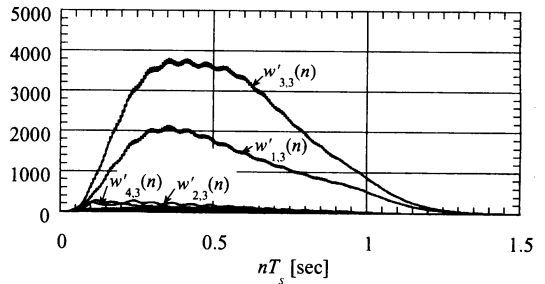


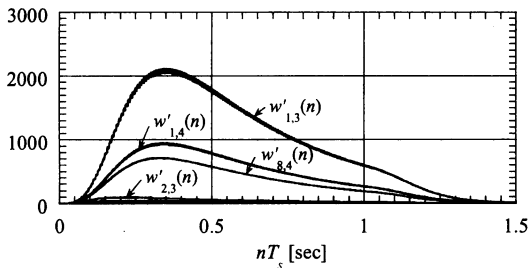
図9 C3, D3 piano複音入力時の出力レベル検出特性

最後に、第3と第4の2オクターブにわたるレベル検出結果について述べる。入力楽音としては、検出が最も難しい、(a)C3単音 (b)C4単音 (c)C3, C4複音の3つの場合について行った。これらの場合の出力レベル検出特性を、図10(a), (b), (c)にそれぞれ示す。

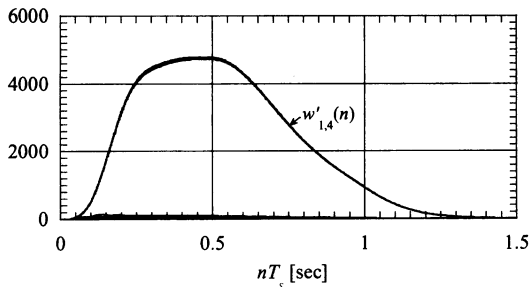
まず、(b)の場合は、 $w'_{1,4}(n)$ のみに大きなレベルが検出され、音階(1,4)のみの存在が正しく判定されることが分かる。次に、(a)の場合を考える。この場合、 $w'_{1,3}(n)$ に大きなレベルが検出されているので、単音(1,3)の存在は正しく判定できる。しかし、音階(1,3)の2倍音および3倍音の周波数が、それぞれ、音階(1,4)および音階(8,4)の基本周波数に、一致しない極めて近くなるため、 $w'_{1,4}(n)$ および $w'_{8,4}(n)$ にもはっきりとしたレベルが現れている。このため、音階(1,4)と(8,4)が存在するかどうかの判定が難しくなる。この場合は、 $w'_{1,3}(n)$ に比べて $w'_{1,4}(n)$, $w'_{8,4}(n)$ のレベルが、1/2程度以下と低いこと

から、倍音の影響と考えると、音階(1,4), (8,4)が存在しないと判定することになる。しかし、曖昧さが残るという問題がある。

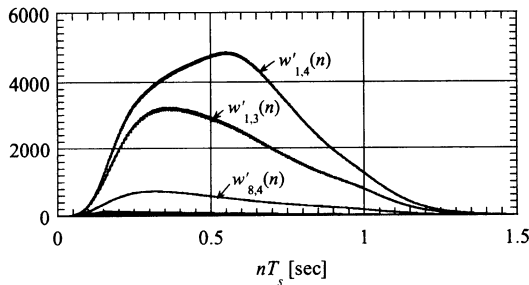
最後に、(c)の場合を考える。この場合、 $w'_{1,3}(n)$, $w'_{1,4}(n)$ に大きなレベルが現れており、特に、 $w'_{1,4}(n)$ のレベルが高いので、音階(3,1)と(1,4)がともに存在すると正しく判定できる。なお、上に述べた理由から $w'_{8,4}(n)$ にも若干のレベルが現れているが、 $w'_{1,3}(n)$ に比べて十分に低いことから音階(8,4)は存在しないと正しく判定できる。



(a) C3 単音入力



(b) C4 単音入力



(c) C3, C4 複音入力

図 10 第 3, 4 オクターブの出力レベル検出特性

5. むすび

従来、くし形フィルタなどフィルタ形のピッチ検出法では、検出する音階の基本周波数および倍音周波数で零出力とする遮断形のくし形フィルタを用いているため、音階検出手順が煩雑で、検出精度や処理量の点で問題があると考えられる。

本論文では、この問題を解決するため、各音階の基本周波数及び倍音周波数でシャープなバンドパス特性を持つ通過形のくし形フィルタである音階分離フィルタを用いたピッチ検出法を提案した。しかし、平均率音階では、ある音階の倍音周波数が他の音階の基本周波数や倍音周波数に極めて近い値をとる場合があり、単に音階分離フィルタ出力のレベル検出では誤検出する危険があることを指摘した。これに対処するため、音階分離フィルタ出力とその基本周波数成分のレベルを用いた改良形音階出力レベル検出法を提案した。本構成法について計算機シミュレーションを行い、音階検出特性が大幅に改善されることを示した。

しかし、ある音階の倍音周波数が他の音階の基本周波数に一致するというピッチ検出にとって最も難しい場合には、音階検出に若干の曖昧さが残ることを示した。この場合には、楽器の倍音構造も考慮したさらに高度な音階レベル検出法の検討が必要であり、今後の課題と考えている。

文 献

- [1] 井口征士昭, “音楽情報の処理—電子計算機を用いた自動採譜,” 計測制御, vol.19, no.3, pp.314-319, 1980.
- [2] 原裕一郎, 井口征士昭, “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] 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.
- [4] 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.32, no.12, pp.2005-2010, 1995.
- [5] 三輪多恵子, 田所嘉昭, 斎藤 努, “零出力に着目したくし形フィルタによる音階検出,” 電学論 C, vol.118-C, no.1, pp.57-64, 1998.
- [6] 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, 1965.
- [7] 松本 弥, 山口 満, 田所嘉昭, “適応くし形フィルタの縦続接続による採譜のための音高検出,” 信学技報, DSP2001-148, pp.23-30, (2002-01)
- [8] A. Nehorai, B. Porat, “Adaptive Comb Filtering for Harmonic Signal Enhancement,” IEEE Trans. ASSP, vol.ASSP-34, no.5, pp.1124-1138, Oct. 1986.