

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

瀬戸康裕, 夏井雅典, 田所嘉昭
豊橋技術科学大学 情報工学系

あらまし ある音の基本周波数成分が他音の高調波成分と重なる和音（音高推定困難和音）において，実楽器音の場合，異なる音の成分同士の間になぜか周波数差が存在する．このことに注目して，一音をくし形フィルタで消去し，残った成分の周期を自己相関関数より求めることで，重なった各音の音高を推定する音高推定法を提案する．提案法によって，実楽器音の音高推定困難和音（2和音）の音高を推定できることを確認した．

Pitch estimation method for difficult consonance using comb filters and autocorrelation functions

Yasuhiro Seto, Masanori Natsui, Yoshiaki Tadokoro
Dept. of Information and Computer Sciences, Toyohashi University of Technology

Abstract This paper presents a pitch estimation method for difficult consonance whose frequency components overlap. We notice the small frequency difference of overlapping components for real musical sounds. Using a comb filter, we eliminate one component and measure the period of the other one by calculating an autocorrelation function. We confirmed that the proposed method can estimate each pitch of difficult consonance composed of two tones.

1. まえがき

マイクなどの入力装置，あるいはCD再生などによって入力された音響信号を楽譜に直すシステム（採譜システム）として，これまでいくつかのシステムが報告されている^{1)~5)}．また，いくつかの楽器メーカーによって商品化されている．しかし，商品化されているシステムのほとんどは独唱や単独の楽器によって演奏される単音を対象にしたものであり，複数の音源によって構成された和音までを扱うことが可能な本格的なものは，いまだ実用化されていない．この大きな理由の一つとして，2和音を例に述べると，ユニゾン（基本周波数の等しい複数音によって形成される和音，例： $A3+A3$ ），オクターブ（基本周波数の比率が1:2の関係にある和音，例： $A3+A4$ ）や3倍音（基本周波数の比率が1:3の関係にある和音，例： $A3+E5$ ）といったある音の基本周波数成分が他音の高調波成分と重なる和音（図1参照）の処理に関する技術が未だ確立されていないことが挙げられる．本稿ではこのような和

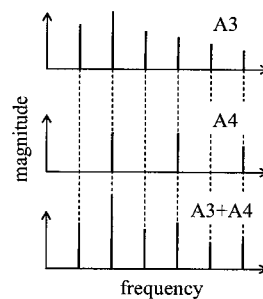


図 1. 音高推定困難和音になるスペクトルの例

音のことを音高推定困難和音と呼ぶ．

音高推定困難和音は，一般的な音楽演奏における最も基本的な和音として多く用いられるが，離散フーリエ変換（DFT: Discrete Fourier Transform）による周波数解析に代表される振幅スペクトルに注目する従来の手法⁶⁾では，和音において生じる周波数成分の重複を短時間で精度良く分離することが困難である．これは音高の認識精度の低下，あるいは分離された音の品質の

低下へとつながり、高精度な採譜システムを実現する上で大きな妨げとなる。したがって、和音を含む楽音の音高を精度良く処理することが可能な高性能な採譜システムを構築するためには、音高推定困難和音の音高を精度よく推定する音高推定手法が必須となる。

和音の採譜に関しては多くの研究が行われており、本研究で対象にしている音高推定困難和音に関しては、遺伝的アルゴリズム (GA: Genetic Algorithm) による方法⁷⁾、スペクトルの和を仮定した方法⁸⁾、音楽規則等に従った推定法²⁾などが知られている。GA による方法は前もって各楽器の時間波形を記憶しておく必要があり、また、他の二つの方法についても、位相の関係でスペクトルの和は必ずしも成立しないこと、また従来の音楽規則に従わない現代音楽などには対応できないなどの問題がある。

本稿では、時間領域から音高推定困難和音を推定できる可能性を検討し、音高推定困難和音 (2 和音) の重なる 2 成分にわずかな周波数差があることに注目した方法を提案する。すなわち、1 音をくし形フィルタで除去し、残りの信号の自己相関を測定してその周期よりその音の音高を推定する。そして推定された音高を除去するくし形フィルタを使用して、もう一方の音高を推定する方法である。

実楽器音の解析対象として、RWC 研究用音楽データベース⁹⁾のうち、トランペット、ホルン、アルトサクソ、クラリネット、ピアノ、エレキギターの 6 楽器の単音についてオクターブ 3 から 5 (サンプリング周波数 $f_s=44.1$ [kHz]) と、それらを計算機上で加算して生成した 2 和音を使用する。また、和音の最低音高については我々が既に提案しているくし形フィルタを利用した音高推定システム⁴⁾で既知とする。

2. くし形フィルタと楽音処理

本研究では楽音の処理にくし形フィルタを用いる。ここでは、その具体的な方法について述べる。くし形フィルタの入出力特性は、以下の線形差分方程式で表される。

$$y_p(n) = x(n) - x(n - N_p) \quad (1)$$

また、伝達関数は以下の式で表される。

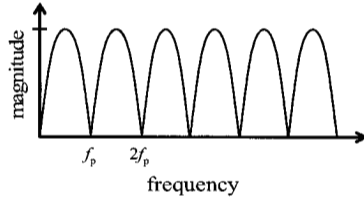


図 2. くし形フィルタ $H_p(z) = 1 - z^{-N_p}$ の振幅特性

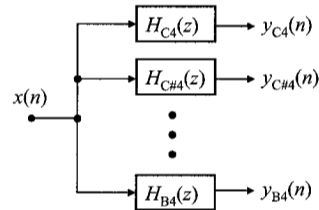


図 3. 並列構成くし形フィルタによる音高推定

$$H_p(z) = 1 - z^{-N_p} \quad (2)$$

この振幅特性を図 2 に示す。くし形フィルタを用いた音高推定では、くし形フィルタのもつ振幅特性と楽音の調波構造を利用する。図 3 のように、各音高 p に対応させたノッチ周波数をもつくし形フィルタ $H_p(z)$ を、1 オクターブに相当する 12 個並列接続し、それらの出力 $y_p(n)$ が最小となるくし形フィルタから音名を推定し、その後、違うオクターブへくし形フィルタの次数を変えてオクターブを推定して音高を決定する。このとき、フィルタの次数 N_p は $N_p = [f_s / f_p]$ で与えられる。

以上に示したように、くし形フィルタは、フィルタの次数 N_p に対応する周波数 f_p と、その定数倍の周波数を有する信号成分を同時に消去する。そのため、和音の他の音の基本周波数がある音の定数倍の関係にある和音、すなわちユニゾン (例: C4+C4), オクターブ (C4+C5), 3 倍音 (C4+G5) といった和音 (協和音) を音高推定の対象とした場合、くし形フィルタは複数の音高を同時に消去してしまうため、音高の分離推定が困難となる。

しかし、実際の音楽信号においては、二つの音楽信号が形成する調波構造に、楽器の構造や信号の発生原理に起因する非定常性に基づくわずかなずれが存在する。したがって、実楽器音による協和音をくし形フィルタに入力した場合、たとえ入力信号の最低音高がく

し形フィルタの対象とする音高と一致していたとしても、もう一方の信号がくし形フィルタによって完全に消去されず、微弱な振幅を有する出力として検出される。

本稿では、このくし形フィルタの出力信号に残留する信号成分を検出することで、もう一方の音楽信号の音高を推定する手法を検討する。自動採譜という応用を考える上で、解析アルゴリズムの時間分解能は十分高いものである必要がある。一方、本稿の推定対象である協和音が有する周波数のずれは極めて小さいものであるため、同時に高い周波数分解能も実現する必要がある。以上の理由により、本稿では、時間領域における信号の自己相関を周波数推定に用いる。

3. くし形フィルタと相関関数による音高推定困難和音の処理

2章で述べた理由により、本研究では、くし形フィルタの出力信号の周期を測定するために自己相関関数を用いる。ここでは、くし形フィルタと自己相関関数を組み合わせた音高推定困難和音の具体的な処理法について述べる。

提案する音高推定法の手順を以下に示す。なお、くし形フィルタを用いることにより、協和音に含まれる最も低い音高の基本周期 $T_0 = 1 / f_0$ はあらかじめ既知であるとする。また、初期条件として $m=1, i=1$ とする。

- 1) 入力信号 $x_0(n)$ の自己相関関数 $R_{x_0x_0}(k)$ を求め、基本周波数 f_0 に対応するインデックス $N_0 = [f_s / f_0]$ 付近におけるピークを探索し、そのインデックスを P_1 とおく。
- 2) $R_{x_0x_0}(P_1) \ll R_{x_0x_0}(0)$ のとき、入力された和音は協和音ではないとし、処理を終了する。そうでない場合、3)へ進む。
- 3) $x_0(n)$ をフィルタ次数 P_1 のくし形フィルタに通し、その出力を $x_1(n)$ とする。
- 4) $x_1(n)$ の自己相関関数 $R_{x_1x_1}(k)$ を求め、インデックス N_0 付近におけるピークを探索し、そのインデックスを P_{i+1} とおく。
- 5) $P_i = P_{i+1}$ 、もしくは $R_{x_1x_1}(P_{i+1}) \ll R_{x_1x_1}(0)$ のとき、入力信号には基本周波数 $m \times (f_s / P_i)$ の音が含まれると推定し、処理を終了する。そうでない場合、

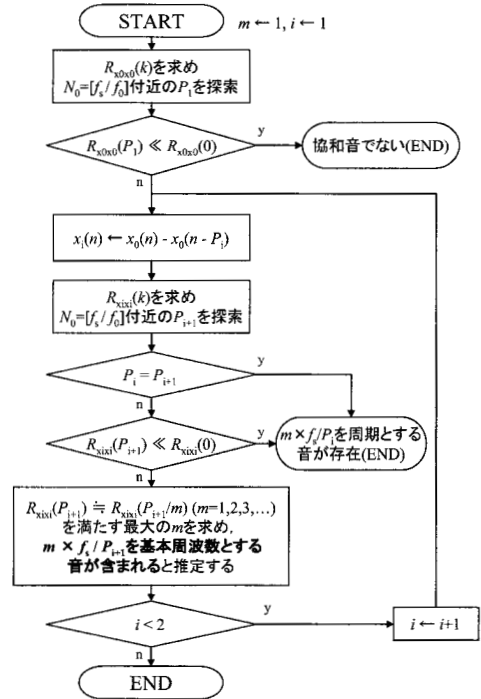


図 4. 自己相関を用いた周期推定の流れ

6)へ進む。

- 6) $R_{x_1x_1}(P_{i+1}) \doteq R_{x_1x_1}(P_{i+1} / m)$ ($m=1,2,3,\dots$) を満たす最大の m を求め、入力信号には基本周波数 $m \times (f_s / P_{i+1})$ の音が含まれると推定する。
- 7) $i < 2$ なら i をインクリメントし、3) に戻る。そうでない場合、処理を終了する。

次数が異なるくし形フィルタを通した後の残余信号に含まれる周期の差は微小であり、高速フーリエ変換でこの変化を検出するためには、長時間のデータが必要となる。一方、自己相関を用いた方法では、1回の推定に必要なデータ長は、和音の最低音高の2周期分でよい。例えば、200Hzと402Hzの和音においては、高速フーリエ変換を用いた方法では2Hzの周波数分解能を得るために500ms程度のデータ長が必要となる。一方、自己相関を用いた方法では、200Hzの2周期分のデータ長、10ms程度のデータでよい。すなわち自己相関を用いることで、必要な周波数分解能を維持したまま、高い時間分解能を得ることができる。

また、時間領域の信号をそのまま処理することと比べると、信号が周期を持っているかどうかの判断が容

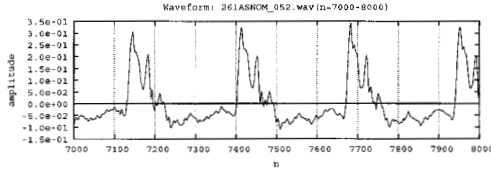
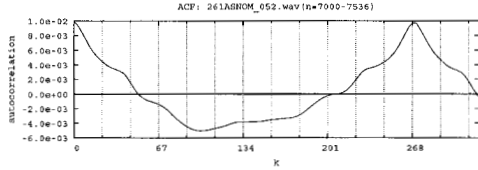
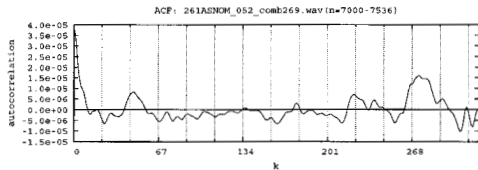


図 5. alto-sax E3 の時間波形



(a) くし形フィルタ通過前



(b) くし形フィルタ通過後 ($N_{P1}=269$)

図 6. alto-sax E3 の自己相関関数

易であることや、周期の計測が、相関関数のピークをみるだけで可能であるという利点がある。

4. 実楽器音による測定結果

実楽器音を用いて作成した(i)単音, (ii)ユニゾン, (iii)オクターブおよび(iv)3倍音それぞれに対し, 3章で示した手法を適用したときの具体的な推定の流れを以下に示す. なお, 自己相関関数を求める際には, あらかじめくし形フィルタによって推定されている和音の最低音高の2周期分のデータを使用している.

4.1. 単音

図 5 に入力が alto-sax E3 の単音である場合の時間波形, 図 6 にその音高推定過程に用いる自己相関関数を示す. 図 6 (a) のインデックス k において, $N_0 = [f_s / f_0] = [44100 / 164.81] = 268$ 付近におけるピークは $k = 269$ にある. この k を P_1 とすると, $R_{x_0x_0}(P_1) \ll R_{x_0x_0}(0)$ は成立しない. よって, この信号を $N_{P1} = 269$ のくし形フィルタに通し, その自己相関関数を計算すると, 図 6 (b) のような結果が得られる. $k = 275$ にピークが存在

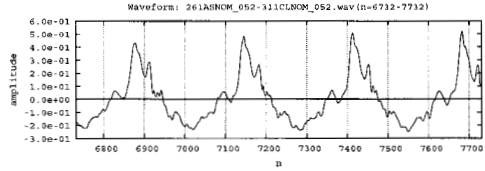
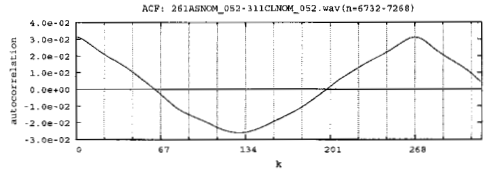
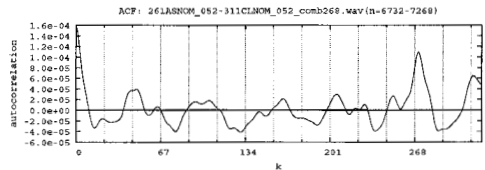


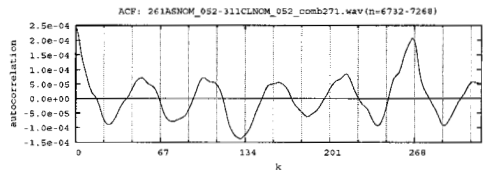
図 7. alto-sax E3 + clarinet E3 の時間波形



(a) くし形フィルタ通過前



(b) くし形フィルタ通過後 ($N_{P1}=268$)



(c) くし形フィルタ通過後 ($N_{P2}=271$)

図 8. alto-sax E3 + clarinet E3 の自己相関関数

するが, $R_{x_{1x_1}}(275) \ll R_{x_{1x_1}}(0)$ なので, くし形フィルタで楽音がすべて消去されたとし, 基本周波数 $f_s / P_1 = 163.9\text{Hz}$ の単音と推定できる.

4.2. 和音

ここでは, E3 を最低音高とする和音についての測定結果を示す.

図 7 に alto-sax E3 + clarinet E3 (ユニゾン) の時間波形, 図 8 にその自己相関関数を示す. 図 8 (a) のインデックス k において, $N_0 = 268$ 付近におけるピークは $k = 268$ にある. この k を P_1 とすると, $R_{x_0x_0}(P_1) \ll R_{x_0x_0}(0)$ は成立しない. よって, この信号を $N_{P1} = 268$ のくし形フィルタに通し, その自己相関関数を計算すると, 図 8 (b) のような結果が得られる. 図 8 (b) において, $k = 268$ 付近におけるピークは $k = 271$ にある. この k を P_2 とすると, $P_1 = P_2$ と $R_{x_{1x_1}}(P_2) \ll R_{x_{1x_1}}(0)$

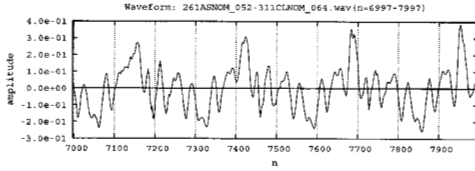


図 9. alto-sax E3 + clarinet E4 の時間波形

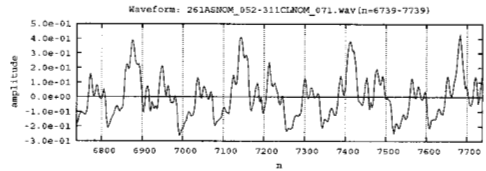
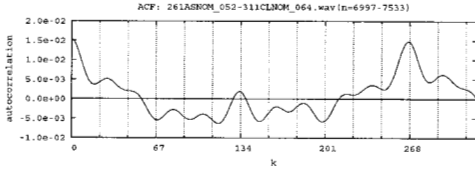
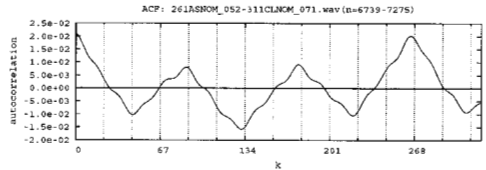


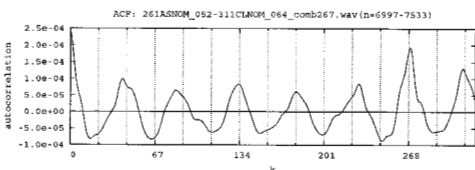
図 11. alto-sax E3 + clarinet B4 の時間波形



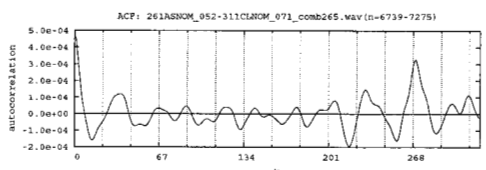
(a) くし形フィルタ通過前



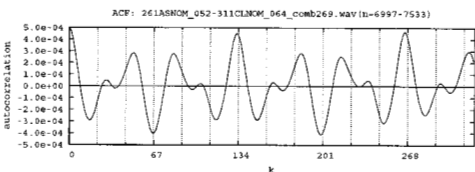
(a) くし形フィルタ通過前



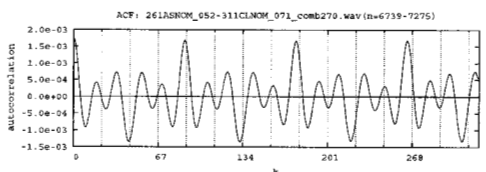
(b) くし形フィルタ通過後 ($N_{p1}=267$)



(b) くし形フィルタ通過後 ($N_{p1}=265$)



(c) くし形フィルタ通過後 ($N_{p2}=269$)



(c) くし形フィルタ通過後 ($N_{p2}=270$)

図 10. alto-sax E3 + clarinet E4 の自己相関関数

図 12. alto-sax E3 + clarinet B4 の自己相関関数

は成立しないので、 $R_{x_1x_1}(P_2) \approx R_{x_1x_1}(P_2/m)$ を満たす最大の m を求める。その結果 $m=1$ が得られ、ここでまず、基本周波数 $f_s/P_2=162.7\text{Hz}$ を周期とする音が存在すると推定できる。

次に、ここで推定した楽音を消去するために、 $N_{p2}=271$ のくし形フィルタに通すと、図 8 (c) のような自己相関関数が得られる。これは、完全に消去されていないもう一方の楽音の自己相関関数に近いものだと考えられる。図 8(c) において、 $N_0=268$ 付近におけるピークは $k=267$ にある。この k を P_3 とすると、 $P_2=P_3$ と $R_{x_2x_2}(P_3) \ll R_{x_2x_2}(0)$ は成立しないので、 $R_{x_2x_2}(P_3) \approx R_{x_2x_2}(P_3/m)$ を満たす最大の m を求める。その結果 $m=1$ が得られ、基本周波数 $f_s/P_3=165.2\text{Hz}$ を周期とする音が存在すると推定できる。結果として、 162.7Hz 、 165.2Hz を基本周波数とする二つの音が存在すると推

定できる。

図 9 に alto-sax E3 + clarinet E4 (オクターブ) の時間波形、図 10 にその自己相関関数を示す。ユニゾンの場合と比べると、図 10 (c) のように、一方の楽音が消去された場合に、最低音高の周期付近に相当する自己相関 ($k=266$) と、最低音高の 2 倍音の周期付近に相当する自己相関 ($k=266/2$) がほぼ等しくなっている。すなわち $R_{x_2x_2}(P_3) \approx R_{x_2x_2}(P_3/m)$ ($m=2$) が成立する。結果として、 163.9Hz 、 331.6Hz の二つの音が存在すると推定できる。

図 11 に alto-sax E3 + clarinet B4 (3 倍音) の時間波形、図 12 にその自己相関関数を示す。この場合も同様の手法で、 163.3Hz 、 501.1Hz の二つの音が存在すると推定できる。

5. まとめ

音高推定困難和音（2 和音）に対し、くし形フィルタと自己相関関数を用いた音高推定法を提案し、実際に、2 成分にわずかな周波数差がある楽器音同士の音高推定困難和音の音高推定が可能であることを確認した。

提案法は、基本周期を自己相関関数によって測定することによって、くし形フィルタ通過後の信号に含まれている成分の変化をみている。このことから、基本周波数に相当する成分が弱い音色では、本推定法では誤推定されることがある。また、それぞれの和音を構成する調和成分以外の成分が強いと、自己相関関数による周期推定が困難になるという問題があり、システムのロバスト性についても検討する必要がある。このため、今後さらに多数の音高推定困難和音に対する提案法の効果を検討していく。また、3 和音以上の音高推定困難和音についても検討していく。

なお、本研究の一部は、文部科学省 21 世紀 COE プログラム「インテリジェントヒューマンセンシング」の援助により行われた。

参考文献

- 1) 片寄晴弘, 井口征士: 知的採譜システム, 人工知能学会誌, Vol.5, No.1, pp.59-66, (1988).
- 2) 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, 電子情報通信学会論文誌, Vol.J79-D-II, No.11, pp.1762-1770, (Nov. 1996).
- 3) 山口 満: くし形フィルタに基づく歌唱音の自動採譜に関する研究, 豊橋技術科学大学博士学位論文, pp.66-72, (Jan. 2005).
- 4) 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, 電子情報通信学会論文誌, Vol.J87-D-II, No.12, pp.2271-2279, (2004).
- 5) 後藤真孝: 実世界の音楽音響信号を対象としたメロディーとベースの音高推定, 情報処理学会 音楽情報科学研究会 研究報告, 99-MUS-31-16, Vol.99, No.68, (Aug. 1999).
- 6) Curtis Roads: コンピュータ音楽 歴史・テクノロジー・アート, 東京電機大学出版局, p.421, (Jan. 2001).
- 7) 小野徹太郎, 斎藤英雄, 小沢慎治: 自動採譜のための GA を用いた混合音推定法, 計測自動制御学会論文集 Vol.33, No.5, pp.417-423, (1997).
- 8) 植田 護, 橋本周司: 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, 情報処理学会論文誌, Vol.38, No.1, pp.146-157, (1997).
- 9) 後藤真孝, 橋口博樹, 西村拓一: RWC 研究用音楽データベース: 音楽ジャンルデータベースと楽器音データベース, 情報処理学会 音楽情報科学研究会 研究報告, 2002-MUS-45, pp.19-26, (2002).