

高調波成分の畳み込みによる多重音解析法の有効性の検証 Verification of the Effectiveness of Multiple Sound Analysis by the Convolution of the Harmonic Overtone Component

岩元 建悟†
Kengo Iwamoto

安達 季†
Minoru Adachi

大田健紘†
Kenko Ota

1. はじめに

近年、単一音からの基本周波数抽出・推定精度は高精度に行えるようになってきている。従来から用いられている基本周波数推定手法として、ゼロ交差法、自己相関法、ケプストラム法などがあるがこれらの手法は、多重音を対象としたものではないため、多重音においては精度や安定性が悪く、正確に基本周波数を抽出することは困難である。

そこで、楽器や音声等の多重音解析に用いる手法として、Specmurt 法[1]が提案された。この手法は、対数周波数軸上において解析対象の音の調波構造は共通と仮定し、その調波構造を解析対象の音の振幅スペクトルに逆畳み込みすることにより基本周波数を抽出する。しかし、すべての音の調波構造は共通と仮定するため、時間的に調波構造の変動が大きい音に適用することは困難とされている。

そこで、調波構造を一致させる必要のない手法として Specmurt 法を元にした多重音のピッチ抽出法[2]が提案された。この手法は、対数周波数軸上において高調波成分を強調したスペクトルを分析対象の多重音の振幅スペクトルから減算することで基本周波数分布を得る手法である。これにより、共通調波構造を分析対象の音の調波構造と一致させることなく、様々な多重音から基本周波数を抽出することが可能となるが、その有効性に関して十分な検討がなされていなかった。

本稿では、多重音のピッチ抽出法について、合成音、実際の楽器音、歌声など様々な音を用いて有効性及び適用範囲について明らかにする。

2. Specmurt 法概要

調和性の単音の振幅スペクトルの場合、基本周波数と第 2 倍音、第 3 倍音、…、第 n 倍音の周波数関係は整数倍である。線形周波数軸の場合、基本周波数が変化すると倍音間の間隔も変化してしまうが、対数周波数(log スケール)軸に変換することで、基本周波数が変化しても倍音間の間隔は常に一定となる。そのため、線形周波数軸を対数周波数軸に変換することで、基本周波数の異なる音のスペクトルを周波数軸方向への平行移動のみで表現可能となる。

次に単音の性質として、すべての音の調波構造は共通と仮定する。この共通と仮定した調波構造を、基本周波数に対応するスペクトルが原点にくるように平行移動したものを共通調波構造 $h(x)$ と呼ぶ。異なる基本周波数の単音が重なった多重音の場合、これらの基本周波数の対数位置にそれぞれの共通調波構造 $h(x)$ を配置するとその和に

より多重音のスペクトルを表せる。

基本周波数分布を $u(x)$ とした時、共通調波構造が $h(x)$ であるなら、多重音の振幅スペクトル $v(x)$ はこれらを畳み込み、

$$v(x)=h(x)*u(x) \quad (1)$$

となる。

ここで共通調波構造 $h(x)$ が既知ならば、基本周波数分布 $u(x)$ は式(1)を逆畳み込みすることにより求めることができる。

$$u(x)=h^{-1}(x)*v(x) \quad (2)$$

これにより基本周波数分布 $u(x)$ を求めることができる。この逆畳み込みの操作は、 $u(x)$, $v(x)$, $h(x)$ の逆フーリエ変換を $U(y)$, $V(y)$, $H(y)$ とすると、

$$U(y)=V(y)/H(y) \quad (3)$$

となる。この y 領域からフーリエ変換をすることで $u(x)$ を求めることが可能である。

3. 提案法概要

Specmurt 法では、共通調波構造 $h(x)$ を分析対象の音の調波構造と一致させる必要があるため、調波構造が不安定な音に用いることは非常に困難である。この課題を解決するために、共通調波構造を学習データから推定する手法[3]も提案されているが、性能は学習データに依存する。

そこで、本稿では逆畳み込みによる手法以外で基本周波数の抽出を行い、調波構造を解析対象の音のものと一致させる必要のない手法を提案する。この手法は、対数周波数軸上において高調波成分を強調した振幅スペクトルを作成し、それを多重音の振幅スペクトルから減算することで基本周波数成分を抽出している。

共通調波構造に存在する原点上の基本周波数成分を 0 とし、高調波成分だけの高調波構造 $h'(x)$ を使用する。ただし、この高調波構造 $h'(x)$ は分析対象の音の高調波と一致している必要はない。図 1 に高調波構造 $h'(x)$ を示す。この高調波構造 $h'(x)$ を多重音の振幅スペクトル $v(x)$ に畳み込みを行うと、

$$v'(x)=h'(x)*v(x) \quad (4)$$

が得られる。 $v'(x)$ が高調波成分を強調した振幅スペクトルに当たる高調波強調スペクトル $v'(x)$ となる。式(4)の関係

† 日本工業大学, Nippon Institute of Technology

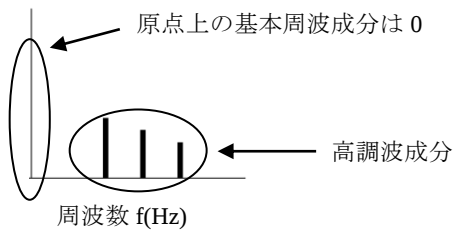
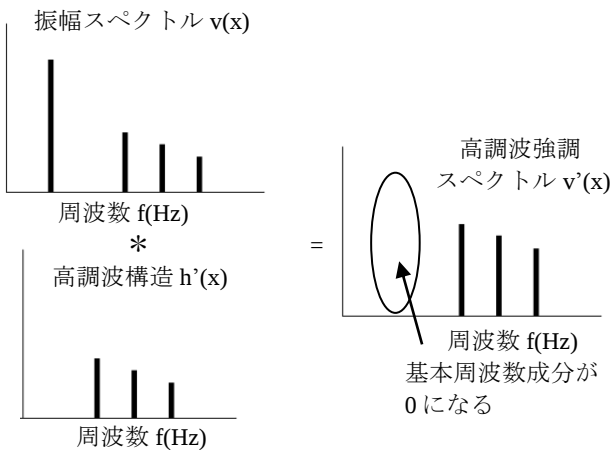
図1 高調波構造 $h'(x)$ 

図2 高調波構造の畳み込み

を図2に示す。ただし、説明は簡単のため $v(x)$ を単音としている。

畳み込みによって得た高調波強調スペクトル $v'(x)$ を多重音の振幅スペクトル $v(x)$ から減算すると、

$$u(x) = v(x) - v'(x) \quad (5)$$

となる。この時、基本周波数分布 $u(x)$ において 0 以下になる成分は 0 として処理する。これにより基本周波数成分の抽出が可能となる。式(5)の関係を図3に示す。

Specmurt 法では調波構造を一致させなければならなかったものを、提案法では分析対象の音の調波構造の高調波成分が、使用する高調波構造の高調波成分よりも弱い、もしくは一致する場合において基本周波数推定が可能となる。

4. 評価実験

4.1 評価方法

提案法により基本周波数成分のみを抽出できているかを確認するため、基本周波数成分を信号、その他の周波数成分を雑音とした SN 比(Signal to Noise Ratio)を求めた。SN 比は以下のように定義した。

$$S/N = 10 \log_{10} \frac{\sum_t \sum_f s(t, f)^2}{\sum_t \sum_f (s(t, f) - s'(t, f))^2} [\text{dB}] \quad (6)$$

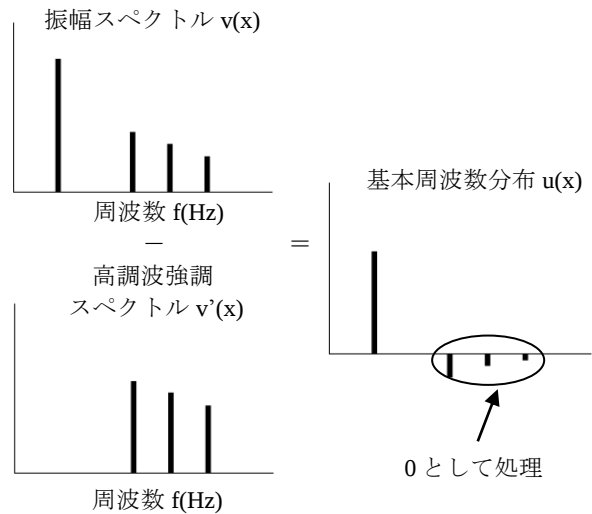


図3 高調波強調スペクトルの減算

表1 Specmurt 法及び提案法の SN 比[dB]

	倍音成分	Specmurt 法	提案法
振幅 スペク トル	基本～10 倍音	27.232	∞
	基本～9 倍音	15.144	∞
	基本～8 倍音	16.122	∞
	基本～7 倍音	16.779	∞
	基本～6 倍音	16.850	∞
	基本～5 倍音	14.971	∞
	基本～4 倍音	11.340	∞
	基本～3 倍音	36.370	∞
	基本～2 倍音	-0.166	∞
	基本周波数	-3.346	∞

ただし、 $s(t, f)$ は理想的な基本周波数分布であり、 $s'(t, f)$ は処理の結果得られた基本周波数分布である。SN 比を計算する際には、最大値を一定値(255)に正規化した。理想的な基本周波数分布は、基本周波数成分として 255 を代入し、それ以外の周波数成分には 0 を代入することで作成した。

評価では、共通調波構造が解析対象の音の調波構造と一致していない状況を想定して行った。Specmurt 法の共通調波構造は基本周波数の振幅スペクトルの大きさを 1、2 倍音から 10 倍音までの大きさを 10 とした。提案法の高調波構造は基本周波数の振幅スペクトルの大きさを 0 とし、2 倍音から 10 倍音までの大きさを 10 とした。

4.2 提案法の評価

合成音、実際の楽器音及び歌声を用いて Specmurt 法及び提案法による基本周波数抽出性能の比較を行った。

4.2.1 合成音による評価

ここではまず単音に対する性能の比較を行う。合成音の作成条件はサンプリング周波数 44.1kHz、量子化ビット数 16 ビット、チャンネルはモノラル、データ長は 10s とした。440Hz を基本周波数とし、基本周波数成分のみの音（純音）、基本周波数に 2～10 倍音までを足し合わせた複合正弦波の単音を 10 種類作成した。

表1は評価に用いた倍音成分数の異なる 10 種類の単音に対する Specmurt 法及び提案法の SN 比の結果である。表

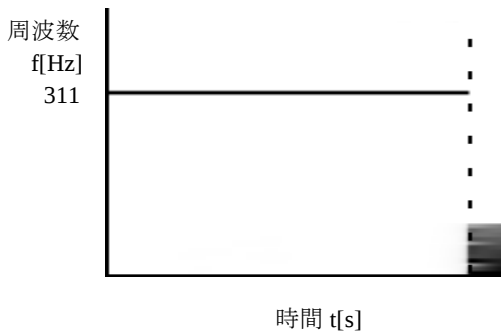


図 4 トランペット(D#3)の基本周波数分布

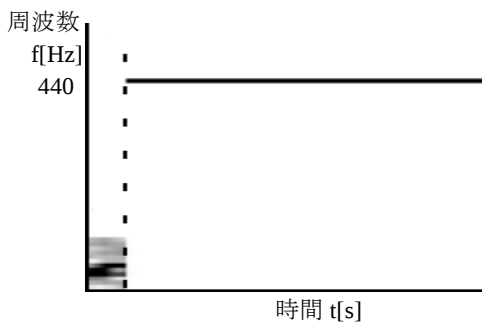


図 5 フルート(C4)の基本周波数分布

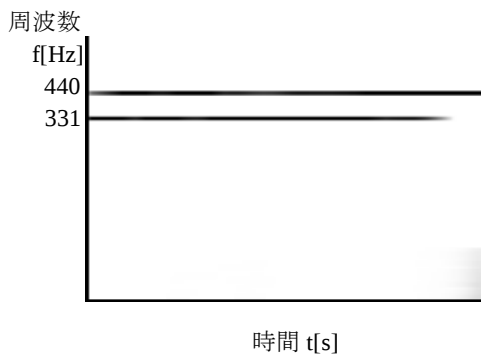


図 6 トランペットとフルートの多重音の基本周波数分布

からわかるように、Specmurt 法に比べ提案法のほうが SN 比は高く、提案法の SN 比は ∞ [dB]のため、理想的な基本周波数分布と一致している。

4.3 楽器音による評価

合成音だけでなく実際に収録した異なる楽器音が混合した多重音からも基本周波数の抽出が可能であるか、フルートとトランペットを用いて評価実験を行った。

用いた楽器音は、フルートはラ音(440Hz)、トランペットはレ#音(311Hz)の音をそれぞれ個別に収録し、計算機上で多重音とした。

図 4 にトランペットの基本周波数分布、図 5 にフルートの基本周波数分布を示す。まず、単音ではそれぞれの基本周波数成分がほぼ正確に抽出されていることが確認できた。図 6 はトランペット音とフルート音の多重音からの基本周波数成分の抽出結果である。どちらの音も基本周波数が抽出できていることが確認できる。

4.4 歌声による評価

歌声でも基本周波数抽出が可能であるかの評価を行った。曲は「カエルの合唱」の冒頭 10 秒の音声を収録した。収録条件はサンプリング周波数 44.1kHz、量子化ビット数 16 ビット、チャンネルはモノラル、データ長は 10s とした。3 人の音声(低音、中音、高音)は個別に収録し、後に計算機上で加算し多重音とした。3 人の歌声のテンポをそろえるためにメトロノームを用いて収録を行った。

図 7 は基本周波数分布を示す。図 7(a)は低音の場合の基本周波数分布であり、基本周波数は 130Hz 前後である。図 7(b)は中音の場合の基本周波数分布であり、基本周波数は 261Hz 前後である。図 7(c)は高音の場合の基本周波数分布であり、基本周波数は 465Hz 前後である。この結果から、メロディーのある歌声からでも基本周波数抽出が可能であることが確認された。図 7(d)は個別に収録した音声を計算機上で加算して多重音とした場合の基本周波数分布である。この結果より、低音と中音の歌声は基本周波数がオクターブ違いの音であるため、130Hz の倍音にあたる 261Hz の成分は減算され抽出できないことがわかる。しかし、倍音関係にある基本周波数が抽出できないのは他の多重音解析手法にもある課題である。

5. 検討

前節までで、様々な調波構造を持つ音に対しても単一の高調波構造を用いて基本周波数成分が抽出できることを確認した。ここでは提案法において、どの程度高調波構造の設定に自由度があるかを高調波構造に含まれる倍音数を変化させて明らかにする。

用いるデータは前節で用いた合成音と同じく、440Hz を基本周波数とし、基本周波数+2 倍音、基本周波数+2 倍音+3 倍音のように、2 倍音から 10 倍音までをそれぞれ足し合わせた複合正弦波を 9 種類作成した。

高調波構造も同じく 2 倍音まで含むものから 10 倍音まで含むものまでの 9 種類を作成し、81 通りの組み合わせについて SN 比を用いて評価した。

表 2 は高調波構造と解析対象の音の振幅スペクトルを変化させたときの基本周波数成分抽出結果に対する SN 比を表している。7 倍音までを足し合わせた複合正弦波から 10 倍音まで足し合わせた複合正弦波は、高調波構造に含まれる倍音数を変化させても結果は同じであった。表 2 より、分析対象の音の振幅スペクトルの倍音数よりも高調波構造に含まれる倍音数を多くすれば基本周波数成分の抽出が可能であると考えられる。また、高調波構造に含まれる倍音数が少ないと基本周波数以外の周波数成分が抽出されることが確認できた。

6. 結論

本稿では、高調波成分の量み込みによる基本周波数分析法について Specmurt 法と比較し、提案法による基本周波数成分抽出の有効性及び適用範囲を明らかにした。合成音などの多重音による評価の結果、Specmurt 法では他の周波数も抽出してしまうが、提案法では高調波構造の倍音数を振幅スペクトルの倍音数よりも多く設定すれば基本周波数以外の成分が抽出されないことが確認できた。Specmurt 法の場合、解析対象の音の調波構造と共通調波構造を一致させる必要があるが、一方で提案法の場合、

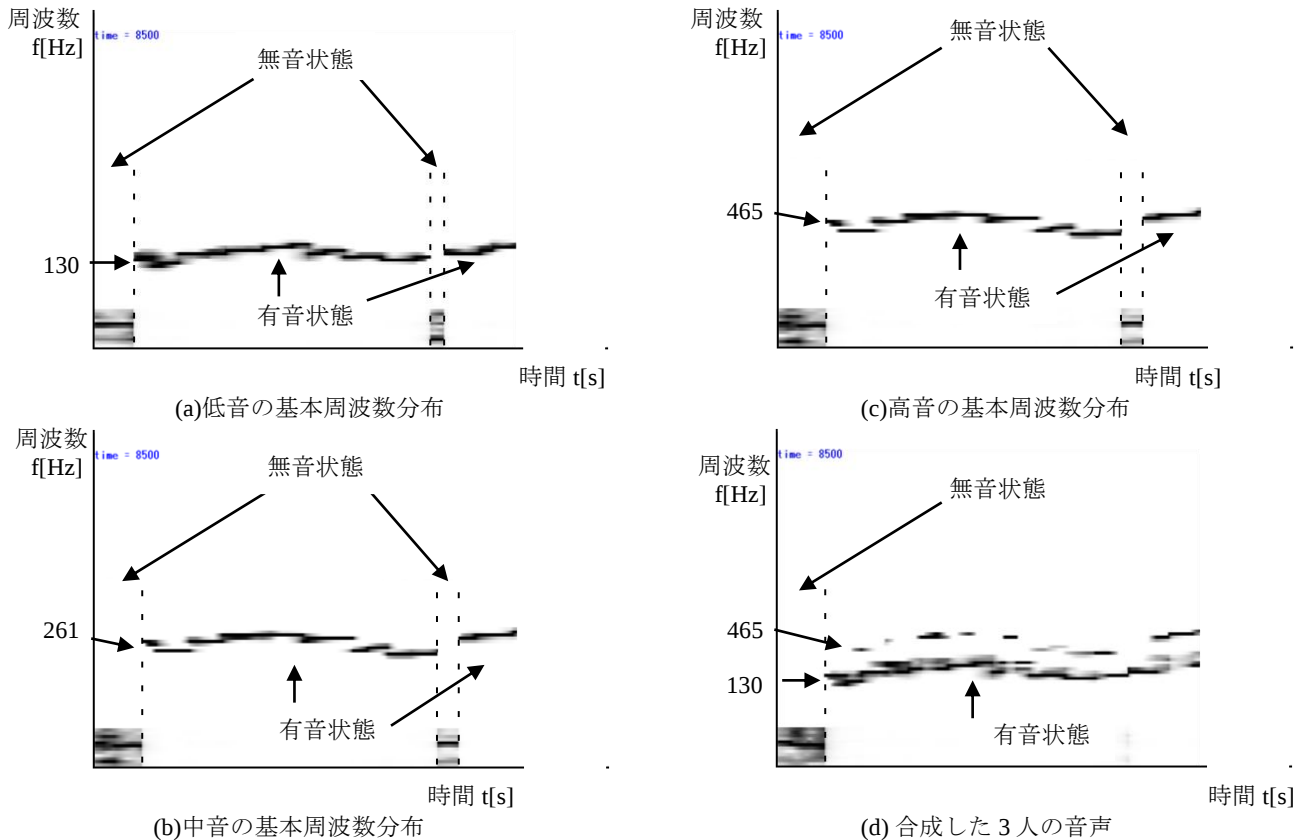


図7 歌声からの基本周波数抽出結果

表2 高調波構造と倍音の関係

単位：dB		高調波構造					
		10～7倍音	6倍音	5倍音	4倍音	3倍音	2倍音
振幅スペクトル	基本～10倍音	∞	1.126	1.087	-2.193	-2.193	-5.310
	基本～9倍音	∞	1.087	1.087	-2.193	-2.193	-5.310
	基本～8倍音	∞	1.126	1.087	-2.193	-2.193	-4.167
	基本～7倍音	∞	1.126	1.087	-2.193	-2.193	-4.167
	基本～6倍音	∞	∞	∞	0.563	0.563	-2.629
	基本～5倍音	∞	∞	∞	0.563	0.563	-2.629
	基本～4倍音	∞	∞	∞	∞	∞	0.207
	基本～3倍音	∞	∞	∞	∞	∞	0.207
	基本～2倍音	∞	∞	∞	∞	∞	∞

共通調波構造を高調波構造に変えることで調波構造を一致させる必要がなくなり、Specmurt法よりも柔軟にかつ正確な基本周波数分析が可能であることが確認できた。そのため、収録した楽器音や歌声にも提案法を使用することで基本周波数の抽出が可能であることも確認された。今後の課題として、ギターなどのコード推定に応用することを考える、倍音の関係にある音の基本周波数成分が抽出できないことは問題となるため、倍音の関係にある音の基本周波数成分の抽出方法を検討する。また、音楽理論等を適用することで、不完全な抽出結果からでも、コード推定が可能となる仕組みを検討していく。

参考文献

- [1] 高橋佳吾，西村卓也，嵯峨山茂樹，“対数周波数逆畳み込みによる多重音の基本周波数解析”，情報処理学会研究報告，2003-MUS-53，pp.61-66，Dec. 2003
- [2] 岡部俊一，塚原悠太，大田健紘，青木正和，“高調波成分の畳み込みによる多重音に対する基本周波数推定に関する研究”，日本音響学会 2014 年春季研究発表会，pp. 435-436，3月，2014.
- [3] 亀岡弘和，齊藤翔一郎，西本卓也，嵯峨山茂樹，“Specmurtにおける準最適共通調波構造パターンの反復推定による多声音楽信号の可視化とMIDI変換”，情報処理学会研究報告，Vol. 2004-MUS-56，pp. 41-48.