

# 基本周波数が欠落した調波信号の基本周波数推定法の検討

城下 修一 苜木 禎史 宇佐川 毅 江端 正直  
(熊本大・工)

## 1. はじめに

音声処理システムにおいて、ピッチ情報を用いることはかなり有効な手段である。これまでにピッチを検出に関する多くのアルゴリズムが、提案されてきた。また近年、伝送路は広帯域化されているが、音声情報がすべて相手に伝送されているわけではない。例えば、電話回線においてピッチ周波数の存在する帯域の信号成分は減衰している。GHA、FFTなどの解析法では減衰してしまったピッチ周波数を直接検出するのは難しい。そこで筆者らは、調波信号の解析を効率的に行う調波 Wavelet 変換を提案してきた [1][2]。

本論文では、調波 Wavelet 変換の特性を決める Signal Specific Analyzing Wavelet に調波型 Analyzing Wavelet を導入する。そして、その Analyzing Wavelet に観測信号の振幅情報を与えて、伝送路によって帯域制限された信号のピッチ周波数の推定を行う。またその他の解析法と比べて、シミュレーションにおいて有効性を検討する。

## 2. システム

まず調波 Wavelet 変換の特性を決める Signal Specific Analyzing Wavelet を次のように定義した。

$$g(t, f, x) = \sqrt{f_0} \cdot \left( \sum_{k=1}^n \alpha_k e^{j(\omega_k(T-t) + \phi_k)} \right) \cdot w(T) \quad (1)$$

$$-\frac{N}{2f_0} \leq T \leq \frac{N}{2f_0}$$

ここで、 $\omega$  は角周波数であり、 $n$  は高次成分数、 $N$  は分析フレーム長に相当するパラメータであり、基本周期  $1/f_0$  に関する周期数である。また  $\omega_k = 2\pi f_k$  であり、窓関数  $w(T)$  は式(3)で定義されている区間  $T$  内で窓処理を施し、それ以外は0である。 $\phi_k, \alpha_k, f_k$  はそれぞれ  $k$  次調波成分での位相、振幅、周波数である。

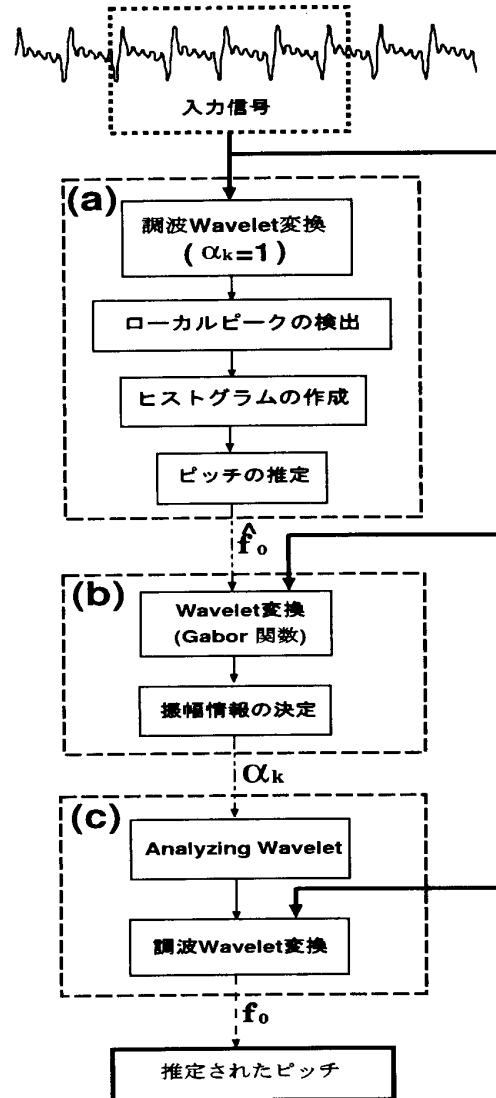


図 1. 処理の流れ

提案するシステムは大きく分けて3つのブロックから構成されている。図1にブロック図を示す。ブロック (a) では、大まかなピッチ  $\hat{f}_0$  の推定を行い、ブロック (b) で振幅情報  $\alpha_k$  を決定し、ブロック (c) でピッチ  $f_0$  の決定を行う。

$\hat{f}_0$  の推定は、フォルマント情報がわかる帯域幅で観測信号に対して調波 Wavelet 変換を行う。このときの振幅係数はすべて  $\alpha_k = 1$  である。そこで計算したパワースペクトルからピークを求めるために移動平均をとり、ヒストグラムを作成する。ヒストグラムの作成は、まずピッチ周波数が存在すると思われる  $500\text{Hz}$  以下とその周波

\* Consideration for Method of Virtual Pitch Detection

By S.Shiroshita, Y.Chisaki, T.Usagawa and M.Ebata(Kumamoto Univ)

数よりも高い周波数で2つに分割する。それぞれの周波数帯域において移動平均によって見つかるピークの数をも5個以内、つまり合計10個以内までしぼる。次に500Hz以下のピークについて、 $k$ 次高調波の周波数 $f_k$ に対して $0.9f_k$ から $1.1f_k$ の範囲でマッピングを行いヒストグラムを作成する。そのヒストグラムから最頻度のピークを $\hat{f}_0$ とする。次に振幅情報を決定するブロック(b)では、ブロック(a)で推定したピッチ周波数 $\hat{f}_0$ の $k$ 次調波成分においてそれぞれWavelet変換を行う。ここでのAnalyzing Waveletは、 $k$ 次調波成分でのパワを求めるので調波型ではなく、Gabor関数を用いた。これから求めた全ての調波成分のパワを振幅情報 $\alpha_k$ とした。

最後にブロック(c)では、ブロック(b)から求めた振幅情報 $\alpha_k$ をAnalyzing Waveletに与え、ブロック(a)で求めた $\hat{f}_0$ について

$$0.9\hat{f}_0 \leq f_0 \leq 1.1\hat{f}_0$$

の範囲で0.1Hzステップで分析を行い、ピッチ周波数 $f_0$ を推定する。

### 3. シミュレーション

#### 3.1 シミュレーション条件

ここでは、提案する手法を用いて基本周波数が欠落した信号から、基本周波数を推定できることを確認し、雑音が付加された場合に推定精度を評価する。また比較対象アルゴリズムとして変形相関法を用いる。

ここで使用する観測信号として疑似音声<sup>[3]</sup>を用いる。それは女声/a/からフォルマント特性のフィルタを作成し、基本周期50ms(200Hz)間隔のパルス列にフィルタ処理を施した。またその帯域は300Hzから3kHzである。雑音は白色雑音を観測信号と同様の帯域で帯域制限したものを用いた。サンプリング周波数、量子化精度は10kHz、16bitsである。すべてのシミュレーションで基本周波数は200Hzである。

Analyzing Waveletのフレーム長は解析する基本周波数の3周期とし、窓関数はガウス窓を用いた。また位相はどの調波成分においても $-\pi$ から $\pi$ の間で正規分布に従う乱数で与えた。解析する信号のフレーム長は25.6msでシフト量は1msである。ブロック(a)で $\hat{f}_0$ を求めるための、周波数の解析の範囲は50Hzから $\frac{1}{2}$ 1.6kHzである。

まず図2はAnalyzing WaveletにGabor関数を用いて帯域制限された信号を解析した結果である。図から判断できる様に、200Hz付近にピークはみつかるが最大のパワが500Hz以上にあるので、ピッチを検出することできない。つまり単峰性の通過型フィルタでは、欠落したピッチ

チを推定できないことがわかる。

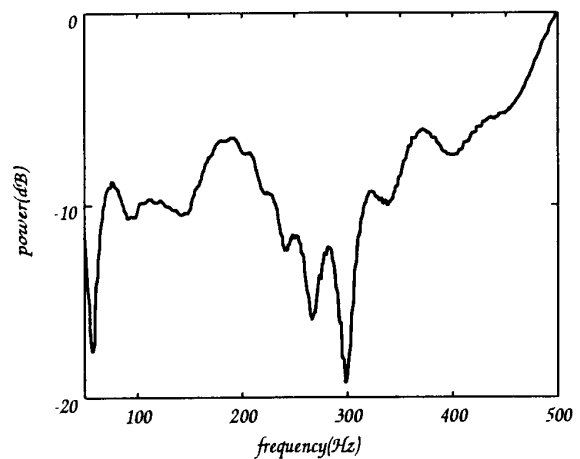


図2. 帯域制限信号のGabor関数でWavelet変換 (SNR = ∞)

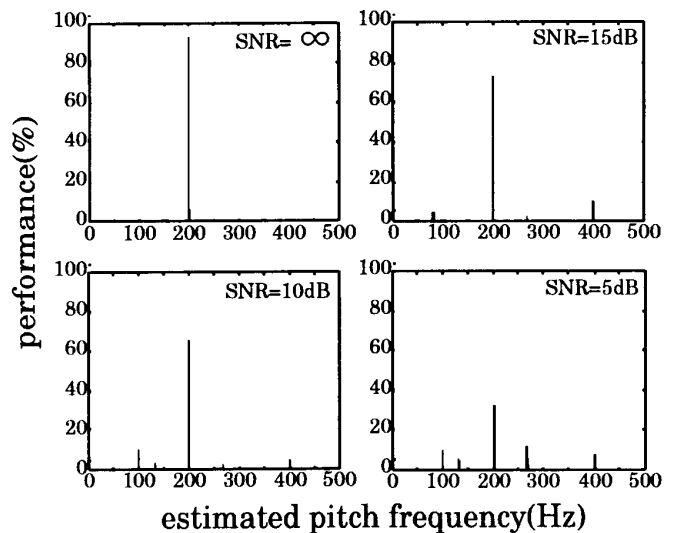


図3. 各SNRでのピッチ周波数推定のヒストグラム

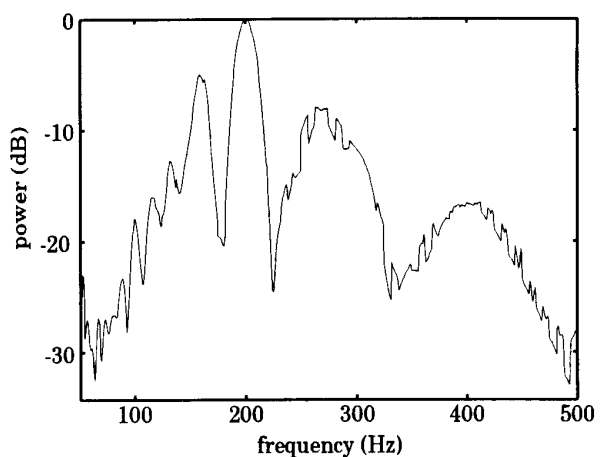


図4. 帯域制限信号の調波Wavelet変換 (SNR = ∞)

### 3.2 $\hat{f}_0$ に対する性能評価

次に雑音を付加した場合の推定精度について検討する。図3は  $SNR$  が  $5dB, 10dB, 15dB, \infty$  の場合において、推定した周波数の頻度を示している。この図で横軸の  $\Delta f$  は  $1Hz$  とした。この結果は、 $SNR$  が低くなるにつれてピッチの推定精度が悪くなっているが、悪くなる原因として  $\hat{f}_0$  の時点で倍ピッチ、半ピッチなどを推定してしまうことが原因だと考えられる。

### 3.3 振幅情報に対する性能評価

まず図4は、振幅係数  $\alpha_k$  に観測信号の情報を与えて、帯域制限信号を  $50Hz$  から  $500Hz$  を解析した結果である。この図からわかるように、基本周波数  $200Hz$  付近に最も大きなピークがあり、基本周波数を検出するのは容易であることがわかる。

次の図5では帯域制限をしていない信号、つまり基本周波数が存在している信号について、観測信号の情報を与えて解析した結果である。図4と比べて、ほぼ同じ波形が見られる。つまりこのシステムでは振幅情報を Analyzing Wavelet に与えることによって、観測信号が帯域制限に関係なく、基本周波数の推定ができた。

図6では、振幅係数がすべて一定の場合の解析結果である。基本周波数  $200Hz$  付近にピークがみられるが、最大パワを持つピークは倍ピッチ付近に現れている。これから正確なピッチの推定は困難である。

図7は  $SNR = 10dB$  のとき、解析した結果である。これは当然観測信号の情報を与えての結果である。つまり雑音環境化においても、振幅情報が正しければ、基本周波数の推定が行えるということを示している。前のシミュレーションにおいて雑音を付加した場合の推定精度について検討したが、このときは  $\hat{f}_0$  が倍ピッチなどの間違った推定を行っていた。つまり振幅情報が正しいものでない為、最終的に間違った推定を行ってしまった。よってこのシステムでは、振幅情報が正しいものであれば、雑音環境化においても精度よくピッチを求めるができる。

### 3.4 他のアルゴリズムとの比較

図8は  $SNR = 15dB$  のときの結果である。破線が変形相関法を用いて推定したピッチ、実線が調波 Wavelet 変換を用いて推定したピッチである。またこのときの変形相関法の分析フレーム長は Wavelet 変換と同じ  $25.6ms$  である。変形相関法では  $0ms$  から  $7ms$  と  $10ms$  から  $13ms$  の間では基本周波数である  $200Hz$  付近を推定しているが、その他のフレームでは倍ピッチである  $400Hz$  付近を推定している。しかし、調波 Wavelet 変換ではすべてフレームにおいて  $200Hz$  付近を推

定した。変形相関法においては  $SNR = 10$  においても  $SNR = 15dB$  と同様に倍ピッチの推定がみられた。

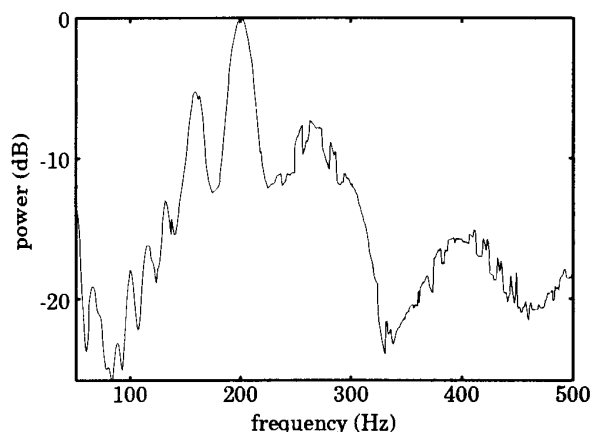


図5. 帯域制限されていない信号の調波 Wavelet 変換 ( $SNR = \infty$ )

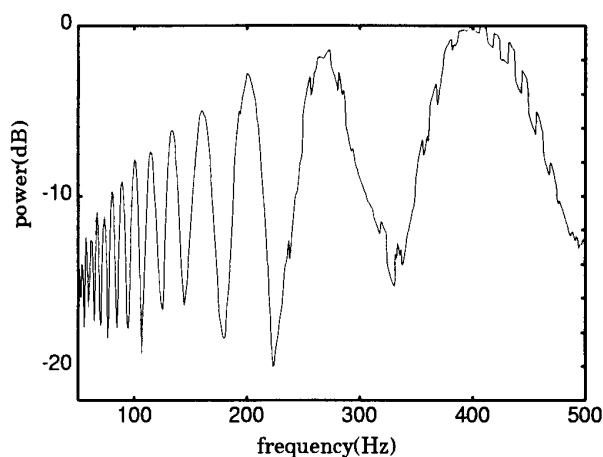


図6. 振幅情報が全て一定のときの調波 Wavelet 変換 ( $SNR = 15dB$ )

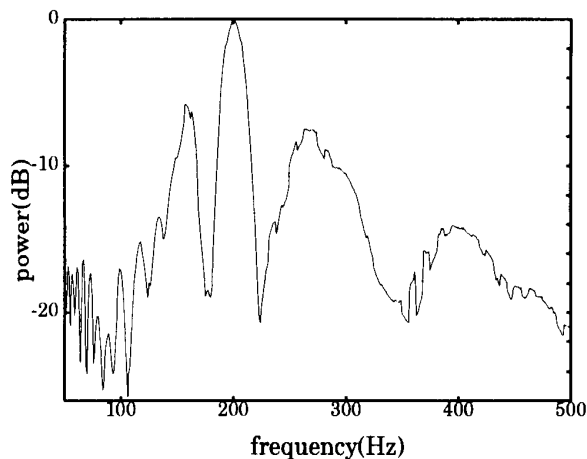


図7. 帯域制限信号の調波 Wavelet 変換 ( $SNR = 10dB$ )

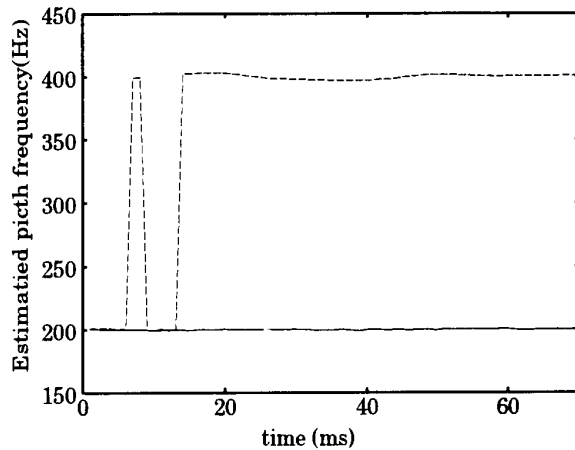


図 8. 推定されたピッチ (実線:調波 Wavelet 変換, 破線:変形相関法、 $SNR = 15dB$ )

#### 4. まとめ

今回、基本周波数が欠落した調波信号から基本周波数を推定する方法を提案し、推定精度について検討した。この結果、 $SNR = 15dB$  以上において基本周波数の推定は可能であり、また比較対象とした変形相関法よりも良い精度で基本周波数が推定が行われた。しかし  $\hat{f}_0$  の時点で、半ピッチ、倍ピッチを推定すること  $SNR = \infty$  以外において起きてしまった。今後の課題として、倍ピッチ、半ピッチといったまちがった推定をどう抑えるかである。

#### 参考文献

- [1] 菅木 禎史 宇佐川 毅 江端 正直 : "調波型 Analyzing Wavelet を用いた Wavelet 変換の時間周波数分解能の検討" :日音議論 2-6-11(1999.3)
- [2] Y.Chisaki: "Spectral analysis of harmonics signal using modified wavelet transform" :*Proc. of WESTPRAC VI* 299-304(1997)
- [3] Yoav Medan,Eyal Yair,Dan CHazan : "Super Resolution Pitch Determinaiotn of Speech Signals" :*IEEE TRANSACTION ON SIGNAL PROCESSING*. vol.39 no.1 40-47(1991)