

## TANDEM-STRAIGHT に基づく基本周波数の抽出と評価について

河原 英紀<sup>†</sup> 森勢 将雅<sup>†</sup> 高橋 徹<sup>†</sup> 西村 竜一<sup>†</sup> 坂野 秀樹<sup>††</sup>

入野 俊夫<sup>†</sup>

<sup>†</sup> 和歌山大学システム工学部  
〒640-8510 和歌山市栄谷 930

<sup>††</sup> 名城大学理工学部  
〒468-8502 名古屋市天白区塩釜口 1-501

E-mail: †{kawahara,s055068,tall,nisimura,irino}@sys.wakayama-u.ac.jp, ††banno@ccmfs.meijo-u.ac.jp

**あらまし** 簡単な処理で、周期信号のパワースペクトルの時間方向の変動を取り除くことのできる方法 (TANDEM 窓法) を用いて、音声分析変換合成法 STRAIGHT の再構築を進めている。ここでは、TANDEM スペクトルと STRAIGHT スペクトルの比を利用することにより、基本周期のみを選択的に抽出する方法を提案する。複数の基本周期を仮定し、それぞれの基本周期の成分を選択的に検出する機構を用意して組み合わせることにより、広い範囲にわたって、同じ基準で基本周波数の確からしさを評価することが可能となる。さらに、こうして求められた基本周波数情報を用いて、周波数軸上の直交位相信号を用いて非周期成分を推定する方法を提案する。提案した方法は、二重音声のような従来の基本周波数抽出法では扱いが難しい信号に対しても、合理的な分析結果を与える。

**キーワード** 周期信号、STRAIGHT、基本周期、非周期成分、確率

## Fundamental frequency estimation based on TANDEM-STRAIGHT and its evaluation

Hideki KAWAHARA<sup>†</sup>, Masanori MORISE<sup>†</sup>, Toru TAKAHASHI<sup>†</sup>, Ryuichi NISIMURA<sup>†</sup>, Hideki

BANNO<sup>††</sup>, and Toshio IRINO<sup>†</sup>

<sup>†</sup> Wakayama University  
930 Sakaedani, Wakayama, 640-8510 Japan

<sup>††</sup> Meijo University  
1-501 Shiogamaguchi Tenpaku0ku, Nagoya, 468-8502 Japan

E-mail: †{kawahara,s055068,tall,nisimura,irino}@sys.wakayama-u.ac.jp, ††banno@ccmfs.meijo-u.ac.jp

**Abstract** TANDEM method, a power spectrum estimation method for periodic signals was proposed to provide temporally stable representation and has been applied to reformulate STRAIGHT, a system for speech analysis modification and synthesis. This article proposes a fundamental period estimation method based on a ratio between TANDEM spectrum and STRAIGHT spectrum. By providing specialized F0 detectors for multiple F0 candidates and integrating individual clues, the proposed method selectively detects fundamental components and yields a probability measure for each estimate. It also provides a method to estimate aperiodicity in each frequency band by making use of estimated fundamental frequency information to design a quadrature signal on the frequency axis for filtering periodic spectral component due to the signal periodicity. The proposed method is capable of representing pathological speech signals more precisely than conventional methods.

**Key words** periodic signal, STRAIGHT, fundamental period, aperiodic component, probability

## 1. はじめに

音声分析変換合成法 STRAIGHT [5], [10] では、周期性に起因する時間方向と周波数方向の周期的構造をスペクトログラムから選択的に取り除くために、相補的時間窓と周波数軸上での逆フィルタに基づく方法が用いられていた。これらの複雑で見通しの悪い定式化に基づく STRAIGHT の実装は、森勢らによって発明された TANDEM 法 [11] と、標準化定理の新しい定式化である consistent sampling [12], [13] を用いることにより、より見通しが良く効率が良いものに置き換えられた [6]。しかし、基本周波数の推定と非周期成分の推定 [9] には、同様に調整を要する見通しの悪い部分が残っている。基本周波数の推定については、より精密な方法 [8] を提供しているものの、複雑なコード (Matlab で 3000 行を超える) は、見通しと保守性が悪く、実行速度の遅さも問題となっていた。非周期性についても、精密な推定の試みでは、同様に複雑さと実行速度の遅さが問題となっていた [4]。本報告では、TANDEM スペクトルとそれに基づく STRAIGHT スペクトルから導かれる周期性を表すスペクトル (変動スペクトル) を利用して [7]、これらの問題を解決する方法について説明する。

## 2. 分析位置に依存しないパワースペクトル

基本周波数  $f_0$  およびそれ以上の周波数成分に対する応答が十分に減衰している低域通過型の周波数特性  $H(\omega)$  を有し、時間領域での定義域が有限である非負の時間窓  $w(t)$  を考える。この窓を用いて、基本周期が  $T_0 = 1/f_0$  である周期信号  $x(t)$  を、時刻  $\tau$  において分析して求めたパワースペクトルを  $P(\omega, \tau)$  と表すこととする。このとき、以下で求められる平均パワースペクトル  $P_T(\omega)$  は、分析時刻  $\tau$  に依存しない。

$$P_T(\omega) = \frac{1}{N} \sum_{k=0}^{N-1} P(\omega, \tau + kT_0/N) \quad (1)$$

このパワースペクトルを、TANDEM スペクトルと呼ぶことにする。(以下では、 $N = 2$  とし、 $(P(\omega, \tau - T_0/4) + P(\omega, \tau + T_0/4))/2$  を考えることにする。)

TANDEM スペクトルには、周波数方向に正弦波状に周期  $f_0$  で振動する成分が存在する。周波数軸上で、幅が  $f_0 = \omega_0/2\pi$  である矩形の平滑化関数 (積分が 1 となるように正規化されている) と TANDEM スペクトルを畳むことにより、この振動成分を完全に消去することができる。具体的には、以下により上記と等価な計算ができる。

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda) d\lambda$$

$$P_{TST}(\omega) = [C(\omega + \omega_0/2) - C(\omega - \omega_0/2)]/\omega_0. \quad (2)$$

このようにして求めたパワースペクトル  $P_{TST}(\omega)$  を STRAIGHT スペクトルと呼ぶことにする。これらの簡単な処理により、信号の周期性による影響が時間方向と周波数方向の双方において取り除かれたスペクトログラムが得られる。

## 3. 周期性の抽出

ここでは、次式で定義される変動スペクトル  $P_C(\omega)$  が、周波数軸上で基本周波数の周期の正弦波状 ( $\cos$  位相) の変動を示す性質を利用する。

$$P_C(\omega) = \frac{P_T(\omega)}{P_{TST}(\omega)} - 1. \quad (3)$$

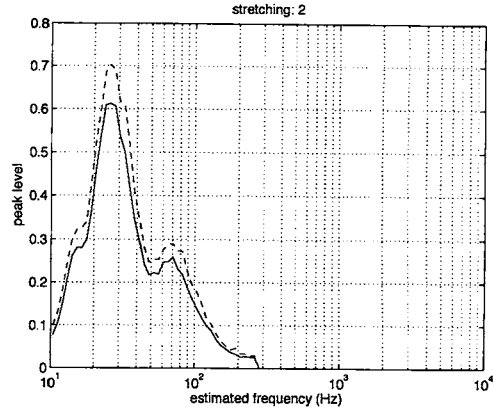


図1 Equal probability contour of peak levels of  $A(\tau; T_C)$ . Solid line represents 90% and dashed line represents 95%. (Blackman window with  $4T_C$  in length.  $T_C = 40$  ms)

ここで、 $P_T(\omega)$  と  $P_{TST}(\omega)$  は、それぞれ前節で説明した TANDEM スペクトルと STARIGHT スペクトルを表す。

基本周波数を求めるときには、基本周波数は既知ではない。また、音声の基本周波数は一般的には時間的に変化しており、パワースペクトルの周波数軸上での周期構造には、分析に用いた時間窓内での基本周波数の変化に比例した空間周波数の振幅変調が加わる。したがって、変動スペクトルの成分に基づいて基本周波数を求めようとする場合には、この振幅変調よりも狭い範囲の成分を用いる必要がある。具体的には、例えば以下のような関数を用いて周波数範囲を選択する。

$$w_{\omega_0, N}(\omega) = c_0 (1 + \cos(\pi\omega/N\omega_0)), \quad (4)$$

なお、 $c_0$  は  $w_{\omega_0, N}(\omega)$  の積分を 1 とするための正規化係数である。この周波数重みをかけて Fourier 変換することで、ある基本周波数を仮定した場合の変動成分を抽出することができる。以下では、 $A(\tau; T_C)$  のように仮定した基本周期  $T_C$  を明記することとする。(ここでは、 $\omega_0 = 2\pi/T_C$  とする)

$$A(\tau; T_C) = \int_{-\infty}^{\infty} w_{\omega_0, N}(\omega) P_C(\omega; T_C) e^{-j\omega\tau} d\omega, \quad (5)$$

定数  $N$  は、実際の音声の基本周波数の変化速度の統計や許容する誤り率に基づいて決められる。

仮定した基本周期により求められる  $A(\tau; T_C)$  が分析対象の信号の実際の基本周期で最大にピークとなる範囲は限られているため、帯域を分けて複数の基本周期を仮定する必要がある。このようにして求められるそれぞれの帯域での計算結果を、時間遅れ領域での荷重  $w_{LAG}(\tau; T_C)$  を掛けて統合することとする。

$$\bar{A}(\tau) = \frac{1}{M} \sum_{k=1}^M w_{LAG}(\tau; T_L 2^{\frac{k-1}{L}}) A\left(\tau; T_L 2^{\frac{k-1}{L}}\right), \quad (6)$$

ここで、 $w_{LAG}(\tau; T_C)$  は、周期  $T_C$  で値が 1 となるような単峰性の重み関数である。 $L$  は、オクターブあたりに設ける帯域の数を表す。

### 3.1 TANDEM 用パラメタの決定

$T_C$  を仮定して求めた  $A(\tau; T_0)$  は、抽出したい成分以外のランダムな成分に起因するパワースペクトルの周波数軸上での変動も抽出する。抽出されるこの不要な成分と、本来の目的である周期成分との S/N 比が最大になるように、TANDEM 分析

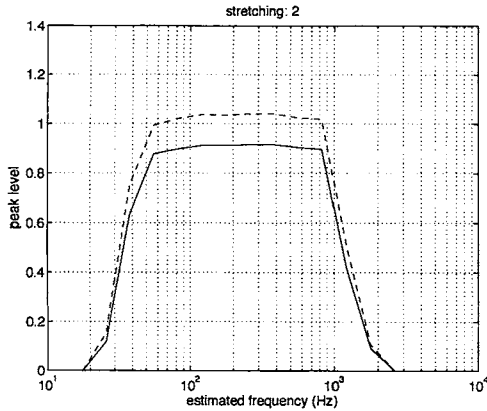


図2 Equal probability contour of peak levels of  $\bar{A}(\tau)$ . Solid line represents 90% and dashed line represents 95%.

に用いる時間窓のサイズを調整する。例えば、Blackman 窓を用いた場合には、窓長を仮定する周期  $T_C$  の 4 倍に設定すると、ランダム入力に対する分布のピークの位置を  $T_C$  に一致させることができる。周期  $T_C$  の周期成分に対する応答もこの場合に最大となり、S/N 比が最大となる。この様子を図 1 に示す。

図 1 には、本来検出すべき周期性に対応する 25 Hz のピークの外に、幾つかのピークが表れている。これらは元の窓のサイドローブや、長過ぎる時間窓を用いることでパワースペクトル上での空間周波数成分に生ずる高調波歪みに起因する。これらの不要なピークを、以下のような重み関数  $w_{LAG}(\tau; T_C)$  を用いて抑圧する。

$$w_{LAG}(\tau; T_0) = 0.5 + 0.5 \cos\left(\pi \log_2\left(\frac{\tau}{T_0}\right)\right), \quad (7)$$

このような設定の下で式 6 により求められた  $\bar{A}(\tau)$  のピークの分布を図 2 に示す。ここでは、仮定する基本周波数を 40 Hz から 600 Hz の範囲に 1/2 オクターブ毎に置いた。注目する帯域では、ランダム入力に対するピーク値は周波数に依存していないことが分かる。

図 2 に示したように、 $\bar{A}(\tau)$  のピークの分布は、注目する帯域では周波数に依存しない。したがって、入力がランダムであると仮定した場合のピークの出現確率をピーク値の関数として表すことができる。図 3 に、その結果を示す。図より、ランダムな揺らぎによって生じたピークを周期信号が存在する証拠であると誤判定する危険率として 1% を許容する場合には、判定の閾値を 1.23、0.1% の場合には、1.43、0.01% の場合には、1.53 に設定すれば良いことが分かる。

### 3.2 瞬時周波数を用いた推定値の改良

十分に狭い帯域通過フィルタの出力の瞬時周波数を、帯域フィルタの中心周波数の関数と見なすこととする。フィルタの帯域内に顕著な正弦波成分が含まれている場合には、その正弦波の周波数が、中心周波数から出力の瞬時周波数への写像の不動点を与える [1], [2], [9]。この性質を利用して、 $\bar{A}(\tau)$  のピーク位置を初期値として、線形補間により基本周波数の推定値を改良する。

この方法に必要なのは、初期値の近傍の二点での瞬時周波数の値である。ここでは以下に示すような直交位相信号を用いて、Flanagan の計算法 [3] を用いて瞬時周波数を求めることとする。

$$h_N(t; T) = w_{T,N}(t) \exp(-2\pi j\omega t). \quad (8)$$

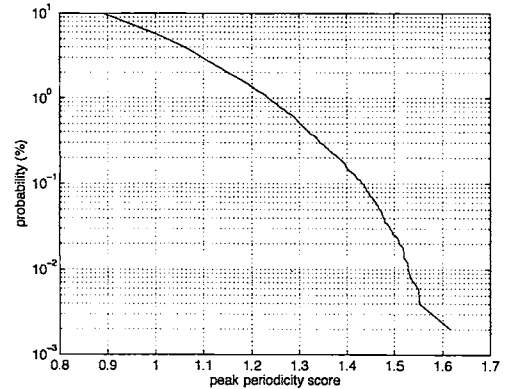


図3 Probability of  $\bar{A}(\tau)$  peaks for random inputs.

振幅包絡  $w_{T,N}(t)$  としては、例えば長さが  $NT$  の Blackman 窓を用いる。

こうした準備の下で、基本周波数の初期推定値  $\omega_{r,0}$  の近傍の 2 つの周波数  $\omega_0, \omega_1$  における瞬時周波数  $\lambda_0, \lambda_1$  を求める。これらは、周波数-瞬時周波数平面上の 2 点を指定する。この二点が直線で結ばれているとすると、以下の関係が成立する。

$$\begin{bmatrix} \lambda_0 \\ \lambda_1 \end{bmatrix} = \begin{bmatrix} \omega_0 & 1 \\ \omega_1 & 1 \end{bmatrix} \begin{bmatrix} u_0 \\ u_1 \end{bmatrix} \quad (9)$$

ここで、不動点の条件  $\lambda(\omega) = \omega$  を用い、改良された基本周波数の推定値  $\omega_{r,1}$  を次式により求める。

$$\omega_{r,1} = \frac{u_1}{1 - u_0} \quad (10)$$

こうして求められた改良された基本周波数の推定値  $\omega_{r,1}$  を初期値として、更に、この初期値を挿むように上下の周波数における瞬時周波数を求めて式 9 と式 10 を計算することで、更に改良された推定値  $\omega_{r,2}$  を求めることができる。なお、周波数から瞬時周波数への写像の形状より、窓長が  $3T_C$  の Blackman 窓を用いることとした。また、瞬時周波数の計算には、第一から第三までの調波成分を用いることとした。

## 4. 非周期成分の抽出

音声信号は、有声音であっても完全には周期的ではない。基本周波数は時間とともに変化し、スペクトルも、調音器官の運動により変化する。しかし、これらの変化は、基本周波数や STRAIGHT スペクトルとして、既に推定されている。分析合成系を構成することを目的とするのであれば、周期性からの外れの中から、基本周波数の変化やスペクトルの変化として抽出できる成分によるものについては、除外することが必要となる。実際には、以下の議論により、基本周波数の変化の影響のみを除去することとした。

**基本周波数の変化:** 基本周波数の変化の影響は、周波数に比例して拡大する。TANDEM の計算に含まれる  $1/2T_0$  ( $T_0$  は、基本周期) の時間差における周波数変化  $\Delta f_0$  は、調波の次数を  $k$  とするとそれに比例して拡大され、 $k\Delta f_0$  となる。実際の音声でも、TANDEM スペクトルの高い周波数領域には処理の妨害となるモアレが生じている。

**スペクトルの変化:** スペクトルの変化には、基本周波数のような周波数に比例した拡大機構は存在しない。TANDEM の計算に含まれる  $1/2T_0$  ( $T_0$  は、基本周期) の時間差は、そのような拡大機構が存在しない場合には、小さな時間間隔である。ま

た、もともと声帯振動がカオス状の成分を含むため、調音器官の運動によるスペクトル変動と音源の非周期性による変動を精度良く区別することは困難である。

#### 4.1 基本周波数の変化の影響の排除

時間的に変化する信号  $s(t)$  の基本周波数の瞬時周波数を  $f_0(t) = \omega_0(t)/2\pi$  とすると、(振幅を無視した) 基本波成分の波形  $s_0(t)$  は、以下のように表される。なお、ここでは基本波の位相  $\varphi(t)$  の初期値を 0 と置いた。

$$\begin{aligned} s_0(t) &= \sin \varphi(t) \\ \varphi(t) &= \int_0^t \omega_0(\tau) d\tau \end{aligned} \quad (11)$$

ここで、位相が一定の速度  $2\pi f_C$  で変化する場合の時間軸と解釈できる次の量  $\lambda(t)$  を求めておく。

$$\lambda(t) = \frac{\varphi(t)}{2\pi f_C} \quad (12)$$

この時間軸を用いて  $s_0(t)$  を  $\lambda$  の関数として表すと、その瞬時周波数は定数  $f_C$  となる。同様な議論は、調波複合音を構成する調波成分についても成立する。

#### 4.2 直交位相信号による基本周期成分の評価

前の節で説明した処理により、基本周波数が定数  $f_C = \omega_C/2\pi = 1/T_C$  となるように信号が変換されているため、この周期の成分のみの周期性を評価すれば良い。まず、基本周波数成分による周波数軸上での周期構造の強さを調べるために、次のような直交位相信号を作成する。

$$h_N(\omega; T_C) = w_{\omega_C, N}(\omega) \exp(2\pi j\omega/\omega_C). \quad (13)$$

ここで、 $w_{\omega_C, N}(\omega)$  は、周期構造を調べる場合に用いる空間周波数方向での振幅包絡である。例えば、F0 の場合と同様に、以下のような raised cosine 型の関数を用いることができる。

$$w_{\omega_C, N}(\omega) = c_0 (1 + \cos(\pi\omega/N\omega_C)), \quad (14)$$

この直交位相信号を用いて、F0 の推定で定義した、変動スペクトル  $P_C(\omega; T_C)$  の中に含まれる  $\omega_C$  の速度で変動する成分の強さ  $Q_C^2(\omega; T_C)$  を、次式により計算する。

$$Q_C^2(\omega; T_C) = \left| \int_{-\infty}^{\infty} h_N(\lambda; T_C) P_C(\omega - \lambda; T_C) d\lambda \right|^2 \quad (15)$$

#### 4.3 非周期成分の割合の推定

この変動成分の大きさから、信号に含まれている周期成分と非周期成分の相対的な割合を推定することが解くべき課題である。まず、直交位相信号による畳込みを、記号  $\circ$  を用いて表すことにする。この表記を用いて式 15 を書き換える。

$$\begin{aligned} Q_C^2 &= |h_N \circ P_C(\omega; T_C)|^2 \\ &= \left| h_N \circ \frac{P_T(\omega; T_C)}{P_{TST}(\omega; T_C)} - 1 \right|^2 \\ &= \left| h_N \circ \frac{P_T(\omega; T_C) - P_{TST}(\omega; T_C)}{P_{TST}(\omega; T_C)} \right|^2 \end{aligned} \quad (16)$$

TANDEM スペクトルは、STRAIGHT スペクトルに  $h_N$  で選択的に除去される周期的な変動分が加えられたものであること、また、その周期的な変動分には、信号の周期性に起因するものと、信号のランダムな変動に起因するものが含まれることに注意する。ここで、 $\Delta P_P$  を、信号の周期性に起因する変動分、 $\Delta P_R$  をランダムな変動に起因する変動分、 $P_P$  を、周期成分の STRAIGHT スペクトル、 $P_R$  を、ランダムな成分の

STRAIGHT スペクトルと表記することとする。

ここで、 $h_N$  の定義域の幅の中では  $P_P(\omega; T_C)$  と  $P_R(\omega; T_C)$  が定数と見なせると仮定期待値を計算する。

$$Q_C^2 = \frac{V[h_N \circ \Delta P_P]}{P_P^2 + P_R^2} + \frac{V[h_N \circ \Delta P_R]}{P_P^2 + P_R^2} \quad (17)$$

ここで  $V[x]$  は、 $x$  の分散を表す。

周期信号の場合には、窓関数が決まれば、 $V[h_N \circ \Delta P_P]$  の値は、 $P_P$  の定数  $C_P$  倍として一意に決まり、ランダムな成分の値  $V[h_N \circ \Delta P_R]$  についても、窓関数と  $h_N$  が決まれば実効的な TB 積から、 $P_R$  の定数  $C_R$  倍として一意に決まる。結局、以下が得られる。

$$Q_C^2 = \frac{C_P^2 P_P^2}{P_P^2 + P_R^2} + \frac{C_R^2 P_R^2}{P_P^2 + P_R^2} \quad (18)$$

周期成分の自乗平均値の意味での平均振幅を  $a_{PRD}(\omega)$ 、非周期成分の平均振幅を  $a_{RND}(\omega)$  とすると、以下のように表されることが分かる。

$$\begin{aligned} a_{RND}(\omega) &= \sqrt{\frac{C_P^2 - Q_C^2}{C_P^2 - C_R^2}} \\ a_{PRD}(\omega) &= \sqrt{\frac{Q_C^2 - C_R^2}{C_P^2 - C_R^2}} \end{aligned} \quad (19)$$

#### 4.4 係数の決定と実装

$a_{PRD}(\omega)$  と  $a_{RND}(\omega)$  の計算のためには、測定で求められる値  $Q_C^2$  の他に、TANDEM で用いる窓により決まる  $C_P$  と、分析条件により変化する  $C_R$  の統計的な性質が必要となる。

##### 4.4.1 周期成分に対する係数 $C_P$

基本周期の 2.4 倍の Blackman 窓を用いた今回の分析では、シミュレーションの設定により若干の差があるが、 $C_P = 0.56$  が得られた。

##### 4.4.2 ランダムな成分に対する係数 $C_R$

この係数は、直交位相信号  $h_N(\omega; T_C)$  の周波数方向の広がりを表す  $N$  に依存する。図に  $N = 2$  の場合と  $N = 16$  の場合の観測値  $Q_C$  の分布を示す。図より明らかのように  $N = 2$  の場合には、分布が大きく広がる。この問題を避けるには、複数の分析フレームでの結果を平均することにより TB 積を大きくすることが必要となる。

ここでは、実際に利用される可能性のある範囲をカバーするように、分析フレーム周期、周波数方向の広がり  $N$ 、統合する分析フレーム数との組合わせについてシミュレーションを行い  $Q_C$  を求め、その平均値と分散を 3 次元のテーブルとして格納した。必要な  $C_R$  の値は、このテーブルから線形補間により求めることとした。実際の計算では、 $C_R$  の値は、該当する条件の  $Q_C$  の平均値に、 $Q_C$  の標準偏差の定数倍を加えたものとする。定数の具体的な値は、主観評価実験および評価値の consistency の条件を最適化する客観評価を用いたシミュレーション等により決定する。

##### 4.4.3 柔らかな飽和特性を用いた実装

式 19 の  $Q_C$  にはランダムな成分が含まれているため実際に観測される値は確率的に揺らいでいる。従って、 $C_P$  と  $C_R$  をそのまま用いると、周期成分が少ない領域では、周期成分が過大に推定される。また、スペクトル変化を無視しているため、周期成分が多い部分では、周期成分が過小に推定される。前述の揺らぎとこれらのバイアスの補償のため、このままでは根号内が負となったりそれぞれの成分が 100% を超えるなどの不合理な値を持つ場合が生ずる。ここでは式 19 の根号内の値  $x$  を次式により変換することで、それらの問題を回避する。

$$g(x) = \frac{1}{\alpha} \log \frac{1 + \exp(-\alpha x)}{1 + \exp(-\alpha(x-1))} + 1 \quad (20)$$

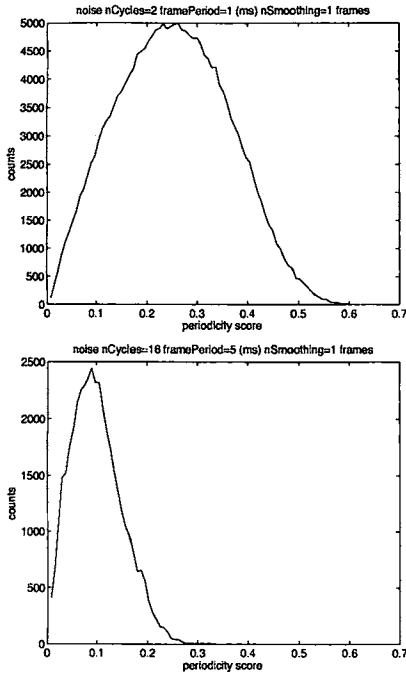


図4 Histogram of periodicity rates. Frequency spreads are  $N-2$  for upper plot and  $N=16$  for lower plot.

ここで soft さを定める値  $\alpha$  は、受聴試験などにより定める必要がある。

## 5. 分析例

ここでは提案した方法の動作を、試験用信号と実音声を用いて調べる。

### 5.1 チャープ信号による試験

1 サンプルづつ周期が毎回長くなる信号の分析例を図5に示す。結果を図5に示す。図では、得られた  $\hat{A}(\tau)$  の正のピークを、値の大きなものから順に5個、表示している。最も値の大きな候補の周波数は、各時刻でのパルス間隔の逆数として表される基本周波数の真値と一致している。図の下に示すように、第一位の候補の値は1.8付近にあり、0.0001%以下の危険率に対応している。二位以下の候補の値は0.8以下であり、ランダムな揺らぎによるものと区別できない。すなわち、提案する方法には、実質的に半ピッチや倍ピッチの誤検出の問題が存在しない。それらに相当する大きな値のピークが存在する場合には、むしろ、分析対象とした振動自体に、それらのピークに対応する構造があると解釈すべきであろう。

チャープ信号は、局所的には周期的な信号である。したがって、正しい分析が行われた場合には、全ての時間-周波数領域において周期成分が支配的となる。図6に示すように、提案した方法の出力は、ほとんど全ての部分で周期性が飽和していることを示している。これは、周期性の観測値  $Q_C$  のヒストグラムを計算すると、先に説明した  $C_P$  の値付近に分布が集中していることに対応する。このように、提案した方法は、基本周波数が変化する信号に対しても適切な分析結果を与えるものであることが分かる。

### 5.2 実音声の分析例

図7に、実際の音声の分析例を示す。図の上段の太く見える部分が、瞬間周波数を用いて修正された基本周波数に相当する。

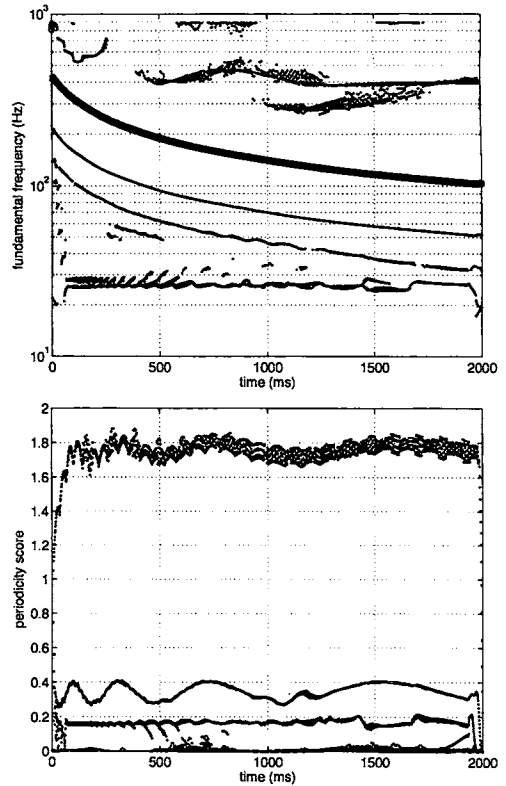


図5 F0 candidates (upper plot) and their periodicity scores (lower plot). The most likely F0 candidate at each frame is represented using a circle. (Test signal: pulse train with increasing fundamental period)

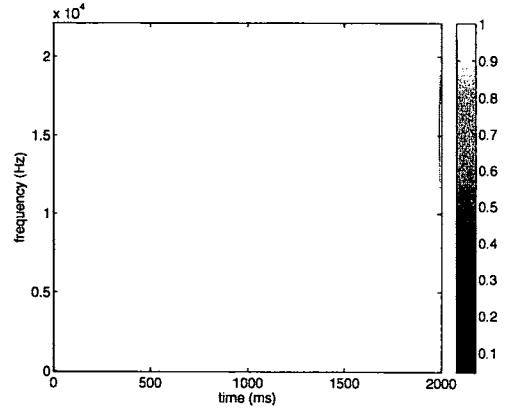


図6 Map of periodicity rates for the signal analyzed in Figure 5

この例では、下段のピーク値が示すように、最も高い評価値を示すピークを、基本周波数の候補としている。図より明らかにように、隣接フレームでの値の連続性等を併せて後処理を行うことで、より信頼できる基本周波数軌跡を得ることができる。

なお、下段の400ms付近では、第二位の候補の評価値が先に述べた1%の危険率のレベルを超えていることが分かる。これは、声門の閉止時点だけではなく基本周期の中央付近にも声門の開き始めによる音響的イベントが存在していることを示唆

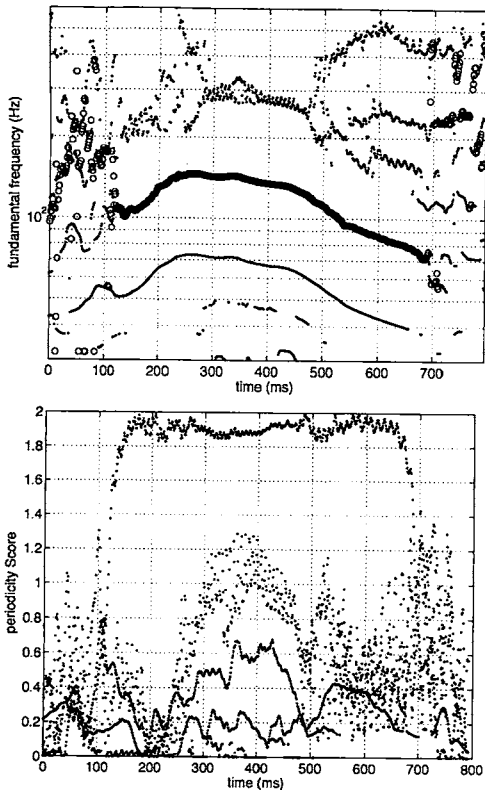


図7 実音声の分析例。上段に基本周波数の候補、下段にそれぞれの候補のピーク値を示す。上段では、瞬時周波数を用いて改良した最終的な推定値を○印で示している。音声は、男性の発声した日本語連続母音「あいうえお」。

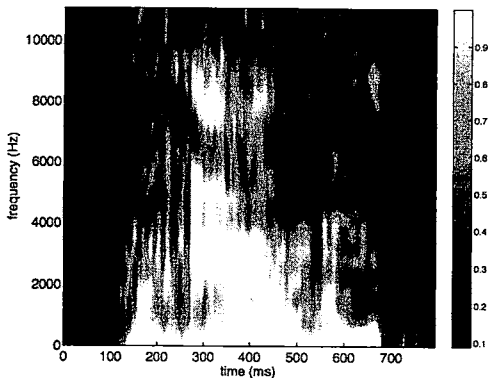


図8 男性の発声した母音連鎖の周期成分のマップ。

していると解釈すべきであろう。

基本周波数抽出で用いたチャープ信号と、男性による連続母音の分析結果をそれぞれ図6と図8に示す。

図8は、実音声の場合の分析結果を示す。4kHz以下で、有声音部分の周期性が支配的であることが分かる。図7において、第二位の候補の周期性のレベルが危険率1%のラインを超える部分の存在を指摘した。図8においても、対応する部分において、高い周波数まで周期成分が支配的となっている。これは、声道と声帯振動の相互作用に基づく拘束の存在を示唆するもの

であり、人間の発声機構の解明の観点からも非常に興味深い。

## 6. まとめ

周期信号に対して時間的な変動の無いパワースペクトルを推定する TANDEM 法を用い、基本周波数および非周期性指標を求めるための新しい方法を提案した。これらの方法は、実質的に基本波成分だけを抽出し、しかも確率として解釈できる評価指標が得られるという特徴がある。これらの方法では、実装に高速フーリエ変換が多用されており、従来の STRAIGHT の1/5~1/10で、分析合成を遂行することができる。また、フレーム毎の分析に基づく可読性の高い実装となっている。今後は分析パラメタの最適化および主観評価実験によるシステムとしての従来の STRAIGHT との比較と検証が課題となる。

### 謝 辞

本研究の一部は、科学研究費補助金基盤 (A)19200017 による支援を得て行われた。本資料の草稿について、討論して頂いた名工大の徳田研究室の方々に感謝する。

### 文 献

- [1] T. Abe, T. Kobayashi, and S. Imai. The IF spectrogram: A new spectral representation. In *Proc. ASVA-97*, pp. 423-430, Tokyo, 1997.
- [2] F. J. Charpentier. Pitch detection using the short-term phase spectrum. In *Proc. ICASSP 1986*, pp. 113-116, 1986.
- [3] J. L. Flanagan and R. M. Golden. Phase vocoder. *Bell Syst. Tech. J.*, Vol. 45, pp. 1493-1509, 1966.
- [4] 河原英紀, 森勢将雅, 高橋徹, 入野俊夫, 坂野秀樹, 藤村靖. STRAIGHT スペクトルに基づく音源信号の抽出と非周期成分の評価について. *信学技報*, Vol. 36, No. SP2006-83, pp. 43-48, 2006.
- [5] 河原英紀. Vocoder のもう一つの可能性を探る—音声分析変換合成システム STRAIGHT の背景と展開—. *日本音響学会誌*, Vol. 63, No. 8, pp. 442-449, 2007.
- [6] 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 入野俊夫, 坂野秀樹. 時間方向および周波数方向の周期性の影響を除去した周期信号のパワースペクトルの表現について. *信学技報 SP2007-26*, Vol. 37, No. 6, pp. 395-400, 2007.
- [7] 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 坂野秀樹, 入野俊夫. 分析位置に依存しない周期信号のパワースペクトル推定法に基づく音声分析変換合成法 STRAIGHT の再構成について. *信号処理シンポジウム*, 2007.11. [CD-ROM].
- [8] H. Kawahara, A. de Cheveigné, H. Banno, T. Takahashi, and T. Irino. Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT. In *Interspeech '05*, pp. 537-540, Lisboa, 2005.
- [9] H. Kawahara, H. Katayose, A. de Cheveigné, and R. D. Patterson. Fixed point analysis of frequency to instantaneous frequency mapping for accurate estimation of F0 and periodicity. In *Proc. Eurospeech '99*, Vol. 6, pp. 2781-2784, 1999.
- [10] H. Kawahara, I. Masuda-Katsuse, and A. de Cheveigné. Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction. *Speech Communication*, Vol. 27, No. 3-4, pp. 187-207, 1999.
- [11] 森勢将雅, 高橋徹, 河原英紀, 入野俊夫. 窓関数による分析時刻の影響を受けにくい周期信号のパワースペクトル推定法. *電子情報通信学会誌 D*, 2007. (12月号に掲載予定).
- [12] Michael Unser. Sampling - 50 years after Shannon. *Proceedings of the IEEE*, Vol. 88, No. 4, pp. 569-587, 2000.
- [13] Michael Unser and Akram Aldroubi. A general sampling theory for nonideal acquisition devices. *IEEE Trans. Signal Processing*, Vol. 42, No. 11, pp. 2915-2925, 1994.