

TANDEM-STRAIGHT に基づく 基本周波数抽出法の最適化に関する一検討

板垣 英恵[†] 森勢 将雅^{††} 西村 竜一[†] 入野 俊夫[†] 河原 英紀[†]

[†] 和歌山大学システム工学部 〒640-8510 和歌山市栄谷 930

^{††} 関西学院大学理工学部 〒669-1337 兵庫県三田市学園 2-1

E-mail: †{s090008,nisimura,irino,kawahara}@sys.wakayama-u.ac.jp, ††morise@kwansei.ac.jp

あらまし 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である TANDEM と、それに基づくスペクトル包絡の推定法 STRAIGHT とを組み合わせた、基本周波数抽出法が提案されている。本資料では、この TANDEM-STRAIGHT に基づく基本周波数抽出法に含まれている設計パラメタの役割と抽出器の性能との関係を調べ、それらの最適化を試みた。取上げた設計パラメタは、周波数軸上での周期性の評価に用いる調波の個数と、それぞれ異なった基本周波数に特化した複数の基本周波数抽出器を統合する際の、それぞれの抽出器の守備範囲を定める重み関数の幅である。まず、それぞれのパラメタがどのように抽出性能に影響するかを説明した後、EGG 信号を同時記録した音声データベースを用いて、パラメタの値と抽出器の性能との関係を定量的に評価した。その結果、Gross error を評価指標とした場合の最適値として、調波の個数は 3、重み関数の幅は、特化した基本周波数の $1/1.2$ が採用された。これらの最適値を用いた場合の性能は、これまでの STRAIGHT [7] の基本周波数抽出器、および広く引用されている YIN と比較しても、遜色ないことが確認された。

キーワード 周期信号 基本周波数 TANDEM STRAIGHT 基本周期

Parameter optimization for a fundamental frequency extractor based on TANDEM-STRAIGHT

Hanae ITAGAKI[†], Masanori MORISE^{††}, Ryuichi NISHIMURA[†], Toshio IRINO[†], and Hideki
KAWAHARA[†]

[†] Wakayama University, 930 Sakaedani, Wakayama, 640-8510 Japan

^{††} Kwansei Gakuin University, 2-1 Gakuen, Mita, Hyogo, 669-1337 Japan

E-mail: †{s090008,nisimura,irino,kawahara}@sys.wakayama-u.ac.jp, ††morise@kwansei.ac.jp

Abstract A fundamental frequency extractor based on a temporally stable power spectral representation for periodic signals (TANDEM spectrum) and a spectral envelope derived from the representation (STRAIGHT spectrum) is proposed. This article describes roles of system parameters of the proposed method and their effects on system performance and reports results of preliminary optimization of them. System parameters investigated are; number of harmonic component for detecting hypothesized periodicity peak and weighting width on the log-lag domain for integrating specialized individual F0 detectors. Detailed descriptions of these parameters and their impact on F0 extraction performance are presented and they were further investigated using a database consisting of simultaneous recording of speech and EGG (electroglottogram) signals. Test results indicated that the proposed method has comparable performance with F0 extractors used in STRAIGHT and other popular F0 extractors such as YIN, when three harmonic components are used and weighting with a width of $1/\sqrt{2}$ of the center lag is used.

Key words periodic signal fundamental frequency STRAIGHT TANDEM

1. はじめに

音声の分析には、短時間フーリエ変換が用いられる [5]。音声周期信号（時間的に変化しないインパルス応答と周期的パルス列の畳み込みによって表される信号）として近似すると、短時間フーリエ変換は、窓関数を用いてその一部を抽出したもののフーリエ変換を求めることに相当する。このようにして求められたパワースペクトルは、元のインパルス応答が時間的に変化しないものであるにもかかわらず、時間窓の位置とパルスの相対位置に応じて、時間方向で周期的に変化するものとなる。周期信号のパワースペクトルには、周波数方向にも基本周波数を周期とする周期性があるため、スペクトログラムは、時間方向と周波数方向の双方に周期に基づく構造が重なったものとなる。これらの周期性に起因する構造を選択的に除去する方法が STRAIGHT [6] である。

TANDEM [3] は、基本周期の半分だけの時間間隔を隔てた二つの時間窓により求められるパワースペクトルを平均するという簡単な操作により、分析位置に依存しないパワースペクトルを求める方法である。TANDEM を用いることにより、STRAIGHT は、TANDEM-STRAIGHT として根本的に再構成された。

TANDEM-STRAIGHT に基づく基本周波数抽出法 [1], [2] は、TANDEM により求められるスペクトルの性質と、TANDEM-STRAIGHT により求められるスペクトルの性質を利用した方法である。この方法には、基本周波数の抽出性能に影響を与える設計パラメタが含まれている。しかし、それらの設計パラメタの値と抽出性能の関連についての詳細な定量的評価は行われておらず、様々な条件で収録された音声の分析に最適な設定方法は、明らかではない。本資料では、どのような設計パラメタが含まれており、それらがどのように抽出性能に影響するかを定性的に説明した後、複数のデータベースを用いた定量的な評価により、それらの最適値を明らかにする。最後に、求められた最適値を用いた場合の性能を、幾つかの代表的な基本周波数抽出法と比較し、TANDEM-STRAIGHT に基づく方法の特長を明らかにする。

2. TANDEM-STRAIGHT と周期性

ここでは、TANDEM とその STRAIGHT への応用についての簡単な説明を行い、TANDEM-STRAIGHT に基づく基本周波数抽出法の原理と、具体的な構成について説明する。次いで、それらに含まれる設計パラメタの意味と、基本周波数の抽出性能に与える影響を、模式図や例を交えて定性的に説明する。

2.1 TANDEM スペクトル

音声波形を $x(t)$ 、TANDEM を構成する窓関数を $w(t)$ とし、基本周波数 f_0 およびその逆数として定義される $T_0 = 1/f_0$ は、既知であるとする。ここでは、基本周波数は時間的に変化しないものとする。すると、ある時刻 t_0 における TANDEM スペクトル（パワースペクトル） $P_T(\omega, t_0)$ は、次のように求められる。

$$P_T(\omega, t_0) = (P(\omega, t_0 - T_0/4) + P(\omega, t_0 + T_0/4))/2 \quad (1)$$

$$P(\omega, t) = \left| \int_{-\infty}^{\infty} x(\lambda)w(\lambda - t)e^{-j\omega\lambda}d\lambda \right|^2,$$

なお、 $w(t)$ の値は、定義域 $(-T_w \leq t \leq T_w)$ の外では常に 0 である。この仮定の下では、TANDEM スペクトルには、時刻に依存した成分が含まれないため、しばらくの間、時間に関する変数 t を取り除き、 $P_T(\omega)$ と表すこととする。

TANDEM スペクトルの計算に用いるための時間窓は、以下の条件を満たす必要がある。1) 窓関数の周波数領域の表現の値が実質的に 0 以外の値となる周波数領域の幅は、 $2f_0$ 以下であること。2) それ以外の領域での値（窓関数の周波数領域の表現のサイドロープのレベル）は、十分に小さいこと。（例えば、メインロープよりも 40dB 以上レベルが低いこと。）

窓の選択と TANDEM の性能の関係については、論文 [10] を参照されたい。

2.2 TANDEM-STRAIGHT への適用

前節で求めた TANDEM スペクトルには、周波数方向に f_0 の周期で振動する成分が存在する。そこで、幅が $f_0 = \omega_0/2\pi$ 、面積が 1 となる矩形の平滑化関数を TANDEM スペクトルに畳みこむことで、この周期的振動を選択的に取り除く。ここでは、実装を意識し、TANDEM スペクトルの周波数に関する積分（累積和として実装）と線形補間を用いる。まず、積分を求める。始点 ω_L は、 $-\omega_0/2$ よりも小さく、上限は $f_s/2 + \omega_0/2$ よりも大きければ良い。なお、 f_s は、標準化周波数を表す。

$$C(\omega) = \int_{\omega_L}^{\omega} P_T(\lambda)d\lambda \quad (2)$$

離散時間系での実装では、周波数も離散化されるため、正確に ω_0 の幅の矩形の関数を作成することができない。そこで、TANDEM スペクトルの周波数方向の累積和（同様に離散化されている）を線形補間し、以下のように ω_0 だけ離れた位置での値の差を用いて、矩形の関数を用いた畳み込みを近似する。

$$P_{TST} = \left[C(\omega + \omega_0/2) - C(\omega - \omega_0/2) \right] / \omega_0 \quad (3)$$

このようにして求めた、時間周波数両方向での周期性の影響を取り除いたスペクトル $P_{TST}(\omega)$ を、STRAIGHT スペクトルと呼ぶ。

2.3 TANDEM-STRAIGHT を用いた周期性検出

TANDEM スペクトルは、信号の周期性を反映し、周波数方向に f_0 の周期で振動している。STRAIGHT スペクトルでは、この周期性を反映した周波数方向の振動だけが、取り除かれている。両者を重ね合わせた際の模式図を図 1 に示す。

ここで、TANDEM スペクトルの STRAIGHT スペクトルとの比を求めると、模式図の図 2 に示すように、値 1 を中心として変動する成分が得られる。

基本周波数の抽出に必要なものは、この変動する成分だけであるので、バイアス分に相当する 1 を引いて、以下のように $P_C(\omega)$ を定義する。 $P_C(\omega)$ を、以下では変動スペクトルと便宜的に呼ぶこととする。

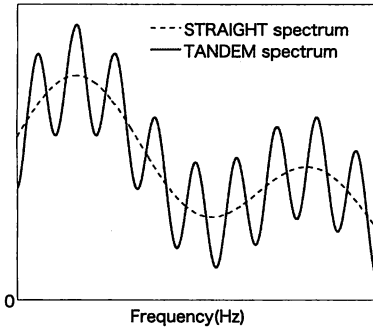


図 1 TANDEM スペクトルと STRAIGHT スペクトル

Fig.1 TANDEM spectrum and STRAIGHT spectrum

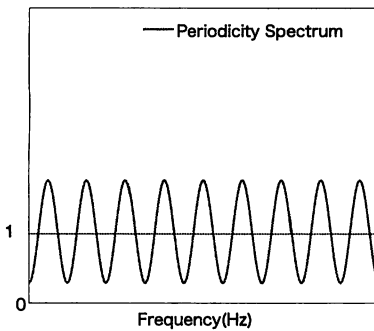


図 2 正規化されたスペクトルは、周波数方向に f_0 の周期を持つ

$$P_C(\omega) = \frac{P_T(\omega)}{P_{TST}(\omega)} - 1 \quad (4)$$

ここで注意しなければならないのが、変動スペクトル $P_C(\omega)$ に至るこれまでの定式化は、基本周期 T_0 が既知、すなわち基本周波数 f_0 も既知であると仮定した上での導出結果だということである。実際の周期信号 $x(t)$ からの F0 抽出においては、基本周期は不明である。そこで、異なった基本周波数を仮定した多数の検出器を用意し、それらの出力を合成して、基本周波数を推定することとする。

2.3.1 特定の周期を仮定した検出器の設計

まず、ある基本周波数を仮定した検出器を作成する。信号の基本周波数と区別するために、仮定した基本周期を T_C 、対応する基本周波数を f_C と表す。変動スペクトルの表記にあたっては、この仮定した基本周波数を $P_C(\omega; T_C)$ のように明記することとする。変動スペクトルには、周期成分のみが含まれているため、以下のように、逆フーリエ変換を行うことにより、F0 検出器を設計することができる。

以下の式で定義される指標 $A(\tau; T_C)$ は、パワースペクトルから導かれる量の逆フーリエ変換として定義されているため、自己相関関数と同様に、時間差 τ の関数となる。まず、この指標が、仮定した基本周期 T_C と同じ周期を有する周期信号を分析した場合に最大になるように、検出器を設計する。

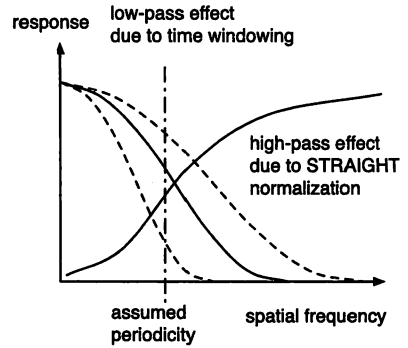


図 3 Time window size effect on F0 extractor

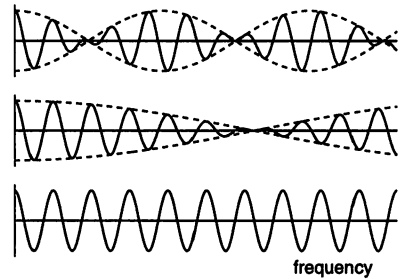


図 4 Effect of FM in fundamental frequency

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

$$w_{\omega_C, N}(\omega) = c_0 \left(1 + \cos\left(\frac{\pi\omega}{N\omega_C}\right) \right) \quad (-N\omega_0 < \omega < N\omega_0) \quad (6)$$

ここで用いられている周波数領域での重み $w_{\omega_C, N}(\omega)$ については後で説明することとし、まず、 $A(\tau; T_C)$ の性質を定性的に説明する。

図 3 に、STRAIGHT スペクトルによる割り算と、TANDEM スペクトルを求める際の時間窓のサイズの影響を模式図を用いて示す。STRAIGHT スペクトルは、滑らかに変化する成分が主体であり、それによる割り算は、空間周波数に対する高域通過フィルタとして働く。時間窓は、逆に空間周波数に対する低域通過フィルタとして働く。図中の破線で示すように、時間窓のサイズが大きい場合には、空間周波数の遮断周波数は高く、小さな場合には、遮断周波数は低い。検出器がピークとなる位置は、これらの影響により変化する。

逆フーリエ変換の際に用いる周波数重み $w_{\omega_C, N}(\omega)$ は、これまで無視して来た基本周波数の時間変化による影響を避けるための仕組みである。基本周波数の時間変化は、周波数軸上のスペクトルの周期的な変化の振幅変調として表れる。この振幅変調の空間周波数は、図 4 に模式的に示すように、基本周波数の変化速度に比例する。基本周波数が時間的に変化しない場合には、図 4 の一番下の図のように、スペクトルの周期的な変動の幅は、周波数に依存せず一定である。しかし、基本周波数が時間的に変化する場合には、図 4 の中央と上に示すように、ス

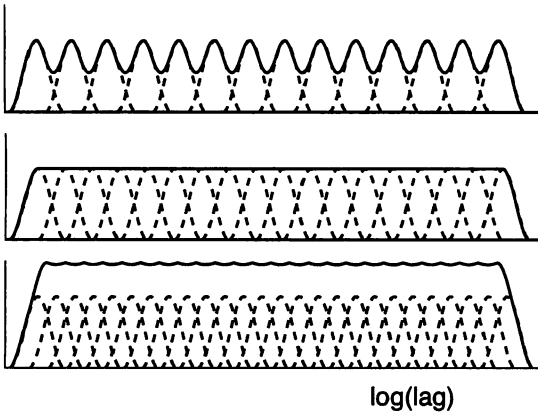


図 5 Integration of specialized detectors

ペクトルの周期的な変動の幅は、余弦状の振幅変調を受ける。基本周波数の時間的な変化が速い場合には、図 4 の上に示すように、振幅変調の空間周波数が高くなる。このように振幅変調されたスペクトルの逆フーリエ変換は、双峰性の特性となり、キャリアに相当する基本周期ではなく、その両側にピークが生ずる。

上記 $w_{\omega_0, N}(\omega)$ の重みは、この振幅変調による問題を避けるため、振幅変調による最初の節が生ずるまでの部分を取り出すために用いられる。

2.3.2 検出器の統合

特定の基本周期 T_C を仮定して設計した抽出器は、 T_C と近い基本周期を有する信号の周期を検出することができる。検出器の出力は、 T_C を中心とする選択性を有する。放送等の番組に出現する音声の基本周波数を抽出するためには、40 Hz から 800 Hz の範囲の基本周波数に対応する必要がある。ここでは、その範囲に、対数周波数軸上等間隔に T_C を仮定して設計した複数の抽出器を配置し、出力を統合することとする。

統合された抽出器の出力を $\bar{A}(\tau)$ と表す。1 オクターブに L 個の抽出器を配置することとすると、仮定する T_C は、守備範囲の基本周波数の下限に対応する周期を T_L として、その $2^{-k/L}$ 倍 (ただし、 $k = 0, 1, \dots, M$) の $M+1$ 個となる。 M は、基本周波数の守備範囲の上限の周波数 f_U と下限の周波数 f_L から $L \log_2(f_U/f_L) + 1$ よりも大きい整数に設定する。これらの検出器を、以下のように統合する。

$$\bar{A}(\tau) = \sum_{k=0}^M w_{LAG}(\tau; T_L 2^{-\frac{k}{L}}) A\left(\tau; T_L 2^{-\frac{k}{L}}\right) \quad (7)$$

ここで用いられている重み $w_{LAG}(\tau)$ の役割を、図 5 に示す模式図を用いて説明する。図では、個別の検出器の出力を破線で表し、統合された検出器の出力を実線で表している。重み関数は、以下の形をしている。 b_w は重み関数の鋭さを決定し、この値を大きくすることでより鋭い窓となる (窓幅は短くなる)。なお、 $b_w \log_2 \frac{\tau}{T_C}$ の絶対値が 1 以上となる τ における重み関数の値は 0 である。

$$w_{LAG}(\tau; T_C) = 0.5 + 0.5 \cos\left(\pi b_w \log_2 \frac{\tau}{T_C}\right) \quad (8)$$

3. パラメタの最適化

前章で説明した TANDEM-STRAIGHT を用いた基本周波数抽出法は、依然、最適化されていないパラメタを含んでいる。未チューニングのパラメタを以下に挙げる。

式 (6) における N

変動スペクトル $P_C(\omega)$ から検出器 $A(\tau; T_C)$ を求める際に、周波数方向において、調波成分をどこまで考慮して切り出すかを決定するためのパラメタである。速くかつ深いピブラートを正しく分析するためには、 N を小さくすべきである。しかし、 N を小さくすると、基本周波数の情報を抽出するために利用出来る調波の個数が少なくなるため、雑音などの影響を受け易くなり、誤差が増加する可能性がある。

式 (8) の重みの広がり方 (鋭さ) を決定する b_w

対数周波数軸上において、等間隔の T_C を仮定して設計された複数の抽出器を統合する際に用いる重み関数である。個別の検出器が、 τ に依存しない場合には、図 5 の中段の図に示すように、重み関数が 50% ずつ重複する場合に、統合された検出器は守備範囲において一定の応答を示す。しかし、個別の検出器自体が選択性を有しているため、重みによる選択性が二重に加わり、上の図に示すように、統合された検出器の応答が一定ではなくなる。この問題を回避するには、設計パラメタ L を大きくして検出器を密に配置すれば良い。検出器を密に配置し過ぎた場合には、図の下段に示すように、統合された検出器の出力は、個別の検出器の出力よりも大きくなる。この場合には、統合された検出器の応答の変動の増加は少なく、検出性能に与える影響は少ない。ただし、 L の増加は計算量の増加を招く。従って、 L は、性能を劣化させない範囲で、少ない方が良い。 L を少なくするためには、重み関数 $w_{LAG}(\tau)$ の設計パラメタである b_w を小さくする必要がある。しかし、重み関数の選択性を甘くすると、個別の検出器の応答が実際には単峰性ではない影響により、統合された検出器の性能が劣化する。

上記の 2 つのパラメタは、分析対象とする音声の性質に依存するため、データに基づいて決定すべきである。

本手法の実用化にあたり、これらのパラメタを最適値にチューニングすることは必須である。以降でそのチューニング方法について解説する。

3.1 チューニング方法

3.1.1 パラメタ変化の影響を考慮した推定精度

まず、各パラメタの変化が F0 抽出に及ぼす影響を調べるため、音声と EGG (Electro glottal graph) 信号を同時収録したデータベースを用いて、客観的評価を行う。声門を経由する経路のインピーダンスの変化を記録した EGG 信号には、声門の開閉が明瞭に記録されている。この信号を時間微分することにより得られる DEGG 信号から基本周波数を求めることにより、正解として用いることのできる信頼出来る値が得られる。以下に具体的な実験手法についてまとめる。

(1) F0 の抽出

TANDEM-STRAIGHT により、音声、EGG 信号からそれぞれ F0 抽出を行う。以下、前者を F0-sp、後者を F0-EGG と呼称する。

(2) ヒストグラムを求める

有声区間において、フレーム（分析時刻）ごとに F0-sp と F0-EGG の一致率を調べる。両者が完全に一致する場合を 100% としてヒストグラムを求め、 $\pm 20\%$ 以内を正解、それ以外を誤検出と見なす。全フレームの個数に対し、誤検出となったフレームの個数が占める割合を gross error とし、テキストごとに gross error を求める。

(3) gross error の平均を求める

全テキストに対する gross error より、男女別の平均、性別を無視した全員の平均を求める。

3.1.2 使用データベースと実験条件

実験は 2 つの F0 推定法評価用データベースについて行う。使用するデータベースに関しては、以下を参考されたい。

なお、実験条件として、フレームシフトは 5(ms)、基本周波数の探索範囲は 40-800 (Hz) とする。また、はじめに調波数 N を変化させ評価を行う。この際、もう一つのパラメタである、重み関数の幅 b_w の値は、仮に 1 として固定する。 b_w に関する評価の際には、評価実験結果より決定された N の最適値を用いる。

DB1:阿竹のデータベース [4]

22 歳から 36 歳までの男女各 14 名による日本語短文章 30 テキストを収録。EGG 波形より判定した有声/無声ラベルが付加されている。16kHz サンプリング。総時間 36 分 25 秒（有声区間のみならば 21 分 32 秒）、総フレーム数 1,293,452。

DB2:Bagshow のデータベース

男女各 2 名による英語短文章 50 テキストを収録。20kHz サンプリング、12-bits。

3.2 実験結果およびパラメタの最適値の決定

各パラメタに対する実験結果を示し、結果より、パラメタのチューニングを行う。

3.2.1 調波成分の個数 N

N ごとのグロスエラーの推移を表したグラフを、図 6 に示す。 $N = 1$ から $N = 3$ の区間の変化に対しては精度の向上が見られ、 $N = 3$ における gross error の値は 0.77% (男性 0.94%、女性 0.61%) となった。 $N > 3$ の区間では、 N の増加に対して gross error の値は減少していくが、その変化はほとんど見られない。先にも述べたように、歌唱音声等の分析を考慮するならば、 N の値はできるだけ小さい方が望ましい。よって、3 がその最適値であると考えられる。

3.2.2 重み関数 $w_{LAG}(\tau)$ の鋭さを決定する b_w

b_w ごとのグロスエラーの推移を表したグラフを、図 7 に示す。 $b_w = 2$ 以降の区間で、著しく精度が低下した。また、 $b_w = 1$ 付近の値に対して、gross error の値は比較的小さく、安定している。この区間にて更に細かく b_w を指定して評価を行ったところ、 $b_w = 1.2$ の際に、最も gross error が小さくなった (0.77%、男性 0.93%、女性 0.61%)。よって、 $b_w = 1.2$ が最適値であると考えられる。

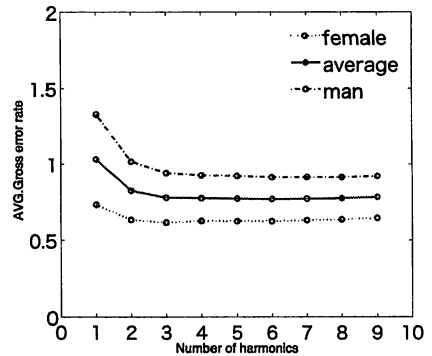


図 6 N の値の変化に対する gross error の推移
Fig. 6 Evaluation findings for N

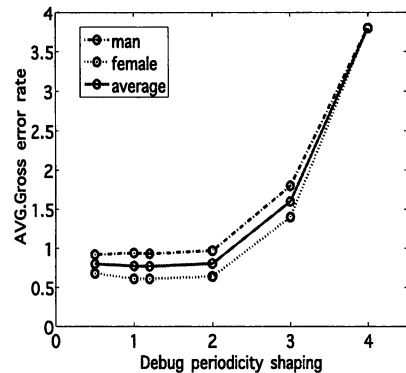


図 7 b_w の変化に対する gross error の推移
Fig. 7 Evaluation findings for b_w of $w_{LAG}(\tau)$

4. 性能評価

$N = 3$ を最適値と考え、チューニング後の TANDEM-STRAIGHT 法による F0 抽出結果精度と、従来法による精度と比較した。結果を以下の表にまとめる。従来法は、旧 STRAIGHT の基本周波数抽出法（既定値）[8] と YIN 法 [9] を対象とした。精度の比較には、前章の性能評価実験同様 gross error の平均値を用いる。

	DB1	DB2
proposed method	0.77%	0.69%
STRAIGHT (default extractor)	0.77%	2.8%
YIN	0.29%	1.4%

TANDEM-STRAIGHT を用いた F0 抽出法の DB1 に対する精度は、TEMPO 法と同程度であった。また、DB2 に対しては、より優れた結果を得ることができた。

5. まとめ

TANDEM とは、分析時刻に依存しないスペクトル推定を、ごく簡単なプロセスで行うことができる手法である。本

報告では、TANDEM に基づいて再構成された TANDEM-STRAIGHT において、新たに設計された TANDEM スペクトルと STRAIGHT スペクトルに基づく基本周波数抽出器の、基本周波数抽出精度に関する検討を行った。まずは、各種パラメータの変化が抽出精度に及ぼす影響を確かめるため、音声、EGG 波形の収録されたデータベースを用い、評価実験を行った。検討すべきパラメータは、周波数軸上において考慮する調波数を決定する N と、時間遅れ領域において F0 検出器を統合する際の、重み窓の鋭さを決定する b_w である。評価の際は、EGG から求めた F0 を真値として、音声の F0 との比をとり、完全一致 (100%) から $\pm 20\%$ 以上はなれている場合を gross error とみなす。この gross error となった回数が、全分析点中に占める割合を求める。評価の結果、各パラメータの最適値は、 $N = 3$ 、 $b_w = 1.2$ であることが分かった。次に、これらの最適値により TANDEM-STRAIGHT をチューニングし、従来法 (STRAIGHT の既定値、YIN 法) との比較評価を行った。その結果、通常発話音声の場合ならば、従来手法に劣らない精度を持つことがわかった。

謝 辞

本研究の一部は、科学研究費補助金 基盤 (A) 19200017、基盤 (B) 18300060 の支援を受けた。

文 献

- [1] 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 坂野秀樹, 入野俊夫. "TANDEM-STRAIGHT に基づく基本周波数の抽出と評価について". 信学技報 Vol.107, pp.259-264, 2007.
- [2] 河原英紀, 森勢将雅, 高橋徹, 西村竜一, 坂野秀樹, 入野俊夫. "TANDEM および STRAIGHT スペクトルに基づく基本周波数および非周期性の表現について". 日本音響学会 2008 年春季研究発表会講演論文集, pp.563-564, 2008-3.
- [3] 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 論文誌 D, Vol.J 90-D, No.12, pp.3265-3267, 2007.
- [4] 阿竹義徳, 入野俊夫, 河原英紀, 陸金林, 中村哲, 鹿野清宏. "調波成分の瞬時周波数を用いた基本周波数推定方法". 電子情報通信学会誌, D-II, J83-D-II, 11, pp.2077-2086, 2000.
- [5] 板橋秀一, 赤羽誠, 石川泰, 大河内正明, 粕谷英樹, 桑原尚夫, 田中和世, 新田恒雄, 矢頭隆, 渡辺隆夫. "音声工学". 森北出版, pp.82-83, 2005-2.
- [6] 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.
- [7] Hideki Kawahara, Ikuyo Masuda-Katsuse and Alain de Cheveigne. "Restructuring speech representations using a pitch-adaptive time-frequency smoothing and an instantaneous-frequency-based F0 extraction: Possible role of a repetitive structure in sounds". Speech Communication, 27, 3, pp.187-207 (1999).
- [8] 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", Proc. Eurospeech'99, Vol.76, pp. 2781-2784 (1999).
- [9] Alain de Cheveigne, Hideki Kawahara, YIN, "a fundamental frequency estimator for speech and music" Journal of the Acoustical Society of America, Vol.111, No.4, pp.1917-1930 (2002)
- [10] 森勢将雅, 高橋徹, 河原英紀, 入野俊夫: , 電子情報通信学会