

経験的モード分解による主構造抽出を介した 雑音環境下における音声信号の基本周波数推定

松田 徹也[†] 広瀬 啓吉[†] 峯松 信明[†]

[†] 東京大学大学院情報理工学系研究科

〒113-0033 東京都文京区本郷 7-3-1

E-mail: †{matsuda,hirose,mine}@gavo.t.u-tokyo.ac.jp

あらまし これまで提案された基本周波数推定手法の多くは、音声信号の線形性や定常性を短時間区間内において仮定した処理を行っていた。しかし、この仮定は、厳密に言えば事実と反するものであり、推定精度低下の原因ともなり得る。一方、経験的モード分解は、非線形・非定常な信号を扱うのに適した解析手法であるが、基本周波数に関わる信号成分までをも複数の信号に分解してしまう問題を抱えており、基本周波数推定への応用は困難であった。本研究では、遅れ時間空間においては信号のエネルギーが特定の帯域ごとに集中するため、先述の問題が解消され、経験的モード分解の解析結果と周波数帯域との対応付けが可能となる点に着目し、信号の周期性によって生じる自己相関関数中の主たる構造の経験的モード分解による抽出を介した雑音にロバストな基本周波数推定手法を提案する。また、雑音環境における基本周波数の推定実験を通して、提案手法のロバスト性を評価する。

キーワード 基本周波数推定, 雑音環境, 経験的モード分解, 自己相関関数

Robust estimation of fundamental frequencies of speech signals in noisy environments based on signal filtering by empirical mode decomposition

Tetsuya MATSUDA[†], Keikichi HIROSE[†], and Nobuaki MINEMATSU[†]

[†] Graduate School of Information Science and Technology, The University of Tokyo

7-3-1 Hongo, Bunkyo-ku, Tokyo, 113-8656, Japan

E-mail: †{matsuda,hirose,mine}@gavo.t.u-tokyo.ac.jp

Abstract A number of methods have already been developed for the estimation of fundamental frequencies of speech signals. However, almost all the methods observe the signals in a short time span and assume them being linear and stationary in the span, which is not correct strictly speaking. This situation limits precision of the estimation. While the empirical mode decomposition (EMD) is a method for signal analysis not assuming linear and stationary features of signals, it is rather difficult to apply it to the pitch estimation, since components corresponding to fundamental frequencies may spread to several functions after decomposition. This paper introduces a new method of pitch estimation based on EMD, where the problem is solved by applying EMD on the auto-correlation function of the signal. Since, in the lag-time domain, the signal energy concentrates to frequencies corresponding to signal structures, robust estimation of fundamental frequencies is realized by properly selecting a function from decomposed ones. Through experiments on pitch estimation in noisy environments, the robustness of the proposed method was shown.

Key words pitch estimation, noisy environment, empirical mode decomposition, autocorrelation

1. はじめに

音声信号からの基本周波数推定は、音声処理を行う上で欠か

すことの出来ない技術であり、雑音除去、音声認識と音声理解、韻律的特徴の解析やモデル化、低ビットレートでの音声符号化といった様々な応用分野において重要な役割を果たしている。

そのため、これまで数多くの基本周波数推定法が提案されてきた。しかし、雑音環境下や、周期性が低下した波形部分でも、安定に基本周波数を推定できる方法は未だ開発されておらず、その実現は依然として困難な課題とされている [1]。

基本周波数の推定方法には、自己相関関数法や、ケプストラム法など様々なものが存在する。こうした手法の多くは、短時間区間内において音声信号の線形性や定常性を仮定した解析を行っている。しかし、この仮定は事実と反しており、基本周波数の推定精度を低下させる原因となり得る。特に、感情音声、歌唱、母音の開始部や終了部等、声帯振動が不規則になる部分で顕著な問題となる [2]。

信号の定常性を仮定しない解析手法も存在する。ヒルベルト変換を用いた瞬時周波数の計算による方法がそのひとつである。しかし、この方法には瞬時周波数が意味を持つための条件が不明確であるという問題があった。そこで考案されたのが経験的モード分解 (EMD; Empirical Mode Decomposition) である [3]。EMD とは、信号を固有モード関数 (IMF; Intrinsic Mode Function) と呼ばれる狭帯域かつトレンド成分を持たない信号の和に分解する手法のことであり、IMF から得られる瞬時周波数と IMF の基本周波数はよく一致する。また、その分解方法から、非線形・非定常な信号の解析に適している。こうした特徴は信号処理を行う上で都合が良く、既に様々な応用が考案されている [4][5]。だが一方で、EMD はフーリエ変換のような基底関数を持たないため、IMF の帯域制御が難しいという欠点も抱えている。音声信号から基本周波数を推定する上で、これは大きな障害となる。しかし、直接 EMD を用いて基本周波数を推定するのが難しくとも、様々な工夫を凝らすことで、EMD の長所を音声信号からの基本周波数推定に生かすことは可能である。

EMD を用いた基本周波数推定手法として Huang らの提案した手法 [6] がある。この手法は、瞬時周波数ではなく瞬時エネルギーに着目し、声帯におけるパルスの発生間隔を推定することで基本周波数を計算する。しかしこの方法は、信号の線形性や定常性は仮定しないものの、雑音耐性の低さに起因したパルス位置の誤検出が多く、実用に耐えうるまでの精度には達していない [1]。改善策としては、瞬時周波数と組み合わせる方法や [7]、Wavelet 変換と組み合わせる方法 [8] などが検討されている。

一方、EMD を自己相関関数 [9] と組み合わせる手法 [1] も考案されている。この手法は、EMD を時間空間ではなく遅れ時間空間において適用する。遅れ時間空間では音声信号の帯域幅が安定するため、その結果として、帯域幅の安定した IMF が得られるようになる。これは、線形性・定常性を仮定しない雑音除去実現の可能性を示している。こうした時間空間以外においての EMD の利用はほとんど試みられておらず、研究の余地を残している。

本研究の目的は、雑音環境下における音声信号の基本周波数推定精度を向上させることである。そのために、音声信号の非線形性・非定常性を考慮した雑音除去を行えば推定精度が向上するという仮説のもと、遅れ時間空間において音声信号の帯域幅が安定する点に着目し、EMD を用いた雑音除去を試みる。さら

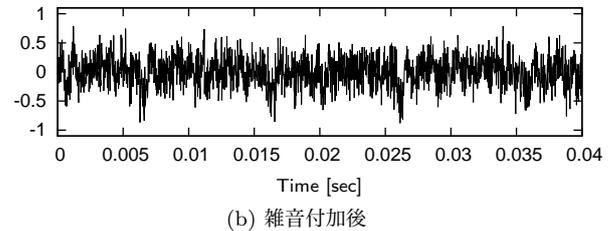
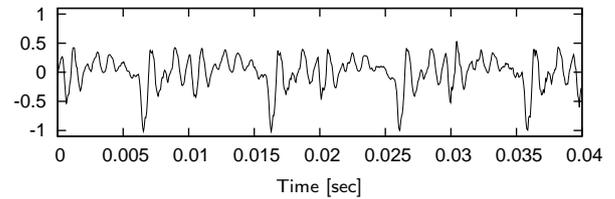


図 1 音声信号 (母音/a/, 男性, 基本周波数 101 [Hz], SN 比 -5 [dB], ガウス雑音)

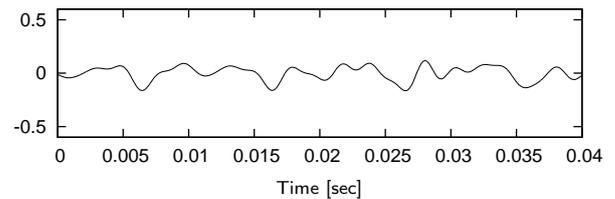


図 2 帯域通過処理の例 (幅 50-500 [Hz], 母音/a/, 男性, 基本周波数 101 [Hz], SN 比 -5 [dB], ガウス雑音)

に、その結果得られた音声信号の自己相関関数の主構造から、実際に基本周波数推定を行い、仮説の妥当性について検討を行う。

2. EMD を用いた基本周波数推定

まず、先行研究である EMD と自己相関関数を組み合わせて用いる基本周波数推定法について説明を行う。以降、図 1 の音声信号を解析対象として用いる。図 1(a) は男性話者、母音/a/、基本周波数 101 [Hz] の音声信号であり、図 1(b) はこれに SN 比が -5 [dB] となるようガウス雑音を付加した信号である。

2.1 正規化自己相関関数

この手法の特徴は、信号に対し直接 EMD を行うのではなく、信号の自己相関関数に対し間接的に EMD を用いる点である。こうして得られた IMF 中より、基本周波数に関わる信号成分である基本波を選び出すことで推定精度の向上を図る。この際、より基本波成分にエネルギーが集中するよう、いくつかの工夫を凝らす。

まず、自己相関関数を計算する前に、信号を帯域通過フィルタに通す。この処理は、信号の定常性と線形性を仮定しているため、周波数推定に際し誤差の要因ともなるが、目的の帯域以外の信号成分を取り除くことで、雑音による影響を低減させることにつながる。図 2 は、図 1(b) に 50 - 500 [Hz] の帯域通過フィルタをかけた結果である。

EMD の対象となる自己相関関数は式 (1) により計算される。このような自己相関関数を正規化自己相関関数 (NACF; Normalized Auto-Correlation Function) と呼ぶ。NACF は、

自己相関関数を式 (2) の様に定義される信号のエネルギー $e(\tau)$ で正規化したものである。

$$\text{NACF}(\tau) = \frac{\sum_{n=0}^{N-1-|\tau|} s(n)s(n+|\tau|)}{\sqrt{e(0)}\sqrt{e(|\tau|)}} \quad (1)$$

$$e(\tau) = \sum_{n=0}^{N-1-|\tau|} s(n+|\tau|)^2 \quad (2)$$

これにより、信号の周期性の有無だけに着目した自己相関関数の計算が可能になる。微弱なノイズ等も強調してしまうため、ノイズの影響を受けやすいという問題を抱えているが、周期性の強調という点では極めて高い効果をもたらす。

2.2 経験的モード分解

EMD は、信号 $s(t)$ を式 (3) のように IMF と呼ばれる狭帯域かつトレンド成分を持たない信号 $c_n(t)$ とトレンド成分 $r(t)$ に分解する。

$$s(t) = \sum_{n=0}^N c_n(t) + r(t) \quad (3)$$

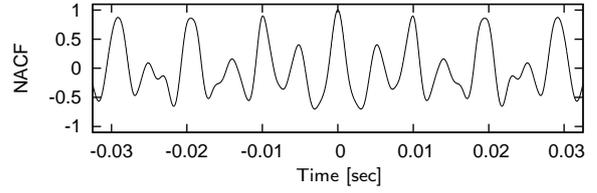
このとき、 $c_n(t)$ は信号を構成する信号成分の周波数帯域に対応し、添数の小さいものほど高周波数帯域の信号成分となる。分解を行う際、EMD は、信号が狭帯域かつトレンド成分を持たないことの基準を、それぞれ、信号中の極の数が零交差数と同じであるか 1 つしか変わらないこと、信号中の極大値を結ぶエンベロープと極小値を結ぶエンベロープの平均値が任意の点において 0 であることと定義している。

信号中から、IMF を抽出する手段は複数存在するが、最も一般的なのは以下のアルゴリズムである。

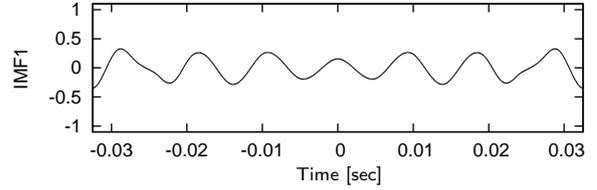
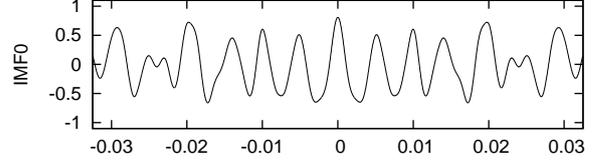
1. $n = 0, k = 0$ とし、 $h_{nk}(t) = s(t)$ と置く。
2. $h_{nk}(t)$ 中の極大値を結ぶエンベロープ $l(t)$ と極小値を結ぶエンベロープ $m(t)$ を 3 次スプライン補間により求め、 $h_{nk}(t)$ のトレンド成分 $\mu(t)$ を $\mu(t) = [l(t) + m(t)]/2$ とし計算する。
3. $h_{n(k+1)} = h_{nk} - \mu(t)$ とし、トレンド成分を取り除く。
4. 先述した狭帯域かつトレンド成分を持たない信号の条件を $h_{n(k+1)}(t)$ が満たしていれば、 $c_n(t) = h_{n(k+1)}(t)$ とする。条件を満たしていなければ $k = k + 1$ とし、ステップ 2 に戻る。
5. $h_{(n+1)0}(t) = h_{n0}(t) - c_n(t)$ とする。
6. $h_{(n+1)0}(t)$ 中からエンベロープの計算に必要な極が無くなったか、 $h_{(n+1)0}(t)$ のエネルギーが適当な閾値を下回った際に、 $r(t) = h_{(n+1)0}(t)$ とし、処理を打ち切る。それ以外の場合は、 $n = n + 1, k = 0$ とし、ステップ 2 に戻る。

ステップ 2, 3 で行う処理をシフト処理という。実際問題として、信号中の極大値を結ぶエンベロープと極小値を結ぶエンベロープの平均値が任意の点において 0 となるまでシフト処理を繰り返すのは難しいため、式 (4) により、シフト処理によるトレンド成分除去の効果を計算し、その値を元にシフト処理の打ち切りを決める。

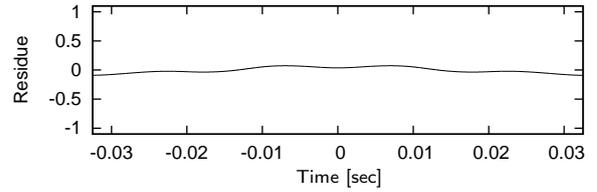
$$\text{SD} = \sum_{t=0}^T \left[\frac{|h_{n(k+1)}(t) - h_{nk}(t)|^2}{h_{n(k+1)}^2(t)} \right] \quad (4)$$



(a) 解析対象となる正規化自己相関関数



(b) 固有モード関数



(c) トレンド成分

図3 経験的モード分解の例 (母音/a/, 男性, 基本周波数 101 [Hz], SN 比 -5 [dB], ガウス雑音)

ここでは、シフト処理の打ち切りを SD が 0.3 以下となったときとする。

このように、EMD は信号の線形性や定常性を仮定することなく解析を行うため、非線形・非定常な信号成分のエネルギーが誤差の要因となることが起き難い。しかし、フーリエ変換などと異なり、EMD には基底関数が存在しないため、基本波成分が複数の IMF 中に分散してしまう可能性がある。信号の自己相関関数に対して EMD を用いることで、この問題は低減される。

2.3 基本波成分の抽出

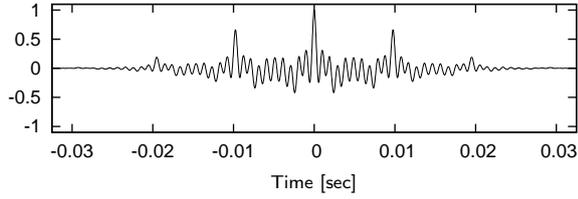
基本波成分の抽出には、式 (5) で定義される重み付け自己相関関数 (WAC; Weighted Auto-Correlation function) [11] を用いる。

$$\text{WAC}(\tau) = \frac{\sum_{n=0}^{N-1-|\tau|} s(n)s(n+|\tau|)}{\text{AMDF}(\tau) + \epsilon} \quad (5)$$

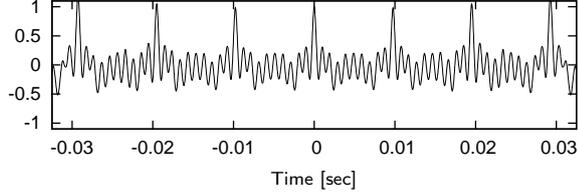
$$\text{AMDF}(\tau) = \sum_{n=0}^{N-1-|\tau|} |s(n) - s(n+|\tau|)| \quad (6)$$

WAC の定義に用いられている式 (6) の関数は、平均振幅差関数 (AMDF; Average Magnitude Difference Function) と呼ばれる、信号の周期性の強さを表す関数である。ただし、自己相関関数とは雑音からの影響の受け方が異なるため、この 2 つを掛け合わせると、通常の自己相関関数に比べ、雑音の影響は小さくなり、信号の周期性は逆に強調される。

基本波成分抽出は次の手順で行われる。まず、音声信号から



(a) 窓関数の自己相関関数で除算を行う前



(b) 窓関数の自己相関関数で除算を行った後

図 4 窓関数の自己相関関数で除算した自己相関関数の例 (母音/a/, 男性, 基本周波数 101 [Hz], 雑音付加無し)

NACF を計算し, EMD によって IMF に分解した後, IMF の基本周波数を波形のピーク間隔からそれぞれ求める. そして, WAC より求めた基本周波数と IMF の基本周波数を比較する. IMF は NACF 中の特定の帯域に対応した信号成分であるため [10], WAC の基本周波数と最も近い基本周波数を持つ IMF を選び出すことで NACF の基本波を得ることができる. こうして得られた基本波の基本周波数は音声信号の基本周波数とよく一致する.

図 3 は NACF に対して EMD を行った例である. 図 3(a) は図 2 より求めた NACF であり, これに対し EMD を行うと, IMF (図 3(b)) とトレンド成分 (図 3(c)) が得られる. IMF の添数は算出された順を表す. この例の場合, IMF1 が基本波として選び出される.

3. 主構造抽出を介した基本周波数推定

このように, 音声信号に対し EMD を間接的に用いることで, IMF の帯域制御の難しさが軽減され, 信号の非定常性や非線形性に対応した基本周波数推定が実現できることがわかった.

ところで, EMD より得られる IMF は添数が小さいほど SN 比が下がることがガウス雑音を用いた数値的な検証により定量的に確認されている. この性質を利用し, 雑音を含む自己相関関数中から SN 比の高い IMF だけを取り出すことで雑音除去を行い, 基本周波数推定精度を向上させることを考える.

3.1 窓関数の自己相関関数で除算した自己相関関数

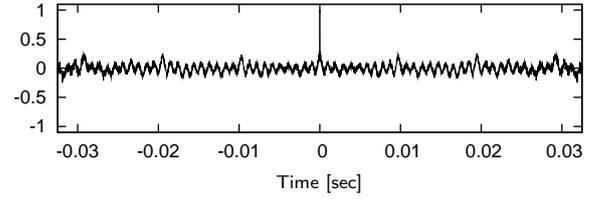
まず, 解析の対象となる自己相関関数について説明する.

信号の自己相関関数を計算するに際し, 式 (7) で定義されるハン窓を用いて式 (8) のように窓掛けを行う.

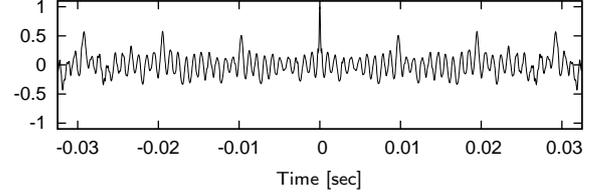
$$w(t) = \begin{cases} \frac{1}{2} - \frac{1}{2} \cos \frac{2\pi t}{T} & (0 \leq t \leq T) \\ 0 & (t < 0) \text{ or } (T < t) \end{cases} \quad (7)$$

$$a(t) = \left(x(t_{mid} - \frac{1}{2}T + t) - \mu_x \right) w(t) \quad (8)$$

t_{mid} は窓の中心位置, T は窓幅, μ_x は区間 T における $x(t)$ の



(a) 解析対象となる自己相関関数



(b) 主構造抽出を行った結果

図 5 EMD による主構造抽出の例 (母音/a/, 男性, 基本周波数 101 [Hz], SN 比 -5 [dB], ガウス雑音)

平均値である. $a(t)$ の自己相関関数 $R_{aa}(\tau)$ は式 (9) により求めることができるが, この時ある問題が生じる.

$$R_{aa}(\tau) = \frac{\sum_t a(t)a(t+\tau)}{\sum_t a^2(t)} \quad (9)$$

それは, 窓関数を用いたことによる低域成分の減衰に起因する, 相対的な高域成分の強調である. これは自己相関関数を用いた周期性の検出に共通する問題であり, 倍ピッチエラーと呼ばれる推定誤差をもたらす. そこで, この問題による影響を低減させるため, 式 (10) から求まる窓関数の自己相関関数を用いて, 式 (11) のように自己相関関数の低域成分の強調を行う [12].

$$R_{ww}(\tau) = \frac{\sum_t w(t)w(t+\tau)}{\sum_t w^2(t)} \quad (10)$$

$$R_{xx}(\tau) = \frac{R_{aa}(\tau)}{R_{ww}(\tau)} \quad (11)$$

図 4 は共に図 1(a) の自己相関関数であるが, 式 (9) より求めた自己相関関数 (図 4(a)) と比べ, 式 (11) より求めた自己相関関数 (図 4(b)) の方が低域の周期性が強調されていることが見て取れる. ただし, それぞれ $R_{aa}(0)$, $R_{xx}(0)$ を用いて正規化を行っている.

3.2 EMD による主構造抽出

EMD より得られる IMF は信号を構成する信号成分の周波数帯域におよそ対応することはすでに説明した. このとき, 添数が小さい IMF ほど SN 比が下がることが数値的な検証により定量的に確認されている [13]. この性質を利用し, 雑音を含む自己相関関数から SN 比の高い IMF だけを取り出すことで雑音除去を行う. これより, 自己相関関数中の音声信号によって構成される主たる構造を抽出することができる.

まず, 雑音を含まない理想的な信号を y , 雑音を含む信号中から推定した y を \tilde{y} としたとき, 雑音除去とは, 式 (12) の平均二乗誤差を最小とすることであるといえる.

$$\text{MSE}(y, \tilde{y}) = \frac{1}{K} \sum_k [y(k) - \tilde{y}(k)]^2 \quad (12)$$

EMD を用いた雑音除去における \tilde{y} は、添数が小さいほど SN 比が下がるという仮定より、式 (13) のように定義される。

$$\tilde{y}_i(k) = \sum_{n=i}^N c_n(k) + r(k) \quad (13)$$

しかし、 y を得ることは出来ないため、式 (12) を計算することは不可能である。そこで、 y も推定値とし、式 (14) のように連続平均二乗誤差と呼ばれる距離尺度を定義する。

$$\begin{aligned} \text{CMSE}(\tilde{y}_i, \tilde{y}_{i+1}) &= \frac{1}{K} \sum_k [\tilde{y}_i(k) - \tilde{y}_{i+1}(k)]^2 \\ &= \frac{1}{K} \sum_k c_i^2(k) \end{aligned} \quad (14)$$

CMSE を最小とする i を求めることは、雑音が支配的な添数の小さい IMF と、音声信号が支配的な添数の大きい IMF の境界に位置する IMF を探すことに相当する。また、雑音除去を行う際に CMSE と \tilde{y} の比を計算し、CMSE が \tilde{y} の 25 [%] を上回る場合は抽出処理を中断する。これは、無音区間における誤った信号成分の除去を防ぐための処理である。

以上の処理を式 (11) より求めた自己相関関数に対して行うことで自己相関関数の主構造が得られる。そのピーク間隔より求めた基本周波数は音声信号の基本周波数とよく一致する。図 5 に主構造抽出の例を示す。図 5(a) は式 (11) より求めた図 1(b) の自己相関関数であり、図 5(b) は図 5(a) から主構造 $\tilde{R}(\tau)$ を抽出した結果である。ただし、それぞれ $R_{xx}(0)$, $\tilde{R}(0)$ で正規化を行っている。主構造抽出を行ったことで、音声信号のみによって形作られる図 4(b) の自己相関関数に近い形状が得られることがわかる。

4. 基本周波数推定精度の評価実験

基本周波数の推定実験を通して提案手法のロバスト性を評価する。また、短時間区間内において音声信号の線形性や定常性を仮定した手法や先行研究との比較を行い、提案手法の優位性について考察する。

4.1 実験条件

基本周波数の推定には、提案手法 (Proposed) の他に、基準となる手法として、式 (11) の自己相関関数 (Default)、短時間区間内において音声信号の線形性や定常性を仮定した手法として、Wavelet 変換による雑音除去 [14] を行った後に手法 Default を用いる方法 (Wavelet)、先行研究 (Previous) [1] の 3 つを用いた。

基本周波数の推定は、窓幅 40 [ms] のハン窓をフレームシフト 5 [ms] で使用し、探索範囲を 75-500[Hz] として行った。

音声試料には、以下のデータベースを利用した。

阿竹らのデータベース [15]

成人男女各 14 名による日本語短文書 30 テキストを収録。EGG 波形と EGG 波形より判定された有声/無声ラベルが付属。16 [kHz] / 16 [bit] サンプリング。

基本周波数推定には、データベースから男女各 2 名ずつ (M01, M02, F01, F02) を選び出し、その有声区間のみを用いた。このとき、音声信号は 8 [kHz] にダウンサンプリングした。

雑音を含む音声信号は、Octave の randn 関数を用いて生成

したガウス雑音を、上記の音声試料に、SN 比が、5[dB], 0 [dB], -5 [dB] となるように加えることで人工的に作成した。SN 比の基準となる信号のパワーは、信号全体の平均パワーの 1/30 より大きくなる部分の平均パワーとした。

正答とする基本周波数は、Electro Glottal Graph(EGG) より計算した。EGG の微分信号のピーク間隔の逆数をその区間の基本周波数として求めた。

推定精度の評価指標には、Gross Error Rate を用いた。正答との相対誤差が 5 [%] 以上のものを Gross Error とした。

4.2 実験結果

Default, Proposed, Wavelet, Previous の 4 つの手法を用いて、用意したすべての音声試料について、それぞれ基本周波数の推定を行ったところ、表 1, 図 6 の結果となった。Gross Error Rate は、男性のみ (Male)、女性のみ (Female)、合計 (Total) のそれぞれで計算した。このとき、基本周波数の推定を行ったフレーム数は、それぞれ、17467, 18554, 36021 となった。

Default と Proposed の Gross Error Rate を比較したところ、SN 比が 0 [dB] 以下の場合では、いずれについても Proposed の方が下回った。さらに危険率 1 [%] で片側検定を行ったところ、SN 比が -5 [dB] の場合では、いずれについても有意差が確認された。また、Wavelet と Proposed について比較したところ、SN 比が -5 [dB] の場合では、いずれについても Proposed の方が下回り、有意差も確認された。最後に、Previous と Proposed について比較したところ、SN 比が -5 [dB] の Male については、Previous の方が Proposed を有意差を持って下回ったが、SN 比が -5 [dB] の Female については、逆の結果となった。

4.3 考察

まず、低 SN 比時において、Proposed の基本周波数推定精度が Default を上回っていることから、提案手法を用いたことで基本周波数推定のロバスト性が向上していることがいえる。

また、Wavelet に対しても Proposed の推定精度が同様に上回っていることから、基本周波数推定を行う際に音声信号の非線形性・非定常性を考慮したことで推定精度の向上が得られたと考えられる。

ところで、Previous の推定精度の評価結果をみると、Male については Proposed 以上のロバスト性が得られているのに対し、Female については Default を下回ってしまっている。この傾向は WAC の倍ピッチエラーに起因する基本波成分の選択ミスによって生じていると考えられる。提案手法ではこうした問題が起きにくいと、ロバスト性については先行研究を上回っているといえる。

5. まとめ

これまで提案された基本周波数推定手法の多くは、音声信号の線形性や定常性を短時間区間内において仮定した処理を行っていた。しかし、この仮定は事実と反するものであり、推定精度低下の原因ともなり得る。

そこで本研究では、非線形・非定常な信号を扱うのに適した解析手法である EMD に着目し、EMD による主構造抽出を介した雑音にロバストな基本周波数推定手法の提案を行った。

表 1 基本周波数推定実験の結果 (Gross Error Rate [%])

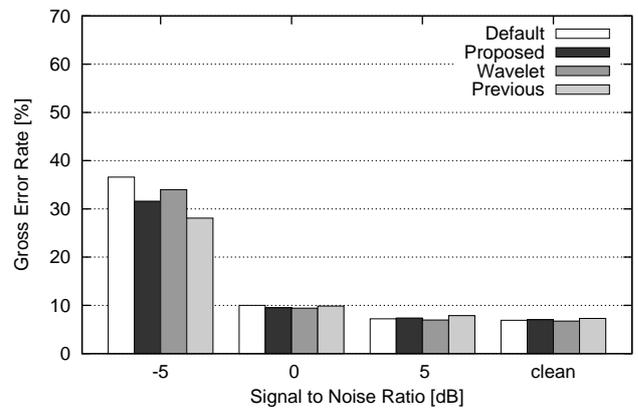
SNR [dB]		-5	0	5	clean
Male	Default	36.6	10.0	7.20	6.90
	Proposed	31.6	9.48	7.36	7.01
	Wavelet	34.0	9.41	6.96	6.72
	Previous	28.1	9.86	7.85	7.28
Female	Default	34.8	9.28	7.35	7.30
	Proposed	30.0	8.97	7.37	7.31
	Wavelet	32.9	9.05	7.36	7.32
	Previous	60.3	20.9	12.2	10.0
Total	Default	35.7	9.64	7.28	7.10
	Proposed	30.7	9.22	7.37	7.17
	Wavelet	33.4	9.23	7.16	7.03
	Previous	44.7	15.6	10.1	8.69

さらに、提案手法のロバスト性を検証するため、男性 2 名、女性 2 名の発話文章計 120 テキスト分の音声信号と EGG データを用いて、人工的にガウス雑音を加えながら、基本周波数の推定実験を行ったところ、提案手法を用いたことによる推定精度の向上が確認された。

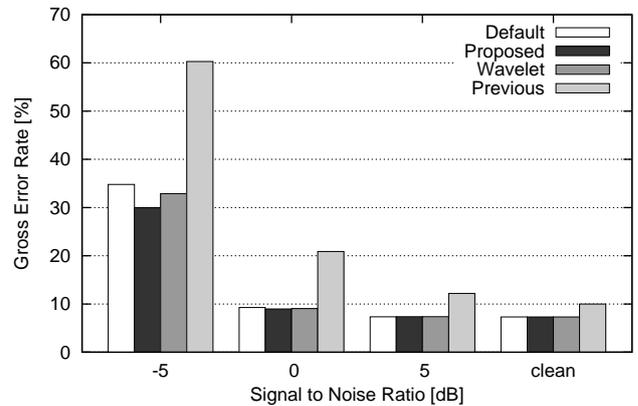
謝辞 基本周波数推定の精度評価には、和歌山大学システム工学部の河原英紀博士のご厚意により研究用 EGG データベースを使用させていただいた。

文 献

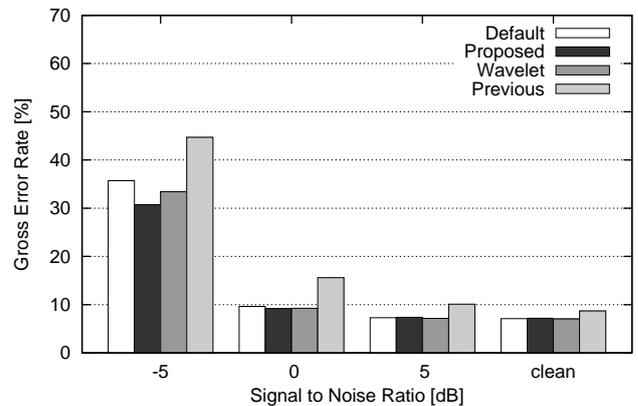
- [1] Molla, Md. K. I., Hirose, K., Minematsu, N. and Hasan, Md. K., "Pitch estimation of noisy speech signals using empirical mode decomposition", Interspeech 2007, pp.1645-1648, 2007.
- [2] Scalart, P., Filho, J. V., "Speech enhancement based on a priori signal to noise estimation.", ICASSP-96. Conference Proceedings., vol. 2, pp. 629-632, 1996.
- [3] Huang, N. E. et. al., "The empirical mode decomposition and Hilbert spectrum for nonlinear and non-stationary time series analysis", Proc. Roy. Soc. London A, Vol.454, pp.903-995, 1998.
- [4] Yu, D., Cheng, J., Yang, Y., "Application of EMD method and Hilbert spectrum to the fault diagnosis of roller bearings", Elsevier. Mechanical Systems and Signal Processing, 2005.
- [5] Adriano, O. A., Peter, K. and Slawomir, J. N., "The application of the Hilbert spectrum to the analysis of electromyographic signals", Elsevier. Information Sciences, 2008.
- [6] Yang, Z., Huang, D. and Yang, L., "A novel pitch period detection algorithm based on Hilbert-Huang transform", LNCS 3338, pp.586-593, Sinobiometrics, 2004.
- [7] Weiping, H., Yaling, L., Minghui, D. and Xiuxin, W., "A Novel Pitch Period Detection Algorithm Bases on HHT with Application to Normal and Pathological Voice", IEEE-EMBS 2005., pp.4541 - 4544, 2005.
- [8] Zhou, C., Yang, L., Liu, Y., Yang, Z., "A novel method for computing the Hilbert transform with Haar multiresolution approximation", Elsevier. Journal of Computational and Applied Mathematics, 2008.
- [9] Talkin, D., "A robust algorithm for pitch tracking", Elsevier Science. Speech Coding and Synthesis, 1995.
- [10] Flandrin, P., Rilling, G., and Goncalves, P., "Empirical mode decomposition as a filter bank", IEEE signal processing letters, Vol. 11, No. 2, pp.112-114, 2004.
- [11] Shimamura, T., Kobayashi, H., "Weighted autocorrelation for pitch extraction of noisy speech", IEEE Trans. Speech and Audio Proc., 9(7), pp.727-730, 2001.
- [12] Boersma, P., "Accurate short-term analysis of the funda-



(a) Gross Error Rate の推移 (男性のみ)



(b) Gross Error Rate の推移 (女性のみ)



(c) Gross Error Rate の推移 (合計)

図 6 基本周波数推定実験の結果

- mental frequency and the harmonics-to-noise ratio of a sampled sound", Proc. Institute of Phonetic Sciences 17, pp.97-110, 1993.
- [13] Boudraa, A. O., Cexus, J. C., "EMD-Based Signal Filtering" Instrumentation and Measurement, IEEE Transactions on Vol. 56, Issue 6, pp.2196-2202, 2007.
 - [14] Qiang, F., Eric A. W., "A novel speech enhancement system based on wavelet denoising.", 2003.
 - [15] "調波成分の瞬時周波数を用いた基本周波数推定方法", 阿竹 義徳, 入野 俊夫, 河原 英紀, 陸 金林, 中村 哲, 鹿野 清宏, 電子情報通信学会論文誌, J83-D-II, 11, pp.2077-2086, 2000.