Vol.2016-SLP-110 No.5
2016/2/6

基本波検出に基づくF0推定法の耐雑音性向上

森勢 将雅[1,a)]

概要：本稿では，筆者らが 2010 年に提案した基本波検出に基づく基本周波数（F0）推定法の耐雑音性向上手法について述べる．2010 年に提案した F0 推定法は，周期信号の調波構造における基本波を低域通過フィルタにより抽出し，基本波の周波数を求める．F0 が未知であるため，カットオフ周波数の異なる複数の低域通過フィルタを用意し，各フィルタにより処理された信号から F0 候補と信頼度を求め，全ての候補中最も信頼できる候補を選択していた．基本波検出に基づく方法は，低域に雑音が混入する環境では充分な SNR の確保が困難であるため，高 SNR 環境で収録された音声を対象としていた．提案法では，滑らかな F0 軌跡を描くよう候補を再選択するアルゴリズム，および推定結果に対し瞬時周波数により結果を補正する処理を導入することで雑音に対する頑健性を向上させる．本稿では，耐雑音性向上手法について述べ，耐雑音性に限定した評価から提案法が期待通り動作することを示す．

キーワード：音声分析，基本周波数，基本波，瞬時周波数，耐雑音性

Improvement of noise robustness in the F0 estimator based on fundamental component extraction

Morise Masanori[1,a)]

Abstract: This article represents an algorithm for improving the noise robustness in the fundamental frequency (F0) estimator based on the fundamental component extraction. The conventionally proposed estimator requires the low-pass filter for extracting fundamental component from the periodic signal. Since F0 is unknown, many filters with difference cutoff frequencies are used, and then F0 candidates and their reliabilities are obtained. The estimated F0 is selected by their reliabilities. This estimator has required high-SNR speech because it depends on the SNR in lower frequency band. In this research, we introduce an algorithm for compensating for the estimated result by instantaneous frequency. This compensation can improve the noise robustness. This article shows the algorithm and carried out an evaluation in the noise robustness. The effectiveness of the proposed algorithm was discussed based on the evaluation results.

Keywords: Speech analysis, fundamental frequency, fundamental component, instantaneous frequency, noise robustness

1. はじめに

音声情報処理において基本周波数 (F0) とスペクトル包絡はもっとも基礎的なパラメータであり，高精度な推定法は，例えば音声分析合成 [1] や歌声合成をはじめとする音声分析が必要な分野に恩恵を与える．音声の F0 は，声帯振動が生じる周期のうち最も短い周期（基本周期）の逆数として定義される．F0 推定に関しては，実音声の F0 は時間とともに変化することに加え，毎回の声帯振動も一定ではなく，収録環境に起因する雑音が含まれるなど，推定を困難にする要因が複数存在する．様々な要因に対して頑健に推定可能であることが必要となるが，万能な方法は提案されていないのが現状である．

実環境で音声認識を行うスマートホンアプリケーション

[1] 山梨大学
University of Yamanashi, 4-3-11, Takeda, Kofu, 400–8511, Japan
[a)] mmorise@yamanashi.ac.jp

や，個人が自分の声を利用して歌唱合成を行う UTAU[*1] などの歌声合成システムは，実環境で収録された雑音を含む音声から高精度な F0 を推定する技術を必要としている．筆者らは，主に防音室やスタジオレコーディングで収録された音声を対象に高速かつ高精度な F0 推定を行う方法 [2] を提案してきた．この方法には，低域に雑音が存在する環境，すなわち，一般的な室内で収録された音声に対して高精度な F0 を推定することが難しいという問題がある．計算速度を犠牲にすることで高精度な推定を行う方法は存在する [3] が，実用性の観点からは実時間で動作する程度の計算コストであることが望ましいといえる．本研究では，文献 [2] で提案された方法をベースに耐雑音性を向上させるための後処理と補正を導入し，低い計算コストと高い精度を両立する F0 推定法の確立を目指す．

2. F0 推定法の関連研究

F0 推定法の歴史は古く，時間波形における周期性に着目し相関を用いる方法 [4] や，Cepstrum [5], [6] を代表とするパワースペクトルの特徴に着目した方法などが提案されている．時間波形の周期性に着目した方法では，YIN [7] が広く利用されており，2014 年には，YIN に改良を加えた pYIN [8] が提案されている．パワースペクトルの特徴を用いた方法では，SWIPE′ が高精度な方法として提案されている [9]．筆者らが提案した基本波検出に基づく方法 [2]（以下では基本波検出法と呼称する）は，基本波フィルタリングによる方法 [10] をベースにしている．また，本稿で課題としている耐雑音性に特化した F0 推定法が検討されており [11], [12]，SNR が 5 dB 程度でも高い精度で推定可能であることが示されている．

基本波検出法は，音声が高 SNR であれば，計算時間を数 10 分の 1 に圧縮しつつ，state-of-the-art となるいくつかの方法と比較して遜色ない推定精度を達成可能であることがすでに示されている [2]．一方，基本波が存在する低域には空調雑音などの低域の雑音が存在するため，一般的な室内などの環境で収録された音声からの F0 推定には不向きという問題があった．本稿では，この問題に対処するため，基本波検出法で推定された F0 に対する後処理と瞬時周波数を用いた補正を導入することで，計算速度の大幅は増加を避けつつ推定精度を向上させることを目指す．

3. 基本波検出法のアイディアと問題点

基本波検出法のアイディアは文献 [2] に示されているため，本稿では概要について述べる．基本波検出法は，以下の 3 ステップにより F0 軌跡を推定する．

(1) 複数の低域通過フィルタによるフィルタリング
(2) F0 候補と信頼性の計算
(3) F0 候補から最終的な F0 の選択

[*1] http://utau2008.web.fc2.com/

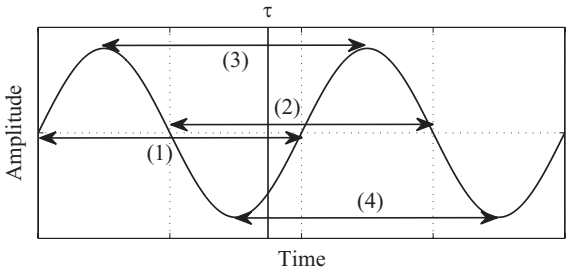

図 1 時刻 τ において F0 を計算するために用いる 4 つの間隔

3.1 ステップ 1 : 複数の低域通過フィルタによるフィルタリング

ステップ 1 では，複数のカットオフ周波数を持つ複数の低域通過フィルタにより入力信号を処理する．適切な低域通過フィルタを用いることで基本波のみを取り出すことが可能であるが，F0 が未知であることから基本波のみを含むようカットオフ周波数を設定することはできない．カットオフ周波数の異なる複数の低域通過により信号を処理し，それぞれについて F0 候補と信頼度を求めることでこの問題点に対処する．現状では，1 オクターブにつき 2 個のフィルタを設定することで，充分な精度を達成可能なことが示されている．

3.2 ステップ 2 : F0 候補と信頼性の計算

ステップ 2 では，ステップ 1 で N 個のフィルタにより処理された N 個の信号に対し，図 1 に示される 4 つの間隔を計算する．図中の τ を任意の時刻毎に与えることで，各時刻の F0 を計算可能となる．フィルタ処理後の信号が基本波のみを含む場合処理後の信号は正弦波となるため，4 つの間隔は全て同一の値を示す．よって，4 つの間隔の平均の逆数が F0 候補，標準偏差が F0 の信頼性となる．標準偏差が小さいほど，その F0 候補の信頼性が高いことを意味する．

3.3 ステップ 3 : F0 候補から最終的な F0 の選択

ステップ 3 では，各時刻について得られた N 個の F0 候補と信頼度から，最も信頼度の高い（標準偏差の小さい）候補を最終的な F0 として選択する．本手法は，原理的に真値の倍や半分を誤推定するエラーは発生しにくいが，F0 より低い周波数に雑音が存在して SNR が低下する場合，原理的に対処が困難である．雑音による影響が少ない場合でも，F0 の時間変化が速い場合など，特定のフレームにおいて本来の F0 と全く異なる値を誤推定する可能性がある．全候補のうち信頼度にのみ基づいて決定するのではなく，前後の候補を用いて滑らかな F0 軌跡となるよう補正することで，この問題に対処する．その後，瞬時周波数を

用いて F0 を修正することで，耐雑音性の向上を試みる．

4. 提案する耐雑音性向上法

本稿では，全 F0 候補から滑らかな F0 軌跡を得るための F0 の再選択を行い，その後瞬時周波数による補正を行うことで耐雑音性を向上させる方法を示す．F0 の再選択については，声帯振動が周期的に生じる音声において，基本波成分は短時間で大きく跳躍せず滑らかに遷移する，という仮説に基づく．なお，以下で説明する方法は，筆者が公開している音声分析合成方式 WORLD[*2] [13], [14], [15] の実装に準ずる．WORLD では，基本波検出法と F0 の再選択を合わせた方法を DIO と呼称し，瞬時周波数による補正を StoneMask とし別関数で実装している．これは，防音室やレコーディングスタジオなどの特殊な環境で収録された音声の場合，瞬時周波数による補正を行わずとも DIO のみで充分な精度を達成できることに起因する．StoneMask は，各時刻の結果を補正するために瞬時周波数を計算するため，計算コスト面においてはフレーム単位での FFT などの演算を行わない DIO よりも不利であるが，ノート PC 程度の環境で実時間処理が可能であることは確認済みである．

4.1 F0 候補から滑らかな軌跡を得るための後処理

F0 候補の再選定を行う後処理は，以下の 4 つのステップから構成される．

(1) 前後の時刻の F0 に基づく有声無声区間の分離
(2) 有声区間の調整
(3) 候補の再選定（前向き）
(4) 候補の再選定（後向き）

4.1.1 ステップ 1：前時刻の F0 に基づく有声無声区間の分離

ステップ 1 では，F0 は短時間で大きく跳躍しない，というルールに基づいて有声区間・無声区間の判定を行う．具体的に，n 番目の F0 が $n-1$ 番目の F0 と比較して何%変化しているかを計算し，それが閾値以上であれば無声音と判定する処理を全ての時刻の F0 について計算する．ステップ 1 により，F0 が跳躍している区間の F0 は無声と判定される．なお，実験に用いるプログラムでは，この閾値を前時刻の F0 の 10%としている．

4.1.2 ステップ 2：有声区間の調整

ステップ 1 により有声無声区間の判定がなされるが，ステップ 2 ではこの結果に基づいて有声無声区間の調整を行う．有声区間は声帯振動が連続して生じている区間であることから，数 ms 程度の有声区間は声帯振動が連続して発生している区間とは言い難い．ステップ 2 では，この性質に着目し，連続して有声音と判定された区間の長さを計算し，その長さが閾値を下回る区間を無声区間と修正する処理を行う．プログラムでは，F0 の下限から求めた基本周期の倍の区間を閾値とする．

ここまでの処理により，ある程度連続して滑らかな F0 軌跡が得られることとなるが，有声区間において局所的に生じた推定誤差については，その前後の時刻を巻き込んで一定区間が無声区間となる．また，有声音は始まりと終わりにおいて声帯振動が不安定になるため，これらの区間も無声区間とみなされることがある．

4.1.3 ステップ 3, 4：候補の再選定

ステップ 2 において無声区間と判定された区間には，局所的な誤推定が原因で無声区間と判定された，本来有声音の区間も存在する．ステップ 3 では，ステップ 2 で有声区間と判定された F0，および基本波検出法により得られた全 F0 候補を用いて，F0 の連続性を加味して候補の再選定を行う．ステップ 4 はステップ 3 とほぼ同様であるため，ここで併せて説明する．

図 2 は，$n+1$ 番目の F0 を再選定を行う際の基準値を決めるイメージ図を示す．F0 軌跡は，瞬時に跳躍しないが，ビブラートのような歌唱法では正弦波的な高速な振動を含む．つまり，前時刻の F0 からの範囲で探索するのではなく，前々時刻と前時刻から変化量も含めて探索範囲を決定することが望ましいといえる．提案法では，n 番目と $n-1$ 番目の F0 を利用して $n+1$ 番目の F0 の基準値を求める．図 2 における基準値 f は，以下の式で与えられる．

$$f = \frac{3f(n) - f(n-1)}{2}. \tag{1}$$

f は，直前の値と $n-1$ と n 番目から線形補間により与えた $n+1$ 番目の値の平均としている．これは，F0 には雑音に起因する揺らぎが入るため，直前の値に対する重みを与えることで，雑音の影響を軽減する狙いがある．再選定は，基本波検出法により得られた全候補から f に最も近い候補を選定することで行われる．ただし，F0 は短時間で跳躍しないというルールに則り，$f \pm a$ の範囲に存在しない候補は選択されず，範囲に候補が無い場合は無声音と判定する．a の値は，ステップ 1 と同様に基準値 f の 10%としている．

ステップ 3 では，推定値を過去の時刻から未来の時刻に向かって再選定し，ステップ 4 では，ステップ 3 とは逆に未来の情報を使い過去の F0 を再選定する．この両ステップは有声区間毎に行い，各有声区間の起点は有声区間の中央時刻とする．ステップ 2 で無声区間と判定された有声音の開始・終了区間は，これらのステップで滑らかな F0 軌跡として再選定され，同時に有声音・無声音の判定も同時に行われる．こうして得られた滑らかな F0 軌跡について，次節で述べる瞬時周波数に基づく補正を行うことで，各時刻における F0 の正確性を向上させる．

[*2] http://ml.cs.yamanashi.ac.jp/world/

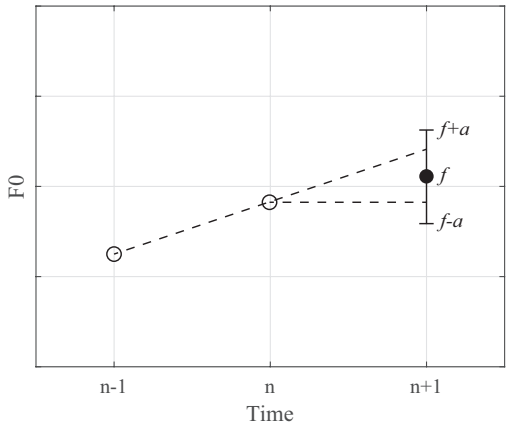

図 2 $n+1$ 番目の値を決定する方法．$n-1$ 点と n 点の F0 から中心となる周波数を決定し，そこから特定の幅に存在 F0 のうち最も中心周波数に近い F0 を新たな候補とする．

4.2 瞬時周波数による F0 軌跡の再推定

DIO により得られた F0 軌跡には雑音量に依存した誤差が混在するため，文献 [16] による瞬時周波数を用いた方法により補正される．瞬時周波数は基本波検出と比較して耐雑音性に優れたいるため，耐雑音性を向上させる効果が期待される．

瞬時周波数は，以下に示す Flanagan の式 [17] により計算する．

$$\omega_i(\omega,t) = \frac{\Re[S(\omega,t)]\Im[\frac{dS(\omega,t)}{dt}] - \Im[S(\omega,t)]\Re[\frac{dS(\omega,t)}{dt}]}{|S(\omega,t)|^2}, \tag{2}$$

$$S(\omega,t) = \int w(\tau-t)x(\tau)e^{-j\omega\tau}d\tau \tag{3}$$

$$\frac{dS(\omega,\tau)}{d\tau} = \int \frac{dw(\tau-t)}{dt}x(\tau)e^{-j\omega\tau}d\tau, \tag{4}$$

ここで，$w(t)$ は切り出しに用いる窓関数，信号 $x(t)$ は入力信号を表す．波形の切り出しは，DIO により推定された基本周期の 3 倍の周期を有する Blackman 窓により行う．切り出された波形から瞬時周波数とパワースペクトルを計算し，以下の式により修正された F0 を計算する．

$$\hat{\omega}_0(t) = \frac{\sum_{k=1}^{K}|S(k\omega_0,t)|\omega_i(k\omega_0,t)}{\sum_{k=1}^{K}k|S(k\omega_0,t)|}, \tag{5}$$

ここで，$\omega_0(t)$ と $\hat{\omega}_0(t)$ は，それぞれ修正前の F0 と修正された F0 の角周波数表現を表す．

WORLD での実装では，文献 [16] と同様に，式 (5) による補正を 2 度行う．1 回目は，DIO により推定された F0 を ω_0 とし，K は 2 で $\hat{\omega}_0(t)$ を求める．2 回目は，1 回目の補正で求められた $\hat{\omega}_0(t)$ を ω_0 とし，K を 6 として行う．

5. 評価

DIO と StoneMask による補正の効果を確認するための実験を行う．実験は，F0 の再選択が基本波検出法による評価結果に悪影響を及ぼさないこと，および，StoneMask が耐雑音性を向上しているかの 2 点に着目して行う．本実験では，音声データベースを用いた評価は行わず，人工的に生成した F0 が既知の調波複合音により相対的な性能を比較する．

5.1 評価用信号と評価指標

まず，両評価で共通して用いる調波複合音を，以下の式により定義する．

$$x(t) = n(t) + \sum_{m=1}^{M}\cos\left(2\pi m\int_0^t f_0(\tau)d\tau\right), \tag{6}$$

ここで $n(t)$ は加算性の雑音，$f_0(t)$ は F0 の時系列を示す．M は調波数に対応し，$Mf_0(t)$ がナイキスト周波数を超えない範囲での最大値に設定される．

実験に用いる F0 軌跡 $f_0(t)$ は，基本的にターゲットとなる基本周波数で固定される．ただし，人間の音声には揺らぎが含まれるため，F0 軌跡は完全な固定値にするのではなく，Klatt により提案された，以下の式により定義される揺らぎ [18] を加える．

$$\Delta f_0(t) = \frac{FL}{50}\frac{f_c}{100}(\sin(2\pi 12.7t) + \sin(2\pi 7.1t) + \sin(2\pi 4.7t)), \tag{7}$$

FL はフラッターに相当するパラメータであり，文献 [18] に倣い 25 で固定する．f_c は F0 の標準値であり $f_0(t) = f_c$ とする．これに，$\Delta f_0(t)$ を加えることで最終的な F0 軌跡とする．

評価指標は，Fine pitch error (FPA)，Gross pitch error (GPA) [19] や Gross error [7] ではなく，真値と推定値との RMS 誤差により求めることとする．これは，真値が明確であり，大きく外れた値が推定された場合は純粋に RMS 誤差が増加するため，SNR を段階的に変化させて評価して傾向を確認することで，充分比較可能であるという判断による．

5.2 比較に用いる従来法と共通する実験条件

F0 推定法には，Baseline となる基本波検出法，DIO，DIO の後に瞬時周波数で補正する DIO+StoneMask の他に，state-of-the-art の方法として SWIPE′[9]（以下では単に SWIPE とする）と TANDEM-STRAIGHT で利用される XSX[20], [21] を利用する．SWIPE については，論文の著者が Matlab のソースコードを公開しているため，それを利用した．

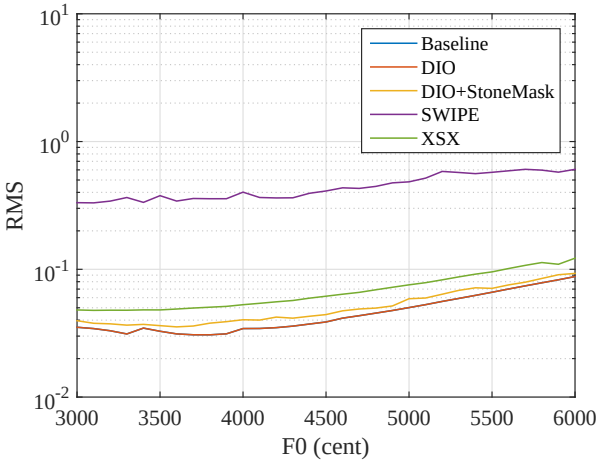

図 3 各方法により推定された F0 の推定誤差．F0 は 3000 cent (92.5 Hz) から 6000 cent (523.3 Hz) まで 100 cent 刻みとしている．Baseline の結果は，DIO と完全に一致している．

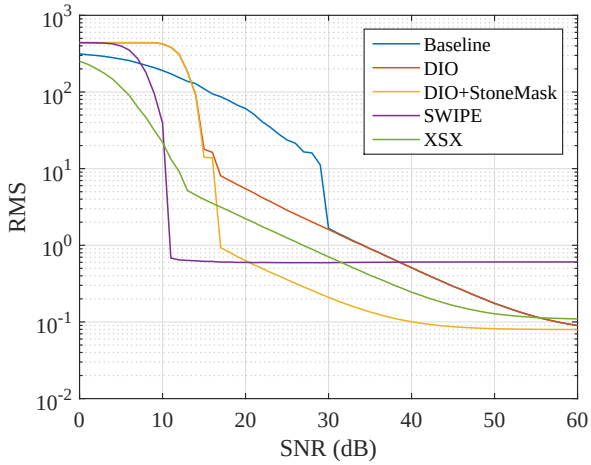

図 4 SNR と推定誤差との関係．F0 は 440 Hz に固定している．

評価に用いる調波複合音の信号長は 1.2 s とし，評価用信号のサンプリング周波数は 48 kHz とする．$f_0(t)$ は 1 ms 毎に求めることとする．評価には 0.1 から 1.1 s の区間で得られる 1000 サンプルの結果を利用し，RMS 誤差を計算する．これは，音源の開始や終了時刻での F0 を推定することは原理的に困難であることに起因する．

5.3 F0 の高低と推定精度との関係

第 1 の評価では，F0 の高さと推定精度との関係を調査する．F0 は 3000 cent (約 92.5 Hz) から 6000 cent (約 523.3 Hz) まで 100 cent 刻みで変更し，評価用の調波複合音を生成した．評価結果を図 3 に示す．図の横軸は F0，縦軸は推定精度を対数表示で示し，値が小さいほど優れた性能であることを示す．本実験では雑音が無く大幅な推定ミスが生じていないため，Baseline と DIO の結果は完全に一致している．StoneMask は，DIO により推定された結果と比較して，わずかに誤差を拡大することが分かる．しかしながら，TANDEM-STRAIGHT で利用される XSX や SWIPE よりは高い性能を達成していることも確認できる．

5.4 耐雑音性の評価

第 2 の評価では，ホワイトノイズを雑音として SNR を変えながら RMS 誤差を計算することで，耐雑音性を検証する．SNR は 0 dB から 60 dB までとし，F0 は 440 Hz に固定した．また，雑音のランダム性の影響を低減するため，異なる雑音を用いて 100 回評価を行い，その中央値を最終的な結果とした．本実験では各調波の振幅を固定し，雑音源をホワイトノイズとしているため，全帯域において SNR は等しいこととなる．

結果を図 4 に示す．横軸は SNR で，縦軸は RMS 誤差の対数表示である．図から明らかに，Baseline は耐雑音性において他の方法より明らかに劣り，30 dB 以上の SNR を確保できない場合推定性能が大幅に低下していることが分かる．一方，後処理による再選定を含む DIO では，SNR の低下に伴い精度も低下するが，SNR が 17 dB 程度までは大幅な推定精度の低下が生じていないことが確認できる．StoneMask による補正は，DIO で大幅に誤差が増加する 17 dB 以上であれば効果があり，20 dB 以上あれば SWIPE より高い精度を達成可能である．これは，ある程度推定値が真値に近い値にすることができれば瞬時周波数による補正で誤差が抑圧できることを意味するが，補正には限界があり SNR が 20 dB 以下になると原理的に SWIPE のほうが有利であることを示す．

6. 考察

評価結果より，瞬時周波数による補正は耐雑音性を向上できることが示された．以下では，性能や今後の展望に関して考察する．

6.1 性能に関する考察

基本波検出法は高 SNR の音声を対象として提案された方法であり，SNR が 38 dB 以上で SWIPE より高い性能を，55 dB 以上で XSX より高い性能を達成している．F0 再選定の後処理と瞬時周波数に基づく補正を行うことで，SNR が 20 dB 程度確保できれば SWIPE のような最先端の方法よりも優れた性能を達成可能であることが示された．現在，ヘッドセットマイクロホンを用いた音声収録や，歌声合成において自身の声を収録し合成に利用することが行われているが，提案法を用いることで，それらの環境に耐えうる性能が達成可能であることが期待される．瞬時周波数による補正は，雑音が全く存在しない前提で評価した場合において DIO よりも劣ることが示されたが，この条件はむしろ特異であり，実環境で収録された音声分析において悪影響は無いと考えて良いと思われる．

本実験の信号と雑音の条件では，全帯域で SNR が均一

となるため，この SNR を実環境で収録された音声に対して適用することはできない．ピンクノイズのように低域にパワーが偏った信号では，傾向は同様である一方誤差が大幅に増大しはじめる境界となる SNR は異なる可能性が存在する．雑音源の種類を変えて詳細な評価を行うことは，今後の重要な課題である．

6.2 提案法の展望

瞬時周波数による補正は，推定精度の向上という利点はあるものの，SNR が 20 dB 以下では SWIPE のほうが高い性能を達成していることが確認された．耐雑音性の向上と，高 SNR 音声に対する性能の改善は異なる方針となるが，使用者が目的に応じて利用する F0 推定法の選択肢を増やすことは有益であると考えられる．今後は，その両面について，個別の方法を探る予定である．例えば SNR の高い帯域の調波のみ補正に利用することで，更に低域の雑音に頑健にすることなどが考えられる．

7. おわりに

本稿では，筆者らが提案した高 SNR の音声を対象とした F0 推定法について，候補の再選定による滑らかな F0 軌跡を得る後処理，および瞬時周波数による補正による耐雑音性の向上法について述べた．後処理は，耐雑音性を向上させ，後処理無しの方法と比較して真値に近い F0 軌跡を推定可能となることを示した．瞬時周波数による補正は，音声の SNR を 20 dB 程度確保することで，SWIPE など最先端の方法と遜色のない性能と耐雑音性を実現した．

今後は，瞬時周波数による補正に用いる調波の数や帯域の最適化による性能向上について検討する．例えば，帯域毎のパワーを計算し，フォルマントなど強いパワーを有する帯域の調波を選択的に用いることで，さらに耐雑音性を向上できる可能性がある．

謝辞 本研究は，科研費 15H02726，26540087，および東北大学電気通信研究所 共同プロジェクト（H25/A08）の支援を受けて実施された．

参考文献

[1] Dudley, H.: Remaking speech, *J. Acoust. Soc. Am.*, Vol. 11, No. 2, pp. 169–177 (1939).
[2] 森勢将雅，河原英紀，西浦敬信：基本波検出に基づく高 SNR の音声を対象とした高速な F0 推定法，電子情報通信学会論文誌 D， Vol. J93-D, No. 2, pp. 109–117 (2010).
[3] Kawahara, H., Cheveigné, A., Banno, H., Takahashi, T. and Irino, T.: Nearly defect-free F0 trajectory extraction for expressive speech modifications based on STRAIGHT, *in Proc. Interspeech2005*, pp. 537–540 (2005).
[4] Ross, M., Shaffer, H., Cohen, A., Freudberg, R. and Manley, H.: Average magnitude difference function pitch extractor, *IEEE Transactions on acoustic, speech, and signal processing*, Vol. ASSP-22, No. 5, pp. 353–362 (1974).
[5] Noll, A.: Short-time spectrum and “cepstrum” techniques for vocal pitch detection, *J. Acoust. Soc. Am.*, Vol. 36, No. 2, pp. 269–302 (1964).
[6] Noll, A.: Cepstrum pitch determination, *J. Acoust. Soc. Am.*, Vol. 41, No. 2, pp. 293–309 (1967).
[7] Cheveigné, A. and Kawahara, H.: YIN, a fundamental frequency estimator for speech and music, *J. Acoust. Soc. Am.*, Vol. 111, No. 4, pp. 1917–1930 (2002).
[8] Mauch, M. and Dixon, S.: PYIN: A fundamental frequency estimator using probabilistic threshold distributions, *in Proc. ICASSP2014*, pp. 659–663 (2014).
[9] Camacho, A. and Harris, J. G.: A sawtooth waveform inspired pitch estimator for speech and music, *J. Acoust. Soc. Am.*, Vol. 124, No. 3, pp. 1638–1652 (2008).
[10] 大村浩，田中和世：基本波フィルタリング法による精細ピッチパターンの抽出，日本音響学会誌，Vol. 51, No. 7, pp. 509–518 (1995).
[11] Shimamura, T. and Kobayashi, H.: Weighted autocorrelation for pitch extraction of noisy speech, *IEEE Transactions on speech and audio processing*, Vol. 9, No. 7, pp. 727–730 (2001).
[12] Nakatani, T. and Irino, T.: Robust and accurate fundamental frequency estimation based on dominant harmonic components, *J. Acoust. Soc. Am.*, Vol. 116, No. 6, pp. 3690–3700 (2004).
[13] Morise, M.: CheapTrick, a spectral envelope estimator for high-quality speech synthesis, *Speech Communication*, Vol. 67, pp. 1–7 (2015).
[14] Morise, M.: Error evaluation of an F0-adaptive spectral envelope estimator in robustness against the additive noise and F0 error, *IEICE Trans. Inf. & Syst.*, Vol. E98-D, No. 7, pp. 1405–1408 (2015).
[15] 森勢将雅：目指せ音声分析合成マスター！，日本音響学会聴覚研究会資料， Vol. 45, No. 8, pp. 1–7 (2015).
[16] 河原英紀，森勢将雅，西村竜一，入野俊夫：基本波の FM と AM 成分に基づく高速な基本周波数推定法について，日本音響学会聴覚研究会資料，Vol. 41, No. 9, pp. 679–684 (2011).
[17] Flanagan, J. and Golden, R.: Phase vocoder, *The Bell System Technical Journal*, Vol. 45, No. 9, pp. 1493–1509 (2009).
[18] Klatt, D. and Klatt, L.: Analysis, synthesis, and perception of voice quality variations among female and male talkers, *J. Acoust. Soc. Am.*, Vol. 82, No. 2, pp. 820–857 (1990).
[19] Rabiner, L., Cheng, M., Rosenberg, A. and McGonegal, C.: A comparative performance study of several pitch detection algorithms, *IEEE Transactions on acoustic, speech, and signal processing*, Vol. ASSP-24, No. 5, pp. 399–418 (1976).
[20] Kawahara, H., Morise, M., Takahashi, T., Nisimura, R., Irino, T. and Banno, H.: TANDEM-STRAIGHT: A temporally stable power spectral representation for periodic signals and applications to interference-free spectrum, f0, and aperiodicity estimation, *in Proc. ICASSP2008*, pp. 3933–3936 (2008).
[21] Kawahara, H. and Morise, M.: Technical foundations of TANDEM-STRAIGHT, a speech analysis, modification and synthesis framework, *SADHANA - Academy Proceedings in Engineering Sciences*, Vol. 36, No. 5, pp. 713–728 (2011).