

Title	雑音環境における基本周波数推定とこれを用いた雑音抑圧に関する研究
Author(s)	石本, 祐一
Citation	
Issue Date	2000-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1362
Rights	
Description	Supervisor:赤木 正人, 情報科学研究科, 修士

雑音環境における基本周波数推定と これを用いた雑音抑圧に関する研究

石本 祐一

北陸先端科学技術大学院大学 情報科学研究科

2000年2月15日

キーワード： 基本周波数推定、雑音環境、くし形フィルタ、雑音抑圧.

1 はじめに

音声認識は一部実用化されているものの、雑音環境における認識精度は大幅に低下する。また、補聴器では、雑音が補聴器使用者にうるさい感覚を引き起こし、会話理解を妨害する。その他、音声分析合成や音源分離等も含めて音声信号処理の実用には、雑音環境において音声波形や音声の特徴を抽出することが重要となる。

河原らはフィルタの中心周波数からフィルタ出力の瞬時周波数への写像の不動点を用いて基本周波数を抽出する方法 (TEMPO2) を提案している [1]。TEMPO2 はクリーンな音声から高精度の基本周波数を得ることができるが、雑音のある環境においては推定精度が低下する。雑音に強い基本周波数推定法として、鶴木は周波数軸上の Comb Filtering による方法を提案している [2]。瞬時振幅の Comb Filtering による方法は信号対雑音比 (SNR) が 0dB 程度の雑音が付加された音声に対しても基本周波数を推定することができるが、クリーンな音声に対しては nTEMPO2 よりも推定精度の点で劣る。

本論文では、帯域幅可変くし形フィルタによる雑音抑圧と TEMPO2 と Comb Filter による基本周波数推定法を組み合わせることにより、雑音環境においても頑健な基本周波数推定法を提案する。また、これによって推定された高精度な基本周波数を基にした帯域幅可変フィルタによる雑音抑圧を行なう。

2 提案法の概要

提案法の処理の流れを図 1 に示す。

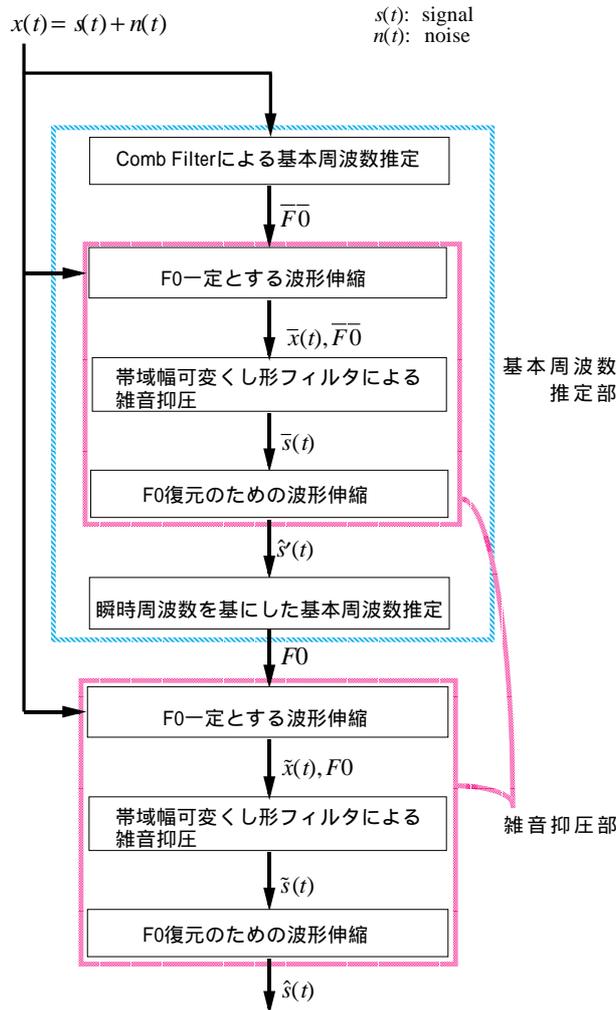


図 1: 雑音抑圧アルゴリズムの概要

調波構造をもつ目的音 $s(t)$ と雑音 $n(t)$ が混ざり合った混合音 $x(t)$ が、ただ一つの受音源で観測される。基本周波数推定部において、

1. 雑音に強い瞬時振幅の Comb Filtering による基本周波数推定法 [2] を用いて、観測された混合音 $x(t)$ から目的音の基本周波数を大まかに推定 ($\bar{F}0$) する
2. 推定された基本周波数に合わせた帯域幅可変くし形フィルタによって雑音を抽出し、その雑音を元の混合音から引き去ることにより雑音の抑圧を行なう
このとき、基本周波数を一定にするような波形伸縮を行なうことにより、雑音抑圧時の誤差を低減させる。
3. 雑音を抑圧した信号 $\hat{s}'(t)$ に対して、TEMPO2[1] を用いる

という操作によって高精度の基本周波数 F_0 が推定できる。また、さらに次の雑音抑圧部において、

4. この高精度の F_0 を用いた帯域幅可変くし形フィルタで雑音を推定し除去することで、雑音抑圧した目的音 $\hat{s}(t)$ を得ることができる。

3 帯域幅可変くし形フィルタによる雑音抑圧

目的音 $s(t)$ を基本周期 $T(t)$ の調波複合音とすれば、雑音を $n(t)$ として、混合音 $x(t)$ は、

$$x(t) = s(t) + n(t) \quad (1)$$

$$= \sum_m a_m e^{j(m\omega_0(t)t + \theta_m)} + \sum_k b_k e^{j(\omega_k t + \theta_k)} \quad (2)$$

$$(\omega_0(t) = 2\pi/T(t))$$

と表わされる。 $T(t)$ を一定値 $T (= 2\pi/\omega_0)$ と仮定し、式 (2) を $\pm T$ だけ時間軸でずらして混合音から引き去る信号 $g(t)$ を計算すれば、

$$g(t) = \frac{2x(t) - x(t - T) - x(t + T)}{4} \quad (3)$$

$$= \sum_k b_k e^{j(\omega_k t + \theta_k)} \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (4)$$

となる。 $n(t)$ のフーリエ変換を $N(\omega_k)$ とすれば、 $g(t)$ のフーリエ変換 $G(\omega_k)$ は、

$$G(\omega_k) = N(\omega_k) \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (5)$$

となる。よって、雑音スペクトル $N(\omega_k)$ は、

$$N(\omega_k) = G(\omega_k) / \sin^2 \frac{\omega_k}{\omega_0} \pi \quad (6)$$

となる。 $N(\omega_k)$ を逆フーリエ変換した雑音 $n(t)$ を、元の混合音から引き去ることにより雑音が除去され、目的音 $s(t)$ が推定できる。

しかし、このままでは ω_k/ω_0 が整数のときに式 (6) の雑音スペクトル $N(\omega_k)$ が無限大となるため、実際の使用にはある値 ϵ を設定し、

$$\tilde{N}(\omega_k) = \begin{cases} G(\omega_k) / \sin^2 \frac{\omega_k}{\omega_0} \pi & \left| \sin \frac{\omega_k}{\omega_0} \right| \geq \epsilon \\ G(\omega_k) & \left| \sin \frac{\omega_k}{\omega_0} \right| < \epsilon \end{cases} \quad (7)$$

とする。

これは、 ϵ の値によって通過帯域幅が制御できるくし形フィルタとなる。

また、式 (3) では基本周期を一定と仮定して計算を行なっているが、実音声では基本周期は時間的に変動するため、このままでは推定雑音 $\tilde{n}(t)$ に誤差が生じる。そこで、基本周期を一定にするために音声波形の時間軸での伸縮を行なう。

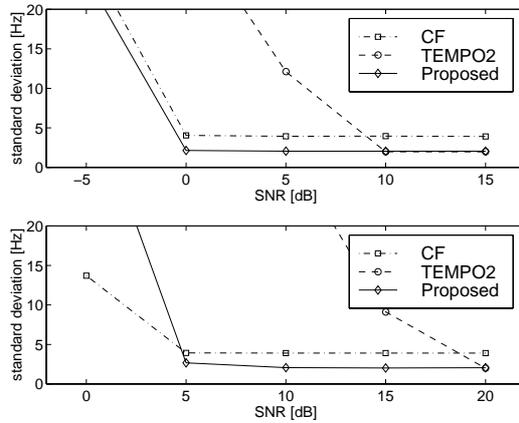


図 2: 雑音付加音声における提案法の基本周波数推定精度：白色雑音 (上)、ピンク帯域雑音 (下)

4 雑音環境における基本周波数推定法

帯域幅可変くし形フィルタによる雑音抑圧の実際の使用においては雑音のある環境から目的音の基本周波数を推定しなければならない。そこで、まず、従来手法の対雑音性能を調べた。

調査の結果から、雑音が白色雑音、ピンク帯域雑音のどちらの場合にも

- 雑音が小さい場合は TEMPO2 が最も精度が良いが、雑音が大きくなると推定精度が急激に低下する
- 雑音が大きい場合は瞬時振幅の Comb Filtering による基本周波数推定法が精度が良いが雑音が小さい場合では TEMPO2 程の推定精度が得られない

という傾向が見られた。

そこで、先に述べた雑音抑圧アルゴリズムを用いて、雑音環境において頑健で高精度な基本周波数推定法を作成する。提案法のアルゴリズムは図 1 の基本周波数推定部に示すように、TEMPO2 と Comb Filter による方法の 2 種類の基本周波数推定法と帯域幅可変くし形フィルタから構成される。図 2 に提案法の推定精度 (推定誤差の標準偏差) と雑音強度の関係を示す。

図 2 から、提案法は SNR が 5dB と大きな雑音が存在する環境においても雑音のない環境と同程度の基本周波数を推定することができる頑健で高精度な基本周波数推定法であるといえる。

5 雑音抑圧シミュレーション

表 1 に示す実験データを用いて、提案法による目的音の抽出を試みた。雑音は 0dB から 20dB まで 5dB 刻みで加え、評価尺度として SNR、スペクトル歪み尺度 (SD)、聴覚特

表 1: 実験データ

Sim. No.	目的音	雑音
1	/a/ /i/ /u/ /e/ /o/ /aoi/ (話者 mht,mau,fsu,fkn : ATR speech database)	白色雑音
2	/a/ /i/ /u/ /e/ /o/ /aoi/ (mht,mau,fsu,fkn)	ピンク帯域雑音 (60 ~ 6000kHz)
3	/a/ (fsu)	/a/ (mht)

性を考慮した歪み評価尺度 (ASD)[3] を用いて評価を行なった。

シミュレーションの結果、雑音抑圧前に比べて SNR で 5 ~ 7dB 程度の向上が見られ、本研究で提案した雑音環境における基本周波数推定法及び雑音抑圧法が、雑音が白色雑音であるときや二重母音の場合において、有効であることが分かった。しかし、雑音がピンク帯域雑音の場合には、音声の高調波成分の除去や目的音と同じ周波数帯域の雑音の影響があり、改良の余地がある。

6 おわりに

帯域幅可変くし形フィルタによる雑音抑圧と雑音に頑健な瞬時振幅の Comb Filtering による基本周波数推定法、雑音には弱いが高精度の瞬時周波数を基にした基本周波数推定法を組み合わせた雑音環境における頑健で高精度な基本周波数推定法を提案した。提案法により、SNR が 5dB と大きな雑音が存在する環境においても雑音のない環境と同様の基本周波数を推定することができた。

また、これによって得られる高精度の基本周波数を用いた帯域幅可変くし形フィルタによる雑音抑圧法を提案した。本方法では、雑音抑圧前に比べて SNR で 5 ~ 7dB 程度の向上が得られた。

参考文献

- [1] 河原 英紀, Parham Zolfaghari, Alain de Cheveigné, Roy D. Patterson, 周波数から瞬時周波数への写像の不動点を用いた音源情報の抽出について, 電子情報通信学会技術報告, SP99-40, July 1999.
- [2] 鶴木 祐史, 赤木 正人, 聴覚の情景解析に基づいた雑音下の調波複合音の一抽出法, 電子情報通信学会論文誌 (A), vol.J82-A, No.10, pp.1497-1507, 1999.
- [3] Mitsunori Mizumachi, Masato Akagi, An Objective Distortion Estimator for Hearing Aids and its Application to Noise Reduction, EUROSPEECH'99, pp.2619-2622, 1999.