論　文

対数スペクトルにクリッピングと帯域制限を用いる基本周波数抽出法

小林　　載[†*]　　島村　徹也[†]

An Extraction Method of Fundamental Frequency Using Clipping and Band Limitation on Log Spectrum

Hajime KOBAYASHI[†*] and Tetsuya SHIMAMURA[†]

あらまし　音声の基本周波数は，音声処理の幅広い分野で必要とされる特徴パラメータである．音声信号から基本周波数を抽出する手法は過去に多数提案されているが，あらゆる条件に対し有効な手法はいまだに確立されていない．本論文では，ケプストラム法を改良することにより雑音環境下における音声に有効となる基本周波数抽出法を提案する．本手法の特色は，対数スペクトルのうち特に雑音の影響を受けやすい高周波数成分とスペクトルの谷の部分を除去し，音声信号の調波構造を明確にした上でケプストラムを求める点にある．計算機シミュレーション実験の結果，従来法に比べ，本手法における抽出精度は gross pitch error を改善することができた．特に，周期性を有する雑音が混入された音声の場合に，本手法により顕著な効果が得られた．

キーワード　基本周波数，調波構造，クリッピング，帯域制限，ケプストラム法

1. まえがき

音声における基本周波数は音の高さを表す特徴パラメータであり，話者，心理，環境などの様々な条件で変化する．このため，基本周波数を抽出することは，非常に重要なことである [4].

過去に，数多くの基本周波数抽出法が提案されている．従来の基本周波数抽出法は，時間波形に対する処理，相関関数による処理，そしてスペクトル領域の処理に大別することができる．しかし，抽出に対する精密さ，耐雑音性，信頼性等，そして高速処理のすべての条件を満たす手法はいまだに存在しない [1]〜[3].

ケプストラム法（CEP）[5] は，広く知られる基本周波数抽出法である．この手法は，音声信号の対数スペクトルを波形として見ると，一つの周期波として観測することができる点に着目し，それをスペクトル分析することによって抽出を行うものである．この手法はホルマントの影響を容易に取り除くことができる長所があり，対数スペクトルにおいて周期性が明確であれば，高精度の抽出を可能にする．しかし，雑音によって対数スペクトルの周期性が失われると，抽出において，信頼性，精密さの双方が大幅に劣化する欠点を有する．

雑音が付加された音声信号の対数スペクトルを見ると，スペクトルの谷と高周波数帯域において，雑音のひずみが顕著に現れる．しかし，離散フーリエ変換（DFT）を行うことによって基本周波数を抽出するときは，周期波のうち 2 周期以上をデータとして用いれば，抽出が可能である．そこで本論文は，対数スペクトルにクリッピング，高周波数帯域の除去を施し，それらを CEP の処理に加える基本周波数抽出法を提案する．これは，対数スペクトルの周期性が明確な部分のみを利用することによって耐雑音性を実現する CEP といえる．

本論文では，まず **2.** で提案法の原理を述べ，具体的な手法を記述する．続く **3.** では実音声に対する従来法（自己相関関数法，ケプストラム法）と提案法の比較評価を行う．音声の信号対雑音比（SNR）を変化させ，基本周波数の抽出を行い，抽出精度を検討する．そして，最後に，**4.** で結論を述べる．

2. 提案法の原理

2.1 抽出原理

有声音の対数スペクトルは，図 1 のように調波構造を有するため，これを波形として考えれば，一つの周

[†] 埼玉大学工学部情報システム工学科，浦和市
Dept. of Information and Computer Sciences, Saitama University, Urawa-shi, 338–8570 Japan
[*] 現在，パイオニア（株）

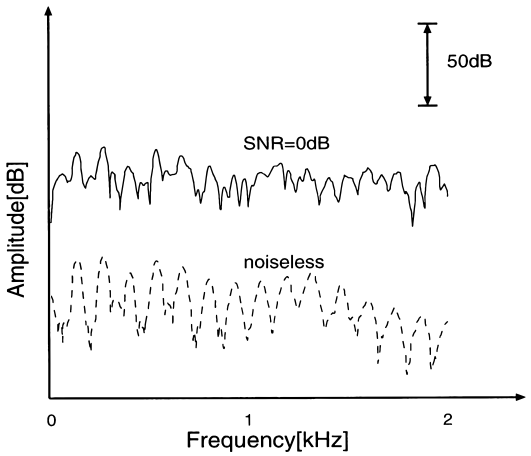


図 1　調波構造の様子
Fig. 1　Harmonics structure of speech signal.

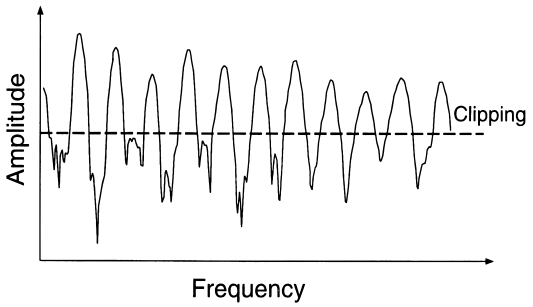


図 2　クリッピングの様子
Fig. 2　Clipping on log spectrum of speech signal.

期波として観測することができる．CEP は，対数スペクトルの有するその周期性に着目し，基本周波数を抽出する方法である．具体的には，対数スペクトルを DFT によってスペクトル分析することによってケプストラムを求め，その横軸にあたるケフレンシー上に現れるピークの位置を調べることによって基本周波数を抽出する．そのため，CEP の抽出精度は，調波構造の明確さに依存する抽出法といえる．

しかし，低 SNR における音声信号の対数スペクトルは雑音の影響を大きく受けるため，周期性が失われる．そのため，ケフレンシー上に現れるべきピークを抑え，音声信号が有する周期とは無関係な部分に振幅値を分散させてしまう．そこで，この対数スペクトルを眺めると，大きく雑音の影響を受ける部分が次に述べる 2 点存在することに気づく．

まず，第 1 点は，対数スペクトルの谷の部分で雑音のひずみが生じている点である．パワースペクトルに対数をとったとき，その振幅値が高い部分ほど雑音の影響を圧縮できるのに対し，振幅値が低い部分ほど雑音の影響を圧縮できなくなる．そのため，振幅値の小さい部分は，雑音の影響を受けやすくなる．それは，スペクトルの谷の部分で複雑なピークを生成し，本来有する周期の半分の周期波を形成する．そのため，半ピッチエラーが多発し，抽出精度において支障をきたす．

この理由により，次の処理 1 を加え，雑音の影響を軽減することを考える．

［処理 1］

対数スペクトルにクリッピングを施し，調波構造の谷の部分を除去する．

この様子を図 2 に示す．ただし，対数スペクトルは声道の影響を受けており，この状態でクリッピングを行っても効果は薄い．そのため，対数スペクトルにリフタ処理を行い，声道の影響を除去することによって声帯音源信号のみを取り出した後にクリッピングを行った．なお，リフタ処理は，対数スペクトルから DFT を計算することによって一度ケプストラムを求め，声道特性の部分（0〜2.5 [ms]）をすべて 0 に置き換え，逆離散フーリエ変換（IDFT）を計算することによって行った．

第 2 点は，高周波数帯域の部分に雑音の影響を受ける点である．対数スペクトルのもつ周期性は周波数が高くなるにつれ明確さを失う．特に，雑音の混入された音声の対数スペクトルは周期性の明確な低周波数部分より，高周波数部分で雑音の影響を受けることに気づく．そのため，これをスペクトル分析しても，その周期の不明確さはケフレンシー上で現れるべきピークを抑え，結果として抽出精度を劣化させる．この理由により，次の処理 2 を CEP に加え，雑音の影響を軽減することを考える．

［処理 2］

周波数 F_B をあらかじめ設定し，F_B よりも高い周波数でのスペクトル値を 0 に置き換える．

この様子を図 3 に示す．

これは，一般に，周期信号を DFT することによりその周波数を求めるとき，その周期信号の 2 周期分のデータさえあれば，周波数推定が可能である点を考慮したものである．

以上の二つの処理を，CEP において，対数スペク

トルを求めた後に適用することにより，対数スペクトルの周期性が明確な部分のみを残す．そして，ケフレンシー上に現れるピークを強調することによって，抽出精度の向上を図る．

低 SNR の音声信号であっても，ケフレンシー上のピークの位置を正確に検出できれば，基本周波数を正確に決定することができる．これが提案法の原理である．

2.2 流 れ 図

提案法の流れを図 4 に示す．まず，標本化周波数 10 kHz の音声信号をハミング窓により 512 点で切り出す．分析精度を高めるために 0 を付加して 2048 点で DFT を行い，対数スペクトルを求める．しかし，この対数スペクトルはホルマントの影響を受けている．そこで，その影響を除去するリフタ処理を行い，声帯音源信号成分を取り出す．そのあと，雑音の影響を受けているスペクトルの谷の部分を除去するためにクリッピングを施す．これにより対数スペクトルは，その周期性がより明確になる．また，設定された F_B よりも高い周波数部分を 0 に置き換えることにより，乱れた周期波形部分を除去する．その後で IDFT によりケプストラムを求め，ケフレンシー軸上においてピークを検出する．しかし，このケフレンシー軸上における基本周期の値は，10000/2048 = 4.88 Hz ごとに求められ，このままでは分析精度が低いため，ピークの位置とその前後の 3 点を利用しラグランジュ補間を行い，基本周波数を求める．

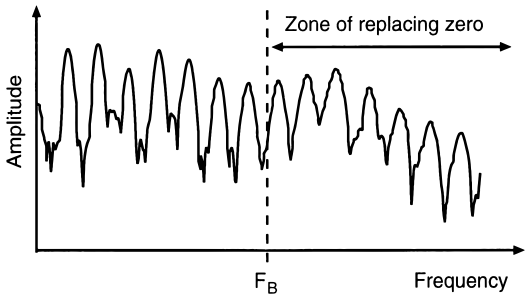


図 3 帯域制限の様子
Fig. 3 Band limitation on log spectrum of speech signal.

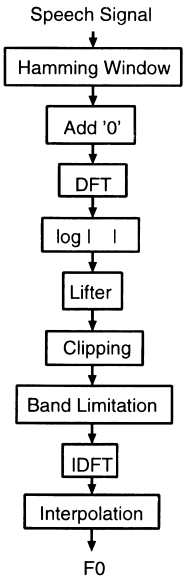


図 4 提案法の流れ
Fig. 4 Flowchart of the proposed method.

3. 実 験

本章では，提案法の抽出精度を調べるため，従来法との比較評価を行い，検討する．

3.1 実験条件

使用した音声データは，NTT アドバンステクノロジー（株）の「20ヶ国語音声データベース」に収録されている日本人話者男女各 4 名が発声した約 10 秒間の短文である．付加雑音は，白色雑音及び電子協騒音騒音データベースに収録されている “No.1 走行自動車内 2000 cc クラス”，“No.9 幹線道路・交差点”，“No.14 計算機室（ワークステーション）” の雑音を使用した．雑音データ，音声データは，ともに標本化周波数 10 kHz で標本化され，3.4 kHz で帯域制限されている．本実験では，SNR が ∞ dB，10 dB，5 dB，0 dB，-5 dB である音声データを用いた．なお，SNR は有声音区間だけを用い，以下の式 (1) から計算した．

$$\mathrm{SNR} = 10\log_{10}(P_S/P_N) \qquad (1)$$

ただし，P_S は音声信号の平均電力，P_N は雑音の平均電力である．

基準とする基本周波数は，音声の時間波形の視察によって基本周期を求め，その逆数をとったものを用いた．また，各音声データの有声音区間は，窓関数によって囲まれたフレーム内の最大振幅が信号全体の最大振幅の -30 dB 以上のときとし，それ以外は無声音区間として扱った．

抽出精度における評価法には，Rabiner らの手法 [6] をもとに，以下のように設定した．まず，抽出時間 n において基準となる基本周波数を $F_S(n)$，各手法に

よって抽出された基本周波数を $F_d(n)$ とし，抽出誤差 $e(n)$ を式 (2) によって求める．

$$e(n) = F_d(n) - F_S(n) \tag{2}$$

そして，$|e(n)| \geqq 10$ Hz のとき，gross pitch error（GPE）とし，総フレーム数のうち GPE の発生する割合を求めた．また，$|e(n)| < 10$ Hz のとき，fine pitch error とし，標準偏差を求めた．

なお，リフタ処理は，高速フーリエ変換（FFT）を用いて処理した．具体的には，対数スペクトルを IFFT し，一度ケプストラムを求めてから，0〜25 ポイント，2024〜2048 ポイントの低ケフレンシー部の実部，虚部の両方を 0 に置き換え，FFT を行う方式をとった．クリッピングは，クリッピングレベル C を用いた式 (3) からしきい値 L を設定し，L 以下の振幅値の部分を 0 に置き換えることによって行った．

$$L = C(d_{max} - d_{min}) + d_{min} \tag{3}$$

ここで，d_{max}，d_{min} はそれぞれ対数スペクトルの最大値及び最小値を表している．抽出範囲は，音声がもつ基本周波数の範囲といわれる 50〜400 Hz で処理した．

3.2 クリッピング及び帯域制限の効果

本実験では，クリッピングと帯域制限の効果を調べるため，提案法の抽出アルゴリズムのうちクリッピングあるいは帯域制限の一方を省略し，抽出精度を評価した．

［クリッピングによる効果］

まず最初に，帯域制限の処理を省略し，クリッピングのみの効果を調べた．クリッピング処理は，クリッピングレベル C を 0〜0.95 に変化させることによって抽出精度を調べた．その結果を図 5〜7 に示す．この特性グラフは男女各 4 名の音声データの総フレーム数のうち，GPE が発生した件数の割合を上述の四つの雑音をその音声データに SNR $= \infty$ dB，10 dB，0 dB でそれぞれ付加した場合それぞれにおいてまとめたものである．

これによると，自動車の雑音や交差点の雑音が付加された音声においては $C = 0.65$〜0.75 の間で設定したときに最も GPE が改善され，白色雑音や計算機室の雑音が付加された音声においては $C = 0.85$〜0.90 の間で設定したときに最も GPE が改善されている．

しかし，これら 4 種類の雑音が付加されたそれぞれの音声の GPE の平均を見ると，C=0.70〜0.80 と設定したときの抽出結果が最良の結果を与えている．

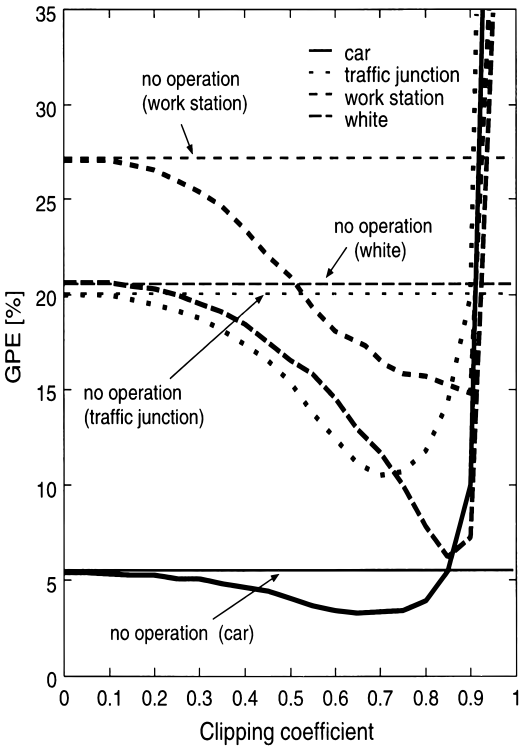


図 5　SNR = 0 dB の場合のクリッピングの効果
Fig. 5　Effect of clipping operation in case of SNR = 0 dB.

［帯域制限による効果］

次に，クリッピングの処理を省略し，帯域制限のみの効果を調べた．本実験では，F_B を 500〜3400 Hz に変化させることによって抽出精度を調べた．

その結果を図 8〜10 に示す．この特性グラフは，クリッピングによる効果の場合の特性グラフと同様の手段で求めた．

白色雑音が付加された音声においては $F_B = 500$〜1500 Hz のとき，計算機室の雑音，交差点の雑音，自動車の雑音がそれぞれ付加された音声においては，$F_B = 1400$〜1600 Hz のときに最も GPE が改善されている．

これら 4 種類の雑音が付加されたそれぞれの音声の GPE の平均を見ると，$F_B = 1400$〜1600 Hz と設定したときの抽出結果が最良の結果を与えている．

3.3 実験結果

提案法（MCEP）と従来法による抽出精度の評価の結果を図 11, 図 12 に示す．図 11 は GPE, 図 12 は

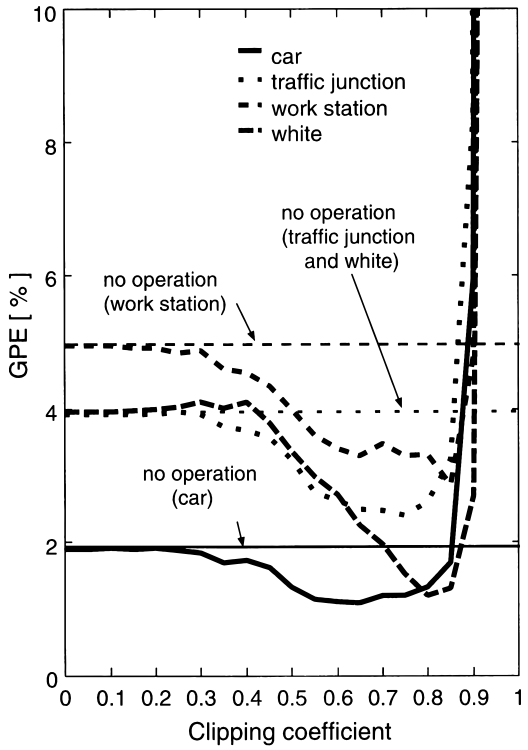


図 6　SNR = 10 dB の場合のクリッピングの効果
Fig. 6　Effect of clipping operation in case of SNR = 10 dB.

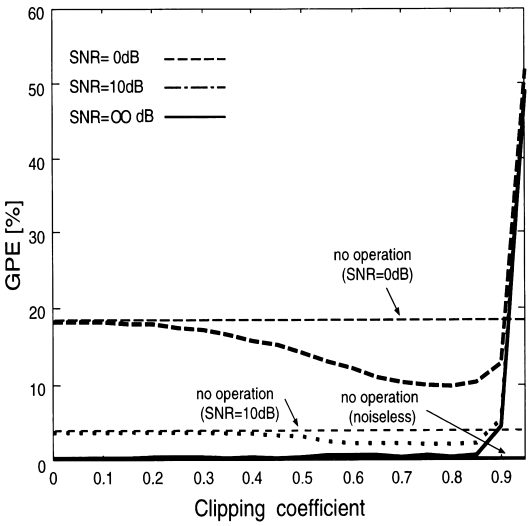


図 7　全体的なクリッピングの効果
Fig. 7　Total effect of clipping operation.

FPE を示してある．

従来法に用いた手法は自己相関関数法（AUTOC），

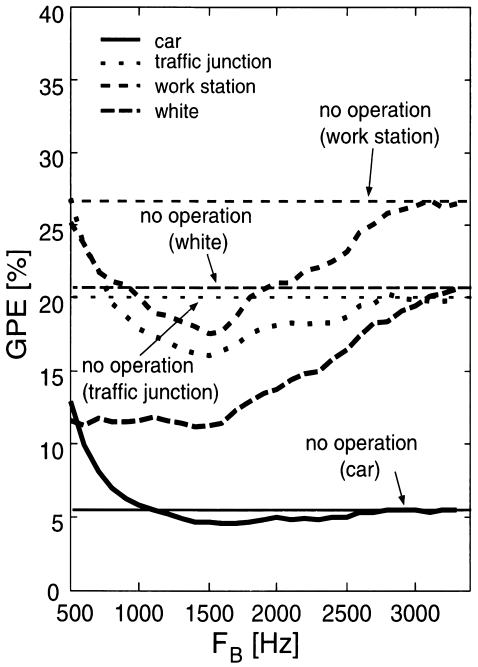


図 8　SNR = 0 dB の場合の帯域制限の効果
Fig. 8　Effect of band limitation operation in case of SNR = 0 dB.

ケプストラム法（CEP）である．K.A.Oh らは，八つの基本周波数抽出アルゴリズムで雑音の混入された音声の基本周波数を抽出させたところ，耐雑音性に関しては，AUTOC が最も優れていると結論づけている [3]．一方で，Hess は，無雑音の条件下においては CEP が優位である報告をしている [2]．

なお，AUTOC と CEP は，MCEP と分析精度を等しくするため，基本周波数抽出の手掛りとなるピークの位置とその前後の計 3 点において，ラグランジュ補間を行った．CEP は DFT, IDFT ともに 2048 点で行った．AUTOC において，自己相関関数 $\phi(\tau)$ は以下の式 (4) で計算した．ただし，音声信号を $s(n)$，フレームデータ長を N とする．

$$\phi(\tau) = \sum_{n=0}^{N-1} s(n)s(n+\tau) \qquad (4)$$

ラグ数 τ は 0〜500 の範囲で求めた．ここで評価に用いる MCEP は，先の実験，すなわち **3.2** の結果から，$C = 0.8$，$F_B = 1500$ Hz とした．

無雑音音声において評価すると，GPE においては AUTOC，CEP，MCEP の 3 手法ともに抽出精度は同等であるが，FPE については MCEP の方が AU-

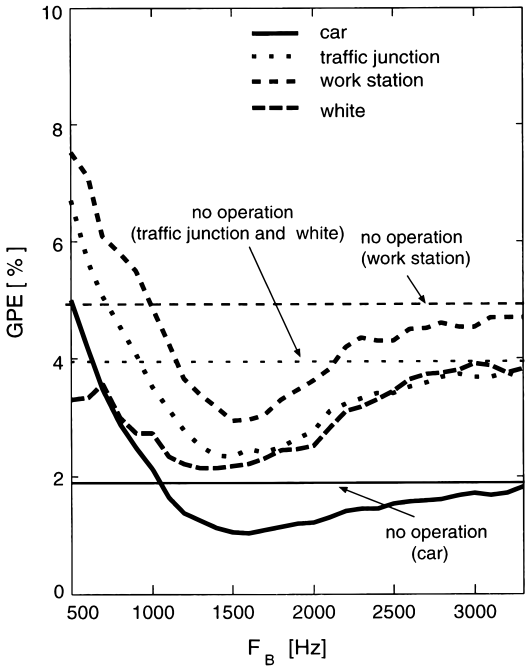


図 9　SNR = 10 dB の場合の帯域制限の効果
Fig. 9　Effect of band limitation operation in case of SNR = 10 dB.

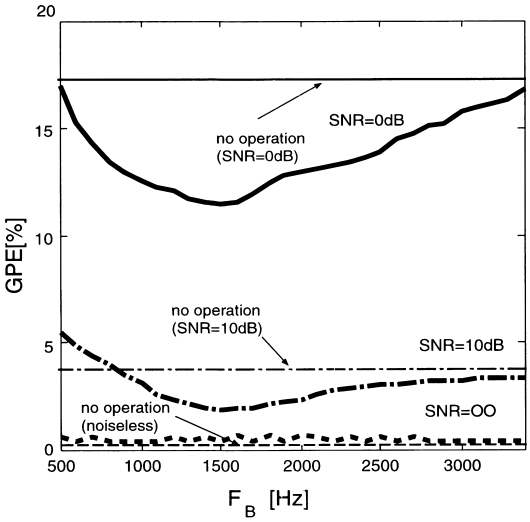


図 10　全体的な帯域制限の効果
Fig. 10　Total effect of band limitation operation.

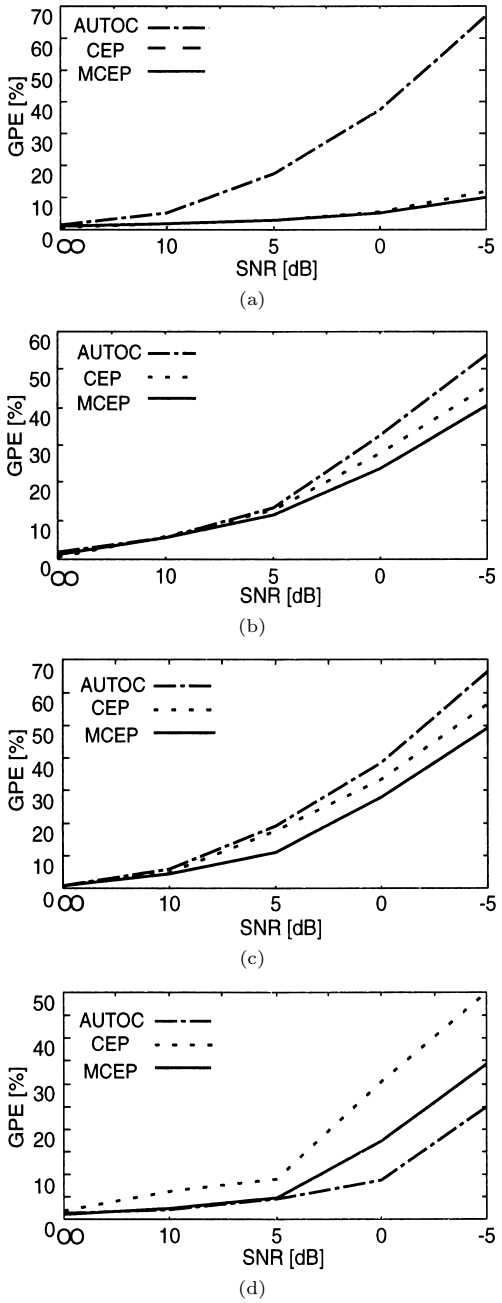


図 11　Gross pitch error の割合：(a) 自動車内の雑音，(b) 交差点の雑音，(c) 計算機室の雑音，(d) 白色雑音
Fig. 11　Parcentage of gross pitch error: (a)car noise, (b) noise in a traffic junction, (c) noise in a computer room, (d) white noise.

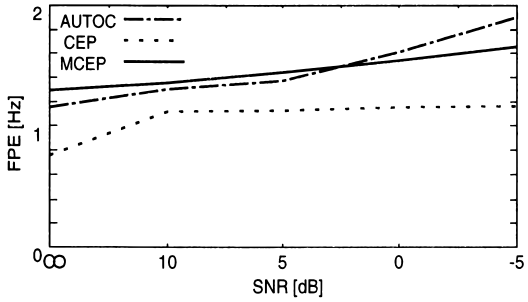


図 12　Fine pitch error の標準偏差：(a) 自動車内の雑音，(b) 交差点の雑音，(c) 計算機室の雑音，(d) 白色雑音

Fig. 12　Standard deviation of fine pitch error: (a) car noise, (b) noise in a traffic junction, (c) noise in a computer room, (d) white noise.

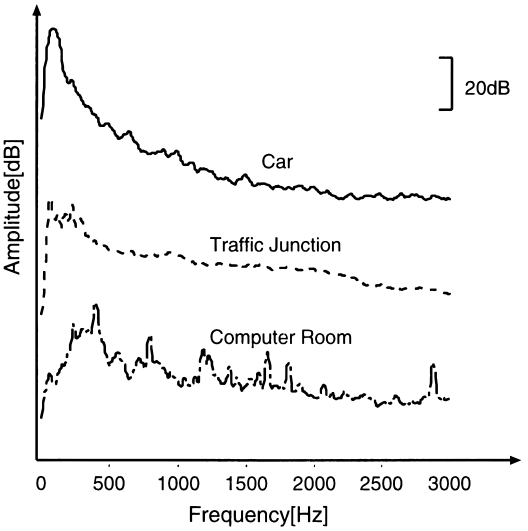


図 13　自動車内，交差点，計算機室の雑音の長時間スペクトル

Fig. 13　Long time spectrum of noise in a car, in a traffic junction and in a computer room.

TOC，CEP よりも抽出精度が低い．この場合において，対数スペクトルの谷や高周波数帯域の部分には，調波構造を乱す成分がないので，これらは基本周波数を抽出する際に必要な情報となる．MCEP のように，対数スペクトルにクリッピングや帯域制限を処理の一部として加えると，情報が処理前よりも減るため，必然的に抽出精度の劣下の要因となる．この理由により，無雑音音声から基本周波数を抽出するときは，クリッピングや帯域制限が有効ではないことがわかる．

次に雑音が付加された音声について評価する．まず，自動車内の雑音を付加した場合について評価する．図 13 に示すように，雑音自体の長時間スペクトルにおいて，周波数成分は 50〜400 Hz の抽出範囲よりも低い部分に強く有する．そのため，抽出範囲から考え，AUTOC よりは大いに有効であるが，MCEP，CEP においての雑音の影響は小さく，MCEP についての改善度は小さい．しかし，雑音が付加された音声について若干 CEP との優位性が見られるのは，スペクトルの谷の雑音によるひずみを除去するクリッピングの効果が現れたものであると考えることができる．

交差点の雑音については，図 13 が示すように，雑音自体の長時間スペクトルにおいて，周波数成分が基本周波数が有するとされる 50〜400 Hz 全体に強く有する．そのため，帯域制限の効果は期待できない．し

かし，低SNR下においてAUTOC，CEPとの優位性が見られるのは，スペクトルの谷の部分に生じる雑音によるひずみがクリッピングによって除去されているため，効果が現れていると考えることができる．

計算機の雑音は，図13が示すように，雑音自体の長時間スペクトルが抽出範囲よりも高い部分で強く有する．しかも，この雑音は調波構造に似たスペクトルを有するため，クリッピング以上に帯域制限が効果的である．そのため，このような雑音に対しては，MCEPはAUTOC，CEPよりも有効であると考えることができる．

白色雑音についても，雑音の特性が周波数全体に一様に分布するため，雑音の影響が現れやすい高周波数帯域を除去することは効果があると考えられる．白色雑音はランダム雑音であり，雑音成分を低遅延部分に圧縮するAUTOCの長所が抽出精度において最も効果を発揮した．このため，AUTOCに対するMCEPの優位性は見られなかったが，MCEPの土台となるCEPに対する優位性は顕著に見られる．

以上の結果から，本手法は雑音の付加された音声に対し，有効な手法であることを示すことができた．周期雑音が付加された場合は，特に，従来法と比較し，優位であることを示すことができた．

4. むすび

本論文では，雑音が混入された音声の対数スペクトルのうち，調波構造が乱れやすい部分を除去し，ケプストラムを求める基本周波数抽出抽出法を提案した．そして，計算機シミュレーションを介し抽出精度について評価した．その結果，雑音環境下における音声において有効な基本周波数抽出結果が得られた．

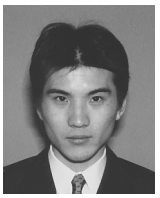
具体的には，提案法では自己相関関数法やケプストラム法よりも全体的にGPEを改善することができた．交差点や自動車内の雑音のように音声が有するとされる基本周波数の範囲及びそれ以下の周波数帯域で調波構造に影響を与えるような雑音においてはクリッピングが有効であり，計算機室内や白色雑音のように，高周波数帯域で影響を与えるような雑音については帯域制限が有効であることを確認することができた．

謝辞 本研究において，予備的検討を行った本研究室の國枝伸行氏（現在，松下通信工業（株））に感謝致します．また，日ごろから御討論していただく日本工業大学の鈴木誠史教授，並びに本学の八嶋弘幸助教授に感謝致します．

文　献

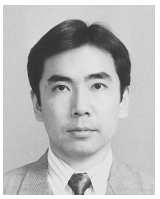
[1] W.J Hess, Pitch determination of speech signals, Springer-Verlag, Berlin, 1983
[2] W.J Hess, "Pitch and Voicing Determination," in Advances in speech signal Processing, ed. S.Furui and M.M.Sondhi, Marcel Dekker, 1992.
[3] K.A. Oh and C.K. Un, "A performance comparison of pitch extraction algorithms for noisy speech," Proc. IEEE Inter. Conf. Acoustics, Speech and Signal Processing, pp.18B4.1–18B4.4, 1984.
[4] 古井貞熙，ディジタル音声処理，東海大学出版会，1985.
[5] A.M. Noll, "Cepstrum pitch determination," J. Acoust. Soc. Amer., vol.41, no.2, pp.293–309, Feb. 1967.
[6] L.R. Rabiner, M.J. Cheng, A.E. Rosenberg, and C.A. McGonegal, "A comparative performance study of several pitch detection algorithms," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-24, no.5, pp.399–417, Oct. 1976.

（平成10年8月28日受付，11年1月11日再受付）



小林　載

平9埼玉大・工・情報卒．平11同大大学院博士前期課程了．現在，パイオニア（株）勤務．音声信号処理に関する研究を進めている．日本音響学会会員．



島村　徹也　（正員）

昭和61慶大・理工・電気卒．平3同大大学院博士課程了．工博．同年埼玉大・工・助手．現在同助教授．この間，平7ラフバラ大学，平8ベルファーストクイーンズ大学（共に連合王国）客員研究員．スペクトル解析及び適応信号処理に関する研究に従事．計測自動制御学会，日本音響学会，IEEE各会員．