論　文

帯域制限をかけた振幅スペクトルのべき乗に基づく基本周波数抽出法

島村　徹也[†a)]　　高木　浩司[†]

Fundamental Frequency Extraction Method Based on the p-th Power of Amplitude Spectrum with Band Limitation

Tetsuya SHIMAMURA[†a)] and Hiroshi TAKAGI[†]

あらまし　音声の基本周波数抽出は，多くの音声処理分野で必要とされる重要な課題である．本論文では，基本周波数の存在する帯域内で振幅スペクトルのべき乗を施す基本周波数抽出法を提案する．本法は，分析フレームごとに，入力音声信号の信号雑音比に依存して振幅スペクトルのべき乗数を変化させる．これにより，白色雑音環境下での，高い信号雑音比から低い信号雑音比までの幅広い音声信号への対処を可能とする．実母音を用いた予備実験により，振幅スペクトルのべき乗数と信号雑音比の関係式が導き出される．それを実連続音声にも適用し，従来法との比較実験において，提案法の優位性が明らかにされる．

キーワード　基本周波数抽出，べき乗振幅スペクトル，信号雑音比

1. まえがき

基本周波数は，声の高さやイントネーションを表す特徴パラメータであり，音声認識，音声合成，音声強調などの様々な音声処理システムにおいて用いられる．したがって，音声信号からの基本周波数抽出は，音声処理における重要な研究テーマの一つである．

これまでに数多くの基本周波数抽出法が提案されているが，抽出の精密さと雑音耐性を兼ね備えた確約される抽出法は存在しないと思われる [1]〜[3]．しかし中でも，自己相関関数法 (AUTOC) [4] はランダム性雑音に頑強であり，特に白色雑音環境下においては最も有力な方法の一つとされている．白色雑音環境は通信システム系に多く見られ，その対策が望まれている．しかし，AUTOC といえども基本周波数の抽出誤りは発生し，その割合は声道特性によって大きく左右されることが指摘されている [2]．

上記の AUTOC の性質は，周波数領域における入力音声信号の形状から解釈できる．本論文では，このような周波数領域での AUTOC の振舞いを考慮して，定常白色雑音環境下での特性改善の観点から新たな基本周波数抽出法を導出する．提案法は入力信号の振幅スペクトルに対して，入力信号の信号雑音比 (SNR: Signal-to-Noise Ratio) に対応させたべき乗処理，すなわちスペクトル変形を行う．このとき，振幅スペクトルには，基本周波数が存在する範囲で帯域制限がかけられる．これにより，声道特性の影響を低減し，べき乗数の可変領域を拡大することが可能となる．

実母音を用いた予備実験において，スペクトル変形に用いられるべき乗数と入力信号の SNR との関係がまず考察される．そして，その予備実験において得られた関係式をもとに，実連続音声に対して提案法と従来法を施し，それらの実行精度を比較し，検討する．

本論文の構成は，**2.** において，AUTOC について述べ，その特徴を明らかにする．そして **3.** で，**2.** での見解をもとに提案法を導出する．続く **4.** においては，予備実験並びに比較実験の結果を示し，考察を行う．そして最後に **5.** にて，本論文を結ぶことにする．

2. 周波数領域解釈

本章では，AUTOC に対して周波数領域からの解釈を与える．

音声信号の周期性を強調するために，AUTOC では入力信号の自己相関をとり，その得られた波形より基本周期を抽出する [4]（基本周波数は基本周期の逆数として得られる）．一般に自己相関関数は時間領域で定義

† 埼玉大学工学部情報システム工学科，さいたま市
Faculty of Engineering, Saitama University, Saitama-shi, 338–8570 Japan
a) E-mail: shima@sie.ics.saitama-u.ac.jp

されるが，Wiener–Khintchine の定理に基づきパワースペクトルの逆フーリエ変換として算出することも可能である [5]．このとき周波数領域での処理に着目すると，図 1 に示すように，AUTOC は入力信号の振幅スペクトルを 2 乗することによりスペクトルを拡張することになる．このスペクトル拡張操作により，もし雑音特性が周波数上で平たんなら，音声の基本周波数とその高調波成分を雑音成分よりも相対的に高くすることが可能となる．これが白色雑音の低減へとつながる．しかし一方で，声道特性により，音声の基本周波数のある高調波成分がホルマント周波数付近で強調される場合，上記のスペクトル拡張操作はそれを更に強調してしまう．これは，基本周波数の抽出誤りを引き起こす結果となる．したがって，AUTOC は声道特性の影響を大きく受けることになる．

AUTOC が受ける声道特性の影響は，基本的に音声パワースペクトルにレベル差が生じることに起因する．したがって，音声パワースペクトルを平たん化，すなわちスペクトル圧縮することにより，問題解決されることになる．しかし，従来試みられた予測誤差フィルタリングを前処理とする方法 [6] では，無雑音環境下において基本周波数の抽出精度が向上されるものの [7]，白色雑音環境下においては大きく特性劣化してしまう [2]．これは，予測誤差フィルタリングが，本来低いレベルにある雑音成分までも音声成分のレベルまで引き上げてしまうためである．

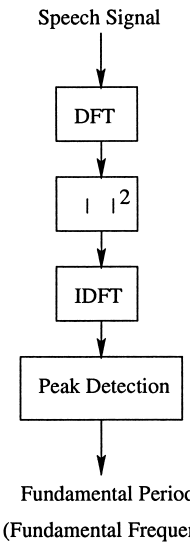

図 1　AUTOC の概念図（DFT と IDFT はそれぞれ離散フーリエ変換，逆離散フーリエ変換を表す）
Fig. 1　Concept of AUTOC. (DFT and IDFT denote Discrete Fourier Transform and Inverse Discrete Fourier Transform, respectively)

そこで，入力信号に含有される雑音量によって，スペクトル拡張とスペクトル圧縮を使い分けることが考えられる．入力信号の SNR に着目すると，低 SNR の場合には，スペクトル拡張をして耐雑音性を向上させ，高 SNR の場合には，スペクトル圧縮して声道特性への頑強性を向上させると，特性改善が得られると期待できる．本論文では，このような観点から，入力信号の SNR によってスペクトル変形の処理を施す新しいタイプの基本周波数抽出法を導出することにする．

3. 提 案 法

本章では，前章での周波数領域解釈をもとに，提案する基本周波数抽出法を導出する．処理対象としては，定常白色雑音環境下での音声信号を基本的に考える．

3.1 スペクトル拡張

まず低 SNR の環境を考えることにする．前章でのスペクトル拡張の考え方からすると，AUTOC による振幅スペクトルの 2 乗のスペクトル拡張に処理を固定する必要はなく，p 乗のスペクトル拡張をすることでより一般化できるはずである．ここで，p はある正の整数を表すことにする（原理的に p は整数に限定されることはないが，ここでは簡単のためにこのように考えることにする）．

SNR が下がるにつれて，雑音成分のスペクトルレベルは高くなっていく．しかし，このような場合においても，白色雑音環境下では，上記のべき乗数 p を増大することにより，音声と雑音のスペクトルレベルの差を拡大できる可能性があると考えられる．現にこれは事実である．例を示すことにする．

図 2 は，10 kHz のサンプリング周波数で得られた（3.4 kHz の帯域制限を伴う），低 SNR 白色雑音下 (SNR-5 dB) でのある連続音声の有声音部のスペクトルを表している．$p = 1$ が振幅スペクトルに対応している．$p = 2, 3, 4$ はその振幅スペクトルの 2，3，4 乗の特性をそれぞれ表している．この図より，音声信号が有する基本周波数の高調波成分が，べき乗数 p を増大するにつれて明確になっていく様子がわかる．これは，べき乗数の増大が，雑音成分の低減を導いた結果と解釈することができる．

図 3 は，図 2 のそれぞれのべき乗振幅スペクトルを逆フーリエ変換し，時間波形に変換したものを示して

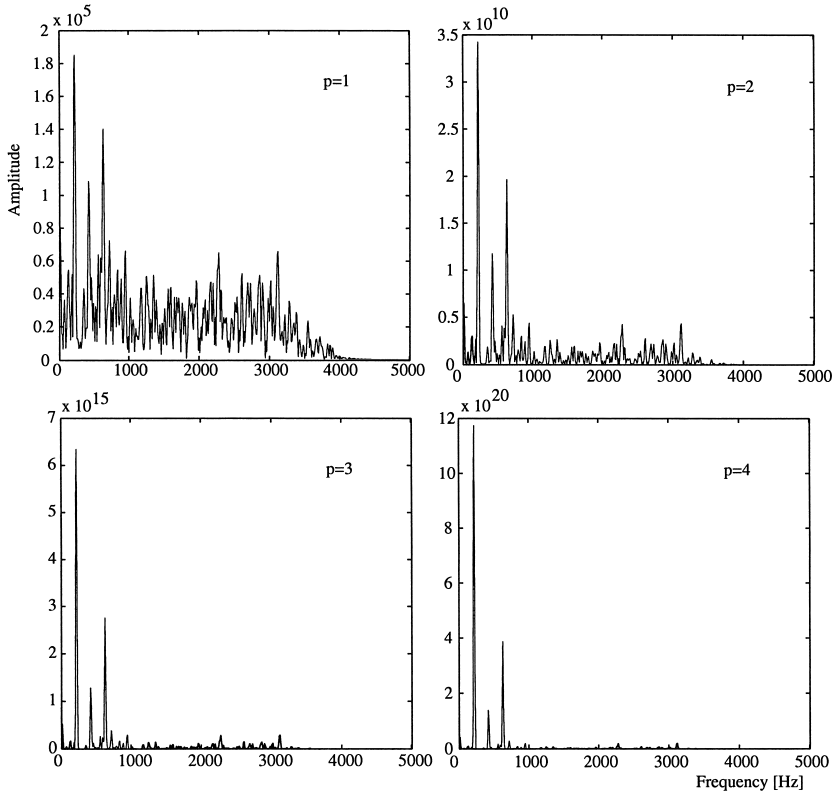

図 2　異なるべき乗数の振幅スペクトル
Fig. 2　Amplitude spectra with different exponents.

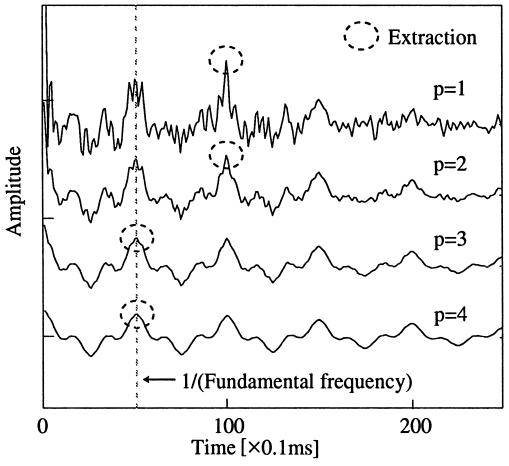

図 3　図 2 におけるスペクトルの時間領域波形
Fig. 3　Time-domain waveforms of spectra in Fig.2.

いる．点線の楕円で囲った部分は，零時間成分付近を除く範囲において，振幅値の最大値から抽出されることが期待される基本周期抽出部分を表している．すなわち，図 3 では，$p = 1$ 及び $p = 2$ の場合において，実際の基本周期の 2 倍の抽出がなされてしまい，また正確な基本周期の抽出は，$p = 3$ 及び $p = 4$ の場合にのみなされることを表している．ここで，$p = 2$ の場合が従来の AUTOC に対応することに留意すると，AUTOC では抽出誤りが発生してしまう場合においても，べき乗数を向上させることにより，抽出誤りを回避できる可能性があることを図 3 は示唆している．

3.2　帯域制限

上記のように，スペクトル拡張は雑音成分を低減する性質を有する．しかし，このスペクトル拡張においては，声道特性の影響をも考慮することが望まれる．なぜなら，振幅スペクトルを 3 乗，4 乗と拡張する場合，もし声道特性の影響が多大であるなら，その原理からして，ホルマント周波数付近のスペクトルピークが過度に強調されてしまう可能性があるためである．そこで，本法では，べき乗処理の前に振幅スペクトル

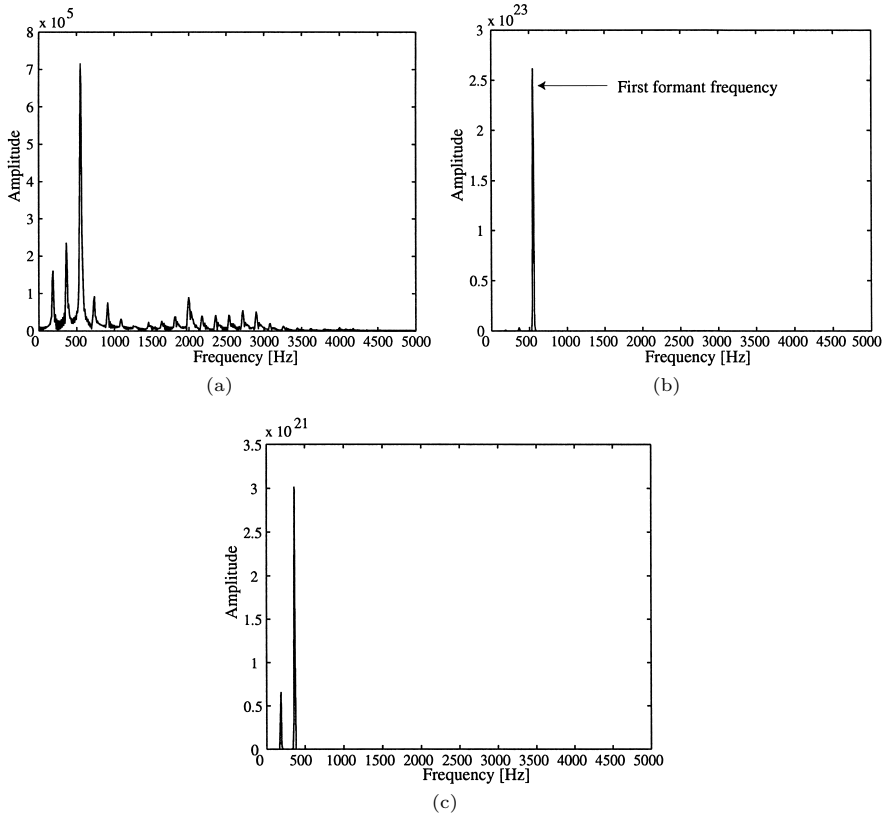

図 4　4 乗振幅スペクトル．(a) 振幅スペクトル，(b) 4 乗振幅スペクトル，(c) 帯域制限された 4 乗振幅スペクトル

Fig. 4　4-th power amplitude spectra. (a) amplitude spectrum, (b) 4-th power amplitude spectrum, (c) 4-th power amplitude spectrum with band limitation.

に帯域制限を施し，声道特性の影響を抑える工夫を講じることを考えることにする．

図 4 は，本法で施される帯域制限処理の例を示している．ここでの音声信号は，やはり 10 kHz のサンプリング周波数で得られた（3.4 kHz の帯域制限を伴う）連続音声のある有声音部であるが，無雑音環境下のものである．図 4 (a)，(b) は，それぞれ振幅スペクトルとその 4 乗特性を表している．この図 4 (a) より，ここでの音声信号の基本周波数は 200 Hz 付近にあるとわかるが，明らかに図 4 (b) では，それとは関係ない，ホルマント周波数に対応するスペクトルピークを強調してしまっている．

この図 4 (a) の振幅スペクトルに，50 Hz から 400 Hz までの帯域通過処理を施した後，4 乗した特性が図 4 (c) に示してある．ここでの帯域通過処理が，本法でのスペクトル拡張に用いられる帯域制限処理である．図 4 (c) での振幅スペクトルに施された帯域制限処理は，人の声の基本周波数が存在するといわれる範囲（例えば文献 [7] など参照）に対応させてある．図 4 (c) では，明らかに基本周波数に対応するスペクトルピークが現れている．

図 5 (a) は，図 4 (b) を時間領域に変換した波形であり，図 5 (b) は，図 4 (c) を時間領域に変換した波形である．200 Hz の基本周波数の基本周期が 50 ms であることを考慮すれば，図 3 においての抽出処理と同様の抽出処理を施すとき，図 5 (a) は明らかに抽出誤りを与えるが，図 5 (b) は正確な抽出結果をもたらすことがわかる．ここで注目すべき点は，図 4 (c) では，基本周波数に対応する第 1 高調波でのスペクトルピークレベルより第 2 高調波でのスペクトルピークレベルが高くなっているにもかかわらず，図 5 (b) では，基本周波数に対応する基本周期が正確に抽出され得るこ

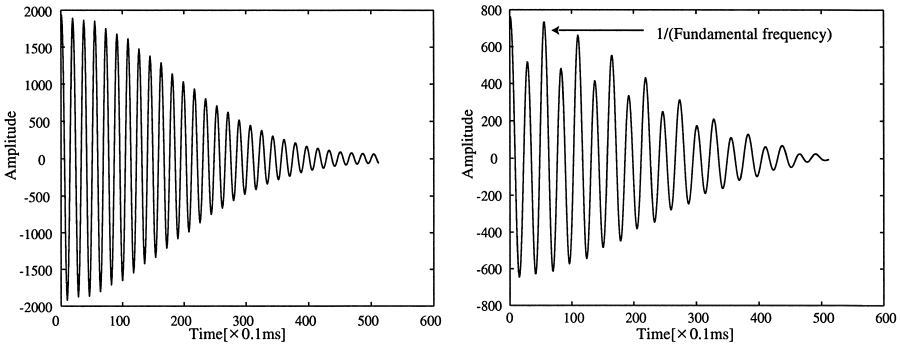

図 5　図 4 におけるスペクトルの時間領域波形（(a)，(b) はそれぞれ図 4 の (b)，(c) に対応する）
Fig. 5　Time-domain waveforms of spectra in Fig.4. ((a) and (b) correspond to (b) and (c) in Fig.4, respectively)

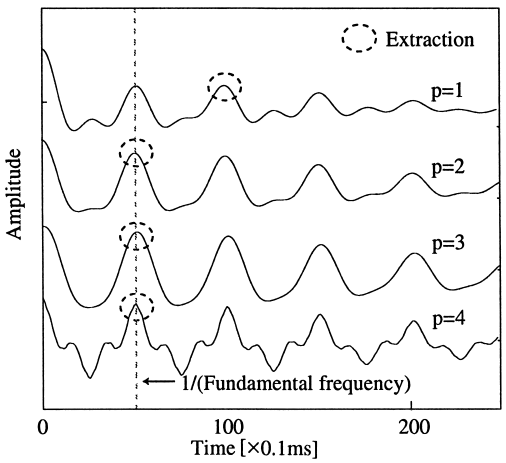

図 6　図 2 におけるスペクトルの時間領域波形（帯域制限がある場合）
Fig. 6　Time-domain waveforms of spectra in Fig.2. (with band limitation)

とである．高調波成分の和として形成される時間波形には，このように基本周波数抽出に適した性質がある．

図 6 は，図 2 で用いた雑音付加音声に上記の帯域制限を施し，その振幅スペクトルから図 2 と同様の $p=1$ から $p=4$ までのべき乗処理を行った後，それぞれ逆フーリエ変換をした結果を示している．図 6 では，$p=3$ 及び $p=4$ の場合に図 3 とほぼ同様の特性を与え，帯域制限処理がスペクトル拡張による雑音低減効果を保持できることを裏づけている．一方，$p=2$ の場合，図 6 では抽出誤りを与えておらず，図 3 の場合に比べて特性を改善している．これは，帯域制限の処理自体も，逆フーリエ変換による時間波形への変換の際に，SNR を向上させる効果を有するためである．図 6 での処理波形は図 3 での処理波形より全体的にスムーズであることに気づく．これは，帯域制限による雑音低減効果の現れとみなせる．このように帯域制限処理は，声道特性の影響を抑えるばかりでなく，結果的に雑音低減をももたらす．

本法での帯域制限処理は，スペクトル拡張を施す前提において，帯域制限される帯域内で雑音特性が平たんであれば効果的に働くと考えられる．したがって，それ以外の帯域で任意の特性を有する雑音であれば，容易に対処可能となる．

3.3　スペクトル圧縮

次に高 SNR の環境を考えることにする．**2.** での周波数領域解釈からすれば，高 SNR になるにつれて，振幅スペクトルを圧縮することが望まれる．このとき，振幅スペクトルの p 乗処理を行うとすると，p はある小さな正の実数値に近づけられるべきである．しかし，前節で述べた帯域制限を事前に振幅スペクトルに施せば，その帯域制限された周波数の範囲内での振幅スペクトルの最大及び最小のレベル差は，もとの振幅スペクトルにおける最大及び最小のレベル差より一般に小さくなるはずである．これは，一種のスペクトル圧縮効果に相当する．したがって，帯域制限のこのようなスペクトル圧縮効果を考慮して，本法では，1 より小さい値でのべき乗処理は施さないことにする．すなわち，高 SNR になるにつれて，べき乗数 p は 1 に近づけることにする．

3.4 提 案 法

以上のことを考慮し，本論文では，分析フレームごとに基本周波数が存在する帯域内，すなわち 50〜400 Hz で SNR を求め，その大きさに応じて，振幅スペクトルのべき乗数を変化させる基本周波数抽出法を提案する．以降では，この基本周波数が存在する 50〜400 Hz 内での SNR を SNR_{band} と記述し，信号全体としての SNR と区別して用いることにする．

本法における基本周波数抽出のための基本関数は，帯域制限された振幅スペクトルのべき乗を逆フーリエ変換することにより得られる．ここではそれを

$$R_x^p(\tau) = \mathrm{IDFT}[(|X(f)|B(f))^p] \qquad (1)$$

と表すことにする．ここで，$X(f)$ は入力信号の振幅スペクトル，$B(f)$ は上記の基本周波数存在帯域で大きさ 1，それ以外で 0 となる帯域通過型の零位相フィルタである．また，式 (1) 左辺の τ は時間変数に対応している．

基本周波数抽出においては，抽出の処理対象を 0〜900 Hz のように拡大して考える場合もあるが [8]，本論文では人の声の基本周波数が統計的にほぼカバーされる最低限の周波数帯域を保持しつつ，それ以外の帯域からの成分を基本的に除去する思想に基づき，基本周波数の処理対象を 50〜400 Hz と設定することにする．

3.5 流 れ 図

図 7 は，提案法の流れ図を示している．本アルゴリズムは，大別すると以下の三つの処理部分からなる．

- 雑音推定
- べき乗数決定
- 基本周波数抽出

以下にこれらの処理部分をそれぞれ説明する．ただしここでは，窓掛けされた入力信号が

$$x(m) = s(m) + n(m) \qquad (2)$$

で表されるとする．ここで，$x(m)$ はサンプリング時刻 m での入力信号，$s(m)$ はもとの音声信号，$n(m)$ は付加雑音に対応している．また，式 (2) のフーリエ変換は

$$X(f) = S(f) + N(f) \qquad (3)$$

で表されるとする．

[雑音推定]

雑音推定部では，あらかじめ既知の無音区間の分析

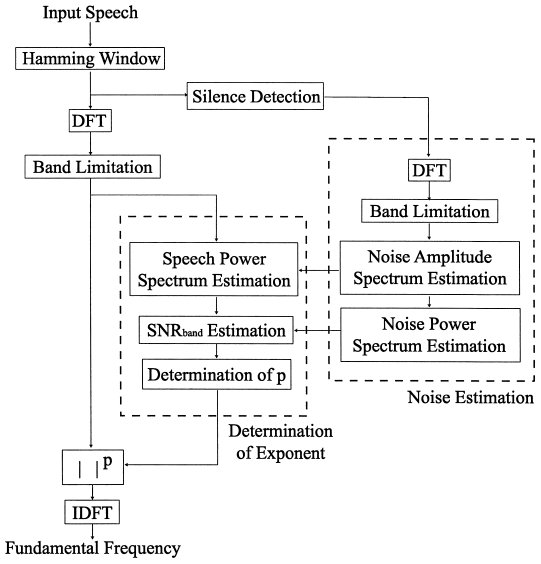

図 7 提案法の流れ図
Fig. 7 Flowchart of the proposed method.

フレームから（例えば，無音区間を含む音声信号の先頭部分などから），まず帯域制限された雑音振幅スペクトル $|\hat{N}_i(f)|B(f), i = 1, 2, \ldots, L$ を求める（ここで，i は分析フレームの番号を表している）．そして，それらの平均値として雑音振幅スペクトル $|\hat{N}_{band}(f)|$ を算出する．また，その 2 乗値を計算することにより，雑音パワースペクトル $|\hat{N}_{band}(f)|^2$ を求めておく．

[べき乗数 p の決定]

べき乗数決定部では，音声強調法で知られるスペクトル引き算法 [9] を適用し，分析フレーム内での音声信号のパワースペクトルをまず求める．具体的には，入力信号の帯域制限された振幅スペクトル $|X_{band}(f)| = |X(f)|B(f)$ を求め，それから先の雑音推定部で算出した雑音振幅スペクトル $|\hat{N}_{band}(f)|$ を引き去る．このとき，引き算の結果として現れる負の値は 0 に置き換える．このようにして得られた音声振幅スペクトル $|\hat{S}_{band}(f)|$ は 2 乗され，音声パワースペクトル $|\hat{S}_{band}(f)|^2$ が求められる．

上記の音声パワースペクトルと雑音推定部で求められた雑音パワースペクトルを用いて

$$S\hat{N}R_{band} = 10\log_{10}\frac{\sum|\hat{S}_{band}(f)|^2}{\sum|\hat{N}_{band}(f)|^2} \qquad (4)$$

のように SNR_{band} の推定値をデシベル表示で求める．そして，その算出された $S\hat{N}R_{band}$ 値をもとに，**4.2**

で後述する予備実験により得られる関係式から，べき乗数 p を決定する．

［基本周波数抽出］

べき乗数決定部で求められる p を用いて，式 (1) より基本関数を算出する．そして，基本周波数存在範囲 50〜400 Hz に対応する時間領域内で最大ピークの探索をし，基本周期を抽出する．そして，その得られた基本周期の逆数を算出し，基本周波数を求める．

4. 実　　験

本章では，実母音を用いた予備実験と実連続音声を用いた比較実験について述べる．そして，それらの結果を考察する．

4.1 予備実験と比較実験

［予備実験］

予備実験として，提案法で用いる振幅スペクトルのべき乗数と SNR_{band} との関係をまず求める．ここでは，ある男性話者によって発声された五つの母音/a/，/i/，/u/，/e/，/o/を用いる．各母音からは，それぞれ典型的な 1 分析フレームが切り出される．そして，その各音声波形に白色雑音を加え，それぞれ基本周波数を抽出する．本実験では，独立した付加白色雑音を各母音に対して 10^4 回用いる．そして，それぞれの場合において，設定した任意のべき乗数 p とともに提案法を実行し，基本周波数を抽出する．ここでは抽出された基本周波数 $\hat{F}$ が真の基本周波数 F_0 と $|F_0 - \hat{F}| < 10$ Hz の関係となったとき，基本周波数抽出は成功とみなす．そして，その成功した割合を抽出成功率とし，それを評価量とする．ここでの基本周波数の真値には，母音波形から測定された基本周期の逆数を用いる．本実験の結果得られた基本周波数の抽出成功率をグラフにまとめ，抽出成功率の高い部分を通るような曲線を考えることにする．そしてその曲線から，振幅スペクトルのべき乗数と SNR_{band} との関係式を求める．ここでの SNR_{band} には，付加前の雑音と音声信号を帯域制限し，それぞれのパワースペクトルから算出される値を用いる．表 1 には，本実験の諸条件がまとめてある．

表 1　実験の諸条件
Table 1　Conditions of experiments.

サンプリング周波数	10 kHz
帯域制限	3.4 kHz
窓関数	ハミング窓
分析フレーム長	51.2 ms
FFT (IFFT) 長	1024 ポイント

［比較実験］

予備実験にて得られた関係式を，SNR_{band} からの振幅スペクトルのべき乗数決定式とみなし，連続音声において実験をする．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」に収録されている短文である．これらの連続音声は男女それぞれ 4 人により発声されており，10 秒程度の長さがある．これらの連続音声信号には，それぞれ，全区間での SNR が −5 dB，0 dB，5 dB，10 dB，∞ dB となるように白色雑音が加えられる．

本実験は，提案法の従来法との実行精度の比較を目的とする．従来法としては，AUTOC，CEPST [10] を用いる．また，ごく最近提案された重み付き自己相関関数法 (WAUTOC) [11] も実行される．実験の諸条件は，基本的に表 1 と同じである．ただし，連続フレーム処理におけるフレームシフトは 10 ms とした．これらの実験諸条件は，提案法と従来法に共通である．

本実験での評価量としては，Rabiner ら [8] に基づく Gross Pitch Error (GPE) と Fine Pitch Error (FPE) を用いる．これらは

$$e(j) = F_{true}(j) - \hat{F}(j) \tag{5}$$

に基づいて算出される．ここで，$F_{true}(j)$ は j 番目の分析フレームでの基本周波数の真値，$\hat{F}(j)$ は抽出された基本周波数値，$e(j)$ はその誤差である．もし $|e(j)| \geq 10$ Hz となれば，それを GPE とみなし，全体に占めるその割合を求める．一方，$|e(j)| < 10$ Hz となれば，それを FPE とみなし，真値との差の標準偏差を求める．

基本周波数の真値は，予備実験と同様に音声波形より直接的に基本周期を求め，その逆数をとったものとする．ただし，連続音声信号からの有声/無声判別は，あらかじめ視察によって行った．

4.2 結果と考察

図 8 は予備実験で得られた振幅スペクトルのべき乗数と SNR_{band} との関係を示している．この図には，SNR_{band} が −5 dB，0 dB，5 dBd，10 dB であるとき，それぞれの場合にべき乗数 p を 1 から 6 まで変化させたときの基本周波数抽出結果が示されている．色が塗られた領域は基本周波数抽出の抽出成功率を表している．黒が抽出成功率 100%に相当し，以下薄くなるにつれて灰色が 50%，白が 0%となっている．図 8 は，それぞれの SNR_{band} で，五つの母音から得られ

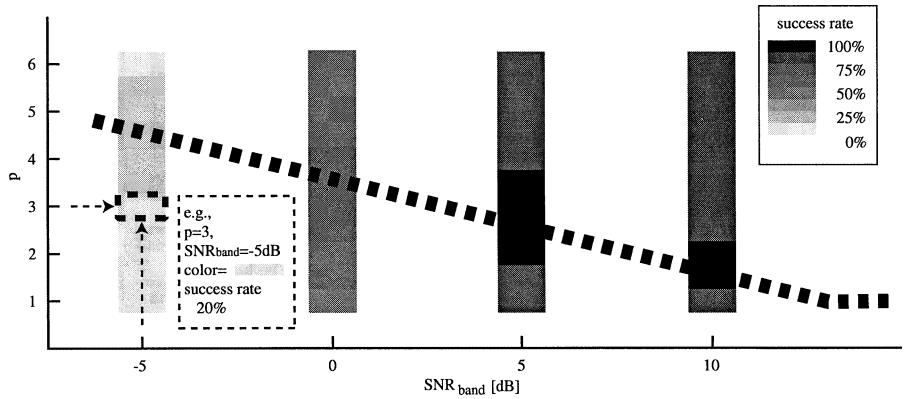

図 8　母音におけるべき乗数と SNR_{band} との関係
Fig. 8　Relation between exponent and SNR_{band} for the case of vowels.

た結果の平均値をとってある．例えば，SNR_{band} が -5 dB で，p が 3 ならば，それに対応する領域の色は “明るい灰色” となる．この色は抽出成功率 20%，すなわち，五つの母音にそれぞれ 10^4 回独立な白色雑音が用いられ，全体として 5×10^4 回基本周波数抽出をし，1.0×10^4 回抽出成功したことを表している．したがってここでは，できるだけ黒い部分を通るように曲線を決めれば，提案法は各 SNR_{band} において良好な結果を与えることになる．しかし，この図からもわかるように，抽出成功率が良い結果を示す領域はある程度幅がある．そこで簡単のため，ここでは直線で近似することにする．直線近似でのべき乗数 p と SNR_{band} との関係式は，p を SNR_{band} の関数として下記のように求められる．

$$p(SNR_{band}) = \begin{cases} 1, & 14 \leq SNR_{band} \\ -\frac{1}{5}SNR_{band} + 3.8, & SNR_{band} < 14 \end{cases} \quad (6)$$

図 9，図 10 は，連続音声に本法を適用したときの，各 SNR において得られた女性のみの GPE の結果と男性のみの GPE の結果を示している．比較のために，AUTOC，WAUTOC と CEPST の結果が示してある．ただし，ここでの Proposed (ideal) は，提案法において，あらかじめ付加前の雑音と音声信号から各分析フレームごとの SNR_{band} を求めておき，それを利用して上記の関係式からべき乗数を決定する方法を意味している．Proposed (noise estimation) は，**3.** で述べた提案法に直接対応している．

図 9 において，全体的に提案法が良好な雑音耐性を

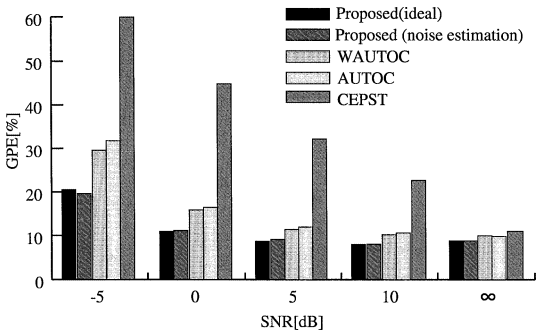

図 9　GPE の結果（女性）
Fig. 9　Results of GPE (female).

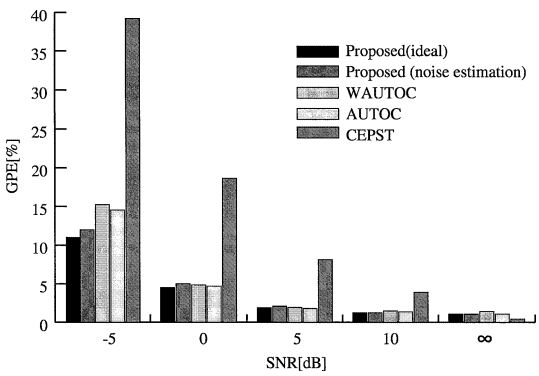

図 10　GPE の結果（男性）
Fig. 10　Results of GPE (male).

示している．特に SNR が -5 dB の場合には GPE が 10%以上も改善されるなど，顕著な改善効果がある．一方，図 10 においては，低 SNR 環境下において雑

音耐性は向上していると考えられるが，高 SNR 環境下では，従来の AUTOC とほぼ同等である．これは男性の場合，女性に比べ比較的基本周期が長く，すなわち周波数領域で基本周波数が低くなるためと考えられる．このような場合には，基本周波数に対応して調波構造が密になり，そのような調波スペクトルから逆フーリエ変換により得られる時間波形は，雑音の影響を受けにくくなる．AUTOC でも，既に SNR10 dB 程度で抽出誤差は 2%以下となっている．すなわち，まだ雑音低減の余地が残っていた低 SNR 環境下において，本法の能力が発揮される結果となったとここでは解釈できる．

図 9 及び図 10 からは，共通して Proposed (ideal) と Proposed (noise estimation) にほとんど差がないことを見てとれる．提案法は各分析フレームで決定されるべき乗数によって特性変化されるため，この結果は，提案法における SNR_{band} の推定方法が大きな推定誤りを与えることなく，べき乗数決定のためにほぼ

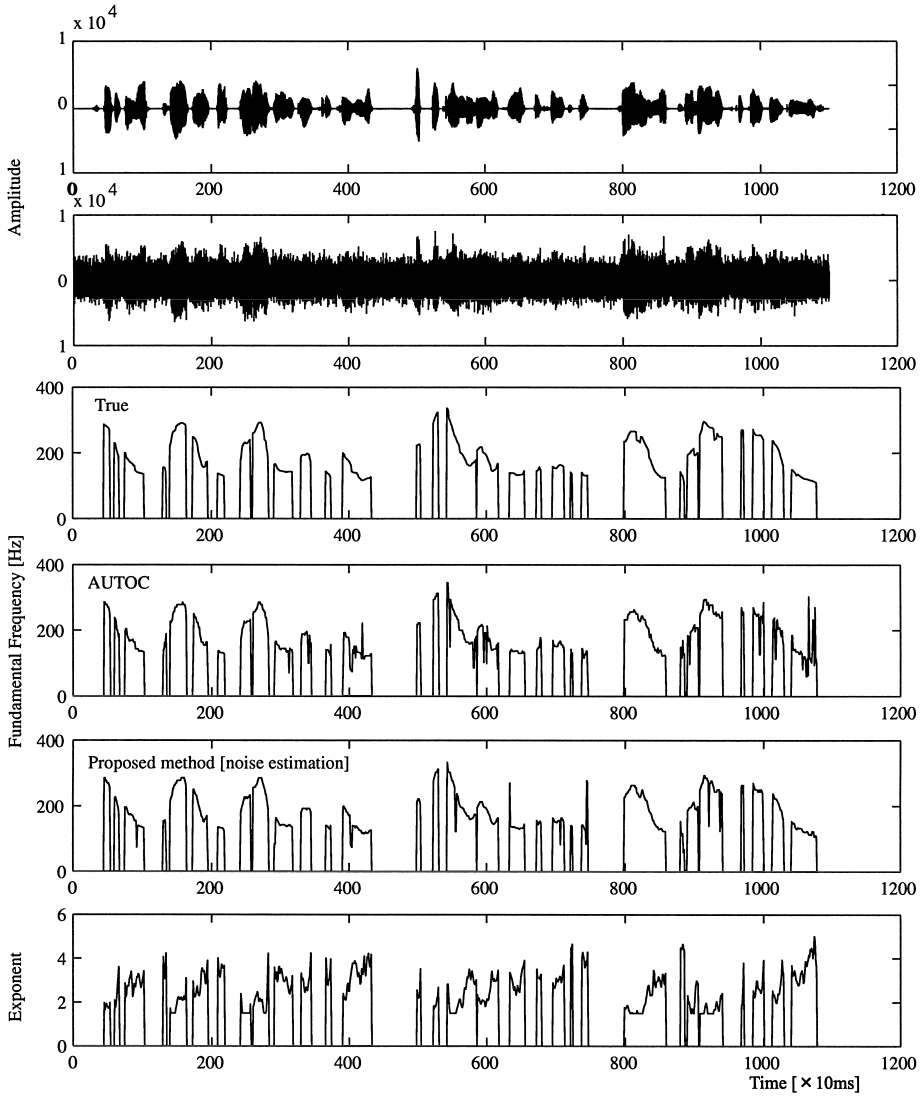

図 11　基本周波数抽出の比較
Fig. 11　Comparison of fundamental frequency detections.

満足されるものであることを裏づけているといえる．一方，WAUTOC に関しては，文献 [11] で示されている実験結果より AUTOC に対する改善を本実験では与えていないことに気づく．これは，WAUTOC が短い分析フレームにおいてより効果を発揮する方法であることを示唆している．文献 [11] では，本実験と同様白色雑音環境下で実験されているが，10 kHz のサンプリング周波数で 25.6 ms の分析フレーム長で WAUTOC が用いられたことに注意すべきである．

予備実験の結果の図 8 を図 9 及び図 10 と比べると，図 8 では，p の値を変えても SNR の向上とともに抽出成功率がほぼ 100%に近づいており，また図 9 及び図 10 では，AUTOC で GPE がたかだか女性の場合に 10%程度で抑えられ，ほぼ抽出に成功している．したがって，結果がほぼ合致していることがわかる．しかし，低 SNR 環境になるにつれ，特に SNR-5 dB の場合には，結果が大きく異なっている．図 8 では抽出成功率が 25%以下であるが，図 9 及び図 10 では AUTOC で GPE が女性の場合に 30%程度，すなわち抽出成功率は 70%程度となっている．この原因は，双方の実験での基本周波数抽出における実質的な SNR の違いにあると考えられる．白色雑音が周波数上で平たんなため，帯域制限された SNR，SNR_{band} と全帯域を考慮する SNR に本質的な差異は生じないとみなせる．しかし，図 8 で使用した音声，母音の中には無音区間が含まれていないが，図 9 及び図 10 で使用した連続音声の中には多くの無音区間が含まれている（一例が図 11 に示される）．SNR は，全音声信号パワーに対する全雑音パワーの比で算出されるため，考察対象の有声音区間での実質的な SNR は，図 9 及び図 10 の場合において比較的高くなっているはずである．これが低 SNR の場合に顕著に現れ，ここでの抽出成功率の差異につながったと考えられる．

図 11 は評価音声データ中のある女性話者に対する基本周波数抽出の結果を示している．SNR は 0 dB で，発話内容は “人々の屏風絵と如来像に対する興味は，800 年の年月によって生じた表面の微妙な色彩変化にある” である．図 11 は，上から雑音が付加されていない音声波形，雑音が付加された音声波形，本来の基本周波数，AUTOC による基本周波数抽出，雑音推定をする提案法による基本周波数抽出を表している．AUTOC と提案法の GPE はそれぞれ 24.38%と 12.28%である．処理波形及びこれらの数値より，明らかに提案法が優れていることがわかる．

図 11 の最下段は処理されたべき乗数の変化を示している．ここでは，抽出誤差が改善されているフレームでの処理が $p > 2$ となっていることに気づく．すなわち，提案法は，従来法である AUTOC よりも大きくスペクトル拡張を行っているわけである．したがって，大きくスペクトル拡張を行うことで更なる雑音低減を可能としたことがわかる．しかし一方で，多少ではあるが，新たな抽出誤差を招いている部分も見受けられる．これはホルマント強調による悪影響だと考えられる．

図 12 は連続音声における FPE の男女の平均結果を示している．CEPST は，無雑音環境下で良好な結果を与えるが，低 SNR になるにつれて大きく特性劣化する．その変化の幅は，SNR の -5 dB から ∞ dB の範囲で 5〜6 Hz になる．AUTOC，WAUTOC においても，これとほぼ同様の傾向がある．しかし，提案法は，高 SNR 環境下において CEPST 及び AUTOC，WAUTOC に劣るものの，SNR の -5 dB から ∞ dB の範囲での変化の幅は 2 Hz 程度である．また，低 SNR 環境下においては，従来法に比べ良好な結果を与えている．したがって，提案法は，SNR の広い範囲にわたって効果的であると考えられる．

提案法の，特に高 SNR 環境下での FPE の増大は，振幅スペクトルに施される帯域制限の処理の結果と解釈できる．なぜなら，この帯域制限処理は，振幅スペクトル上での周期性情報を大幅に損失させるためである．したがって，抽出精度をより向上させるには，帯域制限の幅を広げることが考えられる．しかし，この

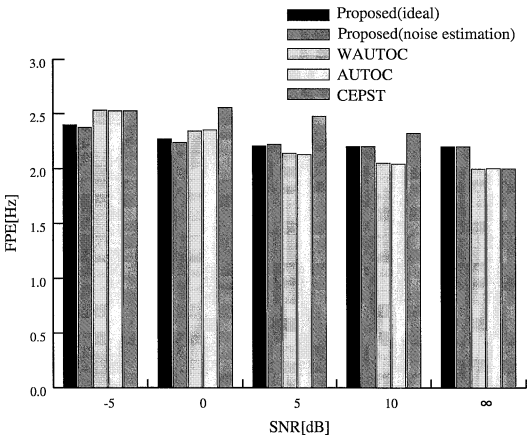

図 12　FPE の結果
Fig. 12　Results of FPE.

ようにすることにより，処理する帯域内の雑音量が増大することから，上記までの雑音耐性は多少なりとも劣化することが予想される．したがって，トレードオフを考慮に入れる必要があると思われるが，そのような提案法における帯域制限処理と実行精度のより詳細な検討に関しては，今後の課題としたい．

5. むすび

本論文では周波数領域におけるスペクトル変形に着目し，帯域制限されたべき乗振幅スペクトルに基づく基本周波数抽出法を提案した．本法は，分析フレームごとに入力信号の帯域制限された SNR に応じてべき乗数を決定する．低 SNR 環境下ではべき乗数を大きくとり，スペクトル拡張し，高 SNR 環境下ではべき乗数を小さくとり，帯域制限を利用してスペクトル圧縮を行う．これらの処理により，幅広い範囲の SNR で抽出精度を保ちつつ，良好な雑音耐性が得られるようになる．

提案法は，スペクトルの拡張圧縮のためのべき乗数決定に，音声強調法であるスペクトル引き算法を利用する．この場合，対象雑音の定常性が前提とされる．また，スペクトル拡張における雑音のスペクトル平たん性の制約から，対象雑音は白色雑音となる．したがって提案法は，定常白色雑音環境下で効果的な基本周波数抽出法となり得る．しかし，通信システム系でしばしば見られるような，白色雑音と正弦波周期雑音が同時に混在する場合などにも，提案法は直接利用できる．正弦波周波数が帯域制限される 50〜400 Hz 以外に存在するなら，その振幅レベルが比較的大きくても，提案法はそれを除去し得ると期待できる．今後は，種々の有色雑音に関して検討を行う予定である．

本論文では有声/無声判別を直接的に施さなかったが，自己相関関数法の延長上にある提案法は，有声/無声判別における特性改善も期待できる．今後詳しく調べていく予定である．

文　献

[1] W.J. Hess, Pitch Detemination of Speech Signal, Spring-Verlag, Berlin, 1983.

[2] W.J. Hess, “Pitch and voicing determination,” in Advances in Speech Signal Processing, ed. S. Furui and M.M. Sondhi, pp.3–48, Marcel Dekker, 1992.

[3] K.A. Oh and C.K. Un, “A performance comparison of pitch extraction algorithms for noisy speech,” Proc. IEEE Int. Conf. Acoustics Speech and Signal Processing, pp.18B4.1–18B4.4, 1984.

[4] L.R. Rabiner, “On the use of autocorrelation analysis for pitch detection,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-25, no.1, pp.24–33, Feb. 1977.

[5] L.R. Rabiner and R.W. Shafer, Digital Processing of Speech Signals, Prentice-Hall, New Jersey, 1978.

[6] 板倉文忠，斎藤収三，“最ゆうスペクトル推定法をもちいた音声情報圧縮，” 音響誌，vol.27, no.9, pp.463–472, 1971.

[7] 古井貞熙，ディジタル音声処理，東海大学出版会，1985.

[8] 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.

[9] S.F. Boll, “Suppression of acoustic noise in speech using spectrum subtraction,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-27, no.2, pp.113–129, April 1979.

[10] A.M. Noll, “Cepstrum pitch determination,” J. Acoust, Soc. Am., vol.41, pp.293–309, Feb. 1967.

[11] T. Shimamura and H. Kobayashi, “Weighted autocorrelation for pitch extraction of noisy speech,” IEEE Trans. Speech Audio Process., vol.9, no.7, pp.727–730, Oct. 2001.

（平成 14 年 11 月 19 日受付，15 年 5 月 6 日再受付，
7 月 9 日最終原稿受付）

島村　徹也　（正員）

昭 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．平 10 同助教授，現在に至る．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（ともに連合王国）客員研究員．ディジタル信号処理とその音声，通信システムへの応用に関する研究に従事．IEEE，EURASIP 各会員．

高木　浩司　（正員）

平 12 埼玉大・理・物理卒．平 14 同大大学院前期博士課程了．同年（株）NTT ドコモ入社，現在に至る．在学中，音声信号処理に関する研究に従事．