

ピッチ同期加算処理を用いた雑音低減に基づく LPC 分析

島村 徹也^{†a)} 黒岩世進伸^{†*}

LPC Analysis Based on Noise Reduction Using Pitch Synchronous Addition

Tetsuya SHIMAMURA^{†a)} and Yoshinobu KUROIWA^{†*}

あらまし 音声信号に白色雑音が入り込んだ場合、自己相関法による LPC 分析の精度を向上させるには、自己相関関数上で雑音パワーを除去する雑音補正法が有効と考えられる。しかし、この方法では、雑音パワーの引き過ぎにより、得られる AR フィルタが不安定になる場合がある。この問題を解決するために、本論文では、音声の短時間区間における定常的性質に着目したピッチ同期加算処理を提案し、この処理を音声のピッチ同期 LPC 分析の雑音低減に適用する。そしてその有効性を計算機シミュレーションで確認するとともに、ピッチ周期の異なる音声に適用した結果を検討し、提案するピッチ同期 LPC 分析法が、とりわけピッチ周期の短い音声に対して優れた効果を発揮することを明らかにする。また連続音声へも適用し、信号対雑音比が 20 dB 以下の場合に特に有効であることを示す。更には、その提案法にピッチ周期性の拡張を施すことにより、雑音環境下における推定精度が更に向上する可能性があることを示唆し、その拡張法の有効性をも明らかにする。

キーワード LPC, 定常性, ピッチ周期, 同期加算, S/N 改善

1. ま え が き

線形予測符号化 (LPC) [1] は、少ないパラメータ数で音声信号を正確に表現することを目的として開発された方式である。スペクトル分析をはじめ、フォルマント周波数抽出、ピッチ抽出など、その幅広い有用性から、LPC は現在音声処理において広く用いられている [3]。

LPC 分析は、基本的には線形予測モデルにおける予測係数を求めることに帰着される。雑音を含まない音声の場合には、自己相関法 [2] を用いて、効率的かつ高精度に予測係数を求めることができる。また、自己相関法は、形成される相関行列の正定値性により、得られる自己回帰 (AR) フィルタの安定性を保証することが可能である。この場合、次式で表される AR フィルタの伝達関数

$$H(z) = \frac{G}{1 - \sum_{i=1}^M a_i z^{-i}} \quad (1)$$

のすべての極は単位円内におさめられる。式 (1) で、 a_i と G はそれぞれ求められる予測係数と利得に対応している。分析合成系などに代表されるような、音声合成を伴う音声処理システムにおいては、この AR フィルタの安定性を保証する性質は極めて重要である。

しかし、音声に雑音が入り込んだ場合には、自己相関法の分析精度は大幅に低下してしまう [4]。この問題を解決するために様々な方法 [5] ~ [9], [21] が提案されているが、自己相関法が有する計算の効率性を損なわずに、雑音耐性の改善と推定される AR フィルタの安定性の保証を両立する、確約されるべき方法はいまだ存在しないと思われる。

具体的に自己相関法を用いて予測係数を求める際には、音声の自己相関関数 $R(k)$ のラグ数が $k = 0 \sim M$ (ここで M は予測次数に対応する) の部分を利用する。ここで、 $R(k)$ から雑音成分の除去ができれば、LPC 分析の特性改善が期待できる [6], [7]。従来の LPC 分析における雑音低減の有力な手法は、このような雑音補正の考え方に基づくものである。特に、雑音が広帯域かつランダム性を有する白色雑音の場合には、音声信号と白色雑音が互いに無相関となることから、雑音パワーは $R(0)$ 付近に集中し、 $R(0)$ から雑音パワーを引くことにより、原理的に LPC 分析の精度が改善

[†] 埼玉大学工学部情報システム工学科, さいたま市

Department of Information and Computer Sciences, Saitama University, 255 Shimo-Okubo, Saitama-shi, 338-8570 Japan

* 現在, (株) コルグ

a) E-mail: shima@sie.ics.saitama-u.ac.jp

できる [8]。しかし、このような雑音補正による手法では、雑音パワーの推定を正確に行わない限り、雑音パワーの引き過ぎで相関行列の正定値性を失い、結局得られる AR フィルタを不安定性にしてしまうことがある。現実問題、付加される雑音の非定常性により、無音区間からの雑音パワーの推定値と分析フレーム内に混入する雑音パワーの大きさは多くの場合異なると考えられる。そこで、文献 [10] では、分析フレーム内で雑音パワーを推定する方法を提案している。しかし、このような方法を用いたとしても、雑音補正に基づく LPC 分析の後に得られる AR フィルタの安定性は、必ずしも保証されない。

一方、LPC の次数を増大させる文献 [5] の方法では、AR フィルタの安定性は容易に保証できる。しかし、本質的に雑音環境下での特性改善はあまり望めない。また、ピッチ周期を手掛りに、自己相関関数のラグ数の低い部分を取り除く文献 [9] の方法では、AR フィルタの安定性を保証しようとするほど、耐雑音性が低減される結果となる欠点を有する。更に、本来音声強調手法として導出された方法ではあるが、ウィナーフィルタを音声信号の AR スペクトルを用いて反復的に設計する方法がある [21]。この方法では、間接的に LPC 分析を行うことが可能であり、自己相関法を AR スペクトル推定に用いれば、AR フィルタの安定性も保証できる。しかし、基本的に分析フレーム外の無音区間から雑音パワースペクトルを求める必要があり、分析フレームのみでの処理は困難である。また、反復処理を繰り返すことにより、得られる分析結果はむしろ悪化する可能性があることが報告されている [22]。このような従来の LPC 分析の雑音対策の背景に鑑み、本論文ではピッチ周期を利用する新しい雑音低減手法、ピッチ同期加算処理法を提案し、それを従来の自己相関法に適用することにより、LPC 分析における安定性の問題を解決する。そして、効率的な計算を保持しつつ、雑音環境下での特性改善を図る。

本論文では、音声分析の中でも特に重要とされる有声音の分析を考えることにする。有声音分析は、ピッチ周期を利用するピッチ同期型分析とピッチ周期を利用しないピッチ非同期型分析に大別されるが、本論文では、ピッチ同期分析に LPC を用いるピッチ同期 LPC 分析を考察対象として取り上げることとする。これは、ピッチ同期 LPC 分析が声道及び音源分析に適し [1]、また音声通信システムにおけるより効率的な伝送を実現し得る [11] ためである。ピッチ同期 LPC 分

析に特化した雑音問題に関する報告は従来見られないが、上記の雑音下における LPC 分析の問題点は当然のことながらここでも当てはまる。したがって、本論文では、現実問題として考慮すべき、ピッチ同期 LPC 分析における付加雑音への一対応策を提示することを目的とする。

ピッチ同期分析においては、ピッチ周期の同期化が必要とされる。このピッチ同期化の方法としては、ピッチ周期の中の最大振幅値を与えるデータサンプルからの 1 ピッチ周期、またはそれ以内を分析フレームとする方法がしばしばとられる [11] ~ [13]。したがって、本論文でもこれを適用し、特に 1 ピッチ周期内の全音声データを積極的に利用する立場をとることとする。

本論文の構成は、本まえがきに続き、2. で提案するピッチ同期加算処理法及びそれを LPC 分析に適用する新しいピッチ同期 LPC 分析法を述べる。3. においては、類似性の観点から、ピッチ同期加算処理法を適応コムフィルタと比較する。続く 4. では、計算機シミュレーション実験による結果から提案法の有効性を示し、その特徴を明らかにする。また 5. では、ピッチ周期性を更に有効利用し、提案法をより改善する方法へと発展し、6. でいくつかの検討事項を述べる。そして最後に、7. で本論文をむすぶことにする。

2. ピッチ同期加算処理を用いた雑音低減

2.1 原 理

図 1 は、音声波形の一例であり、有声音の基本的な性質を示している。この図より、音声波形は時間とともに変化するが、適当な短い時間区間に分割すれば、

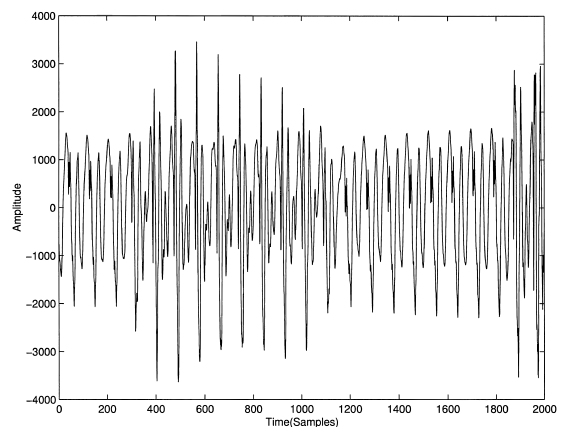


図 1 音声波形

Fig.1 Speech waveform.

それぞれの区間内において性質がほぼ一定であることが分かる．すなわち，それはピッチ周期に対応した周期波形となっている．このような音声の定常的な性質は，おおよそ 20~25[ms] 程度の短時間区間で成り立つといわれている [1]．一方，音声信号に付加される雑音は，音声通信システムなどにおいては，広帯域性を有するランダム雑音，すなわち白色雑音が典型的である [4]．このような付加雑音は，一般的には非定常確率過程と考えられるが，短時間区間内では定常確率過程と考えてよい．したがって，上記の音声信号の定常性が成立する区間，すなわち 20~25[ms] 程度の短時間区間内においては，雑音低減の基本的手法の一つである同期加算法 [19] を音声のピッチ周期に同期して適用すれば，音声波形の性質を保持できると同時に，各ピッチ周期ごとの雑音の無相関性を利用して，効率良く雑音成分を減少できると考えられる．そして，その処理された音声信号に自己相関法を適用すれば，より正確な予測係数の算出と AR フィルタの安定性の保証を同時に満足する結果を得ることができる．更に，このようなピッチ同期加算による雑音低減は，基本的に分析フレーム内で処理されるため，時間的に緩やかに特性が変化する非定常の雑音に対しても有効に働き，精度的に安定した分析結果を得ることが可能となる．

2.2 ピッチ同期加算処理

具体的にピッチ同期加算処理を以下に記述する．

今，音声信号を $s(n)$ ，付加される雑音（以下では，白色雑音を仮定することにする）を $w(n)$ で表すことにする．このとき，雑音が加わった音声信号は

$$x(n) = s(n) + w(n) \quad (2)$$

と表せる．この雑音混入音声信号 $x(n)$ が，ある定常区間内で与えられるとする．このとき，1 ピッチ周期ごとに波形を分割し，それぞれを 1 ブロックとすると

$$x_i(j) \quad i = 0, 1, 2, \dots, K-1, j = 0, 1, 2, \dots, P-1 \quad (3)$$

と表すことができる．ここで， j は 1 ブロック内に含まれる音声データの時間に対応しており， P は 1 ブロック内に含まれるデータサンプルの数を表している．また， i は，分析フレーム内におけるブロックの番号に対応しており， K は分析フレーム内に含まれるピッチ周期の数を表している．このとき，ピッチ同期加算処理は

$$x_0(j), x_1(j), x_2(j), \dots, x_{K-1}(j) \quad (4)$$

の集合平均で表され

$$z(j) = \frac{1}{K} \sum_{i=0}^{K-1} x_i(j) \quad j = 0, 1, 2, \dots, P-1 \quad (5)$$

となる．式 (5) は， K による割算を含んでいるため，実質的には加算平均を求めていることになる．

2.3 加算回数と S/N 改善度の関係

式 (5) によるピッチ同期加算処理によって得られる S/N 改善度を考えてみる．

一般に，信号対雑音比 (SNR) は次式で定義される．

$$SNR = \frac{\sum_{i=0}^{N-1} s(n)^2}{\sum_{i=0}^{N-1} w(n)^2} \quad (6)$$

ただし，ここでは N 個のデータサンプルからなる分析フレームを考えている．

ところで， $x_i(j)$ を音声と雑音に分けて表現すれば

$$x_i(j) = s_i(j) + w_i(j) \quad (7)$$

と表すことができる．ここで， $s_i(j)$ ， $w_i(j)$ における i と j は， $x_i(j)$ におけるそれと同じである．このブロック化された信号の SNR を SNR_1 とおくと

$$SNR_1 = \frac{\sum_{j=0}^{P-1} \{s_i(j)\}^2}{\sum_{j=0}^{P-1} \{w_i(j)\}^2} \quad (8)$$

と表すことができる．また，ピッチ同期加算処理は， $x_i(j)$ の $i=0$ から $K-1$ までの K 回の平均処理であり，その平均値 $x_{ave}(j)$ は

$$\begin{aligned} x_{ave}(j) &= \frac{1}{K} \sum_{i=0}^{K-1} x_i(j) \\ &= \frac{1}{K} \sum_{i=0}^{K-1} \{s_i(j) + w_i(j)\} \end{aligned} \quad (9)$$

で表せる．音声にピッチ周期ごとの完全な周期性があるとすれば， $s_i(j)$ は i に関係なく一定となる．一方， $w_i(j)$ は白色雑音の瞬時値であるから，全くランダムな値をとる．したがってこのような状況下で，式 (9) は

$$\lim_{K \rightarrow \infty} x_{ave}(j) = s_i(j) \quad (10)$$

となり、至るところ、無限回の同期加算によって雑音を含まない音声信号が求まることになる。しかし実際には、音声の定常性は有限時間においてのみ成立するため、分析フレーム内における無限回の加算は不可能である。結局、 K 回の同期加算処理が施されるとすると、音声信号は波形変化がないのに対して、雑音は、統計的な標本平均の確率分布の性質から、パワーが $1/K$ 倍に変化すると考えられる。この場合、SNR は

$$\begin{aligned} SNR_2 &= \frac{\sum_{j=0}^{P-1} \{s_i(j)\}^2}{\frac{1}{K} \sum_{j=0}^{P-1} \{w_i(j)\}^2} \\ &= K \frac{\sum_{j=0}^{P-1} \{s_i(j)\}^2}{\sum_{j=0}^{P-1} \{w_i(j)\}^2} \end{aligned} \quad (11)$$

となり、 S/N 改善度は

$$\Delta SN = \frac{SNR_2}{SNR_1} = K \quad (12)$$

となる。したがって、 K 回の同期加算処理により S/N は K 倍改善されることになる。

式 (12) は、分析フレーム内の音声信号が完全なピッチ周期ごとの周期性を有する前提のもとに導かれる結果である。現実には、音声はピッチ周期ごとに微妙な波形変化を伴うと考えられる。したがって、厳密には式 (12) は成り立たないかもしれない。しかし、4. で示されるシミュレーション結果は、LPC 分析におけるピッチ同期加算処理の有効性を明らかに裏づけている。

2.4 ピッチ同期加算処理の LPC 分析への適用

本論文では、従来のピッチ同期 LPC 分析法に、2.2 でのピッチ同期加算処理を定常とみなせる区間内で施す方法を提案する。その具体的な手順は以下のとおりである。

(1) 与えられる音声信号のピッチ周期をもとに、20~25[ms] 相当の区間内で、最大振幅を与えるデータサンプルからのピッチ周期ごとに音声信号を分割する。

(2) (1) の結果得られた信号に、ピッチ同期加算処理を施す。

(3) (2) の結果得られた 1 ピッチ周期信号に自己相関法を施し、予測係数 a_i を求める。

パワースペクトルを求める際には、自己相関法の具体的な解法である Levinson-Durbin アルゴリズムで得られる予測誤差を G^2 と対応させる。そして、式 (1) から

$$P(f) = |H(e^{j2\pi f/f_s})|^2 \quad (13)$$

のようにして求める。ここで f_s はサンプリング周波数を表している。

3. ピッチ同期加算処理と適応コムフィルタとの関係

前章で述べたピッチ同期加算処理は、音声強調法である適応コムフィルタ [20] と類似性を有する。適応コムフィルタは

$$y(n) = \sum_{k=-L}^L b_k x(n + Q_k) \quad (14)$$

のように雑音混入音声信号 $x(n)$ をフィルタリングする方法である。式 (14) で、 b_k はフィルタ係数、 Q_k はピッチ周期、 L は後ろ向き及び前向きに考慮されるピッチ周期の数を表している。

式 (14) は、 $x(n)$ の 1 サンプル入力ごとに $y(n)$ の 1 サンプルを出力する適応処理方式である。その適応化は、 Q_k をピッチ周期の変化に応じて調整することによってなされる。このとき、 b_k はあらかじめ決められたフィルタ長 $2L + 1$ において、固定されて用いられる。文献 [20] では、大きさが正規化されたハミング窓関数係数が b_k として用いられた。

式 (14) の特別な場合として、 $Q_k = Q$ (一定) において、 b_k を

$$b_k = \frac{1}{2L + 1} \quad (15)$$

とする場合を考えてみる。このとき、適応コムフィルタにはもはや音声の時間変動性への追従能力がなく、音声強調法としては実行精度が期待できないが、数式上では擬似的にピッチ同期加算処理とみなせる。この点に関して、前章でのピッチ同期加算処理は適応コムフィルタと類似性を有することになる。しかし、適応コムフィルタには分析フレームの概念がないために、算出されるピッチ周期波形は、ピッチ同期加算処理の場合とは基本的に異なる。また、ピッチ同期加算処理は本来デジタル信号処理の同期加算処理に動機されるものであり、音声信号に対して可能な範囲で同期加

算を行うことを目的とする。ゆえに、適応コムフィルタにおいて $Q_k = Q$ (一定) が成り立つ範囲で、最大限 L を増大することに対応することになる。

ピッチ同期加算処理は短時間区間での一括処理法である。この性質により、ピッチ同期 LPC 分析に望まれる 1 ピッチ周期のみの音声信号を直接的に与えることができる。しかし、適応コムフィルタではこれは不可能である。この点がピッチ同期加算処理と適応コムフィルタとにおいて根本的に異なる。

4. 計算機シミュレーション

本章では、提案する音声分析法の有効性を明らかにする目的で行った計算機シミュレーション実験の結果を示す。実験は次の二つに大別される。

1. 基本的性質の評価
2. 音声信号全体に対する評価

実験 1 では、単母音の実音声及び合成音声を用いて、提案法の同期加算回数及び SNR への依存性を調べた。また実験 2 では、連続音声を用いて、提案法の SNR への依存性を調べた。

4.1 シミュレーション条件

実験 1, 2 に共通した実験の諸元を表 1 に示しておく。また、雑音混入音声は式 (2) に基づき生成した。このとき、各雑音混入音声信号の SNR は、付加雑音の振幅値を変化させることにより調整した。

4.1.1 実験 1

実験 1 で用いた音声信号は、男性及び女性によって発声された母音/a/の 2 種類である。それぞれのピッチ周期は以下のとおりである。

- 男声：6.9[ms]
- 女声：3.7[ms]

また、実験結果に対する洞察を得るために、AR フィルタをインパルス列で励振して生成する合成母音（文献 [12]）も用いた。

4.1.2 実験 2

実験 2 では、連続音声として、NTT アドバンステクノロジー (株) の「20 カ国語音声データベース」に収録されている、男性と女性それぞれによる発声データ

“人々の屏風絵と如来像に対する興味は、八百年の年月によって生じた表面の微妙な色彩変化にある。”を用いた。それぞれの音声データより算出されたピッチ周期の平均値は

- 男声：8.1[ms]
- 女声：6.2[ms]

である。

4.2 シミュレーション手順

4.2.1 実験 1

実験 1 の手順は以下に示すとおりである。

まず、用意された雑音混入音声信号の定常とみなせる区間で分析フレームを取り出した。このとき、ピッチ周期内の最大値がその分析フレームの最初のデータサンプルになるようにし、また、その分析フレームの長さはピッチ周期の整数倍になるようにした。具体的にここで用いた分析フレームのデータサンプル数は

- 男声： $69 \times 3 = 207(20.7[\text{ms}])$
- 女声： $37 \times 6 = 222(22.2[\text{ms}])$

である。上記の括弧内にはサンプル数に対応する時間が表されている。

このように切り出された分析フレーム内の最初の 1 ピッチ周期分のデータサンプルに対して提案法を適用し、その分析精度を分析フレーム内の全データサンプルに対して適用された従来法の分析精度と比較した。このとき、提案法を実行する際には、分析フレーム内でピッチ同期加算処理をピッチ周期の整数倍の回数だけ実行した。具体的には、男声データにおいて 3 回、女声データにおいて 6 回の同期加算をここでは実行した。

4.2.2 実験 2

実験 2 の手順は実験 1 と基本的に同じである。しかし、連続音声には無声区間が含まれるため、分析する際にはあらかじめ有声/無声区間の分別を行い、無声区間を除去しておいた。また、連続音声のピッチ周期は時間とともに変化するので、各分析フレームごとにあらかじめピッチ周期を求めておき、実験 1 と同様の同期加算処理を行うことにした。

4.3 シミュレーション結果

4.3.1 実験 1 の結果

図 2 に男声の場合におけるシミュレーションの結果を示す。ここで $S(f)$, $X(f)$, $Z(f)$ は、原音声信号、雑音混入音声信号 (SNR=5[dB])、同期加算処理後の音声信号に対してそれぞれ自己相関法を適用した結果、得られたパワースペクトルを表している。パワースペ

表 1 実験の諸元

Table 1 Conditions for experiment.

標本化周波数	10 kHz
LPC 次数	10
分析窓	方形窓
付加雑音	白色性ガウス雑音

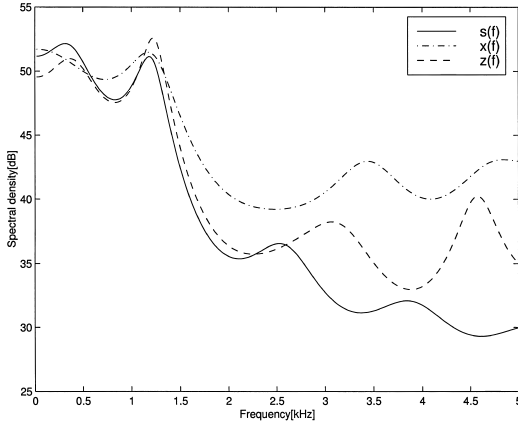


図 2 男性母音/a/における LPC スペクトルの比較
Fig. 2 Comparison of LPC spectra for male vowel /a/.

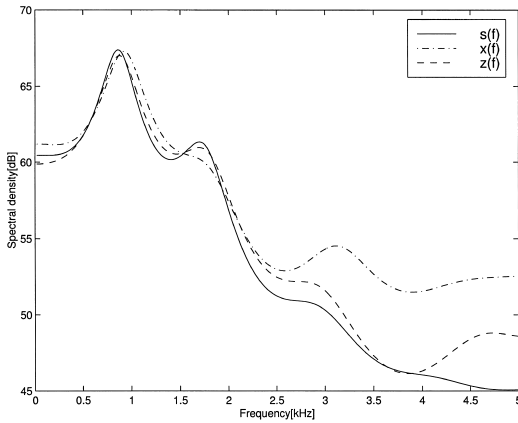


図 3 女性母音/a/における LPC スペクトルの比較
Fig. 3 Comparison of LPC spectra for female vowel /a/.

クトルの算出には式 (13) が用いられている。図 2 においては、明らかに $Z(f)$ が $X(f)$ に比べ $S(f)$ に近づいている。また、雑音混入音声信号からは正しく検出されていない第 1 ホルメントが、同期加算処理により検出されるようになったことが分かる。

次に女声での場合の結果を図 3 に示す (ここでの場合、雑音混入音声信号の SNR は 10[dB] である)。図 3 においても、 $Z(f)$ が $X(f)$ に比べ $S(f)$ に近づいている。また、雑音混入音声信号からは正しく検出されていない第 2 ホルメントが、同期加算処理により検出されるようになったことが分かる。

これらの分析結果を定量化するために、次のスペクトル誤差

表 2 スペクトル推定誤差 [dB]
Table 2 Spectral estimation errors [dB].

	E_x	E_z
男声 (図 2)	16.92	13.54
女声 (図 3)	12.39	7.44

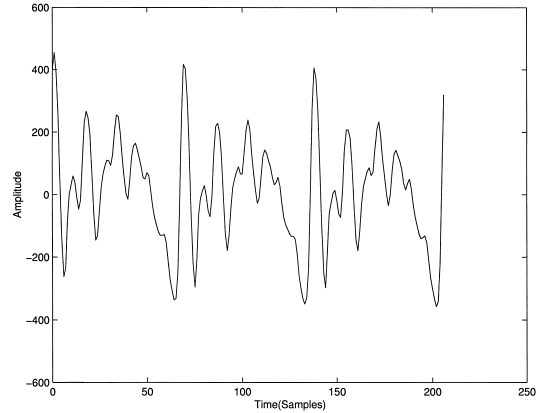


図 4 男性母音/a/
Fig. 4 Male vowel /a/.

$$E_x = 10 \log_{10} \left(\int_0^{f_s/2} \frac{|X(f) - S(f)|}{S(f)} df \right) \quad (16)$$

$$E_z = 10 \log_{10} \left(\int_0^{f_s/2} \frac{|Z(f) - S(f)|}{S(f)} df \right) \quad (17)$$

を算出してみた。図 2 及び図 3 でのそれぞれのスペクトル推定誤差が表 2 にまとめてある。表 2 より、男声、女声ともに、提案法により推定誤差が低減されているのが分かる。また、男声より女声の方が推定精度がより改善されていることが見てとれる。

図 4~図 6 に、図 2 のスペクトル結果を与えた原音声信号、雑音混入音声信号、ピッチ同期加算処理後の音声信号を、図 7~図 9 に、図 3 のスペクトル結果を与えた原音声信号、雑音混入音声信号、ピッチ同期加算処理後の音声信号をそれぞれ示しておく。ただし、図 6 及び図 9 には、算出された 1 ピッチ周期の波形をピッチ周期ごとに繰り返し使い、分析フレーム長に復元した波形を示している。これらの図 4~図 6 及び図 7~図 9 の音声波形図より、ピッチ同期加算処理を施すことにより、原音声の波形は保持されつつ、雑音低減がなされていることが確認できる。

表 2 の結果に見られるところの、提案法が男声より女声の場合に改善度が大きい理由は、ピッチ同期加算の回数が女声の場合の方が多いためと考えられ

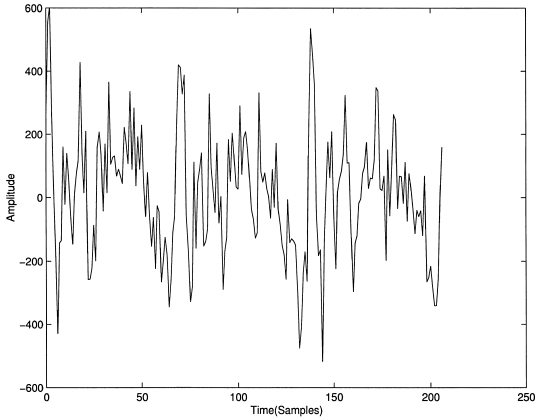


図 5 雑音が含まれた男性母音 /a/
Fig. 5 Male vowel /a/ corrupted by noise.

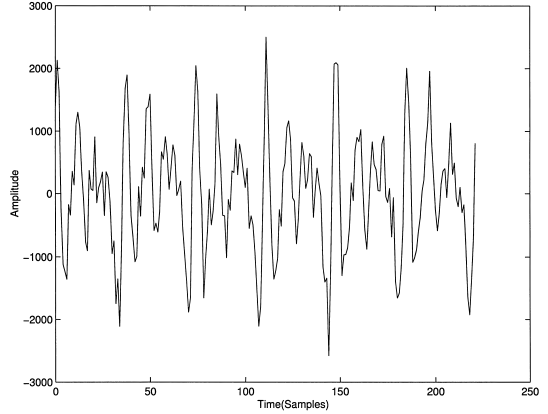


図 8 雑音が含まれた女性母音 /a/
Fig. 8 Female vowel /a/ corrupted by noise.

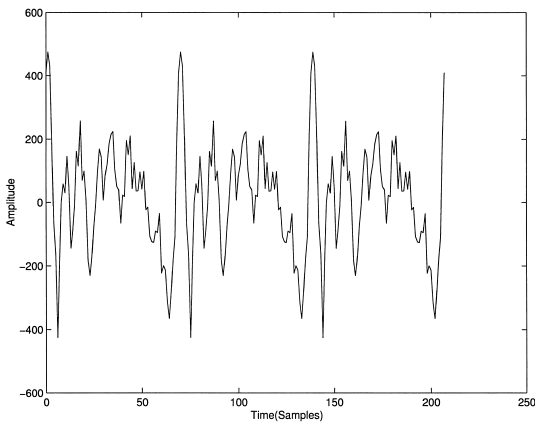


図 6 ピッチ同期加算処理後の男性母音 /a/
Fig. 6 Male vowel /a/ after pitch synchronous addition.

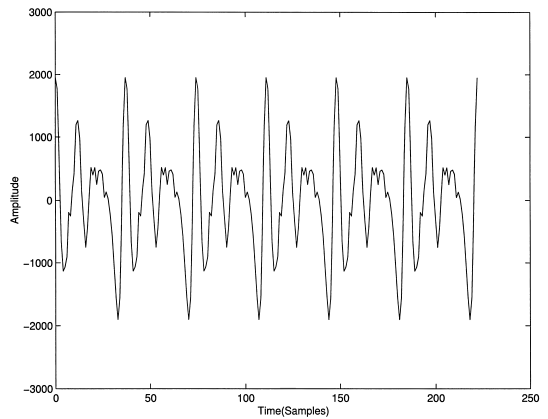


図 9 ピッチ同期加算処理後の女性母音 /a/
Fig. 9 Female vowel /a/ after pitch synchronous addition.

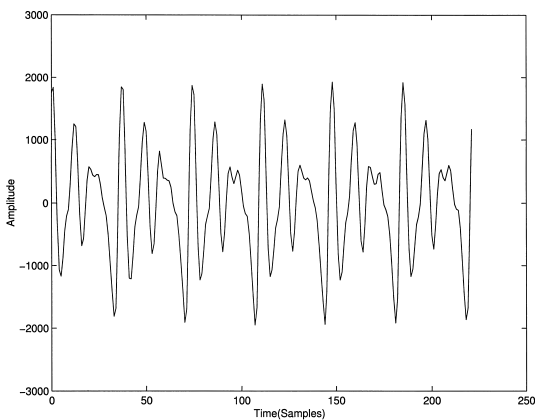


図 7 女性母音 /a/
Fig. 7 Female vowel /a/.

よう。実際、10 kHz のサンプリング周波数において、文献 [12] に示される方法で合成音 /o/ を生成し、定常とみなせる 240 データサンプル (24[ms]) において、ピッチ周期を

$$\{20, 30, 40, 60, 80, 120\}(\text{サンプル数})$$

と変化させた場合のそれぞれの S/N 改善度と加算回数との関係を調べてみた。図 10 がその結果を示している。この図 10 からは、ピッチ周期が短いほど S/N 改善度が大きいことが分かる。これは明らかに、ピッチ周期が短くなるにつれて、同期加算可能な最大回数が増加するためである。

図 10 の詳細な S/N 改善度 ΔSN の数値データは

$$\Delta SN(2) = 1.98, \Delta SN(3) = 3.01,$$

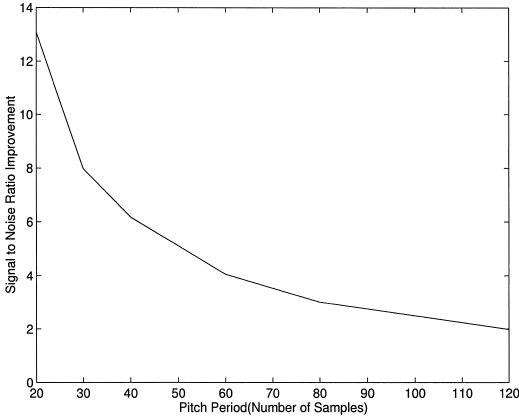


図 10 ピッチ周期と S/N 改善度の関係 (合成母音 /o/)
 Fig.10 Relation between the pitch period and the improvement in SNR (synthetic vowel /o/).

$$\Delta SN(4) = 4.05, \Delta SN(6) = 6.17, \\ \Delta SN(8) = 7.98, \Delta SN(12) = 13.08$$

である。ここで、上記の括弧内には分析フレーム内での同期加算回数が表してある。これらを、算出された S/N 改善度と比べてみれば、近似的にほぼ等しいことが分かる。このような同期加算回数と S/N 改善度の関係は、明らかに式 (12) を反映している。

次に、提案法の SNR 依存性を調べた結果を示しておく。ここでは、次式で算出される LPC ケプストラム距離で実行精度を評価した。

$$CD = \frac{10}{\ln 10} \sqrt{2 \sum_{i=1}^M (c_i - \hat{c}_i)^2} \quad (18)$$

ここで、 c_i は原音声信号から得られる予測係数のケプストラム、 \hat{c}_i は雑音混入音声信号から分析される予測係数のケプストラムである。雑音混入音声信号から直接 LPC 分析した場合と、ピッチ同期加算処理した後に LPC 分析した場合とで上記の LPC ケプストラム距離 CD をそれぞれ算出し、無雑音の場合での結果とどれだけ近くなるかをここでは評価した。先に用いた男声の場合において、120 個の分析フレームから得られた CD 値の平均値が図 11 に示してある。図 11 より、SNR に依存せず、提案法によってほぼ一定の CD 値の改善が得られていることが分かる。

注意深く図 11 を観察すると、SNR が 20[dB] に近づくとつれ、改善度が多少低下する現象が見受けられるが、これは、音声信号がピッチ周期ごとに、多少波

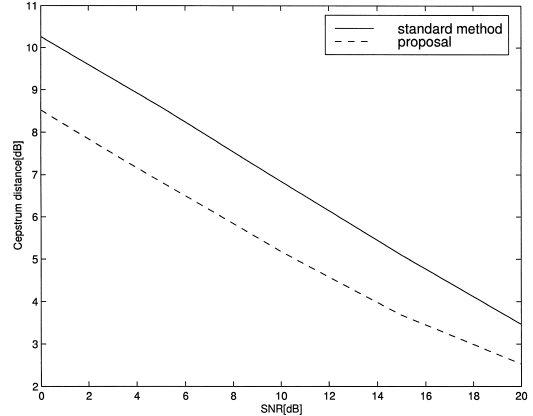


図 11 男性母音 /a/ の場合の LPC ケプストラム距離の比較
 Fig.11 Comparison of LPC cepstrum distances in case of male vowel /a/.

表 3 同期加算処理と適応コムフィルタの比較
 Table 3 Comparison of pitch synchronous addition with adaptive comb filter.

	SNR[dB]		
	0	5	10
CD (ピッチ同期加算処理)	7.527	6.068	4.677
CD (適応コムフィルタ)	7.745	6.231	4.728

形変化を起こしたために生じた結果と考えられよう。表 3 は、提案法に用いられるピッチ同期加算処理が適応コムフィルタと異なる結果を与えることを示す目的で行った実験結果である。先の男声のある 1 分析フレームを処理対象とした。適応コムフィルタには、式 (14) での L を $L=1$ と設定し、文献 [20] と同じフィルタを用いた。ただし、分析フレームを無視し、あらかじめ対応する音声データに適応処理を施した。SNR を変化させたときの、算出された CD 値が表 3 である。いずれの条件においても、ピッチ同期加算処理が良好な結果を与えていることが確認できる。

4.3.2 実験 2 の結果

図 11 と同様にして、図 12、図 13 には、連続音声に従来法と提案法を適用したときの、男声と女声の場合での LPC ケプストラム距離の算出結果をそれぞれ示している。ここでは、各分析フレームごとに式 (18) の LPC ケプストラム距離 CD を算出し、全フレームから得られた結果をすべて平均した値が SNR に対して示されている。図 12、図 13 より、男声、女声の双方の場合において、提案法を用いることによって LPC ケプストラム距離が小さくなっていることが分かる。

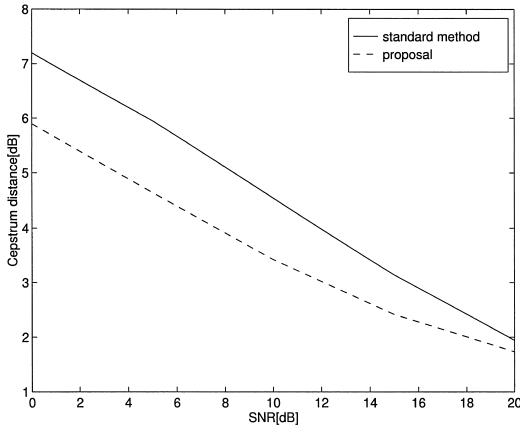


図 12 男性の連続音声の場合の LPC ケブストラム距離の比較

Fig. 12 Comparison of LPC cepstrum distances in case of male continuous speech.

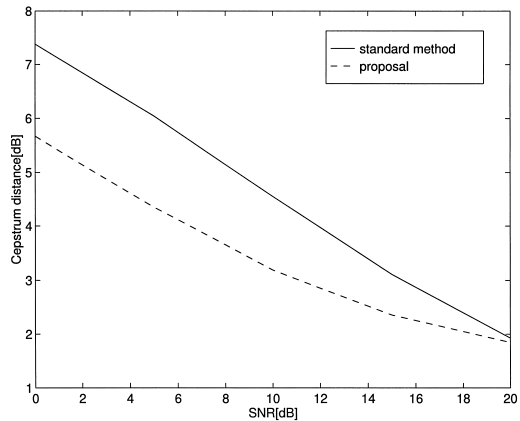


図 13 女性の連続音声の場合の LPC ケブストラム距離の比較

Fig. 13 Comparison of LPC cepstrum distances in case of female continuous speech.

したがって、提案法は連続音声に対しても有効であるといえよう。

図 12 を図 13 と比較すると、再び提案法は女声の場合に改善度が大きいことが分かる。これは、特に低 SNR において顕著である。SNR=0[dB] のとき、男声では約 1.3[dB] の従来法に対する改善があるのに対して、女声では約 1.8[dB] ほどの改善がある。

一方、図 12、図 13 には共通して、図 11 で見られた高 SNR になるにつれて改善度が少なくなる現象が、より顕著に現われている。これは、連続音声中に含まれる語頭、語尾や、音の移り変わりでの波形レベル及びその周期の長さの変化により、分析フレーム内の

ピッチ周期波形の周期性が乱されたためと考えられる。しかし、図 12、図 13 の結果は、SNR が 20[dB] 以下であれば、提案法は明らかに有効であることを示しているといえよう。

5. 周期性の拡張

2. で述べた提案法は、基本的に 1 ピッチ周期のデータサンプルに対して LPC 分析する方法である。3. ではその提案法の従来法に対する優位性が示されている。しかし、この 1 ピッチ周期データサンプルに基づく提案法は、文献 [12] の方法と同様にピッチ周期性を利用して更に高精度化され得る。本章では、その発展法について述べる。

5.1 特性改善法

文献 [12] では、音声信号 $s(n)$ の有するピッチ周期を P としたとき、 $s(n)$ に

$$s(n) = s(mP + n) \quad (19)$$

(m はある整数) の関係があることを利用し、ピッチ周期を拡張する LPC 分析法を導出した。具体的には、 $s(n)$ がピッチ周期に対応した長さ P の切り出しをうけるとき、通常の自己相関法を適用する際の自己相関関数の計算

$$R_{ss}(k) = \frac{1}{P} \sum_{n=0}^{P-1-k} s(n)s(n+k) \quad (20)$$

を施すより、その切り出された 1 ピッチ周期以外のデータサンプルも利用し、自己相関関数を

$$R'_{ss}(k) = \frac{1}{P} \sum_{n=0}^{P-1} s(n)s(n+k) \quad (21)$$

と求めることにより、LPC 分析の精度が向上することを示した。本章では、式 (2) で与えられる雑音混入音声信号 $x(n)$ に対して、2. での同期加算処理を施した後、上記と同様に自己相関関数を

$$R(k) = \frac{1}{P} \sum_{n=0}^{P-1-k} x(n)x(n+k) \quad (22)$$

と算出する代わりに

$$R'(k) = \frac{1}{P} \sum_{n=0}^{P-1} x(n)x(n+k) \quad (23)$$

と算出する方法を提案する。このとき、 $R'(k)$ から形

成される相関行列の Toeplitz 性は、 $R'_{ss}(k)$ のときと同様に保持されることになる。したがって、 $R'(k)$ を用いたとしても、Levinson-Durbin アルゴリズムにより予測係数を算出すれば、AR フィルタの安定性は保証されることになる。

5.2 付加雑音の影響

ここでの場合、分析される信号自体に完全なピッチ周期性がなく、

$$x(n) \neq x(mP + n) \quad (24)$$

の関係性を有する。したがって、式 (19) の性質を有する音声信号 $s(n)$ に対して式 (21) を施す場合と、式 (24) の性質を有する雑音混入音声信号 $x(n)$ に対して式 (23) を施す場合とは、その性質が異なると考えられる。そこで、以下に式 (23) が有する性質を解析的に調べておくことにする。

式 (23) に式 (2) を代入し、展開すると

$$\begin{aligned} R'(k) &= \frac{1}{P} \sum_{n=0}^{P-1} [s(n)+w(n)][s(n+k)+w(n+k)] \\ &= \frac{1}{P} \sum_{n=0}^{P-1} [s(n)s(n+k) + s(n)w(n+k) \\ &\quad + w(n)s(n+k) + w(n)w(n+k)] \\ &= R'_{ss}(k) + R'_{sw}(k) + R'_{ws}(k) + R'_{ww}(k) \end{aligned} \quad (25)$$

となる。ここで両辺の期待値をとれば、 $x(n)$ の中に含まれる音声信号 $s(n)$ と付加雑音 $w(n)$ の無相関性により、

$$E[R'(k)] = E[R'_{ss}(k)] + E[R'_{ww}(k)] \quad (26)$$

となる。一方、式 (22) に式 (2) を代入し、展開して期待値をとれば、同様にして

$$E[R(k)] = E[R_{ss}(k)] + E[R_{ww}(k)] \quad (27)$$

となる。ここで、付加雑音の項をそれぞれ算出すると、 $w(n)$ は白色雑音を仮定しているため、その分散 σ_w^2 を用いることにより

$$\begin{aligned} E[R'_{ww}(k)] &= E[w(n)w(n+k)] \\ &= \begin{cases} \sigma_w^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \end{aligned} \quad (28)$$

$$\begin{aligned} E[R_{ww}(k)] &= \frac{P-k}{P} E[w(n)w(n+k)] \\ &= \begin{cases} \sigma_w^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \end{aligned} \quad (29)$$

となる。したがって、付加雑音から受ける影響は、 $R'(k)$ 、 $R(k)$ とともに同程度となる。しかし、文献 [12] で示されているように、本質的に $E[R'_{ss}(k)]$ は $E[R_{ss}(k)]$ よりピッチ周期を有する音声信号に対して確からしい性質を有する。したがって、結局、付加雑音が存在する場合にも、ピッチ周期性を利用して式 (23) の自己相関関数推定を施すことにより、より高精度な結果が得られることになる。

5.3 シミュレーション

本法の有効性を明らかにするために、3. での実験 1 における合成音声を用いたシミュレーション結果を示しておく。ここでは、実音声のピッチ周期性は完全ではなく、推定結果の真値が不明確になる場合があると考え、ピッチ周期性が完全で、真値が明らかな合成音声を用いることにした。図 14 は、SNR が 7dB の音声信号からピッチ同期加算処理を施した後に、1 ピッチ周期のデータサンプルに対して、式 (23) から自己相関関数を求めたときの LPC 分析結果と、式 (22) から自己相関関数を求めたときの LPC 分析結果を比較して示している。また、与えられる音声データに直接 LPC 分析した結果も示している。いずれも、LPC 次数 $M = 10$ 、分析データ数 $N = 240$ 、ピッチ周期 $P = 80$ の条件下で求められている。結果として、ピッチ周期の拡張を施さなかった場合は真値とのスペク

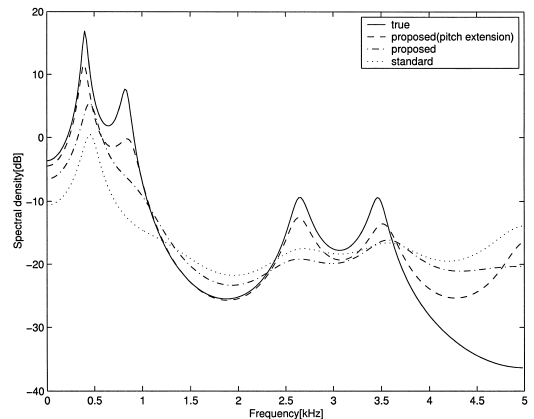


図 14 合成母音/o/における LPC スペクトルの比較
Fig. 14 Comparison of LPC spectra for synthetic vowel /o/.

トル推定誤差が 20.04[dB] で、ピッチ周期の拡張を施した場合は 16.17[dB] であった(ここでは式 (17) の $S(f)$ を真のスペクトルとして算出した)。また、従来の LPC 分析ではスペクトル推定誤差が 21.25[dB] であった(ここでは式 (16) の $S(f)$ を真のスペクトルとして算出した)。これらの結果から、ピッチ同期加算処理のみを施すことによって 1.21[dB] の改善が得られ、ピッチ周期の拡張により更に 3.87[dB] の改善が得られたことが分かる。したがって、自己相関関数の算出に式 (23) を適用することの効果の大きさが伺える。

6. 検 討

6.1 ピッチ周期の抽出について

本論文では、ここまでに音声データからの具体的なピッチ周期の自動抽出について述べてこなかった。これは、従来多くの報告があるためである。雑音環境下におかれるとしても、例えば咽喉マイクなどを用いてピッチ周期を正確に求めることは可能と思われる [14]。また、特にピッチ周期の短い女声に対して良好な結果を与える ACLOS [15] や平方根及び 4 乗根パワースペクトルを用いるピッチ抽出法 [16] などを併用すれば、提案法の有効性は更に発揮されると思われる。一方、最近では雑音に口バスタな声門閉鎖点の探索方法も研究されている [17], [18]。これらも、本論文での提案法を支援してくれるものと期待できる。

6.2 雑音特性について

対象とする雑音は白色雑音を考慮してきたが、文献 [7] にもあるようなプレホワイトニングを施すことにより、本論文での提案法は有色性の雑音にも対処可能となり得る。しかし、その具体的な実験結果等については別の機会に報告したい。

7. む す び

本論文では、雑音環境下におけるピッチ同期 LPC 分析の改善を図る目的において、短時間分析フレーム内における音声信号の定常性に着目したピッチ同期加算処理法を提案した。そして、自己相関法を直接利用することにより、雑音成分を低減しつつ、効率的に良好な分析結果を与え、同時に得られる AR フィルタの安定性を保証することに成功した。また、ピッチ周期性を更に有効利用することにより、雑音環境下における分析精度を更に改善する発展法を導出した。そして、その有効性を明らかにした。

実験結果では、ピッチ周期の短い音声に対して提案

法は効果的であることが明らかとなった。今後は、より多くの音声資料に基づく実験的検証及び特性のより複雑な付加雑音への対応策の検討などを行っていく予定である。

文 献

- [1] B.S. Atal and S.L. Hanauer, "Speech analysis and synthesis by linear prediction of the speech wave," *J. Acoust. Soc. Am.*, vol.50, no.2, pp.639-640, Aug. 1971.
- [2] J. Markel, "Digital inverse filtering—A new tool for formant trajectory estimation," *IEEE Trans. Audio Electroacoust.*, vol.AU-20, no.2, pp.129-137, June 1972.
- [3] J.R. Deller, Jr., J.G. Proakis, and J.H.L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan, 1993.
- [4] M.R. Sambur and N.S. Jayant, "LPC analysis/synthesis from speech inputs containing quantizing noise or additive white noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-24, no.6, pp.488-494, 1976.
- [5] J. Tierney, "A study of LPC analysis of speech in additive noise," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-33, no.6, pp.389-397, Aug. 1980.
- [6] S.M. Kay, "Noise compensation for autoregressive spectral estimates," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-28, no.3, pp.292-303, 1980.
- [7] S. Maitra, "Reducing the effect of background noise for low-bit-rate voice digitizers," *Proc. IEEE ICASSP*, pp.696-698, 1980.
- [8] H. Hu, "Noise compensation for linear prediction via orthogonal transformation," *Electron. Lett.*, vol.32, no.16, pp.1444-1445, 1996.
- [9] T. Shimamura, N. Kunieda, and J. Suzuki, "A robust linear prediction method for noisy speech," *Proc. IEEE ISCAS*, pp.IV257-IV260, 1998.
- [10] K.K. Paliwal, "Estimation of noise variance from the noisy AR signal and its application in speech enhancement," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-36, no.2, pp.292-294, 1988.
- [11] S. Chandra and W.C. Lin, "Linear prediction with a variable analysis frame size," *IEEE Trans. Acoust. Speech Signal Process.*, vol.ASSP-25, no.4, pp.322-330, 1977.
- [12] K.K. Paliwal and P.V.S. Rao, "A modified autocorrelation method of linear prediction for pitch-synchronous analysis of voiced speech," *Signal Process.*, vol.3, no.2, pp.181-185, 1981.
- [13] K.K. Paliwal, "Performance of the weighted Burg methods of AR spectral estimation for pitch-synchronous analysis of voiced speech," *Speech Commun.*, vol.3, no.3, pp.221-231, Dec. 1984.
- [14] 高杉敏男, 鈴木誠史, "高騒音下の音声の性質と雑音レベルの低減," *電波研究所季報*, vol.22, no.119, pp.197-213,

- 1976.
- [15] 國枝伸行, 島村徹也, 鈴木誠史, “対数スペクトルの自己相関関数を利用したピッチ抽出法,” 信学論 (A), vol.J80-A, no.3, pp.435-443, March 1997.
 - [16] 吉尾重治, 趙 奇方, 島村徹也, 鈴木誠史, “平方根及び 4 乗根パワースペクトルの自己相関に基づくピッチ抽出,” 信学論 (A), vol.J84-A, no.3, pp.436-440, March 2001.
 - [17] B. Yegnanarayana and R. Smits, “A robust method of determining instants of major excitations in voiced speech,” Proc. IEEE ICASSP, pp.776-779, 1995.
 - [18] P.S. Murthy and B. Yegnanarayana, “Robustness of group-deley-based method for extraction of significant instants of excitation from speech signals,” IEEE Trans. Speech Audio Process., vol.7, no.6, pp.609-619, Nov. 1999.
 - [19] 宮川 洋, ほか, デジタル信号処理, 電子情報通信学会編, 1975.
 - [20] J.S. Lim, A.V. Oppenheim, and L.D. Braida, “Evaluation of an adaptive comb filtering method for enhancing speech degraded by white noise addition,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-26, no.4, pp.354-358, Aug. 1978.
 - [21] J.S. Lim and A.V. Oppenheim, “All-pole modeling of degraded speech,” IEEE Trans. Acoust. Speech Signal Process., vol.ASSP-26, no.3, pp.197-210, June 1978.
 - [22] J.H. Hansen and M.A. Clements, “Constrained iterative speech enhancement with application to speech recognition,” IEEE Trans. Signal Process., vol.39, no.4, pp.795-805, April 1991.
- (平成 14 年 12 月 11 日受付, 15 年 7 月 1 日再受付,
12 月 12 日最終原稿受付)



黒岩世進伸

平 11 埼玉大・工・情報システム卒。同年 (株) コルグ入社, 現在に至る。在学中, 音声信号処理に関する研究に従事。



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 (ともに英国) 客員研究員。デ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。