



# HOKKAIDO UNIVERSITY

|                  |                                                                                   |
|------------------|-----------------------------------------------------------------------------------|
| Title            | 解析信号を用いた適応的ARMA-Dモデリング                                                            |
| Author(s)        | 堀田, 英輔; 宮永, 喜一; 栃内, 香次                                                            |
| Citation         | 北海道大學工學部研究報告, 168, 39-48                                                          |
| Issue Date       | 1994-02-21                                                                        |
| Doc URL          | <a href="https://hdl.handle.net/2115/42418">https://hdl.handle.net/2115/42418</a> |
| Type             | departmental bulletin paper                                                       |
| File Information | 168_39-48.pdf                                                                     |



## 解析信号を用いた適応的 ARMA-D モデリング

堀 田 英 輔    宮 永 喜 一    栃 内 香 次

(平成 5 年 10 月 21 日受理)

### An adaptive ARMA-D modeling using an analytic signal

Eisuke HORITA, Yoshikazu MIYANAGA and Koji TOCHINAI

(Received October 21, 1993)

#### Abstract

In this paper, a complex adaptive spectral estimation method of speech signals is proposed. Speech signals are transformed in this method to analytic signals. A pre-processing is applied to compensate the discontinuity of a spectrum in decimated analytic signals. Furthermore, tracking ability of the complex adaptive method is discussed. It is shown from the results of experiments that the proposed method has better tracking ability than conventional adaptive methods and estimates higher accurate power spectra of speech signals than conventional real adaptive methods.

#### 1. ま え が き

音声信号の特徴を抽出する手法が現在までに数多く提案されている<sup>(1)-(7)</sup>これらの手法は音声信号を線形確率モデルの出力であるとみなしている。音声分析に線形予測理論を利用するときには、母音の声道の伝達関数が全極フィルタに近似的に等価であるという理由から、主に AR モデルが用いられている<sup>(1)-(4)</sup>さらに、ARMA モデルを用いたいくつかの分析手法も子音の音声生成モデルの極と零点を推定するために議論されている<sup>(4)-(7)</sup>

線形予測モデルのパラメータを推定するために一括処理と適応的処理の両方が議論されてきた。適応的手法のほうが特徴の非定常表現に適していると考えられる。つまり、線形予測モデルの非定常パラメータを推定する適応的手法がいくつか提案されている<sup>(5)-(7)</sup>正確な音声情報処理システムを構築するときには、いくつかの問題がまだ残っている。その問題のひとつとして、時変音声スペクトルを高精度に推定することも考えられる。我々はこのことを目標にした手法の一つとして時変 ARMA-D モデルを用いた拡張 MIS 法を提案している<sup>(17),(18)</sup>

一方、スペクトルの推定精度の向上を期待できる別のアプローチとして、解析信号を用いる手法がある<sup>(10),(14)</sup>文献(10)では、この方針に基づき複素線形予測法を提案して、実数の線形予測法より精度の高い音声スペクトルを推定している。

本論文では、複素数の拡張 MIS 法が複素信号を分析するために議論される。そして、提案する複素アルゴリズムが従来の手法より音声信号のパワースペクトルを正確に推定することが示され

る。

以下、本論文では2. で提案する適応的複素アルゴリズムを示す。ここで提案するアルゴリズムは推定音声スペクトルの高域での誤差を減少させ、かつ、従来の、信号のデシメーションを伴う適応的複素処理に存在していたトラッキング能力の悪化という問題を改善している。そして、3. では合成音声と実音声に対して実験を行い、提案する手法は従来の適応的複素処理手法より優れたトラッキング能力をもち、従来の拡張 MIS 法より精度の高いスペクトルを推定できることを示す。

## 2. 解析信号を用いた時変モデリング

本節では、複素信号の特徴を適応的に抽出する手法を述べる。2. 1 では複素信号について簡単に述べる。そして、2. 2 では実信号の最高周波数  $\omega_{\max}$  のパワースペクトルの成分を周波数  $\omega=0$  での値に一致させる適応的前処理の手法について述べる。2. 3 では複素信号のパワースペクトルを推定する適応的手法について述べる。複素信号を用いる適応的手法は高精度なスペクトルを推定できるが、そのトラッキング能力は解析信号にデシメーションを行うため一般的に悪くなる。本論文で述べる手法ではそのトラッキング能力が改善される。

### 2. 1. 解析信号

本論文で導入される複素信号は解析信号である。解析信号は虚部に実部信号の Hilbert 変換値をもつ。いま、観測される実信号を  $\{y(k)\}$ 、その Hilbert 変換値を  $\{y_h(k)\}$  とすると、解析信号  $\{z(k)\}$  は

$$z(k) = y(k) + jy_h(k) \quad (1)$$

で表される。式(1)において、 $\{y_h(k)\}$  と  $\{y(k)\}$  は

$$y_h(k) = \sum_{i=-\infty}^{\infty} h(i)y(k-i)$$

で関係づけられる。ここで、 $\{h(i)\}$  は Hilbert 変換特性のインパルス応答を表し

$$h(i) = \begin{cases} \frac{2}{\pi i} \sin^2\left(\frac{\pi i}{2}\right), & i \neq 0 \\ 0, & i = 0 \end{cases}$$

で与えられる<sup>(8)</sup>

いま、実信号  $\{y(k)\}$  のフーリエ変換を  $Y(e^{j\omega})$  とすると、解析信号  $\{z(k)\}$  のフーリエ変換  $Z(e^{j\omega})$  は  $2\pi$  周期内で次のように表せる<sup>(8)</sup>

$$Z(e^{j\omega}) = \begin{cases} 2Y(e^{j\omega}), & 0 \leq \omega < \pi \\ 0, & -\pi \leq \omega < 0 \end{cases} \quad (2)$$

式(2)より、解析信号には  $2\pi$  周期内で負の周波数成分が存在しないので、比2のデシメーションを行なうことができる。それゆえ、解析信号にデシメーションを施した信号のスペクトル  $Z'(e^{j\omega})$  は周波数  $\omega$  上で元のスペクトル  $Z(e^{j\omega})$  をエリアジング効果なしに2倍に引き伸ばしている<sup>(10)</sup>ので、推定スペクトルの精度の改善が期待できる<sup>(10)</sup>

### 2. 2. 適応的前処理

最初に、AR 次数  $n$ 、MA 次数  $m$  の ARMA モデルの伝達関数のパワースペクトルを考える。

$$|H_{ARMA}(e^{j\omega})|^2 = \left| \frac{b_0^* + b_1^* e^{-j\omega} + \dots + b_m^* e^{-jm\omega}}{1 + a_1^* e^{-j\omega} + \dots + a_n^* e^{-jn\omega}} \right|^2 \quad (3)$$

ここで、\*は複素共役を意味する。もし、あるパワースペクトルが変数  $\omega$  の連続関数であれば、次数  $n$  と  $m$  を適切に選んだ式(3)の関数によって、そのパワースペクトルが精度よく近似されることは、よく知られている<sup>(9)</sup>。しかしながら、2. 1 で述べた解析信号  $\{z(k)\}$  にデシメーションを施した  $\{z(2k)\}$  のスペクトル  $Z'(e^{j\omega})$  は

$$Z'(e^{j\omega}) = Z(e^{j\omega/2}), \quad 0 \leq \omega < 2\pi$$

であるから、スペクトル  $Z'(e^{j\omega})$  は  $\omega = 2\pi i$  ( $i$  は整数)において一般に不連続となる。従って、信号  $\{z(2k)\}$  のパワースペクトルを式(3)の関数を用いて推定すると、その推定量は高周波数領域で誤差をもつことになる<sup>(10)</sup>。

従って、本論文では観測実信号  $\{x(k)\}$  に前処理を行い、パワースペクトルの最高周波数成分  $\omega_{\max}$  が周波数  $\omega = 0$  の成分にはほぼ等しい実信号  $\{y(k)\}$  を作り、それを解析信号  $\{z(k)\}$  に変換させて処理を行う。最終的には適応的に信号の特徴抽出をすることを目的としているため、前処理も適応的に行う。

次に本論文で行う前処理を具体的に説明する。この前処理は二つの手順からなる。

[手順 1]

実信号  $\{x(k)\}$  のスペクトルの大局的な特徴を適応的にとらえるため、AR 次数 1, MA 次数 1 の ARMA モデルを用いた RLS 法<sup>(11),(12)</sup> で特徴抽出を行う。RLS 法のアルゴリズムとしては文献(18)の拡張 MIS 法を簡便化したものを用いる。その結果、時刻  $k$  の推定 AR 係数  $\hat{a}_{pre}(k)$ , MA 係数  $\hat{b}_{pre}(k)$  を得る。

[手順 2]

手順 1 から得られる ARMA 係数を用いて逆フィルタリング

$$y(k) = -\hat{b}_{pre}(k)y(k-1) + x(k) + \hat{a}_{pre}(k)x(k-1) \quad (4)$$

を行い、パワースペクトルの最高周波数成分  $\omega_{\max}$  が周波数  $\omega = 0$  の成分にはほぼ等しい信号  $\{y(k)\}$  を得る。ただし、この逆フィルタリングの安定性を保証するため

$$\text{if } |\hat{b}_{pre}(k)| > 1.0 \text{ then } \hat{b}_{pre}(k) = 1.0 / \hat{b}_{pre}(k) \quad (5)$$

なる処理を式(4)のフィルタリングを実行する前に行う。仮に、式(5)の処理が行われても、その ARMA モデルの対数パワースペクトルの形は変わらない。

### 2. 3. 複素時変モデリング

本節では、2. 2 の手法によって前処理を行った結果得られた実信号  $\{y(k)\}$  を解析信号変換した複素信号  $\{z(k)\}$  の特徴抽出について考える。本論文では、音声信号の適応的な特徴抽出をする手法として文献(18)の拡張 MIS 法を複素化する。ただし、ここでは、デシメーションを行わない信号  $\{z(k)\}$  を対象にアルゴリズムの複素化を行う。

最初に、パラメータベクトルのプロセス式と時変 ARMA-D モデルからなる観測式を定義する。

$$\begin{aligned} \bar{\mathbf{p}}(k+1) &= \bar{\mathbf{p}}(k) + \mathbf{v}(k) \\ z(k) &= \bar{\mathbf{p}}^H(k) \mathbf{h}(k) + \bar{u}(k) + \bar{w}(k) \end{aligned} \quad (6)$$

ここで、添え字  $H$  はベクトルのエルミート転置を意味する。 $\bar{\mathbf{p}}(k)$  はパラメータベクトルで次式で定義される。

$$\bar{\mathbf{p}}^T(k) = [\bar{a}_1(k) \cdots \bar{a}_n(k) \bar{b}_1(k) \cdots \bar{b}_p(k) \bar{c}_1(k) \cdots \bar{c}_q(k)]$$

ここで、添え字  $T$  はベクトルの転置を意味する。そして、次数  $\bar{n}$ ,  $\bar{p}$ ,  $\bar{q}$  はそれぞれの真の次数  $n$ ,  $p$ ,  $q$  に対して以下の不等式を満足する。<sup>(17),(18)</sup>

$$\hat{n} > n, \hat{p} > p, \hat{q} > q$$

また、式(6)のベクトル  $\mathbf{h}(k)$  は

$$\mathbf{h}^T(k) = [-z(k-1) \cdots -z(k-\hat{n}) \hat{u}(k-1) \cdots \hat{u}(k-\hat{p}) \hat{w}(k-1) \cdots \hat{w}(k-\hat{q})] \quad (7)$$

で表される。ここで、 $\{\hat{u}(k)\}$ 、 $\{\hat{w}(k)\}$  はそれぞれ、推定疑似周期パルス列入力、推定白色雑音入力である。入力  $\{\hat{w}(k)\}$  は次の統計的特徴を満足する。

$$E[\hat{w}(k)] = 0, E[\hat{w}(k)\hat{w}^*(t)] = \delta_{k,t} \hat{W}^2(k) \quad (8)$$

式(8)の分散  $\hat{W}^2(k)$  は区間  $T_\alpha$  で計算される。従って

$$\hat{W}^2(k) = \hat{W}^2(k-i) \quad (i=1, 2, \dots, T_\alpha)$$

を得る。もし、

$$\begin{aligned} \hat{\mathbf{z}}(k|k-1) &= \hat{\mathbf{p}}^H(k|k-1)\mathbf{h}(k) \\ \hat{\mathbf{p}}^T(k|k-1) &= [\hat{a}_1(k|k-1) \cdots \hat{a}_{\hat{n}}(k|k-1) \hat{b}_1(k|k-1) \cdots \hat{b}_{\hat{p}}(k|k-1) \\ &\quad \hat{c}_1(k|k-1) \cdots \hat{c}_{\hat{q}}(k|k-1)] \end{aligned}$$

なる推定値を定義して、文献(18)のアルゴリズムの複素化を行うと次の更新式が得られる<sup>(12),(18)</sup>

[複素拡張 MIS 法]

$$\begin{aligned} \nu_w(k) &= z(k) - \hat{\mathbf{p}}^H(k|k-1)\mathbf{h}(k) \\ \text{if } |\nu_w(k)|^2 > C_N \hat{W}^2(k-1)(1 + \lambda^{-1}(k-1)\mathbf{h}^H(k)\mathbf{F}(k-1)\mathbf{h}(k)) \text{ then } v &= v+1 \end{aligned}$$

else  $v=0$  endif.

if  $v=N$  then

$$\begin{aligned} \hat{\sigma}_q^2(k-1) &= \{|\nu_w(k)|^2 - \hat{W}^2(k-1)(1 + \lambda^{-1}(k-1)\mathbf{h}^H(k)\mathbf{F}(k-1)\mathbf{h}(k))\} / \mathbf{h}^H(k)\mathbf{h}(k), \\ v &= 0, \tau_i = k \end{aligned}$$

else  $\hat{\sigma}_q^2(k-1)=0$  endif. (9)

$$\text{if } k \geq \tau_i + \tau_0 \text{ then } \lambda(k-1) = \lambda_0$$

$$\text{else } \lambda(k-1) = 1 \text{ endif.} \quad (10)$$

$$\mathbf{F}(k, k-1) = \lambda^{-1}(k-1)\mathbf{F}(k-1) + \frac{\hat{\sigma}_q^2(k-1)}{\hat{W}^2(k-1)} \mathbf{I} \quad (11)$$

if  $|\nu_w(k)|^2 > C \hat{W}^2(k-1)(1 + \mathbf{h}^H(k)\mathbf{F}(k, k-1)\mathbf{h}(k))$  and  $k > t_i + t_0$  then  $\hat{u}(k) = \nu_w(k)$ ,  $\nu(k)=0$ ,  $t_{i+1}=k$

else

$$\hat{u}(k)=0, \nu(k)=\nu_w(k)$$

endif

$$\hat{w}(k) = \frac{\nu(k)}{1 + \mathbf{h}^H(k)\mathbf{F}(k, k-1)\mathbf{h}(k)}$$

$$\mathbf{G}(k) = \mathbf{F}(k, k-1)\mathbf{h}(k)(1 + \mathbf{h}^H(k)\mathbf{F}(k, k-1)\mathbf{h}(k))^{-1}$$

$$\hat{\mathbf{p}}(k+1|k) = \hat{\mathbf{p}}(k|k-1) + \mathbf{G}(k)\nu^*(k)$$

$$\mathbf{F}(k) = \mathbf{F}(k, k-1) - \mathbf{G}(k)\mathbf{h}^H(k)\mathbf{F}(k, k-1)$$

$$\begin{aligned} \hat{W}^2(k) &= \hat{W}^2(k-1) + \{|\nu(k)|^2 / (1 + \mathbf{h}^H(k)\mathbf{F}(k, k-1)\mathbf{h}(k)) - \hat{W}^2(k-1)\} \\ &\quad / \min(k, T_\alpha) \end{aligned} \quad (12)$$

$$\hat{\mathbf{p}}^s(k) = \hat{\mathbf{p}}^s(k-1) + (\hat{\mathbf{p}}(k+1|k) - \hat{\mathbf{p}}^s(k-1)) / \min(k, T_s)$$

以上の式において、式(9)、(11)、(12)の  $\hat{\sigma}_q^2(k-1)$  と  $\hat{W}^2(k-1)$  は式(13)のイノベーション  $\nu(i)$  ( $i=$

1, 2, ..., k)の尤度関数<sup>(12)</sup>

$$L = \sum_{i=1}^k \log \left[ \frac{1}{\pi \sigma_i^2} \exp \left( -\frac{|\nu(i)|^2}{\sigma_i^2} \right) \right] \quad (13a)$$

$$\sigma_i^2 = E[|\nu(i)|^2] \quad (13b)$$

に最尤推定法を用いて得られる。また、式(10)の定数  $\tau_0$  はアルゴリズムの実行に先立って与えられるもので

$$\tau_0 = 1/(1 - \lambda_0) \quad (14)$$

で計算される。ここで、定数  $\lambda_0$  はRLS法の重み係数に相当する量で、経験的に決定される。式(14)の定数  $\tau_0$  はRLS法のメモリの尺度に相当するとみなせる<sup>(11),(12)</sup>ので、式(9)、(10)によって、ある意味で過去の全てのデータに忘却作用を与える時変量  $\hat{\sigma}_q^2(k-1)$  が零でない値を持ってから  $\tau_0$  区間は過去のデータが重み係数の影響を受けずに記憶されていくことになる。

ここまでで解析信号  $\{z(k)\}$  を複素時変モデリングする手法が示されたので、次に、デシメーションを行った信号  $\{z(2k)\}$  の特徴抽出を考える。従来からよく行われている手法では、2. 1に示したように解析信号  $\{z(k)\}$  にデシメーションを行った信号  $\{z(2k)\}$  に対して処理を行っている。従って、この方針に立てば式(6)、(7)においては  $\{z(k)\}$  ではなくて  $\{z(2k)\}$  を用いることになる。しかしながら、デシメーションの結果、利用できる観測信号のサンプル数が半減するため、適応的手法を用いてその特徴を推定する場合には、その適応的アルゴリズムのトラッキング能力が悪くなる。<sup>(14)</sup> 従って本論文では、このトラッキング能力の問題に対処するため、視点を多少変えて、式(7)のベクトル  $\mathbf{h}(k)$  を次のようにする。

$$\mathbf{h}^T(k) = [-z(k-2) \cdots -z(k-2\hat{n}) \hat{u}(k-2) \cdots \hat{u}(k-2\hat{p}) \hat{w}(k-2) \cdots \hat{w}(k-2\hat{q})] \quad (15)$$

複素拡張MIS法のアルゴリズムの導出においては、イノベーションズ法をカルマンフィルタ理論に応用しているが、この際には式(7)のベクトル  $\mathbf{h}(k)$  の要素にはなんの制限も仮定されていない。<sup>(12),(13),(18)</sup> 従って、式(7)の代わりに式(15)を使っても、複素拡張MIS法のアルゴリズムはベクトル  $\mathbf{h}(k)$  を除いて修正される必要がない。式(7)の代わりに式(15)を用いて式(6)を表せば、推定パラメータベクトル  $\hat{\mathbf{p}}(k+1|k)$  は信号  $\{z(k)\}$  の特徴を推定しているのではなく、信号  $\{z(2k)\}$ 、または、信号  $\{z(2k+1)\}$  の特徴を推定していることになる。つまり、式(6)、(15)を用いてパラメータ  $\hat{\mathbf{p}}(k+1|k)$  を推定するときには、信号  $\{z(2k)\}$  と  $\{z(2k+1)\}$  の両方を交互に、すなわち、 $\{z(k)\}$  を用いてパラメータが更新されるので、単純に信号  $\{z(2k)\}$  のみを用いるときより、パラメータの更新回数が2倍になる。従って、本論文での複素数の適応的アルゴリズムは、従来の拡張MIS法と同じトラッキング能力をもつ。

本論文では、以上の手法で信号  $\{z(2k)\}$  の特徴抽出をするが、実際に音声信号のパワースペクトルを推定してみると、無声子音の場合、推定されるパワースペクトルの高域成分に誤差が存在する。それゆえ、本論文では推定されたパワースペクトルの  $f_M[H_z]$  より上の高域部分から  $f_0/2[H_z]$  ( $f_0$ は信号  $\{x(k)\}$  のサンプリング周波数)までを利用しないことにする。これにより、利用できるスペクトルの帯域幅が狭くなるが、予め、 $f_0[H_z]$  より高いサンプリング周波数でサンプリングを行って本手法を用いれば、この問題は解決されと考えられる。

以上の議論より、本手法では観測実信号  $\{x(k)\}$  のパワースペクトルの推定値  $\hat{S}_x(k, \omega)$  は次のように表現される。

$$\hat{S}_x(k, \omega) = \hat{S}_z(k, \omega) + |\hat{H}_{pre}(k, e^{j\omega/2})|^2, [dB] \quad (0 \leq \omega \leq 2\pi \frac{f_M}{f_0/2})$$

ここで、 $\hat{S}_z(k, \omega)$ は信号 $\{z(2k)\}$ のパワースペクトルである。また、 $|\hat{H}_{pre}(k, e^{j\omega/2})|^2$ は前処理のために実信号 $\{x(k)\}$ の特徴抽出に用いた実数 ARMA モデルの伝達関数のパワースペクトルである。

### 3. 実験結果

本節では本論文で提案する手法が従来の適応的手法と実験的に比較される。この節の実験で解析信号の生成に使われた Hilbert 変換器は文献(15)の  $2N=40$  次の IIR フィルタである。3. 1 では合成音声に対する実験結果が示される。用いられるデータは入力が Rosenberg 波<sup>(16)</sup>である時変 AR モデルから生成される。定量的な比較より、本論文で提案される手法が従来の拡張 MIS 法より正確なスペクトルを推定でき、かつ、従来のデシメーションを伴う適応的手法よりトラッキング能力が改善されていることを示す。3. 2 では実音声に対する実験結果が示される。FFT との比較より、本手法が前処理を除いた本手法や従来の拡張 MIS 法より精度が高いスペクトルを推定できることを示す。

#### 3. 1. 合成音声に対する実験結果

この実験で使われるデータは入力が Rosenberg 波である時変 AR モデルから生成される。サンプリング周波数は  $f_0=12[kHz]$  であり、Rosenberg 波に関してはピッチ周期が  $8.33\text{ ms}$  で声門閉止区間が  $3.67\text{ ms}$  である<sup>(16)</sup>。時変 AR モデルの係数については、区間  $k=1\sim 2000$  が音素/ $u$ / (定常)で、区間  $k=1001\sim 2000$  では音素/ $u$ /から音素/ $i$ /まで AR 係数を線形的に変化させ(非定常)、区間  $k=2001\sim 2500$  が音素/ $i$ / (定常)である。また、放射特性としては差分特性 $(1-z^{-1})$ を用いた。

次に評価規準であるが、以下の距離を用いる。

[手法 1]

$$d_r(k) = \sqrt{\frac{1}{M} \sum_{i=1}^M (S_r(\omega_i/2, k) - S_0(\omega_i/2, k))^2} [dB]$$

[手法 2]

$$d_c(2k) = \sqrt{\frac{1}{M} \sum_{i=1}^M (S_c(\omega_i, k) - S_0(\omega_i/2, 2k))^2} [dB]$$

[手法 3]

$$d_p(k) = \sqrt{\frac{1}{M} \sum_{i=1}^M (S_p(\omega_i, k) - S_0(\omega_i/2, k))^2} [dB]$$

ここで、 $S_0(\omega_i/2, k)$ は時刻  $k$  での参照スペクトルを意味する。また、 $S_r(\omega_i/2, k)$ ,  $S_c(\omega_i, k)$ ,  $S_p(\omega_i, k)$ は、それぞれ、従来の拡張 MIS 法によって得られたスペクトル(手法 1), 2. 3 の複素数の拡張 MIS 法を 2 のデシメーションを行った信号 $\{z(2k)\}$ に用いた結果得られたスペクトル(手法 2), デシメーションを行わない解析信号 $\{z(k)\}$ に対して式(15)を利用することで複素数の拡張 MIS 法(本論文で提案する手法)を用いた結果得られたスペクトル(手法 3)である。ただし、 $\omega_i$ は  $0 < \omega \leq 2\pi(2f_M/f_0)$ を  $M(=128)$ 等分した離散角周波数である ( $f_M=5[kHz]$ )。

次に、実験に用いた各アルゴリズムの初期値を示す。

[手法 1]

$$\begin{aligned} \hat{\mathbf{p}}(0) &= \mathbf{0}, \quad \mathbf{F}(0) = 100 \mathbf{I}, \quad \lambda_0 = 0.98, \quad C = 12, \\ t_0 &= 24, \quad T_\alpha = 20, \quad T_s = 36, \\ C_N &= 2, \quad N = 3. \end{aligned} \quad (16)$$

[手法 2]

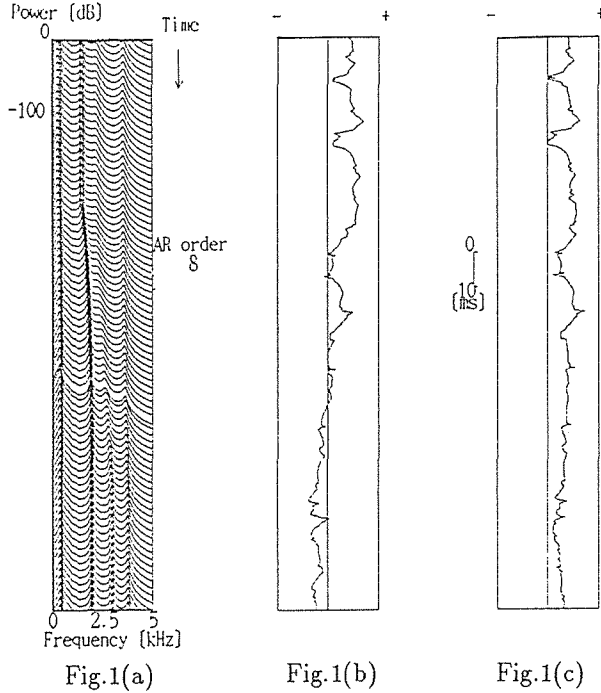
$$\begin{aligned} \hat{\mathbf{p}}(0) &= \mathbf{0}, \quad \mathbf{F}(0) = 100 \mathbf{I}, \quad \lambda_0 = 0.96, \quad C = 6, \\ t_0 &= 12, \quad T_\alpha = 10, \quad T_s = 18, \\ C_N &= 2, \quad N = 3. \end{aligned}$$

[手法 3]

$$\begin{aligned} \hat{\mathbf{p}}(0) &= \mathbf{0}, \quad \mathbf{F}(0) = 100 \mathbf{I}, \quad \lambda_0 = 0.98, \quad C = 6, \\ t_0 &= 24, \quad T_\alpha = 20, \quad T_s = 36, \\ C_N &= 2, \quad N = 3. \end{aligned}$$

ここで、手法 (2), (3) で用いられる 2.2 の適応的前処理の初期値については、手法 (1) の初期値の式(16)を除いたものと同じである。

以上の手法の実験結果が図 1 (a) ~ (c) である。ただし、アルゴリズムの初期学習部分を除くため、区間  $k=501 \sim 2500$  の結果を示している。また、図 1 (b) においては、時間幅を 2 倍に引き伸ばして結果を表示している。図 1 (b) より、最初の定常部分で手法 (2) が手法 (1) より正確なスペクトルを推定できているが、その後の非定常部分では手法 (1) より推定精度が悪く



**Fig. 1** Results of the experiment for the synthesized speech /ui/

(a) Reference spectra of the synthesized speech /ui/

(b) Distance  $d_r(2k) - d_c(2k)$

(c) Distance  $d_r(k) - d_p(k)$

なっていることがわかる。これは、従来の手法ではデシメーションを伴う複素処理を行うことでスペクトルの推定精度は上がるが、その適応的アルゴリズムのトラッキング能力が悪化することを示している。一方、図1(c)の結果を見ると、手法(3)は定常部分でも非定常部分でも手法(1)より精度の良いスペクトルを推定できていることがわかる。すなわち、本手法では、従来のデシメーションを伴う適応的複素信号処理に存在していたトラッキング能力の問題が改善されていることがわかる。

### 3. 2. 実音声に対する実験結果

この実験で使われるデータは、サンプリング周波数  $12[kHz]$  の男性音声  $/o \int i/$  である。実験に用いる各アルゴリズムの初期値は3.1のものと同じであり、比較されるスペクトルの最高周波数は  $f_M = 5[kHz]$  である。

実験結果を図2(a)～(e)に示す。ここで、図2(b)に示してあるスペクトルは、観測実信号をハニング窓を用いた256点FFTで分析した結果である。図2の後半部分の  $/ \int i/$  のスペクトルを比べると、図2(e)の本手法が中低域の推定精度が他の手法よりよいことがわかる。つまり、本手法が前処理を除いた本手法や従来の拡張MIS法より音声スペクトルの高精度推定に有効であることがわかる。

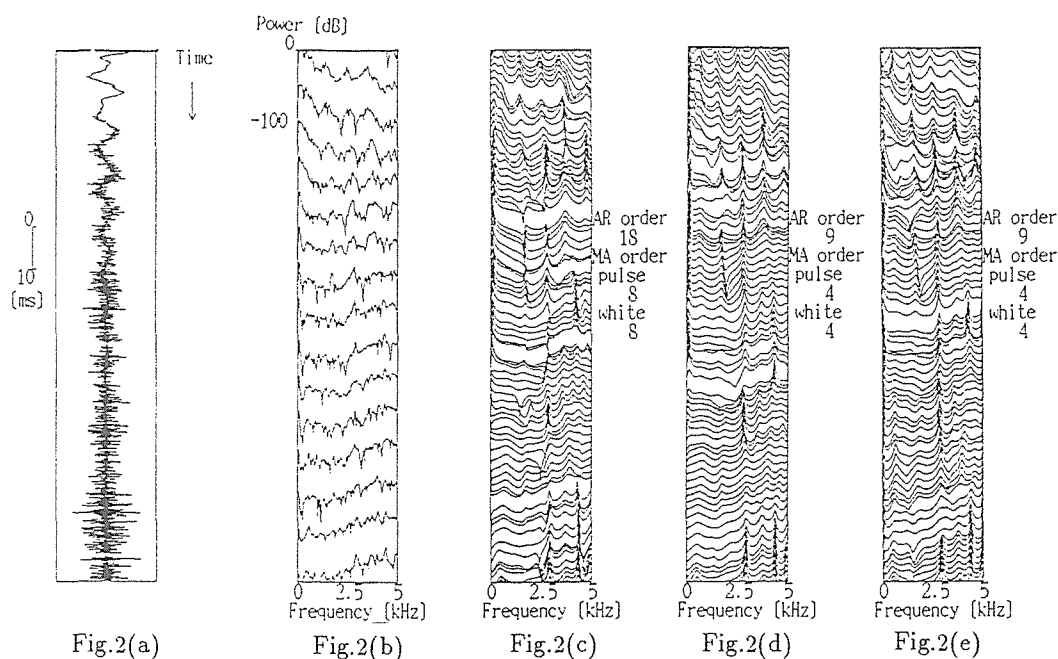


Fig. 2 Results of the experiment for the real speech  $/o \int i/$

- (a) Real speech waveform  $/o \int i/$
- (b) Spectra calculated by FFT of 256 samples
- (c) Estimation spectra of the conventional extended MIS
- (d) Estimation spectra of the proposed method without the pre-processing
- (e) Estimation spectra of the proposed method

#### 4. む す び

本論文では複素数の拡張 MIS 法を提案し、その手法が従来の拡張 MIS 法より正確な音声スペクトルを推定できることを示した。提案する手法では、推定スペクトルの高域での誤差を減少させるため、前処理を導入している。合成音声と実音声に対する実験結果より、提案する手法が従来の適応的複素信号処理手法より優れたトラッキング能力をもち、従来の拡張 MIS 法より精度の高いスペクトルを推定できることを示した。

今後の課題としては、本手法に含まれる前処理の改良等が考えられる。

#### 参 考 文 献

- [1] 板倉文忠, 斎藤収三, “統計的手法による音声スペクトル密度とホルマント周波数の推定”, 信学論 (A), vol. 53-A, pp. 35-42, 1970 年 1 月
- [2] B. S. Atal and S. L. Hanauer, “Speech analysis and synthesis by linear prediction of the speech wave”, J. Acoust. Soc. Amer., vol. 50, no. 2, pp. 637-655, 1971.
- [3] J. D. Markel and A. H. Gray, Jr., Linear Prediction of Speech. New York: Springer, 1976.
- [4] 深林太計志, 鈴木久喜, “極-零形線形モデルによる音声分析”, 信学論(A), vol. 58-A, pp. 270-277, 1975 年 5 月
- [5] H. Morikawa and H. Fujisaki, “Adaptive analysis of speech based on a pole-zero representation”, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, pp. 77-88, Feb. 1982.
- [6] Y. Miyanaga, N. Miki, N. Nagai and K. Hatori, “A speech analysis algorithm which eliminates the influence of pitch using the model reference adaptive system”, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-30, pp. 88-96, Feb. 1982.
- [7] Y. Miyanaga, N. Miki and N. Nagai, “Adaptive identification of a time-varying ARMA speech model”, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 423-433, Jun. 1986.
- [8] A. V. Oppenheim and R. W. Schaffer, Digital Signal Processing, Englewood Cliffs, NJ: Prentice Hall, 1975
- [9] J. A. Cadzow, “Spectral Estimation: An Overdetermined Rational Model Equation Approach”, IEEE Proc., vol. 70, pp. 907-939, Sep. 1982.
- [10] 島村徹也, 高橋進一, “正周波数を推定域とする複素線形予測法”, vol. J 72-A, pp. 1755-1763, 1989 年 11 月
- [11] E. Eleftheriou and D. D. Falconer, “Tracking properties and steady-state performance of RLS adaptive filter algorithms”, IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp. 1097-1110, Oct. 1986.
- [12] S. Haykin, Adaptive Filter Theory, 2nd edition. Englewood Cliffs, NJ: Prentice Hall, 1991.
- [13] T. Kailath, “An innovations approach to least-squares estimation part 1: linear filtering in additive white noise”, IEEE Trans. Automat. Contr., vol. AC-13, pp. 646-655, Dec. 1968.
- [14] 樫本 剛, 高橋進一, “マルチレート処理による適応フィルタの収束速度の向上”, 1993 年信学春期全大, pp. 1-188
- [15] 池原雅章, 田中宏幸, 松尾博文, “Remez アルゴリズムを用いた IIR ヒルベルト変換器の設計”, 信

- 学論(A), vol. J 74-A, pp. 414-420, 1991 年 3 月
- [16] A. E. Rosenberg, "Effect of glottal pulse shape on the quality of natural vowels", J. Acoust. Soc. Amer., vol. 49, no. 2, pp. 583-590, 1971.
- [17] 堀田英輔, 宮永喜一, 栃内香次, "2 入力 ARMA モデルを用いた音声の特徴抽出", vol. J 73-A, pp. 1476-1488, 1990 年 9 月
- [18] 堀田英輔, 宮永喜一, 栃内香次, "時変 ARMA-D モデルを用いた適応的音声分析法", vol. J 75-A, pp. 1324-1332, 1992 年 8 月