論　文

システム同定法を用いた雑音にロバストな音声分析

有馬　由紀[†*]　　島村　徹也[†a)]

Noise Robust Speech Analysis Using System Identification Methods

Yuki ARIMA[†*] and Tetsuya SHIMAMURA[†a)]

あらまし　本論文では，音声分析のための線形予測法の改良法を提案している．二つのシステム同定法，最小2乗法と補助変数法が全極フィルタの係数を推定するために用いられる．線形予測法では観測出力信号である音声信号から全極フィルタの係数を推定するのに対し，システム同定法では観測出力信号と入力信号から全極フィルタの係数を推定する．本論文では，観測出力信号である音声信号から入力信号を高精度かつ付加雑音にロバストに推定する新しい手法を，改良予測誤差信号を生成することにより導出している．そして，有声音を分析する場合を考察対象とし，インパルス列である入力信号を正確に推定できるなら，最小2乗法を利用するとき，フィルタ係数の推定は高精度となり，ピッチ周期の依存性を取り除けることを示す．また，補助モデルを用いた補助変数法を適用すれば，最小2乗法の性質を保持しつつ，雑音環境下でのフィルタ係数の推定精度を大幅に改善できることを示す．音声分析におけるこれらのシステム同定法の有効性は，計算機シミュレーション実験で明らかにされる．

キーワード　線形予測，全極フィルタ，システム同定，入力推定

1. まえがき

音声分析手法の中で，線形予測（LP）法 [1] は最も有力でかつ幅広く利用されている．LP 法はピッチ，フォルマント，スペクトル，声道面積関数などの音声の基本パラメータの推定，また低ビットレート伝送や記憶のための表現方法の主要な技術となっている．LP 法は自己相関法と共分散法に大別されるが [6]，双方とも白色雑音を入力とする自己回帰（AR）モデルを基本とし，導出されている．しかし，音声は一般にある線形時変システムを周期的パルス（有声音の場合）またはランダム雑音（無声音の場合）で励振して得られる出力であるとモデル化される [12]．ここで，線形時変システムを全極フィルタとし，AR モデルに基づく音声生成を仮定した場合，有声音においては入力信号に相違が生じることになる．これが，LP 法を基にした有声音分析の結果が，音声信号のピッチ周期の影響を受ける理由と考えられる [2], [3]．これらの LP 分析における問題に対して，柳田ら [17] や Lee [20] は LP 法に重み付けの処理を加える方法を導出し，有声音分析におけるピッチ周期の影響の軽減を図っている．

一方，音声が雑音を含むとき，LP 法の推定精度は大幅に低下してしまう [7]．この問題を解決する目的においても，LP 法の改良法がいくつか提案されている [8]〜[10], [17], [21]．しかし，Tierney [8] の LP の次数を増大する方法は実質的な耐雑音性の改善に至っていない．自己相関関数のラグの 0 部分から雑音分散を引き去る雑音補正法 [9], [10] は有力な方法と考えられるが，雑音成分の引き過ぎにより相関行列の正定値性が満たされなくなるとき，得られるスペクトルは大幅な推定誤差を含む結果となる．また，雑音成分を含まない自己相関関数のみを利用する高次 Yule-Walker 法（または改良 Yule-Walker 法とも呼ぶ）の適用 [17], [21] は，over-determined 形への拡張を施したとしても，処理される信号の特性及び次数の設定によって大きくそのスペクトル推定精度が左右される [19]．このような背景から，本論文では音声分析に対するシステム同定的アプローチを考察し，特に音声分析において重要とされる有声音のための新しい分析方法を提案する．提案法は，有声音の入力信号をインパルス列として推定する処理を介し，システム同定の原理から，音声信号のピッチ周期の影響を受けずにまた耐雑音性に優れた推

† 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，日本テレコム(株)
a) E-mail: shima@sie.ics.saitama-u.ac.jp

定結果をもたらす．

システム同定は，制御工学の分野において多く研究されている．その重要性から，与えられる入力と出力のデータから未知システムを推定するための様々な方法が今までに提案されている [5], [15], [16]．多くの方法は，入力信号と区別する形で未知システムを推定し，また，出力信号が雑音に乱されていても，いくつかの方法はその未知システムを正確に推定することができる．このようなシステム同定法の特徴は，音声分析において有益と考えられよう．現に，深林 [18] はシステム同定の一括処理法を，また森川ら [13]，宮永ら [14] はシステム同定の逐次処理法を音声分析に適用する試みをしている．しかし，音声分析にシステム同定法を適用するには，観測出力信号から入力信号を得なければならない．そのためには，LP 法の結果として得られる予測誤差信号が，全極フィルタの入力信号に対応するという性質を利用する必要がある．深林 [18] は，予測誤差信号の 5 乗からその平均値を引いて入力信号としている．また，森川ら [13] は，予測誤差信号を直接入力信号とする立場をとったが，宮永ら [14] は，予測誤差信号から入力インパルス列を推定する必要性を示唆している．しかし，いずれにおいても，付加雑音に対する影響及び対策にまでは言及していない．これは，雑音環境下では，入力信号を正確に推定することが容易ではないためと考えられる．そこで，本論文では，観測出力信号と予測誤差信号が等しい周期性を有するという事実に着目し，改良予測誤差信号の生成法を提案し，その改良予測誤差信号から付加雑音に対してロバストにインパルス列を推定する方法を導出する．具体的には，上記の改良予測誤差信号にしきい値を用いてインパルス列を推定する方法を提案する．この方法で入力信号を推定した後，一括処理的にシステム同定法を施したときに得られる推定結果と従来法によって得られる推定結果を比較検討する．本論文では，現在の音声分析の主流が一括処理である事実を鑑み，システム同定法としては一括処理法を積極的に用いる立場をとることにする．

本論文の構成は，まず **2.** で音声生成モデルから導出される線形予測法について簡潔に述べた後，**3.** で提案する改良予測誤差信号生成法及び入力推定法について記述する．続く **4.** では本論文で取り上げる二つのシステム同定法，最小 2 乗法と補助変数法を具体的に記述し，**3.** での入力推定との組合せを明確にする．**5.** では，提案するシステム同定に基づく音声分析法と深林の方法との相違点を明らかにし，**6.** で提案法のシミュレーション結果を示し，従来法との分析精度の比較検討を行う．そして最後に **7.** で結ぶことにする．

2. 音声生成モデルと線形予測

本論文では，音声は図 1 で示すような離散時間モデルによって生成されると仮定する．このとき，声門励振，声道，口唇放射の効果を総合したスペクトル上の特性は，定常状態のシステム関数が次式で与えられるような全極形ディジタルフィルタで表されることになる．

$$H(z) = \frac{G}{1 + \sum_{k=1}^{M} a_k z^{-k}} \tag{1}$$

この全極フィルタは，有声音のときはインパルス列，無声音のときはランダム雑音で励振される．よってこのモデルのパラメータは，有声/無声の分類を基に，有声音のピッチ周期，利得パラメータ G 及びディジタルフィルタの係数 a_k からなる．これらのパラメータは，分析区間において一定と仮定される．

本論文では，特に有声音の分析について考察する．したがって，入力信号はインパルス列とみなし，

$$u(n) = \sum_{j=0}^{\infty} \delta(n - jT) \qquad n = 0, 1, 2, \ldots \tag{2}$$

と表されると仮定する．ここで $\delta(\cdot)$ はデルタ関数を表す．また，T はピッチ周期のサンプル数に対応する．図 1 のモデルにおいて，出力 $s(n)$ は次のような差分方程式によって入力 $u(n)$ と関係している．

$$s(n) = -\sum_{k=1}^{M} a_k s(n-k) + Gu(n) \tag{3}$$

一方，予測係数 α_k をもつ予測器は次式の出力を与えるシステムとして記述される．

$$\hat{s}(n) = -\sum_{k=1}^{M} \alpha_k s(n-k) \tag{4}$$

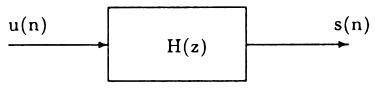

図 1　離散時間音声生成モデル
Fig. 1　Discrete-time speech production model.

ここで，$\hat{s}(n)$ は $s(n)$ の予測値であることを示している．このとき，予測誤差信号 $e(n)$ は次のように与えられる．

$$e(n) = s(n) - \hat{s}(n) = s(n) + \sum_{k=1}^{M} \alpha_k s(n-k) \tag{5}$$

この予測誤差信号 $e(n)$ の 2 乗規範

$$E = \sum_{n=l}^{N-1} e(n)^2 \tag{6}$$

を最小化することにより，我々は式 (1) の係数 $a_k (k = 1, 2, \ldots, M)$ と利得 G を得ることができる．その手段が LP 法である．$l = 0$ のとき自己相関法，$l = M$ のとき共分散法が LP 法として導出される [2]．これら二つの LP 法のうち，本論文では提案法におけるシステム同定法との関連を考慮し，入力インパルス列をより明確に算出し得る共分散法を取り上げることにする．

3. 入力推定

式 (3)，(5) より，予測係数 α_k と全極フィルタの係数 a_k が一致する場合は，

$$e(n) = Gu(n) \tag{7}$$

となる．すなわち，LP 法によって求められる予測誤差信号と入力信号は利得という定数をもって比例することになる．また，式 (2) より $u(n)$ は値が 1 であるインパルス列で表現される．したがって，出力信号である音声信号から，これらの性質を利用し入力インパルス列の推定を行うことが可能と考えられる．

音声が有する性質をより明確に表示するため，ここでは **6.** で用いられる合成音を例にとって説明することにする．図 2 は，データ数 $N = 300$，予測次数 $M = 10$ で得られた予測誤差信号を表している．ただし，この予測誤差信号はあるサンプリング時刻 t に対して，$s(n+t), n = 0, 1, \ldots, N-1$ の分析フレームから共分散法を用いて係数 $a_k, k = 1, 2, \ldots, M$ を推定した後，$\alpha_k = a_k$ とし，式 (5) に基づいて $s(n+t), n = -M, -(M-1), \ldots, 0, 1, 2, \ldots, N-2$ のデータサンプルから算出されている．一般に，音声信号が雑音を含まないとき，得られる予測誤差信号はほぼ図 2 のようになり，インパルスの形状は明確である．しかし，音声信号に雑音（ここでは白色雑音）が混入されると，予測誤差信号はその雑音の影響を大きく受け，図 3 のようになってしまう．図 3 は明らかに図 2 よりも雑音成分とインパルス成分を分離する処理を困難にする．そこで，よりインパルスの成分を強調するため，得られた予測誤差信号から次の改良予測誤差信号を生成する方法をここでは提案する．

出力信号である音声信号と予測誤差には強い相関があり，それらの周期は等しい．しかし，音声信号と予測誤差信号のピークの位置は必ずしも等しくはない．そこで次式のような相互相関関数

$$R_{es}(m) = \frac{1}{N} \sum_{n=0}^{N-m-1} e(n)s(n+m) \qquad m = 0, 1, 2, \ldots \tag{8}$$

を施し，$R_{es}(m)$ の最大値から，ピークの位置のずれ β を求める．そして

$$e'(n) = e(n)^2 s(n+\beta) \tag{9}$$

を算出し，改良予測誤差信号とする．$e(n)$ と $s(n+\beta)$ の周期とそれぞれのピークの位置は一致しているので，式 (9) は図 4 のようにインパルスの形状をより明確にする予測誤差信号となる．

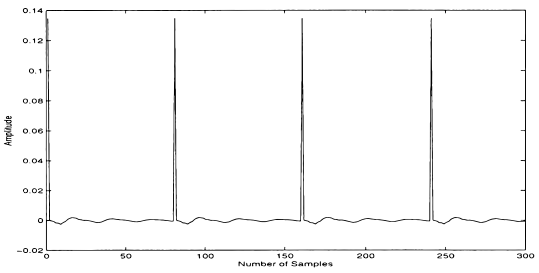

図 2　雑音を含まない場合の予測誤差
Fig. 2　Prediction error in a noiseless case.

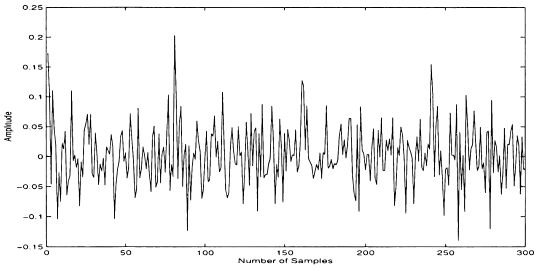

図 3　雑音環境下での予測誤差 $e(n)$ [SN 比=10 dB]
Fig. 3　Prediction error $e(n)$ in a noisy environment. [SNR=10 dB]

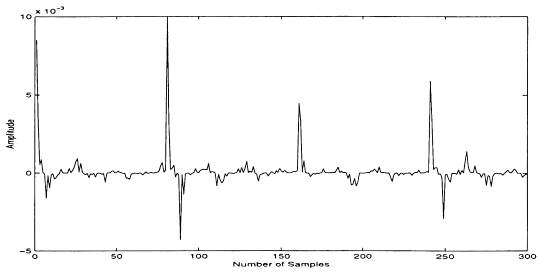

図 4　雑音環境下での改良予測誤差 $e'(n)$ [SN 比=10 dB]
Fig. 4　Improved prediction error $e'(n)$ in a noisy environment. [SNR=10 dB]

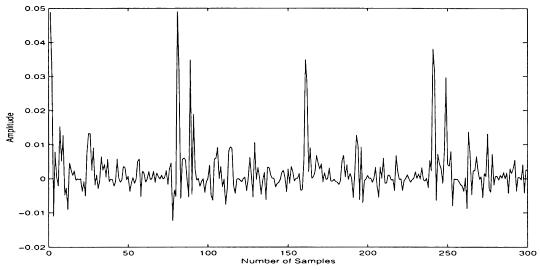

図 5　雑音環境下での改良予測誤差 $e''(n)$ [SN 比=10 dB]
Fig. 5　Improved prediction error $e''(n)$ in a noisy environment. [SNR=10 dB]

一方，式 (9) を導出するにあたり，$e(n)^2$ を $e(n)$ に置き換えた次のような予測誤差信号も考慮した．

$$e''(n) = e(n)s(n+\beta) \tag{10}$$

図 5 は本例における式 (10) の信号を示している．明らかに図 5 は図 4 より多くの雑音成分を含むと考えられる．そこで今回は $e'(n)$ を改良予測誤差として用いることにした．

本論文では，上記の改良予測誤差信号 $e'(n)$ を用いて，次の入力インパルス列推定法を提案する．

[**入力推定法**]

改良予測誤差信号 $e'(n)$ に対してしきい値を設け，

$$\hat{u}(n) = \begin{cases} 1 & \text{for} \quad e'(n) \geqq \gamma E_{\max} \\ 0 & \text{otherwise} \end{cases} \tag{11}$$

のようにインパルス列を求める．ここで，$E_{\max}$ は $e'(n)$ の最大値である．また，γ はある適当な正の実数である．図 6 は，式 (11) の処理を図示している．すなわち本法では，改良予測誤差信号の大きさがあるしきい値以上ならば 1，それ以外ならば 0 と置き換え，インパルス列を推定する．

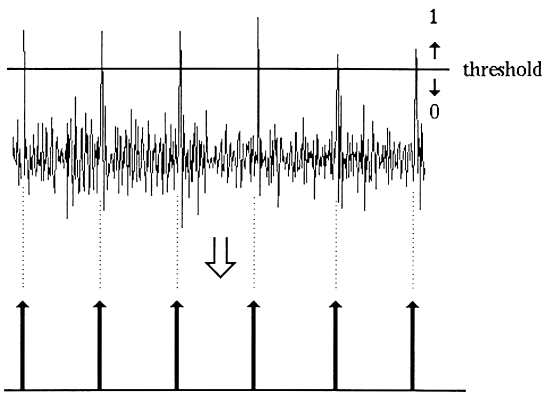

図 6　入力推定法（予測誤差とインパルス列）
Fig. 6　Input estimation method (prediction error and impulse chains).

4. システム同定

3. で述べた方法で全極フィルタの入力信号が得られれば，入出力信号からシステム同定法を利用し，全極フィルタを推定することが可能となる．本章では，システム同定法として代表的な二つの方法，最小 2 乗（LS）法と補助変数（IV）法 [15], [16] を取り上げることにする．そして，その二つの方法を，出力信号すなわち音声信号が雑音を含む場合と含まない場合で使い分けることを考えることにする．しかし実際には，後述するように，LS 法を包含する IV 法のみを適用すれば十分である．だが，雑音を含まない場合における LS 法の性質を明らかにするために，ここでは LS 法も記述しておくことにする．また，**6.** では LS 法を用いた結果も検証のために利用している．

4.1 最小 2 乗法

出力信号が雑音を含んでいないとき，システム同定では LS 法を用いるのが一般的である．

分析フレームにおいて音声サンプル $s(n+t), n = 0, 1, 2, \ldots, N-1$ が与えられていると仮定する．ただしここでは，簡単のため $t = 0$ として記述することにする．このとき，LS 法は

$$J = \sum_{n=M}^{N-1} [s(n) + a_1 s(n-1) + \ldots + a_M s(n-M) - b_0 \hat{u}(n)]^2 \tag{12}$$

を最小とするように係数を決定する．

次のようなベクトルと行列

$$s = [s(M), s(M+1), \ldots, s(N-1)]^T \tag{13}$$

$$\Omega = \begin{bmatrix} s(M-1) & \cdots & s(0) & \hat{u}(M) \\ s(M) & \cdots & s(1) & \hat{u}(M+1) \\ s(M+1) & \cdots & s(2) & \hat{u}(M+2) \\ \vdots & \ddots & \vdots & \vdots \\ s(N-2) & \cdots & s(N-M-1) & \hat{u}(N-1) \end{bmatrix} \tag{14}$$

を用意すると，式 (12) の J は

$$J = (s - \Omega\theta)^T (s - \Omega\theta) \tag{15}$$

と書ける．ここで T は転置を意味し，θ は係数ベクトル

$$\theta = [-a_1, -a_2, \ldots, -a_M, b_0]^T \tag{16}$$

を表している．θ の 2 次形式である J は $\partial J/\partial\theta = 0$ により最小化され，このとき，推定されるべき係数ベクトルは次式によって得られる．

$$\hat{\theta} = [\Omega^T\Omega]^{-1}\Omega^T s \tag{17}$$

式 (16) の b_0 は利得 G に対応する．つまりこの式 (17) によって計算される LS 法では，予測係数と利得を同時に計算することができることになる．出力信号が雑音を含んでいないとき，この LS 法は漸近的に不偏な推定結果を与える．

4.2 補助変数法

観測音声信号が，次のような式で与えられるとする．

$$x(n) = s(n) + w(n) \tag{18}$$

ここで，$w(n)$ は付加雑音を表す．このように出力信号が雑音を含む場合，LS 法は一般に推定値に偏りをもつ．なぜなら，

$$x = [x(M), x(M+1), \ldots, x(N-1)]^T \tag{19}$$

$$\Omega = \begin{bmatrix} x(M-1) & \cdots & x(0) & \hat{u}(M) \\ x(M) & \cdots & x(1) & \hat{u}(M+1) \\ x(M+1) & \cdots & x(2) & \hat{u}(M+2) \\ \vdots & \ddots & \vdots & \vdots \\ x(N-2) & \cdots & x(N-M-1) & \hat{u}(N-1) \end{bmatrix} \tag{20}$$

とおくとき，推定係数ベクトルが

$$\hat{\theta} = \theta + [\Omega^T\Omega]^{-1}\Omega^T v \tag{21}$$

となるためである．ここで

$$v = [v(M), v(M+1), \ldots, v(N-1)]^T \tag{22}$$

は誤差ベクトルを表している．また，$v(n)$ は $w(n)$ と

$$v(n) = w(n) + a_1 w(n-1) + \ldots + a_M w(n-M) \tag{23}$$

の関係を有する．式 (21) の右辺の第 2 項が，LS 法による推定偏りに対応している．IV 法とは，v とは独立であるが，$\Phi^T\Omega$ を正則とする補助変数行列 Φ を導入することにより，この LS 法の推定偏りを補正する方法である．すなわち，IV 法は LS 法の一改良法である．IV 法を用いれば，Φ の要素に適当な補助変数を選ぶことにより，少ない計算量でほぼ正確な係数推定の結果を得ることができる．IV 法で利用される補助変数にはいくつかあるが，ここでは雑音を含まない出力 $s(n)$ の推定値である $\hat{s}(n)$ を補助変数として用いることにする．$\hat{s}(n)$ は図 7 で示されるような補助モデルを用いることで生成される．

一括処理を p 回行ったときの，IV 法の解を $\hat{\theta}^{(p)}$ と表記すると，

$$\hat{\theta}^{(p)} = \left[-\hat{a}_1^{(p)}, -\hat{a}_2^{(p)}, \ldots, -\hat{a}_M^{(p)}, \hat{b}_0^{(p)}\right]^T \tag{24}$$

と表すことができる．この係数ベクトルは，次式のように計算することにより得られる．

$$\hat{\theta}^{(p)} = \left[\Phi^{(p-1)T}\Omega\right]^{-1}\Phi^{(p-1)T}x \tag{25}$$

ここで，

$$\Phi^{(p-1)} =$$

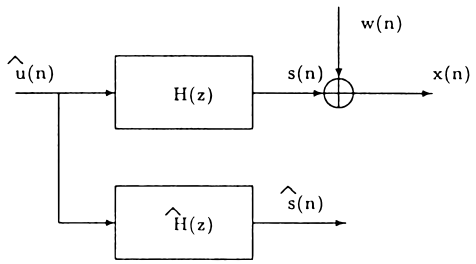

図 7　IV 法の補助モデル
Fig. 7　Instrumental model for the IV method.

$$\begin{bmatrix} \hat{s}^{(p-1)}(M-1) & \cdots & \hat{s}^{(p-1)}(0) \\ \hat{s}^{(p-1)}(M) & \cdots & \hat{s}^{(p-1)}(1) \\ \hat{s}^{(p-1)}(M+1) & \cdots & \hat{s}^{(p-1)}(2) \\ \vdots & \ddots & \vdots \\ \hat{s}^{(p-1)}(N-2) & \cdots & \hat{s}^{(p-1)}(N-M-1) \end{bmatrix} \begin{bmatrix} \hat{u}(M) \\ \hat{u}(M+1) \\ \hat{u}(M+2) \\ \vdots \\ \hat{u}(N-1) \end{bmatrix} \tag{26}$$

である．式 (26) に含まれる補助変数 $\hat{s}^{(p-1)}(n)$ は次のような式で与えられる．

$$\hat{s}^{(p-1)}(n) = -\sum_{k=1}^{M} \hat{a}_k^{(p-1)} \hat{s}^{(p-1)}(n-k) + \hat{b}_0^{(p-1)} \hat{u}(n) \tag{27}$$

$p-1$ 回での係数推定値を基に p 回目の推定結果を与える上記アルゴリズムは，初期値として $p=0$ での係数推定値 $\hat{a}_k^{(0)}$ と $\hat{b}_0^{(0)}$ を必要とする．これらは，文献 [5] での示唆により，入力推定値 $\hat{u}(n)$ と観測出力信号 $x(n)$ に基づく LS 法によって得ることにする．

式 (27) によるフィルタリングでは，全極フィルタがときどき不安定になるという問題があるが，これは次のように解決できる．多項式

$$\hat{A}^{(p-1)}(z) = 1 + \sum_{k=1}^{M} \hat{a}_k^{(p-1)} z^{-k} \tag{28}$$

を因数分解し

$$\hat{A}^{(p-1)}(z) = \prod_{k=1}^{M} (1 - z_k z^{-1}) \tag{29}$$

とする．もし，$|z_k| > 1$ となった場合には，z_k を $z_k/|z_k|^2$ と置き換える．すると，式 (29) を展開して多項式 $\hat{A}^{(p-1)}(z)$ が新しく作り直される．結果として，$\hat{A}^{(p-1)}(z)$ のもつ零点はすべて単位円の中に配置され，推定される全極フィルタの安定性は保証されることになる．

IV 法は，本質的には，付加雑音が存在する場合に適用されるシステム同定法である．しかし，上記のアルゴリズムを用いるなら，付加雑音が存在しない場合には，IV 法の解が $p=0$ での LS 法による解に帰着されることになる．したがって，付加雑音のあるなしにかかわらず，本 IV 法は適用可能である．本 IV 法はその原理からして，基本的には LS 法の性質を保持することになる．

5. 深林の方法との比較

LP 法を用いて入力信号を推定した後にシステム同定法を適用する観点から，**3.**，**4.** で記述した本論文での提案法は深林の方法 [18] と類似性を有する．したがって，本章では，提案法と深林の方法との相違点を明らかにしておく．

深林の方法は，共分散法から予測誤差信号 $e(n)$ を求めた後，

$$e_f(n) = e(n)^5 - \frac{1}{N} \sum_{n=0}^{N-1} e(n)^5 \tag{30}$$

を算出し，これを入力信号とする．そして，得られた入出力信号に対して LS 法を適用し，全極フィルタの係数 a_k を求める．この方法では，$e_f(n)$ にピッチ周期に対応するインパルス列が保存されるため，提案法と同様，有声音分析においてピッチ周期の影響の軽減が期待できる．しかし，$e_f(n)$ で生成されるインパルス列は，その大きさが正規化されないため，利得 G の計算が直接できない．これに対して，提案法ではインパルス列の大きさは 1 に正規化されるため，入出力信号から直接利得 G を求めることができる．

一方，深林の方法は，得られた入出力信号に対して LS 法を適用しているのみであり，システム同定における雑音対策を講じていない．また，$e_f(n)$ 自体も強靭であるとは言い難い．図 8 は，**3.** での合成母音を用いたときに得られた $e_f(n)$ を示している．図 9 は同様にして，付加雑音が混入した場合において得られた $e_f(n)$ を示している．図 8 を図 2 と比べれば，その形状はほぼ同じである．しかし，図 9 は大きく雑音に乱されている．比較のために，**3.** で記した入力推定法によって，図 4 の $e'(n)$ から復元した入力信号を図 10 に示しておく．図 10 を図 9 と比べれば，インパルス列を復元している提案法が，明らかに付加雑音には優位であると言えよう．

ところが，提案法では入力推定においてピッチ抽出誤りを起こす危険性がある．一方で，深林の方法における $e_f(n)$ にも，$e(n)$ に含まれる雑音成分を増大させてしまう欠点がある．次章では，入力推定におけるこれらの特性が，音声分析の結果にどのように反映さ

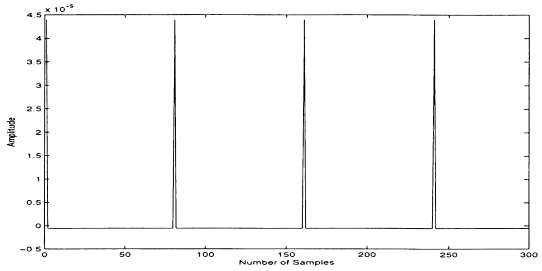

図 8　雑音を含まない場合の深林の方法における入力信号 $e_f(n)$
Fig. 8　Input signal $e_f(n)$ for Fukabayashi's method in a noiseless case.

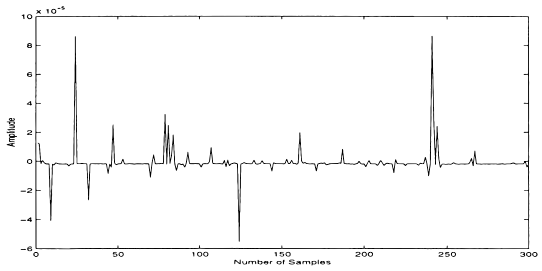

図 9　雑音環境下での深林の方法における入力信号 $e_f(n)$ [SN 比=10 dB]
Fig. 9　Input signal $e_f(n)$ for Fukabayashi's method in a noisy environment. [SNR=10 dB]

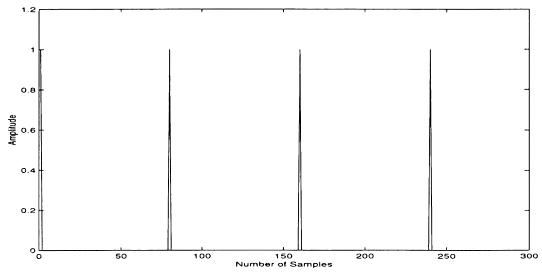

図 10　雑音環境下での提案法における入力信号 $\hat{u}(n)$ [SN 比=10 dB]
Fig. 10　Input signal $\hat{u}(n)$ for proposed method in a noisy environment. [SNR=10 dB]

れるかを更に調べることにする．

6. シミュレーション結果

提案する音声分析手法の有効性を示すためにシミュレーション実験を行った．本章ではそれらの結果を示す．

6.1 合　成　音

提案法の性質を明らかにするために，まず合成音を使用した．音声信号は母音/o/で，式 (2)，(3) に基づいて生成した．各パラメータは次のとおりである．

$G = 0.1354,\ a_1 = -1.53527,\ a_2 = 0.97789,$

$a_3 = -1.48396,\ a_4 = 1.78023,\ a_5 = -0.71704,$

$a_6 = 0.73514,\ a_7 = -0.76348,\ a_8 = -0.12135,$

$a_9 = 0.15552,\ a_{10} = 0.178143,$

サンプリング周波数 $f_s = 10\,\mathrm{kHz}$,

ピッチ周期 $= 8\,\mathrm{ms}(T = 80)$

この合成母音は [4] で有声音分析のために使用されている．

まず，雑音を含まない場合の LS 法についてシミュレーションを行った．ここでは再び，LS 法は IV 法における $p = 0$ の場合に等しいことを付記しておく．図 11 には LS 法によって推定されたスペクトルと，従来法である共分散法 [1] によって推定されたスペクトルが示されている．音声サンプル数は 300 個（$N = 300$）で，共分散法，LS 法ともに予測次数は $M = 10$ とした．LS 法の入力推定においては，インパルス列は $\gamma = 0.35$ をしきい値として得られた．この処理により，入力インパルス列は正確に復元できたことをここに明記しておく．図 11 から，LS 法によって推定されたスペクトルは真のスペクトルと全く一致していることがわかる．また，共分散法によって推定されたスペクトルは真のスペクトルとわずかに異なっている．それぞれのスペクトルの推定誤差は次式で計算される．

$$F = 10\log_{10}\left(\int_0^{f_s/2} \frac{|\hat{P}(f) - P(f)|}{P(f)} df\right) \quad (31)$$

ここで，$P(f)$ は真のスペクトルで

$$P(f) = \left|H\left(e^{j2\pi f/f_s}\right)\right|^2 \quad (32)$$

である．また，$\hat{P}(f)$ はその推定値である．共分散法によるスペクトル推定誤差の値は $F = 11.02\,\mathrm{dB}$，LS 法による値は $F = -93.27\,\mathrm{dB}$ となった．

また，音声信号のピッチ周期が LS 法にどのような影響を与えるかを調べた．入力信号を高ピッチから低ピッチへと変化させ，各々の入力において音声サンプルを生成し，共分散法と LS 法から求められる第 1 フォルマントを評価した．図 12 は，第 1 フォルマントの推定周波数を示している．LS 法はピッチ周期の影響

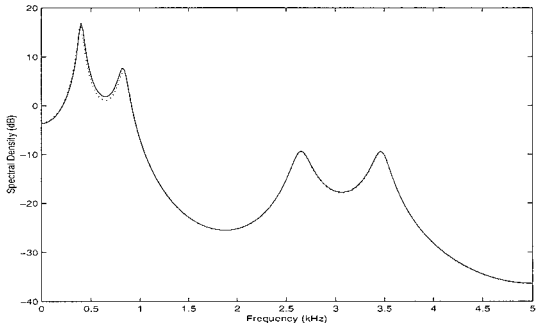

図 11　雑音を含まない場合のスペクトル推定（実線：真のスペクトル，点線：共分散法によって推定されたスペクトル，破線：LS 法によって推定されたスペクトル）

Fig. 11　Spectra estimated in a noiseless case. (Solid line: True spectrum, Dotted line: Spectrum estimated by covariance method, Dashed line: Spectrum estimated by LS method)

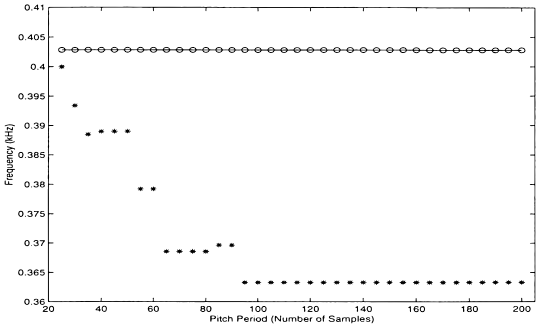

図 12　共分散法と LS 法による第 1 フォルマント周波数の推定（実線：真値，星印：共分散法による推定値，丸印：LS 法による推定値）

Fig. 12　First formant frequencies estimated by the covariance and LS methods. (Solid line: True value, Asterisk plot: Covariance method, Circle plot: LS method)

を全く受けていないが，共分散法はピッチ周期の影響を大きく受けていることがわかる．

次に，音声信号に白色雑音を加え，共分散法と IV 法を施した．SN 比は 10 dB とした．共分散法，IV 法ともに予測次数は $M = 10$ とした．また，IV 法の入力推定においては，再び $\gamma = 0.35$ をしきい値としてインパルス列を得た．IV 法によるスペクトルの推定は 3 回の反復（$p = 3$）で得ることができた．この結果は，反復回数を 10 回（$p = 10$）としたときとほとんど差がなかった．独立した 10 回の試行に対し，図 13 は共分散法によって推定されたスペクトルを重ね合わせて

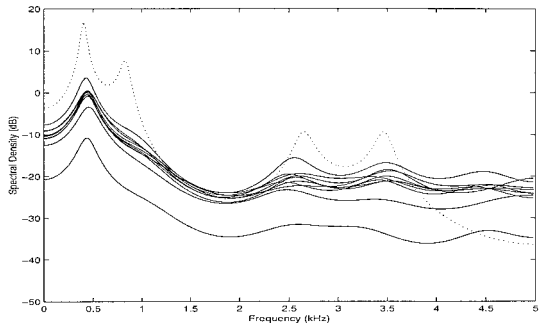

図 13　雑音環境下でのスペクトル推定 [SN 比=10 dB]（点線：真のスペクトル，実線：共分散法によって推定されたスペクトル）

Fig. 13　Spectra estimated in a noisy environment. [SNR=10 dB] (Dotted line: True spectrum, Solid line: Spectrum estimated by covariance method)

示し，図 14 は IV 法によって推定されたスペクトルを示している．また比較のために，同様の条件において LS 法も実行してみた．図 15 はその結果を示している．図 13〜15 のスペクトル結果からは，まず，共分散法によって推定されたスペクトルは雑音の影響を受け，推定精度が大幅に低下していることがわかる．明らかに IV 法は共分散法よりも正確にスペクトルピークを抽出し，良い結果を与えている．また，図 14 と図 15 を比較すると，IV 法によって推定されたスペクトルの方がより高精度であることがわかる．これは，IV 法が，雑音によって受ける LS 法の推定偏りを補正する能力を有するためである．スペクトル誤差の平均は共分散法では $F = 28.70$ dB，IV 法では $F = 25.32$ dB，LS 法では $F = 29.12$ dB となった．

続いて，深林の方法 [18] を調べた．また比較のために，Lee によるロバスト線形予測法 [20] 及び柳田らの重み付き線形予測法 [17] も算出してみた．しかし，これらの方法における利得 G の計算方法は明確に与えられていない．したがって，ここでは予測誤差のみからなる LPC ケプストラム距離を次のように算出し，推定精度を比較した．

$$CD = \frac{10}{ln10}\sqrt{2\sum_{i=1}^{M}(c_i - \hat{c}_i)^2} \tag{33}$$

上式で，c_i は真の全極フィルタ係数のケプストラム，$\hat{c}_i$ は推定された全極フィルタ係数のケプストラムである．$M = 10$ の場合におけるそれぞれの方法の推定

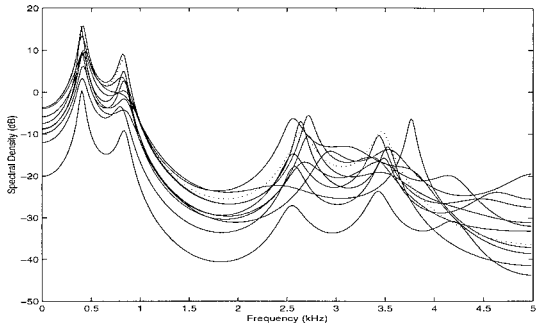

図 14　雑音環境下でのスペクトル推定 [SN 比=10 dB]（点線：真のスペクトル，実線：IV 法によって推定されたスペクトル）

Fig. 14　Spectra estimated in a noisy environment. [SNR=10 dB] (Dotted line: True spectrum, Solid line: Spectrum estimated by IV method)

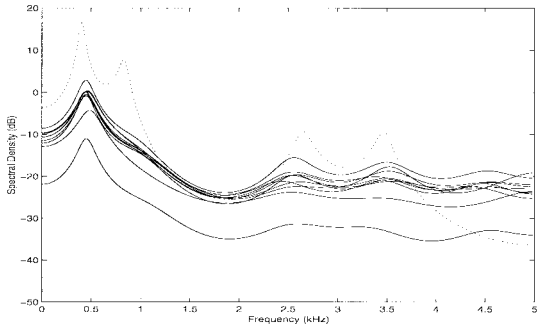

図 15　雑音環境下でのスペクトル推定 [SN 比=10 dB]（点線：真のスペクトル，実線：LS 法によって推定されたスペクトル）

Fig. 15　Spectra estimated in a noisy environment. [SNR=10 dB] (Dotted line: True spectrum, Solid line: Spectrum estimated by LS method)

結果が表 1 にまとめてある．ただし，Lee のロバスト線形予測では，Huber による関数を固定パラメータ $c = 1.5$ とともに重み付けした．また，柳田らの方法では，重み付き線形予測のみを施し，高次 Yule-Walker 方程式及び over-determined 形への拡張は加えなかった．表 1 には，無雑音及び雑音下における結果が示してある．付加雑音は白色雑音である．それぞれにおいて，独立試行 10 回の結果の平均がまとめてある．また，参考として，LS 法の結果も表 1 には示してある．

表 1 より，深林の方法は，無雑音の場合に共分散法より高精度な結果を与えていることがわかる．しかし，

表 1　LPC ケプストラム距離の平均
Table 1　Averages of LPC cepstrum distance.

	SN 比（dB）	
	∞	10
共分散法	0.2583	7.564
深林の方法	0.0878	7.715
LS 法	$1.1466e^{-11}$	7.757
IV 法	$1.1466e^{-11}$	5.123
Lee の方法	0.0471	9.242
柳田らの方法（$\sigma = 0.50$）	0.2403	7.773
柳田らの方法（$\sigma = 0.25$）	0.1936	7.786
柳田らの方法（$\sigma = 0.10$）	0.0423	7.816

雑音下においては逆にわずかに共分散法に劣る．IV 法は，雑音のあるなしにかかわらず，明らかに共分散法及び深林の方法より良好な結果を与えている．Lee の方法は，無雑音下では深林の方法より高精度であるが，雑音下では逆に劣る．柳田らの方法は，重み付けに用いられたパラメータ σ の値に推定精度が左右されるが，$\sigma = 0.1$ の設定において，Lee の方法よりも良好な結果を与えている．全体としては，提案法が従来法より良好な結果を与えていることは明らかである．LS 法は，雑音下において用いる必要はないが，その推定精度は深林の方法と柳田らの方法の中間程度であることを見て取ることができる．

6.2 実 音 声

次に実音声を使用した．音声信号は，サンプリング周波数 f_s=10 kHz，帯域制限 3.4 kHz で得られた男声/a/である．

まず，雑音を含まない場合の LS 法についてシミュレーションを行った．音声サンプル数は 300 個（$N = 300$）で共分散法，LS 法ともに予測次数は $M = 10$ とした．図 16 には LS 法によって推定されたスペクトルと，FFT，共分散法によって推定されたスペクトルがそれぞれ示されている．LS 法におけるインパルス列は $\gamma = 0.35$ をしきい値として得られた．ここでもインパルス列は規則正しく形成され，入力信号は正確に復元できたと思われる．図 16 では，LS 法によって推定されたスペクトルと共分散法によって推定されたスペクトルはほぼ重なっている．また，それらのスペクトルは，FFT によるスペクトルエンベロープを十分に近似している．他の実音声（男声）/i/ /u/ /e/ /o/ についてもシミュレーションを行ったところ，これと同様の結果が得られた．

続いて，音声が雑音を含む場合のシミュレーションを行った．音声信号に 3.4 kHz に帯域制限された白色

雑音を加え，上記の 5 母音に対してそれぞれ SN 比が 0 dB，5 dB，10 dB となるようにした．そして，共分散法と IV 法を施した．共分散法，IV 法ともに予測次数は $M = 10$ とした．また IV 法における入力推定では，インパルス列は $\gamma = 0.35$ をしきい値として得た．ここでも IV 法によるスペクトルの推定は 3 回の反復（$p = 3$）で得ることができた．また，同様にして深林の方法，Lee の方法，柳田らの方法をも実行してみた．5 母音にそれぞれ 10 回の試行を施し，それらの結果の平均値を算出し，表 2 にまとめてある．ここでは，無雑音時における推定結果と，雑音下における推定結果の LPC ケプストラム距離を評価量としている．

ここでの実験においては，柳田らの重み付き線形予測法は $\sigma = 0.5$ を設定したが，音声信号を直接用いた場合，すべてフローティングオーバとなった．これは，音声信号の振幅が数 10 dB であり，相関行列への重み付け及びその逆行列の計算に非常に多くのけた数が必要とされたためである．したがって，音声信号にスケーリングを施し，振幅レベルを先に用いた合成音と同程度としたところ，表 2 にある結果を得ることができた．また，Lee のロバスト線形予測法においても数値的な不安定現象が発生し，結果として大きな推定誤差を含むこととなった．しかし，Lee の方法におけるこの現象は，音声信号のスケーリングによって回避することはできなかった．これは，実音声のピッチ周期の不均一性及び雑音成分の不確定性から，重み付けされた相関行列が悪条件となったためと考えられる．一方，深林の方法は，安定して共分散法に比べ良好な結果を与え，システム同定に基づく手法の数値的安定性を示した．しかし，提案法はその深林の方法を上回る結果を与えた．

提案法の雑音に対するロバスト性を更に検証するために，今度は電子協騒音データベースに収録されている No.12 列車（在来線），No.14 計算機室（ワークステーション），No.9 幹線道路・交差点，No.10 人混みの雑音を音声信号にそれぞれ付加し，同様の実験を行った．ただし，それぞれの雑音は 3.4 kHz に帯域制限して用いた．その結果が表 3〜6 にまとめてある．Lee の方法はいずれも数値的に不安定であったので，ここでは省略してある．また，柳田らの方法では音声信号のスケーリングが施されている．全体的に，深林の方法は共分散法，柳田らの方法より高精度であるが，提案法は更なる改善を与えていることが見て取れる．したがって，これらの結果より，提案法は，従来法に比べ雑音にロバストな音声分析手法であると言えよう．

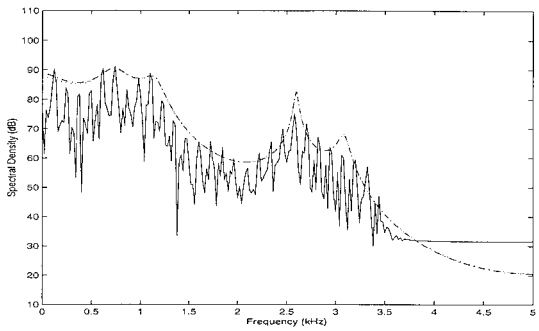

図 16　雑音を含まない場合のスペクトル推定（実線：FFT によるスペクトル，点線：共分散法によって推定されたスペクトル，破線：LS 法によって推定されたスペクトル）

Fig. 16　Spectra estimated in a noiseless case. (Solid line: Spectrum estimated by FFT, Dotted line: Spectrum estimated by covariance method, Dashed line: Spectrum estimated by LS method)

表 2　白色雑音下における LPC ケプストラム距離の平均

Table 2　Averages of LPC cepstrum distance in white noise.

	SN 比（dB）		
	0	5	10
共分散法	17.572	16.541	15.351
深林の方法	17.482	16.438	15.248
IV 法	17.359	16.337	14.835
Lee の方法	$1.0800e^{10}$	$1.2723e^{9}$	$3.7758e^{5}$
柳田らの方法	17.602	16.576	15.376

表 3　列車雑音下における LPC ケプストラム距離の平均

Table 3　Averages of LPC cepstrum distance in train noise.

	SN 比（dB）		
	0	5	10
共分散法	11.337	10.157	8.816
深林の方法	11.306	10.133	8.802
IV 法	10.999	9.804	8.656
柳田らの方法	11.327	10.151	8.824

表 4　計算機雑音下における LPC ケプストラム距離の平均

Table 4　Averages of LPC cepstrum distance in work station noise.

	SN 比（dB）		
	0	5	10
共分散法	11.384	10.204	8.835
深林の方法	11.287	10.126	8.812
IV 法	11.170	9.748	8.458
柳田らの方法	11.383	10.208	8.854

表5　交差点雑音下における LPC ケプストラム距離の平均
Table 5　Averages of LPC cepstrum distance in traffic junction noise.

	SN 比（dB）		
	0	5	10
共分散法	10.537	9.298	7.846
深林の方法	10.394	9.206	7.818
IV 法	9.772	8.878	7.712
柳田らの方法	10.498	9.265	7.820

表6　人混み雑音下における LPC ケプストラム距離の平均
Table 6　Averages of LPC cepstrum distance in crowded people noise.

	SN 比（dB）		
	0	5	10
共分散法	11.339	10.193	8.811
深林の方法	11.181	10.089	8.774
IV 法	11.000	9.539	8.387
柳田らの方法	11.337	10.202	8.830

7. む す び

本論文では，音声分析の問題に対し，二つのシステム同定法，LS 法と IV 法を利用し，特に従来課題とされていた有声音の分析における解を与えた．提案したシステム同定に基づく分析手法は，ピッチ周期依存性及び耐雑音性において従来法を大きく上回る．

改良予測誤差信号から入力インパルス列を推定することにより，LS 法を用いて音声信号のピッチ周期の影響を受けずに全極フィルタの係数を推定できることを示したが，この LS 法の性質は LS 法の改良法である IV 法に包含される．したがって，安定した補助モデルを利用することにより，雑音環境下においても良好な推定結果を与える IV 法は，有力な音声分析手法と考えることができよう．

本論文では，音声分析にシステム同定を適用するためにしきい値処理に基づく入力インパルス列推定法を提案した．これは，シミュレーション結果より，従来法の推定精度を上回るためには十分な入力推定法と思われる．しかし，より高精度なインパルス列推定が行われれば，提案法は更に高精度な結果を与えることが期待できる．

文　献

[1] B.S. Atal and S. Hanauer, “Speech analysis and synthesis by linear prediction of the speech wave,” J. Acoust. Soc. Am., vol.50, no.2, pp.637–655, Aug. 1971.

[2] S. Chandra and W.C. Lin, “Experimental comparison between stationary and nonstationary formulations of linear prediction applied to voiced speech analysis,” IEEE Trans., vol.ASSP-22, no.6, pp.403–415, Dec. 1974.

[3] L.R. Rabiner, B.S. Atal, and M.R. Sambur, “LPC prediction error — analysis of its variation with the position of the analysis frame,” IEEE Trans., vol.ASSP-25, no.5, pp.434–442, Oct. 1977.

[4] K.K. Paliwal and P.V.S. Rao, “A modified autocorrelation method of linear prediction for pitch-synchronous analysis of voiced speech,” Signal Processing, vol.3, no.2, pp.181–185, April 1981.

[5] 相良節男，秋月影雄，中溝高好，片山　徹，システム同定，計測自動制御学会，東京，1981.

[6] J. Makhoul, “Linear prediction: A tutorial review,” Proc. IEEE, vol.63, no.4, pp.561–580, April 1975.

[7] M.R. Sambur and N.S. Jayant, “LPC analysis / synthesis from speech inputs containing quantizing noise or additive white noise,” IEEE Trans., vol.ASSP-24, no.6, pp.488–494, Dec. 1976.

[8] J. Tierney, “A study of LPC analysis of speech in additive noise,” IEEE Trans., vol.ASSP-28, no.4, pp.389–397, Aug. 1980.

[9] S.M. Kay, “Noise compensation for autoregressive spectral estimates,” IEEE Trans., vol.ASSP-28, no.3, pp.292–303, June 1980.

[10] H. Hu, “Noise compensation for linear prediction via orthogonal transformation,” Electron. Lett., vol.32, no.16, pp.1444–1445, Aug. 1996.

[11] W.J. Hess, “Pitch and voicing determination,” in Advanced in Speech Signal Processing, ed. S. Furui et al., Marcel Dekker, New York, 1992.

[12] 古井貞熙，ディジタル音声処理，東海大学出版会，東京，1985.

[13] 森川博由，藤崎博也，“ARMA パラメータの同時推定法による音声分析，” 信学論（A），vol.J61-A, no.3, pp.195–202, March 1978.

[14] 宮永喜一，三木信弘，永井信夫，“ピッチ推定を含めた音声の ARMA パラメータの一推定法，” 信学論（A），vol.J63-A, no.11, pp.737–744, Nov. 1980.

[15] N.K. Sinha and B. Kuszta, Modeling and Identification of Dynamic Systems, Van Nostrand Reinhold, New York, 1983.

[16] T. Soderstrom and P. Stoica, System Identification, Prentice-Hall, Hemel Hempstead, 1989.

[17] 柳田益造，塚田　聡，角所　収，“高次優決定型重み付き線形予測分析，” 信学技報，SP87-14, pp.49–56, 1987.

[18] 深林太計志，“線形予測による音声分析の精度向上，” 信学論（A），vol.J61-A, no.11, pp.1168–1169, Nov. 1978.

[19] S.M. Kay, Modern Spectral Estimation: Theory and Application, Prentice-Hall, Englewood Cliffs, 1988.

[20] C. Lee, “On robust linear prediction of speech,” IEEE Trans., vol.ASSP-36, no.5, pp.642–650, May 1988.

[21] 吉田　誠，松本　弘，“高次 Yule-Walker 方程式による雑音中の母音スペクトルの推定，” 音声研資，S83-81, pp.643–650, 1984.

（平成 12 年 3 月 3 日受付，6 月 28 日再受付）

有馬　由紀　（学生員）

平 11 埼玉大・工・情報システム卒．同年日本テレコム入社．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．