Adaptive Cepstral Analysis of Speech

Keiichi Tokuda, *Member, IEEE,* Takao Kobayashi, *Member, IEEE,* and Satoshi Imai, *Member, IEEE*

Abstract— **This paper proposes an algorithm for adaptive cepstral analysis based on the UELS (unbiased estimation of log spectrum). In the UELS, the model spectrum is represented by cepstral coefficients and the mean square of the inverse filter output is minimized with respect to the cepstral coefficients. By introducing an instantaneous gradient estimate of the criterion in a similar manner of the LMS algorithm, we develop an adaptive cepstral analysis algorithm. In the analysis system, an IIR adaptive filter whose coefficients are given by cepstral coefficients is realized using the log magnitude approximation (LMA) filter. The filter approximates an exponential transfer function and its stability is guaranteed for approximation of speech spectra. To implement the Mth order cepstral analysis, the algorithm requires $O(M)$ operations per sample. It is shown that the algorithm has fast convergence properties in comparison with the LMS algorithm. Several examples of the adaptive cepstral analysis for synthetic signal and natural speech are shown to demonstrate the effectiveness of the algorithm.**

I. Introduction

RECENTLY, many adaptive signal processing algorithms have been proposed (e.g., LMS [1] and RLS [2] algorithms). Linear predictors that utilize these adaptive algorithms are useful in spectral estimation. In fact, the linear adaptive predictors have been applied to speech coding systems successfully [3]. However, they can not estimate zeros in a pole-zero spectral process such as nasalized speech, because they assume an all-pole spectral process for signal generation. Although IIR adaptive filters can estimate zeros, they have two problems: stability of the filter and uniqueness of the solution [1]. On the other hand, since the spectrum represented by a set of cepstral coefficients models poles and zeros with equal weights, the cepstrum [4] is a suitable parameter for representing the speech spectrum. It is, therefore, expected that if we develop an adaptive filter based on cepstral representation, we can use it to overcome problems involved in the IIR adaptive filters.

This paper proposes an algorithm for adaptive cepstral analysis based on the unbiased estimation of log spectrum (UELS) [5]. In the UELS, the model spectrum is represented by cepstral coefficients and a spectral criterion is minimized with respect to the cepstral coefficients. The criterion in the UELS is regarded as minimization of the mean square of inverse filter output. By introducing an instantaneous gradient estimate of the criterion in a similar manner of the LMS algorithm, we develop an adaptive cepstral analysis algorithm. In the analysis system, an IIR adaptive filter whose coefficients are given by the cepstral coefficients is realized using the LMA filter [6]. The LMA filter approximates an exponential transfer function and the stability is guaranteed for approximation of speech spectra.

At each iteration of the adaptive algorithm, the filter coefficients are updated using only the output vector, whereas the input vector and the output are needed to update the filter coefficients in the LMS algorithm. To implement the Mth order cepstral analysis, the algorithm requires $O(M)$ operations per sample. It is shown that the algorithm has fast convergence properties in comparison with the LMS algorithm.

The rest of the paper is organized as follows. In Section II, we give a brief review of the UELS and discuss the properties of the criterion. Based on these preliminaries we derive an algorithm for adaptive cepstral analysis in Section III. To implement the adaptive analysis system we need to realize the exponential transfer function. In Section IV, we describe a realization method of the exponential transfer function using the LMA filter. In Section V, examples of synthesized and natural speech analysis are shown to demonstrate the effectiveness of the algorithm. Conclusions are given in Section VI.

II. Cepstral Representation and Criterion

A. Unbiased Estimation of Log Spectrum

The UELS (unbiased estimation of log spectrum) [5] is a method for obtaining an unbiased log spectral estimator using logarithmic transformation and nonlinear smoothing of a periodogram. We assume that the model spectrum $H(e^{j\omega})$ is represented by the cepstral coefficients $c(m)$ up to the Mth coefficient as follows:

$$H(z) = \exp \sum_{m=0}^{M} c(m)\, z^{-m}. \tag{1}$$

Then the criterion in the UELS is regarded as minimization of

$$E = \frac{1}{2\pi} \int_{-\pi}^{\pi} \{\exp R(\omega) - R(\omega) - 1\}\, d\omega \tag{2}$$

with respect to $c(m)$, $m = 0, 1, \cdots, M$, where

$$R(\omega) = \log I_N(\omega) - \log \left| H(e^{j\omega}) \right|^2 \tag{3}$$

Manuscript received October 29, 1993; revised June 19, 1995. The associate editor coordinating the review of this paper and approving it for publication was Dr. Amro El-Jaroudi.

K. Tokuda is with the Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo, Japan.

T. Kobayashi and S. Imai are with the Precision and Intelligence Laboratory, Tokyo Institute of Technology, Yokohama, Japan.

IEEE Log Number 9414953.

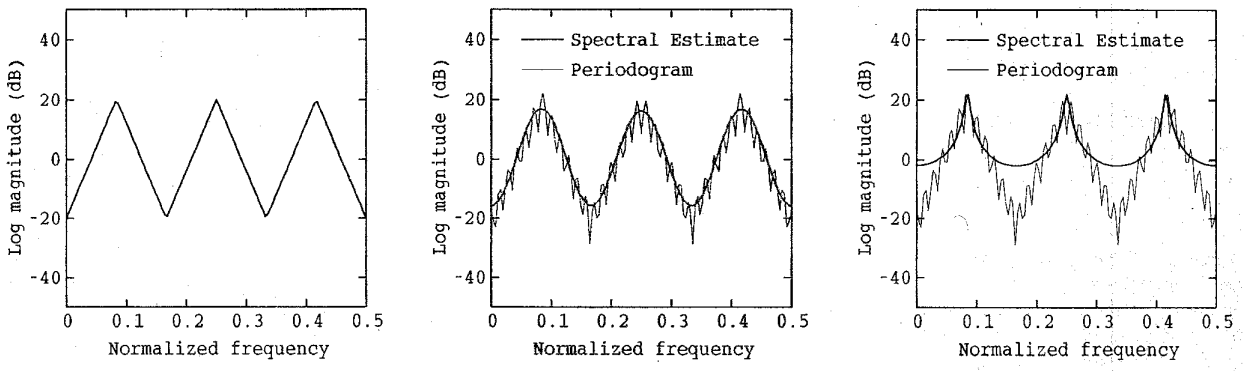


Fig. 1. Spectral estimates for a synthetic signal. (a) True spectrum; (b) UEL's ($M = 10$); and (c) linear prediction ($M = 10$).

and $I_N(\omega)$ is the modified periodogram of a weakly stationary process $x(n)$ with a window $w(n)$ whose length is N:

$$I_N(\omega) = \left| \sum_{n=0}^{N-1} w(n)\, x(n)\, e^{-j\omega n} \right|^2 \Big/ \sum_{n=0}^{N-1} w^2(n). \quad (4)$$

It is noted that the criterion has the same form as that in the maximum likelihood estimation of Gaussian stationary AR process [7]. Therefore, when the model spectrum is given by

$$H(z) = \frac{K}{1 + \sum_{k=1}^{M} a(k)\, z^{-k}} \quad (5)$$

instead of (1), minimizing (2) with respect to K and

$$\boldsymbol{a} = [a(1), a(2), \cdots, a(M)]^T \quad (6)$$

is equivalent to the linear prediction (LP) method [8]. Fig. 1(b) shows an example of spectral estimates for a synthetic signal whose spectrum is given in Fig. 1(a). From the figure it is seen that the UELS extracts both spectral peaks and valleys whereas the LP method cannot extract the valleys as shown in Fig. 1(c). Although the transfer function $H(z)$ is unrealizable as a finite order digital filter, it can be approximated by the log magnitude approximation (LMA) filter [6]. We will discuss the LMA filter in Section IV. Using the LMA filter, we can synthesize high quality speech from the cepstral coefficients obtained by the UELS.

B. Properties of the Criterion

Taking the gain factor $K = \exp c(0)$ outside from $H(z)$

$$H(z) = K\, D(z) \quad (7)$$

$$D(z) = \exp \sum_{m=1}^{M} c(m)\, z^{-m} \quad (8)$$

we rewrite (2) as

$$E = \varepsilon / K^2 + \log K^2 - \frac{1}{2\pi} \int_{-\pi}^{\pi} \log I_N(\omega)\, d\omega - 1 \quad (9)$$

where

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\, D(e^{j\omega})\,|^2}\, d\omega. \quad (10)$$

Thus, minimization of E is equivalent to that of ε with respect to

$$\boldsymbol{c} = [c(1), c(2), \cdots, c(M)]^T \quad (11)$$

and that of E with respect to K. Assuming the number of samples N in the window $w(n)$ is sufficiently large, we interpret (10) as the mean square of $e(n)$:

$$\varepsilon = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{|\, D(e^{j\omega})\,|^2}\, d\omega = E\left[e^2(n)\right] \quad (12)$$

where $P(\omega)$ is the power spectrum of the unknown system and $e(n)$ is the output of the inverse filter $1/D(z)$ driven by $x(n)$.

By setting $\partial E / \partial K = 0$, the gain factor K that minimizes E is obtained by

$$K = \sqrt{\varepsilon_{min}} \quad (13)$$

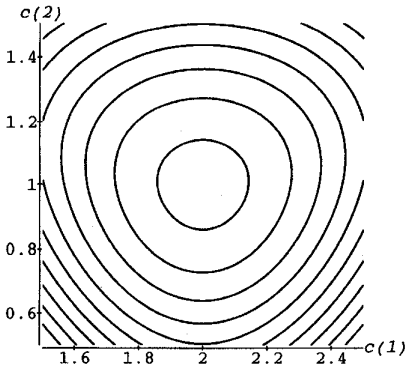
where ε_{min} is the minimized value of ε.

The gradient $\nabla\varepsilon$ and the Hessian matrix $\boldsymbol{H}$ of ε are given by

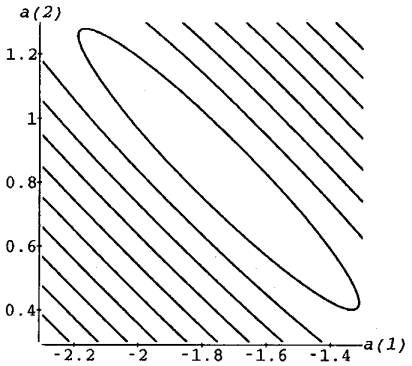
$$\nabla\varepsilon = \frac{\partial \varepsilon}{\partial \boldsymbol{c}} = -2\boldsymbol{r} = -2[r(1), r(2), \cdots, r(M)]^T \quad (14)$$

and

$$\boldsymbol{H} = \frac{\partial^2 \varepsilon}{\partial \boldsymbol{c} \partial \boldsymbol{c}^T} = 2 \left\{ \begin{bmatrix} r(0) & r(1) & \cdots & r(M-1) \\ r(1) & r(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r(1) \\ r(M-1) & \cdots & r(1) & r(0) \end{bmatrix} + \begin{bmatrix} r(2) & \cdots & r(M) & r(M+1) \\ \vdots & \iddots & \iddots & r(M+2) \\ r(M) & \iddots & \iddots & \vdots \\ r(M+1) & r(M+2) & \cdots & r(2M) \end{bmatrix} \right\} \quad (15)$$



(a)



(b)

Fig. 2 Contours of constant ε. (a) UEL's; (b) linear prediction.

respectively, where coefficients $r(m)$ are the autocorrelation coefficients of the inverse filter output $e(n)$:

$$r(m) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{P(\omega)}{|\,D(e^{j\omega})\,|^2}\, e^{j\omega m}\, d\omega = E[e(n)e(n-m)]. \tag{16}$$

Since $\boldsymbol{H}$ is always positive definite (see Appendix A), i.e., ε is convex downward with respect to $\boldsymbol{c}$, there is only a single global optimum. By setting $\nabla\varepsilon = \boldsymbol{0}$, we obtain a set of equations

$$\nabla\varepsilon = -2\boldsymbol{r} = \boldsymbol{0} \tag{17}$$

that gives the optimum.

At the optimum point, from (17) the Hessian matrix is given by

$$\boldsymbol{H} = 2\left\{\begin{bmatrix} r(0) & 0 & \cdots & 0 \\ 0 & r(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ 0 & \cdots & 0 & r(0) \end{bmatrix} + \begin{bmatrix} 0 & & \cdots & 0 & r(M+1) \\ \vdots & & \iddots & \iddots & r(M+2) \\ 0 & & \iddots & \iddots & \vdots \\ r(M+1) & r(M+2) & \cdots & & r(2M) \end{bmatrix}\right\} \tag{18}$$

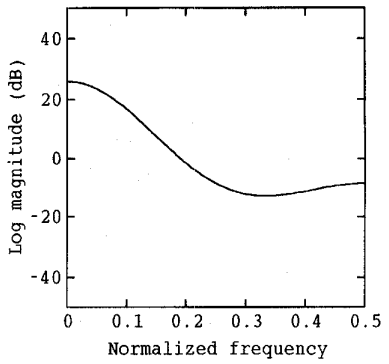


Fig. 3. Log magnitude response of the unknown system.

Furthermore, at the optimum point when $|\,D(z)\,|^2$ is equal to $P(\omega)$, the Hessian matrix becomes diagonal

$$\boldsymbol{H} = \mathrm{diag}[\underbrace{r(0), r(0), \cdots, r(0)}_{M}]. \tag{19}$$

Even in the case where $|\,D(z)\,|^2$ is not equal to $P(\omega)$ at the optimum, since we can assume

$$r(0) >> r(m), \qquad M < m \tag{20}$$

the Hessian matrix is approximated by (19). From the fact that the Hessian matrix $\boldsymbol{H}$ becomes diagonal and the diagonal elements are equal, i.e., the eigenvalues of $\boldsymbol{H}$ are equal, contours of constant ε become circles in the neighborhood of the minimum point, as can be seen in Fig. 2(a). The unknown spectrum for Fig. 2 is shown in Fig. 3. For comparison, contours for the linear predictor are shown in Fig. 2(b). In this case, the model spectrum is given by

$$D(z) = \frac{1}{1 + \sum_{k=1}^{M} a(k)\, z^{-k}} \tag{21}$$

then (12) is minimized with respect to $\boldsymbol{a}$. The contours for the linear predictor are ellipses unless the input signal to the predictor is uncorrelated [1], while the contours for the UELS are always circles in the neighborhood of the optimum, independent of correlation of the input signal.

III. Adaptive Algorithm

A. Derivation of the Algorithm for Adaptive Cepstral Analysis

From the above discussion, we expect that the minimization problem of ε can easily be solved by the method of steepest descent or the Newton–Raphson method [9], [10] (see Appendix B). In the method of steepest descent, from the ith result $\boldsymbol{c}^{(i)}$ the next result is obtained as follows:

$$\boldsymbol{c}^{(i+1)} = \boldsymbol{c}^{(i)} - \mu\,\nabla\varepsilon\Big|_{\boldsymbol{c}=\boldsymbol{c}^{(i)}}. \tag{22}$$

From (14), $\nabla\varepsilon$ is written as

$$\nabla\varepsilon = -2\,E\left[e(n)\,\boldsymbol{e}^{(n)}\right] \tag{23}$$

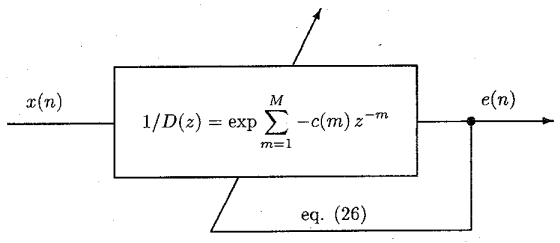


Fig. 4. Block diagram of the adaptive cepstral analysis system.

where $\boldsymbol{e}^{(n)}$ is the output vector

$$\boldsymbol{e}^{(n)} = [e(n-1), e(n-2), \cdots, e(n-M)]^T. \quad (24)$$

To develop an adaptive cepstral analysis algorithm, we introduce an instantaneous estimate [11] in a similar manner of the LMS algorithm [1]:

$$\hat{\nabla}\varepsilon^{(n)} = -2\,e(n)\,\boldsymbol{e}^{(n)}. \quad (25)$$

The mean value of $\hat{\nabla}\varepsilon^{(n)}$ is equal to the true gradient $\nabla\varepsilon$ when the coefficients vector $\boldsymbol{c}$ is held constant. With this estimate of the gradient, we specify the steepest descent type of adaptive algorithm: the coefficients vector $\boldsymbol{c}^{(n)}$ at time n is updated as

$$\boldsymbol{c}^{(n+1)} = \boldsymbol{c}^{(n)} + 2\,\mu\,e(n)\,\boldsymbol{e}^{(n)}. \quad (26)$$

In the neighborhood of the optimum, the function ε can be approximated by a quadratic function whose Hessian matrix is given by the diagonal matrix (19). The sufficient condition for convergence of the method of steepest descent with a quadratic function is

$$0 < \mu < \frac{1}{\mathrm{tr}\boldsymbol{H}}. \quad (27)$$

Thus, from (19), the sufficient condition for convergence of (22) in the neighborhood of the optimum is given by

$$0 < \mu < \frac{1}{M\,r(0)} = \frac{1}{M\,\varepsilon}. \quad (28)$$

On the other hand, it is difficult to examine the convergence of the adaptive algorithm (26) theoretically, since we use an imperfect gradient estimate in the algorithm. However, from the discussion of the criterion surface in section II and the fact that the efficiency of the LMS algorithm approaches a theoretical limit for adaptive algorithm when the eigenvalues of the Hessian matrix are equal [1], the proposed algorithm should have fast convergence properties independent of whether the input signal is correlated.

Instead of (25), other gradient estimates can also be used. For example, the gradient is estimated with an exponential window [10]:

$$\bar{\nabla}\varepsilon^{(n)} = -2\,(1-\tau)\sum_{i=-\infty}^{n}\tau^{n-i}\,e(i)\,\boldsymbol{e}^{(i)} \quad (29)$$

where τ is a constant in a range $0 \le \tau < 1$. Such an estimate can be calculated recursively by

$$\bar{\nabla}\varepsilon^{(n)} = \tau\bar{\nabla}\varepsilon^{(n-1)} - 2(1-\tau)e(n)\boldsymbol{e}^{(n)}. \quad (30)$$

Although we can use the estimate $\bar{\nabla}\varepsilon^{(n)}$ to suppress fluctuation of $\boldsymbol{c}$, this paper uses the instantaneous estimate (25) to simplify the discussion.

Fig. 4 shows the adaptive cepstral analysis system. The coefficients of the adaptive filter that has an exponential transfer function are updated by (26). At each iteration of the adaptive algorithm, the filter coefficients are updated using the output vector $\boldsymbol{e}^{(n)}$ and the output $e(n)$, whereas the input vector and the output $e(n)$ are needed to update the filter coefficients in the LMS algorithm [1]. To implement the Mth order cepstral analysis, the algorithm (26) requires $2M$ operations per sample. The transfer function $1/D(z)$ is minimum phase, but it is unrealizable as a finite order digital filter because it is not a rational function. We will discuss a realization method of the inverse filter $1/D(z)$ and its stability in the section IV.

B. Estimation of the Gain

An estimate of ε at time n is given by

$$\varepsilon^{(n)} = (1-\lambda)\sum_{i=-\infty}^{n}\lambda^{n-i}\,e^2(i) \quad (31)$$

where λ is a constant in a range $0 \le \lambda < 1$. This can be calculated in the same way as (30):

$$\varepsilon^{(n)} = \lambda\,\varepsilon^{(n-1)} + (1-\lambda)\,e^2(n). \quad (32)$$

From (13), we get an estimate of K at time n by

$$K^{(n)} = \sqrt{\varepsilon^{(n)}}. \quad (33)$$

When the gain of the signal is time-varying, μ is normalized by $\varepsilon^{(n)}$ as follows:

$$\mu^{(n)} = \frac{\alpha}{M\,\varepsilon^{(n)}} \quad (34)$$

where α is a constant in a range $0 < \alpha < 1$. This equation is derived by the analogy of the convergence condition (28) of the method of steepest descent (22).

IV. Exponential Transfer Function

A. Realization of Exponential Transfer Function

We realize the exponential transfer functions $D(z)$ and $1/D(z)$ in Fig. 4 using the LMA filter [6]. The complex exponential function $\exp w$ is approximated by a rational function

$$\exp w \simeq R_L(w) = \frac{1+\sum_{l=1}^{L}A_{L,l}\,w^l}{1+\sum_{l=1}^{L}A_{L,l}\,(-w)^l}. \quad (35)$$

For example, if we choose $A_{L,l}$ as

$$A_{L,l} = \frac{1}{l!}\binom{L}{l}\Big/\binom{2L}{l} \quad (36)$$

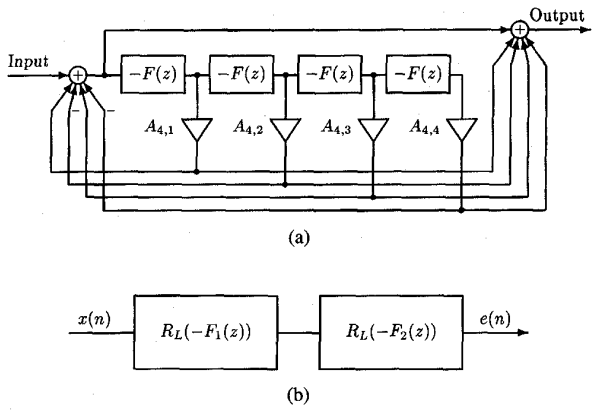


Fig. 5. Realization of the exponential transfer function $1/D(z)$. (a) $R_L(-F9z)) \simeq 1/D(z)L = 4$. (b) Two-stage cascade structure $R_L(-F_1(z)) \cdot R_L(-F_2(z)) \simeq 1/D(z)$.

TABLE I
COEFFICIENTS OF $R_4(w)$

l	$A_{4,l}$
1	4.999273×10^{-1}
2	1.067005×10^{-1}
3	1.170221×10^{-2}
4	5.656279×10^{-4}

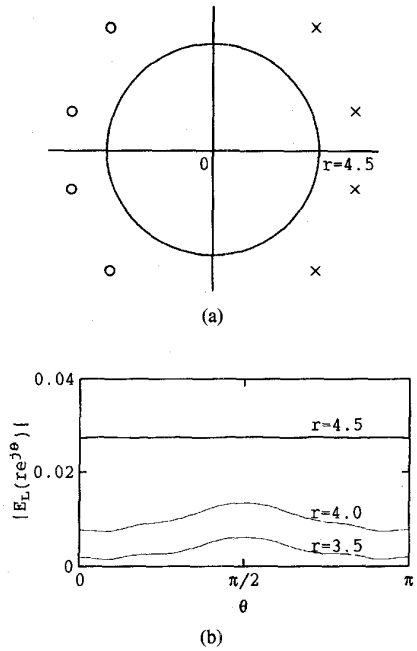


Fig. 6. (a) Pole-zero plot for $R_L(w)$. (b) Error $E_L(r\,e^{j\theta})$. ($L = 4$)

then (35) is the $[L/L]$ Padé approximant of $\exp w$ at $w = 0$. Thus, $D(z)$ and $1/D(z)$ are approximated by

$$R_L(F(z)) \simeq \exp(F(z)) = D(z), \tag{37}$$

$$R_L(-F(z)) \simeq \exp(-F(z)) = 1/D(z) \tag{38}$$

respectively, where $F(z)$ is defined by

$$F(z) = \sum_{m=1}^{M} c(m)\, z^{-m}. \tag{39}$$

Since we can realize $R_L(F(z)) \simeq D(z)$ and $R_L(-F(z)) \simeq 1/D(z)$ in the same manner except for changing the sign of $F(z)$, we will discuss the realization of $1/D(z)$ in the following.

Fig. 5(a) shows the block diagram of the LMA filter $R_L(-F(z)) \simeq 1/D(z)$. Since the transfer function $F(z)$ has no delay-free path, $R_L(-F(z))$ has no delay-free loops, that is, $R_L(-F(z))$ is realizable. The transfer function $R_L(-F(z))$ is also realized in a variety of other structures (for an example, [12]).

B. Approximation Accuracy and Stability

If $c(1), c(2), \cdots, c(M)$ are bounded, $|F(e^{j\omega})|$ is also bounded and there exists a positive finite value r such that

$$\max_{\omega} |F(e^{j\omega})| < r. \tag{40}$$

We can obtain the coefficients $A_{L,l}$, $l = 1, 2, \cdots, L$ which minimize the maximum of the absolute error $\max_{|w|=r} |E_L(w)|$ using a complex Chebyshev approximation technique [13], where

$$E_L(w) = \log(\exp w) - \log(R_L(w)). \tag{41}$$

The coefficients obtained with $L = 4, r = 4.5$ are shown in Table I. From Fig. 6(b), it is seen that the maximum of the error $\max_{|w|=r} |E_L(w)|$ is 0.028 (0.24 dB). The rational function $R_L(w)$ given by Table I has no poles and zeros in the region $|w| < r_{max} = 6.2$ as shown in Fig. 6(a). Therefore, from the maximum principle, the error $|E_L(w)|$ does not exceed 0.028 (0.24dB) in the region $|w| < r = 4.5$. Consequently, when $|F(e^{j\omega})| < r = 4.5$, the error

$$|E_L(-F(e^{j\omega}))| = |\log(1/D(e^{j\omega})) - \log R_L(-F(e^{j\omega}))| \tag{42}$$

does not exceed 0.028 (0.24 dB).

Since $F(z)$ has no poles except $z = 0$, from the maximum principle, (40) is rewritten as

$$\max_{|z|\geq 1} |F(z)| < r. \tag{43}$$

Therefore, since $R_L(w)$ has no poles and zeros in the region $|w| < r_{max} = 6.2$, under the condition that $r < r_{\max}$of (40), $R_4(-F(z))$ has no poles and zeros for $|z| \geq 1$; i.e., it becomes a minimum phase system.

C. Cascade Structure

When $F(z)$ is expressed as

$$F(z) = F_1(z) + F_2(z) \tag{44}$$

the exponential transfer function $1/D(z)$ is approximated in a cascade form

$$\begin{aligned} 1/D(z) &= \exp(-F(z)) = \exp(-F_1(z)) \cdot \exp(-F_2(z)) \\ &\simeq R_L(-F_1(z)) \cdot R_L(-F_2(z)) \end{aligned} \tag{45}$$

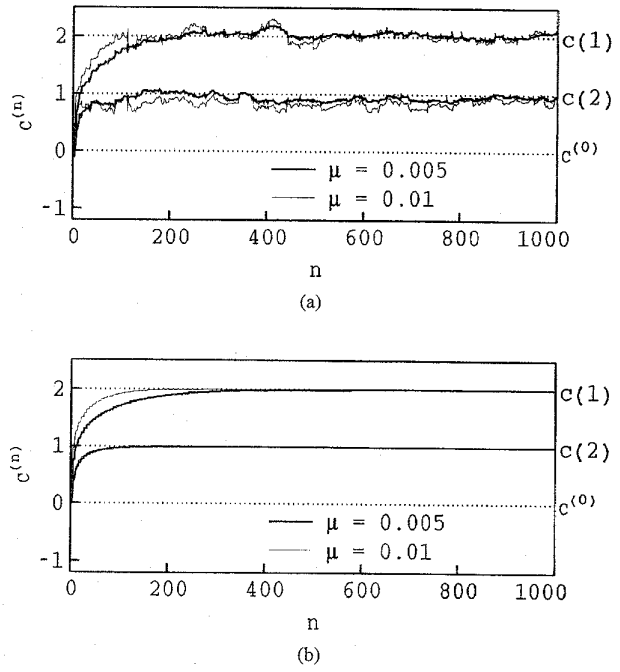


Fig. 7. Convergence characteristics for the adaptive cepstral analysis. (a) White noise. (b) Pulse train.

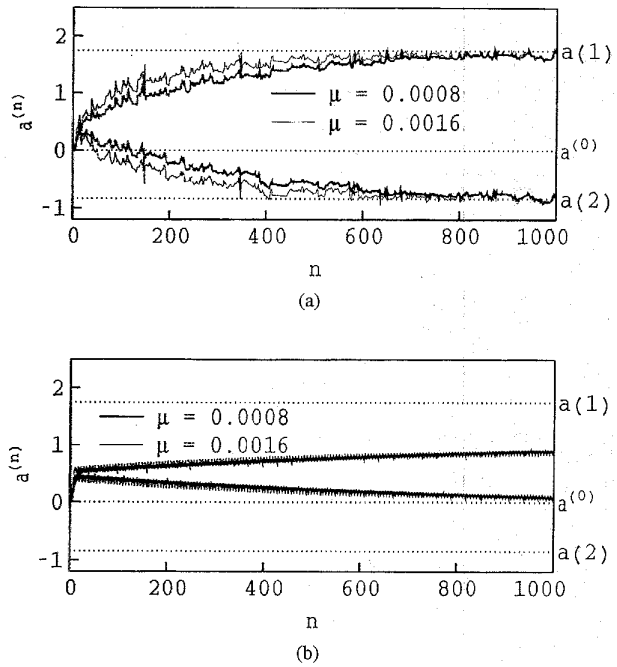


Fig. 8. Convergence characteristics for the LMS linear predictor.

as shown in Fig. 5(b). If

$$\max_{\omega} \left| F_1(e^{j\omega}) \right|, \max_{\omega} \left| F_2(e^{j\omega}) \right| < \max_{\omega} \left| F(e^{j\omega}) \right| \quad (46)$$

it is expected that $R_L(-F_1(e^{j\omega})) \cdot R_L(-F_2(e^{j\omega}))$ approximates $1/D(e^{j\omega})$ more accurately than $R_L(-F(e^{j\omega}))$.

In the following experiments, we let

$$F_1(z) = c(1)\, z^{-1} \quad (47)$$

$$F_2(z) = \sum_{m=2}^{M} c(m)\, z^{-m}. \quad (48)$$

Since we have empirically found that

$$\max_{\omega} \left| F_1(e^{j\omega}) \right|, \max_{\omega} \left| F_2(e^{j\omega}) \right| < r = 4.5 \quad (49)$$

for speech sounds sampled at 10kHz, $R_L(-F_1(z)) \cdot R_L(-F_2(z))$ approximates the exponential transfer function $1/D(z)$ with sufficient accuracy and becomes a minimum phase system.

The LMA filter shown in Fig. 5 requires $4M + O(1)$ multiply-add operations per sample. Thus, the total number of multiply-add operations for the adaptive cepstral analysis is $5M + O(1)$: $M + O(1)$ for the adaptive algorithm (26) and the normalization of the step size (32), (34), and $4M + O(1)$ for the LMA filter.

V. Examples

A. Simulation Results

In order to produce signals to be analyzed, the LMA filter with $M = 2$, which has the log magnitude response shown in Fig. 3, was driven by a white Gaussian noise or a pulse train with unit variance. The period of the pulse train is six samples. The algorithm was implemented with $M = 2$, $\boldsymbol{c}^{(0)} = \mathbf{0}$, and $\mu = 0.005, 0.01$.

The results are shown in Fig. 7. For comparison, the performance of the LMS linear predictor is shown in Fig. 8. The conditions are the same as in Fig. 7 except that $\mu = 0.0008, 0.0016$. Note that $\mu = 0.0016$ is the largest value before divergence with the white noise. It is seen from the figure that the proposed algorithm has fast convergence characteristics, while the LMS linear predictor needs more than 1000 iterations to converge. This result coincides with the discussion in Section III-A.

Fig. 7(b) shows that the coefficients vector $\boldsymbol{c}$ does not vary after convergence when the unknown system is driven by a pulse train. This property is based on the following facts: when $D(z)$ is equal to the transfer function of the unknown system, the inverse filter output $e(n)$ becomes a pulse train; hence, if we assume that the period of the pulse train is greater than M, the estimated gradient (25) becomes zero.

B. Analysis of Natural Speech

Fig. 9 shows the result of the adaptive cepstral analysis for a natural speech. The signal shown in Fig. 9(a) is the natural English speech "two zero eight six" uttered by a male. It was sampled at 10kHz (sampling rate 100μs). The algorithm was implemented with $M = 25$, $\alpha = 0.2$, $\lambda = 0.98$, and μ is normalized by (34). In Fig. 9(b), the thin line shows the coefficients versus the iteration number for the proposed algorithm. For comparison, the coefficients obtained by the UELS are also shown in Fig. 9(b) by the thick line. The UELS was carried out by weighting the signal with a 25.6ms Blackman window with frame shift of 10 ms. Fig. 9(c) shows log magnitude spectra obtained from the cepstral coefficients at intervals of 10ms (100 samples). From

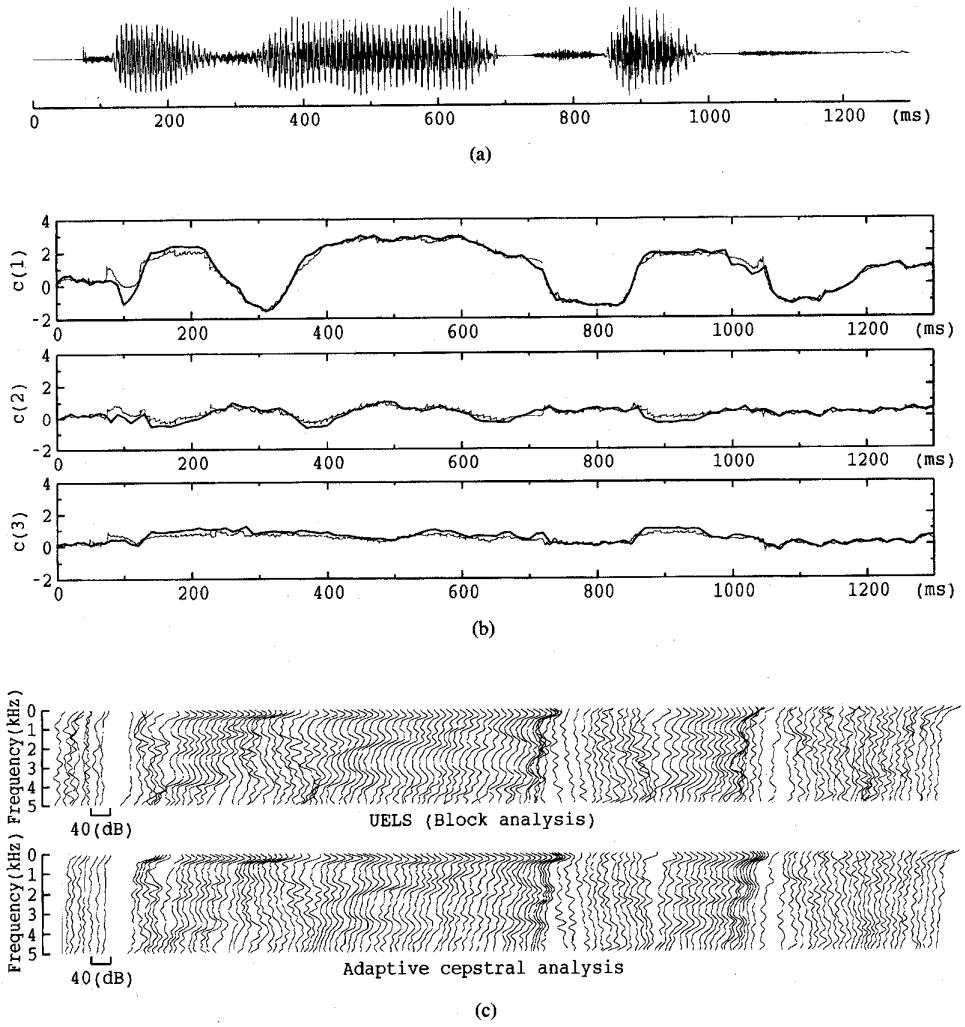


Fig. 9. Example of the adaptive cepstral analysis for natural speech ($M = 25$, $\alpha = 0.2$, $\lambda = 0.98$). (a) Waveform. (b) Cepstral coefficients. (c) Spectral estimates.

Fig. 9, it is seen that the proposed algorithm has sufficiently fast convergence characteristics for speech analysis. The UELS and the conventional cepstrum analysis [4] need several times of FFT to obtain a set of cepstral coefficients. Thus, the number of operations per sample for the adaptive cepstral analysis is considerably small compared with the UELS or the conventional cepstrum analysis, especially, in the case where the frame shift is small. In spite of that, the quality of the synthesized speech based on the adaptive cepstral analysis is only slightly inferior to that based on the UELS.

The adaptive cepstral analysis system has been implemented with a general-purpose 32-bit floating-point DSP (NEC μPD77230). The speed of operation per sample is 59 μs: 30 μs for filtering by the LMA filter, and 29 μs for coefficients update and gain estimation. Consequently, it can easily run at sampling frequency of 10kHz.

VI. Conclusion

In this paper, we have presented an algorithm for adaptive cepstral analysis. The proposed adaptive analysis system is implemented with an IIR adaptive filter whose coefficients are given by cepstral coefficients. The stability of the filter is guaranteed for approximation of speech spectra. The adaptive cepstral analysis requires $O(M)$ operations to obtain the cepstral coefficients up to Mth coefficient sample-by-sample, and has fast convergence properties in comparison with the LMS algorithm. A real-time speech analysis system can easily be implemented with one currently available DSP. The proposed method has been developed to adaptive mel-cepstral analysis [10], and its potential applications to speech recognition [10], speech coding [14], adaptive equalization, echo canceling, etc., are currently investigated. Development of the RLS-type algorithm for adaptive cepstral analysis is also a future research problem.

Appendix A
Proof that H Is Positive Definite

In this section, we show that Hessian matrix $\boldsymbol{H}$ is positive definite, i.e.,

$$\boldsymbol{x}^T \boldsymbol{H} \boldsymbol{x} > 0, \qquad \boldsymbol{x} \neq \boldsymbol{0} \tag{50}$$

where

$$\boldsymbol{x} = [x_1, x_2, \cdots, x_M]^T. \tag{51}$$

The left side of equation (50) is rewritten as follows:

$$\boldsymbol{x}^T \boldsymbol{H}\, \boldsymbol{x} = \frac{1}{2}\, \boldsymbol{y}^T \boldsymbol{R}_{2M}\, \boldsymbol{y} \tag{52}$$

where

$$\boldsymbol{y} = [x_M, \cdots, x_2, x_1, 0, x_1, x_2, \cdots, x_M]^T \tag{53}$$

$$\boldsymbol{R}_{2M} = \begin{bmatrix} r(0) & r(1) & \cdots & r(2M) \\ r(1) & r(0) & \ddots & \vdots \\ \vdots & \ddots & \ddots & r(1) \\ r(2M) & \cdots & r(1) & r(0) \end{bmatrix} \tag{54}$$

It follows from (16) that

$$\begin{aligned} \boldsymbol{x}^T \boldsymbol{H}\, \boldsymbol{x} &= \frac{1}{2}\, \boldsymbol{y}^T \boldsymbol{R}_{2M}\, \boldsymbol{y} \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{P(\omega)}{|\, D(e^{j\omega})\,|^2} \left| \sum_{m=1}^{M} x_m \cos(m\omega) \right|^2 d\omega \\ &> 0, \qquad \boldsymbol{x} \neq \boldsymbol{0}. \end{aligned} \tag{55}$$

Appendix B
Newton-Raphson Method

For the ith result $\boldsymbol{c}^{(i)}$, solving a set of linear equations

$$\boldsymbol{H}\, \Delta \boldsymbol{c}^{(i)} = -\nabla \varepsilon \Big|_{\boldsymbol{c} = \boldsymbol{c}^{(i)}}, \tag{56}$$

we have the values

$$\Delta \boldsymbol{c}^{(i)} = [\Delta c^{(i)}(1), \Delta c^{(i)}(2), \cdots, \Delta c^{(i)}(M)]^T. \tag{57}$$

Then, the next result is obtained as follows:

$$\boldsymbol{c}^{(i+1)} = \boldsymbol{c}^{(i)} + \Delta \boldsymbol{c}^{(i)}. \tag{58}$$

When the approximation of (12) is not used, coefficients $\{r(m)\}_{m=0}^{2M}$ are given by

$$r(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{I_N(\omega)}{|\, D(e^{j\omega})\,|^2}\, e^{j\omega m}\, d\omega \tag{59}$$

and they are calculated efficiently using the FFT. We can use the FFT cepstrum as an initial guess $\boldsymbol{c}^{(0)}$. The convergence is quadratic because the Hessian matrix is positive definite even if coefficients $\{r(m)\}_{m=0}^{2M}$ are given by (59). We have found that typically a few iterations are sufficient to obtain the solution.

Since the matrix $\boldsymbol{H}$ is a symmetric Toeplitz plus Hankel matrix, (56) can be solved using fast recursive algorithms (e.g., [15]) that require $O(M^2)$ arithmetic operations. Further reduction of the computational complexity can be obtained as follows. Equation (56) is rewritten as follows:

$$\boldsymbol{R}_{2M}\, \mathbf{d}_M = \sigma_M^2\, [\underbrace{0, \cdots, 0}_{M}, 1, \underbrace{0, \cdots, 0}_{M}]^T \tag{60}$$

where $\boldsymbol{R}_{2M}$ is given by (54) and

$$\mathbf{d}_M = [d_M(M), \cdots, d_M(1), 1, d_M(1), \cdots, d_M(M)]^T \tag{61}$$

$$d_M(m) = -\Delta c^{(i)}(m), \qquad m = 1, 2, \cdots, M. \tag{62}$$

We can solve (60) with $O(M^2)$ operations, since it has the same form as the normal equation for linear prediction with linear phase [16]. We can also use an algorithm [17] whose number of operations is reduced to about half of [16] using the symmetric property of (60).

References

[1] B. Widrow and S. D. Stearns, *Adaptive Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.

[2] S. T. Alexander, *Adaptive Signal Processing*. Berlin, Vienna, New York: Springer-Verlag, 1986.

[3] J. D. Gibson, "Adaptive prediction for speech coding," *IEEE ASSP Mag.*, vol. 1, pp. 12–26, July 1984.

[4] A. V. Oppenheim and R. W. Schafer, *Descrete-Time Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1989.

[5] S. Imai and C. Furuichi, "Unbiased estimator of log spectrum and its application to speech signal processing," in *Proc. 1988 EURASIP*, Sep. 1988, pp. 203–206.

[6] S. Imai, "Log magnitude approximation (LMA) filter," *Trans. IECE*, vol. J63-A, pp. 886–893, Dec. 1980 (in Japanese).

[7] F. Itakura and S. Saito, "Speech information compression based on the maximum likelihood spectral estimation," *J. Acoust. Soc. Japan*, vol. 27, no. 9, pp. 463–472, Sept. 1971 (in Japanese).

[8] J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*. Berlin, Vienna, New York: Springer-Verlag, 1986.

[9] K. Tokuda, T. Kobayashi, and S. Imai, "Generalized cepstral analysis of speech—Unified approach to LPC and cepstral method," in *Proc. ICSLP-90*, 1990, pp. 37–40.

[10] T. Fukada, K. Tokuda, T. Kobayashi, and S. Imai, "An adaptive algorithm for mel-cepstral analysis of speech," in *Proc. ICASSP-92*, 1992, pp.I-137–I-140.

[11] K. Tokuda, T. Kobayashi, S. Shiomoto, and S. Imai, "Adaptive filtering based on cepstral representation—Adaptive cepstral analysis of speech," in *Proc. ICASSP-90*, 1990, pp. 377–380.

[12] T. Kobayashi, K. Fukushi, K. Tokuda, and S. Imai, "2-D LMA filters—Design of stable two-dimensional digital filters with arbitrary magnitude function," *Trans. IEICE*, vol. E75-A, pp. 240–246, Feb. 1992.

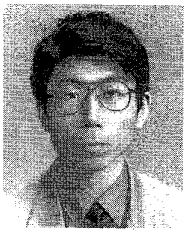
[13] T. Kobayashi and S. Imai, "Complex Chebyshev approximation for IIR digital filters using an iterative WLS technique," in *Proc. ICASSP-90*, 1990, pp. 1321–1324.

[14] K. Tokuda, H. Mastumura, T. Kobayashi, and S. Imai, "Speech coding based on adaptive mel-cepstral analysis," in *Proc. ICASSP-94*, 1994, pp.I-134–I-137.

[15] A. E. Yagle, "New analogs of split algorithms for arbitrary Toeplitz-plus-Hankel matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 39, pp. 2457–2463, Nov. 1991.

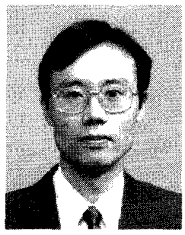
[16] S. L. Marple, Jr., "Fast algorithm for linear prediction and system identification filters with linear phase," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-30, pp. 942–953, Dec. 1982.

[17] K. Tokuda, A. Hashimoto, M. Kaneko, "A fast algorithm for linear prediction with linear phase," *Proc. IEICE Nat. Conf.*, p. 1–137, Mar. 1993.



Keiichi Tokuda (M'89) was born in Nagoya, Japan, in 1960. He received the B.E. degree in electrical and electronic engineering from the Nagoya Institute of Technology, Nagoya, Japan, and the M.E. and Dr.Eng. degrees in information processing from the Tokyo Institute of Technology, Tokyo, Japan, in 1984, 1986, and 1989, respectively.

He is currently a Research Associate at the Department of Electronic and Electric Engineering, Tokyo Institute of Technology. His research interests include speech spectral estimation, speech coding, speech synthesis and recognition, and adaptive signal processing.



Takao Kobayashi (M'82) was born in Niigata, Japan, in 1955. He received the B.E. degree in electrical engineering, and the M.E. and Dr.Eng. degrees in information processing from Tokyo Institute of Technology, Tokyo, Japan, in 1977, 1979, and 1982, respectively.

In 1982, he joined the Research Laboratory of Precision Machinery and Electronics, Tokyo Institute of Technology, as a Research Associate. He is currently an Associate Professor at the Precision and Intelligence Laboratory, Tokyo Institute of Technology. His research interests include digital signal processing, speech analysis and synthesis, speech recognition, and neural networks.

Satoshi Imai (M'75) received the B.E., M.E., and D.Eng. degrees in electrical engineering from the Tokyo Institute of Technology, Tokyo, in 1959, 1961, and 1964, respectively. In 1964, he joined the Research Laboratory of Precision Machinery and Electronics at Tokyo Institute of Technology as a Research Associate, and he became an Associate Professor of Tokyo Institute of Technology in 1968.

He is currently a Professor at Tokyo Institute of Technology. His current research interests are in the area of speaker-independent large vocabulary continuous speech recognition, natural sounding speech synthesis from text, natural language processing for speech recognition, logarithmic spectral estimation of speech, and speech coding.