MEL-GENERALIZED CEPSTRAL ANALYSIS — A UNIFIED APPROACH TO SPEECH SPECTRAL ESTIMATION

Keiichi Tokuda†, *Takao Kobayashi*††, *Takashi Masuko*†† *and* *Satoshi Imai*††

†Department of Electrical and Electronic Engineering, Tokyo Institute of Technology, Tokyo, 152 Japan
††Precision and Intelligence Laboratory, Tokyo Institute of Technology, Yokohama, 227 Japan

ABSTRACT

The generalized cepstral analysis method is viewed as a unified approach to the cepstral method and the linear prediction method, in which the model spectrum varies continuously from all-pole to cepstral according to the value of a parameter γ. Since the human ear has high resolution at low frequencies, introducing similar characteristics to the model spectrum, we can represent speech spectrum more efficiently. From this point of view, this paper proposes a spectral estimation method which uses the spectral model represented by mel-generalized cepstral coefficients. The effectiveness of mel-generalized cepstral analysis is demonstrated by an experiment of HMM-based isolated word recognition.

I. INTRODUCTION

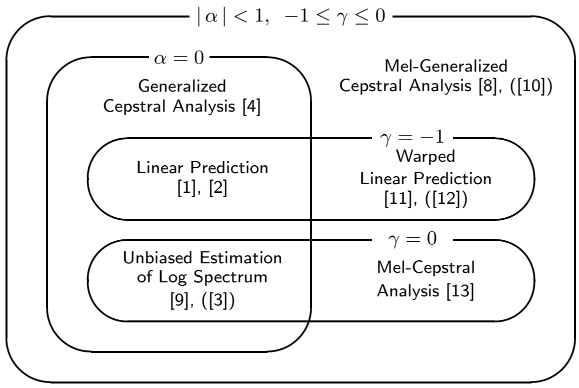
Linear prediction [1], [2] is a generally accepted method for obtaining all-pole representation of speech. However, in some cases, spectral zeros are important and a more general modeling procedure is required. Although cepstral modeling can represent poles and zeros with equal weights, the cepstral method [3] with a small number of cepstral coefficients overestimates the bandwidths of the formants. To overcome this problem, we proposed the generalized cepstral analysis method [4], [5]. The generalized cepstral coefficients [6] are identical with the cepstral and AR coefficients when a parameter γ equals 0 and -1, respectively. Thus, utilizing the generalized cepstral representation, we can vary the model spectrum continuously from the all-pole spectrum to that represented by the cepstrum according to the value of γ.

Since the human ear has high resolution at low frequencies, introducing the similar characteristics to the model spectrum, we can represent speech spectrum more efficiently. The spectrum represented by the mel-generalized cepstrum [7], i.e., frequency-transformed generalized cepstrum, has frequency resolution similar to that of the human ear with an appropriate choice of the value of a parameter α. Hence, it is expected that the mel-generalized cepstral coefficients are useful for speech spectral representation.

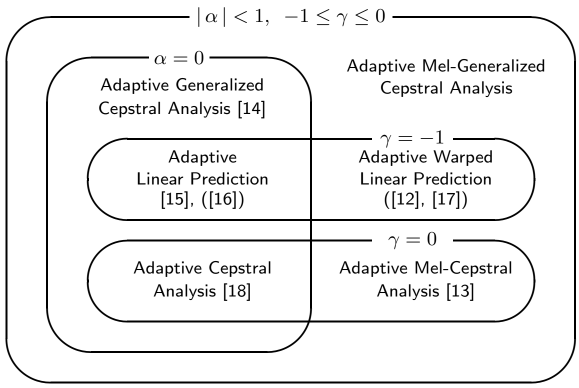
From the above point of view, this paper proposes a mel-generalized cepstral analysis method [8], in which we apply the criterion used in the UELS (unbiased estimation of log spectrum) [9] to the spectral model based on the mel-generalized cepstral representation. It can be shown that the minimization of the criterion is equivalent to the minimization of the mean square of the linear prediction error. As a result, the proposed method can be viewed as a unified approach to speech spectral analysis, which includes several speech analysis methods (see Fig. 1). The detailed discussion is given in Section VI.

Although the method involves a non-linear minimization problem, it can easily be solved by an iterative algorithm [8]. The convergence is quadratic and typically a few iterations are sufficient to obtain the solution. We can also show that the stability of the obtained model solution is guaranteed [8].

Finally, we show some simulation results of natural speech analysis to demonstrate that the characteristic of the obtained spectrum varies according to the values of α and γ. It is shown that we can improve the performance of speech recognition by choosing the values of α and γ.



(a) Block analysis (frame basis)



(b) Adaptive analysis (sample by sample basis)

Fig. 1. A unified view of speech spectral analysis. (References in parentheses are closely related methods.)

II. SPECTRAL MODEL AND CRITERION

The generalized logarithmic function [6] is a natural generalization of the logarithmic function:

$$s_\gamma(w) = \begin{cases} (w^\gamma - 1)/\gamma, & 0 < |\gamma| \le 1 \\ \log w, & \gamma = 0 \end{cases}. \tag{1}$$

The cepstrum $c(m)$ of a real sequence $x(n)$ is defined as the inverse Fourier transform of the logarithmic spectrum, while the mel-generalized cepstrum $c_{\alpha,\gamma}(m)$ is defined as the inverse Fourier transform of the generalized logarithmic spectrum calculated on a warped frequency scale $\beta_\alpha(\omega)$:

$$s_\gamma\big(X(e^{j\omega})\big) = \sum_{m=-\infty}^{\infty} c_{\alpha,\gamma}(m)\, e^{-j\beta_\alpha(\omega)\, m} \tag{2}$$

TABLE I Spectral representation based on mel-generalized cepstrum. (equation (5)).

	$\alpha = 0$	$\mid\alpha\mid < 1$
$\gamma = -1$	all-pole	warped all-pole
$\gamma = 0$	cepstral	mel-cepstral
$\gamma = 1$	all-zero	warped all-zero
$-1 \le \gamma \le 1$	generalized cepstral	mel-generalized cepstral

where $X(e^{j\omega})$ is the Fourier transform of $x(n)$. The warped frequency scale $\beta_\alpha(\omega)$ is defined as the phase response of an all-pass system

$$\Psi_\alpha(z) = \left.\frac{z^{-1} - \alpha}{1 - \alpha\, z^{-1}}\right|_{z=e^{j\omega}} = e^{-j\beta_\alpha(\omega)}, \quad |\,\alpha\,| < 1 \tag{3}$$

where

$$\beta_\alpha(\omega) = \tan^{-1} \frac{(1-\alpha^2)\sin\omega}{(1+\alpha^2)\cos\omega - 2\alpha}. \tag{4}$$

The phase response $\beta_\alpha(\omega)$ gives a good approximation to auditory frequency scales with an appropriate choice of α [12], [19].

In this paper, we assume that a speech spectrum $H(e^{j\omega})$ can be modeled by the $M+1$ mel-generalized cepstral coefficients as follows:

$$H(z) = s_\gamma^{-1}\left(\sum_{m=0}^{M} c_{\alpha,\gamma}(m)\,\Psi_\alpha^m(z)\right) = \begin{cases} \left(1 + \gamma \displaystyle\sum_{m=0}^{M} c_{\alpha,\gamma}(m)\,\Psi_\alpha^m(z)\right)^{1/\gamma}, & 0 < |\,\gamma\,| \le 1 \\ \exp \displaystyle\sum_{m=0}^{M} c_{\alpha,\gamma}(m)\,\Psi_\alpha^m(z), & \gamma = 0 \end{cases}. \tag{5}$$

From (5), it is seen that for $(\alpha, \gamma) = (0, -1)$ the model spectrum takes the form of the all-pole representation and for $(\alpha, \gamma) = (0, 0)$ the model spectrum is identical with the spectrum represented by the cepstrum. The relation between the form of the model spectrum and the values of (α, γ) is summarized in Table I.

We determine $\mathbf{c} = [c_{\alpha,\gamma}(0), c_{\alpha,\gamma}(1), \ldots, c_{\alpha,\gamma}(M)]^T$ in such a way that the following spectral criterion derived in the UELS [9] is minimized:

$$E = \frac{1}{2\pi}\int_{-\pi}^{\pi} \{\exp R(\omega) - R(\omega) - 1\}\, d\omega \tag{6}$$

where

$$R(\omega) = \log I_N(\omega) - \log\left|\,H(e^{j\omega})\,\right|^2 \tag{7}$$

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

$$I_N(\omega) = \left|\sum_{n=0}^{N-1} w(n)\,x(n)\,e^{-j\omega n}\right|^2 \Bigg/ \sum_{n=0}^{N-1} w^2(n). \tag{8}$$

When $(\alpha, \gamma) = (0, 0)$, the proposed method is identical with the UELS, because in the UELS the estimator of the log spectrum is represented by the cepstrum. Furthermore, (6) has the same form as the spectral criterion in the maximum likelihood estimation of Gaussian stationary AR process [1]. Therefore, when $(\alpha, \gamma) = (0, -1)$, the proposed method is identical with the linear prediction method [1], [2]. As shown in Fig. 1, some other spectral estimation methods are also included in the proposed method as special cases.

Since E is convex with respect to $\mathbf{c}$ when $-1 \le \gamma \le 0$,

$$\nabla E = \frac{\partial E}{\partial \mathbf{c}} = \mathbf{0} \tag{9}$$

gives the global minimum [8]. When $\gamma = -1$, (9) can be solved directly, because (9) becomes a set of linear equations. Especially, when $(\alpha, \gamma), = (0, -1)$, (9) corresponds to the normal equations in the linear prediction method. In other cases, to solve the minimization problem, we have given an iterative algorithm [8] whose convergence is quadratic. We have found that typically a few iterations are sufficient to obtain the solution. Furthermore, the stability of the model solution $H(z)$ is guaranteed when $-1 \le \gamma \le 0$ [8].

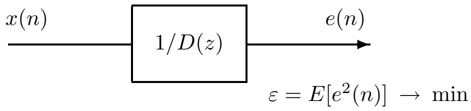


Fig. 2. Interpretation of the proposed spectral analysis as the least mean square of the linear prediction error $e(n)$.

Assuming that $D(z)$ is the gain-normalized version of $H(z)$, i.e., the 0th impulse responses of $D(z)$ and $1/D(z)$ are constrained to unity, the inverse filter output $e(n)$ shown in Fig. 2 is the prediction error [4]. It can be shown that minimizing E with respect to $H(z)$ is equivalent to minimizing

$$\varepsilon = E\left[e^2(n)\right] \tag{10}$$

with respect to $D(z)$ [8]. Thus, the proposed method can be interpreted as the least mean square of the linear prediction error as shown in Fig. 2.

III. EXAMPLE OF SPEECH ANALYSIS

Fig. 3 shows the estimated spectra for a Japanese sentence "naNbudewa" uttered by a male. A sampling frequency of 10kHz was used. The signal was windowed by a 25.6ms Blackman window with a 5ms shift and then analyzed using the mel-generalized cepstral analysis method with $M = 12$ and several values of (α, γ). The spectra for $(\alpha, \gamma) = (0, -1)$, $(0, 0)$, and $(0.35, 0)$ are identical with those obtained by the linear prediction method (LP), the UELS (UELS), and the mel-cepstral analysis method (MCEP), respectively.

From the figure, it is seen that the obtained spectra have high resolution at low frequencies when $\alpha = 0.35$. For vowels (e.g., /e/), it is clear that the resonances are represented accurately when γ is close to -1. On the other hand, for nasal cases (see portion of /N/), it is seen that a more accurate representation of the anti-resonances is provided as γ approaches 0 at the expense of increasing the bandwidths of the resonances. From these facts, we surmise that every phoneme has its own optimal values of (α, γ, M). However, the parameters (α, γ, M) are generally chosen to have constant values. Thus, in the case where the proposed method is applied to speech recognition, it is a reasonable way to choose the values of (α, γ, M) which maximize the speech recognition accuracy.

IV. PARAMETER FOR SPEECH RECOGNITION

We use the mel-generalized cepstrum as parameter for speech recognition. However, the values of (α, γ, M) might be different from those used in spectral estimation. In the following, we assume that the speech signal is analyzed with $(\alpha_1, \gamma_1, M_1)$ and then the obtained mel-generalized cepstral coefficients are transformed to those with $(\alpha_2, \gamma_2, M_2)$, which are given by

$$H(z) = s_{\gamma_2}^{-1}\left(\sum_{m=0}^{\infty} c_{\alpha_2,\gamma_2}(m)\,\Psi_{\alpha_2}^m(z)\right). \tag{11}$$

The coefficients $c_{\alpha_2,\gamma_2}(m)$ can be obtained from $c_{\alpha_1,\gamma_1}(m)$ using the mel-generalized logarithmic transformation [20], which consists of recursion formulas for the frequency transformation [21]

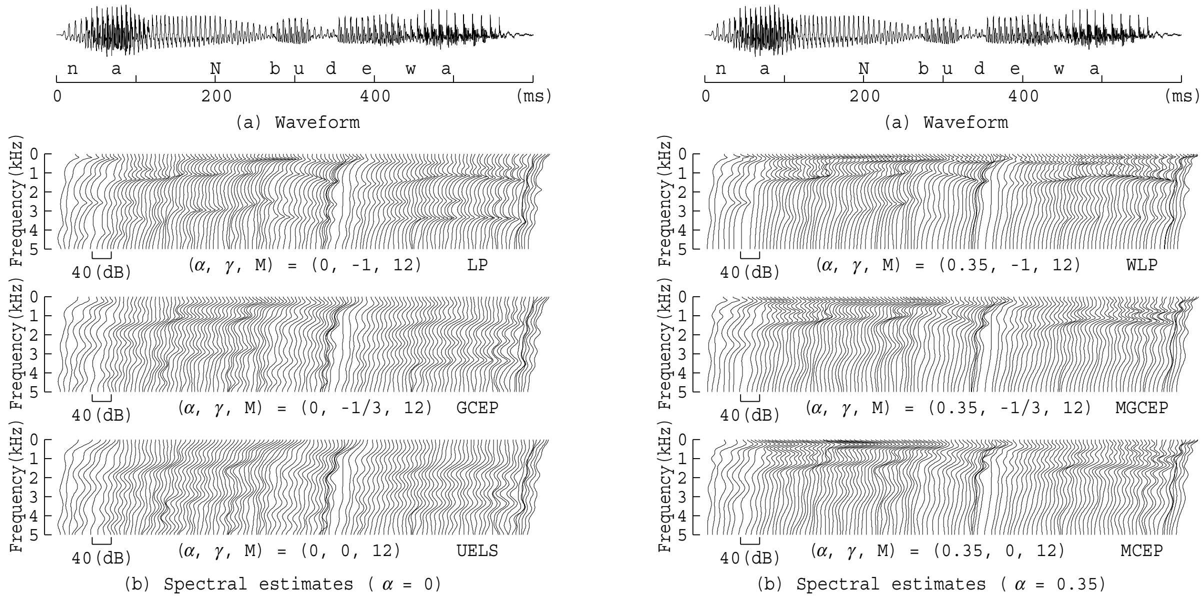


Fig. 3. Spectral estimates of natural speech for several values of $(\alpha,\ \gamma,\ M)$.

and the generalized logarithmic transformation. The algorithm is shown in Table II. Although parameters for speech recognition should be independent on input gain, the values of $c_{\alpha_2,\gamma_2}(m)$ depend on the input gain except for $\gamma_2 = 0$. To avoid this, we normalize $c_{\alpha_2,\gamma_2}(m)$ as

$$H(z) = K_{\alpha_2}\, s_{\gamma_2}^{-1} \left(\sum_{m=1}^{\infty} c'_{\alpha_2,\gamma_2}(m)\, \Psi_{\alpha_2}^{m}(z) \right) \tag{12}$$

where $c'_{\alpha_2,\gamma_2}(m)$ is the normalized mel-generalized cepstrum. The coefficients $c'_{\alpha_2,\gamma_2}(m)$ are also obtained in the recursion shown in Table II.

The LPC-cepstrum [22], which is widely used in speech recognition, is identical with the coefficients obtained by the proposed method with $(\alpha_1,\ \gamma_1) = (0, -1)$ and $(\alpha_, \gamma_2) = (0, 0)$. We can also derive several known recursion formulas as special cases of the mel-generalized logarithmic transformation [20].

V. SPEECH RECOGNITION EXPERIMENT

A preliminary evaluation of the mel-generalized analysis in word recognition was carried out using the ATR 5240 Japanese word data base. Only data from one speaker (speaker MAU) was used with just one realization of each word. Training was conducted on the even words, while testing was conducted on the odd words except 2 words including a phoneme which does not appear in the training words. We used 32 different phoneme models from the standard ATR labeling scheme, plus an additional silence model, i.e., 33 models in all. The grammar used was a simple 2618 word lexicon, mapping phoneme sequences to words. The type of HMM used was a continuous gaussian mixture density model with no explicit duration modeling. All models were 5-state, 3-mixture left to right models with no skips. The first and last states were non-emitting.

All feature vectors comprised of 13 mel-cepstral coefficients and 13 delta mel-cepstral coefficients. Both cepstral and delta cepstral coefficients included 0th coefficients. The signal was windowed by a 25.6ms Blackman window with a 5ms shift. The 12th order mel-generalized cepstral analysis was applied with $(\alpha_1,\ \gamma_1,\ M_1) = (0, -1, 12)$, $(0.35, -1/3, 12)$, $(0.35, 0, 12)$, where the analyses with $(\alpha_1,\ \gamma_1,\ M_1) = (0, -1, 12)$ and $(0.35, 0, 12)$ are equivalent to the linear prediction method and the mel-cepstral analysis method, respectively. The obtained mel-generalized cepstral coefficients were transformed into 13 mel-cepstral coefficients i.e., $(\alpha_2,\ \gamma_2,\ M_2) = (0.35, 0, 12)$.

The recognition results are shown in Fig. 4. As shown in the figure, the mel-generalized cepstral analysis with $(\alpha_1,\ \gamma_1,\ M_1) = (0.35, -1, 12)$ shows the best result. In this experiment, we tried only 3 sets of $(\alpha_1,\ \gamma_1,\ M_1)$ for spectral estimation and one set of $(\alpha_2,\ \gamma_2,\ M_2)$ for recognition parameter. More experiments will give the optimal set of parameters.

VI. DISCUSSION

Several speech spectral estimation methods which use spectral representation other than conventional AR or ARMA representation have been proposed [10], [13], [4], [9], [12], with the intention of introducing the characteristics of the human auditory sensation. Based on the proposed method, we can treat them within a framework as shown in Fig 1.

On the other hand, there exist models for simulating the human auditory system, e.g., [23], [24], [25]. Although the proposed method also takes the characteristics of human auditory sensation into consideration, it is essentially different from the models of auditory system; the proposed method is an approach which applies a cost function, i.e., (6) or (10), to a mathematically well-defined model, i.e., (5). Consequently, in the proposed method, the transfer function of the synthesis filter is clearly defined as the spectral model of (5). As a result, based on the proposed method, we can find many applications of the proposed analysis method in a similar manner of the linear prediction method; analysis/synthesis of speech [7], [26], derivation of adaptive analysis algorithms [18], [13], [14] (Fig. 1(b)), speech coding [11], [27], etc.

The spectral representation obtained as the result of the PLP analysis [10] is similar to that obtained by the mel-generalized cepstral analysis. The difference between the two methods can be summarized as follows. The PLP method first applies auditorylike modification to the power spectrum of speech and then applies the linear prediction method to the modified spectrum. As a result, the PLP method obtains a spectral representation similar to the mel-generalized cepstral representation of (5) with $(\alpha_1,\ \gamma_1,\ M_1) = (0.47, 0, 5)$ (except for the equal loudness pre-emphasis in the PLP analysis). On the other hand, the proposed method first assumes the mel-generalized cepstral model of (5) and then determines the mel-generalized cepstral coefficients in

TABLE II Mel-generalized logarithmic transformation: calculation of $c_{\alpha_2,\gamma_2}(m)$ from $c_{\alpha_1,\gamma_1}(m)$.

$$\alpha = (\alpha_2 - \alpha_1)/(1 - \alpha_1\alpha_2) \tag{13}$$

$$c^{(i)}_{\alpha_2,\gamma_1}(m) = \left\{ \begin{array}{ll} c_{\alpha_1,\gamma_1}(-i) + \alpha\, c^{(i-1)}_{\alpha_2,\gamma_1}(0), & m = 0 \\ (1-\alpha^2)\, c^{(i-1)}_{\alpha_2,\gamma_1}(0) + \alpha\, c^{(i-1)}_{\alpha_2,\gamma_1}(1), & m = 1 \\ c^{(i-1)}_{\alpha_2,\gamma_1}(m-1) + \alpha\left(c^{(i-1)}_{\alpha_2,\gamma_1}(m) - c^{(i)}_{\alpha_2,\gamma_1}(m-1)\right), & m = 2,3,\ldots,M_2 \end{array} \right\}, \quad i = -M_1,\ldots,-1,0 \tag{14}$$

$$K_{\alpha 2} = s^{-1}_{\gamma_1}\left(c^{(0)}_{\alpha_2,\gamma_1}(0)\right), \qquad c'_{\alpha_2,\gamma_1}(m) = c^{(0)}_{\alpha_2,\gamma_1}(m)/\left(1 + \gamma_1\, c^{(0)}_{\alpha_2,\gamma_1}(0)\right), \quad m = 1,2,\ldots,M_2 \tag{15}$$

$$c'_{\alpha_2,\gamma_2}(m) = c'_{\alpha_2,\gamma_1}(m) + \sum_{k=1}^{m-1} \frac{k}{m}\left(\gamma_2\, c_{\alpha_2,\gamma_1}(k)\, c'_{\alpha_2,\gamma_2}(m-k) - \gamma_1\, c_{\alpha_2,\gamma_2}(k)\, c'_{\alpha_2,\gamma_1}(m-k)\right), \quad m = 1,2,\ldots,M_2 \tag{16}$$

$$c_{\alpha_2,\gamma_2}(0) = s_{\gamma_2}(K_{\alpha_2}), \qquad c_{\alpha_2,\gamma_2}(m) = c'_{\alpha_2,\gamma_2}(m)\,(1 + \gamma_2\, c_{\alpha_2,\gamma_2}(0)), \quad m = 1,2,\ldots,M_2 \tag{17}$$

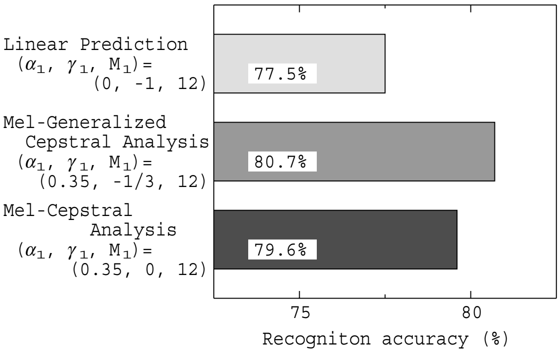


Fig. 4. Recognition accuracy for several values of (α_1, γ_1, M_1).

such a way that the criterion of (6) or (10) is minimized. Therefore, the solution of the proposed analysis is unique and optimal in the sense of the least mean square of the prediction error (Fig. 2).

The motivation of [28] is similar to that of the proposed method. However, [28] does not clearly separate spectral representation, spectral criterion, and parameterization for speech recognition.

VII. CONCLUSION

In this paper, we have described a new spectral estimation method based on the mel-generalized cepstral representation, which can be regarded as a unified approach to speech spectral estimation. We have demonstrated that since the characteristics of the obtained spectrum varies according to the values of α and γ, we can improve the performance of a speech recognition system by choosing the values of α and γ. Investigation of the optimal set of (α_1, γ_1, M_1) and (α_2, γ_2, M_2) for speech recognition is our future work.

REFERENCES

[1] F. Itakura and S. Saito, "A statistical method for estimation of speech spectral density and formant frequencies," *Trans. IECE*, vol. J53-A, pp.35–42, Jan. 1970 (in Japanese). Translation: R. W. Schafer and J. D. Markel, ed., *Speech Analysis*. New York: IEEE Press, 1979, pp.295–302.

[2] B. S. Atal and S. L. Hanauer : "Speech analysis and synthesis by linear prediction of the speech wave", in *J. Acoust. Soc. America*, vol. 50, no. 2, pp.637–655, Mar. 1971.

[3] A. V. Oppenheim and R. W. Schafer, "Homomorphic analysis of speech, " *IEEE Trans. Audio and Electroacoust.*, vol. AU-16, pp.221–226, June. 1968.

[4] 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.

[5] K. Tokuda, T. Kobayashi, R. Yamamoto and S. Imai, "Spectral estimation of speech based on generalized cepstral representation," *Trans. IEICE*, vol. J72-A, pp.457–465, Mar. 1989 (in Japanese). Translation: *Electronics and Communications in Japan (Part 3)*, vol. 73, no. 1, pp.72–81, Jan. 1990.

[6] T. Kobayashi and S. Imai, "Spectral analysis using generalized cepstrum," *IEEE Trans. Acoust., Speech, Signal processing*, vol. ASSP-32, pp.1087–1089, Oct. 1984.

[7] T. Kobayashi, S. Imai and Y. Fukuda, "Mel-generalized log spectral approximation filter," *Trans. IECE*, vol. J68-A, pp.610–611, Feb. 1985 (in Japanese).

[8] K. Tokuda, T. Kobayashi, T. Chiba and S. Imai, "Spectral estimation of speech by mel-generalized cepstral analysis," *Trans. IEICE*, vol. J75-A, pp.1124–1134, July 1992 (in Japanese). Translation: *Electronics and Communications in Japan (Part 3)*, vol. 76, no. 2, pp.30–43, July 1993.

[9] 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.

[10] H. Hermansky, "Perceptual linear predictive (PLP) analysis of speech," *J. Acoust. Soc. America*, vol. 87, pp.1738–1752, Apr. 1990.

[11] E. Krüger and H. W. Strube, "Linear prediction on a warped frequency scale," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-36, pp.1529-1531, Sep. 1988.

[12] H. W. Strube, "Linear prediction on a warped frequency scale," *J. Acoust. Soc. America*, vol. 68, no. 4, pp.1071–1076, Oct. 1980.

[13] 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.

[14] T. Fukada, K. Tokuda, T. Kobayashi and S. Imai, "A study on adaptive generalized cepstral analysis," in *Proc. 1990 IEICE Spring National Convention Record*, Mar. 1988, A-150, p.1-34 (in Japanese).

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

[16] F. Itakura and S. Saito, "On the optimum quantization of feature parameters in the PARCOR speech synthesizer," *Proc. 1972 Conf. Speech Commun. Process.*, 1972, pp.434–437.

[17] M. A. Laugunas-Hernández, A. R. Figueiras-Vidal, J. B. Mariño-Acebal and A.C. Vilanova, "A linear transform for spectral estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-29, pp.989–993, Oct. 1981.

[18] 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.

[19] S. Imai, "Cepstral analysis synthesis on the mel frequency scale," in *Proc. ICASSP-83*, 1983, pp.93–96.

[20] K. Tokuda, T. Kobayashi and S. Imai, "Recursion formula for calculation of mel generalized cepstrum coefficients," *Trans. IEICE*, vol. J71-A, pp.128–131 Jan. 1988 (in Japanese).

[21] A. V. Oppenheim and D. H. Johnson : "Discrete representation of signals", in *Proc. IEEE*, vol. 60, no. 7, pp.681-691, June 1972.

[22] B. S. Atal : "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification", in *J. Acoust. Soc. America*, vol. 55, no. 6, pp.1304–1312, June. 1974.

[23] S. Seneff, "A joint synchrony/mean-rate model of auditory speech processing," *Journal of Phonetics*, vol. 16, pp.55–76, 1988.

[24] O. Ghitza, "Temporal non-place information if the auditory-nerve firing patterns as a front-end for speech recognition in a noisy environment," *Journal of Phonetics*, vol. 16, pp.109–123, 1988.

[25] S. Kajita and F. Itakura, "Subband-autocorrelation analysis and its application for speech recognition," in *Proc. ICASSP-94*, 1994, pp.II-193–II-196.

[26] T. Chiba, K. Tokuda, T. Kobayashi and S. Imai, "Speech synthesis based on mel-generalized cepstral representation", in *Proc. 1990 IEICE Spring National Convention Record*, Mar. 1990, A-243, p.1-243 (in Japanese).

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

[28] P. Alexandre and P. Lockwood, "Root cepstral analysis: a unified view. application to speech processing in car noise environments," *Speech Communication*, vol. 12, pp.277–288, Dec. 1993.