

CELP CODING SYSTEM BASED ON MEL-GENERALIZED CEPSTRAL ANALYSIS

Kazuhito Koishida[†], Keiichi Tokuda[†], Takao Kobayashi[†] and Satoshi Imai[‡]

[†]Precision and Intelligence Laboratory, Tokyo Institute of Technology, Yokohama, 226 Japan

[‡]Dept. of Intelligence and Computer Science, Nagoya Institute of Technology, Nagoya, 466 Japan

ABSTRACT

This paper presents a CELP speech coding system based on mel-generalized cepstral analysis. In the mel-generalized cepstral analysis, we can vary the model spectrum continuously from AR to cepstral modeling by changing the value of a parameter γ and we can choose an appropriate model spectrum. Furthermore, the spectrum represented by mel-generalized cepstrum has frequency resolution similar to that of human ear. Since the perceptual weighting and postfiltering are carried out through the mel-generalized cepstrum, we expect the perceptual performance of the proposed coder to be improved. The subjective performance test indicates that the quality of the proposed CELP coder is about 2 dB higher than that of the conventional one.

1. INTRODUCTION

Code Excited Linear Prediction (CELP)[1] coding has received considerable attention for high quality speech coding at low bit rates. For improvement of quality, reduction of computational complexity, and increase of robustness to channel errors, most of the studies have been focused on excitation structure, e.g., ACELP[2], CS-CELP[3], VSELP[4], PSI-CELP[5] etc. While CELP coders can provide fairly good quality speech at around 8 kbits/s, its performance at below 4 kbits/s is yet unsatisfactory for many applications. For further improvement of speech quality, it is necessary to investigate not only excitation structure but also spectral analysis method.

In CELP coders, AR modeling has been used for short-term predictor. However, it cannot represent spectral zeros. On the other hand, the cepstral modeling can represent spectral poles and zeros with equal weights. Moreover, mel-cepstrum is defined as frequency-transformed cepstrum and the spectrum represented by the mel-cepstral coefficients has frequency resolution similar to that of human ear which has high resolution at low frequencies. From the above point of view, we have proposed speech coders based on mel-cepstral analysis and showed the effectiveness of mel-cepstral representation in speech coding[6] [7].

In this paper, we introduce the mel-generalized cepstral analysis[8] to CELP coding. In the mel-generalized cepstral analysis, we can vary the model spectrum continuously from AR to cepstral modeling by changing the value of γ and we can choose an appropriate model spectrum. Furthermore, the spectrum represented by the mel-generalized cepstrum also has high resolution at low frequencies. Consequently, we can represent speech spectrum more efficiently. In addition, the proposed CELP coder includes the following features:

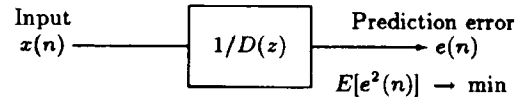


Figure 1: Interpretation of mel-generalized cepstral analysis as the least mean square of the linear prediction error.

(a) spectral representation using mel-generalized cepstrum with ensuring the filter stability after quantization; (b) perceptual weighting and postfiltering carried out through mel-generalized cepstrum; (c) bandwidth expansion in \tilde{z} plane; and (d) adaptive control method for compensating spectral tilt in postfilter. The proposed CELP coder is subjectively compared to the conventional one in the same framework and bit allocations.

2. SPECTRAL ANALYSIS

2.1. Spectral Model and Criterion

For a given order M , we assume that a speech spectrum $H(e^{j\omega})$ is modeled as follows:

$$H(z) = s_{\gamma}^{-1}(\tilde{z}^T c) \quad (1)$$

$$= \begin{cases} \left(1 + \gamma \sum_{m=0}^M c(m) \tilde{z}^{-m}\right)^{1/\gamma}, & -1 \leq \gamma < 0 \\ \exp \sum_{m=0}^M c(m) \tilde{z}^{-m}, & \gamma = 0 \end{cases} \quad (2)$$

where the coefficients $c = [c(0), \dots, c(M)]^T$ are the mel-generalized cepstrum and $\tilde{z} = [1, \tilde{z}^{-1}, \dots, \tilde{z}^{-M}]^T$. The function $s_{\gamma}^{-1}(w)$ is the inverse of the generalized logarithmic function

$$s_{\gamma}(w) = \begin{cases} (w^{\gamma} - 1)/\gamma, & 0 < |\gamma| \leq 1 \\ \log w, & \gamma = 0. \end{cases} \quad (3)$$

An all-pass system \tilde{z}^{-1} is defined by

$$\tilde{z}^{-1} = (z^{-1} - \alpha)/(1 - \alpha z^{-1}) \Big|_{z=e^{j\omega}} = e^{-j\tilde{\omega}} \quad (4)$$

where $|\alpha| < 1$. For a sampling frequency of 8 kHz, the phase characteristic $\tilde{\omega}$ of the system is a good approximation to the mel scale when $\alpha = 0.31$. From (2), it is noted that the spectral model becomes an all-pole model for $(\alpha, \gamma) = (0, -1)$ and cepstral representation for $(\alpha, \gamma) = (0, 0)$.

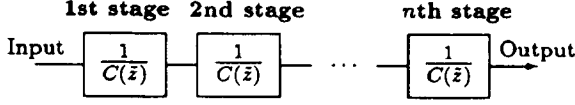


Figure 2: Synthesis filter $D(z)$ ($\gamma = -1/n$).

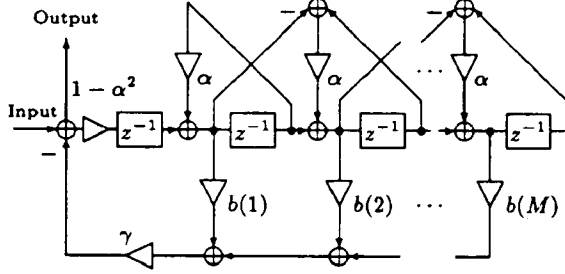


Figure 3: The structure of $1/C(\bar{z})$, avoiding delay-free loop.

To determine the coefficients c , we minimize the spectral criterion derived in the UELS[9]. Now, taking the gain factor K outside from $H(z)$, we have

$$H(z) = K \cdot D(z) = s_{\bar{r}}^{-1}(\alpha^T c) \cdot s_{\bar{r}}^{-1}(\bar{z}^T c') \quad (5)$$

where $\alpha = [1, (-\alpha), \dots, (-\alpha)^M]^T$ and the coefficients $c' = [c'(0), \dots, c'(M)]^T$ are gain-normalized version of c given by

$$c'(m) = \begin{cases} (c(0) - \alpha^T c) / (1 + \gamma \alpha^T c), & m = 0 \\ c(m) / (1 + \gamma \alpha^T c), & 1 \leq m \leq M. \end{cases} \quad (6)$$

Since the gain of $D(z)$ is constrained to be unity, the output $e(n)$ of inverse filter $1/D(z)$ becomes the prediction error [10] as shown in Fig. 1. The minimization of the criterion leads to the minimization of the residual energy. We have given a computationally efficient iterative algorithm for solving the minimization problem and shown that a few iterations are sufficient to obtain the solution[8]. In addition, the stability of the model solution $H(z)$ is guaranteed[8].

2.2. Synthesis Filter

From (2) and (5), the synthesis filter $D(z)$ is not a rational function and, therefore, it cannot be realized directly. However, using MGLSA filter[11], $D(z)$ is approximated with sufficient accuracy and becomes minimum-phase IIR system. In particular, for $\gamma = -1/n$ and n is a natural number, we have

$$D(z) = \{1/C(\bar{z})\}^n \quad (7)$$

$$C(\bar{z}) = 1 + \gamma \bar{z}^T c' = 1 + \gamma \sum_{m=0}^M c'(m) \bar{z}^{-m}. \quad (8)$$

In this case, the filter $D(z)$ is realized by the cascade of $1/C(\bar{z})$ as shown in Fig. 2 without any approximation error.

To remove a delay-free loop from $D(z)$, we modify (8) as follows:

$$C(\bar{z}) = 1 - \gamma \bar{z}^T A A^{-1} c' = 1 + \gamma \Phi^T b \quad (9)$$

$$= 1 + \gamma \sum_{m=1}^M b(m) \Phi_m(z) \quad (10)$$

where

$$A = \begin{bmatrix} 1 & \alpha & \dots & 0 \\ 0 & 1 & \dots & \vdots \\ \vdots & \vdots & \ddots & \alpha \\ 0 & \dots & 0 & 1 \end{bmatrix} \quad (11)$$

$$A^T \bar{z} = [1, \Phi^T]^T = [1, \Phi_1(z), \dots, \Phi_M(z)]^T \quad (12)$$

$$\Phi_m(z) = \frac{(1 - \alpha^2) z^{-1}}{1 - \alpha z^{-1}} \bar{z}^{-(m-1)}. \quad (13)$$

The coefficients b are given by

$$A^{-1} c' = [0, b^T]^T = [0, b(1), b(2), \dots, b(M)]^T. \quad (14)$$

where A^{-1} is the inverse matrix of A . Instead of matrix multiplication $A^{-1} c'$, a recursive algorithm can be used:

$$b(m) = \begin{cases} c'(M), & m = M \\ c'(m) - \alpha b(m+1), & 0 \leq m < M. \end{cases} \quad (15)$$

From (6), the first component of $A^{-1} c'$ is equal to zero, i.e., $\alpha^T c' = 0$. This indicates that the gain of $D(z)$ is unity. The structure of $1/C(\bar{z})$ based on (10) is shown in Fig. 3.

2.3. Stability Problem

For $\gamma = 0$, i.e., cepstral model, the minimum-phase property of $D(z)$ is always preserved for any c' [8]. However, for $-1 \leq \gamma < 0$, it does not hold and, therefore, the filter may become unstable after quantization of c' .

Assuming that $F(z) = F_1(z) + F_2(z)$ is a real polynomial and $F_1(z)$ and $F_2(z)$ are symmetric and antisymmetric polynomials, respectively. For $F(z)$ to be a stable polynomial, it is necessary and sufficient that[12] [13]: (a) The zeros of $F_1(z)$ and $F_2(z)$ are located on the unit circle; (b) They are simple; and (c) They separate each other. If $F(z)$ is the LPC polynomial, the angular positions of the zeros of $F_1(z)$ and $F_2(z)$ correspond to the LSP coefficients.

Since the all-pass system \bar{z}^{-1} maps the inside of the unit circle of the \bar{z} plane to the inside of the unit circle of the z plane if and only if $|\alpha| < 1$, the stability problem with respect to z plane is equivalent to that with respect to \bar{z} plane. Consequently, when the polynomial $C(\bar{z})$ is decomposed into symmetrical and antisymmetrical polynomials, they have the roots on the unit circle of \bar{z} plane. The roots of interest are $e^{j\bar{\omega}_i}$ for $i = 0, \dots, M+1$. These parameters $\{\bar{\omega}_i\}_{i=1, \dots, M}$ are interpreted as the angular positions of the roots on warped frequency scale. (It is noted that $\bar{\omega}_0 = 0$ and $\bar{\omega}_{M+1} = \pi$ are fixed roots.) The minimum-phase property of $1/C(\bar{z})$ is ensured after quantization if the parameters $\bar{\omega}_i$ satisfy the above three conditions.

Fig. 4 shows the spectra estimated by mel-generalized cepstral analysis and fluctuation of the proposed spectral parameters for $(\alpha, \gamma, M) = (0.31, -1/3, 10)$. The LPC spectra with tenth order and fluctuation of the LSP parameters are also shown. Compared with LPC spectra, it is seen that the obtained spectra has high resolution at low frequencies. Although the distribution of each parameter is more limited than that of each LSP parameter, there exists a tradeoff between distribution range and spectral sensitivity. It is shown that the quantization and interpolation performance of the proposed parameters is slightly better than that of the LSP parameters[14].

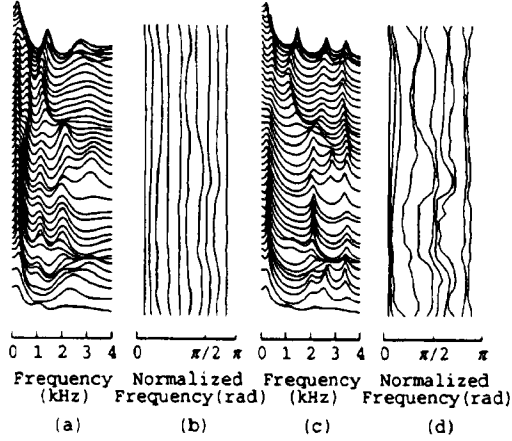


Figure 4: Spectral estimates and fluctuation of the spectral parameters. (a) mel-generalized cepstral analysis. (b) proposed parameter. (c) LPC analysis. (d) LSP.

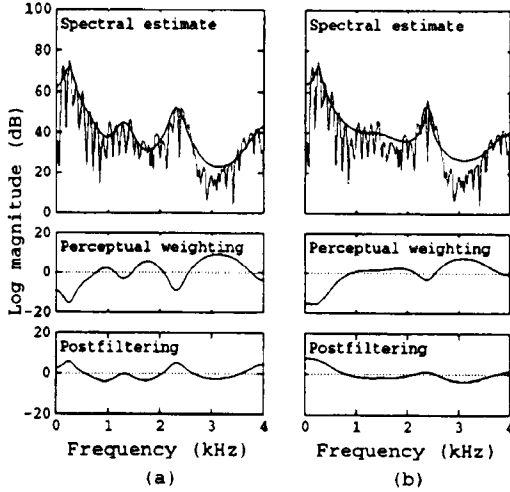


Figure 5: Example of perceptual weighting and postfiltering. (a) proposed coder. (b) conventional coder.

3. CODER STRUCTURE

In this section, we will describe the structure of the proposed CELP coder. Here, we let $(\alpha, \gamma, M) = (0.31, -1/3, 10)$ for spectral analysis. In this case, the synthesis filter is realized by three-stage cascade structure.

3.1. Perceptual Weighting Filter

The perceptual weighting filter $H_w(z)$ is defined by

$$H_w(z) = C(\bar{z})C(\bar{z}) / C(\bar{z}/\beta_1) \quad (16)$$

where \bar{z}/β represents bandwidth expansion in \bar{z} plane. Since this bandwidth expansion operation replaces the coefficients $c'(m)$ by $\beta^m c'(m)$, the filter after bandwidth expansion can be written in the form

$$C'(\bar{z}/\beta) = 1 + \gamma \bar{z}^T B c' \quad (17)$$

where matrix B is diagonal

$$B = \text{diag} [1, \beta, \dots, \beta^M]. \quad (18)$$

We modify (17) in a similar manner to (9):

$$C'(\bar{z}/\beta) = 1 + \gamma \bar{z}^T A A^{-1} B c' \quad (19)$$

$$= 1 + \gamma [1, \Phi^T] A^{-1} B c' \quad (20)$$

$$= 1 + \gamma [1, \Phi^T] b'_\beta \quad (21)$$

Again, the recursive algorithm can be used instead of the matrix multiplication $A^{-1} B c'$:

$$b'_\beta(m) = \begin{cases} \beta^M c'(M), & m = M \\ \beta^m c'(m) - \alpha b'_\beta(m+1), & 0 \leq m < M. \end{cases} \quad (22)$$

The gain of $C'(\bar{z}/\beta)$ is not equal to zero, i.e., $b'_\beta(0) \neq 0$, due to bandwidth expansion in \bar{z} plane. Taking the gain K' outside from $C'(\bar{z})$, we have

$$C'(\bar{z}/\beta) = K' \cdot C(\bar{z}/\beta) = (1 + \gamma b'_\beta(0))(1 + \gamma \Phi^T b_\beta). \quad (23)$$

The filter $C(\bar{z}/\beta)$ is the gain-normalized version of $C'(\bar{z}/\beta)$ and its coefficients $b_\beta = [b_\beta(1), \dots, b_\beta(M)]$ are given by

$$b_\beta(m) = b'_\beta(m) / (1 + \gamma b'_\beta(0)), \quad 1 \leq m \leq M. \quad (24)$$

From (23), replacing the coefficients b by b_β in Fig. 3, we can realize the filter $1/C(\bar{z}/\beta)$ in the same manner as $1/C(\bar{z})$.

3.2. Postfilter

The postfilter $H_p(z)$ has the transfer function

$$H_p(z) = \{C(\bar{z}/\beta_2) / C(\bar{z})\} (1 - \mu z^{-1})^p. \quad (25)$$

The postfilter consists of two parts: formant postfilter $C(\bar{z}/\beta_2)/C(\bar{z})$ and spectral tilt compensation filter $(1 - \mu z^{-1})^p$. The parameter μ controls the global spectral tilt in postfilter. For example, in conventional postfilter, the parameter μ was chosen to be 0.5, or $0.15k_1$ where k_1 is the first reflection coefficient and $p = 1$. Here, we propose a new adaptive control method for compensating global spectral tilt. This method is realized by setting the first mel-cepstral coefficient of $H_p(z)$ to be zero. The first mel-cepstral coefficient is calculated using the recursive formulas[15]. Under such constraint, μ is given by

$$\mu = \frac{-\gamma(1 - \beta_2)c'(1)}{-\alpha\gamma(1 - \beta_2) + p(1 - \alpha^2)}. \quad (26)$$

To restrict absolute of μ to within unity, the tilt compensation filter has p -stage cascade structure.

3.3. Frequency Responses of Perceptual Weighting and Postfiltering

Fig. 5 shows the frequency responses of perceptual weighting filter and postfilter in proposed coder whose tunable parameters are $\beta_1 = 0.7, \beta_2 = 0.0, p = 2$. For comparison, those of conventional coder are also shown in Fig. 5. Their transfer functions are defined by

$$H_w(z) = A(z/0.9) / A(z/0.4) \quad (27)$$

$$H_p(z) = \{A(z/0.5) / A(z/0.8)\} (1 - 0.15k_1 z^{-1}) \quad (28)$$

where $A(z)$ is LPC polynomial. From these figures, the spectral envelope obtained by mel-generalized cepstral analysis has high resolution at low frequencies; accordingly the spectra of perceptual weighting and postfiltering retain fine structure at low frequencies.

