Recursive Calculation of Mel-Cepstrum from LP Coefficients

Keiichi Tokuda†, Takao Kobayashi††, and Satoshi Imai††,

†Department of Computer Science
Nagoya Institute of Technology, Nagoya, 466-8555 Japan
††Precision and Intelligence Laboratory
Tokyo Institute of Technology, Yokohama, 227 Japan

1 April 1994

Abstract

The mel-cepstral coefficients are often calculated from the linear prediction coefficients by using recursion formulas. However, the obtained mel-cepstral coefficients have errors caused by truncation in the quefrency domain. The purpose of this report is to point out that the mel-cepstral coefficients can be calculated from the LP (Linear Prediction) coefficients using the recursion formulas without the truncation error.

1 INTRODUCTION

The mel-cepstrum is useful parameter for speech recognition, and have widely been used in many speech recognition systems [1]. There exist several methods to obtain mel-cepstral coefficients. In the case using recursion formulas, the mel-cepstral coefficients, i.e., frequency-transformed cepstral coefficients, are calculated from the LP (Linear Prediction) coefficients as follows: the LP coefficients are first transformed to the cepstral coefficients by using the recursion formula of [2], then the cepstral coefficients transformed to the mel-cepstral coefficients by using the recursion formula of [3]. However, since the cepstrum calculated from the LP coefficients is an infinite sequence, it must be truncated in the quefrency domain. As a result, we cannot obtain exact values of mel-cepstral coefficients.

This report points out that the mel-cepstral coefficients can be calculated from the LP coefficients without the truncation error when we first apply the recursion formula for frequency-transformation, and then apply the recursion formula of LPC-cepstrum.

2 Definition of The Mel-Cepstrum

The cepstrum[1] $c(m)$ of a real sequence $x(n)$ is defined as

$$c(m) = \frac{1}{2\pi j} \oint_C \log X(z)\, z^{m-1}\, dz \tag{1}$$

[1]Strictly, the complex cepstrum defined by A. V. Oppenheim and R. W. Schafer [4].

$$\log X(z) = \sum_{m=-\infty}^{\infty} c(m)\, z^{-m} \tag{2}$$

where $X(z)$ is the z-transform of $x(n)$, and C is a counterclockwise closed contour in the region of convergence of $\log X(z)$ and encircling the origin of the z-plane. Frequency-transformed cepstrum, so-called mel-cepstrum [5], is defined as

$$\tilde{c}(m) = \frac{1}{2\pi j} \oint_C \log X(z)\, \tilde{z}^{m-1}\, d\tilde{z} \tag{3}$$

$$\log X(z) = \sum_{m=-\infty}^{\infty} \tilde{c}(m)\, \tilde{z}^{-m} \tag{4}$$

where

$$\tilde{z}^{-1} = \Psi(z) = \frac{z^{-1} - \alpha}{1 - \alpha\, z^{-1}}, \qquad |\,\alpha\,| < 1. \tag{5}$$

The phase response $\tilde{\omega}$ of the all-pass system $\Psi(e^{j\omega}) = e^{-j\tilde{\omega}}$ is given by

$$\tilde{\omega} = \beta(\omega) = \tan^{-1} \frac{(1-\alpha^2)\sin\omega}{(1+\alpha^2)\cos\omega - 2\alpha}. \tag{6}$$

Thus, evaluating (3) and (4) on the unit circle of the $\tilde{z}$-plane, we see that $\tilde{c}(m)$ is the inverse Fourier transform of $\log X(e^{j\omega})$ calculated on a warped frequency scale $\tilde{\omega}$:

$$\tilde{c}(m) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \log \tilde{X}(e^{j\tilde{\omega}})\, e^{j\tilde{\omega}m}\, d\tilde{\omega} \tag{7}$$

$$\log \tilde{X}(e^{j\tilde{\omega}}) = \sum_{m=-\infty}^{\infty} \tilde{c}(m)\, e^{-j\tilde{\omega}m} \tag{8}$$

where

$$\tilde{X}(e^{j\tilde{\omega}}) = X(e^{j\beta^{-1}(\tilde{\omega})}). \tag{9}$$

The phase response $\tilde{\omega} = \beta(\omega)$ gives a good approximation to auditory frequency scales with an appropriate choice of α. An example of 16kHz sampling is shown in Fig. 1. In the figure, dashed lines show the mel [6] and Bark[2] frequency scales normalized by 8kHz. Table 1 shows examples of α for approximating the mel and Bark scales at several sampling frequencies.

The system function obtained by the LP method has the form

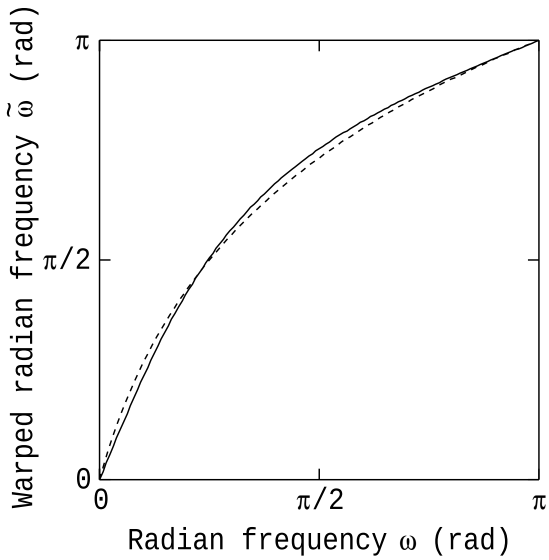
$$H(z) = \frac{K}{1 + \displaystyle\sum_{m=1}^{M} a(m)\, z^{-m}} \tag{10}$$

where we assume that $H(z)$ is a minimum phase system[3] . The problem is to calculate the mel-cepstral coefficients $\tilde{c}(m)$, $m = 0, 1, \ldots, N$ given by

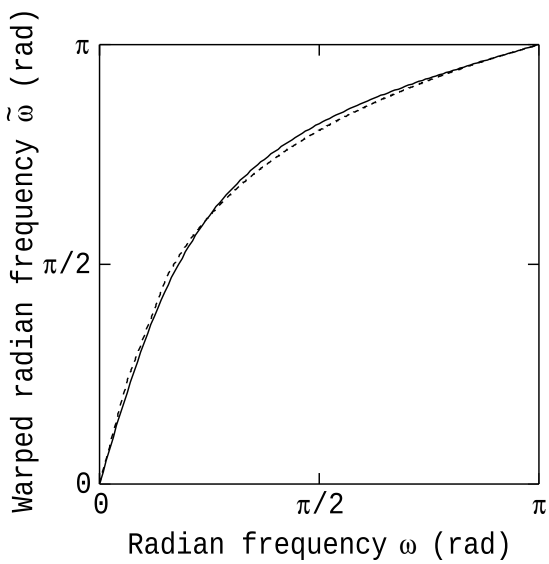
$$\log \frac{K}{1 + \displaystyle\sum_{m=1}^{M} a(m)\, z^{-m}} = \sum_{m=0}^{\infty} \tilde{c}(m) \tilde{z}^{-m} \tag{11}$$

[2]A piecewise function shown in [7] is used for plotting the Bark scale.

[3]The system $H(z)$ obtained by the autocorrelation method is always minimum phase.



(a) mel scale ($\alpha = 0.42$)



(b) Bark scale ($\alpha = 0.55$)

Figure 1: Phase responses $\tilde{\omega} = \beta(\omega)$ of $\Psi(e^{j\omega}) = e^{-j\tilde{\omega}}$ (real lines) and auditory frequency scales (dashed lines). (16kHz of sampling frequency)

Table 1: Examples of α.

Sampling Frequency	8kHz	10kHz	12kHz	16kHz	20kHz	22.05kHz
mel scale	0.31	0.35	0.37	0.42	0.44	0.45
Bark scale	0.42	0.47	0.50	0.55	—	—

from the LP coefficients $a(m)$, $m = 1, 2, \ldots, M$ and K. It is noted that when the system $H(z)$ is of minimum phase, the mel-cepstrum becomes a causal and stable sequence.

In the conventional recursive calculation method, the mel-cepstral coefficients are calculated as follows:

1. The cepstral coefficients $c(m)$'s given by

$$\log \frac{K}{1 + \sum_{m=1}^{M} a(m)\, z^{-m}} = \sum_{m=0}^{\infty} c(m)\, z^{-m} \tag{12}$$

   are calculated by using the recursion formula of [2].

2. The mel-cepstral coefficients $\tilde{c}(m)$, $m = 1, 2, \ldots, N$ given by

$$\sum_{m=0}^{\infty} c(m)\, z^{-m} = \sum_{m=0}^{\infty} \tilde{c}(m) \tilde{z}^{-m} \tag{13}$$

   are calculated by using the recursion formula for frequency transformation [3].

Unfortunately, the cepstrum obtained by (12) is an infinite sequence and we require the infinite sequence to calculate finite number of the mel-cepstral coefficients given by (13). Practically, the mel-cepstrum is calculated from truncated cepstrum by an approximation:

$$\sum_{m=0}^{L} c(m)\, z^{-m} = \sum_{m=0}^{\infty} \tilde{c}(m) \tilde{z}^{-m}. \tag{14}$$

It is noted that to reduce the truncation error sufficiently, L should be chosen in such a way that $L >> M$. The complexity to calculate $\tilde{c}(m)$, $m = 0, 1, \ldots, N$ from $a(m)$, $m = 1, 2, \ldots, M$ and K is $O(ML) + O(LN)$.

3 Recursive Calculation of The Mel-Cepstrum

The following procedure can avoid the trancation error:

1. The frequency transformed LP coefficients are first calculated based on the relation:

$$\frac{K}{1 + \sum_{m=1}^{M} a(m)\, z^{-m}} = \frac{\tilde{K}}{1 + \sum_{m=1}^{\infty} \tilde{a}(k) \tilde{z}^{-m}} \tag{15}$$

   by the recursion formula for frequency transformation [3].

2. The mel-cepstrum given by

$$\log \frac{\tilde{K}}{1+\sum_{m=1}^{\infty}\tilde{a}(k)\tilde{z}^{-m}} = \sum_{m=0}^{\infty}\tilde{c}(m)\tilde{z}^{-m} \tag{16}$$

are calculated by the recursion formula of [2].

Although $\tilde{a}(m)$ is an infinite sequence, the mel-cepstral coefficients $\tilde{c}(m)$, $m = 0, 1, \ldots, N$ given by (16) can be calculated from the finite sequence $\tilde{a}(m)$, $m = 1, 2, \ldots, N$ and $\tilde{K}$. Therefore, it is not necessary to calculate infinite number of coefficients $\tilde{a}(m)$, $m = 1, 2, \ldots, \infty$.

The recursion formulas for calculation of mel-cepstrum from LP coefficients are written as follows:

$$\tilde{a}^{(i)}(m) = \begin{cases} a(-i) + \alpha\,\tilde{a}^{(i-1)}(0), & m = 0 \\ (1-\alpha^2)\,\tilde{a}^{(i-1)}(0) + \alpha\,\tilde{a}^{(i-1)}(1), & m = 1 \\ \tilde{a}^{(i-1)}(m-1) + \alpha\,(\tilde{a}^{(i-1)}(m) - \tilde{a}^{(i)}(m-1)), & m = 2, 3, \ldots, N \end{cases}$$
$$i = -M, \ldots, -1, 0 \tag{17}$$

$$\tilde{K} = K/\tilde{a}^{(0)}(0), \qquad \tilde{a}(m) = \tilde{a}^{(0)}(m)/\tilde{a}^{(0)}(0), \quad 1 \le m \le N \tag{18}$$

$$\tilde{c}(m) = \begin{cases} \log \tilde{K}, & m = 0 \\ -\tilde{a}(m) - \sum_{k=1}^{m-1}\frac{k}{m}\tilde{c}(k)\,\tilde{a}(m-k), & 1 \le m \le N \end{cases} \tag{19}$$

where $a(0) = 1$.

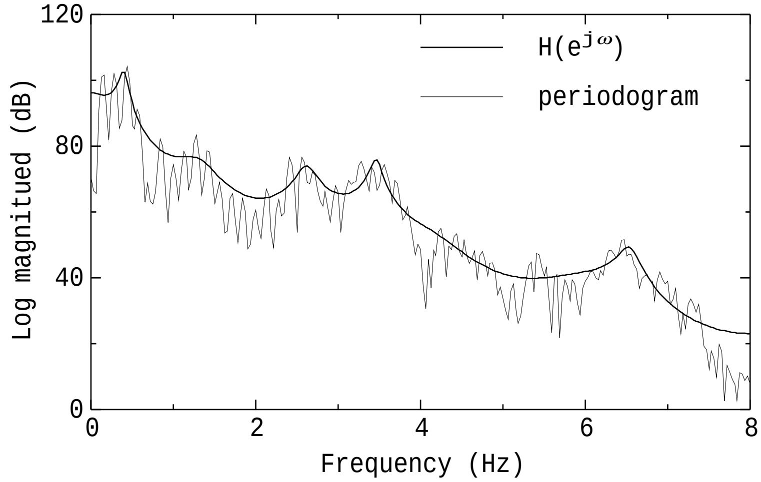
The above algorithm calculates $\tilde{c}(m)$, $m = 1, 2, \ldots, N$ from $a(m)$, $m = 1, 2, \ldots, M$ and K with the complexity of $O(MN)+O(N^2)$. Note that the above algorithm is a special case of the recursion formula of [8], [9].

4 Example

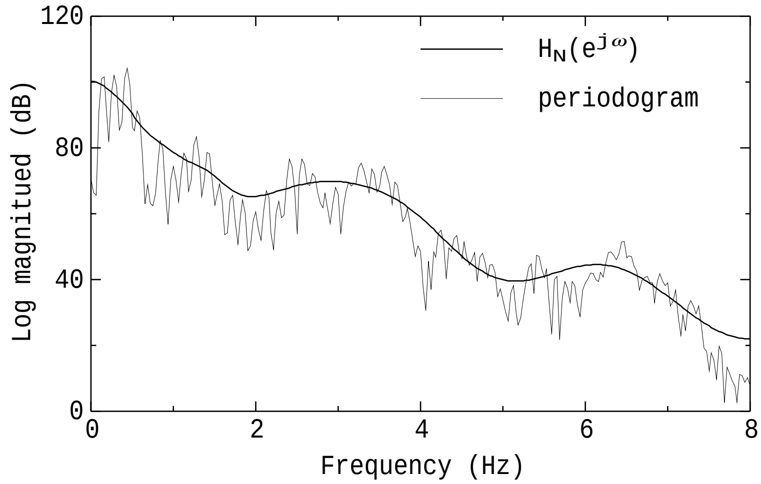
Fig. 2(a) shows an example of the spectrum obtained by the LP method. Fig. 2(b) and (c) show the spectra represented by the mel-cepstra which obtained the conventional and proposed methods, respectively. The spectrum represented by the mel-cepstral coefficients $\tilde{c}(m)$, $m = 0, 1, \ldots, N$ is given by

$$H_N(e^{j\omega}) = \exp \sum_{m=0}^{N}\tilde{c}(m)\tilde{z}^{-m}. \tag{20}$$

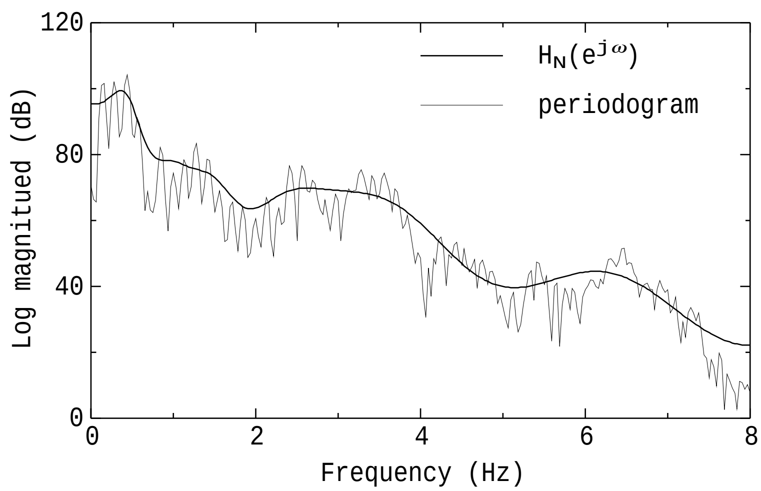
From the figures, it is seen that the spectrum obtained by the conventional method loses the feature at the low frequencies because of the truncation in the quefrency domain. On the other hand, the proposed method is independent from the truncation in the quefrency domain. In this simulation, $L = 15$ $(= N = M)$ is used for the conventional calculation method. Although the truncation error can be reduced sufficiently with L greater than about $2M$, we cannot obtain exact values of mel-cepstral coefficients unless $L \to \infty$.



(a) Original spectrum ($M = 15$)



(b) Conventional method ($L = N = 15$)



(c) Proposed method ($N = 15$)

Figure 2: Spectra represented by the mel-cepstra which are calculated from the LP coefficients by using the conventional and proposed methods. ($\alpha = 0.42$)

5 CONCLUSIONS

In this report, we have shown a recursive calculation method of mel-cepstral coefficients from LP coefficients. Our approach first applies the frequency transformation and then applies the recursion formula of [2]. Using the recursion formulas, we can obtain the mel-cepstral coefficients without the error caused by the truncation in the quefrency domain.

In the LP method, the LP coefficients are determined to minimize the mean square of linear prediction error. However, N-th order mel-cepstral coefficients, which are obtained from the LP coefficients by the recursion formulas, do not minimize the mean square of linear prediction error. To avoid this problem, we have already proposed a mel-cepstral analysis method [10], [11], in which we can extract mel-cepstral coefficients which minimizes the mean square of linear prediction error directly.

References

[1] J. W. Picone, “Signal modeling techniques in speech recognition,” in *Proc. IEEE*, vol. 81 no. 9, pp.1215–1247, Sep. 1993.

[2] 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.

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

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

[5] S. Imai, K. Sumita, C. Furuichi, “Mel log spectral approximation filter for speech synthesis”, *Trans. IECE*, vol. J66-A, pp.122–129, Feb. 1983 (in Japanese).

[6] G. Fant, *Speech sound and features.* Cambridge: MIT Press, 1973.

[7] S. Seneff, “A computational model for the peripheral auditory system: application to speech recognition research,” in *Proc. ICASSP’86*, 1986, pp.1983–1986.

[8] 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).

[9] K. Tokuda, T. Masuko, T. Kobayashi and S. Imai, “Mel-generalized cepstral analysis —a unified approach to speech spectral estimation,” in *Proc. ICSLP-94*, 1994, pp.1043–1046.

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

[11] K. Tokuda, T. Kobayashi, T. Fukada, H. Saito and S. Imai, “Spectral estimation of speech based on mel-cepstral representation,” *Trans. IEICE*, vol. J74-A, pp.1240–1248, Aug. 1991 (in Japanese).