Recursive Calculation of Mel-Cepstrum from LP Coefficients

Gheorghe RADU, Gheorghe ANTON

***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.**

***Index Terms*—Estimation and detection, Pattern recognition, Signal processing, Speech recognition.**

***2000 Mathematics Subject Classification*: 94A12, 93E10, 68T10.**

I. 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 frequency 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.

II. Definition of the Mel-Cepstrum

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

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

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

where $X(z)$ is the z-transform of $x(n)$ and C is a counter clock wise 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, is defined as

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

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

where

$$\tilde{z}^{-1}=\psi_\alpha(z)=\left.\frac{z^{-1}-\alpha}{1-\alpha z^{-1}}\right|_{z=e^{j\omega}}=e^{-j\beta_\alpha(\omega)};\ \ |\alpha|<1 \tag{5}$$

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

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

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

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

The phase response $\tilde{\omega}=\beta(\omega)$ gives a good approximation to auditory frequency scales with an appropriate choice of α. An example of 16 kHz sampling is shown in Fig. 1.

In the figure, dashed lines show the Mel [4] and Bark frequency scales normalized by 8 kHz. Table1 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

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

where we assume that $H(z)$ is a minimum phase system.

The problem is to calculate the Mel-cepstral coefficients $\tilde{c}(m)$, $m = 0, 1, \ldots, N$ given by

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

from the LP coefficients a(m), m = 1, 2, . . .,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.

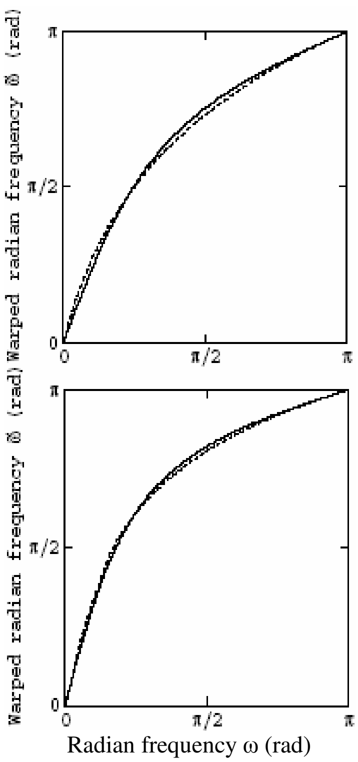

Fig. 1. Phase responses $\tilde{\omega} = \beta(\omega)$ of $\psi(e^{j\tilde{\omega}}) = e^{-j\tilde{\omega}}$ (real lines) and auditory frequency scales (dashed lines). (16 kHz of sampling frequency)

Table 1: Examples of α.

Sampling Frequency	8kHz	10kHz	12 kHz	16kHz	20 kHz	22 kHz
Mel scale	0.31	0.35	0.37	0.42	0.44	0.44
Bark scale	0.42	0.47	0.50	0.55	--	--

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

1. The cepstral coefficients c(m) is given by

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

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

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

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 $O(ML) + O(LN)$.

III. RECURSIVE CALCULATION OF THE MEL-CEPSTRUM

The following procedure can avoid the truncation 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:

$$\tilde{a}_m^{(i)} = \begin{cases} a_{-i} + \alpha \tilde{a}_0^{(i-1)}, & m = 0 \\ (1-\alpha^2) \tilde{a}_0^{(i-1)} + \alpha \tilde{a}_1^{(i-1)}, & m = 1 \\ \tilde{a}_{m-1}^{(i-1)} + \alpha (\tilde{a}_m^{(i-1)} - \tilde{a}_{m-1}^{(i-1)}), & m = 2,3,\ldots \end{cases} \tag{17}$$

$$i = -M, -1, \ldots, 0$$

$$\tilde{K} = \frac{K}{\tilde{a}_0^{(0)}}; \quad \tilde{a}_m = \frac{\tilde{a}_m^{(0)}}{\tilde{a}_0^{(0)}}, \quad m = 1, \ldots, N; \tag{18}$$

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

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 [5, 6].

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] G. Fant, Speech sound and features, Cambridge, MIT Press, 1973.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

Gheorghe RADU is with "Henri Coanda" Air Force Academy, #160, M.Viteazul Street, 500183 Brasov, gh.radu@gmail.com.
Gheorghe ANTON is with Trades and Services High School, Buzau, #14 bis, Bazalt Street, g10.anton@gmail.com.