EFFICIENT ENCODING OF MEL-GENERALIZED CEPSTRUM FOR CELP CODERS

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

Precision and Intelligence Laboratory, Tokyo Institute of Technology, Yokohama, 226 Japan
†Department of Computer Science, Nagoya Institute of Technology, Nagoya, 466 Japan

ABSTRACT

In this paper, the performance of several algorithms for the quantization of the mel-generalized cepstral coefficients is studied. First, the objective and subjective performance of two-stage vector quantization (VQ) is measured. It is shown that subjective quality for the mel-generalized cepstral coefficients is higher than that for LSP. Secondly, interframe prediction is introduced in the encoding of mel-generalized cepstral coefficients. By utilizing interframe moving average (MA) prediction, the mel-generalized cepstral coefficients can be encoded more efficiently than LSP in terms of cepstral distortion. Finally, we implement a CELP coder based on mel-generalized cepstral analysis in which mel-generalized cepstral coefficients are quantized using MA prediction. This coder has higher objective quality than conventional CELP.

1. INTRODUCTION

Many spectral estimation methods have been proposed for various speech applications. Among these, the mel-generalized cepstral analysis [1],[2] is one of the most effective method. In the method, the model spectrum based on the mel-generalized cepstral representation can be varied continuously from all-pole to cepstral modeling by changing the value of a parameter. Furthermore, the spectrum represented by mel-generalized cepstral coefficients has frequency resolution similar to that of human ear. From the above point of view, we have proposed a CELP coder based on mel-generalized cepstral analysis [1] which achieves an improvement over the conventional CELP [3].

For low bit rate speech coding, it is important to quantize the spectral envelope information using as few bits as possible. Several studies have been done to quantize the LSP parameters efficiently. Since LSP parameters have a strong correlation between frames, the quantization distortion can be reduced by using interframe prediction [4],[5] or discrete cosine transform [6].

In this paper, the performance of several algorithms for the quantization of the mel-generalized cepstral coefficients is studied. First, objective and subjective performance of two-stage VQ is measured. Secondly, we attempt to utilize the correlation between frames to encode the mel-generalized cepstral coefficients. The quantization performance is measured and compared with that for LSP in terms of cepstral distortion measure. Finally, we implement a CELP coder based on mel-generalized cepstral analysis, in which the spectral parameters are quantized using interframe MA prediction, and make a comparison between objective performance of the conventional and proposed CELP.

2. SPECTRAL ESTIMATION

2.1. Mel-Generalized Cepstral Analysis [1],[2]

We assume that a speech spectrum $H(e^{j\omega})$ can be modeled as follows:

$$H(z) = K \cdot D(z) \tag{1}$$

where K is the gain of $H(z)$ and the filter $D(z)$ whose gain is constrained to be unity is defined by

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

The coefficients $c(m)$ are mel-generalized cepstrum and $\tilde{z}^{-1}$ is an all-pass system defined by

$$\tilde{z}^{-1} = \left.\frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}\right|_{z=e^{j\omega}} = e^{-j\tilde{\omega}}, \qquad |\alpha| < 1. \tag{3}$$

For a sampling frequency of 8 kHz, the phase characteristics $\tilde{\omega}$ of the system gives a good approximation to the mel scale when $\alpha = 0.31$. It should be noted that the spectral model of (2) becomes an all-pole model for $(\alpha, \gamma) = (0, -1)$ and cepstral representation for $(\alpha, \gamma) = (0, 0)$.

To find an optimum set of coefficients $c(m)$, we minimize the spectral criterion derived in the UELS [7]. It is shown that the minimization of the criterion leads to the minimization of the residual energy [8]. We can solve the minimization problem using efficient iterative algorithm based on FFT and recursive formulas [2]. In addition, the stability of the model solution $H(z)$ is guaranteed [2].

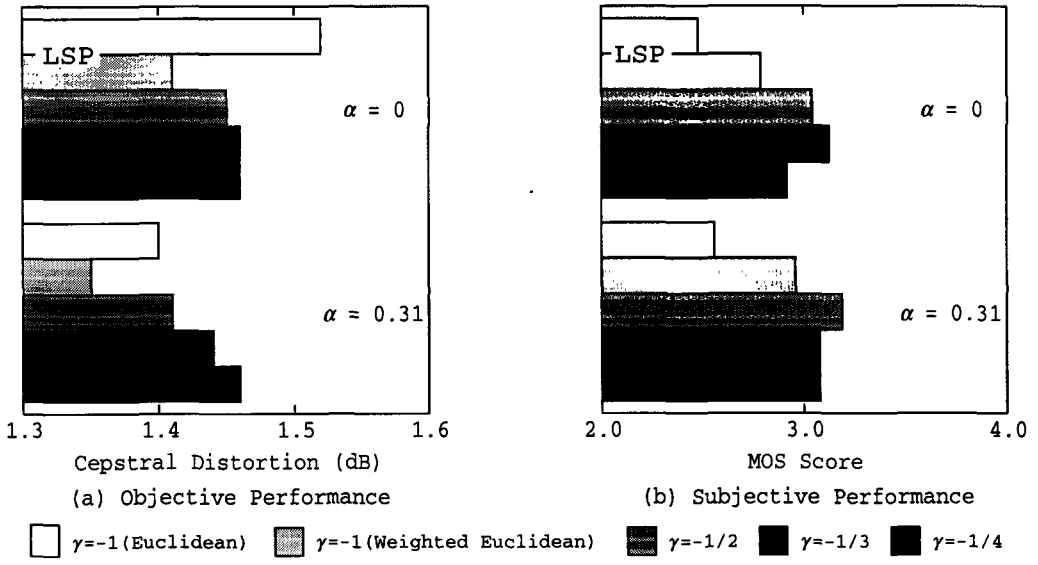

Fig. 1. Encoding performance of two-stage VQ.

Table 1. Analysis conditions.

Sampling Frequency	8 kHz
Order of Analysis	10
Window	32ms Hamming
Frame Period	10 ms

2.2. Spectral Parameters for Quantization

When $\gamma = 0$, i.e., the cepstral representation, minimum-phase property is preserved for any $c(m)$. Hereinafter, therefore, we discuss the stability after quantization for $-1 \leq \gamma < 0$.

When $-1 \leq \gamma < 0$, the synthesis filter of (2) is expressed as

$$D(z) = \left\{ \frac{1}{C(\tilde{z})} \right\}^{-1/\gamma} \tag{4}$$

where

$$C(\tilde{z}) = 1 + \gamma \sum_{m=0}^{M} c(m) \tilde{z}^{-m}. \tag{5}$$

Direct quantization of $c(m)$ may cause unstable synthesis filter. To avoid this problem, we decompose the polynomial $C(\tilde{z})$ into symmetrical and antisymmetrical polynomials, i.e., $C(\tilde{z}) = C_1(\tilde{z}) + C_2(\tilde{z})$. The polynomials $C_1(\tilde{z})$ and $C_2(\tilde{z})$ have the following properties: (a) all roots of $C_1(\tilde{z})$ and $C_2(\tilde{z})$ are located on the unit circle in $\tilde{z}$-plane, and (b) roots of $C_1(\tilde{z})$ and $C_2(\tilde{z})$ are interlaced with each other. Since the unit circle of $\tilde{z}$-plane maps onto the unit circle of the z-plane, the polynomial $C(\tilde{z})$ has the minimum-phase property if the roots of $C_1(\tilde{z})$ and $C_2(\tilde{z})$ satisfy these two properties. Thus the stability of $C(\tilde{z})$ can be ensured after quantization by representing the spectral information as the roots of $C_1(\tilde{z})$ and $C_2(\tilde{z})$. In [9], some properties of the roots are shown in detail.

In the following, we use $\tilde{\omega}_n = [\tilde{\omega}_{n,1}, \tilde{\omega}_{n,2}, \cdots, \tilde{\omega}_{n,M}]$ to denote the roots of $C_1(\tilde{z})$ and $C_2(\tilde{z})$, associated with the n-th frame of speech.

3. QUANTIZATION PERFORMANCE

3.1. Experimental Conditions

A speech database of 20 females and 20 males is used for training. The analysis conditions are shown in Table 1. In the experiments, two-stage VQ [10] (12 bits are allocated to each stage) is used and designed by LBG algorithm [11]. The proposed parameters are quantized with Euclidean distance measure. For comparison, LSP parameters are quantized with the weighted Euclidean distance measures [12].

3.2. Objective Performance

For evaluating the quantization performance objectively, we use the cepstral distortion measure with a order of 128. Fig. 1(a) shows the results for 16 sentences (spoken by 8 males and 8 females speakers). In the case of Euclidean distance measure, the proposed parameters have smaller distortion than LSP. Compared to LSP with weighted Euclidean distance measure, the distortion performance is much the same or slightly worse.

3.3. Subjective Performance

Subjective quality evaluation is done through a mean opinion score (MOS) test for 6 listeners and 6 sentences (spoken by 3 males and 3 females speakers). We use here a analysis-synthesis framework to generate synthesized speech. In the test, only spectral parameters are quantized and the other parameters such as pitch and gain are not quantized. The spectral parameters are interpolated sample by sample. The result

Table 2. Auto-covariance coefficients $\psi_i(j)$ of the proposed parameters ($\alpha = 0.31, \gamma = -1/3$).

	j=1	2	3	4	5	6	7
i=1	0.97	0.92	0.88	0.83	0.78	0.74	0.70
2	0.89	0.74	0.62	0.52	0.44	0.37	0.32
3	0.92	0.81	0.71	0.62	0.53	0.46	0.40
4	0.92	0.80	0.70	0.61	0.54	0.47	0.42
5	0.91	0.78	0.67	0.59	0.51	0.45	0.41
6	0.94	0.84	0.76	0.67	0.60	0.53	0.47
7	0.94	0.85	0.76	0.67	0.59	0.52	0.45
8	0.94	0.84	0.74	0.65	0.57	0.50	0.43
9	0.92	0.80	0.70	0.60	0.52	0.45	0.39
10	0.91	0.80	0.70	0.61	0.53	0.46	0.41

Table 3. Auto-covariance coefficients of LSP.

	j=1	2	3	4	5	6	7
i=1	0.94	0.86	0.79	0.73	0.68	0.63	0.59
2	0.89	0.74	0.62	0.52	0.43	0.36	0.31
3	0.90	0.77	0.66	0.57	0.50	0.43	0.38
4	0.92	0.80	0.70	0.61	0.54	0.48	0.44
5	0.94	0.86	0.77	0.69	0.61	0.54	0.48
6	0.94	0.84	0.75	0.66	0.58	0.50	0.44
7	0.92	0.81	0.70	0.61	0.53	0.45	0.39
8	0.91	0.78	0.68	0.58	0.50	0.43	0.38
9	0.90	0.77	0.67	0.58	0.50	0.44	0.39
10	0.89	0.75	0.65	0.57	0.50	0.44	0.39

of subjective test is shown in Fig. 1(b). From the figure, it is seen that the proposed parameters have better subjective quality than LSP.

4. SPECTRAL QUANTIZATION USING INTERFRAME PREDICTION

In this section, we will show the interframe correlation and quantization performance for the proposed parameters and also compare them with those for LSP. In the following, we set $(\alpha, \gamma) = (0.31, -1/3)$. These values are the same as those used in the CELP coder which is proposed in [3].

4.1. Interframe Correlation

To investigate the interframe correlation, we computed the auto-covariance coefficients $\psi_i(j)$ between $\tilde{\omega}_{n,i}$ and $\tilde{\omega}_{n-j,i}$. The result is shown in Table 2. For comparison, the auto-covariance coefficients of LSP are also shown in Table 3. The results indicate that the proposed parameters have a stronger correlation in neighboring frames than LSP parameters.

4.2. VQ Using Interframe MA Prediction

Fig. 2 shows the cepstral distortion of LSP parameters and the proposed parameters versus the order of MA prediction. In the experiment, two-stage VQ (8 bits are assigned to each stage) is used. LSP parameters and the proposed parameters are quantized with weighted Euclidean and Euclidean distance measures, respectively. It can be seen from the figure that, as the order of MA prediction increases, the cepstral distortion for the proposed parameters becomes smaller than that for LSP.

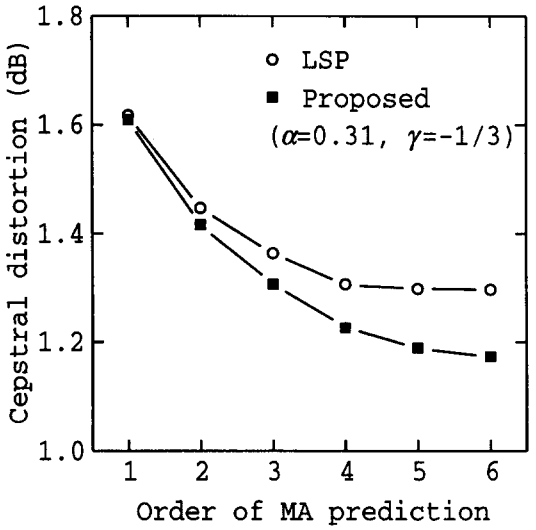

Fig. 2. Effect of MA interframe prediction.

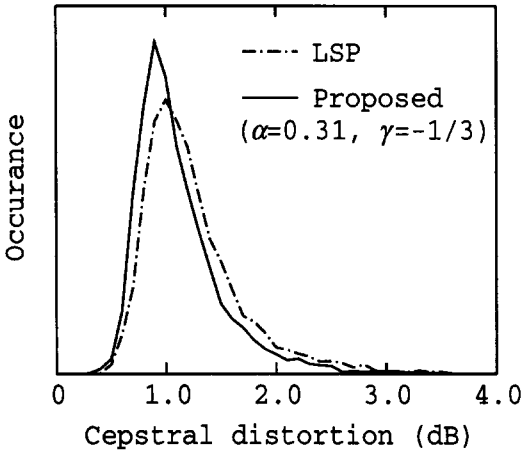

Fig. 3. Cepstral distortion histograms for sixth-order MA prediction.

An important issue in encoding the spectral parameters is that of distribution of the cepstral distortion. The histograms of the spectral distortion for sixth-order MA prediction are presented in Fig. 3. This figure also indicates that utilizing interframe prediction enables us to efficiently encode the proposed parameters.

5. CELP SPEECH CODING

We have implemented two CELP coders, one is based on LPC (conventional CELP) and another is mel-generalized cepstral analysis (proposed CELP).

Table 4. Bit allocations.

	bits/frame
Spectral Parameters	16
Power	5
Adaptive Codebook	8
Algebraic Codebook	21
Gain Codebook	7
Total	57 (5.7 kbits/s)

5.1. Coder Structure

The frame length is 10 ms and the bit allocation is summarized in Table 4. The coders compute spectral information every 10ms frame. The spectral parameters are coded by two-stage VQ (8 bits are assigned to each stage) with interframe MA prediction. The power calculated in each frame is quantized in the μ-law domain. The algebraic codebook [13] is adopted for excitation codebook. The excitation vector contains four non-zero pulse whose signs and positions is restricted. A pitch sharpening procedure is incorporated by filtering the algebraic codevector through the AR comb filter. The gains of adaptive and excitation codebooks are vector quantized.

5.2. Objective Evaluation

Fig. 4 shows SNR and segmental SNR performance of two coders. The perceptual weighting filter of the conventional and the proposed CELP are defined by $A(z/0.9)/A(z/0.4)$ and $C(\tilde{z})C(\tilde{z})/C(\tilde{z}/0.7)$, respectively, where $A(z)$ is the LPC polynomial. As the prediction order increases, SNR and segmental SNR become higher. It is shown that the proposed CELP coder has slightly better performance than the conventional CELP.

6. CONCLUSIONS

The performance of several algorithms for the quantization of the mel-generalized cepstral coefficients has been studied. First, objective and subjective performance of 24 bits two-stage VQ have been measured. Subjective test has shown that the quantization performance for the mel-generalized cepstral coefficients is higher than that for LSP. Secondly, we have utilized the interframe correlation to encode the mel-generalized cepstral coefficients. As a result, the cepstral distortion is improved over that of LSP. Finally, we have implemented the CELP coder, in which the spectral parameters are quantized using interframe MA prediction, and shown that CELP coder based on mel-generalized cepstral analysis has slightly better performance than conventional CELP in terms of SNR and segmental SNR.

ACKNOWLEGEMENT

This work is supported in part by Research Fellowships of Japan Society for the Promotion of Science for Young Scientists, and in part by Support for International Research of International Communication Foundation.

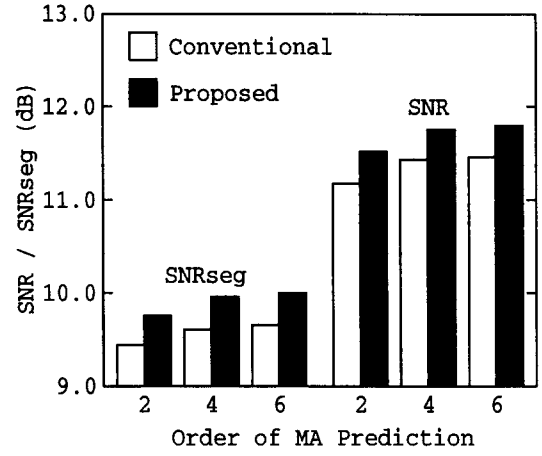

Fig. 4. Objective evaluation.

REFERENCES

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

[2] 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.

[3] K. Koishida, K. Tokuda, T. Kobayashi and S. Imai, "CELP coding system based on mel-generalized cepstral analysis," *Proc. ICSLP'96*, pp.318–321, 1996.

[4] A. Kataoka, T. Moriya and S, Hayashi, "Implementation and performance of an 8-kbit/s conjugate structure CELP speech coder," *Proc. ICASSP'94*, pp.II-93–II-96, 1994.

[5] Y. Shoham, "Vector predictive quantization of spectral parameters for low bit rate speech coding," *Proc. ICASSP'87*, pp.2181–2184, 1987.

[6] N. Farvardin and R. Laroia, "Efficient encoding of speech LSP parameters using the discrete cosine transformation," *Proc. ICASSP'89*, pp.168–171, 1989.

[7] S. Imai and C. Furuichi, "Unbiased estimator of log spectrum and its application to speech signal processing," *Proc. EURASIP'88*, pp.203–206, 1988.

[8] K. Tokuda, T. Kobayashi and S. Imai, "Generalized cepstral analysis of speech —unified approach to LPC and cepstral method," *Proc. ICSLP'90*, pp.37–40, 1990.

[9] K. Koishida, K. Tokuda, T. Kobayashi and S. Imai, "Spectral representation of speech using mel-generalized cepstral coefficients," *3rd Joint Meeting of ASA and ASJ*, pp.963–968, 1996.

[10] B. H. Juang and A. H. Gray, Jr., "Multiple stage vector quantization for speech coding," *Proc. ICASSP-82*, pp.597–600, 1982.

[11] Y. Linde, A. Buzo and R. M. Gray, "An algorithm for vector quantizer design," *IEEE Trans. Communications*, 28, pp.84–95, Sep. 1980.

[12] N. Phamdo, N. Farvardin and T. Moriya, "Combined source-channel coding of LSP parameters using multi-stage vector quantization," *IEICE Technical Report*, SP90-52, pp.63–70, 1990.

[13] R. Salami, C. Laflamme and J-P Adoul, "8 kbit/s ACELP coding of speech with 10 ms speech frame: a candidate for CCITT standardization," *Proc. ICASSP'94*, pp.II-97–II-100, 1994.