5th International Conference on Spoken Language Processing (ICSLP 98)
Sydney, Australia
November 30 - December 4, 1998

# A 16 KBIT/S WIDEBAND CELP CODER USING MEL-GENERALIZED CEPSTRAL ANALYSIS AND ITS SUBJECTIVE EVALUATION

*Kazuhito Koishida* †§,* *Gou Hirabayashi* ‡*, *Keiichi Tokuda* ††, and *Takao Kobayashi* §

†Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106, U.S.A.
§Interdisciplinary Grad. School of Science and Engr., Tokyo Institute of Technology, Yokohama, 226-8502 JAPAN
‡Kansai Research Laboratories, Toshiba Corporation, Kobe, 658-0015 JAPAN
††Department of Computer Science, Nagoya Institute of Technology, Nagoya, 466-8555 JAPAN

10.21437/ICSLP.1998-389

## ABSTRACT

We have proposed a wideband CELP coder, called MGC-CELP, which provides high quality speech by utilizing mel-generalized cepstral (MGC) analysis instead of linear prediction (LP). In this paper, we investigate the performance of the wideband MGC-CELP coder at 16 kbit/s in terms of short-term predictor order, i.e., order of MGC analysis. Subjective tests show that the MGC-CELP coder with a predictor of order 20 gives better performance than ITU-T G.722 at 64 kbit/s. It is also found that the MGC-CELP coder with 12th order achieves comparable quality to the 64 kbit/s G.722, and outperforms the 16 kbit/s conventional CELP coder using 20th-order LP analysis under the same conditions.

## 1. INTRODUCTION

Recently several schemes for high-quality wideband speech coding at low bit rates have been developed. Most of the work in this field uses either transform/subband coding or CELP (Code Excited Linear Prediction) coding. At the bit rates around 16 kbit/s, CELP coding has received much attention since it is able to provide high quality speech with low coding delay. While the performance of CELP coding has been improved by various techniques, they have concentrated in excitation structure. In contrast, our approach to enhance the quality of CELP coding is to incorporate efficient spectral modeling instead of all-pole modeling. By way of example, we have proposed a wideband CELP coder using mel-generalized cepstral (MGC) analysis instead of linear prediction (LP), namely MGC-CELP [1]. A distinguishing feature of the MGC-CELP coder is to adopt frequency warping for encoding fullband of wideband speech signals. From listening tests, it has been shown that frequency warping makes a large contribution to the improvement of subjective quality.

This paper investigates the performance of the wideband MGC-CELP coder at 16 kbit/s in terms of short-term predictor order, i.e., order of MGC analysis. In conventional fullband CELP coders, LP analysis of order 16 to 20 is generally required to obtain sufficient performance. On the other hand, since the most important formants in wideband speech are typically located below 4 kHz, the frequency warping provides efficient representation of wideband speech spectrum using relatively low analysis order. Subjective tests will demonstrate that the MGC-CELP coder can reduce the predictor order of MGC analysis while maintaining high quality. The computational aspects of the wideband MGC-CELP coder will be also discussed.

* This work was performed when the authors were at Tokyo Institute of Technology.

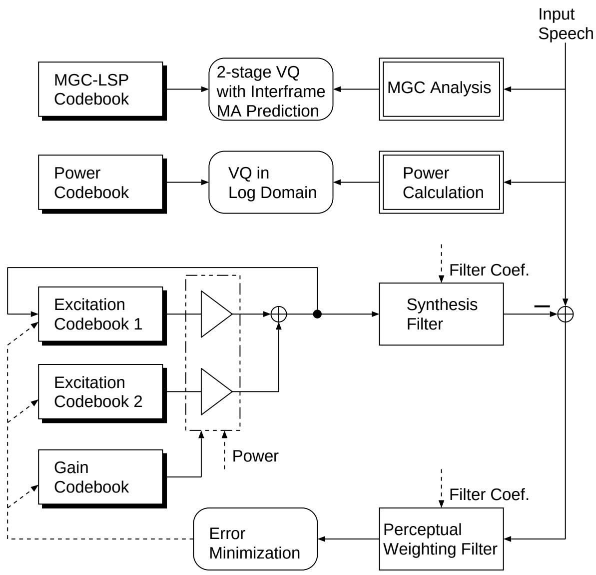



**Figure 1:** Structure of wideband MGC-CELP coding.

## 2. WIDEBAND MGC-CELP CODING

Fig. 1 illustrates the basic structure of the wideband MGC-CELP coder. The basic framework is the same as conventional CELP coder, while the MGC-CELP coder utilizes the MGC analysis instead of linear prediction. The differences between conventional CELP and MGC-CELP coding are therefore the spectral parameters, synthesis filter, perceptual weighting filter and postfilter. We will briefly describe these differences below.

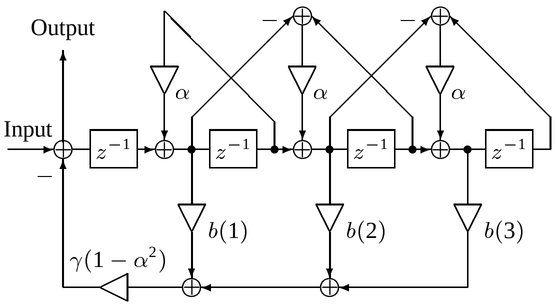



**Figure 2:** Structure of $1/C_1(\tilde{z})$ for $M = 3$.

## 2.1. Spectral analysis and quantization

In MGC analysis [2], we assume that a speech spectrum $H(e^{j\omega})$ is modeled by the MGC coefficients $c(m)$ as

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

where $K$ is a gain and $\tilde{z}^{-1}$ is an all-pass transfer function defined by

$$\tilde{z}^{-1} = \frac{z^{-1}-\alpha}{1-\alpha z^{-1}}, \quad |\alpha| < 1. \quad (2)$$

The parameters $\alpha$ and $\gamma$ control the frequency warping and the weight for pole/zero representation, respectively. The wideband MGC-CELP coding uses a value of $\gamma = -1/2$, and Eq. (1) is therefore reduced to

$$H(z) = \frac{K}{C(\tilde{z})C(\tilde{z})} \quad (3)$$

where

$$C(\tilde{z}) = 1 + \gamma\sum_{m=0}^{M} c(m)\tilde{z}^{-m} \quad (4)$$

and the gain of $1/C(\tilde{z})$ is unity.

The optimum set of the MGC coefficients, for which the residual energy is minimized, can be obtained using an efficient iterative algorithm. In addition, the stability of model solution is always guaranteed.

For quantization and interpolation, MGC coefficients are transformed into MGC-LSP parameters [3]. The MGC-LSP is a frequency-domain representation of speech similar to LSP, and defined on the warped frequency scale.

## 2.2. Synthesis filter

Eq. (3) indicates that the synthesis filter is realized by two-stage cascade structure of $1/C(\tilde{z})$. The structure of filter $1/C(\tilde{z})$ is shown in Fig. 2. The filter coefficients $b(m)$ can be obtained from $c(m)$ using a recursive formula with $M$ multiply-add operations. [1]. It is noted that, since the gain of $1/C(\tilde{z})$ is identical with unity, $b(0)$ always becomes zero [4].

**Table 1:** Bit allocations of 16 kbit/s MGC-CELP coder.

| | Subframe | Frame |
|---|---|---|
| MGC-LSPs | – | 21 |
| Power | – | 7 |
| Excitation codebook 1 | 9 | 9×4 |
| Excitation codebook 2 | 17 | 17×4 |
| Gain codebook | 7 | 7×4 |
| Total | – | 160 bits |

## 2.3. Perceptual weighting

The perceptual weighting filter is defined by the MGC coefficients as

$$S_{pw}(z) = \frac{C(\tilde{z}/\beta_1)}{C(\tilde{z}/\beta_2)} \quad (5)$$

where $\tilde{z}/\beta$ indicates the bandwidth expansion in $\tilde{z}$-plane [4]. The filter $C_1(\tilde{z}/\beta)$ can be realized using the same structure as $C(\tilde{z})$. We set the tunable parameters of the perceptual weighting to $\beta_1 = 1.0$ and $\beta_2 = 0.0$, i.e., $S_{pw}(z) = C(\tilde{z})$.

## 2.4. Short-term postfilter

The short-term postfilter is defined by

$$S_{st}(z) = \frac{C(\tilde{z}/\beta_3)}{C(\tilde{z}/\beta_4)}. \quad (6)$$

The tilt compensation filter has a structure of the form

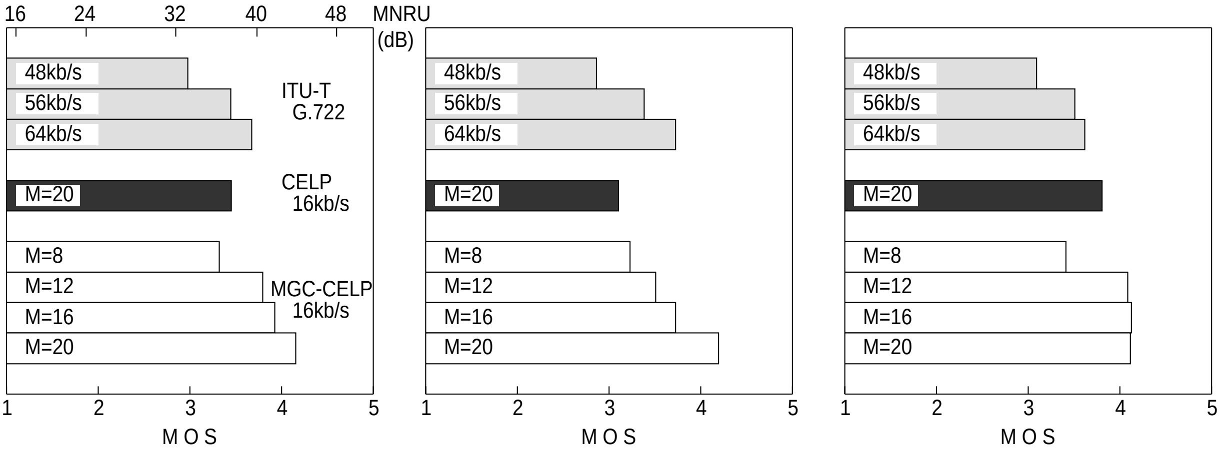
$$S_{tilt}(z) = (1-\mu z^{-1})^p \quad (7)$$

where $\mu$ is a parameter to compensate the global spectral tilt caused by the short-term postfilter. The parameter $\mu$ is adaptively controlled in the mel-cepstrum domain [1]. By informal listening, we set to $(\beta_3, \beta_4, p) = (0.8, 0.9, 2)$.

# 3. 16 KBIT/S WIDEBAND MGC-CELP CODER WITH 10 MSEC FRAME

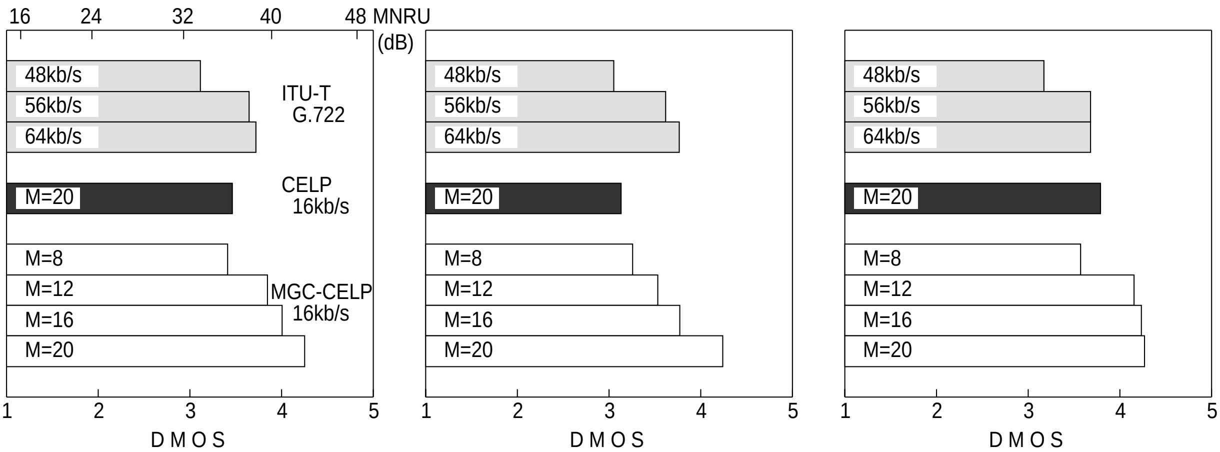
Table 1 shows bit allocations of the MGC-CELP coder at 16 kbit/s. The frame of 10 msec is divided into four subframes. The MGC coefficients are obtained using 32-msec Hamming window centered by the middle of the last subframe.

The MGC-LSP parameters are encoded once per frame using a switched two-stage VQ with moving-average (MA) interframe prediction [5]. The selection of MA predictive coefficients uses 1 bit. In the first stage, the MGC-LSP parameters are quantized using a 8-bit codebook. The vector of the second stage is split




(a) average (b) female (c) male

**Figure 3:** Result of ACR test.




(a) average (b) female (c) male

**Figure 4:** Result of DCR test.

into lower and higher parts, and 6 bits are assigned to each part. The MGC-LSP parameters are quantized with Euclidean distance measure.

The power parameter is calculated on a two-subframe basis, and therefore two dimensional vector of the power parameters is obtained once per frame. The vector is quantized into 7 bits in the logarithmic domain.

The excitation codebook 1 consists of an adaptive codebook and a fixed codebook. The excitation codebook 2 is based on the algebraic codebook structure used in G.729 [6]. The gains of excitation codebook 1 and 2 are vector-quantized using a 7-bit codebook.

In the decoder, postfilter is used to enhance the subjective quality. The postfilter consists of three filters: long-term postfilter, short-term postfilter and tilt compensation filter.

## 4. SUBJECTIVE TESTS

Subjective tests were conducted in a sound-proof booth to evaluate the MGC-CELP coder at 16kbit/s for several values of short-term predictor order $M$. Eight people took part in the tests. The input speech was sampled at 16 kHz and filtered by the sending filter P.341, and the speech level was adjusted at $-26$ dB.

For comparison purpose, the ITU-T G.722 (48, 56 and 64 kbit/s) and conventional CELP (16 kbit/s) are included in the subjective tests. Except for some differences, the configuration of the conventional CELP coder is the same as Fig. 1 and Table 1. The differences are listed below:

- LP method is used for spectral analysis. After obtaining LP coefficients using Levinson-Durbin algorithm, a bandwidth expansion of factor 0.996 is performed.

- The LSP parameters are obtained from LP coefficients and quantized with weighted Euclidean distance [5].
- The synthesis filter is of the form $1/A(z)$ where $A(z)$ is the LP inverse filter.
- The transfer function of perceptual weighting filter is defined by $A(z/0.9)/A(z/0.6)$.
- The short-term postfilter is given by $A(z/0.65)/A(z/0.75)$ and tilt compensation filter is defined by the first order all-zero structure as $(1 - 0.15k_1z^{-1})$ where $k_1$ is the first reflection coefficient.

From informal listening, the frequency warping parameter was set to be $0.3$ in the MGC-CELP coder.

### 4.1. Results

Figs. 3 and 4 show the results of ACR and DCR tests, respectively. It is seen from these figures that the quality of the MGC-CELP coder is enhanced with increasing predictor order $M$, especially for female speech. Note that informal listening tests show that values of $M$ greater than 20 give no improvement of subjective quality.

In the range from $M = 12$ to 20, the MGC-CELP coder produces higher quality speech than conventional CELP coder using the 20th-order LP analysis. The performance of the MGC-CELP coder with $M = 8$ is almost the same as that of the conventional one. These results indicate that MGC analysis is capable of providing efficient representation of wideband speech spectrum.

It is also shown that the MGC-CELP coder with $M = 20$ outperforms the 64 kbit/s G.722. The quality of the MGC-CELP coder is found to be comparable to G.722 at 64 and 48 kbit/s for $M = 12$ and 8, respectively.

## 5. COMPUTATIONAL ASPECTS OF WIDEBAND MGC-CELP CODING

This section discusses the computational aspects of wideband MGC-CELP coder. The excitation codebook search of MGC-CELP coder has high complexity as well as that of conventional CELP coder, and it far exceeds other operations. The excitation search complexity of conventional CELP and MGC-CELP coders is the identical, if it is based on impulse response of the weighted synthesis filter.

As shown in subsection 2.1, Eq. (1) becomes rational function for $\gamma = -1/2$, which makes it possible to further reduce the complexity of MGC analysis. Assuming that the complexity of the 20th-order LP analysis is regarded as unity, the complexity of MGC analysis with $\gamma = -1/2$ is about 4 and 7 for $M = 12$ and 20, respectively[1]. On the other hand, the computation for transforming MGC coefficients into MGC-LSP parameters is less than that for LSP parameters. This is because, in root search procedure, LSP parameters require smaller interval to separate adjacent parameters than MGC-LSPs for $\gamma = -1/2$ [3]. Moreover, since the MGC-LSPs codebook search uses Euclidean distance, a further reduction in the complexity can be obtained. If our attention is restricted to the computational complexity except for the excitation codebook search, the MGC-CELP coder with $M = 20$ still requires about 4 times as high complexity as conventional CELP coder with the 20th-order LP analysis does.

The above discussion leads to the conclusion that, while the overall complexity of the MGC-CELP coder increases compared to conventional CELP coder, the difference is not significant.

## 6. CONCLUSIONS

In this paper, the performance of a wideband CELP coder at 16 kbit/s based on MGC analysis has been investigated in terms of short-term predictor order, i.e., order of MGC analysis. Subjective tests have shown that the MGC-CELP coder with a predictor of order 20 gives better performance than ITU-T G.722 at 64 kbit/s. It has been also found that the MGC-CELP coder of order 12 achieves comparable quality to the 64 kbit/s G.722, and outperforms the 16 kbit/s conventional CELP coder using 20th-order LP analysis under the same conditions.

## ACKNOWLEDGMENT

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

## REFERENCES

[1] K. Koishida, G. Hirabayashi, K. Tokuda and T. Kobayashi, "A wideband CELP speech coder at 16 kbit/s based on mel-generalized cepstral analysis," *Proc. ICASSP'98*, pp.157–160, 1998.

[2] 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.

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

[4] 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.

[5] H. Ohmuro, T. Moriya, K. Mano and S. Miki, "Coding of LSP parameters using interframe moving average prediction and multi-stage vector quantization," *Proc. IEEE Speech Coding Workshop*, pp63–64, 1993.

[6] R. Salami, C. Laflamme, J-P. Adoul, A. Kataoka, S. Hayashi, C. Lamblin, D. Massaloux, S. Proust, P. Kroon and Y. Shoham, "Description of the proposed ITU-T 8kb/s speech coding standard," *IEEE Speech Coding Workshop*, Annapolis, 1995.

[1] The number of iteration is fixed to be 3. We confirmed that 3-iteration gives the sufficient convergence in almost all frames.