IEICE TRANS. INF. & SYST., VOL.E83–D, NO.4 APRIL 2000

PAPER

A 16 kb/s Wideband CELP-Based Speech Coder Using Mel-Generalized Cepstral Analysis

Kazuhito KOISHIDA[†*], *Member*, **Gou HIRABAYASHI**[†**], *Nonmember*, **Keiichi TOKUDA**[††], *and* **Takao KOBAYASHI**[†], *Members*

SUMMARY We propose a wideband CELP-type speech coder at 16 kb/s based on a mel-generalized cepstral (MGC) analysis technique. MGC analysis makes it possible to obtain a more accurate representation of spectral zeros compared to linear predictive (LP) analysis and take a perceptual frequency scale into account. A major advantage of the proposed coder is that the benefits of MGC representation of speech spectra can be incorporated into the CELP coding process. Subjective tests show that the proposed coder at 16 kb/s achieves a significant improvement in performance over a 16 kb/s conventional CELP coder under the same coding framework and bit allocation. Moreover, the proposed coder is found to outperform the ITU-T G.722 standard at 64 kb/s.

key words: *wideband speech coding, mel-generalized cepstral analysis, mel-generalized cepstrum, CELP coding*

1. Introduction

Current telephone networks generally limit the bandwidth of the speech signals to between 300 to 3400 Hz. This bandwidth limitation degrades a speech quality in terms of naturalness, intelligibility and speaker recognition. However, future transmission networks can eliminate the bandwidth limitation of 300–3400 Hz since they are digital end-to-end. Such networks are capable of delivering face-to-face communications quality by increasing the bandwidth of speech signals. Specifically, wideband speech, whose bandwidth ranges from 50 to 7000 Hz, is of increasing interest today because wideband speech offers a much higher quality than telephone speech. Extending the lower frequency range down to 50 Hz increases naturalness, and the higher frequency range of 3400–7000 Hz improves intelligibility. While wideband speech is primarily used for teleconferencing and videoteleconferencing today, it is expected that wideband speech will be employed for a number of applications such as multimedia services and future ISDN voice communication. In most of these applications, wideband speech coding plays a key role in efficient transmission and storage of wideband speech signals.

In wideband speech signals, most of the important formants are typically located at low frequencies, so that the energy in the high frequency region is smaller than that in the low frequency region. Transform and subband coding schemes [1], [2] obtain high-quality speech by exploiting these characteristics. The basic principle of transform and subband schemes is to decompose the speech signal into subbands and separately encode each band. The ITU-T G.722 [3], a two-subband ADPCM coder, is the current standard at 64, 56 and 48 kb/s for wideband applications. On the other hand, code excited linear prediction (CELP) coding [4] has received much attention in recent years due to its great success in the field of telephone-band speech coding. CELP coding schemes achieve high performance by utilizing a parametric model based on speech production and analysis-by-synthesis algorithms.

Wideband CELP coders are traditionally divided into two classes; split-band CELP coding [5]–[8], and fullband CELP coding [9]–[13]. In split-band CELP coding which belongs to the class of subband coders, each subband signal is encoded by a CELP coder. The split-band CELP coders generally suffer from degradation of speech quality in the frequency region where the responses of the filterbanks overlap. In contrast to split-band CELP coding, a fullband coding scheme directly encodes wideband speech signals using only one CELP coder. The fullband CELP coders usually suffer from a high frequency noise in the decoded speech. To increase the quality at high frequencies, several methods have been reported, e.g., a modified perceptual weighting filter [11], a split-band excitation [12] and a multi-band excitation [13].

One promising approach for improving the performance of fullband CELP coding is to incorporate a frequency-warping technique into short-term spectral modeling. This makes it possible to obtain a perceptual frequency scale, in which the frequency resolution at high frequencies is less sharp than at low ones. In addition, since the most energy of wideband speech is concentrated in low frequencies as described before, frequency warping also has an ability to take such properties into consideration. Although several

Manuscript received April 13, 1999.

[†]The authors are with Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama-shi, 226–8502 Japan.

[††]The author is with the Department of Computer Science, Nagoya Institute of Technology, Nagoya-shi, 466–8555 Japan.

[*]Presently, with Signal Compression Laboratory, Department of Electrical and Computer Engineering, University of California, Santa Barbara, CA 93106-9560, USA.

[**]Presently, with Human Interface Laboratory, Research & Development Center, Toshiba Corporation, Kobe-shi, 658–0015 Japan.

studies have shown that the use of frequency warping enhances the perceptual performance of telephone-band speech coders [14]–[17] and audio coders (for a bandwidth of 20.7 kHz) [18], [19], this approach has not yet received much attention in the area of wideband speech coding.

In this paper, we propose a wideband CELP-type speech coder based on a mel-generalized cepstral (MGC) analysis technique [20], called MGC-CELP. The proposed MGC-CELP coder uses a framework of full-band CELP coding, and its distinguishing feature is to exploit MGC analysis instead of linear predictive (LP) analysis. MGC analysis makes it possible to introduce a frequency-warping technique into CELP coding. Moreover, while LP analysis has limitations of representing spectral zeros, MGC analysis can provide more accurate representation of spectral zeros. A major advantage of the MGC-CELP coder is that the benefits of MGC representation of speech spectra can be incorporated into the CELP coding process. We will show that the MGC-CELP coder can generate high-quality speech at 16 kb/s.

This paper is organized as follows: Sect. 2 gives an overview of the wideband MGC-CELP algorithm. In Sect. 3, a specific implementation of this algorithm at 16 kb/s is described. Subjective evaluation of the coder is performed in Sect. 4. Finally, conclusions are given in Sect. 5.

2. Wideband MGC-CELP Coding

Figure 1 shows a simplified block diagram of wideband MGC-CELP coding. The basic framework of MGC-CELP coding is identical to that of conventional CELP coding. However, MGC-CELP utilizes MGC analysis instead of LP analysis and, as a result, the parts related to MGC analysis are completely different from the corresponding parts in conventional CELP. In this section, we focus the discussion on these parts.

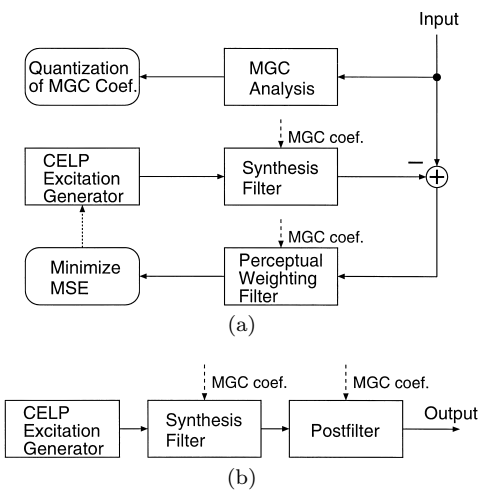


Fig. 1 Simplified block diagram of MGC-CELP coding: (a) encoder, (b) decoder.

2.1 MGC Analysis

In MGC analysis, a speech spectrum $H(e^{j\omega})$ is assumed to be modeled by a set of MGC coefficients $c(m)$ as

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

where K is a gain of $H(z)$ and

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

In the above equation, $\tilde{z}^{-1}$ is an all-pass transfer function defined by

$$\tilde{z}^{-1} = \frac{z^{-1} - \alpha}{1 - \alpha z^{-1}}, \qquad |\alpha| < 1. \tag{3}$$

The parameter α controls the frequency warping. A linear frequency scale is obtained for $\alpha = 0$, and the frequency resolution at low frequencies improves with increasing α at the expense of decreasing the high frequency resolution. When $\alpha = 0.42$, the phase characteristics of the all-pass system give a good approximation to the mel scale for a sampling frequency of 16 kHz. The parameter γ controls the representation accuracy of poles and zeros. The function $H(z)$ becomes all-pole modeling for $\gamma = -1$ and $\alpha = 0$. As the value of γ approaches zero, the accuracy for spectral zeros increases at the expense of formant accuracy. For $\gamma = \alpha = 0$, $H(z)$ is identical to the cepstral representation in which poles and zeros are represented with equal weights.

The optimum set of MGC coefficients, which minimizes the expectation value of the square of the prediction residual, can be obtained using an efficient iterative algorithm based on the FFT and recursive formulas [20]. It has been proven that coefficients obtained with this algorithm result in stable systems [20].

In wideband MGC-CELP coding, the parameter γ is fixed to be $-1/2$. This value gives a good representation of spectral poles and zeros as shown in Fig. 2, and also offers some implementation advantages which will be described later. An appropriate value of α will be determined through listening tests.

2.2 Synthesis Filter

For $\gamma = -1/2$, the synthesis filter is realized by a rational transfer function of the form

$$S(z) = \frac{1}{\{C(\tilde{z})\}^2} \tag{4}$$

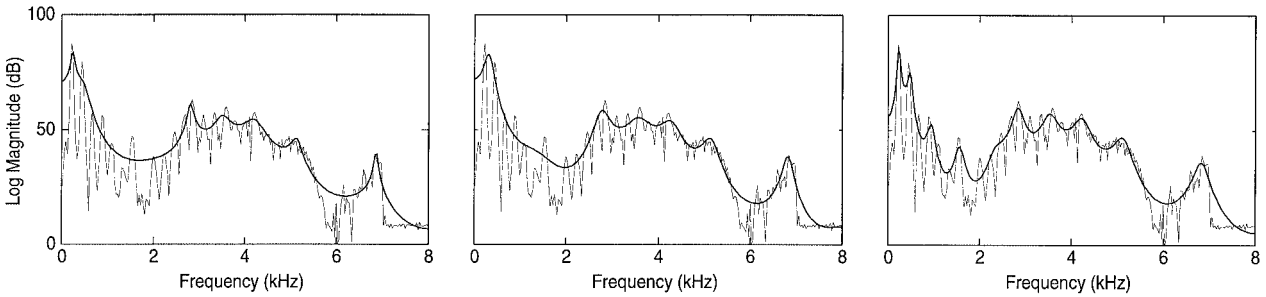


Fig. 2 Example of spectral envelopes obtained by 20th-order analysis using a 32 ms Hamming window: (a) LP analysis, (b) MGC analysis with $\alpha = 0$ and $\gamma = -1/2$, (c) MGC analysis with $\alpha = 0.3$ and $\gamma = -1/2$. Thin lines represent the FFT spectrum.

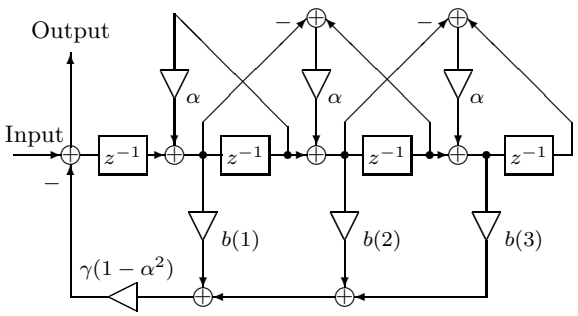


Fig. 3 Filter structure of $1/C(\tilde{z})$ for $M = 3$.

where

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

To remove delay-free loops from $S(z)$, we modify Eq. (5) as follows:

$$C(\tilde{z}) = 1 + \gamma \sum_{m=1}^{M} b(m)\,\Phi_m(z) \tag{6}$$

where

$$\Phi_m(z) = \frac{(1-\alpha^2)z^{-1}}{1-\alpha z^{-1}}\tilde{z}^{-(m-1)}, \qquad m \geq 1 \tag{7}$$

and the filter coefficients $b(m)$ are obtained using a recursive formula given by

$$b(m) = \begin{cases} c(M), & m = M \\ c(m) - \alpha b(m+1), & 0 \leq m < M. \end{cases} \tag{8}$$

Note that, since the gain of $S(z)$ is unity, the coefficient $b(0)$ is always zero [17]. The structure of $1/C(\tilde{z})$ based on Eq. (6) is shown in Fig. 3.

2.3 MGC-LSP Parameters

In MGC-CELP coding, MGC-based line spectrum pair (MGC-LSP) parameters [21] are quantized and transmitted instead of MGC coefficients. The MGC-LSP is an alternative representation of the MGC coefficients and is mathematically equivalent to the MGC coefficient representation. The MGC-LSP representation offers a simple check of the filter stability and gives good interpolation and quantization performance [22].

The MGC-LSP is a frequency-domain representation of speech similar to the LSP, but is defined on the warped frequency scale. The procedure for obtaining the MGC-LSP parameters is as follows: first Eq. (5) is modified as

$$C(\tilde{z}) = (1 + \gamma c(0))C_1(\tilde{z}) \tag{9}$$

where

$$C_1(\tilde{z}) = 1 + \gamma \sum_{m=1}^{M} c_1(m)\tilde{z}^{-m} \tag{10}$$

and the coefficients $c_1(m)$ are calculated from the MGC coefficients using the relationship

$$c_1(m) = \frac{c(m)}{1 + \gamma c(0)}, \qquad 1 \leq m \leq M. \tag{11}$$

Next $C_1(\tilde{z})$ is decomposed into symmetric and antisymmetric polynomials:

$$C_P(\tilde{z}) = C_1(\tilde{z}) + \tilde{z}^{-(M+1)}C_1(\tilde{z}^{-1}) \tag{12}$$

$$C_Q(\tilde{z}) = C_1(\tilde{z}) - \tilde{z}^{-(M+1)}C_1(\tilde{z}^{-1}). \tag{13}$$

Finally, using the same technique as for the LSP case, the MGC-LSP parameters can be obtained as the angular positions of the roots of $C_P(\tilde{z})$ and $C_Q(\tilde{z})$.

Histograms of the LSP and MGC-LSP parameters with $\gamma = -1/2$ are shown in Fig. 4. It is noted that, in Fig. 4 (c), the MGC-LSP parameters are plotted on the warped frequency scale. It can be seen that the variance of the MGC-LSP parameters is smaller than that of the LSP parameters. The reason is that the filter $1/C(\tilde{z})$ has less sharp peaks since the synthesis filter is realized by the two-stage cascade structure of $1/C(\tilde{z})$ as shown in Eq. (4). Consequently, the MGC-LSP parameters require less precision to avoid missing roots of $C_P(\tilde{z})$ and $C_Q(\tilde{z})$ in their root search [21].

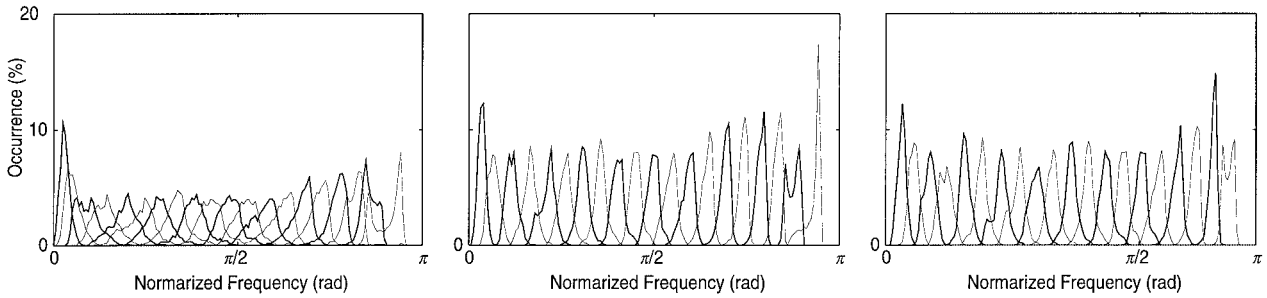


Fig. 4 Histograms of (a) LSP parameters; (b) MGC-LSP parameters with $\alpha = 0, \gamma = -1/2$; (c) MGC-LSP parameters with $\alpha = 0.3, \gamma = -1/2$. These parameters are obtained by 20th-order analysis using a 32 ms Hamming window. Thin and thick lines represent the parameters of even and odd order, respectively.

2.4 Perceptual Weighting Filter

The perceptual weighting filter is defined by the MGC coefficients as

$$S_{pw}(z) = \frac{C(\tilde{z}/\beta_1)}{C(\tilde{z}/\beta_2)} \tag{14}$$

where $\tilde{z}/\beta$ indicates a bandwidth expansion in the $\tilde{z}$-plane. The filter $C(\tilde{z}/\beta)$ can be realized using the same structure as $C(\tilde{z})$ [17]; i.e.,

$$C(\tilde{z}/\beta) = 1 + \gamma \sum_{m=1}^{M} b_\beta(m)\, \Phi_m(z). \tag{15}$$

The filter coefficients $b_\beta(m)$ are obtained from $c(m)$ as follows: first, a set of intermediate coefficients $b'_\beta(m)$ is computed using

$$b'_\beta(m) = \begin{cases} \beta^M c(M), & m = M \\ \beta^m c(m) - \alpha b'_\beta(m+1), & 0 \le m < M \end{cases} \tag{16}$$

and the coefficients $b_\beta(m)$ are then calculated by

$$b_\beta(m) = \frac{b'_\beta(m)}{1 + \gamma b'_\beta(0)}, \qquad 1 \le m \le M. \tag{17}$$

2.5 Postfilter

The short-term postfilter is defined by

$$S_{sp}(z) = \frac{C(\tilde{z}/\beta_3)}{C(\tilde{z}/\beta_4)}. \tag{18}$$

The tilt compensation filter has a structure of the form

$$S_{tl}(z) = (1 - \mu z^{-1})^n \tag{19}$$

where μ is a parameter that adaptively controls the global spectral tilt of the postfilter. The value of μ is obtained in such a way that the first mel-cepstral coefficient of the cascaded filter of $S_{sp}(z)$ and $S_{tl}(z)$ is set to be zero [17]. Under such a constraint, μ is given by

$$\mu = \frac{-\gamma(\beta_4 - \beta_3)c_1(1)}{-\alpha\gamma(\beta_4 - \beta_3)c_1(1) + n(1 - \alpha^2)}. \tag{20}$$

3. Wideband MGC-CELP Coder at 16 kb/s

This section describes a wideband MGC-CELP coder at 16 kb/s. Its structure and bit allocation are shown in Fig. 5 and Table 1, respectively. The coder operates on speech frame of 10 ms corresponding to 160 samples at a sampling rate of 16 kHz. The MGC coefficients and the signal power are extracted and encoded for every 10 ms frame, while the excitation parameters are determined once per subframe of 2.5 ms.

3.1 Encoder

Using a Hamming window of 32 ms duration centered at the middle point of the last subframe, 20th-order MGC analysis is performed once per 10 ms frame. The MGC coefficients are quantized in the MGC-LSP domain using a two-stage VQ with switched fifth-order moving average (MA) prediction. The selection of MA predictive coefficients and the first stage codebook require 1 bit and 8 bits, respectively. The input vector of the second stage is split into a lower dimensional part and a higher dimensional part, and 6 bits are assigned to each part. These codebooks are searched with the Euclidean distance measure.

The power parameters are calculated on a two-subframe basis, i.e., a two-dimensional vector of the power parameter is obtained for each 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 [23]. The adaptive codebook represents the pitch periodicity, in which a pitch delay is used with varying resolution. The resolution is 1/4 in the range 33–96 3/4, 1/2 in the range 97–160 1/2 and integers only in the range 161–224. The fixed codebook which contains 64 random codevectors

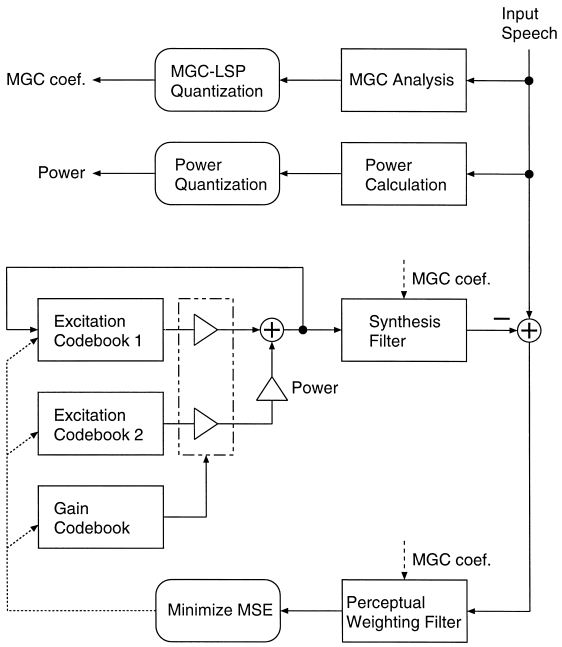


Fig. 5 Structure of wideband MGC-CELP coder at 16 kb/s.

represents the nonperiodic excitation contributions.

The excitation codebook 2 is based on an algebraic codebook structure. In this codebook, each codevector contains four non-zero pulses. These pulses can assume the amplitudes and positions given in Table 2 [24].

The gains of the excitation codebooks are vector-quantized jointly using a 7-bit codebook.

By means of informal listening experiments, we have found that the satisfactory performance is achieved by setting the perceptual weighting parameters to $\beta_1 = 1.0$ and $\beta_2 = 0.0$, i.e., $S_{pw}(z) = C(\tilde{z})$.

3.2 Decoder

In the decoder, the coder parameters are obtained from transmitted indices. The speech is reconstructed by filtering the excitation through the synthesis filter, and is then passed through the postfilter. Besides the short-term postfilter and tilt compensation filter, the postfilter contains a pitch postfilter. We have found that it is necessary to tune the postfilter parameters for different values of the parameter α. Informal listening tests showed that good performance is obtained with $(\beta_3, \beta_4, n) = (0.8, 0.95, 2)$ for $\alpha = 0.0, 0.1, 0.2$ and $(0.8, 0.9, 2)$ for $\alpha = 0.3, 0.4, 0.5$.

4. Subjective Evaluations

4.1 Test Conditions

Table 3 summarizes input speech conditions. The input signal was sampled at 16 kHz and filtered by the sending filter P.341 [25] with 50 to 7000 Hz bandwidth. The speech level was adjusted to −26 dB.

Table 1 Bit allocation of wideband MGC-CELP coder at 16 kb/s. The frame of 10 ms is divided into four subframes.

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

Table 2 Algebraic structure of excitation codebook 2.

Pulse	Amplitude	Positions
1	±1	0, 5, 10, 15, 20, 25, 30, 35
2	±1	1, 6, 11, 16, 21, 26, 31, 36
3	±1	2, 7, 12, 17, 22, 27, 32, 37
4	±1	3, 8, 13, 18, 23, 28, 33, 38 4, 9, 14, 19, 24, 29, 34, 39

Table 3 Input speech conditions for subjective evaluations.

Source Material	Multi-lingual Speech Database for Telephonometry 1994 (NTT-AT)
Sampling Frequency	16 kHz
Bandwidth	50 to 7000 Hz
Word Length	16 bits
Speech Level	−26 dB

Subjective tests were carried out in a sound-proof booth. Eight people took part in the tests. They listened to sixteen Japanese sentences spoken by 4 females and 4 males.

Two types of subjective test were performed for evaluation: an absolute category rating (ACR) test and a degradation category rating (DCR) test. In the ACR test, listeners were presented with the processed speech and gave a 5-point rating. In the DCR test, listeners were presented with the original speech before listening to the processed speech, and gave a 5-point rating according to the degradation. The average rating of ACR and DCR tests for a particular system are referred to as mean opinion score (MOS) and degradation mean opinion score (DMOS), respectively.

4.2 Effectiveness of Frequency Warping

In order to determine an appropriate value of the frequency-warping parameter α, a DCR test was conducted. Figure 6 shows DMOS scores of the 16 kb/s MGC-CELP coder for several values of the parameter α. It is obvious from the figure that frequency warping makes a large contribution to the improvement of the subjective quality. Specifically, it is shown that frequency warping significantly increases the performance of female speech.

From these results, we decided to set the parameter α to 0.3 in the following tests.

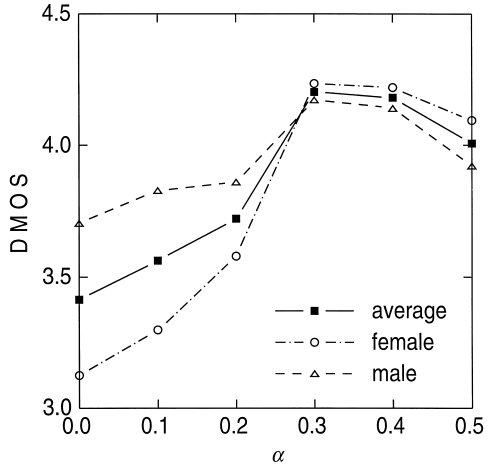


Fig. 6 Speech quality of 16 kb/s MGC-CELP coder as a function of frequency-warping parameter α.

4.3 Comparison with G.722 and Conventional CELP

ACR and DCR tests were carried out to evaluate the 16 kb/s MGC-CELP coder with $\alpha = 0.3$. For comparison purpose, the ITU-T G.722 standard at 64, 56 and 48 kb/s was included in the tests. A conventional CELP coder at 16 kb/s was also included, whose framework and bit allocation are the same as the 16 kb/s MGC-CELP coder. The differences of the conventional CELP coder from the MGC-CELP coder are listed below:

- After the LP coefficients $a(m)$ are computed by 20th-order LP analysis, a bandwidth expansion of 20 Hz is performed.
- The LSP parameters are obtained from the LP coefficients and quantized with a weighted Euclidean distance measure [26].
- The synthesis filter is defined by the LP coefficients $a(m)$ as

$$S(z) = \frac{1}{A(z)} \tag{21}$$

  where $A(z) = \sum_{m=0}^{20} a(m) z^{-m}$ and $a(0) = 1$.
- The perceptual weighting filter is of the form

$$S_{pw}(z) = \frac{A(z/0.9)}{A(z/0.6)}. \tag{22}$$

- The short-term postfilter is given by

$$S_{sp}(z) = \frac{A(z/0.65)}{A(z/0.75)} \tag{23}$$

  and the tilt compensation filter is defined by a first order all-zero structure as

$$S_{tl}(z) = 1 - 0.15k(1)z^{-1} \tag{24}$$

  where $k(1)$ is the first reflection coefficient.

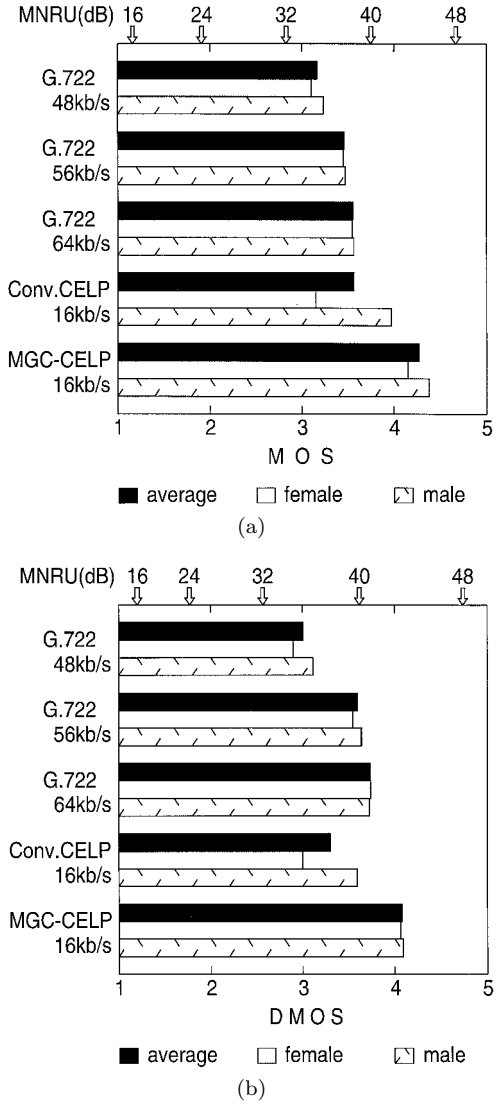


Fig. 7 Results of subjective evaluations: (a) ACR test, (b) DCR test.

Figure 7 shows the results of subjective tests. We see from the figure that, in both tests, the MGC-CELP coder obtains similar scores in terms of equivalent Q values. It is also shown that the MGC-CELP coder at 16 kb/s outperforms the G.722 at 64 kb/s. Moreover, it is clear that a significant improvement is obtained compared to the conventional CELP coder. The improvement is especially noticeable for female speech. As a result, the MGC-CELP coder gives a much smaller difference between male and female speakers than the conventional CELP coder.

5. Conclusions

We have proposed a wideband CELP-type speech coder at 16 kb/s. The proposed coder belongs to the class of fullband CELP coding, and its distinguishing feature is to exploit MGC analysis instead of LP anal-

ysis. MGC analysis makes it possible to incorporate a frequency-warping technique into CELP coding and obtain more accurate representation of spectral zeros. Subjective tests showed that the performance of the proposed coder at 16 kb/s is better than that of the ITU-T G.722 standard at 64 kb/s. It was also found that the proposed coder achieves a significant quality improvement over a conventional CELP coder which has the same coding framework and bit allocation as the proposed coder.

Acknowledgment

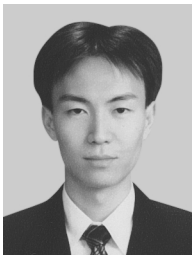
The authors would like to thank Prof. Satoshi Imai, Chiba Institute of Technology, for his valuable discussions.

This work was 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 Communication Research of the International Communication Foundation.

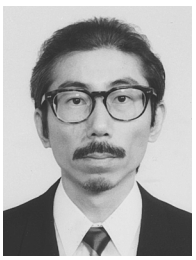
References

[1] R. Zelinski and P. Noll, "Adaptive transform coding of speech signals," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-25, no.4, pp.299–309, Aug. 1977.
[2] J.M. Tribolet and R.E. Crochiere, "Frequency domain coding of speech," IEEE Trans. Acoust., Speech & Signal Process., vol.ASSP-27, no.5, pp.512–531, Oct. 1979.
[3] P. Mermelstein, "A new CCITT coding standard for digital transmission of wideband audio signals," IEEE Commun. Mag., vol.26, no.1, pp.8–15, Jan. 1988.
[4] M. Schroeder and B.S. Atal, "Code excited linear prediction: High quality speech at low bit rates," Proc. ICASSP-85, pp.937–940, 1985.
[5] R. Drago, R. Montagna, F. Perosino, and D. Sereno, "Some experiments of 7 kHz audio coding at 16 kbits/s," Proc. ICASSP-89, pp.192–195, 1989.
[6] A. Kataoka, S. Kurihara, S. Sasaki, and S. Hayashi, "A 16-kbit/s wideband speech codec scalable with G.729," Proc. EUROSPEECH-97, pp.1491–1494, 1997.
[7] J.W. Paulus and J. Schnitzler, "16 kbit/s wideband speech coding based on unequal subbands," Proc. ICASSP-96, pp.255–258, 1996.
[8] A. Murashima and K. Ozawa, "16 kbps wideband speech coding – Subband M-LCELP," Proc. IEICE-95 Spring National Convention Record, vol.D-248, p.251, 1995.
[9] C. Laflamme, J.P. Adoul, S. Morissette, and P. Mabilleau, "16 kbps wideband speech coding technique based on algebraic CELP," Proc. ICASSP-91, pp.13–16, 1991.
[10] S. Sasaki, A. Kataoka, and T. Moriya, "Wideband CELP coder at 16-kbit/s with 10-ms frame," Proc. EUROSPEECH-95, pp.41–44, 1995.
[11] E. Ordentlich and Y. Shoham, "Low-delay code-excited linear-predictive coding of wideband speech at 32 kbps," Proc. ICASSP-91, pp.9–12, 1991.
[12] G. Roy and P. Kabal, "Wideband CELP speech coding at 16 kb/s," Proc. ICASSP-91, pp.17–20, 1991.
[13] A. Ubale and A. Gersho, "A multi-band CELP wideband speech coder," Proc. ICASSP-97, pp.1367–1370, 1997.
[14] E. Krüger and H. Strube, "Linear prediction on a warped frequency scale," IEEE Trans. Acoust., Speech & Signal Process., vol.36, pp.1529–1531, Sept. 1988.
[15] K. Tokuda, T. Kobayashi, T. Fukada, and S. Imai, "Speech coding based on adaptive mel-cepstral analysis and its evaluation," IEICE Trans., vol.J77-A, no.11, pp.1443–1452, Nov. 1994.
[16] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "CELP coding based on mel-cepstral analysis," Proc. ICASSP-95, pp.33–36, 1995.
[17] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "CELP speech coding based on mel-generalized cepstral analysis," IEICE Trans., vol.J81-A, no.2, pp.252–260, Feb. 1998.
[18] Y. Nakatoh, T. Norimatsu, A.H. Low, and H. Matsumoto, "Low bit rate coding for speech and audio using mel linear predictive coding (MLPC) analysis," Proc. ICSLP-98, pp.2591–2594, 1998.
[19] A. Härmä, "Audio coding with warped predictive methods," Licentiate's Thesis, Helsinki University of Technology, 1998.
[20] K. Tokuda, T. Kobayashi, T. Chiba, and S. Imai, "Spectral estimation of speech by mel-generalized cepstral analysis," IEICE Trans., vol.J75-A, no.7, pp.1124–1134, July 1992.
[21] K. Koishida, K. Tokuda, T. Kobayashi, and S. Imai, "Spectral representation of speech based on mel-generalized cepstral coefficients and its properties," IEICE Trans., vol.J80-A, no.11, pp.1999–2006, Nov. 1997.
[22] K. Koishida, "Low bit rate speech coding based on mel-generalized cepstral analysis," Doctoral Thesis, Tokyo Institute of Technology, 1998.
[23] S. Miki, K. Mano, H. Ohmuro, and T. Moriya, "Pitch synchronous Innovation CELP (PSI-CELP)," Proc. EUROSPEECH-93, pp.261–264, 1993.
[24] R. Salami, C. Laflamme, J.-P. Adoul, A. Kataoka, S. Hayashi, T. Moriya, C. Lamblin, D. Massaloux, S. Proust, P. Kroon, and Y. Shoham, "Design and description of CS-ACELP: A toll quality 8 kb/s speech coder," IEEE Trans. Speech & Audio Processing, vol.6, no.2, March 1998.
[25] ITU, Subjective qualification test plan for the ITU-T wideband (7 kHz) speech coding algorithm, version 3.0, Sept. 1996.
[26] H. Ohmuro, K. Mano, and T. Moriya, "Vector-matrix quantization of LSP parameters," IEICE Technical Report, SP91-70, 1991.

Kazuhito Koishida received the B.E. degree in electrical and electronic engineering, and M.E., and Dr.Eng. degrees in intelligence science from Tokyo Institute of Technology, Tokyo, Japan, in 1994, 1995 and 1998, respectively. Since 1996, he has been a research fellow of the Japan Society for the Promotion of Science. He is currently a post-doctoral researcher with Signal Compression Laboratory, University of California, Santa Barbara. His current research interests include speech coding at medium and low bit rates and wideband speech/audio coding. He is a recipient of the 1998 TELECOM System Technology Prize for Student from the Telecommunications Advancement Foundation Award. He is a member of IEEE and ASJ.



Gou Hirabayashi received the B.E. and M.E. degrees from Tokyo Institute of Technology, Tokyo, Japan, in 1996 and 1998, respectively. He joined Toshiba Corporation in 1998. His research interests include speech coding and synthesis.



Keiichi Tokuda was born in Nagoya, Japan, in 1960. He received the B.E. degree in electrical and electronic engineering from the Nagoya Institute of Technology, Nagoya, Japan, the M.E. and Dr.Eng. degrees in information processing from the Tokyo. Institute of Technology, Tokyo, Japan, in 1984, 1986, and 1989, respectively. From 1989 to 1996 he was a Research Associate at the Department of Electronic and Electric Engineering, Tokyo Institute of Technology. Since 1996 he has been with the Department of Computer Science, Nagoya Institute of Technology as Associate Professor. His research interests include speech coding, speech synthesis and recognition and multimodal signal processing. He is a member of IEEE, ASJ and JSAI.

Takao Kobayashi received the B.E. degree in electrical engineering, the M.E. and Dr.Eng. degrees in information processing from Tokyo Institute of Technology, Tokyo, Japan, in 1977, 1979, and 1982, respectively. In 1982, he joined the Research Laboratory of Precision Machinery and Electronics, Tokyo Institute of Technology as a Research Associate. He became an Associate Professor at the same Laboratory in 1989. He is currently a Professor of the Interdisciplinary Graduate School of Science and Engineering, Tokyo Institute of Technology, Yokohama, Japan. His research interests include speech analysis and synthesis, speech coding, speech recognition, and multimodal interface. He is a member of IEEE, ESCA and ASJ.