論文 94-19-4-19

LPC Cepstral 벡터 양자화에 의한 저 전송율 CELP 음성부호기의 스펙트럼 표기

正會員 鄭 在 皓*

Spectrum Representation Based on LPC Cepstral VQ for Low Bit Rate CELP Coder

Jae Ho Jeong* *Regular Member*

요 약

본 논문에서는, 매우 낮은 전송율이 요구되는 음성통신의 환경하에서 CELP 음성 부호기를 사용할 경우, 스펙트럼에 대한 정보를 어떻게 효과적으로 나타낼 것인가에 대하여 고찰하였다. 구체적으로, 스펙트럼에 대한 정보를 나타내는 LPC 파라메타를 cepstrum으로 변형시키고, 변형된 LPC cepstrum 계수들을 효과적으로 벡터 양자화하는 방법을 제시하였다. 벡터 양자화에 사용되는 코드-북의 설계를 위하여, 주파수 대역에서 서로 다른 의미를 갖는 세개의 cepstral distance measure들을 시도하였으며, 각각에 대한 성능이 분석되어졌다. 시뮬레이션을 통하여, 본 논문에서 제시한 LPC cepstral 벡터 양자화 방식이 스펙트럼에 대한 정보를 매우 효과적으로 나타낼 수 있음을 보였다.

ABSTRACT

This paper focuses on how spectrum information can be represented efficiently in a very low bit rate CELP speech coder. To achieve the goal, an LPC cepstral coefficients VQ scheme representing the spectrum information in a CELP coder is proposed. To represent the spectrum information using LPC cepstrums, three different cepstral distance measures having different spectral meanings in the frequency domain are considered, and their performances are compared and analyzed. The experimental results show that spectrum information in low bit rate CELP coders can be represented very efficiently using the proposed LPC cepstral vector quantization scheme.

*仁荷大學校 電子工學科
Dept. of Electronics Eng., InHa Univ.
論文番號 : 9425
接受日字 : 1994年 1月 24日

Ⅰ. INTRODUCTION

Code-Excited Linear Predictive (CELP) coders have been proved to be good low bit rate speech coder[1,2]. For example, CELP coders running at 4800 bits per second (bps) have been reported to produce fairly good synthetic speech[3]. At the present time, one of the major objectives in speech coding is to pull the bit rate further down (for example, form 4.8 Kbps to 2.4 Kbps) while maintaining speech quality. To achieve the goal, efficient representation of the spectrum information is essential.

Many of current low bit rate CELP coders use scalar quantization scheme to represent the spectrum information parameters[3,4]. For example, in federal standard 1016 (US Department of Defense 4.8 Kbps standard) CELP coder, the 10 Line Spectrum Pair(LSP)parameters derived from Linear Predictive Coding(LPC) coefficients are scalar quantized. However, when the bit rate is further pulled down, the luxury of using a simple scalar quantization scheme can not be held any longer. Consequently, a vector quantization scheme is needed for representing spectrum information parameters. In fact, the scheme has been widely used for many systems recently[5,6, 7]. Particularly, many vector quantization methods for LSP parameters are reported.

This paper focusus on how to represent spectrum information efficiently in a very low bit rate CELP speech coder environment. More precisely, vector quantization method using LPC cepstral coefficients is studied. Cepstral coefficients possess desirable characteristics for vector quantization coding. For example, the average of the sum of the cepstral coefficient is equivalent to an averaging of the corresponding log spectra in the frequency domain. In this case, the centroids for the vector quantizer code vectors can be computed simply by averaging all the cepstral vectors in their clusters and can be interpreted in the frequency domain with spectral meaning. To represent the spectrum information using cepstral coefficients, three different cepstral distance measures which have different spectral meaning in the frequency domain are considered, and their performances are compared meanings in the freperformances are compared and analyzed. The results show that spectrum information in a low bit rate CELP coder environment can be represented very efficiently using a cepstral vector quantization scheme.

Section Ⅱ briefly describes the basic structure of the CELP speech coder. Section Ⅲ introduces the LPC cepstral coefficients which are used to represent the spectrum information in this paper. Section IV focuses on how to vector quantize (VQ) the LPC cepstral coefficients. Three different cepstral distance measures considered in the VQ codebook designing procedure will be discussed. Is Section V, the performance results of the proposed cepstral VQ methods are reported and analyzed. Finally, the conclusions of this paper are given in Section VI.

Ⅱ. Basic Structure of a CELP Coder

In the LPC model, a linear filter described by the linear prediction coefficients is used to model the characteristics of the vocal tract. The linear filter is then excited by the excitation signal to make a synthesized signal. In 1982, a method for determining the excitation, called the analysis-by-synthesis excitation algorithm, was proposed and applied to LPC vocal tract model by Atal and Remde[8,9]. In this excitation algorithm, different possible excitations are tried as inputs to the vocal tract model. Then, the best excitaion signal is determined by minimizing a weighted mean-squared error between the synthesized speech and the original. The weighting function is chosen to emphasize factors that are known to be perceptually important. Since the hearing system is less capable of perceiving errors in the frequency bands where the energy is high such as in form-

ant regions[10,11], coding noise energy should be distributed proportionally to the spectral envelope. In general, the weighted mean-squared error E has a form of

$$E = \sum_{n=0}^{N-1} \left((s[n]-h[n]*e[n])*w[n]\right)^2 \quad (1)$$

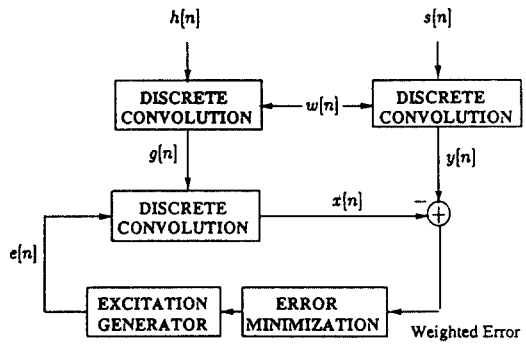
or

$$E = \frac{1}{2\pi}\int_{-\pi}^{\pi} |S(e^{j\omega})-H(e^{j\omega})E(e^{j\omega})|^2\,|W(e^{j\omega})|^2 d\omega, \quad (2)$$

where h[n] and w[n] are the vocal tract impulse response and spectral weighting filter, respectively, $H(e^{j\omega})$ and $W(e^{j\omega})$ are the Fourier transforms of h[n] and w[n], respectively, and N is the excitation frame length. The notations s[n] and e[n] represent speech and excitation sequences, respectively. Eq. (1) can be written as

$$E = \sum_{n=0}^{N-1} (s[n]*w[n]-h[n]*w[n]*e[n])^2, \quad (3)$$

$$E = \sum_{n=0}^{N-1} (y[n]-g[n]*e[n])^2 \quad (4)$$



$s[n]$: pre-emphasized speech
$w[n]$: perceptual weighting filter impulse response
$y[n]$: "weighted" speech
$h[n]$: vocal tract impulse response
$g[n]$: "weighted" vocal tract impulse response
$x[n]$: "weighted" synthetic speech
$e[n]$: excitation signal

Figure 1 : Analysis-by-Synthesis method for obtaining the excitation sequence $e[n]$.

where y[n] = s[n]*w[n] and g[n] = h[n]*w[n]. The signal y[n] is called the weighted speech signal, and g[n] is called the weighted impulse response. Hence, in practice, the excitation signal e[n] is obtained from the weightd speech signal y[n] and the weighted impulse response g[n]. The analysis-by-synthesis excitation coding algorithm is depicted in Figure.1

Different types of excitation models have been studied for the LPC vocoder using the analysis-by-synthesis excitation analysis. In particular, multi-pulse[8,9] and code-excited excitation models[1,2] have been developed successfully. In the case of the code-excited excitation model, i.e., in the CELP model, the excitation signal e[n] is composed of the following two parts: $\beta_0 e[n-\gamma_0]$, which represents a short segment of the past (previously computed) excitation beginning γ_0 samples before the present excitation frame, and $\beta_1 f_{\gamma_1}[n]$, where $f_{\gamma_1}[n]$ is a zero-mean Gaussian codebook sequence corresponding to index γ_1 in the codebook, i.e.,

$$e[n] = \beta_0 e[n-\gamma_0] + \beta_1 f_{\gamma_1}[n] \quad (5)$$

In Eq. (5) β_0 and β_1 are the gain terms, called the self-excitation and codebook gains, respectively. In a given excitation frame, the perceptually weighted synthetic speech x[n] corresponding to e[n] has the form

$$x[n] = \beta_0 x_0[n] + \beta_1 x_1[n] + x_\gamma[n], \quad (6)$$

where $x_0[n] = g[n]*e[n-\gamma_0]$, $x_1[n] = g[n]*f_{\gamma_1}[n]$, $g[n] = w[n]*h[n]$ is the perceptually weighted impulse response, and $x_\gamma[n]$ is the contribution to x[n] from the filter memory.

The optimum excitation parameters, β_0, γ_0, β_1 and γ_1 are determined pairwise. The excitation generator first searches through a fixed interval of the past coded excitation signal looking for an

excitation sequence that minimizes the weighted mean-squared error. Once the best excitation sequence is derived from the history of the excitation signal, the location γ_0 and the scale factor β_0 of the selected excitation signal are stored. A new signal is formed by subtracting the synthetic speech just determined form the speech signal. The excitation generator then searches through the Gaussian codebook looking for an excitation sequence which produces the most highly correlated signal with the residual signal, and the corresponding optimum scale factor β_1 and the index γ_1 of the selected codeword are stored.

Ⅲ. Rrpresentation of the Spectrum Information

In the CELP coder, prior to the analysis-by-synthesis excitation analysis, LPC analysis[12] is performed to obtain the LPC coefficients which contain the time-varying vocal tract information (or, spectrum information). The speech signal $s_0[n]$ is pre-emphasized prior to the LPC analysis, i.e.,

$$s[n] = s_0[n] - \alpha s_0[n-1] \tag{7}$$

or

$$S(z) = (1-\alpha z^{-1})S_0(z) \tag{8}$$

where $S_0(z)$ and $S(z)$ are the z-transforms of the input signal $s_0[n]$ and pre-emphasized signal $s[n]$, respectively. The pre-emphasis allows better modeling of the lower-amplitude, higher-frequency formants. At the final stage of synthesis, the corresponding de-emphasis filter $1/(1-\beta z^{-1})$ is included, and a constant 0.5 is used for α and β in our systems. Each frame of the preemphasized speech signal $s[n]$ is weighted by a Hamming window $v[n]$, and then the corresponding LPC coefficients $a[n]$ are computed by LPC analysis.

Figure 2 shows how the vocal tract and excitation frames are organized with respect to the Hamming window in our system. The vocal tract

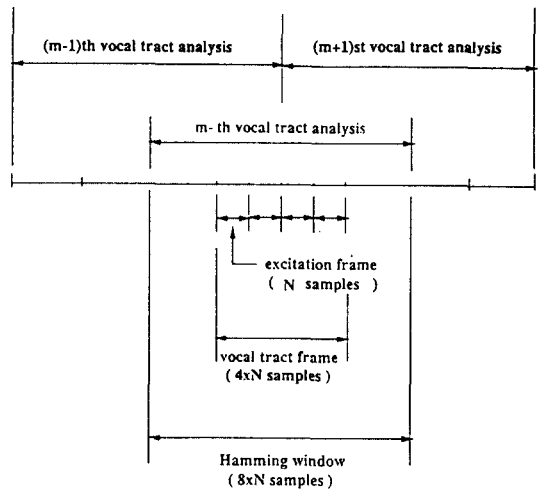


Figure 2 : Organization of the vocal tract and excitation frames with respect to the Hamming window.

frame is moved by half the length of the vocal tract analsis window (the Hamming window), and there are 4 excitation frames per vocal tract frame.

Once the LPC coefficients are obtained, the LPC coefficients are transformed into the form of LPC cepstral coefficients. Thus, in our system, the LPC cepstral coefficients are used to communicate the spectrum information between the transmitter and receiver. The LPC cepstral coefficients $c[n]$ can be computed recursively from the p the order LPC coefficients $a[n]$ using the relationships [12] that

$$c[1] = -a[1], \tag{9}$$

$$c[n] = -a[n] - \sum_{k=1}^{n-1} (1-k/n)a[k]c[n-k], \quad 1<n\leq p. \tag{10}$$

In our systems, the value p (i.e., the LPC order) was set as 11. The zeroeth cepstral coefficient $c[0]$ of each input was set to zero to maintain the corresponding impulse response in normalized form. Figure 3 shows the objective performance

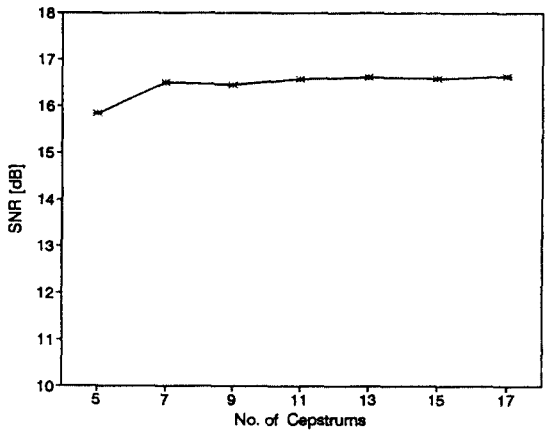


Figure 3 : Objective performance of CELP coder with different number of cepsturms.

of CELP coder described in Section Ⅱ with different number of cepstrums. The parameters used for the experiment are summarized in Table 1. In this experiment, the selfexcitation parameter β_0 and gain parameter β_1 were not quantized. For the experimet, three speech sentences spoken by a female speaker were used. The number of cepstrums were increased from 5 to 17 by amount of 2. The performance of the coder was proportional to the number of used cepstrums. When the used cepstrums were more than 11, however, the increment in coder's performance was not noticeable. On the other hand, when the used cepstrums were 5 or 7, the synthesized speeches were less natural, and distortions were perceived in hearing.

In the speech coding literature, little has been reported about efficient quantization schemes for the set of LPC cepstrum values containing vocal information. In the next section, effective vector quantization schemes of a sequence of 11 LPC cepstral coefficients are studied. In particular, three cepstral distance measures having different desirable spectral meanings in the frequency domain are considered as distortion measures, and performances are analyzed.

Table 1 : Parameters used for the experiment performed in Section Ⅲ

vocal tract frame length	100 samples
excitation frame length	25 samples
self-excitation search range	128 samples
Gaussian codebook size	1024 samples

Ⅳ. LPC Cepstral VQ Codebook Design

To code the cepstrums, a vector quantization scheme was used. To design a quantizer, first, 40 different sentences, obtained from one female speaker were processed generating LPC cepstral training vectors. The LBG algorithm was then used to design a full search 11 dimensional codebook from this training data[13]. (The zeroeth cepstral coefficient of each input were set to zero to maintain the corresponding impulse response in normalized form) The maxmindistance algorithm was used to provide an initial codebook for the LBG algorithm[14].

In the procedure of designing the cepstral VQ codebooks, three different distortion measures were studied, namly, the Euclidean cepstral distance measure d_{ECEP}, the quefrency weighte cepstral distance measure d_{QCEP} and the variance-equalized cepstral distance measure d_{VCEP}[15,16]. The Euclidean cepstral distance measure d_{ECEP} is defined as

$$d_{ECEP}(c, \tilde{c}) = \sum_{i=1}^{11} |c[i]-\tilde{c}[i]|^2 \qquad (11)$$

where c and $\tilde{c}$ are the vectors of original and reproduced LPC cepstrum values. Using the Parseval's theorem, d_{ECEP} can be interpreted as the mean-squared difference between log spectra in the frequency domain, i.e.,

$$d_{ECEP}(c, \tilde{c}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |\log S_c(\omega) - \log S_{\tilde{c}}(\omega)|^2 d\omega, \quad (12)$$

where $\log S_c(\omega)$ and $\log S_{\tilde{c}}(\omega)$ are the log spectra

correspinding to c and $\tilde{c}$, respectively. Hence, d_{EC} $_{EP}$ is an efficient measure to estimate the log spectral distance.

The quefrency weighted cepstral distance measure d_{QECP} is defined as

$$d_{QCEP}(c, \tilde{c}) = \sum_{i=1}^{11} |ic[i]-i\tilde{c}[i]|^2 \quad (12)$$

$$= \sum_{i=1}^{11} i^2|c[i]-\tilde{c}[i]|^2 \quad (13)$$

where c and $\tilde{c}$ are the vectors of original and reproduced LPC cepstrum values. The motivation for examining d_{QCEP} was to reproduce higher cepstral coefficients more faithfully, since the higher cepstral coefficients have small magnitudes relative to lower cepstral coefficients. Using the equality

$$\frac{\partial \log S_c(\omega)}{\partial \omega} = -j\sum_n nc[n]e^{-j\omega n}, \quad (14)$$

the quefrency weighted cepstral distance measure d_{QCEP} can be written as

$$d_{QCEP}(c, \tilde{c}) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \left|\frac{\partial \log S_c(\omega)}{\partial \omega} - \frac{\partial \log S_{\tilde{c}}(\omega)}{\partial \omega}\right|^2 d\omega. \quad (15)$$

Consequently, d_{QCEP} is equivalent to the mean-squared difference between the slopes of log spectra in the frequency domain.

The variance-equalized cepstral distance measure d_{VCEP} is defined as

$$d_{VCEP}(c, \tilde{c}) = \sum_{i=1}^{11} w_i|c[i]-\tilde{c}[i]|^2, \; w_i = \frac{\sigma_1^2}{\sigma_i^2} \quad (16)$$

where c and $\tilde{c}$ are the vectors of original and reproduced LPC cepstrum values, and σ_i^2 is the variance of the ith cepstral coefficient of the training sequences used for designing the codebook. Tohkura showed that the variance of the higher cepstral coefficients is much smaller than the variance of the lower cepstral coefficients [15]. This suggests that some cepstral coeffieients contain more information about the vocal tract than others, and thus those cepstral coefficients should be weighted appropriately. The d_{QCEP} weights the cepstral coefficients based upon their relative importance.

When the weighted distortion measures were used, the cepstral vectors were weighted prior to designing a codebook. Consequently, if a weighted cepstral distance measure was used, a set of cepstrum values to be encoded was also weighted prior to selecting the best codeword from the codebook. Conversely, when the best codeword was chosen, the codeword was deweighted to produce the reproduction vector. The weighting factor used for d_{QCEP} and the variance equalizer $\sqrt{w_i}$ used for d_{VCEP} are plotted in Figure 4.

Three sets of cepstral VQ codebooks were designed. Each set of the codebooks was composed of 15 different codebooks. Among the 15 codebooks, the first five were designed using d_{ECEP} as a distortion measure. The sizes of the five

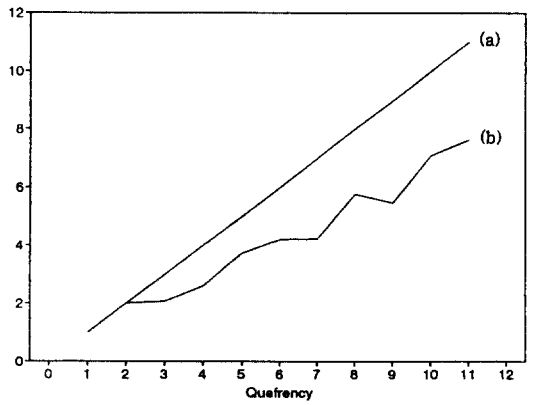


Figure 4 : Weighting factors : (a) quefrency used for d_{QCEP}, and (b) variance equalizer used for d_{VECP}.

VQ codebooks were varied as 64, 128, 256, 512, and 1024. Then, the next five codebooks were designed using d_{QCEP} as a cepstral distance measure. The sizes were also 64, 128, 256, 512, and 1024. Finally, the rest of the codebooks (having the sizes of 64, 128, 256, 512, and 1024) were designed using d_{VCEP} as a VQ measure. Hence, a total of 45 cepstral VQ codebooks were designed . To design the first 15 cepstral VQ codebooks, for speech sampled at 8 KHz, a frame interval of 12.5 ms(i.e., 100 samples)was used as the vocal tract frame length. Similarly, to design the second set, a frame interval of 15.0 ms(i.e., 120 samples) was used. Table 2 summarizes number of training vectors used for the VQ codebook design.

When a 1024 size codebook in the first set is used to vector quantize the LPC cepstra, 800 bps are required to transmit the spectrum information. Similarly, when a 64 size codebook in the third set is used, 300 bits are needed for every second. Bits required for every second to represent the spectrum information versus cepstral VQ codebook sizes are summarized in Tabel 3.

Figure 5 shows the average distortion of d_{ECEP} as a function of the codebook size. The codebook

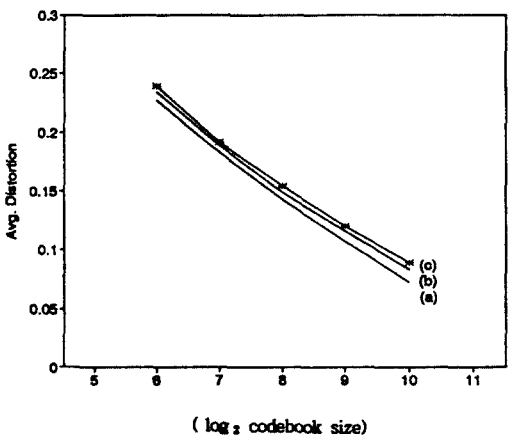


Figure 5 : Average distortions of d_{ECEP} for (a) 20.0 ms, (b) 15.0 ms, and (c) 12.5 ms vocal tract analysis updates.

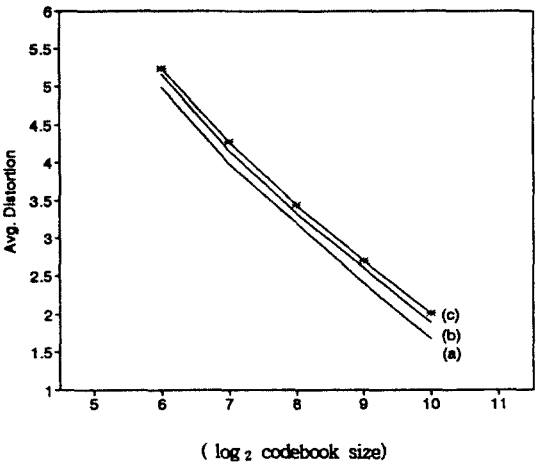


Figure 6 : Average distortions of d_{QCEP} for (a) 20.0 (b) 15.0 ms, and (c) 12.5 ms vocal ttact analysis updates.

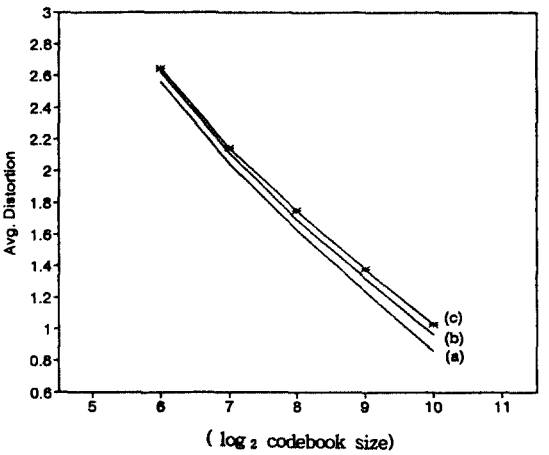


Figure 7 : Average distortions of d_{VCEP} for (a) 20.0 ms, (b) 15.0 ms, and (c) 12.5 ms vocal tract analysis updates.

size was increased from 64 to 1024 by factors of 2. In Figure 5, lines (a), (b), and (c) correspond to the average distortions of d_{ECEP} according to the vocal tract frame intervals. Similarly, Figures 6 and 7 show the average distortions of d_{QCEP} and d_{VCEP} measures. We observe from Figures 5, 6, and 7 that the average distortions of d_{QCEP} and

d_{VCEP} are much greater than the average distortion of d_{ECEP}. This is due to the weighting of the cepstral training vectors prior to designing a codebook when d_{QCEP} or d_{VCEP} is used. Thus, direct comparison of the average distortions among the three measure is not appropriate. Rather, the performances of the designed cepstral VQ codebooks should be analyzed thrugh objective and subjective tests. The next section focusus on that issue.

Table 2 : Number of training vectors used for the VQ codebook design.

	Training vectors used
VQ codebooks in the first set	9,190 vectors
VQ codebooks in the second set	7,651 vectors
VQ codebooks in the third set	5,716 vectors

Table 3 : Bits per second required to represent spectrum information.

		Vocal Tract Frame Interval		
		12.5ms	15.0ms	20.0ms
Cepstral VQ Codebook Size	64	480	400	300
	128	560	467	350
	256	640	533	400
	512	720	600	450
	1024	800	667	500

V. Simulation Results

To analyze the perform of the cepstral VQ codebooks designed in the previous section, several experiments were performed.

First, the LPC cepstral VQ codebooks designed in the previous sections are denoted as

$$C_{D,S,L} \tag{17}$$

where, D indicates the type of distortion measure used, S represents the size of the codebook, and L is the vocal tract frame interval (in me) of the training vectors uned in the codebook design procedure. For example, the notation

$$C_{d_{ECEP},1024,12.5} \tag{18}$$

denotes a LPC cepstral VQ codebook : (i) whose distortion measure is d_{ECEP}, (ii) which has 1024 entries, and (iii) whose vocal tract frame interval was 12.5 ms.

5.1 Experiment I

In the first experiment, for speech sampled at 8 KHz, a frame interval of 12.5 ms (i.e., 100 samples) was used as the vocal tract frame interval. (Consequently, the excitation analysis frame length was 25 samples.) Once the LPC coefficients were obtaind using the LPC vocal tract analysis, the LPC coefficients were converted to LPC cepstral coefficients. Then, the cepstral coefficients were vector quantized using the LPC cepstral VQ codebooks in the first set designed in the previous section. To measure the degradation produced in cepstral VQ quantization, we took the following sequential procedures. First, convert the *quantized* LPC cepstral coefficients into (quantized) impulse response $h[n]$. By applying the analysis-by-synthesis excitaion algorithm depicted in Figure 1, obtain the excitation sequence $e[n]$ from the speech signal $s[n]$ and impulse response $h[n]$. Finally, synthesize the speech signal $x[n]$ by filtering the excitation sequence $e[n]$ through $h[n]$. In our system, the LPC cepstral coefficients were not interpolated in the procedure of generating the excitation signal. In the speech synthesis procedure, the excitation gain parameters, i.e., β_0 and β_1 in Eq. (5), were not quantized to concentrate on the distortion solely caused from the cepstrum quantization. Also, to minimize the distortion due to the restricted size of the Gaussian codebook in excitation sequence searching, the 1024 size Gaussian codebook was used. Thus, the range of the Gaussian codebook sequence index, γ_1 in Eqn. (5) is 1024. The range of the past excitation history used for the position parameter γ_0

was from 32 to 159.

The SNRs of the synthesized speech sentences for different LPC cepstral VQ codebooks are summarized in Table 4. For the experiment, three speech sentences not included in the VQ codebook training sequences were used. The performances of the d_{ECEP} or d_{VCEP} distortion measures turned out to be superior to that of d_{QCEP}. These results were also verified through informal listening tests. The reconstructed speech quality using d_{ECEP} and d_{VCEP} measures was approximately same. One thing to notice from Table 4 is that the SNRs do not exactly reflect the codebook size and its synthesized speech quality. However, the distortion in the synthetic speech related to the codebook size was clearly perceived in the informal listening tests. In other words, the perceived distortion increased as the VQ codebook size decreased in listening test. Overall, however, quality of the synthesized speeches whose spectrum information paramenters were quantized using either $C_{dECEP.S,\ 12.5}$ or $C_{dVCEP.\ S,\ 12.5}$ were very good. Particularly, the performances of $C_{dECEP.\ 1024,\ 12.5}$, $C_{dECEP.\ 512,\ 12.5}$, $C_{dVCEP.\ 1024,\ 12.5}$ and $C_{dVCEP.\ 512,\ 12.5}$ were excellent.

Table 4 : SNRs of the synthesized speech sentences for the Experiment I.

$C_{dECEP.\ 64,\ 12.5}$ 15.51	$C_{dQCEP.\ 64,\ 12.5}$ 14.70	$C_{dVCEP.\ 64,\ 12.5}$ 15.68
$C_{dECEP.\ 128,\ 12.5}$ 15.37	$C_{dQCEP.\ 128,\ 12.5}$ 14.53	$C_{dVCEP.\ 128,\ 12.5}$ 15.96
$C_{dECEP.\ 256,\ 12.5}$ 15.95	$C_{dQCEP.\ 256,\ 12.5}$ 14.89	$C_{dVCEP.\ 256,\ 12.5}$ 15.61
$C_{dECEP.\ 512,\ 12.5}$ 15.62	$C_{dQCEP.\ 512,\ 12.5}$ 14.73	$C_{dVCEP.\ 512,\ 12.5}$ 15.57
$C_{dECEP.\ 1024,\ 12.5}$ 15.38	$C_{dQCEP.\ 1024,\ 12.5}$ 14.78	$C_{dVCEP.\ 1024,\ 12.5}$ 15.50

5.2 Experiment II

The procedures of Experiment II were indentical with the procedures of Experiment I, except the length of the vocal tract frame interval used. In Experiment II, the vocal tract analysis was performed for 15.0 ms frame interval (i.e., for every 120 speech samples). The results of Experiment II are summarized in Table 5. Similar to the results of Experiment I, we observed the following. The performances of d_{ECEP} and d_{VCEP} were about the same, and outperformed the ones of d_{QCEP}. Although the SNRs did not reflect quality degradation exactly against the codebook size, the perceived distortion increased as the VQ codebook size decreased in listening. Overall, the quality of the synthesized speeches whose spectrum information parameters were quantized using either $C_{dECEP.S,\ 15.0}$ or $C_{dVCEP.S,\ 15.0}$ were fairly good.

Table 5 : SNRs of the synthesized speech sentences for the Experiment II.

$C_{dECEP.\ 64,\ 15.0}$ 14.43	$C_{dQCEP.\ 64,\ 15.0}$ 14.79	$C_{dVCEP.\ 64,\ 15.0}$ 13.44
$C_{dECEP.\ 128,\ 15.0}$ 14.81	$C_{dQCEP.\ 128,\ 15.0}$ 14.74	$C_{dVCEP.\ 128,\ 15.0}$ 13.30
$C_{dECEP.\ 256,\ 15.0}$ 14.92	$C_{dQCEP.\ 256,\ 15.0}$ 15.12	$C_{dVCEP.\ 256,\ 15.0}$ 13.39
$C_{dECEP.\ 512,\ 15.0}$ 15.15	$C_{dQCEP.\ 512,\ 15.0}$ 14.71	$C_{dVCEP.\ 512,\ 15.0}$ 13.22
$C_{dECEP.\ 1024,\ 15.0}$ 14.62	$C_{dQCEP.\ 1024,\ 15.0}$ 14.48	$C_{dVCEP.\ 1024,\ 15.0}$ 13.42

5.3 Experiment III

In Experiment III, an interval length of 20 ms (160 speech samples) was used for the vocal tract frame. Thus, in this experiment, the length of an excitation sequence was 40 samples. When the procedures described in Experiment I were applied, we obtained the results shown in Table 6. From the results, we could derive similar conclusions from those made in Experiments I and II. In other words, codebooks of $C_{dECEP.S,\ 20.0}$ and $C_{dVCEP.\ S,\ 20.0}$ performed about same and better than those of $C_{dQCEP.\ S,\ 20.0}$ The overall quality of the synthesized speeches was fair.

Table 6 : SNRs of the synthesized speech sentences for the Experiment III.

$C_{dECEP.\ 64,\ 15.0}$ 12.48	$C_{dQCEP.\ 64,\ 15.0}$ 11.01	$C_{dVCEP.\ 64,\ 15.0}$ 12.40
$C_{dECEP.\ 128,\ 15.0}$ 12.64	$C_{dQCEP.\ 128,\ 15.0}$ 10.96	$C_{dVCEP.\ 128,\ 15.0}$ 12.99
$C_{dECEP.\ 256,\ 15.0}$ 12.91	$C_{dQCEP.\ 256,\ 15.0}$ 11.29	$C_{dVCEP.\ 256,\ 15.0}$ 12.76
$C_{dECEP.\ 512,\ 15.0}$ 12.83	$C_{dQCEP.\ 512,\ 15.0}$ 11.25	$C_{dVCEP.\ 512,\ 15.0}$ 12.98
$C_{dECEP.\ 1024,\ 15.0}$ 12.87	$C_{dQCEP.\ 1024,\ 15.0}$ 11.11	$C_{dVCEP.\ 1024,\ 15.0}$ 12.86

Ⅵ. Conclusions

In this paper, an efficient way of representing the spectrum information in a very low bit rate CELP coder environmont was studied. To achieve this goal, a vector quantization scheme using LPC cepstral coefficents was proposed. To represent the spectrum ifformation using cepstrums, three cepstral distance measures having different desirable spectral meanings in the frequency domain were considered, and performances were compared and anlyzed. The distortion measures considered were the Euclidean cepstral distance measure, quefrency weighted cepstral distance meansure, and variance equalized cepstral distance measure. Three different sets of experiments were performed to analyze the performance of the LPC cepstral VQ method. The experimental results show that spectrum information in al low bit rate CELP coder evvironment can be represented very efficiently using the LPC cepstral VQ method. In particular, two cepstral distortion measures, namely Euclidean and variance equalized cepstral distance measures, turned out to be good distortion measures suitable for LPC cepstral vector quantization.

References

1. M. R. Schroeder and B. S. Atal, "Code-excited linear prediction (CELP) : High-quality speech at very low bit rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp937-940, 1985.
2. B. S. Atal, "High-quality speech at low bit rates : Multi-pulse and stochastically excited linear predictive coders," *Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp.1681-1684, 1986.
3. U.S. Dept. of Defense, "The DoD 4.8 kbps standard (Proposed federal standard 1016), Third draft," August 1990.
4. I. A. Gerson and M. A. Jasiuk "Vector sum excited linear prediction (VSELP)," *Advances in Speech Coding,* pp.69-79. Kluwer Academic Publishers, norwell, MA, December 1990.
5. K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC paramenters at 24 bits/frame," *IEEE Trans. on Speech and Audio Processing,* vol.1. no.1, pp.3-14, Jan. 1993.
6. J. Pan and T. R. Fisher, "Vector quantization of speech LSP parameters using trellis codes and l_1-norm constraints," *Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp. Ⅱ.17-Ⅱ.20, 1993.
7. J. M. Lopez-Soler and N. Farvardin, "A combined quantization interpolation scheme for very low bit rate coding of speech LSP parameters," *Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp. Ⅱ.21-Ⅱ.24, 1993.
8. B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at low bit rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing,* pp614.-617. 1982.
9. B. S. Atal, "New direction in speech coding at low bit rates", *Proc. GLOBECOM,* pp.1083-1086, 1982.
10. M. R. Schroeder, B. S. Atal, and J. Hall, "Optimizing digital speech coders by exploiting masking porperties of the human ear," *Journal Acoust., Soc. Amer.,* vol.66, pp. 1647-1652, Dec. 1979.
11. B. S. Atal and M. R. Schroeder, "Predictive coding of speech and subjective error criteria," *IEEE Trans. Acoust., Speech, and Signal Proc.,* vol.ASSP-27, pp.247-254, June 1979.
12. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals,* Englewood Cliffs, N.J., Prentic-Hall, 1978.
13. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design" *IEEE Trans. Commumications,* vol.COM-28, pp.84-95, Jan. 1980.
14. J. T. Tou and R. C. Gonzalez, *Pattern Recog-*

nition Principles, Addison-Wesley, 1974.

15. Y. Tohkura, "A weighted cepstral distance measure for speech recognition," *IEEE Trans. Acoust., Speech, Signal Processing,* vol.ASSP-35, pp.1414-1422, Oct. 1987.

16. L. R. Rabiner and F. Soong, "Single frame vowel recognition using vector quantization with several distance measures," *AT&T Technical Journal,* vol.64, pp.2319-2330, 1985.

본 연구는 92년도 인하대학교 연구비 지원에 의하여 수행되었음.

鄭 在 皓(Jae Ho Jeong) 정회원

1992년 : 美國 University of Maryland(학사)

1994년 : 美國 University of Maryland(석사)

1990년 : 美國 Georgia Institute of Technology(박사)

1984년~1985년 : 美國 Naval Surface Warfare Center, Electronic Engr.

1991년~1992년 : 美國 AT&T Bell Laboratories, 연구원

1992년~현재 : 인하대학교 공과대학 전자공학과, 조교수