OPTIMIZING GAIN CODEBOOK OF LD-CELP *

Gang Zhang Keming Xie Xueying Zhang Liying Huangfu

Taiyuan University of Technology, 030024, Taiyuan Shanxi, China
E-mail: zhgang. @163.net

ABSTRACT

This paper combines the Jayant adaptive quantization and G.728's gain prediction to propose a scheme fit for designing gain coodbook of LD-CELP, in which the exact gain value was adaptively predicted and adaptively quantized. In this method, for adaptive quantization gain of 3 bit and 4 bit, the SNR have been increased by 0.5 and 6 dB separately than that of G.728. If adopting 4-bit adaptive quantization and the shape codebook of 64 codevectors, the complexity of G.728 could be reduced by 20% while SNR had 1 dB improvement than G.728. When the order of synthesis filter is reduced from 50 to 30, SNR isn't changed and the complexity of G.728 is decreased by 40%.

1. INTRODUCTION

At present, among the toll quality speech coding algorithm with 2ms delay, the ITU G.728 has the lowest code rating. The G.728's structure of gain-shape codebook makes its computing complex. In theory, the computation of the codebook search can be reduced to 1/8 by using the normal shape codebook and scalar quantization gain. An increase of 0.5 dB has been obtained by adjusting the gain offset to G.728 [1][8]. The Jayant adaptive quantization is one of the candidates of G.728 gain filter schemes[2]. This paper presented a method of optimizing gain codebook of LD-CELP by combining adaptive prediction with Jayant adaptive quantization[4]. The better results have been obtained when it was used in G.728. The normal shape codebook and exact denoting of gain are given in section 2. The adaptive gain prediction and adaptive quantization are discussed in section 3 and 4, respectively. The experiment results are discussed in the end.

2. NORMAL SHAPE CODEBOOK WITH EXACT GAIN (NSCEG)

The G.728's codebook search is performed according to the formula (2-1)

$$D_{\min} = \sigma^2(n)\left\|\hat{\boldsymbol{x}}(n) - g_i \boldsymbol{H}(n) g_y \overline{\boldsymbol{y}}_j\right\|^2 \qquad (2\text{-}1)$$

Where, $\boldsymbol{H}(n)$ is the unit impulse respond of short-term filter, g_i is the gain codeword, g_y is the energy of shape codevector and $\overline{\boldsymbol{y}}_j$ is the normalized shape codevector. Let $G_j(n)$ be the energy of excitation vector and $\sigma(n)$ be the estimation of current $G_j(n)$ obtained by gain predicting, then $G_j(n) = \sigma(n) g_j$, and $\boldsymbol{x}(n) = \sigma(n)\hat{\boldsymbol{x}}(n)$ is the target vector, where $\hat{\boldsymbol{x}}(n)$ is the target vector adjusted by $\sigma(n)$. When the energy of shape codebook is normalized (g_y =1) and $\overline{\boldsymbol{y}}_j$ is still denoted y_j, equation (2-1) can be written as

$$D_{\min} = \left\|\boldsymbol{x}(n) - G_j(n)\boldsymbol{H}(n)\boldsymbol{y}_j\right\|^2 \qquad (2\text{-}2)$$

Since minimizing $D_{\min}$ is equivalent to maximizing

$$\widehat{D}_{\max} = 2 g_i \boldsymbol{P}^T(n)\boldsymbol{y}_j - g_i^2 E_j \qquad (2\text{-}3)$$

Where, $\boldsymbol{P}(n) = \boldsymbol{H}^T \hat{\boldsymbol{x}}(n)$ and $E_j = \left\|\boldsymbol{H}\boldsymbol{y}_j\right\|^2$.

For every y_j in equation (2-3) let $\partial\widehat{D}_{\max} / \partial g_j = 0$, then exact gain can be expressed as following

$$g_j = [\boldsymbol{P}^T(n)\boldsymbol{y}_j] / E_j \qquad (2\text{-}4)$$

Suppose $\hat{G}_j(n)$ and $\hat{g}_j(n)$ be the quantization values of $G_j(n)$ and $g_j(n)$, respectively, then $\hat{G}_j(n) = \sigma(n)\ \hat{g}_j(n)$. Quantizing to $G_j(n)$ is equivalent to quantizing $g_j(n)$.

* The research is supported by Chinese National Nature Science Fund (60172070), Shanxi Province Nature Science Fund (991040, 20011033) and Shanxi Province Science Research Fund for researchers returned to China (200224).

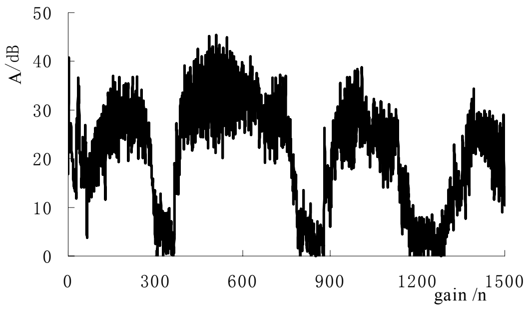

Fig.1 The gain waveform

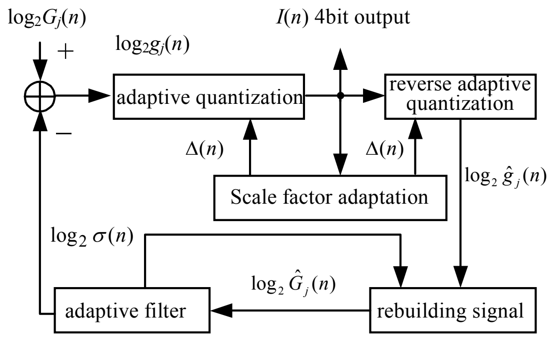

Fig. 2 Encoder block scheme

3. ADAPTIVE GAIN PREDICTOR

Fig. 1 shows a segment waveform of excitation gain. The horizontally axis is time which unit is gain sample (each of them corresponding to five speech samples). The vertical axis is the logarithm value of gain amplitude (dB). Let $\log[G_j(n)]$ be the common logarithm RMS (Root-Mean-Square) of the nth excitation vector, and $\beta(n)$ be the mean of $\log[G_j(n)]$. From Fig.1, it can be seen that the mean oscillates between 0 and 45. The G.728's performance of gain predictor can be enhanced by estimating $\beta(n)$ effectively[1]. The new excitation gain adaptive predictor can be presented as following

$$\log\sigma(n) = \sum_{i=1}^{p} p_i[\log[\hat{G}_j(n-i)] - \beta(n-i)] + \beta(n) \qquad (3\text{-}1)$$

4. GAIN ADAPTIVE QUANTIZATION (GAQ)

4.1. Principles

Fig.2 is the GAQ block scheme. The prediction residual error of $G_j(n)$ is $\log_2 g_j(n) = \log_2 G_j(n) - \log_2 \sigma(n)$ in logarithm domain. The 4-bits index I(n) obtained by quantizing $\log_2 g_j(n)$ will be sent to decoder. At the same time, I(n) is decoded in local to obtain the difference signal $\log_2 \hat{g}_j(n)$ which is added with the gain estimation value $\log_2 \sigma(n)$ to obtain the local rebuild signal $\log_2 \hat{G}_j(n)$. The adaptive predictor is a 12 order AR model, and the quantization step $\Delta(n)$ is controlled adaptively by the input signal. There are three differences from the G.721 standard

* The input signal needn't be transferred to logarithm domain because it has been done before.
* The predictor is a 12-order AR model with adaptive coefficients rather than the G.721's ARMA model with fix coefficients.
* The low speed control is removed from the scaling factor module of quantization, because of no in-band data.

4.2. The optimum quantization parameter

The best result of quantization to $\log_2 g_j(n)$ can be obtained by selecting a set of optimum parameters including ξ_i, η_i and M_i (or W_i) (i=0,1,2,3,4,5,6,7). The distribution $p(x)$ of $\log_2 g_j(n)$ ($-\infty < \log_2 g_j(n) < +\infty$) is shown in Fig. 3. Without considering sign bit, $\log_2 \hat{g}_j(n)$ is only quantizated one of the eight discrete values η_i (i=0,1,2,3,4,5,6,7). The variance of quantization residue is

$$\begin{aligned}\sigma_e^2 &= \sum_{i=0}^{7} E\left[(\eta_i - \log_2 g_j(n))^2\right] \\ &= \sum_{i=0}^{7} \int_{\xi_i}^{\xi_{i+1}} (\eta_i - x)^2 p(x)dx\end{aligned} \qquad (4\text{-}1)$$

The parameters ξ_i and η_i are selected to minimize the σ_e^2. Where $\xi_0 = -\infty$ and $\xi_8 = +\infty$. Let $\partial\sigma_e^2/\partial\xi_i = \partial\sigma_e^2/\partial\eta_i = 0$, we can get

$$\eta_i = [\int_{\xi_i}^{\xi_{i+1}} xp(x)dx] \Big/ [\int_{\xi_i}^{\xi_{i+1}} p(x)dx] \quad (i = 0,1,2,3,4,5,6,7) \qquad (4\text{-}2)$$

$$\xi_i = (\eta_{i-1} + \eta_i)/2 \qquad (i = 1,2,3,4,5,6,7) \qquad (4\text{-}3)$$

The formula (4-2) shows us that the η_i's optimum position locates in the rectangle center between ξ_i and ξ_{i+1}. The formula (4-3) shows that the optimum ξ_i is the average value of η_{i-1} and η_i. Using the iteration method

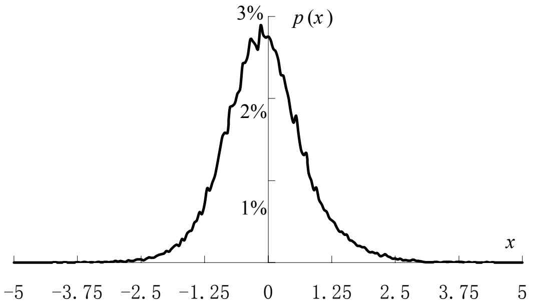

Fig.3 The probability distribution of gain

can find out parameters ξ_i and η_i (i =0,1,2,3,4,5,6,7)[3] as follows. Let Δ be the step of dividing range [ξ_i , ξ_{i+1}]. For each Δ, the number f_i of samples $\log_2 g_j(n)$ is counted and then sample frequency f_i/F (where F is total samples number) can be obtained, which is approximately used as
$p(x)$ in formula (4-1). Let x be the middle value of the step Δ. If Δ is enough small the signal $\log_2 g_j(n)$ is considered as uniform distributing. After {ξ_i }and {η_i } are put initial value we can get the $p(x)$. New {ξ_i } and {η_i } are calculated by formula (4-2) and (4-3). Such iterative calculation to {ξ_i } and {η_i } has been continued until their values are steady.

4.3. The optimum update factor

The adaptive robust multiplying algorithm of quantization step can be described as: $\Delta(n) = M[I(n-1)]\Delta^{\beta}(n-1)$. Where β (0< β <1) is the attenuation factor of increasing the robustness of quantizator, and M is the step update factor. The smaller the signal's probability density the bigger M should be selected. M is the function of last index I (n-1), so we call it 'one-word store adaptive quantization' [3][4]. For getting stable SNR as high as possible, update factor should satisfy $\prod_{i=0}^{7} M_i^{p_i} = 1$, where p_i(i=0,1,2,3,4,5,6,7) is occupancy probability of the ith level when signal energy is kept invariable. During test, when a set of the optimum quantization parameter η_i and ξ_i (i=0,1,2,3,4,5,6,7) have been obtained, the occupancy probability of the ith level could be computed easily. Then we select optimum update factors M_i according to the average segment SNR of quantization gain signal. And a set of the optimum discrete parameters $W[I(k)]$ can be gotten with $W[I(k)] = 2^5 \log_2 M[I(k-1)]$.

5. EXPERIMENT RESULT AND DISCUSSION

5.1. Gain adaptive quantization

The four kinds of gain quantization scheme are performed. Fig. 4 shows us the compare results of them and G.728 quantization. The vertical scale is quantization SNR and horizontal scale is serial number of the sentence. It shows that adaptive quantization is 2dB higher than fix quantization when their quantization bit is same. And 3bit and 4 bit adaptive quantization increased 0.5dB and 6 dB relative to G.728, respectively. The optimum update factors and the optimum quantization parameters of 4bit quantization scheme are listed in Table 1 and Table 2. The average segment SNR of 4bit adaptive gain quantization to g_j is 17.7dB when it is off-line, and 19.45dB on-line.

5.2. Speech coding

After the 128 shape codevectors of G.728 have been normalized we selected the 64 codevetors with the highest use frequency as the shape codebook of new algorithm by training the 61,000 vectors of 30 sentences. During codebook search, for each of input speech vectors, its target vector was first calculated, and then the 64 exact gain g_j were computed by formula (2-4) corresponding to 64 normal shape codevetors y_j. The 64 gains were quantized separately. The $\hat{g}_j$ corresponding to $\hat{D}_{\max}$ is just best gain quantization value and corresponding y_j is best shape codevectors. In practical program, every sub-optimum quantization gain was also calculated. Then a pair of $\hat{g}_j$ and y_j maximizing $\hat{D}_{\max}$ is selected from 64×2=128 codevectors, which are optimum gain and shape codewords. The sub-optimum quantization gain is defined as follows: Suppose g_j is exact gain and $\hat{g}_{ji}$ is optimum adaptive quantization level, the $\hat{g}_{ji-1}$ and $\hat{g}_{ji+1}$ are neighbor of $\hat{g}_{ji}$. If $\hat{g}_{ji-1} < g_j < \hat{g}_{ji}$, then $\hat{g}_{ji-1}$ is the sub-optimum quantization value, whereas if $\hat{g}_{ji} < g_j < \hat{g}_{ji+1}$ then $\hat{g}_{ji+1}$ is the sub-optimum quantization value. In practical program, the probability selecting sub-optimum is about 3%. Testing the same 30 sentences as above, its average segment SNR is 22.16 dB, which is higher 1 dB than G.728 algorithm. When using other 54 sentences to test this method, the similar result was got.

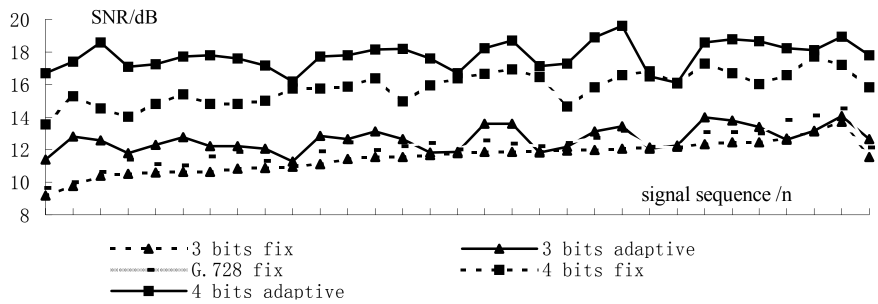

Fig. 4 The compare of five quantization schemes

Table 1 Values of $W(I)$

$\mid I(n) \mid$	7	6	5	4	3	2	1	0
$W(I)$	13.5622	6.59026	-0.84733	0.009999	-0.00847	0.54242	1.53562	1.15441

Table2 Normalizing input/output character of quantization

The input of quantization $\log_2 \mid g_j(n) \mid - \Delta(n)$	$\mid I(n) \mid$	The output of quantization $\log_2 \mid \hat{g}_j(n) \mid - \Delta(n)$
[1.79734, +∞]	7	2.28166
[0.999395, 1.79734]	6	1.31301
[0.448685, 0.999395]	5	0.68578
[0.000695, 0.448685]	4	0.21159
[-0.429295, 0.000695]	3	-0.21020
[-0.909395, 0.429295]	2	-0.64839
[-1.58623,-0.909395]	1	-1.17040
(-∞, -1.58623)	0	-2.00207

5.3. Reducing order experiment

The order of synthesis filter in G.728 is 50 because it performs filtering backwards without long-term pitch prediction. Because the calculation quantity of updating synthesis filter increases non-linearly with its order, the complexity can be decreased heavily if the order of LPC is reduced. The experiments presented that when the order of synthesis filter is 30, SNR is equivalent to that of G.728 and speech quality has not almost change, but the calculation quantity is decreased 6.12 MIPS which is about 24% of total calculation. The G.728's codebook search number is 1024 that is about 40% of total calculation quantity, but new codebook search is 128 that is about 1/8 of G.728. Considering gain adaptive prediction and quantization to need some calculation quantity, with the conservative estimate, the calculation quantity of this algorithm is decreased by 40% and speech quality has no change basically.

6. REFERENCES

[1] X.Y.Zhang, G.Zhang, K.M.Xie and L.Y.Huangfu, "Improving G.728 Algorithm: Hybrid Window and Adaptive Gain Filter".

[2] J.H. Chen, "A Robust Low-Delay Celp Speech Coder at 16 Kb/s", Advances in Speech Coding, pp.30-32, 1991

[3] J. Max., "Quantizing for Minimum Distortion", IRE Transactions on Information Theory, Vol.IT-6, pp.7-12, March 1960

[4] N.S. Jayant, "Adaptive Quantization With a One-word memory", Bell Syst. Tech.J., Vol.52,pp.1119-1144,Sept.1973

[5] Cummiskey P., N.S. Jayant and J.L. Flanagan, "Adaptive Quantization in Differential PCM Coding of Speech", Bell Syst.Teeh, Vol.52

[6] CCITT, Recommendation G.728, Coding of Speech at 16kbit/s Using Low-delay Code Excited Linear Prediction, Geneva, September 1992

[7] J.H. Chen, "High-quality 16kb/s Speech Coding With a One-way Delay Less 2ms", Proc.IEEE ICASSP, April 1990, pp.453~456

[8] Y.X. Qiao, "Improvement on G.728 Recommendation by using adaptive filter method", Taiyuan University of Technology Master's dissertation, 1999