# LSP and Gain Quantization for the Proposed ITU-T 8-kb/s Speech Coding Standard

Akitoshi Kataoka [1], Jotaro Ikedo [2] and Shinji Hayashi [1]

[1]NTT Human Interface Laboratories
3-9-11, Midori-cho, Musashino-shi,
Tokyo 180, Japan

[2]NTT Wireless Systems Laboratories
1-2356 Take, Yokosuka-shi,
Kanagawa 238-03, Japan

## Abstract

*The standardized 8-kbit/s speech coding algorithm approved by the ITU-T SG15 provides toll-quality with a 10-ms frame length. This paper describes methods for quantizing the line spectrum pair (LSP) parameters and the gain of the coder (G.729). The LSP parameters are quantized using multi-stage vector quantization (VQ) with interframe moving-average (MA) prediction. The gain is quantized using a conjugate gain codebook with gain prediction.*

## 1 Introduction

The ITU-T has been discussing standardization of an 8-kbit/s speech coding algorithm for use in the Future Public Land Mobile Telecommunication Systems (FPLMTS)[1]. The ITU-T requires an algorithm that provides toll-quality under error-free conditions and robustness against channel errors. The latter is very important for digital celluar applications. Furthermore, the algorithm should accommodate input speech with a flat frequency response as well as an intermediate reference system (IRS) one. The IRS is designed to imitate the typical sending frequency response of commercial telephone circuits. Flat response is needed to accommodate the possibility of other types of input speech. Therefore, the LSP quantizer has to handle various different frequency responses of input speech.

This paper discribes the methods for quantizing LSP parameters and gain of the CS-ACELP coder. Other papers describe the other modules and their performance [2,3,4,5].

## 2 Quantization of LSP parameters

The CS-ACELP uses a 10th-order synthesis filter. The LP coefficients are converted into LSP parameters and quantized. Since the frame length is 10 ms, we can exploit the redundancy arising from correlation between consecutive frames. The LSP parameters are first quantized using interframe 4th-order MA prediction. The residue of the predicted LSP parameters is then quantized by a two-stage VQ with a split structure at the second stage[6].

The n-th frame LSP parameters, $\Omega_n$, are generated using

$$\Omega_n = \mathbf{G}_0 \mathbf{C}_n + \sum_{i=1}^{M} \mathbf{G}_i \mathbf{C}_{n-i} , \qquad (1)$$

where $M$ is the interframe MA prediction order, $\mathbf{G}_i$ is a diagonal prediction matrix, and $\mathbf{C}_n$ is the output vector from the two-stage VQ at the n-th frame. Residue vector $\mathbf{C}_n$ is obtained from the code in each stage by using

$$\mathbf{C}_n = \mathbf{C}_{1j} + \begin{cases} \mathbf{C}^L_{2j} & j = 0,,,,4 \\ \mathbf{C}^H_{2j} & j = 5,,,,9 , \end{cases} \qquad (2)$$

where $\mathbf{C}_{1j}$ are the decoded values of the first stage, and $\mathbf{C}^L_{2j}$ and $\mathbf{C}^H_{2j}$ are the decoded values of the second stage. The dimension of the first-stage VQ is the same as the LPC analysis order, 10th, and the second stage consists of a combination of two 5-dimensional VQs.

For each frame, the first-stage codebook is searched to find the code vector $\mathbf{C}_{1n}$ that has the least distortion relative to a target vector. The target vector is generated by subtracting the prediction residue from the code vector for the previous frame from input LSP parameters $\Omega_{in}$. The second-stage codebook search selects the code vector that minimizes the weighted distortion between the input LSP parameters and the reconstructed LSP parameters $\Omega_n$:

$$d_{lsp} = (\Omega_{in} - \Omega_n)^T \ \mathbf{W}_n \ (\Omega_{in} - \Omega_n), \qquad (3)$$

where $\mathbf{W}_n$ are the weighting coefficients calculated from the unquantized LSP parameters $\omega_i$. The weights $w_i$ are shown in the following equation;

$$w_i = \begin{cases} 1.0 & \textit{if } \omega_{i+1} - \omega_{i-1} - 1 > 0, \\ 10(\omega_{i+1} - \omega_{i-1} - 1)^2 + 1 & \textit{otherwise} \end{cases} \qquad (4)$$

where $\omega_0$=0.04$\pi$ and $\omega_{11}$=0.92$\pi$. In addition, the weights $w_5$ and $w_6$ are multiplied by 1.2 each. Two sets of diagonal MA predictive coefficients are prepared; the one with the least distortion is selected. The LSP parameters are inspected to see whether they are sequenced appropriately; they are then rearranged to confirm the stability of the synthesis filter.

We investigated the optimum bit allocations of the coder and assigned 18 bits for the LSP quantization. The first stage uses 7 bits, each second stage uses 5 bits, and the selection of the diagonal predictive matrix uses 1 bit. The first-and second-stage LSP codebook and the predictive coefficient codebook are trained by using the generalized Lloyd algorithm. To improve the speech quality under background noise conditions, the LSP parameter quantizer is trained using a mixed database consisting of clean and noisy speech.

## 3 Quantization of gain

### 3.1 VQ gain with gain predictor

Since the bits for the input speech power are sensitive against channel errors, CS-ACELP does not assign bits for the input speech power. VQ gain with backward prediction removes the need to transmit the power information. This scheme is a mechanism that combines vector quantization gain with backward prediction.

The gains of the pitch excitation and the fixed excitation vectors are determined by using the VQ-gain codebook. Although the pitch gain is generally distributed around 1, the fixed codebook gain has a wider range. To compensate for this wider range, a 4th-order MA gain predictor with fixed coefficients is introduced[6,7]. A gain adaptor predicts the fixed codebook gain by considering the sequence of previous fixed-codebook excitation vectors. The predicted gain $g_0$ is expressed;

$$g_0 = 10^{(\tilde{E}^{(m)}+\bar{E}-E)/20}, \qquad (5)$$

where $\tilde{E}^{(m)}$ is the predicted energy, $\bar{E}$=30dB is the mean energy of the fixed codebook excitation and E is the mean energy of the fixed codebook contribution. The fixed-codebook-gain codebook is constructed from the residue of the logarithmic gain prediction with backward prediction. The excitation vector is multiplied by the predicted gain term. The pitch gain is completely defined by the VQ gain codebook, since it is not suitable for backward prediction.

The gain codebook has a conjugate structure with two sub-codebooks[6]. The pitch and fixed codebook gains are the sum of the gains from the two sub-codebooks:

$$g_p = g_{1pi} + g_{2pj}, \qquad g_c = g_0(g_{1ci} + g_{2cj}), \qquad (6)$$

$$\mathbf{G}_1 = (g_{1pi}, g_{1ci}), \qquad \mathbf{G}_2 = (g_{2pj}, g_{2cj}), \qquad (7)$$

where $g_p$ is the pitch gain, $g_c$ is the fixed-codebook gain, $g_0$ is the gain predicted by the gain adapter, and $\mathbf{G}_1$ and $\mathbf{G}_2$ are the VQ gain sub-codebooks. This approach requires less memory (2×(16+8) entries instead of 2×128).

### 3.2 Pre-selection for gain search

To reduce the complexity, pre-selection is applied to the gain codebook search. The gain codebook consists of two sub-codebooks: one is designed with a bias towards the fixed codebook gain $g_c$, and the other is designed with a bias towards the pitch gain $g_p$. That is, the elements of the first codebook distributed around line $L_1$ and the elements of the second codebook distributed around line $L_2$ as shown in Figure 1. Pre-selection is carried out by using the optimum gain $(g_{p\_opt}, g_{c\_opt})$, which are derived from the target vector, the adaptive-codebook contribution and the fixed-codebook contribution. Pre-selection in codebook #1 selects the 4 out of 8 candidates that are close to the fixed codebook gain, $g_{c1}$ (derived from $g_{p\_opt}$, $g_{c\_opt}$ and $g_0$). In the same way, pre-selection in codebook #2 selects the 8 out of 16 candidates that are close to the pitch gain, $g_{p2}$. Pre-selection is done by referring to a table constructed of the gain codebooks in advance. The appropriate gain code index is found by a closed-loop search that minimizes distortion, in which only the pre-selected candidates are used,

$$d_g = \|\mathbf{X} - g_p\mathbf{HP} - g_0 g_c\mathbf{HC}\|^2, \qquad (8)$$

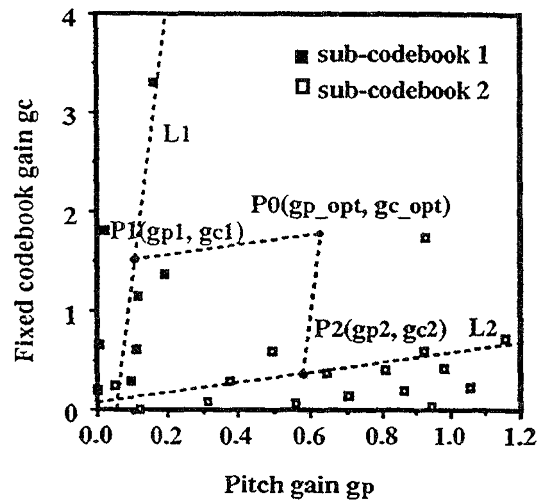


**Figure 1. Distribution of gain sub-codebook.**

where **X** is the target vector, **HP** represents the filtered adaptive-codebook contribution, and **HC** is the filtered fixed-codebook contribution. This pre-selection reduces the complexity by a factor of 4 without degrading quality, compared with the conventional VQ gain search.

## 4 Conclusion

We described methods for quantizing the LSP parameters and the gain of the CS-ACELP coder. The LSP parameters are quantized using multi-stage VQ with interframe MA prediction. The LSP quantizer can handle various different frequency responses of input speech. The gain is quantized using a conjugate gain codebook with gain prediction. This scheme reduces both memory requirements and complexity. These quantizers are very efficient and contribute to the good performance of CS-ACELP.

## Reference


[1] G. Schroeder, "The standardization process for the proposed ITU-T 8-kb/s speech coder standard," IEEE Speech Coding Workshop, Annapolis, 1995.

[2] R. Salami et al., "Description of the proposed ITU-T 8-kb/s speech coding standard", IEEE Speech Coding Workshop, Annapolis, 1995.

[3] D. Massaloux and S. Proust, "Spectral shaping in the proposed ITU-T 8-kb/s speech coding standard," IEEE Speech Coding Workshop, Annapolis, 1995.

[4] P. Usai et al., "Subjective performance of the proposed ITU-T 8-kb/s speech coding standard", IEEE Speech Coding Workshop, Annapolis, 1995.

[5] P. Kroon and Y. Shoham, "Performance of the proposed ITU-T 8-kb/s speech coding standard for a Rayleigh fading channel", IEEE Speech Coding Workshop, Annapolis, 1995.

[6] A. Kataoka et al., "An 8-kbit/s speech coder based on conjugate structure CELP", Proc. ICASSP 1993, pp. 592-595.

[7] R. Salami et al., "A toll quality 8 kb/s speech codec for the personal communications system (PCS)," IEEE Trans. Veh. Technol., vol. 43, no. 3, pp. 808-816, 1994.