PREDICTIVE VQ FOR BANDWIDTH SCALABLE LSP QUANTIZATION

Hiroyuki Ehara†, Toshiyuki Morii‡, Masahiro Oshikiri† and Koji Yoshida†

†Next-Generation Mobile Communications Development Center, / ‡AV Core Technology Development Center,
Matsushita Electric Industrial Co., Ltd.
{ehara.hiroyuki / morii.to / oshikiri.masahiro / yoshida.koji}@jp.panasonic.com

ABSTRACT

Predictive vector quantization (PVQ) for a bandwidth scalable LSP (line spectrum pairs) quantizer was studied. In the quantizer, LSP parameters obtained from a wide-band (16-kHz sampling) speech signal were quantized using LSP parameters encoded in a narrow-band (8-kHz sampling) layer. We designed several configurations of 16-bit quantizer utilizing PVQ and compared their performance. Simulation results showed that effective exploitation of inter-frame and intra-frame prediction improved objective performance. Compared with a baseline algorithm, PVQ using normal inter-frame prediction achieved better SD (spectral distortion) performance by 0.15 dB, and further improvement of 0.05 dB was gained by introducing novel inter-frame and intra-frame prediction. Robustness against frame erasure was considered by adopting switched prediction and "forgetting" capability. Designed memory-based PVQ algorithms were found to be robust up to 20% frame erasure rate.

1. INTRODUCTION

In this paper, we present an algorithm of bandwidth scalable LSP (line spectrum pairs) quantization utilizing a predictive VQ (vector quantization) scheme. Bandwidth scalable LSP quantization using inter-frame and intra-frame prediction was introduced by Nomura et al.[1], and Koishida et al.[2] proposed a joint optimization method for designing such inter/intra-frame predictors. However, there have been few studies on the structure of inter-frame and intra-frame predictors for the bandwidth scalable LSP quantizer. For predicting LSP parameters calculated from a wide-band signal (WB-LSP), we consider exploiting following parameters; 1) quantized LSP parameters obtained from a narrow-band signal (NB-LSP) on the current frame, 2) previously quantized WB-LSP and 3) previously quantized NB-LSP. Some configurations of the predictive VQ (PVQ) scheme are compared for studying efficient exploitation of the above three parameters. Since utilizing parameters 2) and 3) means that the prediction is memory-based and sensitive to transmission errors, switched prediction and "forgetting" capability are adopted in all the configurations in order to increase robustness against errors.

This paper is organized as follows. In Section 2, we explain the baseline PVQ algorithm of the bandwidth scalable LSP quantization. In Section 3, we describe some configurations of PVQ for the LSP quantizer proposed. In Section 4, the training procedure used in our study is outlined. In Section 5, objective performance is presented. We discuss about the results of our experiment in Section 6. Finally, in Section 7, we summarize our findings.

2. BASELINE SCALABLE LSP QUANTIZER

This section explains a scalable LSP quantization algorithm used as the baseline. The baseline algorithm utilizes simple intra-frame prediction and can be hence categorized as a PVQ-based algorithm. The algorithm is memory-less since inter-frame prediction is not used.

2.1. Overview

Figure 1 shows a block diagram of the baseline scalable LSP quantizer. An enhancement layer quantizes a WB-LSP vector in a layered manner using an NB-LSP vector quantized by a narrow-band layer. In the enhancement layer, the quantized NB-LSP vector is upsampled in an autocorrelation domain, and then a residual vector between two vectors of the input WB-LSP and the upsampled NB-LSP is quantized by a 2-mode 3-stage vector quantizer. One mode is designed to quantize LSP in stationary intervals, while the other mode is designed to be robust to a variety of LSP input. In our study, the sampling frequencies of the input wide-band and narrow-band signals are 16-kHz and 8-kHz, the orders of WB-LSP and NB-LSP are 18 and 12, respectively. Both the NB and WB-LSP quantizers are operated on 20 ms frames.

2.2. NB-LSP Quantizer

A 29-bit scalable quantizer is used as the NB-LSP quantizer. It consists of a core layer of a 21-bit quantizer and an enhancement layer of an 8-bit quantizer.

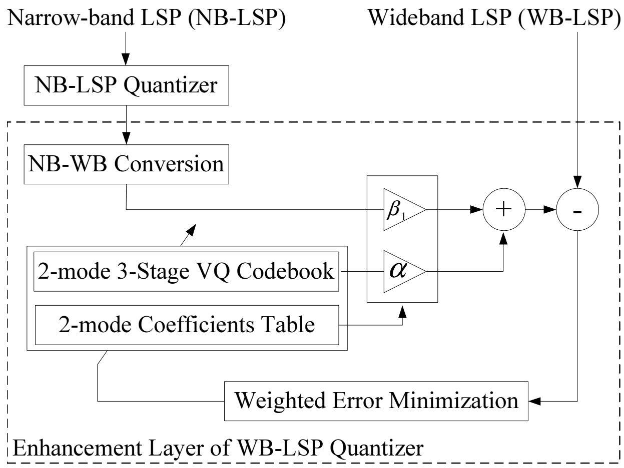

Fig. 1. Baseline scalable LSP quantizer

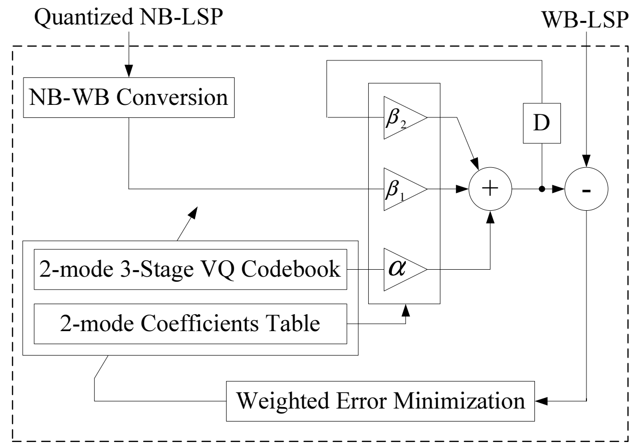

Fig. 2. PVQ1 (the 1st configuration)

2.3. NB-LSP to WB-LSP Conversion

The quantized NB-LSP parameters are transformed into autocorrelation coefficients via linear prediction coefficients. The autocorrelation coefficients are upsampled in a way equivalent to an upsampling process in the time domain. The upsampled autocorrelation coefficients are reconverted to LSP parameters.

2.4. Enhancement Layer Quantization

A quantized WB-LSP vector, $\hat{L}_W$, is given by Equation (1). $\hat{L}_W^{(n)}(i)$ is the i-th element of the quantized WB-LSP vector at the n-th frame. $\hat{L}_N^{(n)}(i)$ is the i-th element of the upsampled NB-LSP vector at the n-th frame. $\hat{C}^{(n)}(i)$ is the i-th element of a residual codevector quantized by the 2-mode 3-stage VQ at the n-th frame. $\alpha(i)$ and $\beta_1(i)$ are the i-the elements of predictive coefficients associated with the mode of the 3-stage VQ codebook.

$$\hat{L}_W^{(n)}(i) = \alpha(i)\hat{C}^{(n)}(i) + \beta_1(i)\hat{L}_N^{(n)}(i) \quad (1)$$

Five bits are assigned to each stage of the 2-mode 3-stage VQ, and one bit is used for switching the modes of the 3-stage VQ codebook and the predictive coefficients table associated with the codebook.

3. COMPARED PVQ

In this section, we present three PVQ configurations compared. Since we use AR (auto-regressive) prediction for PVQ, following two techniques are adopted to all the three configurations; 1) a switched predictive scheme for limiting propagation of effects of transmission errors, 2) "forgetting" capability for decreasing erroneous memory of a predictor. Regarding the switched prediction, the PVQ using AR prediction is introduced to only one of the two modes of the quantizer, so that propagation of errors is stopped when the other mode is selected, since the other mode is memory-less quantizer. Concerning the "forgetting" capability, since a quantized LSP vector is expressed by the sum of a memory-less and a memory-based (AR predictive) components, erroneous memory is diminished by the memory-less component as correct frames are increasingly received. Each of the three configurations is described in the following subsections.

3.1. PVQ1

Figure 2 shows a block diagram of the first configuration of the enhancement layer. A first order AR predictor is introduced to one of the two modes of the baseline quantizer as an inter-frame predictor. A quantized WB-LSP vector, $\hat{L}_W$, is given by Equation (2). In the case of memory-less VQ mode, $\beta_2(i)$ is set to zero in Equation (2) and Figure 2.

$$\hat{L}_W^{(n)}(i) = \alpha(i)\hat{C}^{(n)}(i) + \beta_1(i)\hat{L}_N^{(n)}(i) + \beta_2(i)\hat{L}_W^{(n-1)}(i) \quad (2)$$

3.2. PVQ2

Figure 3 shows a block diagram of the second configuration. A new prediction component is added to the first configuration. The new component is calculated by multiplying the quantized NB-LSP vector on the current frame by the ratio of the WB-LSP to NB-LSP vectors that are quantized on the previous frame. This new component performs both parts of inter-frame and intra-frame prediction. A quantized WB-LSP vector, $\hat{L}_W$, is given by Equation (3). When the

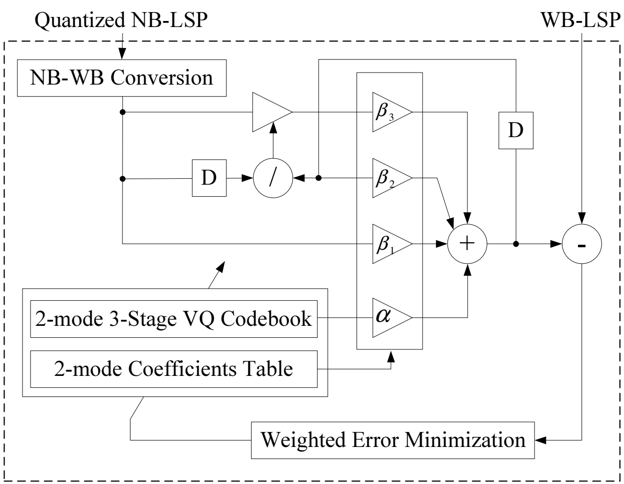

Fig. 3. PVQ2 (the 2nd configuration)

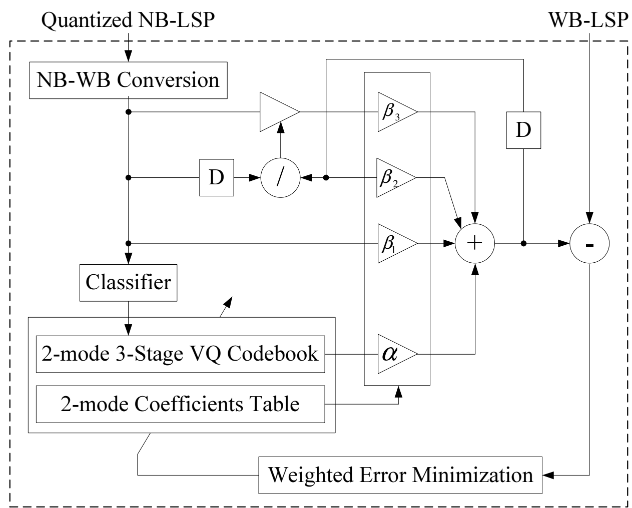

Fig. 4. PVQ3 (the 3rd configuration)

memory-less VQ mode is selected, $\beta_2(i)$ and $\beta_3(i)$ are set to zero in Equation (3) and Figure 3.

$$\hat{L}_W^{(n)}(i) = \alpha(i)\hat{C}^{(n)}(i) + \beta_1(i)\hat{L}_N^{(n)}(i) + \beta_2(i)\hat{L}_W^{(n-1)}(i) + \beta_3(i)\frac{\hat{L}_W^{(n-1)}(i)}{\hat{L}_N^{(n-1)}(i)}\hat{L}_N^{(n)}(i) \quad (3)$$

3.3. PVQ3

The third configuration of the enhancement layer is depicted in Figure 4. An idea of a classified VQ technique [3] is further applied to the second configuration without sending class information. The class information is obtained using NB-LSP quantized and encoded in the narrow-band layer. The quantized NB-LSP is inputted to a 3-bit vector quantizer and classified into one of eight classes. Each of the first stage codebooks of the 2-mode 3-stage VQ codebook consists of eight sub-codebooks that are corresponding to the eight classes. The class information identifies which sub-codebook should be searched in the 3-stage VQ. This technique exploits only the quantized NB-LSP on the current frame; hence it is categorized as intra-frame prediction.

4. CODEBOOK TRAINING

The training procedure for the predictive coefficients and residual codebook is the same as the design algorithm proposed by Koishida et al.[2]. The predictive coefficients and residual codebook were optimized so that the overall distortion, D, was minimized:

$$D = \sum_n \sum_{i=1}^{18} w^{(n)}(i)\left(L_W^{(n)}(i) - \hat{L}_W^{(n)}(i)\right)^2 \quad (4)$$

where $L_W^{(n)}(i)$ is the i-th element of an unquantized WB-LSP vector and $w^{(n)}(i)$ is the i-th weighting coefficient, at the n-th frame. To maintain "forgetting" capability, β_1 was clipped at 0.4 as a minimum value when updated β_1 was below 0.4, while the term of $\beta_2 + \beta_3$ is restricted below 0.6. Regarding the 3-stage codebook, all codevectors over the three stages were jointly optimized using the simultaneous joint codebook design proposed by LeBlanc et al.[4]. Two hundred Japanese short sentences were used as training data, from which 26,600 LSP vectors were generated.

5. OBJECTIVE PERFORMANCE

We evaluated the performance of the four PVQ configurations including baseline using average spectral distortion (SD) between unquantized WB-LSP vectors, L_W, and quantized WB-LSP vectors, $\hat{L}_W$. The performance for clean speech in error-free and frame erasure conditions were tested. Eight Japanese sentence pairs (four males and four females, 64 seconds in total) were used as test data. The test data was not included in the training data. The results are shown in Figure 5. The performance in the error-free condition and the 10% frame erasure condition is also summarized in Table 1 and 2 respectively with percentages of outliers. As shown in the figure and the tables, the three PVQ algorithms described in Section 3 outperformed the baseline algorithm both in error-free and frame erasure (FER$<$20%) conditions.

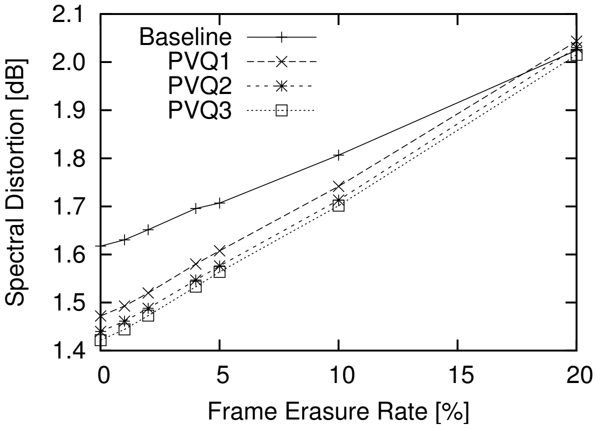

Fig. 5. Comparison of SD performance

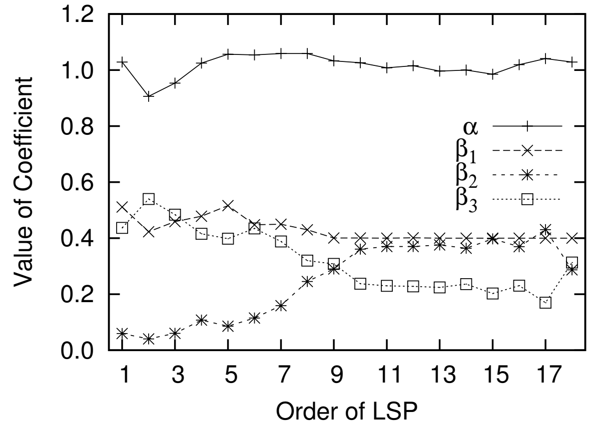

Fig. 6. Example of predictive coefficients

Table 1. Comparison of SD performance (error-free)

	average SD	2≤SD<4dB	4dB≤SD
Baseline	1.62 dB	20.44 %	0.50 %
PVQ1	1.47 dB	14.66 %	0.12 %
PVQ2	1.44 dB	13.84 %	0.22 %
PVQ3	1.42 dB	13.09 %	0.16 %

Table 2. Comparison of SD performance (FER=10%)

	average SD	2≤SD<4dB	4dB≤SD
Baseline	1.81 dB	23.38 %	3.40 %
PVQ1	1.74 dB	20.44 %	3.47 %
PVQ2	1.71 dB	19.81 %	3.38 %
PVQ3	1.70 dB	19.19 %	3.41 %

6. DISCUSSION

Figure 6 shows a typical example of predictive coefficients obtained through the training procedure in the PVQ2 case. The figure suggests that the term of $\hat{L}_W^{(n-1)}(i)$ in Equation (3) contributes to the prediction of higher order LSP while the term of $\frac{\hat{L}_W^{(n-1)}(i)}{\hat{L}_N^{(n-1)}(i)}\hat{L}_N^{(n)}(i)$ is effective in the quantization of lower order LSP. Thus the proposed AR-based prediction (PVQ2) appears to be capable of exploiting inter-frame and intra-frame correlation efficiently.

The classified VQ technique further improves the performance of PVQ2 at the cost of memory requirements and computational loads. PVQ3 scheme appeared to be one way for exploiting intra-frame correlation and realizing efficient quantization.

Regarding the SD performance of PVQ1–3, compared with the past report in [2], the percentages of outliers are relatively high, while the average SD is comparable. Further optimization of VQ codebook might be necessary.

7. CONCLUSION

Predictive vector quantization was applied to a 16-bit LSP (line spectrum pairs) quantizer as an enhancement layer of a bandwidth scalable LSP quantizer. Robustness against frame erasure was realized by adoption of switched prediction and "forgetting" capability of a memory-based predictor. Objective test results showed that better SD performance was achieved by efficient exploitation of inter-frame and intra-frame prediction. Ordinary inter-frame prediction improved SD performance by 0.15 dB, and additional improvement of 0.05 dB was achieved by further exploitation of inter-frame and intra-frame correlation.

8. REFERENCES

[1] T. Nomura, M. Iwadare, M. Serizawa, and K. Ozawa, "A bitrate and bandwidth scalable CELP coder," in *Proc. IEEE ICASSP-98*, June 1998, pp. 341–344.

[2] K. Koishida, J. Lindén, V. Cuperman, and A. Gersho, "Enhancing MPEG-4 CELP by jointly optimized inter/intra-frame LSP predictors," in *Proc. IEEE Workshop on Speech Coding*, 2000, pp. 90–92.

[3] A. Gersho and R. M. Gray, *Vector quantization and signal compression*, chapter 12.5, pp. 423–424, Kluwer Academic Publishers, 1992.

[4] W. P. LeBlanc, B. Bhattacharya, S. A. Mahmoud, and V. Cuperman, "Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding," *IEEE Trans. on Speech and Audio Process.*, vol. 1, no. 4, pp. 373–385, Octorber 1993.