ENHANCING MPEG-4 CELP BY JOINTLY OPTIMIZED INTER/INTRA-FRAME LSP PREDICTORS

Kazuhito Koishida, Jan Lindén, Vladimir Cuperman and Allen Gersho

Signal Compression Laboratory, Department of Electrical and Computer Engineering
University of California, Santa Barbara, CA 93106-9560, USA

ABSTRACT

This paper presents an LSP quantization design method for bandwidth scalable coders such as the MPEG-4 CELP coder. In the enhancement layer of these coders, the LSP parameters are quantized using both interframe and intraframe predictors. The proposed design algorithm enables us to jointly optimize these predictors. Objective and subjective test results show that the quantizer obtained with the proposed algorithm provides better performance than that used in the MPEG-4 CELP.

1. INTRODUCTION

A number of scalable coding algorithms have been developed for packetized speech communication. In these algorithms, the encoder generates the bit-stream in a layered manner so that the decoder can recover the output speech in cases where a part of the entire bit-stream is not available. The bit-stream of the scalable coders consists of a base layer and one or more enhancement layers. The base layer bit-stream provides a minimal quality and the enhancement layers allow a stepwise increase of the bit rate and quality. In some coders [1]–[4], even the transition from narrowband speech to wideband speech can be performed by the enhancement layer. This type of scalability is called bandwidth scalability.

This paper deals with bandwidth scalable CELP coders such as [1] and [2]. Our focus is particularly on LSP parameter quantization for this type of coders, in which intraframe and interframe predictors are used to quantize the LSP parameters of the enhancement layer. A design algorithm for such a quantizer is proposed, and the performance of the two-stage, split vector quantization structure obtained with the proposed method is compared with that of the bandwidth-scalable MPEG-4 CELP (BWS-MP4-CELP).

2. BANDWIDTH SCALABLE MPEG-4 CELP

Figure 1 shows a block diagram of the BWS-MP4-CELP encoder. In the base layer, the input wideband speech is down-sampled and narrowband speech is encoded by the base-layer CELP encoder. The wideband speech is encoded with the enhancement layer CELP encoder while making use of the available information from the base layer. In the decoder side, the narrowband speech is reproduced from the base-layer bitstream, and the wideband speech is obtained from the whole bitstream.

This work was supported in part by Research Fellowships of the Japan Society for the Promotion of Science for Young Scientists. This work was also supported in part by the National Science Foundation under grant MIP-9707764, the University of California MICRO program, Cisco Systems, Inc., Conexant Systems, Inc., Dialogic Corp., Fujitsu Laboratories of America, Inc., General Electric Co., Hughes Network Systems, Lernout & Hauspie Speech Products, Lockheed Martin, Lucent Technologies, Inc., Qualcomm, Inc., and Texas Instruments, Inc.

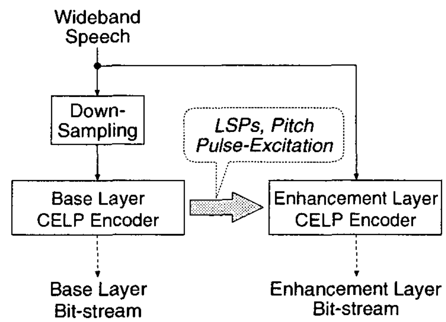


Figure 1: Overview of bandwidth scalable MPEG-4 CELP encoder.

In the enhancement layer, several parameters obtained in the base-layer CELP are used to improve the coding performance. These parameters include the LSP parameters, pitch and pulse excitation. In the next section, our attention will be focused on the LSP quantization of the enhancement layer.

3. LSP QUANTIZATION IN BWS-MP4-CELP

The LSP quantizer for the enhancement layer of the BWS-MP4-CELP coder is shown in Fig. 2. To utilize the LSP parameters obtained in the base layer, intraframe prediction is incorporated into the quantizer. This prediction produces estimated LSP parameters from the base layer's quantized LSP parameters. In addition, moving-average (MA) interframe prediction is employed to remove frame-to-frame redundancy in the enhancement LSP parameters. The prediction error obtained after both interframe and intraframe prediction is quantized using the residual codebook.

The reconstructed LSP parameters at frame n are expressed as

$$\hat{f}_n(i) = \sum_{p=0}^{P} \alpha_p(i)\hat{l}_{n-p}(i) + \beta(i)\hat{g}_n(i), \quad 1 \leq i \leq 20 \qquad (1)$$

where $\hat{l}_n(i)$ is the residual codebook output at frame n, and $\alpha_p(i)$ and $\beta(i)$ are the interframe and intraframe predictive coefficients, respectively. The parameters $\hat{g}_n(i)$ contain the quantized base-layer's LSP parameters in the first ten parameters and zeros in the

0-7803-6416-3/00/$10.00©2000 IEEE

rest, i.e.,

$$\hat{g}_n(i) = \begin{cases} \hat{f}_n^{(\text{base})}(i), & 1 \le i \le 10 \\ 0, & 11 \le i \le 20. \end{cases} \quad (2)$$

The interframe prediction uses two prior frames, so that $P = 2$, in the BWS-MP4-CELP coder.

The LSP quantizer selects the index of each codebook which minimizes the weighted Euclidean distance between the input and reconstructed LSP parameters:

$$D_n = \sum_{i=1}^{20} w_n(i) \left(f_n(i) - \hat{f}_n(i)\right)^2 \quad (3)$$

where $w_n(i)$ are perceptual weighting coefficients that emphasize regions close to the formant frequencies.

4. DESIGN ALGORITHM

In this section, we discuss a design procedure for the LSP quantizer described in the previous section. For simplicity, the algorithm is presented for a single-stage codebook. The predictive coefficients and residual codebook are optimized so that the overall distortion

$$D = \sum_n D_n \quad (4)$$

is minimized. To solve the minimization problem, an iterative training algorithm is used in this paper. The algorithm is based on the generalized Lloyd algorithm and alternately performs two processes. The first process encodes a training set of LSP parameters for given predictors and residual codebook. In the second process, using the indices obtained from the first process, either the predictors or the residual codebook are optimized. These two processes iterate until the overall distortion change since the last iteration is below a threshold.

4.1. Initialization

The intraframe and interframe predictors are initialized by setting $\alpha_0(i) = \cdots = \alpha_P(i) = \beta(i) = 0$. Assuming that all predictors are zero, the initial residual codebook is generated by the LBG algorithm.

4.2. Predictor optimization

The intraframe and interframe predictors are optimized by solving the linear equations:

$$\begin{bmatrix} q^{(i)}(\hat{l}_n, \hat{l}_n) & \cdots & q^{(i)}(\hat{l}_n, \hat{l}_{n-P}) & q^{(i)}(\hat{l}_n, \hat{g}_n) \\ \vdots & \ddots & \vdots & \vdots \\ q^{(i)}(\hat{l}_{n-P}, \hat{l}_n) & \cdots & q^{(i)}(\hat{l}_{n-P}, \hat{l}_{n-P}) & q^{(i)}(\hat{l}_{n-P}, \hat{g}_n) \\ q^{(i)}(\hat{l}_{n-P}, \hat{g}_n) & \cdots & q^{(i)}(\hat{l}_{n-P}, \hat{g}_n) & q^{(i)}(\hat{g}_n, \hat{g}_n) \end{bmatrix}$$
$$\times \begin{bmatrix} \alpha_0(i) \\ \vdots \\ \alpha_P(i) \\ \beta(i) \end{bmatrix} = \begin{bmatrix} q^{(i)}(f_n, \hat{l}_n) \\ \vdots \\ q^{(i)}(f_n, \hat{l}_{n-P}) \\ q^{(i)}(f_n, \hat{g}_n) \end{bmatrix}, \quad i = 1, \cdots, 20 \quad (5)$$

where

$$q^{(i)}(x_{n-j}, y_{n-k}) = \sum_n w_n(i) x_{n-j}(i) y_{n-k}(i). \quad (6)$$

The above equations are obtained by setting the partial derivatives of Eq. (4) with respect to $[\alpha_1(i), \cdots, \alpha_P(i), \beta(i)]$ to zero. An important point here is that the intraframe and interframe predictors are simultaneously optimized.

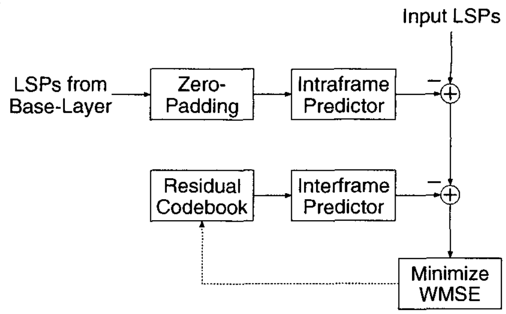


Figure 2: LSP quantizer in enhancement layer of BWS-MP4-CELP.

4.3. Residual codebook update

Taking the partial derivative of Eq. (4) with respect to each entry of the residual codebook and setting the result to zero, we obtain the following equation for updating the j-th centroid of the residual codebook:

$$c_j(i) = \frac{\sum_{n:I_n=j} w_n(i)\alpha_0(i) \left(f_n(i) - \sum_{p=1}^{P} \alpha_p(i)\hat{l}_{n-p} - \beta(i)\hat{g}_n(i) \right)}{\sum_{n:I_n=j} w_n(i)\alpha_0^2(i)} \quad (7)$$

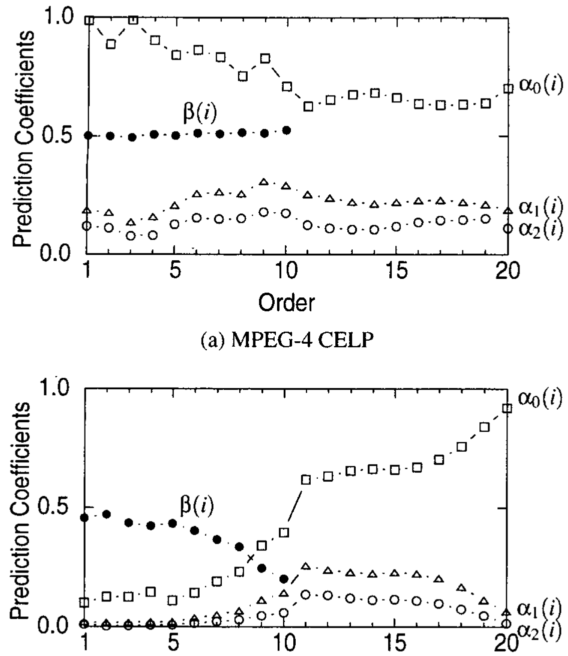
where I_n represents the residual codebook index at frame n that is obtained in the last encoding process. This equation can be easily adapted for a multi-stage vector quantization by including for each stage the effect of all other stages in the numerator term.

4.4. Design example

Fig. 3(b) shows a typical example of the predictive coefficients obtained with the proposed optimization. For comparison purpose, the predictive coefficients of the BWS-MP4-CELP are plotted in Fig. 3(a). It is evident that the coefficients for the proposed optimization show different characteristics from those used in the BWS-MP4-CELP. It is also found from the figure that, in the proposed optimization, the intraframe predictor provides better estimation for lower predictor orders, while the interframe prediction becomes more important at higher orders. An advantage with the predictors generated by the proposed design method is that all components in the residual vector will have values in the same order of magnitude. For example, this means that the vector could be split at any point without causing any problems. For the BWS-MP4-CELP solution, on the other hand, the first ten components in the residual vector will have magnitudes significantly less than those of the last ten components. This is a consequence of the very strong intraframe and interframe predictive coefficients used. For this design the residual vector has to be split at the middle in order to make the MSE meaningful.

5. EXPERIMENTS

Two LSP quantizers for the BWS-MP4-CELP were designed using the proposed design algorithm. The quantizers have similar



(a) MPEG-4 CELP

(b) Proposed

Figure 3: Example of interframe and intraframe predictive coefficients. The order of MA prediction is two.

configuration to that used in BWS-MPEG-4 CELP: second-order MA interframe prediction and two-stage structure for the residual codebook (the input vector is equally divided in each stage, two splits for the first stage and four splits for the second stage). The bit allocations for these quantizers are summarized in Table 1. The proposed Quantizer-I has the same complexity and memory size as that of MPEG-4 CELP. The proposed Quantizer-II employs a switched predictor scheme, in which each set of the intraframe and interframe predictors is associated with a different residual codebook. The Quantizer-II requires almost twice as high complexity and memory size as the BWS-MP4-CELP quantizer.

The LSP parameters for training and testing were obtained with the same pre-processing and LP analysis as in the BWS-MP4-CELP. The training and test sets consisted of 79640 and 3813 LSP vectors, respectively.

5.1. Spectral distortion and outliers

Table 2 lists the objective performance of the three quantizers. The results show that the proposed quantizers achieve a significant improvement over the BWS-MP4-CELP quantizer. Note that the proposed quantizers are able to reduce the number of outliers with SD between 2-4 dB significantly and remove all outliers with SD above 4 dB. From perceptual point of view, this is a promising result. The Quantizer-II with the switched predictors has slightly better performance than the Quantizer-I with the fixed predictor.

Table 1: Bit allocations for three quantizers. The total number of bits for each quantizer is 32.

	Pred.	1st stage	2nd stage
MPEG-4	0	4 7	4 6 7 4
Proposed-I	0	5 7	4 5 7 4
Proposed-II	1	5 7	4 4 7 4

Table 2: Quantization performance (dB).

	SD (dB)	Outliers (%)	
		2-4 dB	4 > dB
MPEG-4	1.64	20.6	0.29
Proposed-I	1.44	8.2	0
Proposed-II	1.41	7.2	0

Table 3: Preference Scores.

MPEG-4	Proposed-I
41 %	59 %

5.2. Listening test

To complement the objective results presented in the previous subsection, an informal listening test was conducted. In this test, the Quantizer-I was integrated into the BWS-MP4-CELP coder, and its subjective quality for wideband speech was compared with that of the conventional BWS-MP4-CELP. The wideband speech was generated from the 6 kb/s base layer and the 10 kb/s enhancement layer. Eleven people listened with headphone to 16 sentences uttered by 8 female and 8 male speakers.

The A-B test result is presented in Table 3. It is shown that the proposed Quantizer-I provides better quality than the BWS-MP4-CELP quantizer without increasing the complexity or the memory requirement.

6. CONCLUSIONS

This paper has presented an LSP quantizer design for bandwidth-scalable coders. The new method is able to jointly optimize the interframe and intraframe predictors and iterates between optimizing the predictors and the residual codebook. It has been shown that the proposed design algorithm yields higher performance than that used in the bandwidth scalable MPEG-4 CELP coder.

REFERENCES

[1] ISO/JTC1, "Final draft international standard FDIS 14496-3: Coding of audiovisual objects, part 3: Audio," ISO/JTC1 SC29 WG11, 1998.

[2] K. Koishida, V. Cuperman and A. Gersho, "A 16-kbit/s bandwidth scalable audio coder based on the G.729 standard," to appear in *Proc. ICASSP 2000*.

[3] A. Kataoka, S. Kurihara, S. Sasaki and S. Hayashi, "A 16-kbit/s wideband speech codec scalable with G.729," in *Proc. EUROSPEECH*, pp.1491–1494, 1997.

[4] ITU-T Recommendation G.722, "7 kHz audio-coding within 64 kbit/s," 1988.