MATRIX PRODUCT QUANTIZATION FOR VERY-LOW-RATE SPEECH CODING

Stefan Bruhn

Institute for Telecommunications, Technical University of Berlin
Berlin, Germany, email: bruhn@ftsu00.ee.tu-berlin.d400.de

ABSTRACT

Efficient block coding methods for LPC information play an essential role in very-low-rate speech coding systems. The subject of this contribution is a new suboptimal matrix quantization scheme for LPC parameters, called *matrix product quantization (MPQ)*, which operates at bit rates between 300 and 700 b/s. MPQ encodes sequences of LPC parameter vectors using a product formulation of two matrices which describe the average parameter vector and the temporal contour. In fixed-rate coding systems for mobile communication, MPQ achieves a very high coding efficiency at a low coding delay. Compared to the multi-frame coding method (MFC) of Kemp et al. [1], which causes a delay of 8 frames, the MPQ scheme operates more efficiently even at a coding delay of only 3 frames. Applying MPQ to a variable-rate segment vocoder, a bit rate reduction of 50% compared to memoryless VQ is obtained at a frame period of 20 ms.

1. INTRODUCTION

The sequence of LPC parameter vectors representing the speech short-term power spectrum has strong statistical interframe dependencies. Therefore, efficient coding methods have to exploit the memory of the LPC vectors. An obvious method is matrix quantization (MQ), i.e. the coding of sequences of succeeding LPC vectors as a matrix. However, due to complexity reasons, its application is restricted to coarse quantization at extremely low bit rates of a few 100 b/s. Very-low-rate mobile speech communication systems require efficient fixed-rate and low-delay coding methods which operate at higher bit rates. Appropriate suboptimal derivates of MQ like multi-frame-coding (MFC) transmit only a few frames out of each matrix as anchor points, whereas the other frames are either differentially encoded or interpolated at the receiver [1], [2]. In storage applications neither a fixed bit rate nor a low coding delay is required. Consequently, the bit rate can be adjusted to the time-varying information rate of speech, which results in an improvement of coding efficiency. Segment vocoders are variable-rate coding systems transmitting speech in acoustical units or segments which represent phonemes or diphones [3-5]. The segments, consisting of a varying number of frames, are time-normalized (depending on the realization) and then encoded using MQ. A variable-rate extension of MFC is the combined quantization-interpolation (CQI) method [6].

This contribution introduces a new coding scheme for LPC parameter matrices, called *matrix product quantization (MPQ)*. It is applicable both in fixed-rate speech coders and in variable-rate segment vocoders. As MPQ is less complex than MQ, it can be used for highly efficient coding at much higher bit rates.

This paper is organized a follows: Section 2 defines MPQ as a fixed-rate coding scheme. The application of MPQ in a variable-rate segment vocoder is discussed in section 3. Section 4 provides simulation results. Section 5 presents conclusions.

2. MATRIX PRODUCT QUANTIZATION (MPQ)

Due to strong temporal dependencies, succeeding speech short-term power spectra remain relatively unchanged for time periods much longer than one frame. Therefore, it is effective to describe each matrix $\boldsymbol{X}$ of m successive LPC parameter vectors $\boldsymbol{x}_i$ to be encoded by the average parameter vector and the temporal contour. The former is represented by a diagonal centroid matrix $\boldsymbol{S}$, while the latter is expressed by a contour matrix $\boldsymbol{V}$ consisting of normalized row vectors. The reconstruction matrix $\boldsymbol{Y}$ is formed by the matrix product:

$$\boldsymbol{Y} = \mathrm{q}(\boldsymbol{X}) = \boldsymbol{S}\boldsymbol{V} \quad , \boldsymbol{S} \in C_S \quad , \boldsymbol{V} \in C_V \quad . \tag{1}$$

The MPQ has to select that pair of centroid and contour matrices $(\boldsymbol{S},\boldsymbol{V})$ out of the corresponding codebooks C_S and C_V which yields the best reconstruction of the input matrix $\boldsymbol{X}$. The codebook indices representing both matrices are transmitted to the receiver. Due to the product formulation, MPQ is substantially less complex than ordinary MQ.

Fig. 1 depicts a basic MPQ system.

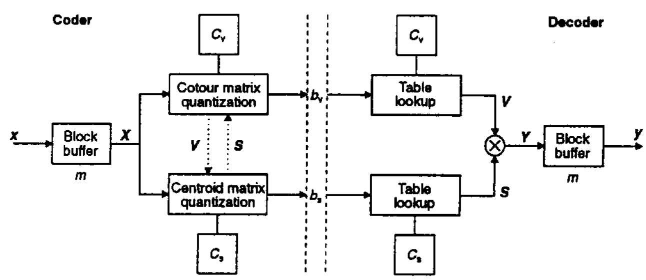

Fig. 1 Basic matrix product quantization system

2.1. Distortion measure

Spectral distortion measures like the likelihood-ratio are very suitable for the speech LPC information. However, in general, they are not applicable in product VQ systems. Instead, MPQ is based on the adaptively weighted squared error:

$$\mathrm{d}(\boldsymbol{X},\boldsymbol{Y}) = \sum_{i=1}^{m} (\boldsymbol{x}_i - \boldsymbol{y}_i)^t \boldsymbol{W}_i (\boldsymbol{x}_i - \boldsymbol{y}_i) \quad . \tag{2}$$

$\boldsymbol{W}_i$ denotes a diagonal weighting matrix for the i-th error vector of dimension p.

In order to accomplish a good approximation of the spectral distortion measures, line spectral frequencies (LSF) are used to represent the LPC information. According to [7], the weights of the squared error measure are adapted to the spectral sensitivities of the LSF vectors to be encoded. Assuming that frames of high energy are particularly important for speech perception, the weights can additionally be adapted to the respective frame energies.

0-7803-2431-5/95 $4.00 © 1995 IEEE

2.2. MPQ algorithms

As a main problem of MPQ, the selections of centroid and contour matrices are mutually dependent. Therefore, the optimal matrix pair can actually only be found by an expensive full-search of all possible matrix combinations. However, a matrix pair which is at least locally optimal can be found using a computationally advantageous iterative algorithm. The procedure alternatingly searches for either the optimal centroid matrix or the optimal contour matrix, while the corresponding contour or centroid matrix, respectively, is fixed. Each step of this algorithm yields a matrix pair causing a distortion which is less than or equal to the distortion of the previous matrix pair. Thus, convergence towards a local optimum is ensured. The algorithm is completed as soon as a further iteration step retains the current matrix pair. The MPQ procedure is given as follows:

(0) Start either with step **I** using an arbitrary centroid matrix S_q, or with step **IIa** using an arbitrary contour matrix V_q.

(I) Find the best contour matrix V_q available, assuming a fixed centroid matrix S_q:

$$V_q = \min_{V \in C_v}^{-1} \left[-2 \sum_{j=1}^{p} s_{q,j} \sum_{i=1}^{m} w_{ji} x_{ji} v_{ji} + \sum_{j=1}^{p} s_{q,j}^2 \sum_{i=1}^{m} w_{ji} v_{ji}^2 \right] . \quad (3)$$

Exit if the previous contour matrix is retained, if not, continue.

(IIa) Calculate the optimal centroid matrix $S_{opt} = S_{opt}(X,V_q)$ and the corresponding weighting matrix W_S, both depending on the current contour matrix V_q:

$$S_{opt} = \mathrm{diag}[s_{opt,1},\ldots,s_{opt,p}] \quad \text{where} \quad s_{opt,j} = \frac{\sum_{i=1}^{m} w_{ji} v_{ji} x_{ji}}{\sum_{i=1}^{m} w_{ji} v_{ji}^2} , \quad (4)$$

$$W_S = \mathrm{diag}[w_{S,1},\ldots,w_{S,p}] \quad \text{where} \quad w_{S,j} = \sum_{i=1}^{m} w_{ji} v_{ji}^2 . \quad (5)$$

(IIb) Find the best centroid matrix S_q available, assuming a fixed contour matrix V_q by minimizing the weighted squared distance between S_q and S_{opt}:

$$S_q = \min_{S \in C_s}^{-1} \mathrm{tr}[W_S (S - S_{opt})^2] . \quad (6)$$

Exit if the previous centroid matrix is retained, otherwise, continue with step **I**.

Experiments have shown that the MPQ procedure converges within only few iterations. Moreover, if the initial matrix is chosen appropriately, a single run of the iteration loop will be sufficient to locate the optimal matrix pair with high certainty. Provided that the quantization of either the centroid or the contour matrices is sufficiently fine, good initial guesses of these matrices are the unquantized optimal matrices. To start the MPQ procedure with step **I**, the optimal centroid matrix according to eq. (4) will be appropriate: $S_q := S_{opt}(X,V)$. Conversely, the MPQ procedure can also begin with step **IIa**, initially using the unquantized optimal contour matrix: $V_q := V_{opt}(X,S)$. However, in the general case of frame by frame distinctive weighting matrices, there is no explicit mathematical solution for the optimal contour matrix.

The numerical complexity of MPQ will be considerably reduced if the weighting matrices are constant during each matrix X. Due to the normalization of the contour matrix row vectors, the calculation of the sums $\Sigma w_{ji} v_{ji}^2$ is superfluous, in this case. Thus, the adaptation to each frame has to be abandoned, and instead, the weighting matrix has merely to depend on the entire input matrix. A suitable choice of such a weighting matrix is the arithmetic mean of the distinctive weighting matrices corresponding to the LPC vectors of X:

$$\bar{W}(X) = \frac{1}{m} \sum_{i=1}^{m} W_i(x_i) . \quad (7)$$

2.3. Stabilization of the LPC synthesis filter

The MPQ of LSFs may affect their monotony resulting in an instable LPC synthesis filter. This problem can be solved by simply sorting the quantized LSFs. However, synthesis filters that are stabilized this way often show sharp spectral peaks, which causes unnatural sounds in the reconstructed speech. Such stabilized synthesis filters can considerably be improved by additionally limiting their pole bandwidths to a minimal value b_{min}. For that purpose, each pole z_x having a bandwidth less than b_{min} is forced to the minimal bandwidth by applying a suitable factor:

$$z_{x,mod} = \frac{z_x}{|z_x|} \exp(-\pi \frac{b_{min}}{f_s}) , \quad (f_s = \text{sampling rate}) . \quad (8)$$

2.4. Extension of MPQ

As the complexity of MPQ increases with the size of the centroid and contour matrix codebooks, there are realization limits for the basic MPQ scheme. However, the application range of MPQ can easily be extended to finer quantizations at higher bit rates by introducing suboptimal VQ methods for the centroid and contour matrices. For its simplicity and high efficiency, the principle of split VQ is applied [8]. The MPQ with split VQ of the centroid matrices *(SMPQ-S)* operates just as the ordinary MPQ, but simply splits up the diagonal vector of the centroid matrix into subvectors. The least complex MPQ with split VQ of both centroid and contour matrices *(SMPQ-SV)* additionally splits up the contour matrices horizontally.

2.5. Comparable methods

MPQ is comparable to a 2-D extension of shape-gain VQ (SGVQ) [9]. However, in contrast to the shape-gain MQ (SGMQ), which is based on the Itakura-Saito distortion measure [10], MPQ applies the adaptively weighted squared error. Although feasible, MPQ does not include the quantization of the LPC model gain. Furthermore, SGMQ is much more complex than MPQ, since the shape quantization part of SGMQ corresponds to an ordinary MQ of the LPC information.

2.6. System design

As the principles of MPQ and SGVQ correspond to each other, centroid and contour matrix codebooks can be designed by the individually-optimized (IO) algorithm, originally established for SGVQ [9]. After the codebook design, the optimal bit allocation to the product codebooks has to be determined. With the optimal bit allocation, MPQ yields a minimum average coding error at a given bit rate.

3. SEGMENT MPQ

When the constraint of a fixed number m of LPC vectors per matrix is dropped, the MPQ system works as a variable-rate segment vocoder. Now, the matrices can be arranged in a way that they represent acoustical units. If the arrangement, i.e. the segmentation, is done appropriately, the matrices will contain separate speech sounds. Since the LPC vectors remain almost unchanged within such segments, the bit rate for the contour matrices is substantially reduced. Additionally, however, the varying matrix length m has to be transmitted.

Fig. 2 depicts the basic *segment MPQ* system including some variants. In the basic system, the input sequence of LPC vectors is first subdivided into segments by an algorithm which detects periods of spectral constancy. After that, the segments are encoded applying a length specific MPQ, i.e. specific codebooks and bit allocations are used. The segment lengths are encoded using a variable-rate Huffman-code.

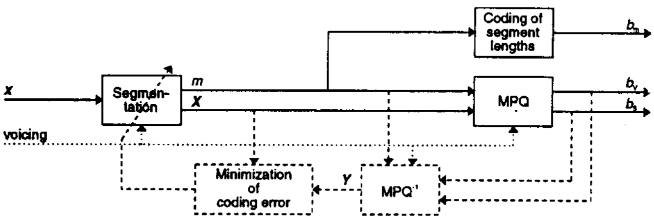

Fig. 2 Basic segment MPQ system and variants

3.1. Joint segmentation and MPQ

A segmentation independent of the following quantization cannot be optimal with respect to the coding error. However, if the coding delay is not constrained, the segmentation and quantization can be performed jointly in a closed-loop scheme. A commonly applied, computationally efficient dynamic programming algorithm allows to optimize the segmentation with respect to the resulting distortion [4]. As a drawback, this algorithm retains the non-optimized segment rate of the initial segmentation. Therefore, it cannot minimize the coding error at a given bit rate, if, according to the bit allocation, segments of different length are encoded with different numbers of bits. To overcome this problem, an extended optimal segmentation algorithm is proposed which is able to merge multiple segments into single segments. Thus, both segmentation and segment rate can be optimized such that a minimal coding error at a given bit rate is achieved. The algorithm consists of three phases: initial segmentation, forward recursion, and backward recursion.

The initial segmentation is done by the segmentation algorithm which is also applied in the basic open-loop segment MPQ. This procedure finds a sequence of j+1 initial segment boundaries $\{n_0^0, n_1^0, .., n_j^0\}$, where n_0^0 and n_j^0 denote the starting and end positions of the LPC vector sequence to be encoded.

After that, the forward recursion optimizes the segment boundaries within search ranges $S_i = [n_i^0 - \Delta, n_i^0 + \Delta]$ around the initial boundaries. Progressing for all initial segment boundaries n_i^0, for each possible segment boundary n of the search range S_i, the corresponding optimal preceding segment boundary n' is determined which minimizes the cost function, i.e. the accumulated coding error $d_\Sigma(n)$ up to the position n. n' is taken out of a predecessor search range $S_i'(n)$, which is defined by the minimal and maximal segment lengths of the segment MPQ. The predecessor search range $S_i'(n)$ is only restricted by the constraint that its members must be possible segment boundaries, already considered in the preceding recursion steps. The predecessor search ranges are independent of the initial segmentation. Furthermore, the search ranges of different initial segment boundaries may overlap. Therefore, in contrast to [4], initial segment boundaries can be removed, and thus the segment rate can be reduced, if advantageous. The forward recursion is completed with the determination of the optimal preceding segment boundary of the end position n_j^0.

Now, for each possible segment boundary the optimal preceding segment boundary is known. Therefore, the final backtracking recursion can easily deduce the optimal segmentation. Starting with the end position of the LPC vector sequence, the procedure determines the optimal predecessor of each already known optimal segment boundary. The algorithm closes when all optimal segment boundaries up to the starting position of the LPC vector sequence are found.

3.1.1. *Control of the resulting bit rate*

The joint segmentation and MPQ algorithm optimizes segmentation and segment rate with respect to the average coding error. However, it does not yet allow a direct control of the resulting bit rate. For this purpose, the optimization must be done under the constraint of a given bit rate. Such a problem can be solved by application of the Lagrangian multiplier method. This leads to a modified cost function which additionally contains the bit rate weighted by a factor λ. Let d be the original accumulated coding error resulting from the MPQ of a segment, and let b be the number of bits needed to encode this segment. Then, the segmentation algorithm has to minimize the modified costs $d_{mod} = d + \lambda b$. Now, the resulting bit rate is taken into account, according to the factor λ. Consequently, with larger λ, an increasing number of initial segment boundaries is removed, and according to the bit allocation, bit-expensive segments become rare.

The numerical value of the factor λ, required to obtain a specific bit rate, has to be found empirically. However, it is also feasible to adjust the resulting bit rate using a variable λ.

3.1.2. *Reduction of the number of outlier frames*

The number of outlier frames with extreme distortion degrading the subjective speech quality significantly can be reduced by an additional modification of the cost function. Applying the ν-th ($\nu > 1$) power d_f^ν instead of the original frame distortion d_f, large coding errors are weighted more heavily than small errors. Thus, the optimization tends to avoid frames with large coding errors since they cause higher than proportional costs.

3.2. Voicing dependent segment MPQ

Another voicing dependent segment MPQ variant operates with a voicing specific LPC filter order. In segments consisting of unvoiced frames only the LPC filter order is reduced from 10 to 4. This extension allows a rate-reduced transmission of unvoiced periods without a degradation of speech quality. Depending on the voicing decision specific codebooks and bit allocations are used.

4. SIMULATION RESULTS

The speech database used for the evaluation of the proposed coding schemes covers 41 min of speech without pauses from several male and female speakers. A subset of 31 min is taken for training. The remaining material of 10 min, produced by speakers not present in the training set, is taken for the assessment. A 10th-order LPC analysis is performed at a frame period of 20 ms using analysis frames of 30 ms duration.

4.1. Fixed-rate MPQ

The fixed-rate MPQ scheme is compared to the MFC method based on split VQ of LSF parameters, as proposed by Kemp et al. [1], and ordinary memoryless VQ of LSF parameters. According to the simulation results shown in Fig. 3, at a matrix length, i.e. a delay, of only m = 3 frames MPQ with split VQ of the centroid matrices (SMPQ-S) slightly outperforms MFC which causes a delay of 8 frames. Furthermore, at the same matrix length

SMPQ-SV with split VQ of both centroid and contour matrices is computationally less complex than MFC, and therefore even operates at higher bit rates. By extending the matrix length, the coding efficiency of MPQ is again improved. However, due to complexity reasons, the application is now limited to lower bit rates. Compared to VQ, the MPQ method yields bit savings of up to 30%.

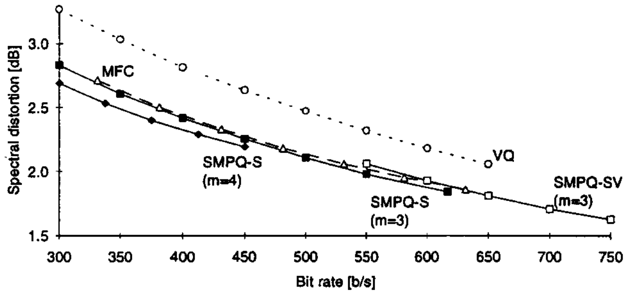

Fig. 3 Performance of MPQ (SMPQ-S, m=3 (■), m=4 (◆); SMPQ-SV, m=3 (□)), MFC (△), and VQ (○)

Regarding the statistical distribution of the spectral distortion (SD), MPQ again outperforms MFC. At a given average SD, MPQ causes considerably fewer outliers, i.e. extremely distorted frames, than MFC. Table 1 shows an example of SMPQ-SV and MFC, both operating at almost identical bit rates and average SD. Compared to MFC, the MPQ method reduces the number of outliers (with SD > 4 dB) by more than 90 %.

Method	Average SD	Bit rate	SD > 4 dB
SMPQ-SV (m=3)	1.85 dB	633 b/s	0.47 %
MFC	1.85 dB	631 b/s	5.28 %

Tab. 1 Relative frequency of outliers of MPQ and MFC

4.2. Variable-rate Segment MPQ

A main result of the simulations of segment MPQ is a further improvement of coding efficiency, due to the dropped requirement of a fixed bit rate (cf. Fig. 4). The principle of optimizing the segmentation as well as the segment rate jointly with the quantization is very powerful. Especially, by additionally performing a voicing dependent quantization, considerable bit rate reductions are achieved. Compared to fixed-rate MPQ and memoryless VQ, bit savings up to 40% and 50%, respectively, are obtained. Even the basic segment MPQ slightly outperforms the CQI method [6] which is applied to VQ of LSFs operating in a range of 10-13 bit per vector.

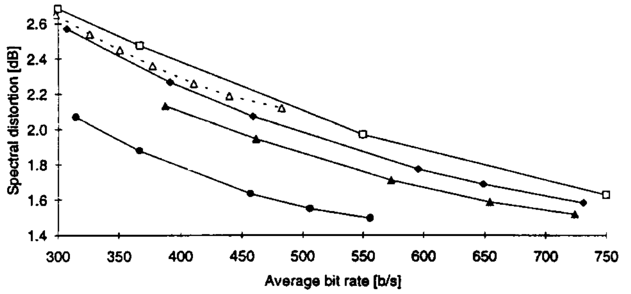

Fig. 4 Performance of basic segment MPQ (◆), closed-loop segment MPQ (▲), closed-loop voicing dependent segment MPQ (●), fixed-rate MPQ (□), and CQI [6] (△)

In the closed-loop segment MPQ the modification of the cost function substantially reduces the number of outliers. Table 2 shows the relative frequencies of extremely distorted frames (with SD > 4 dB) of closed-loop segment MPQ without and with modified cost function (exponent ν = 1, 3), operating at approximately the same bit rate. It is observed that the number of outliers is reduced by 75%, while the average SD is almost unaffected. Furthermore, compared with the results of fixed-rate SMPQ-SV (cf. table 1), despite a lower bit rate the frequency of outliers is again drastically reduced.

Exponent	Average SD	Bit rate	SD > 4 dB
ν=1	1.89 dB	484 b/s	0.20 %
ν=3	1.90 dB	487 b/s	0.05 %

Tab. 2 Relative frequency of outliers of closed-loop segment MPQ depending on exponent ν of cost function

5. CONCLUSION

A new coding principle for LPC parameter vectors is presented which operates at bit rates between 300 and 700 b/s. It can be applied to fixed-rate coding systems as well as to variable-rate segment vocoders. At a fixed bit rate the MPQ method is advantageous especially for very-low-rate mobile communication systems as its computational complexity is relatively low, and it works very efficiently at a low coding delay.

The new method is also applied to a variable-rate segment vocoder for which an extended optimal segmentation algorithm is proposed. In addition to conventional methods, the algorithm optimizes the segment rate, and reduces the number of frames with extreme distortion. Best results are achieved by performing a voicing dependent MPQ.

REFERENCES

[1] D. P. Kemp, J. S. Collura, T. E. Tremain, "Multi-frame coding of LPC parameters at 600-800 bps", Proc. IEEE ICASSP '91, pp. 609-612, Toronto, 1991

[2] B. M. Mouy, P. E. de La Noue, "Voice transmission at a very low bit rate on a noisy channel", Proc. IEEE ICASSP '92, pp. 149-152, San Francisco, 1992

[3] S. Roucos, R. M. Schwartz, J. Makhoul, "Segment Quantization for Very-Low-Rate Speech Coding", Proc. IEEE ICASSP '82, pp. 1565-1568, Paris, 1982

[4] Y. Shiraki, M. Honda, "LPC Speech Coding Based on Variable-Length Segment Quantization", IEEE Trans. ASSP, Vol. 36, No. 9, Sep. 1988

[5] P. A. Chou, T. Lookabaugh, "Variable Dimension Vector Quantization of Linear Predictive Coefficients of Speech", Proc. IEEE ICASSP '94, pp. 505-508, Adelaide, 1994

[6] J. M. López-Soler, N. Farvardin, "A Combined Quantization-Interpolation Scheme for Very Low Bit Rate Coding of Speech LSP Parameters", Proc. IEEE ICASSP '93, Minneapolis, 1993

[7] S. Bruhn, "Efficient Interblock Noiseless Coding of Speech LPC Parameters", Proc. IEEE ICASSP '94, pp. 501-504, Adelaide, 1994

[8] K. K. Paliwal, B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame", IEEE Trans. SAP, Vol 1, No. 1, pp. 3-14, Jan. 1993

[9] M. J. Sabin, R. M. Gray, "Product Code Vector Quantizers for Waveform and Voice Coding", IEEE Trans. ASSP, vol. ASSP-32, pp. 474-488, June 1984

[10] C. Tsao, R. M. Gray, "Shape-Gain Matrix Quantizers for LPC Speech", IEEE Trans. ASSP., Vol. ASSP-34, No. 6, pp. 1427-1439, Dec. 1986