Low-Complexity Predictive Trellis Coded Quantization of Wideband Speech LSF parameters

Yongwon Shin[1], Sangwon Kang[2], Thomas R. Fischer[3], Changyong Son[4], and Yongbeom Lee[4]

[1]Telecommunication R&D Center, Samsung Electronics Co., Suwon, Korea, yw74.shin@samsung.com
[2]School of Elec. Eng. and Comp. Science, Hanyang Univ., Ansan, Korea
[3]School of Elec. Eng. and Comp. Science, Washington State Univ., Pullman, WA99163, USA
[4]Human & Computer Interaction Lab., Samsung Advanced Inst. of Tech., Kiheung-Eup, Yongin, Korea

ABSTRACT

In this paper, low-complexity block-constrained trellis coded quantization (BC-TCQ) structures are introduced, and a predictive BC-TCQ encoding method is developed for quantization of line spectrum frequencies (LSF) parameters for wideband speech coding applications. The performance is compared to the linear predictive coding (LPC) vector quantizers used in the AMR-WB (ITU-G.722.2) speech coding standard, demonstrating reduction in spectral distortion and significant reduction in encoding complexity.

1. INTRODUCTION

Linear prediction is commonly used in low bit-rate speech coding systems. The linear prediction coefficients can be represented as line spectral frequency (LSF) parameters [1] for efficient quantization and coding. Various methods have been proposed to encode LSF parameters using either scalar quantization (e.g., [2]) or some form of vector quantization (e.g., [3, 4]).

Trellis coded quantization (TCQ) [5] is a form of vector quantization that builds the VQ codebook from interleaved constituent scalar quantization codebooks and a trellis structure defined by a convolutional code. The Viterbi algorithm (VA) [6] is used for searching the trellis paths for optimum encoding. The TCQ complexity is modest compared to unstructured VQ. In traditional TCQ, the initial trellis state is encoded as side information, which is negligible additional rate for large-dimensional source vectors. For encoding small-dimensional vectors, it is important, in a rate-distortion sense, to allow the initial trellis state to vary with the source vector, but the overhead of signaling the initial trellis state is no longer of negligible rate.

A "tail-biting" TCQ formulation was recently proposed by Nikneshan and Khandani [7]. This approach requires exactly one bit per source sample to specify the trellis path, and provides rate-distortion performance competitive with VQ for relatively small source vector dimension. Using variations on the tail-biting theme, this paper develops low-complexity block-constrained TCQ (BC-TCQ) structures. A predictive BC-TCQ system is then designed for encoding LSF parameters. The performance is evaluated for wideband speech, and compared to the split and multi-stage VQ used in [8].

Supported by Samsung Advanced Institute of Technology.

2. BLOCK-CONSTRAINED TCQ

2.1. Structure and Complexity

We consider TCQ based on a rate-1/2 convolutional code, with feedback-free encoder, having $N = 2^{\nu}$ trellis states and 2 branches entering/leaving each trellis state [5]. Given a block of n source samples, the VA is used to find the minimum distortion path. This encoding procedure allows the best trellis path to begin in any of N initial states and end in any of N terminal states. In fixed-rate coding of a block of n source samples, the transmitted information would include the initial trellis state, plus one bit per state transition for the n stages through the trellis, for a total of $n+\nu$ bits for the trellis path information. We refer to this situation as basic TCQ encoding.

Nikneshan and Khandani [7] propose a tail-biting TCQ formulation. In the general case, this formulation associates a single terminal state with each initial trellis state. The Viterbi algorithm encoding is then performed N times, once for each of the N initial trellis states, and the minimum distortion initial state and path is selected. Since there is a single terminal state allowed for any given initial state, the final ν path bits are redundant information, because there is a unique path through the trellis from each state at stage $n-\nu$ to a specified terminal state at stage n. The obvious disadvantage of such an encoding method, compared to basic TCQ, is the increase in complexity. We refer to this as the "maximal complexity" block-constrained TCQ structure, since the number of Viterbi algorithm searches of the trellis equals the number of trellis states, and will refer to this encoding method as TB-TCQ in the remainder of the paper. In [7] the experimental results set the initial and terminal states to be identical, yielding the "tail-biting" structure common in convolutional encoding. Fewer than N initial trellis states are used in the numerical evaluations in [7] (typically just two initial states), presumably to reduce the encoding complexity.

For any $0 \le k \le \nu$, consider a block-constrained TCQ (BC-TCQ) structure that allows 2^k initial trellis states, and for each allowed initial trellis state exactly $2^{\nu-k}$ terminal trellis states. A single VA encoding starting from the allowed initial trellis states proceeds in the normal way up to time stage $n-k$. It takes k bits to specify the initial state, and $n-k$ bits to specify the path to time stage $n-k$. A unique terminating path, possibly dependent on the initial trellis state, is pre-specified for each trellis state at time stage $n-k$ through time stage n. Regardless of the value of k, the encoding complexity is only a single VA search of the trellis, and exactly n bits are required to specify an initial trellis state and path through the trellis.

2.2. Encoding Performance

In predictive coding of LSF parameters, the prediction errors are roughly Laplacian distributed [9]. This motivates us to use memoryless Laplacian source for design and evaluation of BC-TCQ. In [5] it was found that a reproduction codebook of size 2^{2+R} is sufficient for near-optimal encoding at a rate of R bits per sample, but a codebook of size 2^{1+R} is not large enough to obtain near-optimum encoding performance. This observation led to fixed-rate TCQ encoding of a Laplacian source using a "quadrupled sized" codebook of size 2^{2+R}. Following that formulation, a 16-state BC-TCQ encoder was designed using 8 codebook subsets, each with 4 codewords, for 3 bit/sample encoding of a memoryless Laplacian source. The trellis was populated with subsets, so that 4 of the subsets were used on 28 of the 32 trellis branches, and the remaining 4 subsets each assigned to a single branch. The codebook was optimized using a training set and the generalized Lloyd algorithm [10] for trellis code design. Using the minimal complexity encoding method outlined above, various combinations of initial and terminal states were studied, with encoding performance summarized in Table 1. The best performing design used 4 initial states, s, with 4 final states, f, for each initial state given by $s = 4i$, $f = 4i + j$, $i = 0,1,2,3$; $j = 0,1,2,3$. This design (using the parameter $k = 2$ in Section 2.1) is used in all subsequent results labeled BC-TCQ in the paper.

The BC-TCQ performance is summarized in Tables 2 for various source block lengths and memoryless Laplacian source at encoding rate of 3 bits/sample. Also presented in Table 2 are the "maximal complexity" TB-TCQ performances (using 16 VA trellis searches). The BC-TCQ performance is consistent with the performance reported in [7] and competitive with the other methods.

TABLE 1 SNR PERFORMANCE OF BC-TCQ OF A LAPLACIAN SOURCE FOR DIFFERENT NUMBER OF INITIAL STATES

SNR (dB)	Block size			
The number of initial states	L=8	L=16	L=32	L=64
1	13.6287	14.4819	15.1030	15.5636
2	14.7567	15.2100	15.5808	15.8499
4	14.9591	15.4942	15.7731	15.9887
8	13.4285	14.5864	15.3346	15.7704
16	11.6558	13.2499	14.4951	15.2912

TABLE 2 TB-TCQ AND BC-TCQ QUANTIZATION PERFORMANCE FOR THE MEMORYLESS LAPLACIAN SOURCE USING RATE 3 BITS/SAMPLE , QUADRUPLED AND OPTIMIZED CODEBOOK, AND 16-STATE TRELLIS. (VALUES ARE LISTED AS SNR IN DB)

SNR (dB)	Block Length				
	8	16	32	64	128
TB-TCQ	15.43	16.16	16.19	16.23	16.26
BC-TCQ	14.96	15.49	15.77	15.99	16.13

3. PREDICTIVE BC-TCQ FOR LSF CODING

There is significant dependence between consecutive LSF vectors, and between elements within an LSF vector. To exploit these dependencies, a predictive BC-TCQ structure is proposed. Methods are presented for designing and optimizing the predictive BC-TCQ for coding of LSF coefficients and the encoding complexity is characterized.

3.1. Design Equations

The highest coding efficiency is achieved only when both inter-frame and intra-frame redundancy is removed. In this paper, a moving average (MA) method, that is more resilient to noisy channels than an auto-regressive (AR) method [11], is used for inter-frame linear prediction in LSF coding systems. In addition, AR method is used to make use of intra-frame dependency existing in the MA inter-prediction error vectors. We use a 4th order MA inter-predictor (orders of 3 to 5 are required to obtain good performance [11]) and 1st order AR intra-predictor, as shown in 'Section A' of Fig. 1.

Joint optimization of MA predictors and VQ codebooks is introduced in [12]. As described in [12], the prediction coefficients and the residuals may exchange a scale factor, regardless of the mean-removed LSF vector. Therefore, the constraint on the prediction coefficients is

$$\sum_{i=0}^{Q} \mu_i = 1 . \quad (1)$$

The optimal BC-TCQ codebooks and intra-prediction coefficients for the prediction error residual are designed for each MA predictor satisfying (1). The best inter-predictor, optimized codebooks, and intra-prediction coefficients that minimize distortion between the LSF vectors and quantized LSF vectors are selected as the final design. As in predictive TCQ [5], intra-predictive BC-TCQ implies trellis-state-dependent prediction must be used. A prediction can be formed for each trellis node using the output symbols specified by the survivor path associated with the particular node.

In order to get more robustness against "outliers"; i.e. input LSF vectors having low correlation within the previous vector, we use a "safety-net" structure [13] with predictive BC-TCQ, as shown in Fig. 1. There still exists intra-frame correlation in the outlier vectors, and so an intra-predictive coding scheme is retained. In the safety-net structure, the input LSF vector is quantized using two modes, one with MA inter-frame prediction and the other without inter-frame prediction.

In order to design and optimize the predictive BC-TCQ with safety-net, we proceed as follows. Assume a given inter-MA predictor. Firstly, design the initial codebooks and intra-prediction coefficients for the two different modes using a training sequence. Secondly, classify the training sequence into two sub-training sequences, depending on which mode produces a lower MSE. Thirdly, update intra-prediction coefficients and codebooks of each mode using these classified sub-training sequences. Fourthly, compute the average distortion $d(k)$ and the fractional change in distortion $Thr = \left|d(k-1) - d(k)\right| / \left|d(k-1)\right|$, where k is the current iteration. If Thr is less than a given criteria, stop the process. Otherwise start again from the generation of sub-training sequence.

The average distortion is not guaranteed to decrease monotonically with each iteration. Typically, a large decrease of distortion is obtained in the first few iterations. The optimization process can be halted by a suitable termination criteria. In this paper, the process is stopped if $Thr < 0.0001$. The codebooks and intra-prediction coefficients which produce the minimum distortion during iterations are selected as the "optimal" codebooks and intra-prediction coefficients. The above design processes are performed for each of a set of allowed MA prediction coefficients. The design generating the least distortion is selected.

3.2. Complexity Analysis

With the predictive BC-TCQ, a single Viterbi algorithm encoding is performed for the allowed trellis paths. The computational complexity of BC-TCQ depends on several design factors, including the number of trellis states, the dimension of the target vector, the number of initial states, and the number of the final states for each initial state. The proposed LSF quantizer uses BC-TCQ with 4 initial states and 4 final states for each initial state. The encoding complexity of the predictive BC-TCQ with safety-net for 16th order LSF vectors is shown in Table 6. The analysis assumes a 16-state trellis with integer R_k bits used to encode the k th vector element.

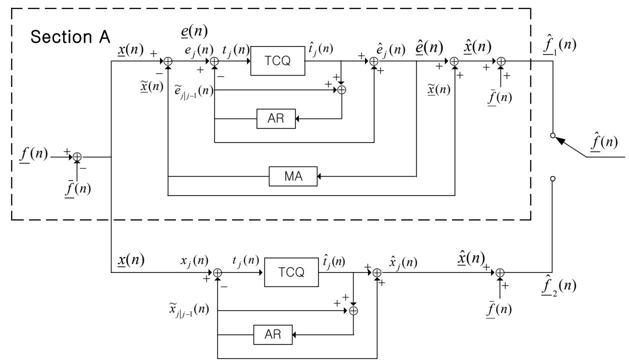

Fig. 1. Predictive BC-TCQ with safety net.

4. CODING RESULTS

The predictive BC-TCQ uses the 16-state quadrupled-output alphabet trellis structure. The database used in the design and testing consists of 26 minutes of sampled speech. The first 13 minutes are used for design, and the last 13 minutes are used for testing. The speech samples were recorded from English female, English male, Korean female, and Korean male speakers. The spectral distortion (SD) [3] is used to evaluate the performance of the LSF quantizer. The weighted Euclidean distance measure presented in [14] is used for trellis encoding. The performance of the predictive BC-TCQ is compared with the LPC quantization performance of a contemporary speech coder: the adaptive multi-rate wideband (AMR-WB) speech coding standard [8]. The predictive BC-TCQ performance is evaluated at different encoding rates. The pre-processing and LP analysis of the AMR-WB coder were used to generate the LSF vectors encoded using the predictive BC-TCQ system, so that the comparison with the AMR-WB S-MSVQ method would be under identical conditions. The bit allocation of the predictive BC-TCQ (at an encoding rate of 46 bits/frame) is shown in Table 3.

Table 4 compares the performance of the predictive BC-TCQ and AMR-WB S-MSVQ methods. Table 5 lists the SD performance of the predictive BC-TCQ at different bit rates. As the bit rate decreases, the bit allocation is modified from that listed in Table 3 by reducing the number of bits allocated to the higher frequency LSF parameters. Table 6 lists the computational complexity and ROM requirement of the predictive BC-TCQ and the AMR-WB S-MSVQ (for an encoding rate of 46 bits/vector).

TABLE 3 BIT ALLOCATION OF THE PREDICTIVE BC-TCQ FOR WIDEBAND SPEECH OF 20MS FRAME

Parameters	Bit allocation
Path information (Initial states + Path + Final states)	2+12+2
Subset codewords	2 X 13(stages 1 to 13) 1 X 3(stages 14 to 16)
Safety net information	1
Total	46

TABLE 4 SD COMPARISON OF THE PREDICTIVE BC-TCQ AND THE AMR-WB S-MSVQ (AT 46 BITS)

	SD		
	Ave SD (dB)	Outliers (%)	
		2dB ~ 4dB	> 4dB
Predictive BC-TCQ	0.6908	0.1453	0
AMR-WB S-MSVQ	0.7933	0.4099	0.0026

TABLE 5 SD PERFORMANCE OF PREDICTIVE BC-TCQ, FOR 16TH ORDER LSF, AT DIFFERENT BIT RATES

Bits	34	36	38	40
Ave SD(dB)	1.0109	0.9424	0.8910	0.8491
2dB~4dB(%)	2.8638	1.9429	1.3100	0.9001
>4dB(%)	0.0052	0.0026	0.0026	0.0026
Bits	42	44	46	
Ave SD(dB)	0.8035	0.7543	0.6908	
2dB~4dB(%)	0.6667	0.4617	0.1453	
>4dB(%)	0	0	0	

TABLE 6 COMPUTATIONAL COMPLEXITY AND ROM REQUIREMENT OF THE PREDICTIVE BC-TCQ AND THE AMR-WB S-MSVQ (AT 46 BITS)

	Operation	Predictive BC-TCQ	AMR-WB S-MSVQ
Complexty	Add.	3784	15624
	Mult.	2968	8832
	Comp.	2335	3570
ROM (words)		928	5280

5. CONCLUSIONS

A "tail-biting" TCQ approach requires exactly one bit per source sample to specify the trellis path, and provides rate-distortion performance competitive with VQ for relatively small source vector dimension. Using variations on the tail-biting theme, this paper developed low-complexity, block-constrained TCQ (BC-TCQ) structures. A predictive BC-TCQ encoding method is developed for quantization of LSF parameters for speech coding applications. The performance is compared to the LPC vector quantizers used in the wideband AMR-WB (ITU-G.722.2) speech coding standard, demonstrating reduction in spectral distortion and significant reduction in encoding complexity. The predictive BC-TCQ can quantize LPC parameters with transparent quality using 34 bits/vector for wide-band speech. The predictive BC-TCQ implementation complexity is 40% of the complexity of the LSF vector quantization methods used in the AMR-WB standard.

6. REFERENCES

[1] F. Itakura, "Line spectral representation of linear predictive coefficients of speech signals," *J. Acoust. Soc. Amer.*, vol. 57, Supplement no. 1, S35, 1975.
[2] N. Sugamura and N. Farvardin, "Quantizer design in LSP speech analysis-synthesis," *IEEE J. Sel. Areas Commun.*, SAC-6, pp. 432-440, 1988.
[3] K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameter at 24 bits/frame," *IEEE Trans. Speech Audio Processing*, vol. 1, no. 1, pp. 3-14, Jan. 1993.
[4] A. Gersho and R. M. Gray, *Vector Quantization and Signal Compression*. Norwell, MA: Kluwer, 1992.
[5] M. W. Marcellin and T. R. Fischer, "Trellis coded quantization of memoryless and Gauss-Markov sources," *IEEE Trans. Communications*, vol. 38, issue 1, pp. 82-93, Jan. 1990.
[6] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE (Invited Paper)*, vol. 61, pp. 268-278, Mar. 1973.
[7] S. Nikneshan and A. K. Khandani, "Soft decision decoding of fixed-rate entropy-coded trellis coded quantizer over a noisy channel," *IEEE Trans. Vehicular Tech.*, to appear.
[8] 3GPP TS 26.190 (V5.1.0.), "AMR Wideband speech codec; Transcoding functions (Release 5)," Dec. 2001.
[9] S. Chi, S. Kang, and C. Lee, "Safety-net pyramid VQ of LSF parameters for wideband speech codecs," Electronics Letters, vol. 37, no. 11, pp. 724-726, May 2001.
[10] S. P. Lloyd, "Least squares quantization in PCM," unpublished memorandum, Bell Laboratories, 1957; see also *IEEE Trans. Inform. Theory*, vol. IT-28, pp. 129-137, 1982.
[11] J. Skoglund. and J. Linden, "Predictive VQ for noisy channel spectrum coding: AR or MA?," *Acoustics, Speech, and Signal Processing, ICASSP-97, 1997 IEEE International Conference*, vol. 2, pp. 1351-1354, 1997.
[12] Y. Shoham, "Coding the line spectral frequencies by jointly optimized MA prediction and vector quantization," *IEEE Worksho. Speech Coding Proceedings,* pp. 46 -48, 1999.
[13] T. Eriksson, J. Linden, and J. Skoglund, "Exploiting inter-frame correlation in spectral quantization: a study of different memory VQ schemes," *Acoustics, Speech, and Signal Processing, ICASSP-96 Conference Proceeding, IEEE International Conference*, vol. 2, pp. 765 -768, 1996.
[14] TIA/EIA/IS641, "TDMA cellular/PCS radio interface-enhanced full-rate speech codec," May 1996.