Block Constrained Trellis Coded Vector Quantization of LSF Parameters for Wideband Speech Codecs

Jungeun Park and Sangwon Kang

ABSTRACT—In this paper, block constrained trellis coded vector quantization (BC-TCVQ) is presented for quantizing the line spectrum frequency parameters of the wideband speech codec. Both a predictive structure and a safety-net concept are combined into BC-TCVQ to develop the predictive BC-TCVQ. The performance of this quantization is compared with that of the linear predictive coding vector quantizer used in the AMR-WB codec, demonstrating reductions in spectral distortion.

Keywords—Speech coding, quantization, trellis coded vector quantization, LSF parameters.

I. Introduction

Linear predictive coding (LPC) parameters, which describe the short-term spectral envelope of speech, can be represented as line spectral frequency (LSF) parameters for efficient quantization and coding. Various methods have been proposed to encode LSF parameters using either scalar quantization or some form of vector quantization (VQ).

Trellis coded quantization (TCQ) [1] is a form of VQ that builds the VQ codebook from an interleaved constituent scalar quantization codebook and a trellis structure defined by a convolutional code. The Viterbi algorithm (VA) [2] is used to search 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 an additional rate for source vectors.

Block constrained TCQ (BC-TCQ) [3] was proposed by Kang and others. This approach requires exactly one bit per source sample to specify the trellis path with low complexity.

Trellis coded VQ (TCVQ) [4] generalizes TCQ to allow vector codebooks and branch labels. The main feature of TCVQ is the partitioning of an expanded set of VQ symbols into subsets and the labeling of the trellis branches with these subsets.

In this paper, TCVQ is combined with BC-TCQ to develop block constrained TCVQ (BC-TCVQ) structured for a vector dimension of 16. The predictive BC-TCVQ system is also designed for encoding LSF parameters.

II. TCVQ

Although the performance of a TCQ coder is in several cases close to the theoretical rate-distortion bound, an improvement is always possible by generalizing its structure to the vector case. For a given rate, TCVQ yields lower distortion than TCQ at the cost of an increase in implementation complexity [4]. In addition, TCVQ allows fractional rates, while TCQ does not. The structure of TCVQ is quite similar to that of scalar TCQ. We consider TCVQ based on a rate-1/2 convolutional code, which has $N=2^{v}$ trellis states and two branches entering/leaving each trellis state. Given a block of m source vectors, 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 m source vectors, the transmitted information would include the initial trellis state, plus one bit per state transition for m stages through the trellis, with a total of $v+m$ bits for the trellis path information.

In TCVQ, the codebook has $2^{(R+\tilde{R})L}$ vector codewords. We refer to $\tilde{R}$ as the (super) "codebook expansion factor" (in bits per dimension) since the codebook has $2^{\tilde{R}L}$ times as many codewords as a nominal rate-R VQ. The encoding is accomplished in two steps.

Step 1. For each input vector, find the closest codeword and

Manuscript received May 1, 2008; revised July 3, 2008; accepted Aug. 20, 2008.

This work was supported by the Basic Research Program of the Korea Science & Engineering Foundation (grant no. R01-2006-000-10885-0).

Jungeun Park (phone: + 82 31 400 4035, email: nesjh@naver.com) and Sangwon Kang (phone: + 82 31 400 5215, email: swkang@hanyang.ac.kr) are with the School of Electrical Engineering and Computer Science, Hanyang University, Ansan, Rep. of Korea.

Table 1. Characteristics of the 16-state quadrupled output alphabet trellis for encoding a memoryless Laplacian source.

Current state	0		1		2		3		4		5		6		7		8		9		10		11		12		13		14		15	
Previous states	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
Associated subsets	4	2	2	0	1	3	3	6	0	2	2	0	1	3	3	1	2	0	0	2	3	1	1	3	7	0	0	2	3	1	1	5

corresponding distortion in each subset.

Step 2. Let the branch metric for a branch labeled with subset S be the distortion found in step 1 and use the VA to find the minimum distortion path through the trellis.

The main steps for the design of the TCVQ encoder are the following: constructing the initial codebook, partitioning the codebook into subcodebooks (subsets), labeling the trellis branches with these subsets, and TCVQ encoding. The initial codebook used in TCVQ plays an important role in the performance of the coder. A good choice for the TCVQ initial codebook is the full search VQ codebook designed by the LBG algorithm [5]. For an L-dimensional TCVQ encoder of rate R (in bits per dimension), the initial extended codebook has $2^{(R+\tilde{R})L}$ codevectors, but only a subset of size 2^{RL} of these codevectors may be used to represent a source vector at any instant. The TCVQ initial codebook which satisfies the rules in [4] is designed in two steps.

Step 1. A basic codebook of size 2^{RL} is designed by the LBG algorithm.

Step 2. An initial extended codebook of size $2^{(R+\tilde{R})L}$ is constructed based on regular points on a circle with radius ∂, where the center is each codeword of the basic codebook.

III. BC-TCVQ

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

In predictive coding of LSF parameters, the prediction errors are roughly Laplacian distributed [3]. In [1], it was found that fixed-rate TCQ encoding of a Laplacian source benefited from using a quadruple-sized codebook. Following that formulation, a 16-state BC-TCVQ encoder was designed using 8 codebook subsets, each with 16 code words, for 3 bit/sample encoding of a memoryless Laplacian source. The trellis was populated with subsets following the method in [1], and the remaining 4 subsets were each assigned to a single branch, as listed in Table 1. The codebook was optimized using a training set and the generalized Lloyd algorithm.

Table 2. SNR comparison of BC-TCQ and BC-TCVQ for a memoryless Laplacian source using a rate of 3 bit/sample.

Quantization methods	Signal-to-noise ratio (SNR)	
	Sample length=16	Sample length=32
BC-TCQ	15.22 dB	15.28 dB
BC-TCVQ	16.44 dB	16.54 dB

The SNR performances of BC-TCQ and BC-TCVQ are compared in Table 2 for two sample lengths for a memoryless Laplacian source with an encoding rate of 3 bit/sample. BC-TCVQ uses a vector dimension of 2. From the table, BC-TCVQ provides better SNR performance than BC-TCQ.

IV. Predictive BC-TCVQ

There is a significant degree of dependency, referred to as interframe dependency, between consecutive LSF vectors. To exploit this dependency, a predictive BC-TCVQ structure is proposed. Methods are presented for designing and optimizing a predictive BC-TCVQ for the coding of LSF coefficients and the encoding complexity is characterized. The predictive BC-TCVQ encoding structure is illustrated in Fig 1. Let $\mathbf{f}(n)$ and $\mathbf{r}(n)$ be the p-dimensional LSF vector and prediction error vector at time n, respectively.

A fourth-order MA predictor is used to compute the prediction error vector [3] and it uses a large correlation of LSF parameter vectors to design the LSF quantization system.

There are occasions, however, when rapid changes in LSF traces are evident; thus, a small interframe correlation is present. Therefore, we suggest the use of quantization methods that exploit memory within an LSF vector but are memoryless between vectors in a *safety-net* structure [6]. In general, the safety-net structure provides better robustness against *outliers*,

ETRI Journal, Volume 30, Number 5, October 2008 Jungeun Park et al. 739

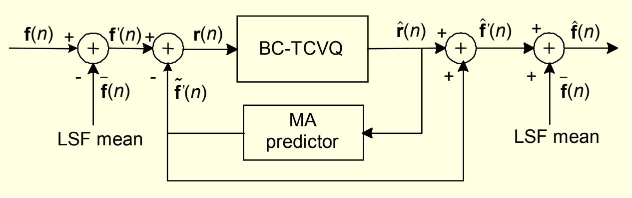

Fig. 1. Predictive BC-TCVQ with a 4th-order MA predictor.

that is, input LSF vectors having a low correlation with the previous vector. We use this safety-net formulation with predictive BC-TCVQ.

In the safety-net structure, the input LSF vector is quantized using both *MA interframe prediction* and *no interframe prediction*. Then, both quantized vectors are compared to the input vector and the better reproduction is selected. The BC-TCVQ information from the selected scheme is transmitted to the decoder, along with a signaling bit that indicates the selected mode.

V. Experimental Results

The database used for the design and test included 26 minutes of sample speech. The first 13 minutes were used for the design, and the last 13 minutes were used for the test. Spectral distortion (SD) was used to evaluate the performance of the LSF quantizers. The performance of predictive BC-TCVQ is compared with the LPC quantization performance of the adaptive multirate wideband (AMR-WB) speech coding standard (ITU-G.722.2) [7]. The AMR-WB speech coder uses split and multistage VQ (S-MSVQ) at 46 bit/vector to encode 16 LSF parameters. The bit allocation for predictive BC-TCVQ (at an encoding rate of 46 bit/frame) is shown in Table 3.

Tables 4 and 5 compare the SD performance and the computational complexity, respectively, of the AMR-WB S-MSVQ and the predictive BC-TCVQ (at an encoding rate of 46 bit/frame). From the results presented in Tables 4 and 5, it is clear that the proposed predictive BC-TCVQ affords better SD performance than the reference method, with a similar degree of computational complexity.

Table 3. Bit allocation of predictive BC-TCVQ for a 20 ms frame of wideband speech.

Parameters	Bit allocation
Path information (initial states + path + final states)	2+4+2
Subset codewords	5×5 (stages 1 to 5) 4×3 (stages 6 to 8)
Safety-net information	1
Total	46

Table 4. SD comparison of the AMR-WB S-MSVQ and the predictive BC-TCVQ (at 46 bit/frame).

	AMR-WB S-MSVQ	Predictive BC-TCVQ
Avg. SD (dB)	0.7933	0.6807
3 dB to 5 dB (%)	0.4099	0.2412
> 5 dB (%)	0.0026	0

Table 5. Computational complexity of AMR-WB S-MSVQ and predictive BC-TCVQ (at 46 bit/frame).

Operation	AMR-WB S-MSVQ	Predictive BC-TCVQ
Addition	15,624	18,624
Multiplication	8,832	6,208
Comparison	3,570	3,400
Total	28,026	28,232

VI. Conclusion

TCVQ was combined with BC-TCQ to develop BC-TCVQ. Predictive BC-TCVQ was proposed for quantizing LSF parameters for wideband speech. The performance of predictive BC-TCVQ was compared to that of the LPC vector quantizer used in the AMR-WB speech coding standard (ITU-G.722.2), and reductions in spectral distortion were demonstrated.

References

[1] M.W. Marcellin and T.R. Fischer, "Trellis Coded Quantization of Memoryless and Gauss-Markov Sources," *IEEE Trans. Comm.*, vol. 38, no. 1, Jan. 1990, pp. 82-93.

[2] G.D. Forney Jr., "The Viterbi Algorithm," *Proc. IEEE*, vol. 61, Mar. 1973, pp. 268-278.

[3] S. Kang, Y. Shin, and T.R. Fischer, "Low-Complexity Predictive Trellis-Coded Quantization of Speech Line Spectral Frequencies," *IEEE Trans. Signal Processing*, vol. 52, no. 7, July 2004, pp. 2070-2079.

[4] T.R. Fischer, M.W. Marcellin, and M. Wang, "Trellis Coded Vector Quantization," *IEEE Trans. Inform. Theory*, vol. 37, no. 6, Nov. 1991, pp. 1551-1566.

[5] Y. Linde, A. Buzo, and R.M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Trans. Comm.*, vol. COM-28, no. 1, Jan. 1980, pp. 84-95.

[6] T. Eriksson, J. Linden, and J. Skoglund, "Exploiting Interframe Correlation in Spectral Quantization: A Study of Different Memory VQ Schemes," *ICASSP*, vol. 2, 1996, pp. 765 -768.

[7] 3GPP TS 26.190 (V5.1.0, 2001-12): *AMR Wideband Speech Codec: Transcoding Functions,* Release 5, 2001.