FIXED-LENGTH SEGMENT CODING OF LSF PARAMETERS

Evgeni Yakhnich and Yuval Bistritz
Department of Electrical Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel
evgeny@comsys.co.il bistritz@eng.tau. ac.il

ABSTRACT

This paper presents a method to attain very low bit-rate compression of speech spectral envelope. It is based on fixed-length segment coding. The method utilizes Temporal Decomposition (TD) technique for the compact representation of segments of *Line Spectrum Frequencies* (LSF) vector followed by split matrix quantization. The TD technique is modified to fit fixed-length segment coding. Computation is reduced by using fixed event functions and because event positions determination requires only a simple search within the short fixed-length segment. Weighted Euclidian distance is used as cost function to better approximate the spectral distance measure. The method achieves low bit rates without significant increase in computation cost. The method has been implemented on coding the spectral envelope for the MELP coder and showed its viability.

1 INTRODUCTION

The main bit budget of modern low rate vocoders goes for spectral envelope parameter coding [1]. Usually, Linear Predictive Coding (LPC) is used to model the speech spectral envelope [2] and most often the LPC vectors are transformed to Line Spectral Frequency (LSF) vectors that exhibit desirable properties for quantization [3]. Quantization of LSF vectors that reach 1 dB average distortion and transparent quality is possible with 34 bits/frame using scalar quantization to as low as 24 bits per frame [3]. These quantization methods use frame by frame quantization. Much research effort has been invested in the recent years to achieve further reduction in bit rates by exploiting interframe redundancies of spectral envelope parameters. Quantization of LSF by Matrix Quantization (MQ) was proposed in [4]. The method codes several LSF vectors from adjacent frames as a single codeword. However, the method requires large codebooks to obtain good performance. Other methods proposed to transmit quantized vectors at selected frames, and interpolation of the missing ones in the decoder [5-6]. More sophisticated methods perform segmentation of input stream of frames into variable length segments and code the segments rather than separate frames. These methods involve sophisticated and computationally heavy search techniques for segmentation and use time warping for transformation of fixed length codewords into variable length segments [7]. Trellis Segmentation-Quantization technique (TSQ) [8] combines variable-length segmentation with interpolation of missed frames in order to provide bit-rate reduction without significant performance degradation.

A different approach to exploit interframe redundancies, called Temporal Decomposition (TD), was introduced by Atal [9]. This technique attempts to represent speech by sounds or events. It seems an attractive attitude to achieve low bit rates because sound rate of human speech is about 10-15 events/sec. Each event is described by an Event Function (E-function) and Event Vector (E-vector), where the E-vector represents spectral feature vector of the event and the E-function describes its temporal behavior. The drawback of the TD technique is that it suffers from both high complexity and processing delay because it involves many iterations and uses large speech fragments.

Kim et al. [10] applied the TD technique for LSF vectors coding with emphasis on LSF order properties.

This paper describes a new method that follows a TD formalism for efficient parameterization of short fixed-length segments of LSF vectors. It uses predefined E-function shapes, in order to avoid iterative refinements and reduce bit rate. Weighted Euclidian Distance (WED) optimization criterion is applied for E-vector calculation, because a weighted Euclidean distance is known to approximate better then regular Euclidean distance the Log-Spectral Distance (LSD) for LSF feature vectors. The obtained E-Vectors are coded using Split Matrix Quantization (SMQ) and transmitted along with event location. The decoder restores the LSF vectors of the segment from the transmitted information.

The proposed approach was tested by incorporating into MELP2400 vocoder [1]. Simulations showed that it can reach 1 dB mean LSD with only moderate complexity increase.

The outline of the paper is as follows. Application of TD to fixed-length segment parameterization is described in Section 2. Section 3 considers SMQ of E-Vectors and some optimality issues. Then, Section 4 presents simulation results. Finally, benefits and drawbacks of the approach are summarized in Section 5.

2 TD OF FIXED-LENGTH SEGMENTS

The main idea of TD is to decompose speech to events, where each event is described by E-function and E-vector pair. The first one characterizes temporal behavior of event, while the second represents its spectral envelope. Moreover, each event has a location, which is the location of the corresponding E-function maximum. Let's denote the i^{th} LSF vector of some segment as $\underline{a}(i)$. Then it can be approximated by M E-functions $\varphi_k(i)$ and E-vectors $\underline{\omega}_k$ in the following way:

$$\underline{a}(i)=\sum_{k=1}^{M}\underline{\omega}_k\varphi_k(i) \qquad (1)$$

$\underline{a}(i)$ is a LSF vector reconstruction at time *i*.

Throughout this paper, linear E-functions will be used. They are equal to 1 at the event location, decline linearly to zero at adjacent event locations and stay zero elsewhere. A segment will be represented by three events: the first event is chosen at the beginning of the segment, the last event is located at the end of the segment and the location of the middle event is chosen such that it minimizes appropriate WED. This way, the first and the last events represent boundary conditions of the segment, while the middle one approximates possible transitions inside the segment. For each middle event location, corresponding E-vectors are found as described later in this section. Figure 1 depicts an original (order 10) LSF vector segment and its reconstruction using TD. E-function and E-vectors are also shown. Note that 5 frame long segment is shown and middle event is located in 3rd frame, however frames 2 and 4 are also valid locations.

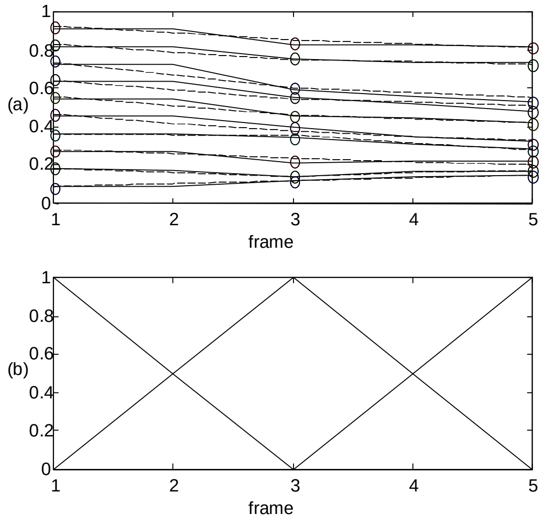

Fig.1. (a) Original LSF segment (solid line), reconstructed one (dashed line) and corresponding Event Vectors (circles); (b) corresponding Event Functions

Let us consider a general case of N-frame long segment. It can be represented by concatenation of its LSF vectors, $\underline{a}(i)$, as denoted by:

$$\underline{\alpha}=\left[\underline{a}^T(1)\ \ \ldots\ \ \underline{a}^T(N)\right]^T \qquad (2.a)$$

The segment length should be chosen to meet performance, bit rate and complexity requirements. We will show later that, for 5 frame long segments, the proposed method can significantly reduce bit rate without sacrificing performance. In a similar way, a concatenated event vector is defined:

$$\underline{\Omega}=\left[\underline{\omega}^T(1)\ \ \ldots\ \ \underline{\omega}^T(3)\right]^T \qquad (2.b)$$

as described above, each segment is decomposed into 3 events.

The reconstructed segment vector $\underline{\alpha}$ can be rewritten, according to equation (1), as follows:

$$\underline{\alpha}=\Phi(l)\underline{\Omega} \qquad (3)$$

where l is the middle segment location. $\Phi(l)$ is a *Np*x*3p* matrix that has the following form

$$\Phi(l)=\begin{bmatrix}\varphi_1(1)\cdot I_{p\times p} & 0 & 0\\ \ldots & \ldots & \ldots\\ \varphi_1(l)\cdot I_{p\times p} & \varphi_2(l)\cdot I_{p\times p} & 0\\ \ldots & \ldots & \ldots\\ 0 & \varphi_2(N)\cdot I_{p\times p} & \varphi_3(N)\cdot I_{p\times p}\end{bmatrix} \qquad (4)$$

where $I_{p\times p}$ is unit matrix of size that corresponds to the length *p* of the LSF vector. Since we use predefined linear E-functions, the E-function matrix, $\Phi(l)$, is completely defined by the middle event location.

The notation described above differs from usual TD formalism [9] in a way that enables the use of weights in the calculation of E-vector.

Define cost function as an accumulated weighted distance of the vectors in the segment:

$$C=\sum_{i=1}^{N}\left(\underline{a}(i)-\underline{a}(i)\right)^T W(i)\left(\underline{a}(i)-\underline{a}(i)\right) \qquad (5)$$

where *W(i)* is a weighting matrix for i^{th} LSF vectors in the segment. Note that diagonal matrixes are used for LSF vector distance definition almost always (*inverse harmonic measure,* etc.). Hence, throughout this paper, the weighting matrix is assumed to be diagonal without loss of generality.

The cost can also be written as

$$C=\left(\underline{\alpha}-\underline{\alpha}\right)^T\tilde{W}\left(\underline{\alpha}-\underline{\alpha}\right) \qquad (6)$$

where

$$\tilde{W}=\begin{bmatrix}W(1) & & 0\\ & W(2) & \\ & & \ddots & \\ 0 & & & W(N)\end{bmatrix} \qquad (7)$$

The determination of the E-vectors becomes a standard weighted least squares problem and its solution is

$$\underline{\Omega}=\left(\Phi^T(l)\tilde{W}\Phi(l)\right)^{-1}\Phi^T(l)\tilde{W}\underline{\alpha} \qquad (8)$$

However, the solution in (8) involves inversion of *3p*x*3p* sized matrix, which is defined below:

$$B=\Phi^T(l)\tilde{W}\Phi(l) \qquad (9)$$

whose inversion is computationally expensive. However, its structure can be exploited in order to reduce the computational complexity of its inversion. The matrix has only few non-zero diagonals, including the main one. They are separated by p-1 zero diagonals. Let us define following matrixes:

$$B_j(i,k) = B(j+(i-1)p,\, j+(k-1));$$
$$j = 1,...,p \ \textbf{ and } i,\ k = 1,2,3 \qquad (10)$$

B_j are 3x3 matrixes, which include all the non-zero elements of B. The following result can be proven:

Proposition 1*. The inverse of the matrix B is composed of inverses of the matrices B_j replacing the locations of B_j in B, and zeros elsewhere.*

Thus the inversion of the *3p*x*3p* matrix can be carried out by *p* inversion of *3*x*3* matrices (complexity reduction by factor of 100 for p=10).

3 SPLIT MATRIX QUANTIZATION

The quantization of the E-vectors has to be carried out such that the WED cost (5) is minimized. A similar problem is addressed in some other techniques such as Trellis Quantization, TQ [7]. There, variable-length segments are coded using fixed-length codewords. The codewords are linearly warped to segment length (in our case, 3 E-vectors are warped to segment length). For a given segment, a corresponding codeword is chosen by warping all codebook entries and calculating accumulated WED for each of them. Here we propose instead an essentially new approach. The next observation will allow the calculation of the WED cost using only the E-vectors without reconstructing the segment using each one of the codewords.

Proposition 2. *Accumulated WED is a sum of TD distortion and quantization distortion.*

Proof*:* Let's say $\underline{\Omega}'$ is a codeword. Then accumulated distortion can be expressed as follows:

$$C = \left(\left(\underline{\alpha} - \underline{\alpha}\right) + \left(\underline{\alpha} - \Phi(l)\underline{\Omega}' \right) \right)^T \tilde{W} \cdot$$
$$\cdot \left(\left(\underline{\alpha} - \underline{\alpha}\right) + \left(\underline{\alpha} - \Phi(l)\underline{\Omega}' \right) \right) =$$
$$= \left(\underline{\alpha} - \underline{\alpha}\right)^T \tilde{W} \left(\underline{\alpha} - \underline{\alpha}\right) + \qquad (11)$$
$$+ \left(\underline{\Omega} - \underline{\Omega}' \right)^T \Phi(l)^T \tilde{W} \Phi(l) \left(\underline{\Omega} - \underline{\Omega}' \right) +$$
$$+ 2\left(\underline{\alpha} - \underline{\alpha}\right)^T \tilde{W} \Phi(l) \left(\underline{\Omega} - \underline{\Omega}' \right)$$

The first term on the right-hand side of (11) is a distortion due to TD, the second one is a distortion due to quantization. However, the last term involves both TD and quantization errors. We will show hereafter that this term is zero.

$$\left(\underline{\alpha}-\underline{\alpha}\right)^T \tilde{W}\Phi(l)\left(\underline{\Omega}-\underline{\Omega}' \right) =$$
$$= \left(\underline{\alpha} - \Phi(l)\left(\Phi^T(l)\tilde{W}\Phi(l)\right)^{-1}\Phi^T(l)\tilde{W}\underline{\alpha} \right)^T \tilde{W}\Phi(l)\left(\underline{\Omega}-\underline{\Omega}' \right) = \qquad (12)$$
$$= \underline{\alpha}^T \left(I - \Phi(l)\left(\Phi^T(l)\tilde{W}\Phi(l)\right)^{-1}\Phi^T(l)\tilde{W} \right)^T \tilde{W}\Phi(l)\left(\underline{\Omega}-\underline{\Omega}' \right) =$$
$$= \underline{\alpha}^T \left(\tilde{W}\Phi(l) - \tilde{W}^T\Phi(l)\left(\Phi^T(l)\tilde{W}\Phi(l)\right)^{-1}\Phi^T(l)\tilde{W}\Phi(l) \right) \cdot$$
$$\cdot \left(\underline{\Omega}-\underline{\Omega}' \right) = \underline{\alpha}^T (\tilde{W}\Phi(l) - \tilde{W}^T\Phi(l))\left(\underline{\Omega}-\underline{\Omega}' \right) = \qquad (12)$$
$$= \underline{\alpha}^T (\tilde{W}\Phi(l) - \tilde{W}\Phi(l))\left(\underline{\Omega}-\underline{\Omega}' \right) = 0$$

The last transition is due to symmetry of the weighting matrix.

Hence, (11) takes the following form:

$$C = \left(\underline{\alpha} - \underline{\alpha}\right)^T \tilde{W} \left(\underline{\alpha} - \underline{\alpha}\right) +$$
$$+ \left(\underline{\Omega} - \underline{\Omega}' \right)^T \Phi(l)^T \tilde{W} \Phi(l) \left(\underline{\Omega} - \underline{\Omega}' \right) \qquad (13)$$

There are two important conclusions from the above proposition:

- TD with following quantization of E-vectors is optimal in accumulated WED sense.
- A proper weighting for E-vector quantization is $\Phi(l)^T \tilde{W}\Phi(l) = B$ (see (9)). We have discussed before some special properties of this matrix in (9)-(10) and proposition 1.

Probably, the best quantization strategy is to quantize 3 E-vectors as a single codeword. However, it will lead to a very large codebook size. In order to overcome this difficulty, the weighting matrix structure may be utilized. The quantization cost function can be written as follows:

$$C_Q = \left(\underline{\Omega} - \underline{\Omega}' \right)^T B \left(\underline{\Omega} - \underline{\Omega}' \right) =$$
$$= \sum_{j=1}^{p} \left[\omega_1(j)\omega_2(j)\omega_3(j) \right] B_j \begin{bmatrix} \omega_1(j) \\ \omega_2(j) \\ \omega_3(j) \end{bmatrix} \qquad (14)$$

where $\omega_i(j)$ is j^{th} element of i^{th} E-vector. Thus, cost function can be split into partial cost functions, e.g., function of first *k* elements of all E-vectors, function of next *k* elements etc. This observation suggests splitting of a codebook in the same way. Equation (14) defines weightings for each part. Finally, the codebook can be trained using LBG algorithm.

4 SIMULATION RESULTS

In order to study the performance of the proposed compression method, it was integrated into the MELP2400 vocoder [1]. Quantization of LSF vectors was replaced by TD with SMQ as follows: Input speech is segmented into 5 frame long segments. After that, each segment of LSF vectors is decomposed using TD, when the middle event can be between 2nd and 4th frames inclusively.

All possible locations of the middle event are examined and the one, which yields the minimal WED, is chosen. Then, E-vectors are quantized using SMQ as described in table 1.

The codeword indexes are transmitted along with event location. Since there are 3 possible locations, 2 bits are needed for event position coding in addition to 53 bits of SMQ. Therefore, the bit rate is 11 bits/frame. Note that standard MELP vocoder quantizes LPC parameters at 25 bits/frame (1110 bits/sec). Decoder uses quantized E-vectors and event position in order to interpolate all the LSF vectors of the segment. Then, it *arranges line spectrum frequencies* into ascending order with minimum separation of 50 Hz as in MELP encoder [1]. The SMQ codebooks were trained using 130000 segments obtained by TD of TIMIT training part, while the performance was assessed using 30000 segments from the test part of TIMIT.

#	parameters	bits
1	1st & 2nd elements of Event Vector	11
2	3rd & 4th	11
3	5th & 6th	11
4	7th & 8th	10
5	9th & 10th	10
	Overall	53

Table 1: SMQ bit count

Three different WED measures were used: Unweighted Euclidian Distance, Inverse Harmonic Measure and theoretically optimal distance [11]. The performance was assessed using LSD integrated over 125-3125 Hz band.

Mean LSD below 1 dB was achieved for all of them. Table 2 summarizes the results. It also includes for comparison performance of TSQ technique [8], that performs at an identical bit-rate for the quantization of the LPC parameters.

	Mean LSD, dB
Unweighted	0.995
Inverse Harmonic	0.986
Optimal	0.987
TSQ [12]	2.11

Table 2: Mean LSD

However, the method doesn't provide transparent coding due to large number of outliers. Table 3 shows statistic of outliers above given distortion level for several levels and WED types.

LSD, dB	**Outliers, %**		
	Unweighted	**Inverse Harmonic**	**Optimal**
1	36.4	37.1	37.3
2	18.1	18.0	17.9
3	7.2	6.6	6.4
4	2.9	2.5	2.4

Table 3: Outlier statistic

5 CONCLUSION

The paper proposed a computationally efficient method for LPC parameter coding based on TD. It attains average LSD less than 1 dB at 11 bits/frame. Thus, it saves 56% of LPC parameter bit bugged and 26% of total MELP2400 bit bugged. The method offers constant delay and complexity. Furthermore, it benefits from WED utilization, which significantly reduces high-distortion outlier probability.

6 REFERENCES

[1] L. M. Supplee, R. P. Cohn and J. S. Collura, "MELP: The New Federal Standard at 2400 BPS," Proc. ICASSP 1997, Vol. 2, pp. 1591-1594,1997

[2] J. R. Deller, J. G. Proakis and J. H. L. Hansen, Discrete-Time Processing of Speech Signals, Prentice Hall, Location, 1993.

[3] K. K. Paliwal and W. B. Kleijn, "Quantization of LPC parameters" in Speech Coding and Synthesis W. B. Kleijn and K. K. Paliwal (Eds) Elsevier 1995 (Ch. 12).

[4] C. Tsao and R. M. Gray, "Matrix Quantizer Design for Speech LPC Using the Generalizer Lloyd Algorithm," IEEE Trans. on Acoustic Speech and Signal Processing, Vol. 32, No. 3, pp. 537-545, June 1985.

[5] D. P. Kemp, J. S. Collura and T. E. Tremain, "LPC Parameter Quantization at 600, 800 and 1200 Bits per Second," Proc. of the Tactical Communications Conf., pp. 71-75, 1992.

[6] E. B. George, A. V. McCree and V. R. Viswanathan, "Variable Frame Rate Parameter Encoding Via Adaptive Frame Selection Using Dynamic Programming," IEEE Trans. on Speech and Audio Processing, Proc. ICASSP 1996, pp. 271-274.

[7] Y. Shiraki and M. Honda, "LPC Speech Coding Based on Variable-Length Segment Quantization," IEEE Trans. on Acoustic Speech and Signal Processing, Vol. 36, No. 9, pp. 1437-1444, September 1988.

[8] R. Mayrench, D. Malah, "Low Bit-Rate Coding Using Quantization of Variable Length Segments", Eurospeech 99.

[9] B. S. Atal, "Efficient Coding of LPC Parameters by Temporal Decomposition," Proc. ICASSP 1983, pp. 81-84.

[10] S. J. Kim, Y. H. Oh, "Efficient Quantization of LSF Based on Restricted Temporal Decomposition", Electronics Letters, Vol. 35, No. 12, pp. 962-963, June 1999

[11] W. Gardner, B. Rao, "Theoretical Analysis of the High-Rate Vector Quantization of LPC Parameters", IEEE Trans. On Speech and Audio Processing, Vol. 3, No. 5, pp 367-381, September 1995.