MULTI CODEBOOK VECTOR QUANTIZATION OF LPC PARAMETERS

Professor Costas S. Xydeas and Thomas M. Chapman
Speech processing research laboratory, Electrical Engineering division,
Manchester school of Engineering, University of Manchester,
Dover st., MANCHESTER M13 9PL, U.K.

I. ABSTRACT

This paper presents a novel and efficient variable bit rate LPC quantization approach. The proposed MCVQ framework allows a Dynamic Programming based minimum quantization distortion partitioning and quantization process to be performed on input LSP vector tracks in time. Variable duration segments of LSP vector tracks are classified into one of a finite number of language related events. Specific codebooks, designed optimally for each event type, are then employed to vector quantize the individual LSP vectors of a given segment. “High quality” LSP quantization can be easily achieved at an average of 700 bits/sec while “transparent” performance is obtained at an average rate of 800 bits/ sec.

II. INTRODUCTION

A common feature of many speech coding algorithms is the necessity to model, encode and transmit sets of LPC parameters that specify the vocal tract shape of the speech production mechanism and thus the “filter” part of a source - filter synthesis model. Furthermore, the transmission of LPC parameters in low bit rate coders [7] often represents a significant proportion of the overall system bit rate and it is therefore of interest to develop efficient methods for coding this short time envelope spectrum information.

In general, LPC quantisation techniques can be classified into two broad categories. Techniques in category I focus on minimizing a spectral distortion formed between *individual* (i.e. single) unquantized and quantized LPC vectors. Well known methods such as Scalar Quantisation, Vector Quantisation (VQ) [1], Split Vector Quantisation (SVQ) [6] or Finite State Vector Quantization (FSVQ) [2] are examples of this type of LPC quantisation. The second category II comprises of techniques which quantize *evolutions* of LPC parameters over time. Methods that fall into this category include Matrix quantisation (MQ) [9], Split Matrix Quantisation (SMQ) [10] and Segment Quantisation (SQ) [4], [8].

In the first category, the perceptual effect of the resulting quantisation distortion is reduced by minimizing the spectral distortion arising from the *individual* quantisation of each LPC vector. In this case, quantization distortion which is introduced into the “tracks” of LPC parameters over time is “random” at each point along the track. In the second category, codebooks are designed in such a way that the quantisation distortion is distributed more smoothly over the evolutions with time of LPC parameters. This more even distribution of quantisation distortion diminishes its perceptual effect on the synthesized signal and leads often to improved subjective results, when compared to techniques of the first category, while operating at the same level of spectral distortion.

Furthermore, studies in Segment Quantisation methods [8] suggest that it is desirable to split the evolution in time of LPC spectra into variable length, language “events”. These events are then quantised through employing a transformation of variable length matrices of LPC parameters to fixed length matrices, followed by Matrix Quantisation.

This paper presents a novel and efficient Multi Codebook Vector Quantisation (MCVQ), category I technique. The concept that underpins MCVQ is the partitioning of the input series of LPC vectors into variable length segments of vectors and the subsequent “Classification” of each segment into one of a finite number of language related events. This allows different codebooks i) to be designed for each type of segment event and ii) be used to vector quantize the individual LPC vectors of a given segment.

The effect of restricting, in this way, the LPC vector space prior to codebook design enables a reduction in codebook sizes while maintaining quantization performance. Of course, a small part of the resulting savings in the number of bits used to transmit codebook indices is canceled by the per segment transmission of “class” information.

Thus the proposed system employs a form of LPC spectral evolution segmentation, a multiplicity of “event” related codebooks and avoids variable to fixed length, SQ type, transformations. Furthermore, MCVQ opens up the possibility of assigning different numbers of bits to different classes of segments, depending upon the “variability” of LPC vectors in a class and the perceptual significance of the class.

The MCVQ framework is generic and allows existing LPC quantization techniques to take advantage of the modality of the spectral evolution of speech, in order to operate at even lower bit rates, whilst retaining their basic concept and structure.

III. MCVQ QUANTISATION SCHEME

The encoding part of the system consists of three processes

i) An LPC analysis that operates at regular 20ms intervals to yield p, $\{a_i\}$ LPC coefficients. These are then transformed into LSP parameters. Successive LSP vectors $\{\overline{LSP}_k\}$, k = 1, 2, 3, … can be thought of as forming a "track" in time, in a p-dimensional space.

ii) A voiced / unvoiced classification that also operates in 20ms frames of input speech and provides V/UN flags.

iii) A dynamic programming (DP) procedure that accepts $\{\overline{LSP}_k\}$ tracks and corresponding V/UN flags, see figure 1a. The DP partitions LSP vector tracks into variable duration segments of $\overline{LSP}_k$ vectors in a way that minimizes the quantization noise produced by the MCVQ quantization of each individual $\overline{LSP}_k$ vector. This track segmentation and quantization of individual $\overline{LSP}$ vectors is performed within this "integrated" DP step. The procedure operates subject to a given constraint; either a minimum acceptable segment length (SL) or a given average number of segments per second. The net effect is to minimize quantization distortion whilst restricting the rate of segments generated per second. Note that a further constraint imposed on the DP process is that the voiced / unvoiced flags assigned to the vectors of a segment must be always of the same value (i.e. a segment may not contain a transition from voiced to unvoiced speech).

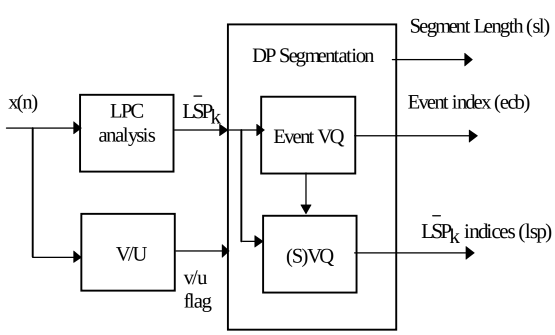

Figure 1a The MCVQ encoding process

Furthermore, the LSP vectors of a given segment are quantized using a two step process:

a) A segment classification VQ process, where a segment is "labeled" as one of a finite number of possible types (event VQ). This results in an event index "ecb" that is transmitted once per segment. Thus an M entry "event" codebook (ECB) that contains "representative" LSP vectors (centroids) for M segment types is employed at this stage. The $\overline{LSP}$ vectors of a given segment are vector quantized by ECB and the index "ecb" of the centroid that produces a minimum weighted Euclidean distortion (WED), summed over all vectors in the segment is selected.

b) A conventional Vector Quantization or Split Vector Quantization process (S)VQ, whose codebook(s) is (are) selected among M possible codebooks CB_i, i=1, …, M, each specifically designed for a particular type of segment. Thus "ecb" obtained from step 1 is used to select the appropriate CB_i which in turn is used to Vector Quantize all K $\overline{LSP}_k$ vectors of the given segment. This results in K "lsp" indices. This step 2 (S)VQ process also operates using the WED measure of step 1. Note that the total WED arising from the CB_i quantization of all K vectors is returned to the DP process, as the distortion produced for a given segment.

The decoding process, see figure 1b, accepts "sl" and recovers the length "SL" of a given segment. "ecb" selects the appropriate (S)VQ codebook(s) from which the recovered $\overline{LSP}_k$ vectors are obtained using SL received "lsp" indices.

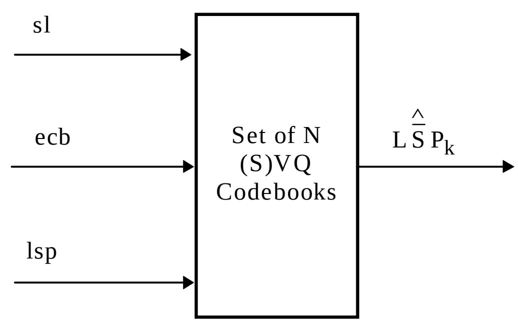

Figure 1b The MCVQ decoding process

IV MCVQ CODEBOOK DESIGN PROCESS

MCVQ codebooks are designed in two stages. Stage 1 determines ECB, whereas the CB_i codebooks are obtained in stage 2.

Stage 1

The process commences with a very large input database of $\overline{LSP}_k$ vectors. Note that more than 5 hours of speech material was used to generate this database. The entire sequence of $\overline{LSP}_k$ vectors can be thought of as providing an "LSP vector track" in a p-dimensional space. An initial 2 entry "event" codebook ECB is used within the DP process. While assuming that each of the CB_i i = 1, 2 codebooks consist of only one vector, which is the same as the corresponding entry of the ECB codebook, the DP process operates on the LSP vector track. This provides track segmentation, an event classification for each output $\overline{LSP}_k$ and a two events ECB vector quantization WED for the input database. The partitioning of the database into two classes of LSP vectors allows the definition of a normalized weighted average vector for each class. These two LSP vectors represent the new entries of the ECB codebook and the operation is repeated iteratively until a two entry minimum WED ($D_m < \varepsilon$) codebook is obtained. The codebook entries are then split into four and the process continues until an optimized M=4 ECB is obtained. In this way an optimal event codebook of size M is derived (see figure 2). At this point, the $\overline{LSP_k}$ vectors of the input database are segmented and labeled according to 1 of M events. This leads to M subsets of LSP vectors each containing the vectors assigned to a specific class.

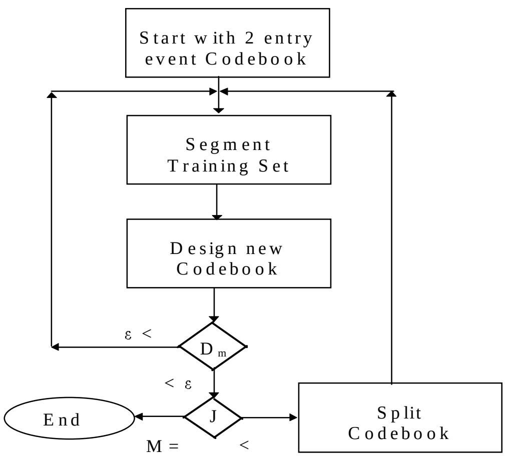

Figure 2 First stage of the training algorithm

Stage 2

Given a subset of LSP vectors, a conventional VQ codebook generation process [3] can be then applied to produce the CB_i i=1,… M codebooks of size N_i. Of course in the case of a j-way split VQ CB_i is defined in terms of j codebooks $SCB_{i,j}$ each of size $N_{i,j}$.

V COMPUTER SIMULATION RESULTS AND DISCUSSION

The proposed MCVQ approach has been simulated using the Burg algorithm on 20ms LPC analysis frames with p=10. LPC coefficients were then converted to LSPs prior to quantization. System performance depends upon the value of M, the method used to quantize the LSP vectors of a specific class (VQ, SVQ were considered although other schemes could be employed) and the number of bits allocated to each class.

Although the scheme is still under investigation, in order to define it's full potential, figure 3 presents typical performance results in terms of a Modified Average Spectral Distortion Measure mASDM [5] and average bit rates, for the following MCVQ configurations

i) MCVQ / VQ where each classified LSP vector is quantised using VQ

ii) MCVQ / 2wSVQ where classified LSP vectors are quantised using a 2 way split VQ and

iii) MCVQ / 3wSVQ where quantization of the classified LPC vectors is performed by 3 way split VQ.

In all of these systems the size M of ECB is 256, with only one entry used to represent unvoiced speech frames. The system bit allocations are shown in tables 1 - 3. Notice that these vary only between "voiced" and "unvoiced" types of segments. Furthermore, the split VQ bit allocation in tables 2 and 3 is not subjectively optimised. In contrast, the bit allocation of the conventional 3 way split VQ system, used as a benchmark in figure 3, is subjectively optimal.

Unvoiced	**Voiced**	**Mean bit rate (bits/sec)**
6	8	535
6	7	503
6	6	469
5	5	418

Table 1 MCVQ/VQ bit allocation for voiced and unvoiced event related VQ codebooks

Voiced LSP's 1 - 5	**Voiced LSP's 6 - 10**	**Unvoiced LSP's 1 - 5**	**Unvoiced LSP's 6 - 10**	**Mean bit rate (bits/sec)**
7	7	5	5	798
6	6	5	5	728
5	5	5	5	658
4	4	5	5	588

Table 2 MCVQ/2wSVQ bit allocation for voiced and unvoiced event related 2 way split VQ codebooks

V LSP's 1 - 3	V LSP's 4 - 6	V LSP's 7 - 10	UV LSP's 1 - 3	UV LSP's 4 - 6	UV LSP's 7 - 10	Av. Bit rate
5	5	5	5	5	5	942
4	4	4	5	5	5	834
3	3	3	5	5	5	732
2	2	2	5	5	5	632

Table 3 MCVQ/3wSVQ bit allocation for voiced and unvoiced event related 3 way split codebooks

While adopting the following criteria [10] for "transparent" LPC quantization performance :

a) LogSegSNR > 10dB

b) mASDM > 1.75dB in the frequency range of 2.4 to 3.4 kHz with less than 3.5 % outliers of a value larger than 4dB

and also using informal subjective tests, MCVQ / 2wSVQ offers "transparency" at an average 16 bits / frame whereas MCVQ / 3wSVQ achieves this at an average 19 bits / frame. Since the 3 way split VQ benchmark system offers transparency with 29 bits / frame, the MCVQ / 3wSVQ and MCVQ / 2wSVQ schemes produce average gains of 10 and 13 bits respectively.

Furthermore, "High quality" LPC quantization , specified by

a) 10dB < LogSegSNR < 9.5dB and

b) 1.75 < mASDM < 2dB with no more than 5% outliers above 4dB

is obtained by MCVQ / 3wSVQ and MCVQ / 2wSVQ operating with an average of 15 to 16 and 14 bits per frame respectively. The SVQ benchmark system achieves an equivalent performance at 27 bits / frame. Notice that MCVQ / VQ outperforms the other two MCVQ schemes, but this is at the expense of impracticable codebook element storage requirements. In contrast, MCVQ / Split VQ schemes lead to efficient systems with realistic storage characteristics which can be controlled via appropriate "vector split" and codebook bit allocation per "event" strategies.

The conputational complexity of the proposed integrated DP segmentation / quantization process varies with time, according to input signal characteristics. Nevertheless as a rough rule of thumb the complexity of the MCVQ systems specified in tables 2 and 3 is no higher than that of a 9 to 10 bits conventional full VQ scheme.

Finally, the MCVQ encoding delay is defined by the DP optimization interval. In the MCVQ schemes and corresponding results outlined in this paper, a 200ms DP optimization interval was employed.

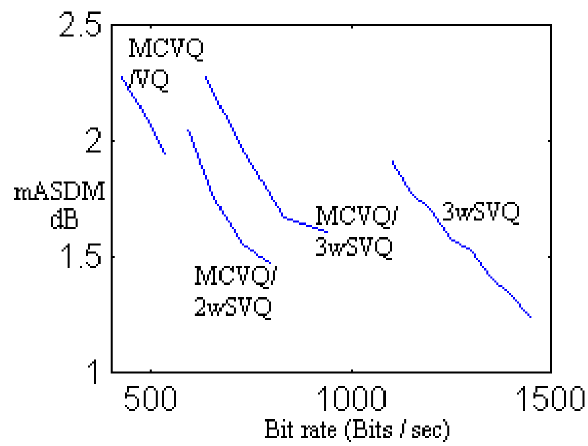

Figure 3 mASDM vs bit rate curves for a standard 3 way split VQ (3wSVQ) and the MCVQ scheme using standard VQ (MCVQ/VQ), 2 way split VQ (MCVQ/2wSVQ) and three way split VQ (MCVQ/3wSVQ)

VI REFERENCES

[1] A. Buzo, A. H. Gray, R. M. Gray, J. D. Markel, *"Speech coding based upon vector quantization"*, IEEE ASSP-28 pp 562 - 574, 1980

[2] M. O. Dunham, R. M. Gray, *"An algorithm for the design of labeled transition Finite State Vector Quantizers"*, IEEE Trans. Com Vol Com-33 pp 83 - 89, 1985

[3] A. Gersho, A. Gray, *"Vector quantization and signal compression"*, Kluwier Academic Publishers London 1992 pp 362 - 369

[4] M. Honda, Y. Shiraki, *"Very Low-Bit-Rate Speech coding"*, in Advances in speech signal processing, Editors S. Furui, M.M. Sondhi, Markel Dekker, 1991 pp 209-230

[5] N. Kitawaki, *"Quality assesment of coded speech"*, in *Advances in speech signal processing*, Editors S. Furui, M. M. Sondhi, Markel Dekker Inc., 1991, pp 357 - 386

[6] K. K. Palival, B. S. Atal, *"Efficient vector quantization of LPC parameters at 24 bits / frame"*, IEEE ASSP pp 3 - 14, 1993

[7] C. Papanstasiou, C. S. Xydeas, *"Efficient mixed excitation models in LPC based prototype interpolation speech coders"*, proc. ICASSP 97 pp 1555 - 1558, 1997

[8] S. Rocous, R. Schwartz, J. Makkhoul , *"Segment quantization for very low rate speech coding"*, proc. ICASSP 1982, pp 1565 - 1568

[9] C. Tsao, R. M. Gray, *"Matrix Quantizer design for LPC speech using the generalized Lloyd algorithm"*, IEEE ASSP-33 pp537-545, 1985

[10] C.S. Xydeas and C. Papanastasiou, *"Efficient coding of LSP parameters using Split Matrix Quantisation"*, in Proc ICASSP 95 pp 740 - 743.