VECTOR QUANTISATION OF THINNED FILTER COEFFICIENTS FOR LOW-BIT-RATE SPEECH CODING

C. F. Chan

Department of Electronic Engineering
City Polytechnic of Hong Kong

Abstract

The results in vector quantization of thinned filter coefficients are presented in this paper. The codeword in the thinned filter codebook is represented by the non-zero coefficients and their corresponding delays. By using a modified likelihood ratio as distortion measure, and employing the generalized Lloyd algorithm for codebook training, we are able to obtain thinned filter codebooks for both the transversal and lattice models with significant lower distortion than conventional LPC-VQ codebooks. Simulation results show that a 7-bit thinned lattice filter codebook has the same distortion level as a 10-bit LPC-VQ codebook.

Introduction

In linear predictive coding (LPC) of speech, the human vocal tract response is modelled by a time-varying all-pole filter. In order to achieve low bit rate, the order of the all-pole filter is usually kept small, say 10. It has been shown in Reference [1] that thinned filter approach to the modelling of speech can achieve better spectrum fitting to speech signals while keeping the data rate for encoding the filter parameters low. Basically, a thinned filter is a high-order filter having only few non-zero coefficients. There are two types of thinned filters, i.e., thinned transversal filter as shown in Fig. 1 and thinned lattice filter as shown in Fig. 2. In thinned filter model, each short-time speech spectrum is represented by the non-zero coefficients and their corresponding delays. Since the delays have to be coded for synthesis, the coding rate would be slightly higher than conventional LPC system. In this paper, we will introduce methods to vector quantize the thinned filter parameters so as to achieve low bit rate speech coding.

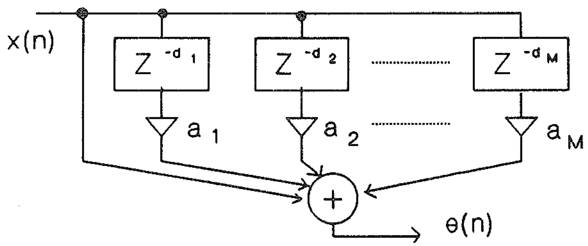

Fig. 1 Thinned Transversal Filter

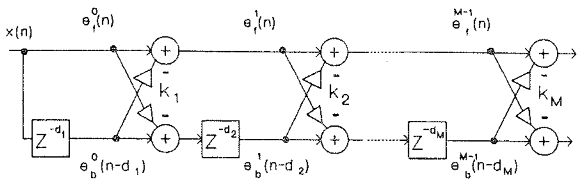

Fig. 2 Thinned Lattice Filter

Thinned Filter Analysis

Methods to obtain the thinned filter parameters for both the transversal and lattice models have been developed and reported in Reference [1]. Without loss of generosity, we are assuming that autocorrelation method is used for thinned filter analysis. Basically, in the thinned transversal model, the non-zero filter coefficients are found by solving a set of linear equations once the delays are known, i.e.,

$$\boldsymbol{R}_M \boldsymbol{a}_M = \boldsymbol{r}_M$$

where $\boldsymbol{a}_M$ is a column vector with entries being the non-zero linear predictive coefficients, $\boldsymbol{r}_M = [r_{d_1} r_{d_2} \ldots r_{d_M}]^T$ is the autocorrelation vector of input speech with each delay d_i corresponding to the non-zero coefficient a_i, and $\boldsymbol{R}_M$ is a symmetric semi-positive matrix with autocorrelation entries $r_{|d_i - d_j|}$ for $1 \le i,j \le M$. Note that $\boldsymbol{R}_M$ is non-Toeplitz as the delays are non-uniformly distributed. A sub-optimum search method similar to multi-pulse analysis can be used to determine the delays one by one[2]. Assume that re-optimization is applied after each delay was found, the complexity of the searching algorithm is $O(M^2T)$, where M is the number of non-zero coefficients and T is the highest order of the thinned filter.

In case of the thinned lattice model, the filter parameters are obtained by minimizing the sum of the forward and backward prediction errors. Define the auto- and cross correlation of the forward and backward prediction errors as:

$$\alpha_t^m = \sum_n e_f^m(n) e_f^m(n-t)$$

and

$$\beta_t^m = \sum_n e_f^m(n) e_b^m(n-t)$$

where $e_f^m(n)$ and $e_b^m(n)$ are the forward and backward prediction errors, respectively. By using autocorrelation analysis, it is straight forward to show that the correlations in the thinned lattice filter satisfies the following recursive equation:

$$\begin{bmatrix} \alpha_t^{m+1} \\ \beta_t^{m+1} \end{bmatrix} = \begin{bmatrix} 1 & -k_{m+1} \\ -k_{m+1} & 1 \end{bmatrix} \begin{bmatrix} \alpha_t^m - k_{m+1}\beta_{d_{m+1}-t}^m \\ \beta_{d_{m+1}+t}^m - k_{m+1}\alpha_t^m \end{bmatrix} \quad 0 \le |t| < T-m \tag{1}$$

with $k_{m+1} = \dfrac{\beta_{d_{m+1}}^m}{\alpha_0^m}$. Note that $\alpha_t^0 = \beta_t^0 = r_t$. Therefore, once the delays are known the non-zero reflection coefficients can be calculated. A dynamic programming algorithm has been given in Reference [3] to optimally partition the range of delays. The algorithm has a complexity of $O(MT^2)$.

To see the effectiveness of the thinned filter models, a short-frame of speech signal is analyzed and the spectrum plots of all filter models are shown in Fig. 3. In this analysis all filters have 10 non-zero coefficients.

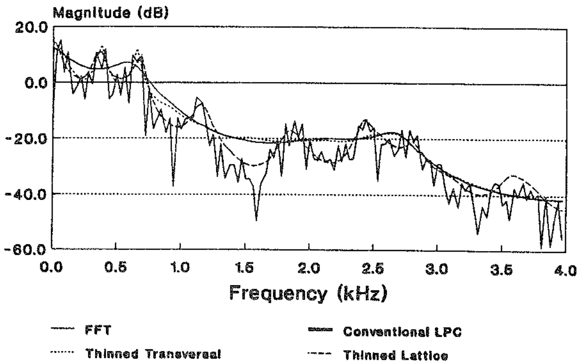

Fig. 3 Spectrum Plots of Various AR Models

In this paper, the relationship between the prediction error energy and the number of non-zero coefficients of the thinned filter is studied. Fig. 4 illustrates plots of average likelihood ratio distortion against number of non-zero coefficients for the conventional LPC, thinned transversal and thinned lattice approaches. For the purpose of comparison, the average likelihood ratio distortion is defined as $d(dB) = 6.142\sqrt{d_{LR}}$ with $d_{LR} = \alpha/\alpha_{64} - 1$, where α_{64} is the prediction error energy obtained from a 64-order LPC analysis. We can consider, say, a 10-order conventional LPC analysis as a 64-order LPC analysis with the last 54 coefficients being truncated to zero. In this analysis, speech signals are band-limited to 3.4 kHz and sampled at 8 kHz. A 32 ms Hamming window is used. We can see from these plots that the thinned filter approaches produce lower distortion levels than conventional LPC approach having the same number of non-zero coefficients. More specifically, for the same distortion levels, a thinned lattice filter with, say, 6 non-zero coefficients is equivalent to an order 10 LPC filter. It is interesting to note that the prediction error energy tends to saturate early in conventional LPC systems while this is not the case for the thinned filter systems. The thinned filter model would have many applications in speech coding especially in low-delay coder where the order of the predictive filter is normally high, eg. 50 in Reference [4].

Vector Quantization of Thinned Filter Parameters

In vector quantization of the LPC parameters, the distortion measure used is a spectrum distortion, for instance, Itakura-Saito distortion[5], and log likelihood ratio distortion[6]. Since a thinned filter is considered as a high-order LPC filter, methods used in vector quantization of the LPC parameters can also be applied on the thinned filter parameters. In this work, we utilize the likelihood ratio d_{LR} as distortion measure. d_{LR} is defined over two gain-normalized filter models, and is normally computed as $d_{LR} = \alpha/\alpha_M - 1$ where α_M is the optimum prediction error energy obtained from an order M LPC analysis. Because the order of the thinned filter depends on the values of the delays, it is rather difficult to define α_M if the analysis result is intended to be compared with conventional LPC VQ. We therefore decide to use a 64-order LPC filter as a reference model so that all distortion measures can be referred to.

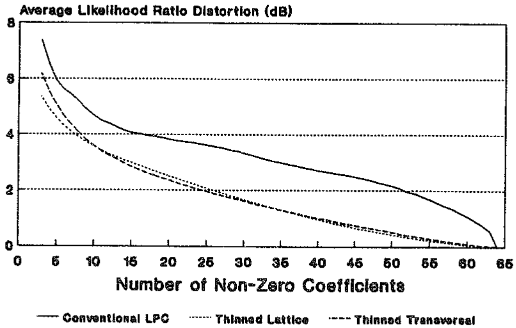

Fig. 4 Distortion vs Number of Non-Zero Coefficients

Distortion Calculation

In case of the transversal model, the prediction error energy can be calculated as

$$\alpha = [1 \ a_M^T] \begin{bmatrix} r_0 & r_M^T \\ r_M & R_M \end{bmatrix} \begin{bmatrix} 1 \\ a_M \end{bmatrix} \tag{2}$$

with a_M being the codeword. The corresponding delays must be used to select the autocorrelation vector. The complexity for evaluating the distortion for each thinned transversal codeword is thus $O(M^2)$. In case of the lattice model, the prediction error energy can be calculated by going through the thinned lattice analysis procedure given in Eqn. 1 using the known codeword represented by $\{k_i, d_i\}_{i=1}^{M}$. The computation required for this distortion evaluation is $O(MT)$.

Codebook Generation

A speech data base contributed by two male and two female speakers and containing over 6000 gain-normalized autocorrelation vectors is used for training the thinned filter codebooks. We employ the generalized Lloyd algorithm with splitting strategy for codebook generation[7]. Since a large amount of time is required for training full-search thinned filter codebook, only binary-tree search codebooks are actually designed. In the codebook generation process, the centriod of a cluster of speech vectors is obtained by performing the thinned filter analysis procedure over the geometric mean of the autocorrelation vectors belong to that cluster[7]. Note that the codebook storage for thinned filter codewords is higher than conventional codebooks as the delays are also needed to be stored.

Simulation Results

Several thinned filter codebooks of different sizes have been designed and the results are compared with conventional LPC-VQ codebooks. Fig. 5 shows the plots of average normalized distortion against codebook size for both conventional and thinned filter codebooks. We see from these plots that the thinned filter codebooks achieve much lower distortion than conventional codebooks where the codebook size is reasonable large. In particular, a 7-bit thinned lattice codebook obtains a distortion level comparable to a 10-bit conventional codebook. It is also interesting to note that all codebooks achieve similar level of distortion when the codebook size is small, however, when the codebook size increases, improvement on the quantization distortion for conventional VQ codebook tends to saturate while thinned codebooks still enjoy favorable improvements. The result is no surprise because the constraint in filter order is relaxed in thinned filter model, the distortion level will not be saturated early as codebook size increases.

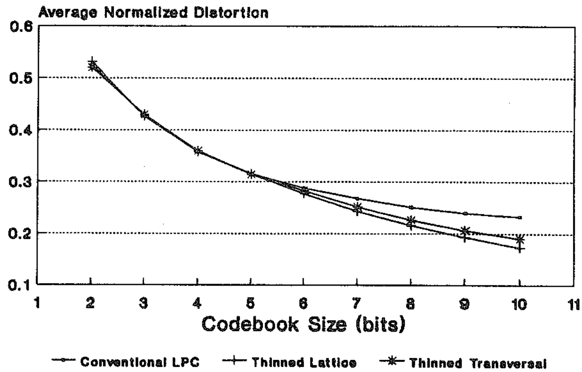

Fig. 5 Quantization Distortion versus Codebook Size

Analysis of Outlier Frames

It is known that occasional large distortion in VQ systems would perceptually reduce the synthetic speech quality to some extent. A measure of the number of outlier frames is important in any VQ systems. We therefore also compare the distribution of distortion for the conventional and thinned filter codebooks to see whether thinned filter codebooks have similar profile of distortion. Fig. 6 shows the probability distribution of distortion for the conventional and thinned lattice VQ systems. In this analysis, all codebooks have 256 codevectors each has 8 non-zero coefficients. All distortions were calculated using the original training vectors. It can seen from these plots that, except for a lower average distortion, thinned filter codebooks have similar profile of distortion as conventional VQ systems.

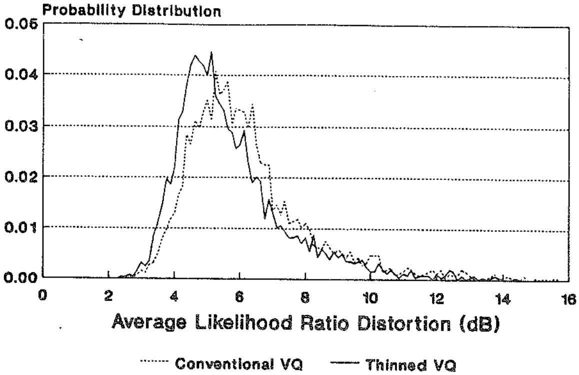

Fig. 6 Probability Distribution of Distortion

Conclusion

By applying vector quantization on thinned filter parameters for linear prediction of speech, the quantization distortion is significantly lower than conventional LPC VQ systems. It has been shown that a 7-bit thinned lattice codebook can achieve a distortion level comparable to a 10-bit LPC-VQ codebook with similar distribution profile of distortion. Future works will be concentrated on developing a multi-stage scheme for vector quantization of the thinned coefficients in order to reduce the complexity and storage requirement.

References

1. C. F. Chan and S. H. Leung, "A Speech Vocoder using High-Order LPC Filter with very Few Non-Zero Coefficients," European Conf. on Speech Communication and Technology, pp.839-842, Genova, Italy, Sept. 1991.
2. B. S. Atal and J. R. Remde, "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp.614-617, 1982.
3. C. F. Chan, "Dynamic Programming Algorithm for Linear Prediction Using Thinned Lattice Filter," IEE Electronics Letters, pp.91-92, Jan. 1992.
4. J.-H. Chen, "High-Quality 16 kb/s Speech Coding with a One-Way Delay Less Than 2 ms," IEEE Int. Conf. on Acoustics, Speech, and Signal Processing, pp.453-456, 1990.
5. F. Itakura, "Analysis Synthesis Telephony Based Upon the Maximum Likelihood Method," Conf. Rec., 6th Int. Congress Acoustics, Y. Yonasi Ed., Tokyo Japan 1968.
6. J. M. Tribolet, L. R. Rabiner, and M. M. Sondhi, "Statistical Properties of an LPC Distance Measure," IEEE Trans. on Acoustics, Speech, and Signal Processing, ASSP-27, pp.550-558, Oct. 1979.
7. Y. Linde, A. Buzo, and R. M. Gray, "An Algorithm for Vector Quantizer Design," IEEE Trans. on Communications COM-28, No. 1, pp.84-95, January 1980.