International Journal of
Applied Mathematics,
Electronics and Computers
ISSN: 2147-8228 http://ijamec.atscience.org
Advanced Technology and Science
Original Research Paper

Residual Lsf Vector Quantization Using Arma Prediction

Selma Ozaydin*[1]

Accepted 3rd September 2016

Abstract: The residual LSF vector quantization yields bit rate reduction in the vocoders. In this work, a residual LSF vector quantization obtained from Auto Regressive Moving Average (ARMA) prediction is proposed for designing codebooks at very low bit rates. This residual quantization method is applied to multi stage vector quantization method and codebooks are designed. For each codebook, the effectiveness and quality are investigated by calculating the spectral distortion and outliers. The proposed quantization method reduced the distortion without any additional complexity.

Keywords: *very low bit rate, speech processing, residual vector quantization, formant tracking, ARMA prediction*

1. Introduction

Speech coding refers to process of reducing the bit rate of digital speech representations for transmission or storage, while maintaining a speech quality that is acceptable for the application. Most of the speech coders reported in the literature are based on linear prediction (LP) analysis [1]. For the LP based vocoders, the bit rate reduction is strongly tied to efficient quantization of the LPC filter coefficients $\{a_j\}$. The Line Spectral Frequencies (LSF) –an equivalent representation of $\{a_j\}$, more suitable for quantization and interpolation– can alternatively be used. In this sense, the Multi-Stage Vector Quantization (MSVQ) of LSF parameters presented in [2] has an efficient quantization performance at 22-24 bits per 20ms frames. Furthermore, the multi stage structure has more flexibility than a single stage VQ in terms of search complexity, codebook storage and channel error protection. Very low rate speech communication systems require efficient fixed-rate and low delay coding methods which operate at lower bit rates.

Generalized vocal tract model consists of the oral tract and nasal tract. On the other hand, the linear predictive coding, which has been widely used in the speech analysis and synthesis, uses all pole type digital filters. Speech signals are assumed to be produced by filtering glottal excitation with these filters. This all pole type filter model approximates the true physical configuration of the human vocal tract, but with the nasal tract left out. The most crucial and well known shortcoming in this assumption that during any voiced pronunciation the velum is always closed and the sound wave proceeds only through the oral tract. So the influence of the nasal tract is ignored in this assumption. There is no big problem when non-nasal sounds are processed but in case of nasal sounds the mismatch of the LP model becomes severe. The zeros during nasal sounds supress the peaks in mid-frequency by flattening the spectrum there but this effect cannot easily fit by all pole modelling. In order to include the effect of both oral and nasal tracts, it is necessary to modify all pole modelling into a pole zero modelling [3]. In order to obtain more efficient speech coding algorithms especially for transmission over noisy channels, differential quantization or predictive quantization of spectrum parameters are used. There are some pole zero modelling approaches in the literature [5-11] but they usually use nonlinear equations or approximations. While all pole modelling is simple, pole-zero modelling requires complex nonlinear calculations. Although a pole-zero algorithm based on adaptive Kalman Filtering presented in [11] linearize the nonlinear components by dividing the frequency range of each formant into four bands, this nonlinear approximation method also requires too many calculations causing a complexity in the pole zero modelling.

In this work, we propose an ARMA prediction model for predictive quantization of spectrum parameters. This ARMA prediction method combines the good features of AR and MA prediction methods while eliminating some drawbacks.

In section 2, basic formulas of MSVQ are given and then in section 2.2, residual vector quantization method using ARMA prediction is described. In section 3, designed codebook results using proposed method are presented.

2. Residual LSF Quantization

In this section, a brief description of the MSVQ method presented in [2] is outlined. The definitions presented in this section are introductory information for the residual MSVQ.

The training technique we used in designing the codebooks is the joint design technique [2]. Representative results of the residual LSF scheme with the joint design technique are presented in section 3.

2.1. Notation and Definition

The MSVQ codebooks are designed using the Generalized Lloyd Algorithm (GLA) to minimize average Weighted Mean Square Error (WMSE) based on a sufficiently rich training sequence.

The training sequence is first partitioned into decision regions or cells for a given set of centroids (or codevectors). Then, for the given partitioning, the codebooks are re-optimized to minimize the distortion over the particular decision regions. In the MSVQ system [2], the parameter vector x consisting of p LSF parameters is approximated as a quantized parameter vector $\hat{x}$ using the minimum distortion rule (all vectors are assumed as column vectors),

$$\begin{aligned}\hat{x} &= y_0^{(l_0)} + y_1^{(l_1)} + \ldots + y_{K-1}^{(l_{K-1})} \\ &= B_0^{(l_0)} c_0 + B_1^{(l_1)} c_1 + \ldots + B_{K-1}^{(l_{K-1})} c_{K-1} \\ &= Bc\end{aligned} \quad (1)$$

where superscripts denote codevector indices from each stage, subscripts denote the stage numbers and K is the number of stages. c_j is the codebook vector for the j[th] stage and created by stacking the codevectors,

[1] *Department of Electronics and Communication, Engineering Faculty, Cankaya University, Campus, 06790, Etimesgut Ankara TURKEY*
** Corresponding Author: Email: selmaozaydin@yahoo.com*
Note: This paper has been presented at the 3rd International Conference on Advanced Technology & Sciences (ICAT'16) held in Konya (Turkey), September 01-03, 2016.

$$c_j = \left[y_j^{(0)T}\ y_j^{(1)T} \ldots . y_j^{(L_j-1)T} \right]^T_{L_j p \times 1} \quad (2)$$

where $y_j^{(k)}$ (p × 1) is the k^{th} codevector for the j^{th} stage and L_j is the size of codebook for that stage. The column vector c is constructed by stacking all codebook vectors from all stages and referred to as the 'codebook' where,

$$c = \begin{bmatrix} c_0 \\ c_1 \\ \ldots \\ c_K \end{bmatrix} = \begin{bmatrix} [y_0^{(0)T}\ y_0^{(1)T} \ldots\ y_0^{(L_0-1)T}]^T \\ [y_1^{(0)T}\ y_1^{(1)T} \ldots\ y_1^{(L_1-1)T}]^T \\ \ldots\ \ldots\ \ldots\ \ldots \\ [y_K^{(0)T}\ y_K^{(1)T} \ldots y_K^{(L_K-1)T}]^T \end{bmatrix}_{Lp \times 1} \quad (3)$$

The selection matrix for the j^{th} stage $B_j^{(k)}$ is a sparse Toeplitz matrix ($p \times L_j p$) constructed such that $y_j^{(k)} = B_j^{(k)} c_j$. The selection matrix B is used for selection of the codevectors from codebook table with using the codebook indices where,

$$B = [B_0^{(l_0)}\ B_1^{(l_1)} \ldots B_{K-1}^{(l_{K-1})}]_{p \times Lp} \quad (4)$$

A WMSE distortion criterion is used for training the codebooks and for the selection of the quantized vector in a codebook [2]. If W is a diagonal matrix which depends on the parameter vector x, the distortion for the whole training sequence is defined as,

$$d_r(x, \hat{x}) = \sum_n ((x^{(n)} - \hat{x}^{(n)})^T W^{(n)} (x^{(n)} - \hat{x}^{(n)})) \quad (5)$$

where the superscript n identifies the nth vector from the training sequence and the subscript r represents the iteration number during the training of a codebook. For details of designing a codebook in MSVQ, the reader is referred to [2].

2.2. Residual LSF vector Quantization with ARMA prediction

In the R_MSVQ method, the residual LSF parameters of current frame are predicted from the quantized LSF parameters of the previous frames using interframe correlation feature of spectrum parameters [5-8] and then residual LSF vectors are coded with a MSVQ codebook. Firstly, the LSF parameter vector is obtained by transforming the 10th order LPC parameter vector. Next, the average LSF vector of the training set x_{DC} is subtracted from the LSF vector $x^{(i)}$ belonging to the i^{th} frame. By defining mean removed LSF vectors ($z^{(i)} = x^{(i)} - x_{DC}$) and its quantized version $\hat{z}^{(i)} = \hat{x}^{(i)} - x_{DC}$), the residual LSF vector $e^{(i)}$ is calculated using,

$$e^{(i)} = z^{(i)} - r^{(i)} \quad (10)$$

where *i=1,2,…and r(0) = 0*. The quantized residual vector $\hat{e}^{(i)}$ is found by quantizing $e^{(i)}$ with a VQ codebook. Depending on how $r^{(i)}$ is computed, various prediction schemes can be proposed. If $r^{(i)} = \alpha(\hat{e}^{(i-1)} + r^{(i-1)})$, a first order Auto Regressive (AR(1)) predictor is obtained [5]. When ($r^{(i)} = \alpha\hat{e}^{(i-1)}$) we have a first order Moving Average (MA(1)) predictor [5]. The research in the literature have focused on these two schemes [6-8] which show that codebooks designed with using AR predictors produce lower distortion than codebooks with MA predictors, however the use of an alternative ARMA model in residual LSF prediction, which is untouched in the literature, can be more advantageous. The ARMA(1,1) predictor is,

$$r^{(i)} = \alpha_1 \hat{e}^{(i-1)} + \alpha_2 r^{(i-1)} \quad (7)$$

Here, we optimize two parameters instead of one in AR(1) and MA(1) predictors. It has been observed that residual LSF codebooks designed by an ARMA(1,1) predictor have lower distortion than AR(1) and MA(1) predictor codebooks. The lowest distortion using an AR(1) model is obtained by using $\alpha=0.5$. In an ARMA(1,1) predictor the lowest distortion is obtained when $\alpha_1=0.3$ and $\alpha_2=0.6$. To see the advantage of the ARMA(1,1) predictor, consider the reconstructed quantized LSF vector $\hat{z}^{(i)}$ in the decoder,

$$\begin{aligned} \hat{z}^{(i)} &= \hat{e}^{(i)} + r^{(i)} \\ &= \hat{e}^{(i)} + \alpha_1 \hat{e}^{(i-1)} + \alpha_1 \alpha_2 \hat{e}^{(i-2)} + \\ &\quad \alpha_1 \alpha_2^2 \hat{e}^{(i-3)} + \alpha_1 \alpha_2^3 \hat{e}^{(i-4)} + \ldots \end{aligned} \quad (8)$$

which is compared to the reconstructed quantized LSF vector of the AR(1) predictor,

$$\begin{aligned} \hat{z}^{(i)} = \hat{e}^{(i)} + \alpha\hat{e}^{(i-1)} + \alpha^2 \hat{e}^{(i-2)} + \\ \alpha^3 \hat{e}^{(i-3)} + \alpha^4 \hat{e}^{(i-4)} + \ldots \end{aligned} \quad (9)$$

It is well-known that [5] AR prediction schemes are susceptible to channel errors due to infinite memory as seen in (9). For example, for $\alpha=0.5$ in AR(1), the weighting of previous quantized residuals decay like *{0.5, 0.25, 0.125, 0.0625, 0.0312, …}*. However, in ARMA(1,1) with $\alpha_1=0.3$ and $\alpha_2=0.6$ the weighting is *{0.3, 0.18, 0.108, 0.0648, 0.0388, …}*. As can be seen, the decay in ARMA(1,1) is faster which means that the susceptibility to channel errors compared to AR(1) is decreased. Hence by using an ARMA predictor not only do we reduce the distortion but also we decrease the effect of channel errors when compared to AR predictors.

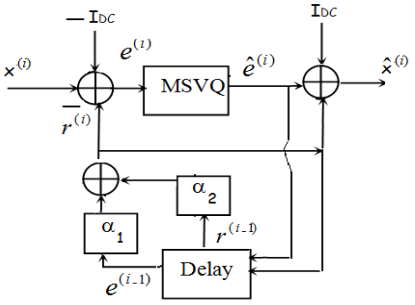

Figure 1. Block diagram of proposed ARMA prediction

Marca [7] suggested an AR predictive scheme in which prediction is performed only on every other frame which limits error propagation at most one adjacent frame. The same approach can be applied to ARMA(1,1) predictors. In practice, AR and ARMA predictive models correspond to higher order MA models as the weighting of previous quantized residuals decay to zero after a sufficiently large previous frame index, say p, which means that the same spectral distortion and outlier results can be obtained by optimizing p coefficients of a MA(p) predictor. However, undue codebook design complexity would be introduced as the search is now over a p dimensional space instead of two in ARMA(1,1).

A weighted Euclidean distance measure is used for training the codebooks and during the search for the best codevector during quantization. The weighted Euclidean distance measure $d(e, \hat{e})$ between the input residual LSF vector e and the quantized residual LSF vector $\hat{e}$ is given by,

$$d(e, \hat{e}) = \sum_{j=1}^{p} w_j (e_j - \hat{e}_j)^2 \quad (11)$$

where p (p=10 in our case) is the number of elements in the residual LSF vector and w_j is the weight coefficients assigned to the j^{th} residual LSF (e_j).

3. Performance Evaluation

In this study some applications of fuzzy logic method in medical LPC/LSF quantization performance can be assessed using subjective tests and/or objective distortion measures. The performance of the codebooks is measured by using spectral distortion method. Hence,

$$A(z)=1+\sum_{j=1}^{p} a_j z^{-j}, \quad \hat{A}(z)=1+\sum_{j=1}^{p} \hat{a}_j z^{-j} \tag{12}$$

$$d_{SD}(A(z),\hat{A}(z)) \approx \sqrt{\frac{1}{n_1-n_0}\sum_{n=0}^{n_1-1}\left[10log_{10}\left[\frac{\left|\hat{A}(e^{j2\pi n/N})\right|^2}{|A(e^{j2\pi n/N})|^2}\right]\right]^2}$$

$$SD = \sqrt{\frac{1}{T}\sum_{i=1}^{T} {d_{SD}}^2(e^{j2\pi n/N}) \quad dB^2} \tag{13}$$

where n_0 and n_1 correspond to 100 Hz and 3800 Hz respectively. A(z) is the optimal p^{th} order linear predictor and $\hat{A}(z)$ is the predictor with quantized coefficients. *N = 256* point FFT is used. The training database (65,685 vectors) is extracted from TIMIT databases, which consists of 630 speakers of 8 major dialects of American English, each reading 10 phonetically rich sentences. The extracted database is low-pass filtered and down sampled to 8 kHz. M-L tree search procedure are used for training and testing [2] and M=8 is taken as the search depth of codevectors.

To further test the efficiency of the R_MSVQ codebook, a novel very low bit rate speech coding-decoding algorithm is designed which is described in detail in [4]. MSVQ codebooks and new residual codebooks with the same bit rate are used respectively. For each designed codebooks, the effectiveness and the limitations of codebooks are investigated by calculating the SD and percentage of outliers. For comparison purposes spectral distortion and outlier results for the codebooks designed with MSVQ and R_MSVQ algorithms are given in Table 1. For lower bit rates, there is approximately 10% bit rate reduction in the R_MSVQ scheme for identical spectral distortion values.

Listening tests and spectral distortion results for test speech data show that a three stage 22 bit/frame R_MSVQ codebook gives the same quality as the four stage 24 bit/frame MSVQ codebook in the new vocoder. We have tried to keep the bit rate of the residual LSF vector quantization as low as possible at an acceptable level as the major contribution to the bit rate of the vocoder comes from the LSF vector quantization.

Table 1. LSF Codebook SD results of AR and ARMA prediction

(a) SD & outliers in LSF joint codebook design,

Bits/frame	SD	%[2-4dB]	%[>4dB]
[765] – 18	1.29	5.86	0.03
[776] – 20	1.17	3.69	0.03
[877] – 22	1.04	2.19	0.01
[888] – 24	0.93	1.28	0.01

(b) SD & outliers in residual LSF codebook design with ARMA prediction

Bits/frame	SD	%[2-4dB]	%[>4dB]
[765] – 18	1.21	4.43	0.04
[776] – 20	1.10	2.94	0.02
[877] – 22	0.95	1.61	0.01
[888] – 24	0.84	0.75	0.00

4. Conclusion

This article has presented an ARMA prediction modelling approach which has been shown to produce lower distortion results. The proposed prediction algorithm improves the performance for all investigated bit rates. Furthermore, the new method has the good features of both AR and MA model. The new residual quantization method reduce the bit rate using residual LSF vectors obtained from ARMA prediction with little calculations in the algorithm. With this method error propagation is limited to a few frames for noisy channels.

It is expected that additional improvement will come with analysing of how adaptive coefficients can be found to model cepstral coefficients. Anti-formant tracking here remains challenging although it has been found better results with the modelling here. On the other hand, the effectiveness of this approach is its ability to model both poles and zeros with a simple algorithm without giving nonlinear complex calculations. Further research is planning to evaluate the effectiveness of this method according to different speech databases and to find an adaptive method to adjust alpha coefficients adaptively.

References

[1] A.V. McCree and T.P. Barnwell III, " A Mixed Excitation LPC Vocoder Model for Low Bit Rate Speech Coding", IEEE Transactions on Speech and Audio Processing, Vol.3, No.4, pp.242-250, July 1995

[2] W.P. LeBlanc, B.Bhattacharya, S.A. Mahmoud, "Efficient Search and Design Procedures for Robust Multi Stage Vector Quantization of LPC Parameters for 4 kbps Speech Coding", IEEE Trans. on Speech and Audio Processing, Vol.1, No.4, pp.373-385, Oct., 1993

[3] I.T.Lim, B.G. Lee, "Lossless pole-zero modelling of speech signals", IEEE Transactions on Speech, Signal and Audio processing, Vol.1, No.3, pp.269-276, July, 1993

[4] S. Ozaydin, B. Baykal, "Matrix quantization based linear predictive speech coding at very low bit rates", Speech Communication, Vol. 41, Issues 2-3, pp:381-392, Oct. 2003

[5] S. Nandkumar, K. Swaminathan, U. Bhaskar, "Robust Speech Model based LSF Vector Quantization for Low Bit Rate Speech Coders", in Proc. IEEE Int. Conf. Acoustics, Speech, Signal Processing, Vol.1, pp.41-44, May 1998

[6] J. Skoglund, J. Lindén, "Predictive VQ for noisy channel spectrum coding : AR or MA?" , Proc. IEEE Int. Conf. on Acoustics, Speech, Signal Processing, May, 1996

[7] J.R. de Marca, "An LSF quantizer for the North-American Half-Rate Speech Coder", IEEE Transactions on Vehicular Technology, Vol.43, No.3, August, 1994

[8] H. Ohmuro, T. Moriya, K. Mano, and S. Miki, "Vector quantization of LSP parameters using moving average interframe prediction", Electronics and Communications in Japan, Part 3, Vol.77, pp.12-26, 1994

[9] B. Wahberg, "ARMA spectral estimation of narrow band processes via model reduction", IEEE Transactions on Acoustics, Speech and Signal Processing, vol. 38, July, 1990

[10] M.G. Kang, B.G. Lee, "A generalized vocal tract model for pole-zero type prediction", Proc. Int.Conf. on ASSP, S14, 10, 1988

[11] L. Deng, L.J. Lee, etc, "Adaptive Kalman filtering and smoothing for tracking vocal tract resonances using a continuous valued hidden dynamic model", IEEE Transactions on Audio, Speech and Language Processing, Vol.15, No.1, 2007