VARIABLE LENGTH CODING OF TRANSFORMED LSF COEFFICIENTS

László Lois

Department of Telecommunications
Tehnical University of Budapest
Sztoczek u. 2, H-1521 Budapest, Hungary
E-mail: lois@tsys.hit.bme.hu

10.21437/Eurospeech.1999-330

ABSTRACT

In this paper, the use of Karhunen-Loeve transform (KLT) and discrete cosine transform (DCT) is studied for encoding of the line spectrum frequency (LSF) parameters at variable bit rate (VBR). For VBR coding, scalar quantization (SQ) is used with Huffman coding. The basic idea in developing these schemes is using linear transform to exploit the strong intraframe and interframe correlation of LSF parameters to increase the performance of the SQ. First, several adaptive KLT schemes are developed to increase the efficiency of the KLT. To reduce the interframe correlation efficiently, two-dimensional transforms and moving average prediction of the transformed coefficients are also investigated. It is shown that these schemes introduce as good as or better performance in the examined bit rates compared to other methods in the field of LSF coding.

1. INTRODUCTION

Linear predictive coding (LPC) parameters are widely used in various speech coding applications for representing the short-time spectral envelope information. In low bit rate speech coding, the line spectrum frequency (LSF) representation is found to be efficient for their good quantization and interpolation characteristics[1]-[4].

Early results from LSF quantization showed that there is a strong correlation between the neighbouring vectors (interframe correlation) as well as within the same frame (intraframe correlation). The strong intraframe correlation could be well exploited either with vector quantization (VQ) or with the Karhunen-Loeve transform (KLT). The interframe correlation may be reduced by using predictive coding[5] or by using two-dimensional (2D) transform[6].

The KLT is considered as an optimal linear transform in the sense that it decorrelates the transform coefficients. The decorrelating property improves the performance of the scalar quantization first of all. The KLT matrix is computed from the estimated autocorrelation (AC) matrix of the signal, hence the performance of the transform coding system depends on the efficiency of the AC matrix estimation[7].

In this paper, the variable length coding scheme is implemented with scalar quantization and Huffman variable lenght coding (VLC). Three schemes are developed to utilize the interframe correlation.

The first scheme based on adaptive KLT. This scheme predicts the AC matrix more efficiently by using the previously decoded frames and other informations such as the voicing and pitch, i.e. this method estimates the optimal KLT matrix of the next LSF frame.

The second scheme is the 2D KLT, similarly to [6], where 2D DCT is used to reduce the intraframe and interframe correlation. To perform a 2D transform, L frames are collected and transformed together. This scheme introduces ancillary $L-1$ frame delay.

The last scheme predicts the transformed coefficients of the next frame by using MA (moving average) pedictor in each coordinate instead of estimating the AC matrix. This idea is similar to the MA prediction of LSF parameters in ITU-T G.729, where two switched predictors are used, and the residual vector is quantized with a two-stage VQ[5].

This paper is organized as follows. In Sect. 2 the statistical properties of the LSF and KL coefficients are described briefly. Based upon this properties, several transform coding schemes are introduced in Sect. 3. The performance of the proposed schemes at different bit rates are presented in Sect. 4, and compared to other methods in the field of LSF coding. Finally, conclusions are given in Sect. 5.

2. PROPERTIES OF THE LSF VECTORS

Let $A_k(z)$ denotes the transfer function of the k-th order LPC analysis filter. By taking sum and difference between $A_k(z)$ and its conjugate function we get the sum filter $P_{k+1}(z)$ and the difference filter $Q_{k+1}(z)$as follows:

$$\begin{aligned} \boldsymbol{P_{k+1}}(\boldsymbol{z}) &= \boldsymbol{A_k}(\boldsymbol{z}) + \boldsymbol{z}^{-(\boldsymbol{k}+1)}\boldsymbol{A_k}(\boldsymbol{z}^{-1}) \\ \boldsymbol{Q_{k+1}}(\boldsymbol{z}) &= \boldsymbol{A_k}(\boldsymbol{z}) - \boldsymbol{z}^{-(\boldsymbol{k}+1)}\boldsymbol{A_k}(\boldsymbol{z}^{-1}) \end{aligned} \qquad (1)$$

and $A_k(z)$ can be reconstructed by using these two filters:

$$A_k(z) = \frac{1}{2}\left(P_{k+1}(z) + Q_{k+1}(z)\right) \qquad (2)$$

The zeros of $P_{k+1}(z)$ and $Q_{k+1}(z)$ are on the unit circle, hence they can be expressed as $e^{j\omega}$, and the ω's are called line spectrum frequencies. Let the LSF vector be denoted by the vector $\boldsymbol{\omega}$ with elements $\omega_1, \omega_2, \ldots, \omega_k$.

Since the zeros of $P_{k+1}(z)$ and $Q_{k+1}(z)$ are interlaced with each other, the zeros of $P_{k+1}(z)$ and $Q_{k+1}(z)$ correspond to the odd and even indices, respectively.

3. TRANSFORM CODING OF LSF VECTORS

3.1. Fixed KL Transform

The fixed KLT can be expressed with its transform matrix $\boldsymbol{T}$. The relationship between the LSF vector $\boldsymbol{\omega}$ and the transformed vector $\boldsymbol{v}$ is the following:

$$\begin{aligned} v &= T \cdot \omega \\ \omega &= T^{-1} \cdot v = U \cdot v \end{aligned} \tag{3}$$

where $\boldsymbol{U}$ is the inverse transform matrix.

The autocorrelation (AC) matrix of the training data can be computed as follows

$$R_{\omega\omega} = E\left\{\omega\, \omega^T\right\} = \frac{1}{N}\sum_{i=1}^{N} \omega_i \omega_i^T \tag{4}$$

where N is the number of training vectors. Let $\boldsymbol{u_i}$ (normalized to unit norm) denotes the i-th eigenvector of $\boldsymbol{R_{\omega\omega}}$, where $i=1,2,\ldots,k$. The columns of the inverse KLT matrix $\boldsymbol{U}$ are the eigenvectors of $\boldsymbol{R_{\omega\omega}}$, i.e. $\boldsymbol{U}=[\boldsymbol{u_1}\,\boldsymbol{u_2}\ldots\boldsymbol{u_k}]$, and the KLT matrix is then $\boldsymbol{T}=\boldsymbol{U^{-1}}=\boldsymbol{U^T}$.

3.2. Adaptive KL Transform

Using one fixed KLT matrix is not optimal for individual frames, hence the KLT matrix must be changed adaptively. The first adaptive method is based on the prediction of the AC matrix and called AC-adaptive KLT.

Let $\boldsymbol{R}(n)$ denotes the estimated AC matrix after n frames, where $\boldsymbol{R}(0)$ is the AC matrix of the training data. The estimated AC matrix after n frames is updated as follows

$$R(n) = \alpha \cdot R(n) + (1-\alpha) \cdot \tilde{\omega}_n \tilde{\omega}_n^T \tag{5}$$

where α is an arbitrary positive real value, and $\tilde{\omega}_n$ is the decoded LSF vector which is also available at the decoded side.

The second adaptive method is based on the classification of the LSF frames using the voicing and pitch and called UV-adaptive (unvoiced-voiced) KLT. The first category is the unvoiced class. For the voiced frames, the range of the pitch value is divided into disjoint intervals so that the probability of each interval is approximately the same. Each interval corresponds to a category and each category has an own KLT matrix, which is determined from the same category of the training data. Let $c(n)$ denotes the category of frame n.

Based upon the experiments, a general speech LSF flow contains long voiced and unvoiced subsequences, and within a voiced sequence the difference between the neighbouring pitch values is generally small or zero, hence $c(n)$ and $c(n\text{-}1)$ are generally the same. The following AC-UV-adaptive method utilizes this fact: if the n-th frame is within a subsequence of category c, i.e. $c=c(n)=c(n\text{-}1)$, then the AC matrix of this category $\boldsymbol{R_c}$ is modified by using Equ. 5, but when $c(n)\neq c(n\text{-}1)$, then $\boldsymbol{R_c}$ is switched to $\boldsymbol{R_c}(0)$, which is the AC matrix of the category c of the training data.

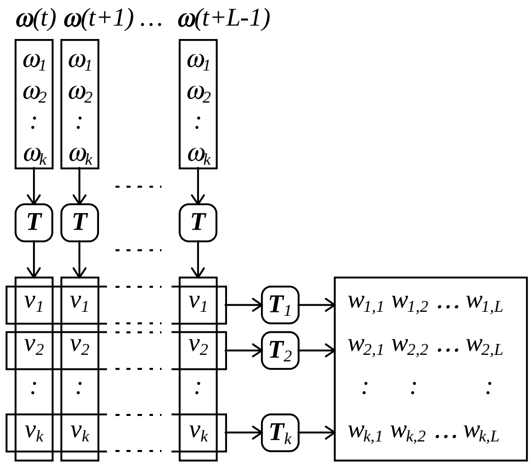

Figure 1 2D linear transform of LSF vectors, where $\boldsymbol{\omega}(n)$ denotes the LSF vector of the n-th frame

The AC-adaptive and hybrid UV-AC-adaptive method predict the AC matrix of the next frame, and the predicted matrix is used to calculate the transform matrix, hence these methods exploit the interframe correlation. In that follows, two methods are introduced for the transformed LSF vectors to reduce the interframe correlation more efficiently.

3.3. Two-Dimensional KLT

The direct way to exploit the interframe correlation of the transformed coefficients is using 2D KLT. As it is shown in Figure 1, L neighbouring k-dimensional (column) vectors are collected and transformed by using the fixed KLT matrix $\boldsymbol{T}$. The L resulting transformed vectors are transformed with k different KLT matrices $\boldsymbol{T_1},\boldsymbol{T_2},\ldots,\boldsymbol{T_k}$, which matrices transform L-dimensional (row) vectors. The resulting $k*L$ coefficients $w_{i,j}$ $(i=1,2,\ldots,k,\ j=1,2,\ldots,L)$ are encoded together, hence this approach introduces $L\text{-}1$ frame delay.

3.4. MA Prediction of the Transformed Parameters

Using MA prediction (MA-PR) in LSF domain is a powerful and popular technique. In this paper, a 4-th order MA prediction scheme is adopted for the transformed vectors $\boldsymbol{v_n}$. At this case, the decoded vector $\tilde{\boldsymbol{v}}_{\boldsymbol{n}}$ for the current frame n is obtained from the weighted sum of the previous 4 decoded vectors $\tilde{\boldsymbol{v}}_{\boldsymbol{n-1}},\ldots,\tilde{\boldsymbol{v}}_{\boldsymbol{n-4}}$ as follows

$$\tilde{v}_{n,i} = \left(1 - \sum_{l=1}^{4} p_{i,l}\right) \cdot \tilde{d}_i + \sum_{l=1}^{4} p_{i,l} \cdot \tilde{v}_{n-l,i} \tag{6}$$

where $\tilde{d}_i$ is the i-th coordinate of the decoded output, and $p_{i,l}$, $l=1,2,3,4$ are the coefficients of the MA predictor of the i-th coordinate.

Table 1 Performance of all schemes below 20 bits/frame for the training and test data

Scheme	Training data			Test data		
	SD	Outliers (%)		SD	Outliers (%)	
	(dB)	2-4	>4	(dB)	2-4	>4
Linear Transform + USSQ + VLC						
DCT	1.85	32.48	0.05	1.85	31.70	0.12
fixed KLT	1.72	23.07	0	1.72	22.65	0.03
UV	1.65	13.21	0	1.66	23.80	0.30
AC	1.72	31.37	0.04	1.72	32.08	0.03
UV-AC	1.64	17.69	0.03	1.62	15.91	0.03
MA prediction + USSQ + VLC						
LSF	1.62	21.64	0.55	1.62	21.38	0.77
DCT	1.53	11.33	0.01	1.54	11.23	0
KLT	1.54	12.40	0	1.54	11.89	0
2D DCT + USSQ + VLC						
L=2	1.72	25.02	0.02	1.71	23.54	0.09
L=4	1.62	18.05	0.02	1.62	18.04	0.03
L=10	1.51	11.86	0.01	1.52	13.13	0
2D KLT + USSQ + VLC						
L=2	1.59	20.31	0.005	1.60	20.25	0.02
L=4	1.45	8.71	0.005	1.45	8.10	0.005
L=10	1.38	6.68	0	1.39	7.10	0

Table 2 Performance of all schemes below 22 bits/frame for the training and test data

Scheme	Training data			Test data		
	SD	Outliers (%)		SD	Outliers (%)	
	(dB)	2-4	>4	(dB)	2-4	>4
Linear Transform + USSQ + VLC						
DCT	1.55	11.67	0	1.55	11.59	0
fixed KLT	1.51	9.80	0.005	1.51	9.79	0
UV	1.41	9.96	0	1.41	10.64	0
AC	1.44	11.12	0.02	1.44	12.57	0
UV-AC	1.36	4.67	0.01	1.36	4.35	0.06
MA prediction + USSQ + VLC						
LSF	1.46	14.70	0.30	1.44	13.69	0.44
DCT	1.31	3.43	0	1.31	4.26	0.03
KLT	1.31	4.67	0	1.31	5.03	0
2D DCT + USSQ + VLC						
L=2	1.55	13.30	0.01	1.54	13.36	0
L=4	1.39	6.59	0.01	1.39	6.86	0
L=10	1.34	5.32	0.01	1.35	6.03	0
2D KLT + USSQ + VLC						
L=2	1.40	5.60	0.005	1.40	5.74	0
L=4	1.25	2.63	0	1.26	2.72	0
L=10	1.19	2.07	0.03	1.20	1.27	0.06

The vector to be quantized for the current frame n is obtained from

$$d_{n,i} = \left(v_{n,i} - \sum_{l=1}^{4} p_{i,l} \cdot \tilde{v}_{n-l,i} \right) \Bigg/ \left(1 - \sum_{l=1}^{4} p_{i,l} \right) \qquad (7)$$

where $\boldsymbol{v}_n$ represents the original transformed vector[5]. Compared to the 2D transform, this scheme requires no ancillary frame delay, while the 2D transform does.

The KLT matrices $\boldsymbol{T}_1, \boldsymbol{T}_2, \cdots, \boldsymbol{T}_k$ and the predictor coefficients are computed from the transformed training data. Since the MA-PR and the 2D KLT use several previously transformed frames, they can not co-operate with adaptive KLT where the transform matrix $\boldsymbol{T}$ alters, hence the previously mentioned AC- and UV-adaptive schemes can not be combined with these methods.

4. EXPERIMENTS AND RESULTS

The speech database used here consists of about 25,000 frames taken from 11 male and 7 female speakers. About 21,500 frames are used for training, the remaining frames are used for test. The sampling rate was 8kHz, and *10*-th order LPC analysis is employed. The voicing and pitch information is also determined for each frame, and the pitch is used to switch between the AC matrices when UV-adaptive KLT is used.

The average spectral distortion (SD) and the number of outliers are computed for each scheme, which parameters are often used as a measure that correlates well with the quality of the reconstructed speech [8]. The results below 20, 22 and 24 bits/frame are shown in Table 1, 2 and 3, respectively.

In this experiments, the Huffman VLC is combined with uniform step scalar quantization (USSQ). Using USSQ, only the clipping level and the number of codepoints must be determined from the training data. Let δ_i denotes the variance of the i-th coordinate where $i=1,2,\ldots,k$, and let B bits are available for the quantization of a k-dimensional vector. The optimal allocated number of bits b_i and number of codepoints n_i can be computed as follows:

$$\boldsymbol{n_i = 2^{b_i}}\ , \quad \boldsymbol{b_i = \frac{B}{k} + log\left(\tfrac{\delta_i}{d}\right)} \qquad (8)$$

where $\boldsymbol{d^{2\cdot k}} = \left(\prod_{i=1}^{k} \delta_i^2 \right)$ [7].

For the UV-adaptive KLT, 5 categories (4 voiced and 1 unvoiced) are used. For AC-adaptive KLT, several α values are simulated and $\alpha=0.8$ was found to be optimal. To demonstrate the effect of the frame delay when 2D transform is used, the simulation results for $L=2,4$ and 10 are presented.

The 2D KLT with large delay ($L=10$) produces the best result and the performance of the 2D KLT is significantly better than the performance of the 2D DCT scheme with the same L. The quality of the 2D schemes increases as the frame delay gets larger. With MA-PR, the DCT and the KLT produces approximately the same results in SD while the DCT produces fewer outliers over 2dB. Despite the adaptive KLT outperforms the fixed KLT and the DCT, the adaptive KLT schemes is worse than the MA-PR and 2D schemes both in terms of SD and number of outliers. The UV-AC scheme insignificantly outperforms the other adaptive schemes at the lower bit rates (the difference is about 0.05 dB), but

Table 3 Performance of all schemes below 24 bits/frame for the training and test data

Scheme	Training data			Test data		
	SD	Outliers (%)		SD	Outliers (%)	
	(dB)	2-4	>4	(dB)	2-4	>4
Linear Transform + USSQ + VLC						
DCT	1.34	4.05	0	1.35	4.94	0
fixed KLT	1.27	2.50	0	1.27	2.66	0
UV	1.19	1.33	0	1.20	1.30	0
AC	1.21	1.95	0.005	1.21	2.25	0
UV-AC	1.18	1.72	0	1.20	2.13	0
MA prediction + USSQ + VLC						
LSF	1.31	10.02	0.10	1.28	8.52	0.15
DCT	1.12	1.08	0	1.11	0.83	0
KLT	1.11	1.26	0	1.11	1.21	0
2D DCT + USSQ + VLC						
L=2	1.33	4.66	0	1.34	5.77	0
L=4	1.20	2.26	0	1.20	2.69	0
L=10	1.13	1.38	0	1.14	1.80	0
2D KLT + USSQ + VLC						
L=2	1.21	1.96	0	1.21	2.19	0
L=4	1.08	0.73	0	1.07	0.98	0
L=10	1.02	0.41	0	1.03	0.83	0.03

at 24 bits/frame the SD is almost the same. Among these VLC schemes, the 2D KLT is the best solution, but when the frame delay is critical, the MA-PR either with KLT or with DCT is the most promising scheme.

The promising schemes are compared to other quantization schemes at 24 bits/frame and their results are tabulated in Table 4. While the results of a SQ-based scheme on the training and test data are approximately the same, a VQ-based scheme produces worse result on test data than on the training data. The best VLC schemes (2D KLT and KLT/DCT with MA-PR) can achieve the SD of the best VQ schemes while the number of outliers are significantly lower, which indicates better speech quality.

The UV-AC adaptive method (without VLC) can achieve approximately the same performance with SQ and with VQ, which indicates that the UV-AC adaptive KLT decorrelates the signal efficiently. However, the split VQ schemes with adaptive KLT introduces no significant improvement in speech quality. The KLT with split VQ and MA prediction produces slightly better quality than the two-stage VQ (2VQ) for LSF-MA-PR scheme, and the 2D KLT with SQ (without VLC) can also achieve this result with L=10.

5. CONCLUSIONS

In this paper, several schemes are introduced to encode the speech LSF parameters with USSQ at variable bit rate. Three KLT schemes are examined at different bit rates. The first family, the adaptive KLT introduces better performance than the simple KLT and the LSF-MA-PR, but cannot achieve the results of the LSF-MA-PR with VQ. Using MA-PR for the transformed coefficients, the performance of the USSQ-VLC scheme surpasses the best VQ schemes even with DCT. The performance of the USSQ-VLC scheme gets better by using 2D KLT, but the disadvantage of this approach is the ancillary frame delay.

Table 4 Performance of other schemes at 24 bits/frame for the training and test data

Scheme	Training data			Test data		
	SD	Outliers (%)		SD	Outliers (%)	
	(dB)	2-4	>4	(dB)	2-4	>4
Split VQ schemes						
LSF	1.23	4.71	0	1.31	6.77	0.03
KLT	1.14	3.30	0.01	1.35	9.67	0.27
UV	1.03	1.86	0.01	1.31	9.05	0.12
AC	1.25	6.56	0.12	1.30	8.10	0.50
UV-AC	1.20	5.55	0.21	1.27	7.13	0.47
MA prediction + Split VQ						
LSF	1.17	3.91	0.01	1.21	5.47	0.03
LSF-2VQ	1.08	2.22	0	1.16	3.99	0
DCT	1.13	4.34	0.02	1.17	5.44	0.12
KLT	1.08	3.13	0.02	1.12	4.26	0.06
KLT + SQ						
UV	1.40	11.57	0.27	1.42	12.77	0.24
AC	1.42	11.90	0.71	1.44	12.39	1.27
UV-AC	1.31	7.57	0.40	1.31	7.72	0.41
2D, L=2	1.34	8.21	0.05	1.35	9.37	0.12
2D, L=4	1.22	5.04	0.02	1.24	6.39	0.03
2D, L=10	1.15	2.91	0.01	1.17	3.11	0.06

REFERENCES

[1] F.K. Soong and B.H. Juang, "Line Spectrum Pair (LSP) and Speech Data Compression," in *Proc. ICASSP*, pp. 1.10.1-1.10.4, 1984

[2] G.S. Kang and L.J. Frosen, "Appication of Line-Spectrum Pair to Low-Bit-Rate Speech Encoders," in *Proc. ICASSP*, pp. 7.3.1-7.3.4, 1985

[3] J.R. Grosmer and T.P. Barnwell III, "A Low Bit Rate Segment Vocoder Based on Line Spectrum Pairs,"in *Proc. ICASSP*, pp. 7.2.1-7.2.4, 1985

[4] F.K. Soong and B.H. Juang, "Optimal Quantization of LSP Parameters," *IEEE Trans. Speech Audio Processing,* Vol. 1, No. 1. Pp. 15-24, Jan. 1993.

[5] ITU Recommendation G. 729, "Coding Speech at 8kbit/s Using Conjugate Structure Algebratic-Code-Excited Linear Prediction (CS-ACELP)," Dec. 1995.

[6] N. Farvadin and L. Laroia, "Efficient Encoding of Speech LSP Parameters Using the Discrete Cosine Transformation," in *Proc. ICASSP*, pp.168-171, 1989.

[7] N. S. Jayant and P. Noll, "Digital Coding of Waveforms: Principes and Applications of Speech and Video," Prentice-Hall, Inc., 1984

[8] K.K. Paliwal and B.S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/Frame," *IEEE Trans. Speech Audio Processing*, Vol. 1., No. 1., pp. 3-14, Jan. 1993