EUROSPEECH '97
5th European Conference on Speech Communication and Technology
Rhodes, Greece, September 22-25, 1997

SPECTRAL SENSITIVITY OF LSP PARAMETERS AND THEIR TRANSFORMED COEFFICIENTS

Hai Le Vu and László Lois
Department of Telecommunications
Technical University of Budapest
Sztoczek 2, 1111 Budapest, Hungary
Tel. +36 1 463 2093, FAX: +36 1 463 3266, E-mail: hai@hit.bme.hu

10.21437/Eurospeech.1997-185

ABSTRACT

In this paper, the optimal transformation and quantization of Line Spectrum Pair (LSP) are accomplished. Based upon the interframe and intraframe correlation properties of the LSPs, the Karhunen-Loeve (KL) transformation is adopted by Principal Component Analysis (PCA) neural network. The spectral sensitivity of the LSP and transformed coefficients are investigated in order to develop better scalar and vector quantizers for these coefficients. Using PCA network with spectral sensitivity guided quantizers we show that this new approach leads to as good as or better distortion compared to other methods for speech coding.

1. INTRODUCTION

While various methods for speech analysis-synthesis are known [1], the Line Spectrum Pair method, first introduced and studied by Itakura [2], is promising and popular methodology of representation of LPC parameters. The strong intraframe correlation has been considered in several coding schemes, for example, differential coding, 2-D DCT and time domain DPCM [3], etc. To exploit the non-uniform property of the differential LSP frames with keeping the subjective quality higher, the spectral sensitivity based quantizer design were adopted in [4].

In our system, we attempt to utilize the correlation between the LSPs with PCA network. The design of the non-uniform quantizers are based upon the spectral sensitivity of the transformed coefficients.
The emphasis of this work is on the efficient transformation of LSPs (also called Line Spectral Frequencies - LSFs) using PCA neural net and the quantization of the coefficient using a data-dependent optimality criterion.

2. THE PRINCIPAL COMPONENT ANALYSIS NEURAL NETWORK

If LSF parameter is a one-dimensional $P\times 1$ vector with autocorrelation matrix R, then the KL transform of LSF coefficients is defined by the fact that its rows are the eigenvectors of R. Since R is a real symmetric matrix, it follows that the eigenvalues are real, and that there are exactly P eigenvectors which are orthogonal and can be chosen to be orthonormal.
Therefore, it turns out that the KL diagonalizes the autocorrelation matrix of the transformed coefficients, i.e., the transformed coefficients are uncorrelated.

The results of investigating the intraframe and interframe correlation property of the LSF parameters [3] indicated that there is a strong correlation between the LSF coefficients of adjacent frames and neighbouring parameters in the same frame.
Then, any compression algorithm that effectively utilizes this correlation can result in improved performance over those that do not use this correlation. As mentioned above the KL transform produces a set of uncorrelated transformed coefficients, and for a given class of signals having the same second-order statistics, the KL transform is shown to be optimal [5].
In our scheme, we perform a one- and two-dimensional KL transformation on the LSF coefficient frame using PCA net [6].

Principal Component Analysis is a standard statistical technique which is defined in terms of the largest eigenvalues and the respective eigenvectors of the autocorrelation matrix of the input data. One-layer feedforward PCA is used with the Gram-Schmidt orthogonality procedure based on the learning rule as follows[7]

$$\underline{\underline{w}}_{k+1} = \underline{\underline{w}}_k + \alpha_k \left[\underline{y}_k . \underline{x}_k^T - LT\left(\underline{y}_k . \underline{y}_k^T\right) . \underline{\underline{w}}_k \right]$$

where $\underline{\underline{w}}_k , \underline{\underline{w}}_{k+1}$ is the weight matrix of the net at learning step *k* and *k+1*, respectively. α_k is the learning rate, x_k is the input sequence of vectors and y_k is the response of the net. LT( ) is the lower triangle matrix function which operates the Gram-Schmidt procedure.

In a 1-D KL coding scheme the intraframe correlation of LSF parameters is removed by the KL transform in the first dimension. The distribution plots of LSFs and KLs are shown in Fig.1 and Fig. 2 respectively. Furthermore, to reduce the interframe correlation between the neighbouring LSF frames, we adopt the 2-D KL transformation in the second dimension by independent PCA networks working parallel.

3. SPECTRAL SENSITIVITY ANALYSIS OF LSF AND KL COEFFICIENTS

We study the statistical property of LSF by using a speech data base consisting about 24,000 frames of male and female speech data, the sampling rate is 8 kHz, each frame is 20 ms long and 10th order LPC analysis is employed. This data base is also used for training the PCA neural network.

The spectral variation with respect to LSF frequency f_i is defined as

$$D_i = \left| \frac{\partial S}{\partial f_i} \right|^2 = \left| \frac{S(f_i + \Delta f_i) - S(f_i)}{\Delta f_i} \right|^2$$

where S is the log spectrum of a given LPC filter. The spectral sensitivity with respect to f_i is defined as

$$SEN_i = \int_{-FS/2}^{+FS/2} \left| \frac{\partial S}{\partial f_i} \right|^2 .df$$

where the magnitude squares of the spectral variation function are integrated and averaged over the whole frequency range (FS is sampling frequency).

The spectral sensitivity with respect to the transformed coefficient is computed via the LSF parameters by transforming forward and backward with the PCA neural network. It can be observed that the spectral changes due to perturbation of any given LSF frequency is highly localized around the specific frequency [4]. Since the LSF parameters are given back from the KL transformed coefficients using inverse transform, any change of a given KL will make change of all LSF frequencies. Thus the resultant spectral changes are spreading on all frequencies, which are summarized of every spectral change related to each LSF frequency weighted by the function of column vectors of the transform matrix.

The scatter diagram of LSFs and transformed coefficients spectral sensitivity were computed. The diagram of 1-D KL coefficients' spectral variations and sensitivity is presented in Fig.3 and Fig. 4 respectively.

Along the frequency domain (horizontal), it can be seen from the Fig. 3 that the spectral change of the most two principal components (KL1, KL2) is less than the changes of others and the spectral variations are not localized.

In Fig. 4 the horizontal axis indicates the transform domain, each diagram in the same line are at the same scale (the fisrt line is in range of 3000 and the next is in range of 1000). The scatter diagrams of KL1, KL2, KL6 and KL7 exhibit a decaying shape with non-uniform density, which property can be well utilized with a non-uniform quantizer. The remaining scatter diagram of the other principal component is more uniform. At the same scale the width of scatter diagram indicates the dynamic range of the coefficient.

The quantizers were designed with the well-known Lloyd-algorithm and with the K-means algorithm in higher dimensions. In both cases the distortion measure was the spectral distortion of the adequate parameters. The bit allocation was based on the relation among the variances and the average sensitivities of the coefficients.

4. EXPERIMENTS AND RESULTS

The performance test was based on a sequence of speech samples taken from 18 speakers (11 male and 7 female). Six short sentences were recorded for each speaker. About 10.000 frames of LPC vectors (both train and test data) were used in the experiments. The LPC Cepstrum Distance Measure (CD) and Log Likelihood Ratio (LR) are used for objective comparison of these encoding schemes. The average CD and LR are defined as follows

$$CD=\frac{1}{N}\sum_{n=1}^{N}\left[\left[c_x(0)-c_y(0)\right]^2+2\sum_{k=1}^{L}\left[c_x(k)-c_y(k)\right]^2\right]^{1/2}$$

and $$LR=\frac{1}{N}\sum_{n=1}^{N}\ln\left(\frac{1}{\pi}\int_0^{\pi}\left|\frac{A_y(e^{j\omega})}{A_x(e^{j\omega})}\right|^2 d\omega\right)$$

where $c_x(k), c_y(k)$ $k=0,1,2,...$ denote the speech cepstral coefficients and $A_x(z)$, $A_y(z)$ denote the analysis filters given by LPC coefficients of the *n*th speech frame of original and distorted speech, respectively. N is the total number of frames.

Method	CD (dB)	LR(dB)
PARCOR, Scalar Quant. 36 bits/frame	10.56	3.80
1-D KL, First 7 Par., Vector Quant. (4 and 3 dim. vectors) 22 bits/frame	10.67	4.05
2-D KL, Scalar Quant. 20 bits/frame	8.92	3.23

Table 1. Average CD and LR Distortion (dB)

The results printed in Table 1. indicate that the spectral sensitivity based quantizers of the transformed coefficients produce better quality in low bit rate.

5. REFERENCES

[1] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signals, Prentice-Hall, Englewood Cliffs, NJ, 1978

[2] F. Itakura, "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals", J. Acoust. Soc. Amer. vol. 57, S35(A), 1975

[3] N. Farvardin and R. Laroia, "Efficient encoding of speech LSP parameters using the discrete cosin transformation", Proc. Int. Conf. Acoust. Speech Signal Processing, pp. 168-171, 1989

[4] F.K. Soong and B-H. Juang, "Optimal Quantization of LSP Parameters", Proc. Int. Conf. Acoust. Speech Signal Processing, pp. 394-397, 1988

[5] N.S. Jayant and Peter Noll, "Digital coding of waveforms, Principles and Applications to Speech and Video", Prentice Hall, 1984

[6] H. L. Vu and L. Lois, "Optimal Transformation of LSP Parameters Using Neural Network", Proc. Int. Conf. Acoust. Speech Signal Processing, pp 1339-1342, 1997

[7] T. Sanger, "Optimal Unsupervised Learning in a Single-layer Linear Feedforward Neural Network", Neural Networks, Vol. 12, pp. 459-473, 1989

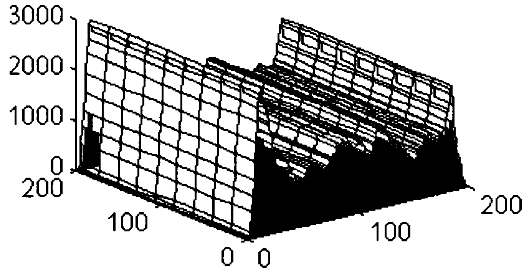

Fig. 1 Histogram of LSFs

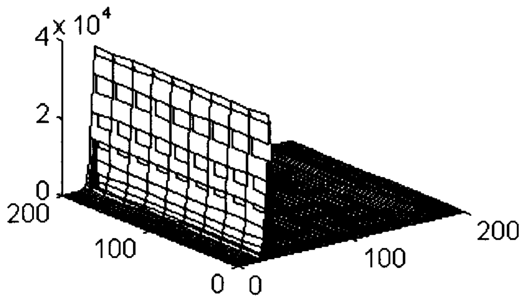

Fig. 2 Histogram of Kls

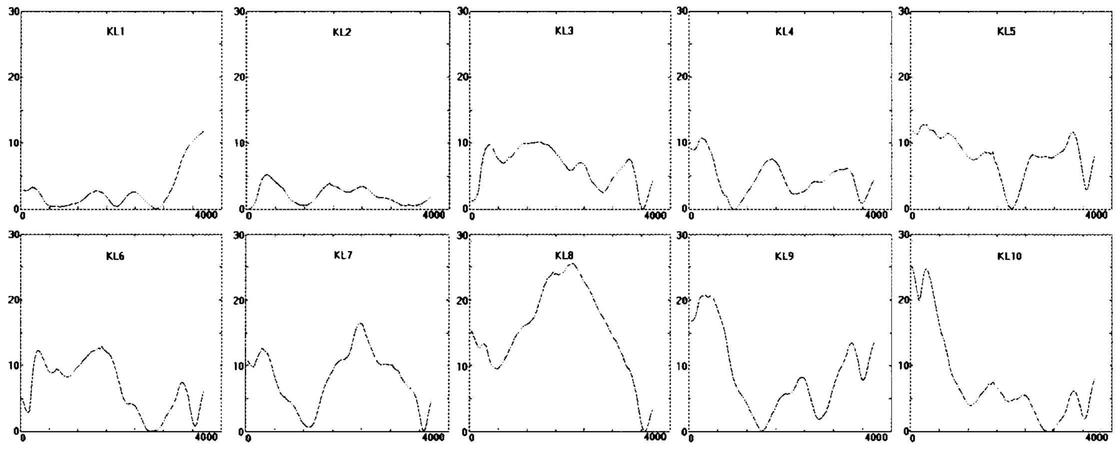

Fig. 3 Spectral Variations of Kls

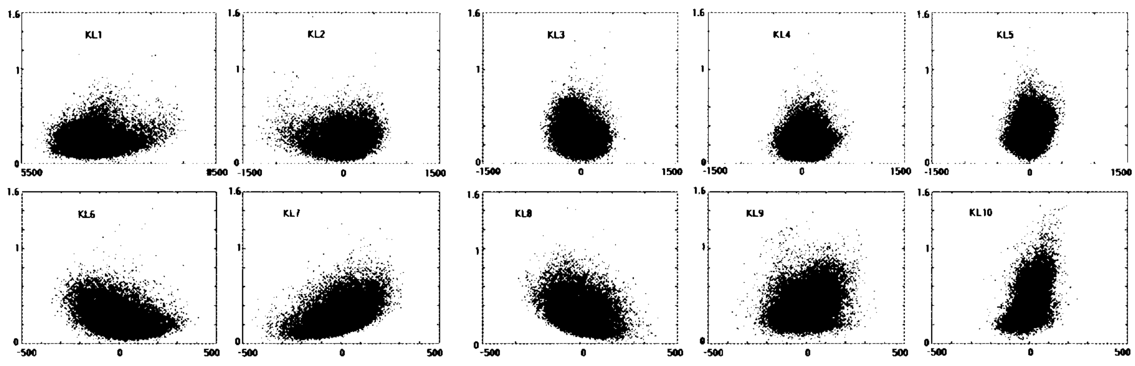

Fig. 4 Spectral Sensitivity Diagrams of Kls