OPTIMAL TRANSFORMATION OF LSP PARAMETERS USING NEURAL NETWORK

Hai Le Vu
hai@hit.bme.hu

Laszlo Lois
lois@hit.bme.hu

Department of Telecommunications
Technical University of Budapest
Sztoczek 2, 1111 Budapest - Hungary

ABSTRACT

In this paper, the intraframe correlation properties of Line Spectrum Pair (LSP) are used to develop an efficient encoding algorithm using the Karhunen-Loeve (KL) transformation. An important nonuniform statistical characteristics of LSP frequencies are investigated. Based upon this nonuniform property the neural network based techniques for generating the transform vectors via system training are studied. Using Principal Component Analysis (PCA) network to decorrelate LSP coefficients, we show that these new approaches lead to as good or better distortion as compared to other methods for speech analysis-synthesis.

Keywords: Speech coding; Low bit rate; Line Spectral Frequencies; Karhunen-Loeve transform; Principal Component Analysis network

1. INTRODUCTION

While various methods for speech analysis-synthesis are known [1], the Line Spectrum Pair (LSP) method, first introduced and studied by Itakura [2], is promising and popular methodology of LPC parameters' representation. The strong intraframe correlation has been considered in several coding schemes, for example, differential coding, 2-D DCT and time domain DPCM [3], etc. In our system, we attempt to utilize the correlation between the LSPs in an effort to reduce the average number of bits/parameter for a given level of quantization distortion.

A comparative study was conducted to investigate the efficiency of the neural net based KL coefficients in comparison with the conventional PARCOR and LSP parameters for both scalar and vector quantization.

The emphasis of this work is on the efficient and fast optimal transformation of line spectral frequencies (LSFs) using PCA neural net [4].
Another important purpose of this paper to develop a new loss encoder-scheme for encoding the LSFs which reduce the high-frequency distortion.

2. LINE SPECTRAL FREQUENCIES

The Line Spectrum Frequencies (LSFs or Line Spectrum Pairs LSPs) transformation of the LPC prediction coefficients was first introduced by Itakura in 1975 [2]. The starting point for deriving the LSFs is the response of the P order prediction error filter

$$A_k(z) = 1 - \sum_{k=1}^{P} a_k z^{-k} \tag{1}$$

The $\{a_k\}$ are the direct form predictor coefficients. In speech coding, the LPC coefficients are known to be inappropriate for quantization because of their relatively large dynamic range and possible filter instability problems. Different set of parameters representing the same spectral information, such as reflection coefficients and log area ratios, etc., were thus proposed for quantization in order to alleviate the above mentioned problems.

LSF parameters have both well-behaved dynamic range and filter stability preservation property, and can be used to encode LPC spectral information even more efficiently than many other parameters.

From Eqs. (1), $A_k(z)$ may be decomposed to a set of two transfer functions, one having an even symmetry, and the other having an odd symmetry. This can be accomplished by taking a difference and sum between $A_k(z)$ and its conjugate function as follows

Difference filter:

$$P_{k+1}(z) = A_k(z) - z^{-(k+1)}A_k(z^{-1})$$

Sum filter:

$$Q_{k+1}(z) = A_k(z) + z^{-(k+1)}A_k(z^{-1})$$

The LPC analysis filter, reconstructed by the use of these two filters, is

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

Three important properties of $P_{k+1}(z)$ and of $Q_{k+1}(z)$ are listed as follows:

- All zeros of $P_{k+1}(z)$ and $Q_{k+1}(z)$ are on the unit circle

- Zeros of $P_{k+1}(z)$ and $Q_{k+1}(z)$ are interlaced with each other

- Minimum phase property of $A_k(z)$ is preserved after quantization of the zeros of $P_{k+1}(z)$ and $Q_{k+1}(z)$.

Since all roots of $P_{k+1}(z)$ and $Q_{k+1}(z)$ are on the unit circle, they can be expressed as $e^{j\omega}$ and ω's are then called the LSP frequencies (LSF). The first two properties are useful for finding the roots of $P_{k+1}(z)$ and $Q_{k+1}(z)$. The third property ensures the stability of the synthesis filter.

3. THE KARHUNEN-LOEVE TRANSFORM

The discrete time Karhunen-Loeve transform is defined below:

Let $\underline{\underline{R}}_X = E\{\underline{X}\underline{X}^T\}$, denote the autocorrelation matrix of the LSF coefficients (column) vector $\underline{X}$. Let $\underline{u}_i$ denote the eigenvectors of $\underline{\underline{R}}_X$ (normalized to unit norm) and λ_i the corresponding eigenvalues. The Karhunen-Loeve transform matrix is then defined as

$$\underline{\underline{T}} = \underline{\underline{U}}^T$$

where $\underline{\underline{U}} = [\underline{u}_1 \underline{u}_2 ... \underline{u}_k]$, that is, the columns of $\underline{\underline{U}}$ are the eigenvectors of $\underline{\underline{R}}_X$.

The transform coefficients can be expressed in the form $\underline{Y} = \underline{\underline{T}}\underline{X}$, as the elements of column vector $\underline{Y}$. Then the autocorrelation matrix of $\underline{Y}$ is given by

$$\underline{\underline{R}}_Y = E\{\underline{Y}\underline{Y}^T\} = E\{\underline{\underline{U}}^T\underline{X}\underline{X}^T\underline{\underline{U}}\} = \underline{\underline{U}}^T\underline{\underline{R}}_X\underline{\underline{U}} = \mathrm{diag}[\lambda_1\lambda_2...\lambda_k]$$

where diag[...] is diagonal matrix with an element in the main diagonal.

The LSF coefficients are given back by inverse transform:

$$\underline{X} = \underline{\underline{T}}^T\underline{Y} = \underline{\underline{U}}\underline{Y}$$

4. STATISTICAL PROPERTIES OF LSF

We study the statistical property of LSF by using a different speech data base of male and female speech data, each frame is 20 ms long and 10th order LPC analysis is employed. The distribution plots of LSFs are shown in Fig. 1

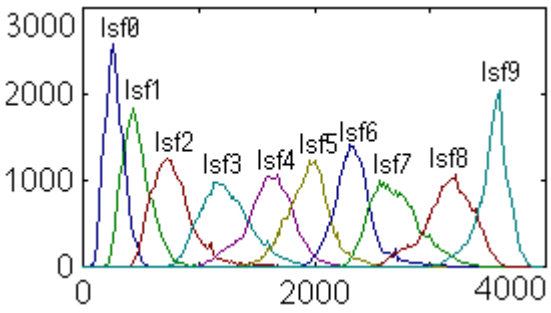

Figure 1: Histogram of LSFs
(Horizontal axis is in Hz)

A very important LSF property is the natural ordering of its parameters, this property indicates that the LSFs within frame are correlated. Thus after decorrelating a set of LSFs by Karhunen-Loeve transform, we will get a set of coefficients that has the following properties:

- Every coefficients are uncorrelated, thus we can quantize them independently

- Some of them have a dominant value, other remaining values are much smaller. Therefore the small value coefficients may not necessarily be quantized and transmitted (loss encoder-scheme).

- The dynamic range of Karhunen-Loeve (KL) coefficients is less than the dynamic range of LSFs, namely, these parameters can be more efficiently quantized.

- However the quantizers can operate on the transformed coefficients separately, but it is possible that the ordering property is violated by this procedure. There is a technique that reorders the quantized LSFs to satisfy the ordering property without increasing the distortion [3].

5. PRINCIPAL COMPONENT ANALYSIS WITH NEURAL NETWORK

In our scheme, we perform a one-dimensional KL transformation on the LSFs associated with each frame using PCA net. A PCA neural network is a one-layer feedforward neural network able to determine the principal components of the input vector stream.

Typically normalized Hebbian type learning rules are used.

The simple learning rule for *m* parallel neurons is introduced by Oja [4]:

$$\underline{\underline{W}}_{k+1} = \underline{\underline{W}}_{k} + \alpha_k \left[\underline{y}_k . \underline{x}_k^T - \left(\underline{y}_k . \underline{y}_k^T \right) \underline{\underline{W}}_k \right] \qquad (2)$$

where $\underline{\underline{W}}_k , \underline{\underline{W}}_{k+1}$ is the $m \times n$ dimensional weight matrix of the net at learning step *k* and *k+1*, respectively. α_k is the learning rate, $\underline{x}_k$ is the *n*-dimensional input sequence of vectors and $\underline{y}_k$ is the *m*-dimensional response of the net.

Learning rule (2) produces a normalized weight matrix, which determines the subspace of the principal vectors of the input stream in the *n*-dimensional space.

To complete learning rule (2) with the Gram-Schmidt orthogonality procedure, the weight matrix will be orthogonal (Sanger) [5]:

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

where LT( ) is the lower triangle matrix function which operates the Gram-Schmidt procedure.

The PCA neural-net is illustrated in Fig. 2 by the identical connections that exist between each input node and each output node. The training sequence of 240.000 samples (24.000 sets of 10-dimensional vector) is used for designing of PCA network.

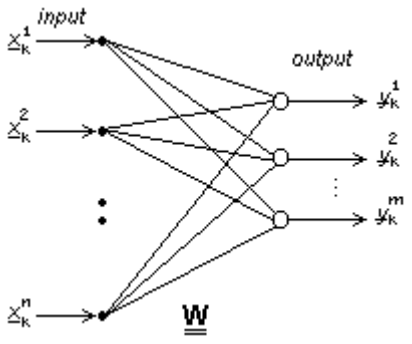

Figure 2: PCA Neural Network

6. OPTIMAL TRANSFORM CODING OF LSFs WITH PCA NET

It is known [6], that the high frequency components of the speech spectral envelope can be quantized even more coarsely, because spectrally less sensitive line spectral influence the all-pole spectrum near the perceptually less critical spectral valleys. In addition, LSFs lend themselves to frame to frame interpolation with smooth spectral changes because of their frequency domain interpretation.

For this reason there is a speech waveform method, in which the high-frequency coefficients are not transmitted and will be regenerated at the receiver from the lower frequency components. In this case the high-frequency parameters are totally lost, thus the high-frequency distortion is significant.

In this paper we propose a new method derived from KL optimal transform approximating with PCA net to spread the high-frequency distortion in frequency domain. Seven resultant dominant KL coefficients of PCA net are encoded and transmitted. At the receiver all LSFs are regenerated using inverse KL transform, therefore the overall distortion is lower.

7. 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 off a microphone. The speech signals were then digitized at an 8 kHz sampling rate. At 10-th order LPC analysis, based upon the autocorrelation method, was performed on the data using a 32 ms Hamming window.

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

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

and

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

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.

The training sequence of 240.000 samples (24.000 sets of 10-dimensional vector) was used for designing the PCA network and related scalar or vector quantizers. The generalized Lloyd algorithm (K-means clustering alg.) with binary split technique is used to design vector quantizer in our comparative studying [7].

Scalar quantization of PARCOR, LSF and KL coefficients (3.6 bits/parameter with {5,4,4,4,4,3,3,3,3,3} bit allocation) and vector quantization of 7 significant LSFs or 7 dominant KLs (22 bits/vector with codebook size of 4096 for first 4 coefficients and 1024 for the remaining 3 coefficients) of loss encoder scheme are studied. The performance results are presented in Table 1.

8. CONCLUSION AND FURTHER STUDIES

This paper presents the use of Karhunen-Loeve transform for the LSF coefficients with Principal Component Analysis Neural Network in low bit-rate speech coding. The basic idea in developing these schemes is using the correlation of LSFs to reduce the bit rate for a given level of fidelity.
All the result indicates that LSFs are strong correlates and we can reduce it efficiently by KL transformation using PCA neural net.

The three 36 bit scalar quantization schemes (M1), (M2) and (M3) produce approximately the same distortion. It can be seen that with a 40% reduction in bit rate requirements (from 36 bits/frame to 22 bits/frame) the method (M5) works almost the same as methods based on scalar quantizers and better than the method with no transformation (M4).

Further study in developing PCA network is possible with more complete learning rule. Replacing 1-D with 2-D Karhunen-Loeve transformation are among the interesting ideas for further research on this subject.

9. 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] E. Oja, "A Simplified Neuron Model as a Principal Component Analyzer", J. Maths. Biol., Vol. 16, pp.267-273, 1982

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

[6] George S. Kang and Lawrence J. Fransen, "Application of line-spectrum pairs to low bit rate speech encoders", Proc. Int. Conf. Acoust. Speech Signal Processing, 7.3.1-7.3.4, 1985

[7] H.L. Vu, "Speech Coding by the Optimal Transformation of LSP Parameters", Proc. IEEE Int. Conf. NORSIG'96, Espoo, Finland, pp. 119-122, 1996

Method (Analysis flow)		CD	LR
(M1) PARCOR → Scalar Q.	36 bits	2.43 (10.56 dB)	0.87 (3.80 dB)
(M2) LSF → Scalar Q.	36 bits	2.44 (10.61 dB)	0.88 (3.82 dB)
(M3) LSF → PCA → Scalar Q.	36 bits	2.44 (10.58 dB)	0.88 (3.81 dB)
(M4) LSF 7 Coeff. → Vector Q.	22 bits	2.61 (11.34 dB)	0.99 (4.30 dB)
(M5) LSF → PCA 7 Coeff. → Vector Q.	22 bits	2.46 (10.67 dB)	0.93 (4.05 dB)

Table 1: Average Cepstrum Distance Measure (CD) and Log Likelihood Ratio (LR)