PERIODICA POLYTECHNICA SER. EL. ENG. VOL. 41, NO. 1, PP. 75–83 (1997)

EFFICIENT ENCODING OF SPEECH LSF PARAMETERS USING THE KARHUNEN–LOEVE TRANSFORMATION

Hai Le VU

Department of Telecommunications
Technical University of Budapest
H–1521 Budapest, Hungary
Phone : + 36 1 463 2093, Fax : + 36 1 463 3266
email : hai@hit.bme.hu

Received: March 24, 1997

Abstract

Line Spectral Frequencies (LSF) provide an alternate parameterization of the analysis and synthesis filters used in Linear Predictive Coding (LPC) of speech. In this paper, a new method of quantization over the LSF vectors using the Karhunen–Loeve transform is presented. The new technique, which operates at about 1.8 kbits/s, has good quality and is not sensitive to quantization noise. When compared with other existing methods, this algorithm is better in terms of robustness and complexity.

Keywords: speech coding, low bit rate, Linear Predictive Coding, Line Spectral Frequencies, Karhunen–Loeve transform.

1. Introduction

Various speech Linear Predictive Coding methods have been studied for speech transmission at low bit rate. The LSF speech analysis method is known as one of the most powerful LPC techniques. The LSF parameters are completely equivalent, in a mathematical sense, to other LPC coefficients, such as the predictor or parcor coefficients. However, the LSF parameters have some interesting properties for quantization and interpolation which make them more attractive than the other LPC coefficients.

The emphasis of this work is on the efficient and accurate quantization of the LSF parameters. However, the improved performance of this system is based on properties of specific Karhunen–Loeve (KL) transform, which is determined from a statistical characteristic of LSF by analysing a speech data base.

The rest of the paper is organized as follows. Section 2 includes a review of the LSF speech analysis method, and we also mentioned the main properties of the LSF parameters. In Section 3, we describe the mathematical form of Karhunen-Loeve transform and we show how to use it to quantize efficiently the LSF coefficients in speech coding. In Section 4 we describe the statistical properties of the LSF coefficients in order to determine the Karhunen–Loeve transform matrix, and the suggested coding scheme using optimized nonuniform Lloyd–Max quantizer is presented. Finally, Section 5 contains our studies, conclusion and suggestions for further research.

2. Line Spectral Frequencies

The Line Spectrum Frequencies (LSF or Line Spectrum Pair LSP) transformation of the LPC predictive coefficients was first introduced by ITAKURA in 1975 [1]. The starting point for deriving the LSF parameters is the response of the order P prediction error filter

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

The coefficients $\{a_k\}$ are the direct form predictive coefficients. In speech coding, the LPC coefficients are known to be inappropriate for quantization because of their relatively large dynamic range. Moreover, one or other pole of LPC filter may exists outside the unit circle in consequence of the quantization error, which cause filter instability problems. Different sets 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. The 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 *Eq.* (1), $A_k(z)$ may be decomposed to a set of two transfer functions, one having even symmetry, and the other one having odd symmetry. This can be accomplished by taking a sum and a difference between $A_k(z)$ and its conjugate function as follows

Sum filter:

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

Difference 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

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

Three important properties of $P_{k+1}(z)$ and of $Q_{k+1}(z)$ are listed as follows [2]:

- 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 $(\omega_1 < \omega_2 < \ldots < \omega_p)$
- The 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 $e^{j\omega_i}$. The ω_i-s are then called the LSP frequencies (LSF). Since the LSF parameters are frequencies, they are quantized in the frequency axis, thereby the LPC filter is stable if the order of the LSF coefficients is valid. 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.

In the LSF representation some intrinsic properties of them permit more efficient quantization than other parameters [3, 4]. One useful property of the LSF is that an error in one line spectrum affects the all-pole spectrum only near that frequency (i.e., frequency selective). Thus, LSF coefficients may be quantized in accordance with properties of auditory perception, namely, coarser quantizer may be applied for high-frequency components of the spectral envelope. Another useful property is that a given LSF set has fixed spectral error sensitivities which can be estimated readily to benefit parameter quantization. We also mentioned that quantization error of the LSF coefficients does not cause filter instability problem because they do not act as a filter coefficient in this representation.

3. Karhunen–Loeve Transform

We used the orthogonal Karhunen–Loeve transform in order to reduce the residual redundancy in a set of the LSF coefficients. The discrete-time Karhunen–Loeve transform is defined below.

Let $\mathbf{R}_x = E\{XX^T\}$, the autocorrelation matrix of the column vector X of the LSF coefficients. Let u_i denote the eigenvectors of $\mathbf{R}_x$ (normalized to unit norm) and λ_i the corresponding eigenvalues. Since any autocorrelation matrix is symmetric and nonnegative definite, there are orthogonal eigenvectors (k vectors) and the corresponding eigenvalues are real and nonnegative.

The Karhunen–Loeve transform matrix is then defined as $\mathbf{T} = \mathbf{U}^T$, where $\mathbf{U} = [u_1 u_2 \ldots u_k]$, that is, the columns of $\mathbf{U}$ are the eigenvectors of $\mathbf{R}_x$.

We note that there is only one transform matrix $\mathbf{T}$, which can be determined from the average $\mathbf{R}_x$ over all speech-frame of the input sequence.

However, the Karhunen–Loeve transform can be applied directly to speech signal, but we will get the impracticable large matrix in this case.

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

$$\mathbf{R}_y = E\{YY^T\} = E\{\mathbf{U}^T XX^T \mathbf{U}\} = \mathbf{U}^T \mathbf{R}_x \mathbf{U} = \text{diag}\,[\lambda_1 \lambda_2 \dots \lambda_k].$$

Thus we see that the Karhunen–Loeve transform does indeed decorrelate the input vector, namely the elements of column vector Y are uncorrelated.

The LSF coefficients are given back by inverse transform:

$$X = \mathbf{T}^T Y = \mathbf{U}Y.$$

It has been known [5] that the Karhunen–Loeve transform is indeed the best possible transform for minimizing the overall distortion of quantization error for a given bit allocation.

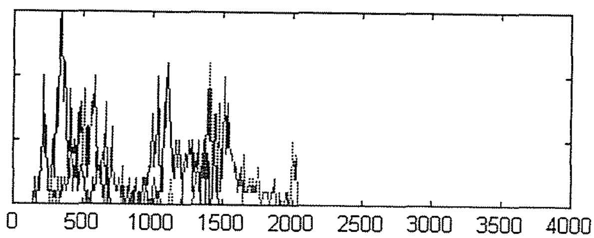

Fig. 1. The distribution plots for the first 4 LSF (horizontal axis is in Hz)

4. Statistical Properties of LSF and Coding Schemes

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 the first 4 LSF coefficients are shown in *Fig. 1* and the last 6 LSF parameters are shown in *Fig. 2*, respectively.

Our investigation indicates that the LSF parameters within frame, and from frame to frame, are correlated. Thus after decorrelating a set of the LSF coefficients by Karhunen–Loeve transform, we get a set of coefficients which has the following properties:

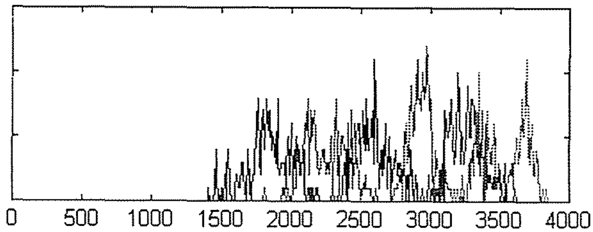

Fig. 2. The distribution plots of the last 6 LSF (horizontal axis is in Hz)

- 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 need not be quantized and transmitted.
- The dynamic range of Karhunen–Loeve coefficients is less than the LSF coefficients' dynamic range, namely, these parameters can be more efficiently quantized.
- Since the LSF coefficients are interlaced with each other, there is a possibility to interpolate from one coefficient to the other within frame or between frames in order to smooth the spectrum of speech.

From our previous discussion (Section 2), the high-frequency components of the speech spectral envelope can be quantized even more coarsely because a spectrally less sensitive line spectrum only influences the all-pole spectrum near the perceptually less critical spectral valleys. In addition, the LSF parameters lend themselves to frame to frame interpolation with smooth spectral changes because of their frequency domain interpretation. For this reason there exists 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.

In this paper we propose a new method derived from the Karhunen–Loeve transform to reduce the high-frequency distortions. The first 4 LSF coefficients are encoded directly into 4 bits for each parameter. The remaining coefficients then are transformed with Karhunen–Loeve matrix, but only three dominant Karhunen–Loeve coefficients are quantized (into 3 bits for each parameter) and transmitted.

The Karhunen–Loeve symmetric transform matrix (6×6) was computed from average $\mathbf{R}_x$ values from speech data bases (both of male and of female). We remark that this transform matrix is quite speaker independent, thus it is not required to transmit in every speech frame.

At the receiver three parameters, which have been chosen by us from our experiment of studying of different speech data bases, and other dominant Karhunen–Loeve coefficients produce the high-frequency components by inverse Karhunen–Loeve transform. The spectral envelope (formans structure) of voiced segments of original speech, reconstructed speech based on LPC and based on LSF are shown in *Fig. 3*, *Fig. 4*, *Fig. 5*, respectively:

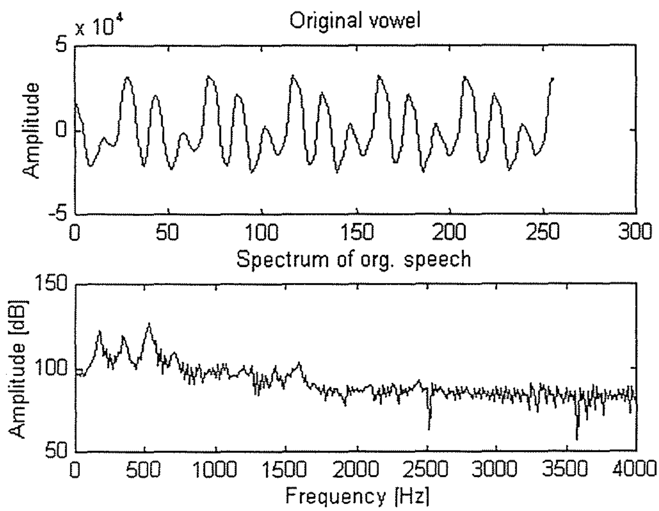

Fig. 3. Voiced segments and their short-time spectrum

It can be seen that with a 30% reduction in bit rate requirements (from 36 bits/frame to 25 bits/frame), the spectrum of reconstructed speech from LSF coders is similar to the spectrum of LPC coders. We have also investigated the assessment of speech quality by objective test, such as the Cepstrum Distance Measure [6]:

$$D = \left[[c_x(0) - c_y(0)]^2 + 2 \sum_{k=1}^{L} [c_x(k) - c_y(k)]^2 \right]^{\frac{1}{2}} \text{ [dB]}$$

where $c_x(k), c_y(k)$, $k = 0, 1, 2, \ldots$ denote the original and distorted speech cepstrum coefficients, respectively.

It is known that this spectral distortion measure is highly correlated with a subjective test result (the correlation has ranged from 0.8 to 0.9)

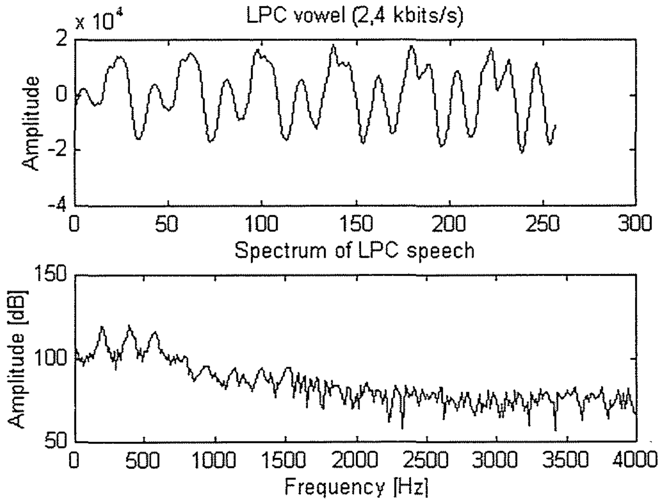

Fig. 4. Voiced segments from LPC coders and their spectrum

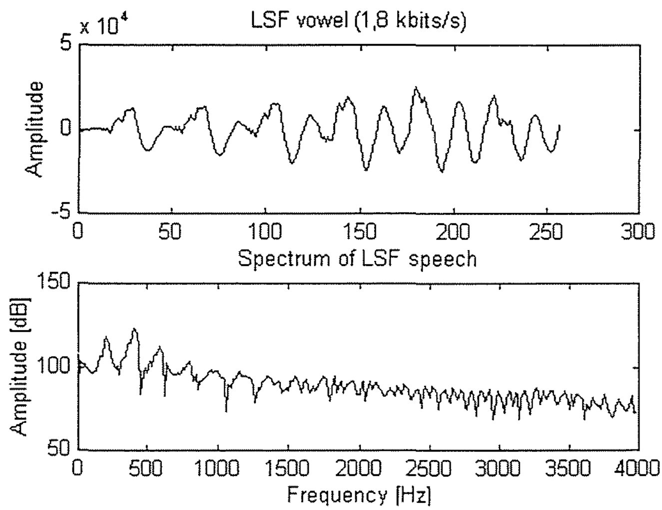

Fig. 5. Voiced segments from LSF coders and their spectrum

[3, 4]. To a great surprise, the resulting cepstrum distortion for LSF coders, when 25 bits are used to represent speech spectrum, is less than the distortion of 36 bits used in representing LPC coders. (The spectral distortions are 12.13 dB and 11.77 dB for LPC and LSF coders, respectively !). We

also mention that optimized nonuniform quantizers are used with Lloyd–Max algorithm to design minimal mean-squared error (distortion) quantization operation.

Our results of reconstructed speech both from coders based on LPC and based on LSF are shown in *Fig. 6*.

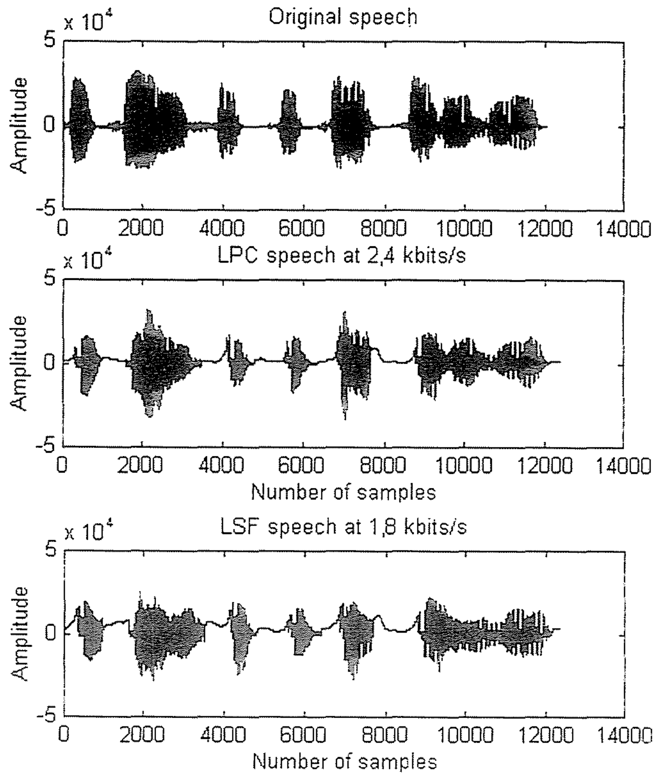

Fig. 6. Original and reconstructed speech in time domain

5. Conclusion and Further Studies

This paper presents the use of Karhunen–Loeve transform for the LSF coefficients in low bit rate speech encoding. The basic idea in developing these schemes has been to reduce the high-frequency distortions by transforming of the LSF coefficients before using quantization of speech parameters.

We note that the distortion caused by the LSF coders was less audible than for other coders with a popular parameter in speech coding. Further study in speech degradation is required, when the Karhunen–Loeve matrix is changed significantly between different speech sources, and we also try to eliminate this effect by interpolating the LSF coefficients within the speech frame.

Replacing the scalar quantization by the vector quantization is among the interesting ideas for further research on this subject [7].

References

1. ITAKURA, F. (1975): Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals, *J. Acoust. Soc. Amer.* Vol. 57, pp. S35(A).
2. SOONG, F. K. – JUANG, B. H. (1984): Line Spectrum Pair (LSP) and Speech Data Compression, *Proc. Int. Conf. Acoust. Speech Signal Processing*, San Diego, pp. 1.10.1–1.10.4
3. KANG, G. S. – FRANSEN, L. J. (1985): Application of Line Spectrum Pairs to Low Bit Rate Speech Encoders, *Proc. Int. Conf. Acoust. Speech Signal Processing*, Florida, pp. 7.3.1–7.3.4.
4. VU, H. L. (1996): A Low Bit Rate LPC Vocoder Model Base on Line Spectrum Frequency, *Proc. Int. Conf. Signal and Image Processing*, Florida, pp. 10–12.
5. JAYANT, N. S. – NOLL, P. (1984): Digital Coding of Waveforms, Principles and Applications to Speech and Video, New York, Prentice Hall.
6. QUACKENBUSH, S. R. – BARNWELL III, T. P. – CLEMENTS, M. A. (1988): Objective Measures of Speech Quality, Atlanta, Prentice Hall.
7. VU, H. L. (1996): Speech Coding by the Optimal Transformation of LSP Parameters, *Proc. IEEE Int. Conf. Nordic Signal Processing*, Helsinki, pp. 119–122.