A Line Spectrum Frequency Pairs Representation for Spectral Envelop Quantization

*Youngho Park, **WonCheol Lee, **MyungJin Bae

1-1 Sangdo-5-Dong, Dongjak-Gu Soongsil University, Seoul, Korea
**Dept. of Information and Telecommunication.
*Dept. of Electronics Engineering

ABSTRICT

This paper introduces a new type of representation of the LSPs as a promising alternative used for transmitting the LPC parameters. Major contribution in this paper is that the vocal track information embedded on the spectral envelope can be represented in terms of the reduced number of LSF compared to the conventional. Hence, it provides a possibility that LPC parameters could be quantized at a reduced bit rate without causing any major spectral distortion. The simulation result illustrates the capability of the proposed LSPs representation as an efficient quantization method via a proper rejection of the redundant pairs of pole and zero along the unit circle.

I. INTRODUCTION

Recently various speech coding schemes pursuing the standardization to application field, specially in mobile communication, have been investigated and developed. In low rate speech coding schemes, depending on how effectively to represent the excitation signal within a reasonable computational complexity, various modified schemes have been proposed so far [1]. But from the viewpoint of the bit rate reduction point, truly none of them hires more efficient procedure for quantizing the spectral envelope component considered as the vocal track information. In fact the vocal track component can be characterized in terms of several formant frequencies which can be imposed through the short-term predictor[2], namely the spectrum modeling filter.

From spectral point of view, it is well-known that the raw LPC parameters are not appropriate to be quantized due to its inherent wide dynamic range. Also the reconstructed modeling filter at receiver side might be unstable. To avoid the deficiency, due to its intrinsic robustness to the quantization error, the LSPs representation becomes a promising and popular methodology as a alternative of the LPC parameters[3, 4]. And its statistical characteristics has been investigated [5], also the efficient quantization schemes for a series of LSF has been introduced in [6]. Conventionally the following designated formula has been utilized as the LSPs representation[1], i.e.,

$$\frac{A_n(z) - zA'_n(z)}{A_n(z) + zA'_n(z)} \qquad (1)$$

In (1), $A_n(z)$ is the first kind Levinson polynomial of degree n and $A'_n(z)$ is the reciprocal polynomial to $A_n(z)$ defined by $A'_n(z) = z^{-n}A^*(1/z^*)$ and $*$ is a complex conjugate operator.

In this paper, a new line spectrum pair representation is introduced on the basis of the theory of a rational positive real function[4]. Towards this, the well established theory relavant to a positive real function together with its parameterization from will be discussed. In the following section II, more details about a positive real function are discussed and introduce a Foster function which is a special kind of the positive real function. And in Section III followed by the simulation result, a new spectrum pair representation is discussed including its feasibility.

II. CHARACTERISTICS OF POSITIVE REAL FUNCTION

In classical circuit theory, a rational function like (1) has been well-established and categorized in a class of Foster function [4], and its special spectral behaviour, i.e., the pole-zero interlacing, is one of its designed intrinsic properties. Furthermore it is well know that a Foster function is a special type of a positive real function. According to its application referred to its theoretical background, a positive function has been investigated and deeply considered for the factorization of any spectrum-like function into maximum- and minimum-phase systems[5]. [5] provides the parametrization form of a class of positive function $Z(z)$ as follows

$$Z(n)=\frac{B_n(z)+z^{-1}\rho(z)B'_n(z)}{A_n(z)-z^{-1}\rho(z)A'_n(z)} \qquad (2)$$

where $A_n(z)$ and $B_n(z)$ are first kind and second kind Levinson polynomials respectively and $\rho(z)$ is a bounded function. Notice that $A_n(z)$ and $B_n(z)$ are always free of zeros in $|z|>1$. Both types of Levinson polynomials are computed in recursive manner [6] from a given finite number of second order moments, namely correction sequences { $\gamma_0, \gamma_1, \gamma_2, \ldots, \gamma_n$}. One of the most interesting observations associated to a positive real function is its interpolation property such that the positive real function as in (2) interpolates the first n+1 correlation sequences. In other words, a rational positive function Z(z) satisfies

$$Z(z)=\gamma_0/2+\gamma_1 z^{-1}+\ldots+\gamma_n z^{-n}+O(z^{-(n+1)}) \qquad (3)$$

where $O(z^{-(n+1)})$are the higher terms above $z^{-(n+1)}$. Refer to the footnote 3, since the unity is an eligible candidate for a bounded function, substituting the unity gives rise to a special type of the positive function shown as

$$\begin{aligned} Z_F(z) &= \frac{B_n(z)+z^{-1}B'_n(z)}{A_n(z)+z^{-1}A'_n(z)} \qquad (4) \\ &= \sum_{m=1}^{M} P_m \frac{1+z^{-1}e^{-j\theta_m}}{1-z^{-1}e^{-j\theta_m}} \end{aligned}$$

therefore, as shown in (4), the Foster function has the degree of M which turns out to be the same as the number of exponential modes. From a practical viewpoint, in speech signal, the voiced signal can be characterized as a combination of pure sinusoidal signals, therefore a part of zeros of denominator polynomial of Foster function updated via (2) indicates the true modes and its degree should be the same as the number of modes. But since these are minor redundant poles and zeros involved, hardly it cannot state that the resulting Foster function has the degree which is the same as the number of modes. But fortunately it can be noticed that if the redundant modes are existing in a voiced signal, they are very closely residing side-by-side. As a result, those can be dismissed by dropping out from the unit circle on the basis of the properly chosen threshold.

III. A MODIFIED LSF REPRESENTATION

In CELP style coding method[1], to transmit the vocal track information, a rational function of a fixed degree as in (1) is updated and then only the angle differences are quantized and sent out. But in many cases, the a priori fixed degree of (1) is unnecessarily large compared to the number of existing dominant formant frequencies. But there is no way to remove the redundant poles and zeros of (1) based on their distribution on the unit circle. Refer to the discussion in section II, after neglecting the redundant poles and zeros by using the form (2), resulting rational function having reduced complexity would have the critical degree implying the number of existing fundamental sinusoidal modes. Thus, due to its intrinsic spectral behaviour, a Foster function as in (2) employed as a new LSF representation provides the improved efficiency rather than the conventional in the spectral envelope quantization. At a sequel, by making use of (2), it becomes possible that the number of bits allocated to LSF for the transmission can be reduced without making any major spectral degradations. To make a further progress, assume that the framed speech signal s(n), n=0→N-1 sampled at 8kHz is processed, the sequence of autocorrelation function are estimated as

$$\gamma_k = \frac{1}{N} \sum_{i=0}^{N-1-k} s(i)s(i+k), \qquad k = 0 \rightarrow M \tag{5}$$

where N is the total number of samples of a frame speech signal, and M is the number of autocorrelation sequence, assume that M<N. Next the first and second kind Levinson polynomials $A_M(z)$ and $B_M(z)$ of degree M are updated in recursive manner as follows[8].

$$\sqrt{1-s_{m+1}^2}A_{m+1}(z) = A_m(z) - z^{-1}s_{m+1}A'_{m(z)},\ m = 1 \rightarrow M, \tag{6}$$

$$\sqrt{1-s_{m+1}^2}B_{m+1}(z) = B_m(z) - z^{-1}s_{m+1}B'_{m(z)},\ m = 1 \rightarrow M, \tag{7}$$

where s_m is the reflection coefficients and the initial values are $A_0(z)=1$ and $B_0(z)=\gamma_0/2$. Then by making use of $A_M(z)$ and $B_M(z)$, the modified LSF representation as in (2) can be formed as

$$\frac{B_M(z) + z^{-1}B'_M(z)}{A_M(z) - z^{-1}A'_M(z)} \tag{8}$$

In the case of voiced speech, according to its spectral characteristics, the dismissible poles and zeros lying closely on the unit circle in pairwise are expected to be involved in (8). By setting a proper threshold in angles within 1% along the unit circle, these redundant pairs of pole and zero can be rejected. The at receiver side, a new Foster function can be constructed from the transmitted LSF. Here the degree of the reconstructed Foster function would be less than the original provided that the number of dominant mode is less than M, i.e., the degree of the original Foster function. Let L+1, L+1≤M+1 be the new degree of the Foster function, at the receiver side, L number of autocorrelations, $\{\gamma'_0, \gamma'_1, \ldots, \gamma'_L\}$ can be computed by solving a system of linear equations from the coefficients of the modified Foster function. And once again, by making the recursion as n (6), the first kind Levinson polynomial of degree L is obtained. Since the information of major dominant modes associated to the vocal track is not contaminated during the transmission, the original shape of the spectral envelope can be recovered without any major degradation. In unvoiced signal, e.g., fricative speech, corresponding framed signal can not be represented in terms of finite dominant modes, the poles and zeros are uniformly distributed along the unit circle. Thus, in the case, there is no efficitive reduction in bits to be allocated for the LSF.

IV. SIMULATION RESULTS

Figure 1 shows the performance of the proposed scheme. The power spectrum shown in Fig. 1(a) of the framed data which generated by passing the white gaussian noise through a rational system where three dominant poles are existing. And the magnitude response of the modeling filter $1/A_{10}(z)$ is depicted in Fig. 1(b). Using the conventional LSF representation as in (1) together with $1/A_{10}(z)$, the distribution of its poles and zeros can be shown in Fig. 1(c). As observing Fig. 1(c), it can be noticed that it is hard to distinguish the dominant modes playing major roles in shaping the spectral envelope. By setting a reasonable threshold value, in this case 0.75 degree, remove the redundant pairs of pole and zero in pairwise. As a result, the remaining poles and zeros are distributed as in Fig 1(d). These reduced number, $14 \equiv 2*(L+1)$, of LSF are quantized and coded through proper steps, and then transmitted to the receiver side. At the receiver side, based on a series of LSF, assume that there is no quantization error, the Foster function of degree L+1 is reconstructed. It is followed that a series of L taylor series coefficients which are the feasible autocorrelation sequencies are updated, then the Levinson polynomial $1/A_L(z)$ of degree L=8 can be computed. Fig. 1(b) shows the spectrum of the output signal of the modeling filter $1/A_L(z)$ excited by the same gaussian noise used to generate the actual signal used for Fig. 1(a). The recovered spectral envelope looks more-or-less the same as the original one as in Fig. 1(b). Noticeably, the three dominant modes are imposed and the contribution of the high frequency components are decayed. According to this result, the more bit rate can be reduced by sending the vocal track information relavant to the major dominant modes instead the overall locations of the poles and zeros.

V. CONCLUSIONS

This paper proposed the are formulation for representing the LSF, and it provides an opportunity considered as a major advantage of reducing the bit rate for transmitting the vocal track information. In many cases, the redundant LSF are involving as a part of information to be sent out. For simplifying the above overloaded process, the more effective quantization scheme was achieved by introducing the new LSF representation formula. And using this formula the redundant pairs of pole and zero inherently located nearby are dismissed and only the dominant ones are quantized for the transmission without making any major degradation on the spectral envelope. The proposed representation formula would be employed in any CELP type of speech coding scheme at the low- or variable bit-rate.

REFERENCES

[1] A. M. Kondoz, Digital Speech, Wiley, 1994

[2] M. R. Schroeder and B.S. Atal, Code-excited linear Prediction(CELP): high-quality speech at very low bit rates, Proc. ICASSP85, pp. 937-940

[3] F. Itakura, Line Spectral Representation of Linear Predictive Coefficients of speech signals, Journ. of Acoust. Soc. Amer., vol.57, Suppl., no.1, S35, 1975

[4] D.C.Youla, " The FEE: A new tunable high-resolution spectral estimator," Paty I, Technical note, no.3, Dept. of Electrical Engr. Polytechnic Inst. of New York, New York, 1993.

[5] T. W. Parks and C. S. Burrus, Digital Filter Design, NewYork : John Wiley, 1985

[6] C. Charalambous and A. Antoniou, " Equalization of recursive digital filters," Proc. Inst. Elec Engr., pt. G, vol.127, pp. 219-225,1980

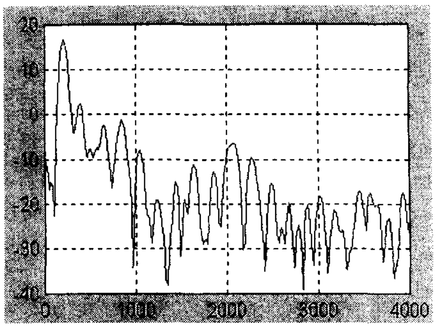

Figure 1(a). Actual speech signal

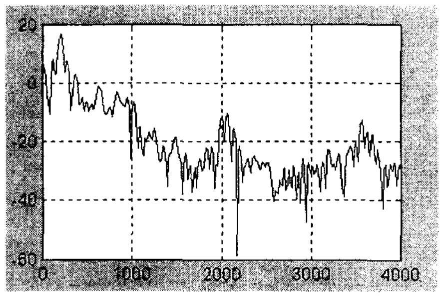

Figure 1(b). Reconstructed speech signal

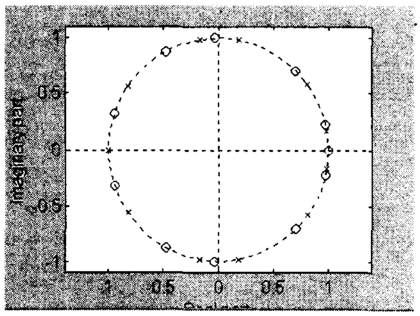

Figure 1(c). the pole and zero of actual speech signal

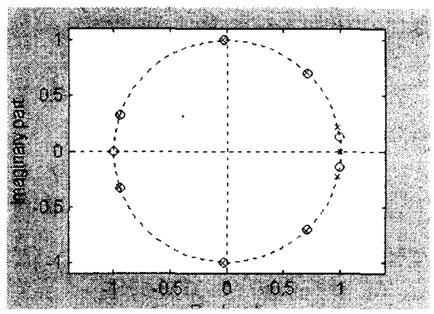

Figure 1(d). the pole and zero of reconstructed speech signal