LINE SPECTRUM PAIR (LSP) AND SPEECH DATA COMPRESSION

FRANK K. SOONG BIING-HWANG JUANG

Acoustic Research Department
Bell Laboratories
Murray Hill, New Jersey 07974

ABSTRACT

Line Spectrum Pair (LSP) was first introduced by Itakura [1,2] as an alternative LPC spectral representations. It was found that this new representation has such interesting properties as (1) all zeros of LSP polynomials are on the unit circle, (2) the corresponding zeros of the symmetric and anti-symmetric LSP polynomials are interlaced, and (3) the reconstructed LPC all-pole filter preserves its minimum phase property if (1) and (2) are kept intact through a quantization procedure. In this paper we prove all these properties via a "phase function." The statistical characteristics of LSP frequencies are investigated by analyzing a speech data base. In addition, we derive an expression for spectral sensitivity with respect to single LSP frequency deviation such that some insight on their quantization effects can be obtained. Results on multi-pulse LPC using LSP for spectral information compression are finally presented.

LINE SPECTRUM PAIR (LSP)

For a given order m, LPC analysis results in an inverse filter

$$A_m(z) = 1 + a_1z^{-1} + a_2z^{-2} + \cdots + a_mz^{-m} \tag{1}$$

which minimizes the residual energy [3]. In speech compression, the LPC coefficients $\{a_1, a_2, \cdots, a_m\}$ 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 [4,5] for quantization in order to alleviate the above-mentioned problems. LSP is one such kind of representation of spectral information. LSP 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.

The LSP representation is rather artificial. For a given mth order inverse filter as in (1), we can extend the order to $(m+1)$ without introducing any new information by letting the $(m+1)^{th}$ reflection coefficients, k_{m+1}, be 1 or -1. This is equivalent to setting the corresponding acoustic tube model completely closed or completely open at the $(m+1)^{th}$ stage. We thus have, for $k_{m+1} = \pm 1$ respectively,

$$A_{m+1}(z) = A_m(z) \pm z^{-(m+1)}A_m(z^{-1}) \ . \tag{2}$$

For convenience, we shall call these two polynomials $P(z)$ (*for* $k_{m+1} = 1$) and $Q(z)$ (*for* $k_{m+1} = -1$), respectively. It is obvious that $P(z)$ is a symmetric polynomial and $Q(z)$ is an anti-symmetric polynomial and

$$A_m(z) = \tfrac{1}{2}\,[P(z) + Q(z)] \tag{3}$$

Three important properties of $P(z)$ and $Q(z)$ are listed as follows:

(1) All zeros of $P(z)$ and $Q(z)$ are on the unit circle;

(2) Zeros of $P(z)$ and $Q(z)$ are interlaced with each other; and

(3) Minimum phase property of $A_m(z)$ is easily preserved after quantization of the zeros of $P(z)$ and $Q(z)$.

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

LSP PROPERTIES

We shall prove LSP properties in this section.

Rewriting equation (2) in a product form we have

$$P(z) = A(z)\left[1 + z^{-(m+1)}\,\frac{A(z^{-1})}{A(z)}\right] \tag{4}$$

$$Q(z) = A(z)\left[1 - z^{-(m+1)}\,\frac{A(z^{-1})}{A(z)}\right] . \tag{5}$$

We define

$$H(z) \triangleq z^{-(m+1)}\,\frac{A(z^{-1})}{A(z)} \tag{6}$$

which can be factored as

$$H(z) = z^{-1}\prod_{i=1}^{m}\frac{(z_i - z^{-1})}{(1 - z_iz^{-1})} = z^{-1}\prod_{i=1}^{m}\frac{(z_iz-1)}{z-z_i} \tag{7}$$

where z_i's are the zeros of $A_m(z)$, $z_i = r_ie^{j\omega_i}$, and $r_i < 1$. Since $|z_iz-1|^2 - |z-z_i|^2 = (1-|z|^2)(1-|z_i|^2)$,

$$|H(z)| \text{ is } \begin{cases} >1 & \text{if } |z|<1 \\ <1 & \text{if } |z|>1 \\ =1 & \text{if } |z|=1. \end{cases}$$

It is also obvious that $H(z)$ is the z-transform of an all-pass filter and we can write

$$H(\omega) = e^{j\phi(\omega)} \tag{8}$$

where $\phi(\omega)$ is the phase function of $H(\omega)$. Because the solution to $P(z) = 0$ or $Q(z) = 0$ requires that $H(z) = \pm 1$, we conclude that $P(z)$ and $Q(z)$ can only have zeros on the unit circle. The phase function $\phi(\omega)$ can further be expressed as

$$\phi(\omega) = -(m+1)\omega - \sum_{i=1}^{m} 2\tan^{-1}\frac{r_i\sin(\omega-\omega_i)}{1-r_i\cos(\omega-\omega_i)} . \tag{9}$$

The group delay of $H(\omega)$, defined as the negative slope of $\phi(\omega)$, is

$$\tau(\omega) \triangleq -\frac{\partial\phi(\omega)}{\partial\omega} = 1 + \sum_{i=1}^{m}\frac{1-r_i^2}{1+r_i^2 - 2r_i\cos(\omega-\omega_i)} \tag{10}$$

Again because all $r_i < 1$, we conclude that $\tau(\omega) > 1$ and $\phi(\omega)$ is a monotonically decreasing function. A typical phase function of $H(\omega)$ is depicted in Fig. 1. Also, $\phi(0) = 0$ and $\phi(2\pi) = -2(m+1)\pi$. Therefore, $\phi(\omega)$ crosses each $\phi = n\pi$ line exactly once resulting in $2(m+1)$ crossing points for $0 \leqslant \omega < 2\pi$. These crossing points constitute the total $2(m+1)$ zeros of $P(z)$ and $Q(z)$ alternately on the unit circle. Properties (1) and (2) are thus proved.

We want to show that in quantizing the LSP frequencies the reconstructed all-pole filter preserves its minimum phase as long as properties (1) and (2) are observed. To show this, it suffices to show that if $A(z)$ is non-minimum phase, properties (1) and (2) will be violated. We first show that $A(z)$ can not have zeros on the unit circle. Assume it has one at $e^{j\omega}$, i.e., $A(\omega) = 0$. Since $A(\omega) = A(-\omega)$, $e^{j\omega}$ is also a zero of $P(z)$ and $Q(z)$. But this is impossible because of property 2, which implies that $P(z)$ and $Q(z)$ have no common zeros. Next we show that zeros of $A(z)$ can not reside outside the unit circle either. Again, we assume $A(z)$ has some zeros outside the unit circle. We write $A(z)$ as a product of two polynomials,

$$A(z) = A_1(z)A_2(z) \tag{11}$$

CH1945-5/84/0000-0010 $1.00 © 1984 IEEE

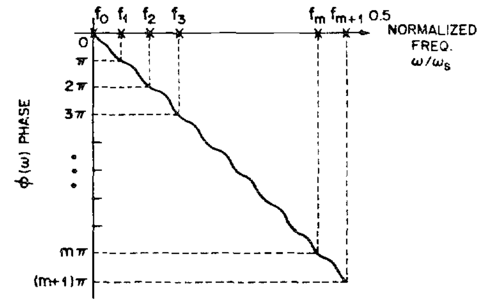

Fig. 1 A typical phase function of $H(\omega)$

where $A_1(z)$ is an ℓth order polynomial and $A_2(z)$ is an $(m-\ell)$th order polynomial representing the minimum and maximum phase parts of $A(z)$, respectively. The corresponding $P(z)$ and $Q(z)$ are

$$\left.\begin{matrix} P(z) \\ Q(z) \end{matrix}\right\} = A(z)\left[1 \pm z^{-1}\underbrace{\frac{A_1(z^{-1})z^{-\ell}}{A_1(z)}}_{H_1(z)}\;\underbrace{\frac{A_2(z^{-1})z^{-(m-\ell)}}{A_2(z)}}_{H_2(z)}\right] \tag{12}$$

Obviously, both $H_1(z)$ and $H_2(z)$ are all-pass functions with monotonically decreasing and increasing phase characteristics, correspondingly. The combined phase characteristics $\phi(\omega)$ traverses from 0 to $(m-2\ell-1)2\pi$ when ω goes around the unit circle. Thus, for $P(z)$ and $Q(z)$ to have together $2(m+1)$ zeros on the unit circle, the phase function $\phi(\omega)$ cannot be either monotonically decreasing or monotonically increasing because $\phi(\omega)$ would otherwise have only $|m-2\ell-1|$ crossings with $\phi = n\pi$ lines for $0 \leqslant \omega < 2\pi$, less than the $2(m+1)$ crossings required. (Note that if $m=\ell$, $A(z)$ is minimum phase.) Therefore, $\phi(\omega)$ has to either intersect some $\phi = n\pi$ lines more than once consecutively or be tangent to them to make the total crossings $2(m+1)$ as shown in the conceptual diagrams in Fig. 2(a) and (b).

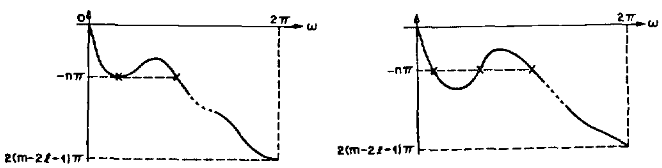

Fig. 2 (a) (b) Several possible intersections of $\phi(\omega)$ and $n\pi$ lines

But the case of multiple intersection violates the property (2) because if $\phi(\omega)$ crosses any integral π line more than once consecutively the corresponding zeros of $P(z)$ and $Q(z)$ are not interlaced. Furthermore, if $\phi(\omega)$ is tangent to $n\pi$ line at $\omega = \omega_\ell$, n odd, then

$$\left.\frac{dP(\omega)}{d\omega}\right|_{\omega=\omega_\ell} = \frac{d}{d\omega}\left[A(\omega)(1+e^{j\phi(\omega)})\right]_{\omega=\omega_\ell} \tag{13}$$

$$= \left\{[1+e^{j\phi(\omega)}]\frac{dA(\omega)}{d\omega} + A(\omega)\,e^{j\phi(\omega)}\frac{d\phi(\omega)}{d\omega}\right\}_{\omega=\omega_\ell}$$

$$= 0$$

because $e^{j\phi(\omega_\ell)} = -1$ and $\left.\frac{d\phi(\omega)}{d\omega}\right|_{\omega=\omega_\ell} = 0$. (Use $Q(\omega)$ for even n.) Let $e^{j\omega_i}$, $i = 1,2,\ldots, m+1$, be the zeros of $P(\omega)$. Clearly,

$$\left.\frac{dP(\omega)}{d\omega}\right|_{\omega=\omega_\ell} = \frac{d}{d\omega}\left[\prod_{i=1}^{m+1}(1-e^{-j(\omega-\omega_i)})\right]_{\omega=\omega_\ell}$$

$$= \left[\sum_{n=1}^{m+1} je^{-j(\omega-\omega_n)}\prod_{\substack{i=1\\ i\neq n}}^{m+1}(1-e^{-j(\omega-\omega_i)})\right]_{\omega=\omega_\ell}$$

$$= j\prod_{\substack{i=1\\ i\neq \ell}}^{m+1}(1-e^{-j(\omega_\ell-\omega_i)}) \neq 0 \tag{14}$$

Therefore $\phi(\omega)$ cannot be tangent to any $n\pi$ line to accumulate $2(m+1)$ crossings for $0 \leqslant \omega < 2\pi$ and we conclude our proof of property (3) that the reconstructed prediction polynomial $A(z)$ must be of minimum phase.

SINGLE-PARAMETER LOG SPECTRAL SENSITIVITY

Due to the interaction between LSP frequencies, it is difficult to analyze the log spectral sensitivity with respect to all LSP parameter variations in a closed form. We investigated only the single parameter log spectral sensitivity to gain some insight on the quantization effects of LSP frequencies.

The frequency response of the inverse filter is

$$A(\omega) = \tfrac{1}{2}[P(\omega) + Q(\omega)]$$

where

$$P(\omega) = \left(1+e^{-j\omega}\right)\prod_{i=1}^{m/2}\left(1-2\cos\omega_{2i-1}e^{-j\omega}+e^{-2j\omega}\right)$$

$$Q(\omega) = \left(1-e^{-j\omega}\right)\prod_{i=1}^{m/2}\left(1-2\cos\omega_{2i}e^{-j\omega}+e^{-2j\omega}\right) \tag{15}$$

are, respectively, the two LSP polynomials evaluated on the unit circle. A first order approximation of the log spectral deviation due to $\Delta\omega_{2i-1}$ is obtained as

$$(\Delta d)^2 \simeq \int_{-\pi}^{\pi}\frac{|\Delta A|^2}{|A|^2}\frac{d\omega}{2\pi}$$

$$= (\Delta\omega_{2i-1})^2\,\frac{\sin^2\omega_{2i-1}}{2}\int_{-\pi}^{\pi}\frac{1+\cos\phi(\omega)}{(\cos\omega_{2i-1}-\cos\omega)^2}\frac{d\omega}{2\pi} \tag{16}$$

where

$$\Delta A(\omega) \triangleq A(\omega_{2i-1}+\Delta\omega_{2i-1}) - A(\omega_{2i-1})$$

Similarly for even LSP frequencies ω_{2i} (i.e., LSP frequencies of $Q(\omega)$), the spectral deviation is

$$(\Delta d)^2 \simeq (\Delta\omega_{2i})^2\,\frac{\sin^2\omega_{2i}}{2}\int_{-\pi}^{\pi}\frac{1-\cos\phi(\omega)}{(\cos\omega_{2i}-\cos\omega)^2}\frac{d\omega}{2\pi} \tag{17}$$

We computed the spectral sensitivity with respect to each LSP frequency resulted from a 10th order LPC analysis. The speech data base, consisting of 4 sentences spoken by 4 male and 4 female speakers were recorded in a sound booth through microphones, and digitized at 8 kHz with 16 bit A/D. Over the whole data base, a time average of each single parameter spectral sensitivity was calculated. The results, which verify the analysis of (16) and (17) and are similar to those obtained by Sugamura and Itakura [2] by using a different approximation, are listed in Table I.

SPECTRAL SENSITIVITY (DB/HZ)

	mean	standard deviation
F1	.0188	.0070
F2	.0238	.0066
F3	.0135	.0046
F4	.0135	.0053
F5	.0144	.0046
F6	.0116	.0053
F7	.0130	.0043
F8	.0117	.0039
F9	.0108	.0026
F10	.0133	.0048

TABLE I

STATISTICAL PROPERTIES OF LSP FREQUENCY AND CODING SCHEMES

We studied the statistical properties of LSP frequencies by using a different speech data base which consists of 37000 frames of male and female speech data. Each frame is 20 ms long and the analysis rate is 100 frames/sec. A 10th order LPC analysis is employed and the resulting histograms of the ten different LSP frequencies are plotted in Fig. 3. As shown in the figure, the LSP frequencies are distributed orderly along the frequency axis. With these statistical characteristics it may be already apparent that simple coding schemes can be easily employed to efficiently

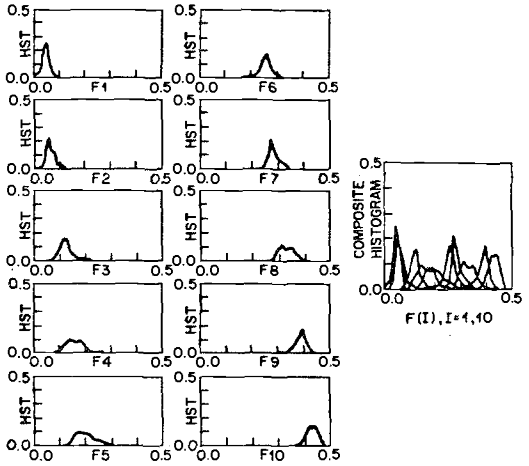

Fig. 3 Histograms of LSP frequencies

encode the LSP frequencies. However, we found that speaker characteristics, analysis as well as recording conditions contributed significant variation in the corresponding LSP frequency statistics. It then appears to be desirable to find another transformation of the LSP frequencies that are less susceptible to variations in speaker characteristics and analysis/recording conditions so that even more efficient quantization can be accomplished.

Using the same data base, we investigated the distributions of the differences between adjacent LSP frequencies i.e., $\omega_i - \omega_{i-1}$. The results are plotted in Fig. 4. The range of distributions are highly limited. About 99%

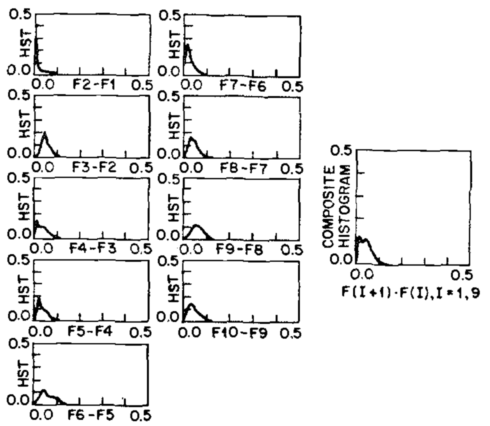

Fig. 4 Histograms of LSP frequency difference

of the frequency difference samples are within the range of 0 and 0.1 normalized frequency for the 10th order LPC analysis used. We found that the statistical characteristics, particularly the range, of the $\Delta\omega$ distributions remain practically the same for different speakers as well as various analysis and recording conditions. More efficient scalar encoding of speech spectral parameters can thus be achieved by quantizing the LSP frequency differences. We demonstrate below how this can be done with a simple encoding strategy that is similar to DPCM in waveform coding:

(1) Begin by quantizing ω_1 to $\hat{\omega}_1$ and setting $i = 1$;

(2) form the difference between ω_{i+1} and $\hat{\omega}_i$, $\Delta\omega_i = \omega_{i+1} - \hat{\omega}_i$;

(3) quantize $\Delta\omega_i$ to $\Delta\hat{\omega}_i$;

(4) reconstruct $\hat{\omega}_{i+1}$ as $\hat{\omega}_{i+1} = \hat{\omega}_i + \Delta\hat{\omega}_i$;

(5) if $i = m$ (analysis order), stop; otherwise, go to (2).

The corresponding block diagram is depicted in Fig. 5. We further compared

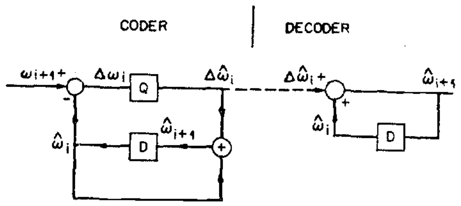

Fig. 5 Block diagram to encode LSP frequency difference

the coding efficiency of this proposed quantization scheme with one existing LPC parameter quantization scheme, namely, uniform quantization of log area ratios (LAR) with optimal (integer) bit allocation. We use the likelihood ratio distortion [6] as our objective measure. The resulting histograms and the corresponding mean and standard deviations of the distortions due to quantization are shown in Fig. 6. The total number of bits

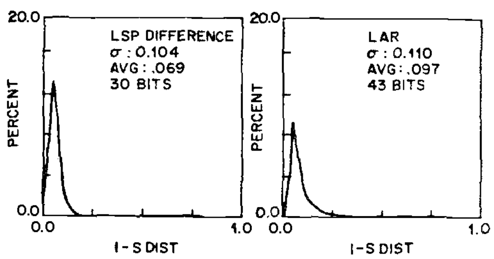

Fig. 6 Coding efficiency of LSP frequency difference and LAR.

used for LSP in 30, 3 bits for each LSP frequency difference, and the total number of bits used for LAR is 43 where the number of bits for each corresponding parameter is allocated as {6,5,5,4,4,4,4,4,4,3}. It can be seen that with a 30% reduction in bit rate requirements, the LSP difference scheme achieves an even lower average distortion level, compared to the LAR scheme. Also as a comparison, Itakura and Sugamura [7] reported a 25% reduction in bit rate requirements when uniform quantization of the straight LSP frequencies is employed.

COMPUTATIONAL ASPECTS OF LSP FREQUENCY FINDING

Computationally it is not a trivial task to find the roots of a polynomial in general. Fortunately, the LSP polynomials have some nice properties that allow us to alleviate the complicated root finding problem. First, the symmetric and the anti-symmetric polynomials have two real zeros, namely 1 and -1. Therefore, the $(m+1)$th order polynomials can be reduced to mth order symmetric polynomials:

$$\bar{P}(z) = P(z)/(1+z^{-1}) = 1 + p_1 z^{-1} + \cdots + p_1 z_{-(m-1)} + z^{-m}$$

$$\bar{Q}(z) = Q(z)/(1-z^{-1}) = 1 + q_1 z^{-1} + \cdots + q_1 z^{-(m-1)} + z^{-m} \qquad (18)$$

Second, since all the roots are on the unit circle, we need only to evaluate polynomials $\bar{P}(z)$ and $\bar{Q}(z)$ on the unit circle; in particular,

$$\bar{P}(z)|_{z=e^{j\omega}} = 2\, e^{-j\frac{m}{2}\omega}[\cos\frac{m\omega}{2} + p_1 \cos\frac{(m-1)\omega}{2} + \cdots + p_{m/2}]$$

$$= 2\, e^{-j\frac{m}{2}\omega} P'(\omega)$$

and,

$$\bar{Q}(z)|_{z=e^{j\omega}} = 2\, e^{-j\frac{m}{2}\omega}[\cos\frac{m\omega}{2} + q_1 \cos\frac{(m-1)\omega}{2} + \cdots + q_{m/2}]$$

$$= 2\, e^{-j\frac{m}{2}\omega} Q'(\omega) \qquad (19)$$

Given the coefficient sets, $\{1, p_1, \ldots, p_{m/2}\}$ and $\{1, q_2, \ldots, q_{m/2}\}$, the discrete cosine transform can now be used to find the roots of $P(z)$ and $Q(z)$. Before elaborating this approach, we shall discuss the case of an 8th order LPC. For telephone speech signals sampled at 6.67 kHz sampling rate, an 8th order LPC is probably adequate in spectral analysis and representation. The resulting $P'(\omega)$ and $Q'(\omega)$ are

$$P'(\omega) = \cos 4\omega + p_1 \cos 3\omega + p_2 \cos 2\omega + p_3 \cos\omega + p_4$$

$$Q'(\omega) = \cos 4\omega + q_1 \cos 3\omega + q_2 \cos 2\omega + q_3 \cos\omega + q_4 \tag{20}$$

which can be rewritten as two fourth order polynomial as

$$P'(x) = [8x^4 + 4p_1x^3 + (2p_2-8)x^2 + (p_3-4p_1)x + (1+p_4-p_2)]$$

$$Q'(x) = [8x^4 + 4q_1x^3 + (2q_2-8)x^2 + (q_3-4q_1)x + (1+q_4-q_2)] \tag{21}$$

In the above expression, $x = \cos\omega$. It is well known that any polynomial with order 4 or less can be solved through its radicals [8] in a closed-form. The root finding cost is then only nominal. For any LPC with order higher than 8, the root finding is more complicated. However, efficient algorithms are still available. Here we propose a sequential way to find them.

With an adequately fine grid, $\Delta\omega$, we evaluate the discrete cosine transform of the sequence $\{1, p_1, \ldots, p_{m/2}\}$, at $\omega = 0$ and $\omega = \Delta\omega$ first. If $P'(0)$ and $P'(\Delta\omega)$ are of the same sign, then by Descarte's rule there exists only an even number of zeros (i.e., 0, 2, 4, etc.) Since the grid is assumed to be adequately fine, no zero should exist between 0 and $\Delta\omega$. Such a procedure is repeated till the signs of $P'(i\Delta\omega)$ and $P'((i+1)\Delta\omega)$ are different. The interval $(i\Delta\omega, (i+1)\Delta\omega)$ is then dissected into two equi-distance subintervals and then the same sign comparison is used to determine which subinterval the zero is located. Repeat the dissecting procedure further till the prescribed accuracy (i.e., $|P'(\omega)| < \epsilon$) or the assigned quantization tolerance in encoding the LSP frequency is reached. After all zeros of $P(\omega)$ have been found, zeros of $Q(\omega)$ can be found in an even more efficient manner. Due to the interlace property of LSP frequencies, the search for zeros of $Q(\omega)$ is done in a prescribed interval, i.e., intervals between adjacent zeros of $P(\omega)$, and only one zero needs to be found in each interval. This algorithm is simple and effective. No FFT or fast DCT is necessary because the coefficient sequences $\{1, p_1, p_2, \ldots, p_{m/2}\}$ and $\{1, q_1, q_2, \ldots, q_{m/2}\}$ are relatively short for commonly used LPC order m. Also cosine table can be stored beforehand to speed up the computation.

LINE SPECTRUM PAIR AND MULTI-PULSE LPC

Multi-pulse LPC has been proposed by Atal and Remde [9] as a high quality, moderate bit rate voice coding technique. At 9.6 kbits/sec rate, the multi-pulse LPC requires around 2 kbits/sec for encoding the spectral information if reflection coefficients or log area ratio coefficients are used. If LSP frequency difference is adopted for encoding the spectral information only 1.5 kbits/sec is needed to achieve the same level of spectral distortion. The extra 500 bits/sec thus saved can be allocated to improve the encoding of excitation signals of the speech. Quality improvement was achieved in our simulation results.

CONCLUSION

We presented both our proof of the deterministic properties and findings of the statistical properties of the LSP representation of speech signals. With all the proven and found properties and efficient root finding schemes proposed, LSP is very well suited for efficient speech data compression. A 30% higher coding efficiency of LSP than LAR has been demonstrated.

ACKNOWLEDGEMENT

We would like to thank N. Sugamura and F. Itakura at NTT Masashino ECL in Japan for providing us with some useful references of their LSP research work.

References

[1] Itakura, F., "Line Spectrum Representation of Linear Predictive Coefficients of Speech Signals," J. Acoust. Soc. Am., 57, 535(A), 1975.

[2] Sugamura, N. and Itakura, F., "Speech Data Compression by LSP Speech Analysis-Synthesis Technique," Trans. IECE '81/8 Vol. J 64-A, No. 8, pp. 599-606.

[3] Markel, J. and Gray A., *Linear Prediction of Speech*, Springer-Verlag, 1976.

[4] Viswanasthan, R. and Markoul, J. "Quantization Properties of Transmission Parameters in Linear Predictive Systems," IEEE ASSP Trans., Vol. ASSP-23, pp. 309-321, June 1975.

[5] Gray, A. and Markel, J. "Quantization and Bit Allocation in Speech Processing," IEEE ASSP Trans., Vol. ASSP-24, pp. 459-473, Dec. 1976.

[6] Itakura, F. and Saito, S. "A Statistical Method for Estimation of Speech Spectral Density and Formant Frequencies," Electron. and Commun., Vol. 53-A, pp. 36-43, 1970.

[7] Itakura, F. and Sugamura, N. "LSP Speech Synthesizer," Tech. Rept. 5, Speech Group, Acoustical Society of Japan, Nov. 1979.

[8] Weisner, L., *Introduction to the Theory of Equations,* NY, Macmillan, 1938.

[9] Atal, B. and Remde, J. "A New Model of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates," Proc. ICASSP, Paris, France, 1982, pp. 614-617.