Coding of LPC Parameters for Low Bit Rate Speech Coders

C.C. Chu and P. Kabal

INRS-Télécommunications
3 Place du Commerce
Ile des Soeurs, Qué.
CANADA H3E 1H6

March 1987

Rapport technique de l'INRS-Télécommunications no. 87-19

Coding of LPC Parameters for Low Bit Rate Speech Coders

Abstract

This report summarizes the results of a study of the use of line spectral frequencies (LSF's) for the low bit rate coding of the linear predictive (LPC) parameters for use in a speech coder. Different forms of quantization using LSF's for the LPC coefficients are examined. An LSF based scheme allows the quantizer to take into account the perceptual impact of spectral distortion. One of the schemes considered takes advantage of the frame-to-frame correlation of the LSF parameters. The LSF based coding scheme is compared to a quantization based on a reflection coefficient representation. The application of this quantization scheme to the low bit rate, 4800 bits/sec, Code Excited Linear Predictive (CELP) coder is considered. In this context, adaptive gain factors for the differential (frame-to-frame) LSF quantizers are useful. In addition, a frame-to-frame interpolation scheme is proposed. With these modifications, LSF coding of 10 LPC parameters requires 1150 bits/sec.

Contents

Figures

Tables

Coding of LPC Parameters for Low Bit Rate Speech Coders

1. Introduction

In speech compression and transmission problems using linear predictive coding (LPC), the predictor parameters $\{a_i | i = 1, 2, \ldots, p\}$, of a linear formant predictor of order p have to be efficiently coded so as to keep the number of bits required small, while at the same time maintaining an acceptable amount of spectral distortion. While the LPC coefficients can be quantized directly, many other representations or equivalent transformations are available. Examples are (1) autocorrelation coefficients of the input samples, (2) direct form predictor coefficients, (3) predictor filter zero locations, (4) reflection coefficients, and (5) line spectral frequencies. These representations differ in the efficacy of coding the LPC parameters. For example, it has been found that a large number of bits is required to achieve good perceptual quality if the LPC coefficients are directly quantized. Two of the representations have the property that the stability of the corresponding synthesis filter can be guaranteed after quantization. These are the reflection coefficients [1] and the line spectral frequencies [2]. The performance of these two representations will be compared.

2. Reflection coefficients vs. line spectral frequencies

Both the reflection coefficients and line spectral frequencies are equivalent representations of the direct form linear predictor coefficients. While a single LPC coefficient does not have a one-to-one direct relationship with a reflection coefficient or a line spectral frequency, a complete set of the LPC coefficients can be transformed into a unique set of reflection coefficients and/or line spectral frequencies and vice versa.

In the past, a considerable amount of attention has been given to the study of reflection coefficients and its alternative representations. Non-uniform quantization of the reflection coefficients has become the standard method to code the LPC parameters for use in low-bit rate speech coders. An example of a coding algorithm for low-bit rate speech coding is the standard LPC-10 algorithm which implements a 2400 bits/sec LPC vocoder. A strong motivation for using the reflection coefficient representation is that these parameters can guarantee that the prediction error filter remains minimum phase after quantization. Since the synthesis filter in a speech coder is the inverse of the prediction error filter, the minimum phase condition is equivalent to stability of the synthesis filter. The reflection coefficients, $\{k_i | i = 1, 2, \ldots, p\}$, can be computed as a byproduct of the solution of the autocorrelation equations using the Levinson recursion, or directly from the direct form coefficients using a backward recursive formulation.

Reflection coefficients have been found to have non-uniform U-shaped spectral sensitivity. Non-linear quantization of these coefficients is necessary to efficiently make use of the bits assigned for quantization. Non-uniform quantizers can be represented as a cascade of a non-linear function (compressor), a uniformly spaces quantizer, and the inverse non-linear function (expander). Among many non-linear functions used for this purpose, log-area ratios and inverse-sine transformations have been used extensively [3] for reflection coefficient quantization. The log-area ratio coefficients g_i and the inverse-sine coefficients are defined in terms of the reflection coefficients k_i as follows.

$$\begin{aligned} g_i &= \log \frac{1+k_i}{1-k_i} \qquad && i = 1, 2, \ldots, p \\ \theta_i &= \sin^{-1}(k_i) \qquad && i = 1, 2, \ldots, p\ . \end{aligned} \tag{1}$$

These log-area ratio coefficients or inverse-sine coefficients are quantized using a uniform quantizer.

The line spectral frequency (LSF) representation is another transformation of the predictor filter parameters [2]. One of the important properties of the LSF's is that stability of the resulting synthesis filters is guaranteed upon quantization provided the natural ordering of the LSF's is maintained. An important benefit of an LSF representation is the ability to interpret the coefficients as frequency domain parameters. Kang and Fransen [4] show that the quantization noise due to the coding of a particular LSF manifests itself as distortion of the spectrum which is primarily local to the corresponding frequency.

The formant frequencies are the resonances of the synthesis filter and model the resonances of the vocal tract. Accurate modelling of the formant regions is known to be important for good quality speech reproduction. It has been found, both from theoretical arguments and experiments, that formant frequencies correspond to closely spaced line spectral frequencies. LSF's in close proximity to one another indicate the presence of formants, while isolated LSF's affect the spectral tilt. This provides a way to closely monitor the spectral distortion during quantization of the line spectral frequencies. A line spectral frequencies quantizer can be designed in such a way as to minimize the perceptual impact of spectral distortions.

Two basic quantizer design rules can be postulated. First, closely positioned line spectral frequencies should be quantized with more bits (using a fine quantizer). In this way, quantization error can be minimized around the formants. Second, high frequency LSF's can be quantized more coarsely than low frequency LSF's because frequency spectral distortion at high frequencies is more tolerable than at low frequencies. Stability of the synthesis filter is ensured if the LSF's remain correctly ordered after quantization.

Although the computation of a set of LSF's from a set of LPC coefficients requires solving for the roots of two polynomials in the z-plane, the roots are constrained to lie on the unit circle. Using a Chebyshev transformation, the search for roots can be carried out on the real line using a computationally efficient algorithm given by Kabal and Ramachandran [5]. With this algorithm, the computational burden is manageable and the use of LSF's becomes more attractive.

Both the reflection coefficients and LSF representations are useful transformations of the LPC filter parameters. The stability of synthesis filters based on quantized reflection coefficients or quantized LSF's can be ensured. Furthermore, linear interpolation of the parameters of either representation results in stable filters. However, because of the differences in spectral sensitivities and in the level of spectral information contained in each representation, they have to be quantized in different ways to maintain a perceptually low distortion. The perceptually important distortion can be controlled directly and effectively with the use of LSF's. In addition, the properties of the LSF's can be exploited to allow some spectral information to be regenerated at the receiver without explicit transmission of that information. For example, knowing the decoded value of penultimate LSF, the last LSF can be reinserted in between the penultimate LSF and the cutoff frequency using a fixed absolute or relative spacing. The resulting spectral distortion will be local to the upper part of the spectrum and is in general perceptually tolerable. In the case of reflection coefficients, although the higher indexed reflection coefficients are not as important as the lower indexed ones, the distortion they contribute to the spectrum is distributed across the spectrum. From the point of view of being able to easily control spectral distortion, quantization of LSF's is to be preferred to quantization of the reflection coefficients.

In the sequel, the performance of different line spectral frequencies quantizers will be studied and compared with the performance of the standard reflection coefficient quantizer. Later the LSF scheme will be adapted for use in a code-excited linear predictive (CELP) coder.

3. Line spectral frequencies quantizer designs

Line spectral frequency quantization of the LPC parameters has been used by a number of workers in the recent past. Among them are the designs by Kang and Fransen [4] and Crosmer and Barnwell [6]. These two designs were tested and reported to be superior to the conventional reflection coefficient quantizer in terms of a higher perceptual quality in the resulting synthesized speech and a lower bit rate requirement.

3.1 Scheme I: Center and offset frequencies quantization

The design by Kang and Fransen is studied first. As documented in their report for their 4800 bits/sec encoder/decoder design [4], the LSF quantizer codes the center and offset frequencies of successive pairs of LSF's. Let $l_i(n)$ be the i^{th} line spectral frequency of the n^{th} LPC analysis frame. Then the center and offset frequencies $\bar{l}_i(n)$ and $\delta l_i(n)$ are defined as follows.

$$\begin{aligned} \bar{l}_i(n) &= \frac{l_{2i}(n) + l_{2i-1}(n)}{2} \\ \delta l_i(n) &= \frac{l_{2i}(n) - l_{2i-1}(n)}{2} \end{aligned} \qquad i = 1, 2, \ldots, p/2 \; . \tag{2}$$

It is obvious from the definitions of the center and offset frequencies above that the $(\bar{l}_i(n), \delta l_i(n))$ pair and $(l_{2i}(n), l_{2i-1}(n))$ pair have a one-to-one relationship for $i = 1, 2, \ldots, p/2$. The quantized values of l_{2i}, l_{2i-1} can be recovered from the quantized values of $\bar{l}_i$ and δl_i. The advantage of quantizing the pairwise transformations of the LSF's is that it allows for efficient quantization and preserves spectrally important information. However a disadvantage with the quantization of center and offset frequencies is that each pair of LSF's is quantized without utilizing information about adjacent pairs. Unless the pairs of LSF's are far apart, the quantized frequencies may overlap or even cross over. The occurrence of crossovers violates the ordering requirement which guarantees the stability of the synthesis filter.

It is found that the quantizer design reported has stability problems. Assuming a ten pole ($p = 10$) linear formant predictor and an 8 kHz sampling frequency for the speech coding process, a set of center and offset frequencies are quantized using 21 bits with a 6% frequency resolution. As shown in Table 1 taken from [4], the center frequencies are quantized with the bit assignment $(4, 3, 3, 2, 2)$ and the offset frequencies with the bit assignment $(2, 2, 2, 1, 0)$. Note that the ranges of quantized values for $\bar{l}_1$ and $\bar{l}_2$ overlap. The natural ordering of the LSF's may not be preserved after pairwise quantization. It is noted from the table that the minimum quantized offset frequency for all LSF pairs is 300 Hz. As a consequence of this, a decoded pair of LSF's will be at least 600 Hz apart. Crossover of two adjacent line spectral frequencies from two successive pairs of LSF's

Filter Parameters		Frequency	No. of Bits
Center Frequency of LSF Pair	1	400, 420, 450, 480, 500, 530, 570, 600, 640, 670, 710, 760, 800, 850, 900, 950	4
	2	900, 950, 1010, 1070, 1130, 1200, 1270, 1350	3
	3	1430, 1510, 1600, 1700, 1800, 1900, 2020, 2140	3
	4	2260, 2400, 2540, 2690	2
	5	3020, 3200, 3390, 3590	2
Offset Frequency of LSF Pair	1	300, 350, 420, 480	2
	2	300, 350, 400, 420	2
	3	300, 350, 400, 420	2
	4	300, 350	1
	5	300 (fixed)	0
			Total ... 21

Table 1 Quantized center and offset frequencies used by Kang and Fransen.

can occur easily. This occurs with high probability for the low frequency LSF's since the center frequencies tend to be close together.

Experimentally, the LSF quantizer was evaluated in a simple configuration shown in Fig. 1. The input signal is sampled with an 8 kHz sampling frequency and a 10th order formant predictor is used as reported in [4]. The LPC coefficients are updated every 160 samples using the autocorrelation method with a 200 sample Hamming window. Pre-emphasis is turned off. The formant predictor filter and the synthesis filter are realized using the unquantized and quantized LPC coefficients, respectively. This is implemented by feeding the residual signal from the prediction error filter (unquantized coefficients) directly to the synthesis filter (quantized coefficients). The test configuration does not include the effects of residual signal quantization. The processing can be viewed as removing the formant structure of the speech and then reinserting a quantized formant structure.

As anticipated, crossovers of quantized LSF's do occur resulting in unstable synthesis filters.[†] Attempts were made to reorder the quantized LSF's based on their order before quantization. This strategy stabilizes the synthesis filter, but the resulting synthesized signals are very distorted.

3.1.1 Modification to Scheme I

In order to test the applicability of the basic model of the center and offset LSF quantizer, the quantizer output levels are redesigned in two different ways on the basis of the long time statistical

[†] Kang and Fransen [4], use a pre-emphasis filter with pre-emphasis factor $\mu = 15/16$. In our experiments, the crossover problem occurs with or without the pre-emphasis filter.

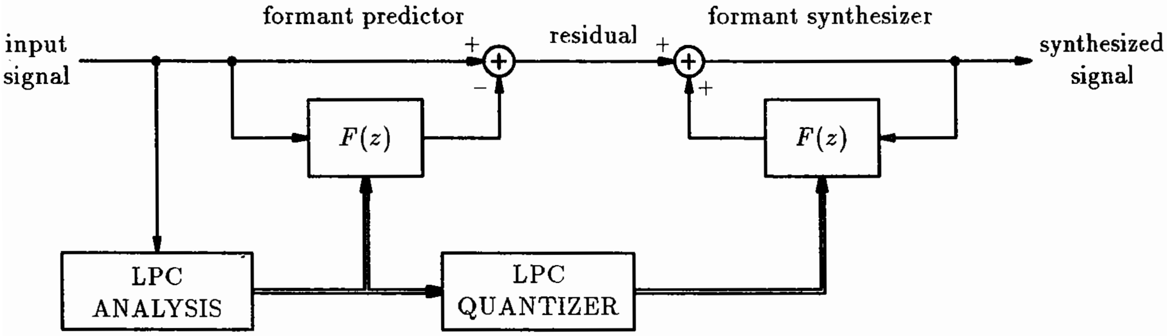


Fig. 1 A simple LPC vocoder model for the evaluation of the performance of LSF quantizers.

distribution of the center and offset frequencies. The distributions of the LSF's were determined for 8 utterances.

1. PIPM8: (male) The pipe began to rust while new.
2. DEPM8: (male) It's easy to tell the depth of a well.
3. CATM8: (male) Cats and dogs each hate the other.
4. CANM8: (male) The red canoe is gone.
5. TOMF8: (female) Tom's birthday is in June.
6. OAKF8: (female) Oak is strong and also gives shade.
7. HOGF8: (female) The hogs were fed chopped corn and garbage.
8. THVF8: (female) Thieves who rob friends deserve jail.

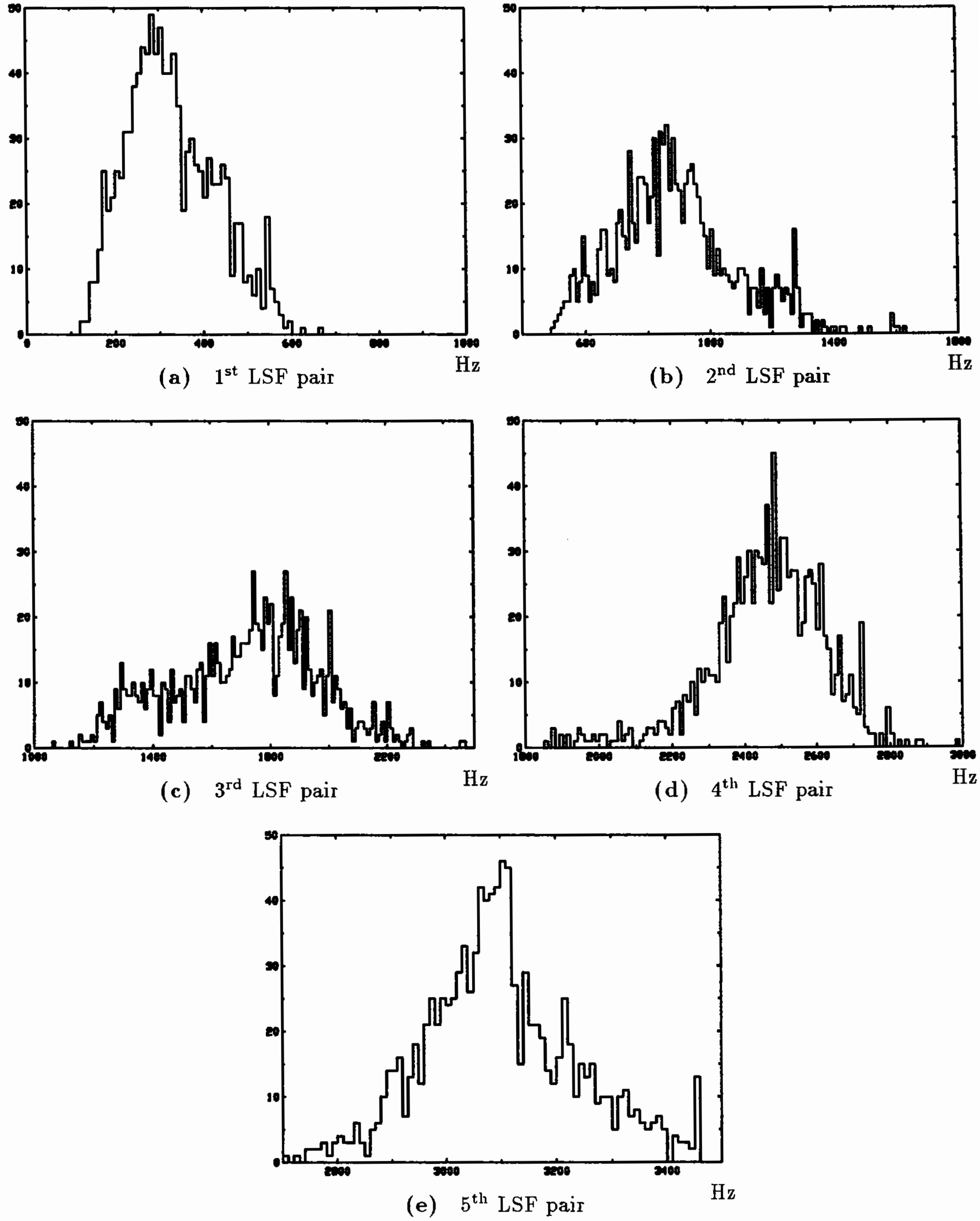
The resulting histograms of the corresponding center and offset frequencies are shown in Fig. 2 and Fig. 3 respectively.

In the first new design, the quantizers for the offset and center frequencies of LSF pairs are specified in a way similar to the original design in [4]. However, instead of using a quantizer table computed with a fixed 6% frequency resolution for LSF's, the required resolution is determined for each center frequency and each offset frequency separately. To find the quantizer output levels for $\bar{l}_i(n)$, the minimum and maximum output levels $\bar{l}_i^{\min}(n)$ and $\bar{l}_i^{\max}(n)$ are first determined from the i^{th} histogram. For a center frequency quantizer with N_i output levels, the frequency resolution γ_i satisfies

$$\bar{l}_i^{\max}(n) = \bar{l}_i^{\min}(n)[1+\gamma_i]^{N_i-1} \ , \tag{3}$$

and the k^{th} quantizer output level $\bar{l}_i^k(n)$ is given by

$$\bar{l}_i^k(n) = \bar{l}_i^{\min}(n)[1+\gamma_i]^k \qquad 0 \le k \le N_i - 1 \ . \tag{4}$$



(a) 1st LSF pair

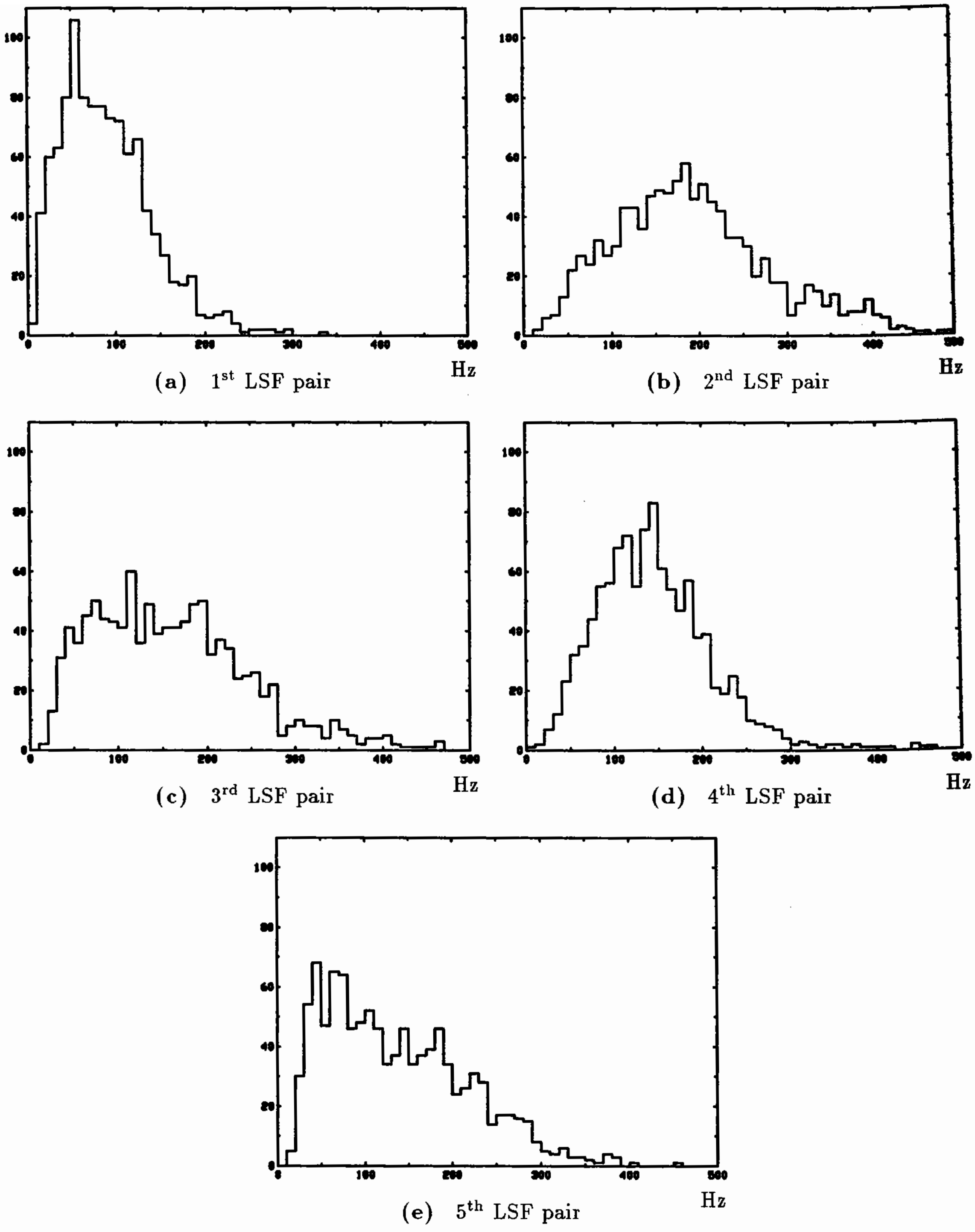
(b) 2nd LSF pair

(c) 3rd LSF pair

(d) 4th LSF pair

(e) 5th LSF pair

Fig. 2 Histograms of LSF center frequencies.



(a) 1st LSF pair

(b) 2nd LSF pair

(c) 3rd LSF pair

(d) 4th LSF pair

(e) 5th LSF pair

Fig. 3 Histograms of LSF offset frequencies.

With this convention, the output levels satisfy $\bar{l}_i^{\min}(n) \leq \bar{l}_i^k(n) \leq \bar{l}_i^{\max}(n)$.

The quantizer output levels for the j^{th} offset frequency are determined from the ranges of the offset frequency histograms. The resolution corresponding to the number of quantizer levels M_j, is computed from

$$\delta l_j^{\max}(n) = \delta l_j^{\min}(n)[1+\gamma_j]^{M_j-1} \ . \tag{5}$$

The k^{th} offset frequency quantizer output level is given by

$$\delta l_i^k(n) = \delta l_i^{\min}(n)[1+\gamma_i]^k \qquad 0 \leq k \leq M_j - 1 \ . \tag{6}$$

The decision levels for each offset frequency quantizer are set to be the mean values of pairs of adjacent output levels. The quantizers for the center frequencies and the offset frequencies have the output ranges and resolutions as listed in Table 2.

Filter Parameters	Indices	Parameter Range	No. of Bits	Resolution γ_i
Center Frequency of LSF Pair	1	100– 600	4	0.13
	2	500–1400	3	0.16
	3	1100–2300	3	0.11
	4	1900–2800	2	0.14
	5	2700–3500	2	0.09
Offset Frequency of LSF Pair	1	35–150	2	0.62
	2	120–300	2	0.36
	3	70–250	2	0.53
	4	100–170	1	0.70
	5	100–100	0	—

Table 2 Center and offset quantizers based on histogram ranges.

A second approach to center frequency / offset frequency quantization is more directly related to the statistical distribution of the LSF parameters. The quantizers are again designed from the histograms. In this case, the quantizer output levels are chosen such that the output levels divide the histograms into equal area regions. As a result, the quantizer output levels are more closely spaced for the frequencies at which LSF's occur with high probability. The decision levels for each quantizer are chosen to be the means of adjacent output levels.

These two new quantizer designs differ from the one reported in [4] mainly in that the offset quantizer output levels have much smaller minimum values. The probability of generating unstable synthesis filters due to the crossover of quantized line spectral frequencies is expected to be reduced. The second design is strongly dependent on the shape of the distribution of the LSF's as well.

These two quantizers were used in the speech coding model shown in Fig. 1. The input signals to the coder are among those used to generate the histograms. As shown in Tables 3 and 4, the quantization processes still create LSF crossover problems and hence instability of the synthesis filter. However, the number of frames with a crossover problem is much reduced from that observed with the Kang and Fransen design. Also, the second modified quantizer gives an improvement in synthesized speech quality.

Input Signal	Subjective Quality	
	Modification I	Modification II
CANM8	Unacceptable	Fair
DOUG3	Unacceptable	Fair
OAKF8	Unacceptable*	Fair
TOMF8	Unacceptable*	Fair

* Integer overflow occurs in the synthesis stage

Table 3 Comparison of subjective quality of synthesized signals using the two modification schemes.

Input Signal	No. of Crossovers in Quantized LSF's	
	Modification I	Modification II
CANM8	1	0
DOUG3	7	0
OAKF8	4*	0
TOMF8	15*	7

* Integer overflow occurs in the synthesis stage

Table 4 Number of crossovers in quantized LSF's resulting from the two modification schemes.

In view of the crossover and system instability problems encountered, three different steps were tested for the first modified quantizer in an attempt to maintain the correct ordering of the LSF's after quantization. The first technique is to exchange positions of two adjacent LSF's which have crossed over. The second technique shifts the quantized center frequency of the higher frequency pair up by one index. In the final fixup technique, the offset frequency of the higher frequency pair is shifted down by one index[†]. In effect, the last technique pulls the pair of quantized frequencies closer together. It is noted that these techniques are designed to reduce the number of crossover in the

[†] The second and third designs do not necessarily give crossover-free quantized LSF's because the extent of a shift in both cases is limited. Furthermore, the second design may create new crossover problems as a pair of LSF's is shifted upward towards another pair at a higher frequency.

quantized LSF's presented to the decoder without regard to any potential spectral distortion caused by them. Experimental results show a slight improvement in perceptual quality in the resulting synthesized signals, especially with the second technique. As expected the number of unstable synthesis filters decreases. Although the reordering mechanisms do slightly improve the coding in terms of an increase in perceptual quality of the synthesized signals, the performance, in general, is still poor with the small number of bits allocated for the quantization of the LSF's.

The major problem with the scheme considered above is the independent quantization of the LSF pairs. In the next section, another type of LSF quantizer which uses both time and frequency differences will be considered.

3.2 Scheme II: Even- and odd-numbered LSF quantization

Another scheme for LSF quantization at low bit rates was reported by Crosmer and Barnwell [6]. Instead of quantizing the LSF's pairwise as center and offset frequencies, this scheme codes the LSF's through a combination of time differences and frequency differences. The odd-numbered LSF's are quantized with differential pulse code modulation (DPCM) and the even-numbered LSF's are quantized relative to the neighboring odd-numbered LSF's. The LSF quantization thus makes use of the time as well as the frequency difference information in the line spectral frequencies.

It is found that the line spectral frequencies in steady speech segments have significant frame-to-frame correlation. A coder can therefore be designed to make use of this information to efficiently code the LSF's. In a DPCM coder, the difference between the input and its predicted value is quantized. Because the dynamic range of this difference signal can be much reduced by exploiting the correlation in successive input samples, the quantization error is reduced for a constant number of output levels. This results in the increase in the coding performance or in the reduction of the required bit allocation. However, if all line spectral frequencies were quantized with an independent DPCM coder, the instability problem encountered by the last model discussed would still be present. It is noted that formant structures are reflected by closely spaced adjacent odd- and even-numbered LSF's. The formant frequencies can be preserved by accurately quantizing the odd-numbered LSF's only. Therefore, it is sufficient to apply the DPCM coding technique to the odd-numbered LSF's. The distance of the even-numbered LSF's from its odd-numbered neighbors is separately quantized. Moreover, fewer bits are required to code the even-numbered frequencies because the information they usually represent, formant bandwidths and spectral tilt, can be quantized coarsely without significant perceptual impact [4][6].

The difference frequency quantizers for the even-numbered LSF's will depend on the frequency at which the LSF occurs. Crosmer and Barnwell [6], code a difference frequency as follows. The magnitude of the difference frequency is determined as the minimum distance to the surrounding odd-numbered LSF's (quantized values). The magnitude of this difference frequency is assigned a quantized value in a predetermined range. However if it is less than the minimum value, the minimum value is used, and if it is larger than the maximum value, it is set to the average of the surrounding odd-numbered LSF's. An appropriate fixed range of the frequency difference quantizers was obtained from subjective listening tests. We will modify this quantization technique, while keeping the essence of the scheme.

For typical speech signals, the line spectral frequencies tend to be more closely spaced at low frequencies and more widely spaced at high frequencies. Thus the range of the difference frequency quantizer should be dependent on the particular line spectral frequencies on either side of the even-numbered LSF being considered.

Kang and Fransen observed that the spectral sensitivity of the line spectral frequencies depends not only on the frequency but also on the relative position with respect to its neighbors. It is therefore natural to consider a relative spacing parameter,

$$D_j(n) = \left(\frac{l_j(n) - \hat{l}_{j-1}(n)}{\hat{l}_{j+1}(n) - \hat{l}_{j-1}(n)} \right) \qquad j = 2, 4, \ldots, p \ . \tag{7}$$

We assume p is even and $\hat{l}_{p+1}(n)$ is the line spectral frequency corresponding to the cutoff frequency (in our case, 4 kHz). Note that for correctly ordered LSF's, $0 < D_j(n) < 1$. The quantizer is chosen so that the quantized value $\hat{D}_j(n)$ also lies between 0 and 1. Because $D_j(n)$ depends on the quantized version of the odd-numbered frequencies, which are available at both the encoder and decoder, the quantized value of an even-numbered LSF can be recovered in the decoder using the decoded odd-numbered line spectral frequencies. No additional side information is required.

Given a set of quantized odd-numbered LSF's, $\hat{l}_i(n)$, $i = 1, 3, \ldots, p-1$, and an unquantized even-numbered LSF, $l_j(n)$, one of the six situations shown in Fig. 4 will occur. The first situation in Fig. 4 is the normal case. The order of the LSF's is maintained after $D_j(n)$ is quantized. In situations 2 and 3, the value of $D_j(n)$ is less than 0 and larger than 1, respectively. (As indicated in the histograms of Fig. 7, these situations can occur for real speech data.) Because the quantized value $\hat{D}_j(n)$ remains between 0 and 1, the order of the quantized LSF's is preserved. With situation 2, after quantization $\hat{l}_j(n)$ will lie just above $\hat{l}_{j-1}(n)$ and with situation 3, after quantization $\hat{l}_j(n)$ will lie just below $\hat{l}_{j+1}(n)$. The perceptual impact of the quantization error is believed to be small. However, in order to minimize the effect of all possible quantization errors, the relative distance is made a function of the *unquantized* LSF's only whenever these situations are detected. So the encoder will code

$$\tilde{D}_j(n) = \frac{l_j(n) - l_{j-1}(n)}{l_{j+1}(n) - l_{j-1}(n)} \ , \tag{8}$$

instead of $D_j(n)$ when $D_j(n)$ is less than zero or bigger than one. Situations 4, 5, and 6 are considered to be anomalies. The probability of two successive odd-numbered LSF's crossing over each other on quantization is almost zero at high frequencies due to the wide spacing of the high frequency LSF's. Also, at low frequencies, if the DPCM coders use fine enough quantizers, these situations should not arise. Therefore, the scheme essentially guarantees no crossovers of the LSF's.

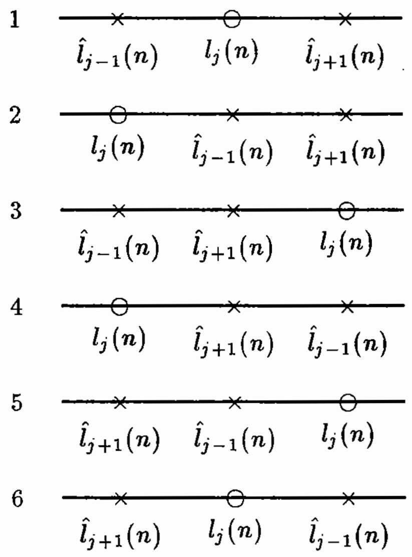

1
$\hat{l}_{j-1}(n)$ $l_j(n)$ $\hat{l}_{j+1}(n)$
2
$l_j(n)$ $\hat{l}_{j-1}(n)$ $\hat{l}_{j+1}(n)$
3
$\hat{l}_{j-1}(n)$ $\hat{l}_{j+1}(n)$ $l_j(n)$
4
$l_j(n)$ $\hat{l}_{j+1}(n)$ $\hat{l}_{j-1}(n)$
5
$\hat{l}_{j+1}(n)$ $\hat{l}_{j-1}(n)$ $l_j(n)$
6
$\hat{l}_{j+1}(n)$ $l_j(n)$ $\hat{l}_{j-1}(n)$

Fig. 4 Six possible sequences of the line spectral frequencies after the odd-numbered frequencies are quantized.

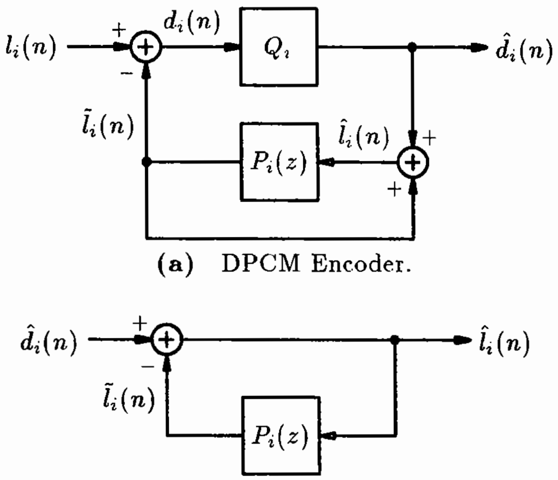
The odd-numbered LSF for the n^{th} analysis frame $l_i(n)$, $i = 1, 3, \ldots, p-1$, is coded and decoded using the DPCM encoder/decoder shown in Fig. 5. In the present case, the predictor is a simple delay, i.e. $\tilde{l}_i(n) = \hat{l}_i(n-1)$. Let $q_i(n)$, $i = 1, 3, \ldots, p-1$, be the quantization error incurred in quantizing the difference value $d_i(n)$, where

$$d_i(n) = l_i(n) - \tilde{l}_i(n) \ . \tag{9}$$

Then it can easily be derived from the block diagram that the error in quantizing $l_i(n)$ is also $q_i(n)$,

$$l_i(n) - \hat{l}_i(n) = q_i(n) \qquad i = 1, 3, \ldots, p-1 \ . \tag{10}$$

It is can be noted that Q_i in Fig. 5 is time invariant. The quantizer Q_i is designed based on the distribution of $d_i(n)$. For the purpose of evaluating the performance of this basic model, the histograms of $l_i(n) - l_i(n-1)$ are generated for $i = 1, 3, \ldots, p-1$ using the 8 speech signals listed before. Given the distributions of this time difference signal (see Fig. 6) the quantizers are designed by dividing each histogram into regions of equal area (probability of occurrence). In view of the differing perceptual impact of spectral distortions at different frequencies, more number of bits are assigned to the lower frequency LSF's than to the high frequency LSF's. Moreover, because of the



(a) DPCM Encoder.

(b) DPCM Decoder.

Fig. 5 DPCM encoder/decoder for LSF coding.

importance of the DPCM encoding/decoding process, more bits are given to the odd-numbered LSF quantization than to the quantizers for the difference frequencies. The bit assignments $(4,3,3,2,2)$ are used for the odd-numbered LSF's.

The relative difference frequency $D_j(n)$, $j = 2,4,\ldots,p$, is quantized using a time invariant quantizer Q_j. The resulting quantization error can be expressed as

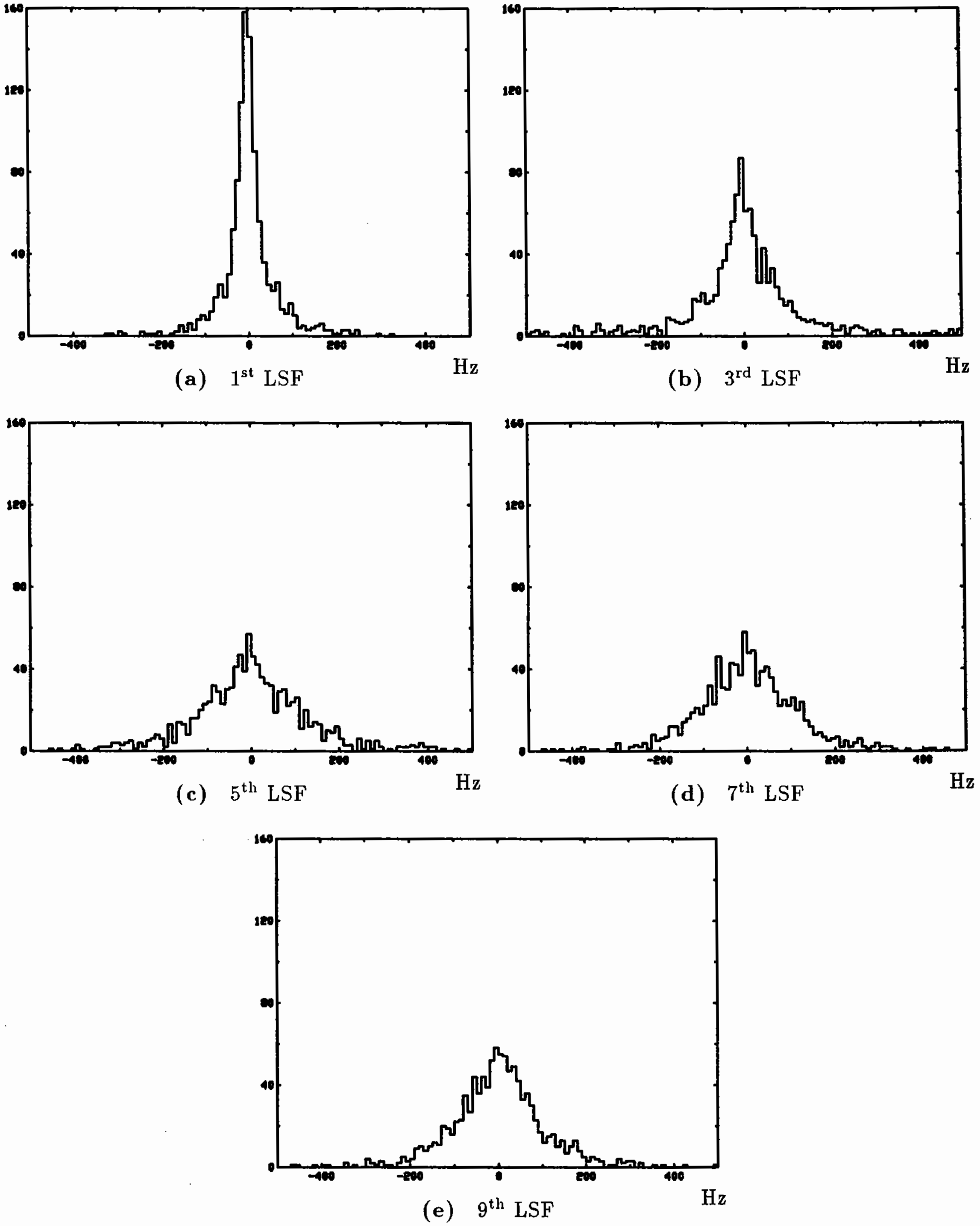
$$l_j(n) - \hat{l}_j(n) = q_j(n)\big(\hat{l}_{j+1}(n) - \hat{l}_{j-1}(n)\big) \qquad i = 2,4,\ldots,p \tag{11}$$

where $q_j(n) = D_j(n) - \hat{D}_j(n)$. Given the quantizers for all odd-numbered frequencies, the histograms of $D_j(n)$, $j = 2,4,\ldots,p$ are obtained as shown in Fig. 7. The quantizers Q_j's are designed by dividing each histogram into regions of equal area. The bit assignments $(2,2,2,1,0)$ are used for the even-numbered LSF's. The p^{th} LSF is given 0 bits, since it can be regenerated at the decoder without significant perceptual impact.

3.3 Computer simulation

The performance of the LSF coding scheme is studied in the context of the system shown in Fig. 1. Four different speech signals are used as input to the system with this LSF quantizer design. The subjective quality of the output signals is evaluated and compared with the original input signals in informal subjective tests. During the course of simulation, the encoding/decoding process occurs without a single incident of LSF crossover.

It is found that in general, the subjective quality of the synthesized output is as good as the original with little noise, chirps, clicks, or other distortions. However, the reproduced speech signals



(a) 1st LSF

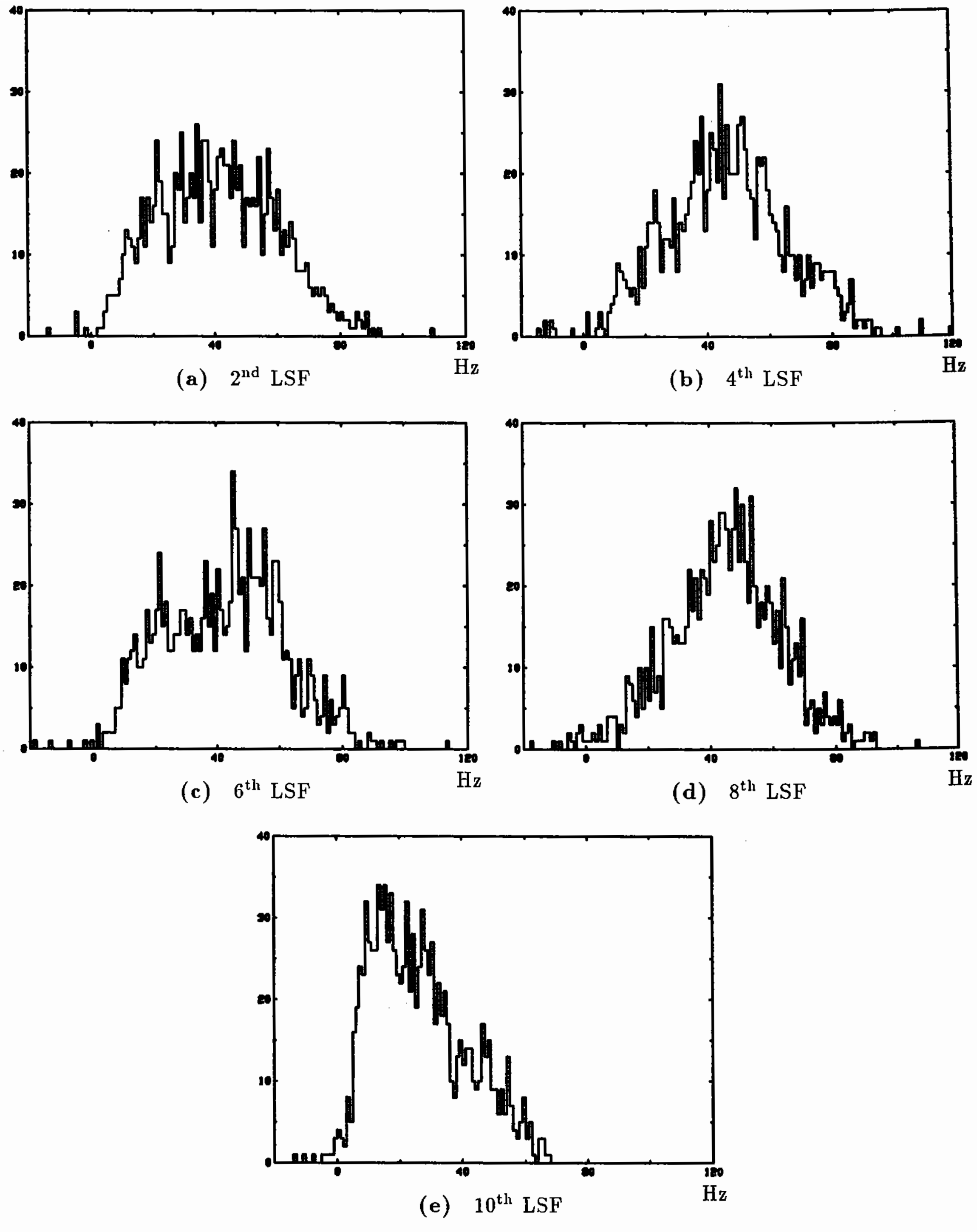
(b) 3rd LSF

(c) 5th LSF

(d) 7th LSF

(e) 9th LSF

Fig. 6 Histograms of LSF time differences.



(a) 2nd LSF

(b) 4th LSF

(c) 6th LSF

(d) 8th LSF

(e) 10th LSF

Fig. 7 Histograms of LSF relative positions. Note the occurrences of values less than zero.

tend to have smaller amplitude than the input speech. It is believed that this is due to the relatively coarser quantization at high frequencies. However, only in one out of the four cases is the reproduced signal found to be slightly buzzy.

In order to compare this quantization scheme with the standard LPC quantization, a set of experiments is performed by replacing the quantizer with a 21 bit log-area ratio quantizer designed for 10th order predictors with a bit assignment $(3,3,3,3,2,2,2,2,1,0)$. Informal subjective listening tests conclude that the subjective quality of the reproduced signals using the line spectral frequency quantizer is much better. In all four cases, the LSF quantizer gives more natural synthesized speech. Increasing the reflection coefficient bit assignment to a total of 41 bits, the quality of the log-area ratio quantizer improves but retains a considerable amount of musical noise. A second 21-bit reflection coefficient quantizer with same distribution of bits, but using non-uniform quantization based on the reflection coefficient histograms was also used. The quantizer divides the histograms into equal area partitions. The reproduced speech signals generated with this quantizer design does not have the musical noise. However, the subjective quality is still poor. The signals are dull rather than clear as in the case with log-area ratio quantization. In some sections, the signal is severely distorted. For example, "shade" in OAKF8 comes out as "fade".

Figures 8 to 15 show the LPC spectra of the quantized and unquantized coefficients. It is clear from the figures that LSF quantizer more closely approximates the true LPC spectrum than does the log-area ratio quantizer. While the error due to the LSF quantization tends to be located in the high frequency part of the spectra, the distortion due to the other quantization is spread along the whole frequency axis. As a result of this, the perceptual effects of the quantization errors from the schemes are very different.

In the following sections, the LSF quantizer will be modified for use in a CELP coder. The changes will bring into effect backward adaptive DPCM systems for the odd-numbered frequencies and a redistribution of the quantizer output levels for the relative frequencies $D_j(n)$.

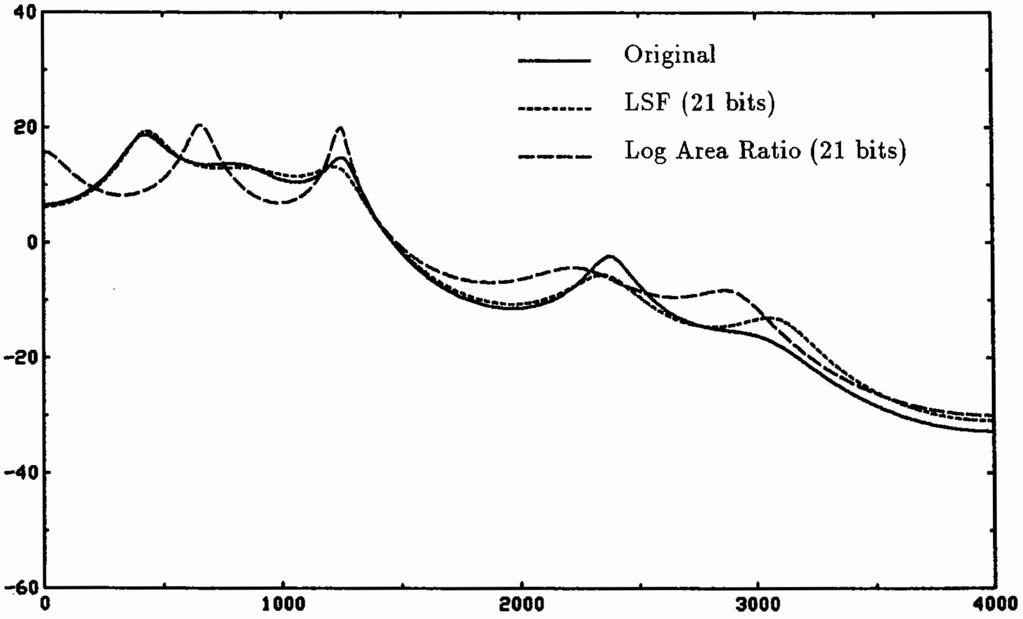


Fig. 8 LPC spectra for frame 12 of TOMF8 (160 sample frames).

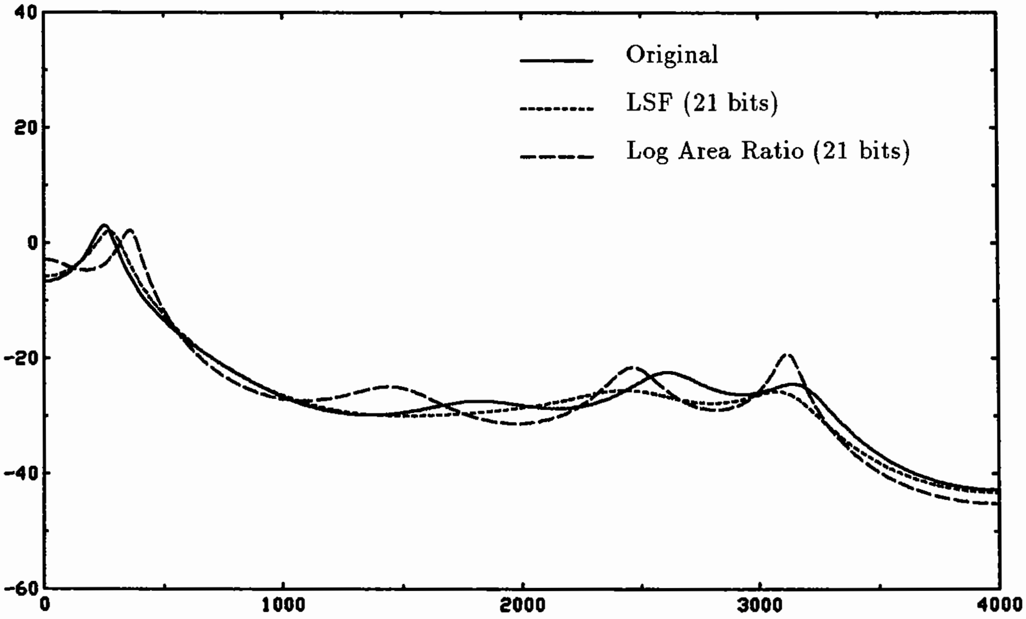


Fig. 9 LPC spectra for frame 66 of TOMF8 (160 sample frames).

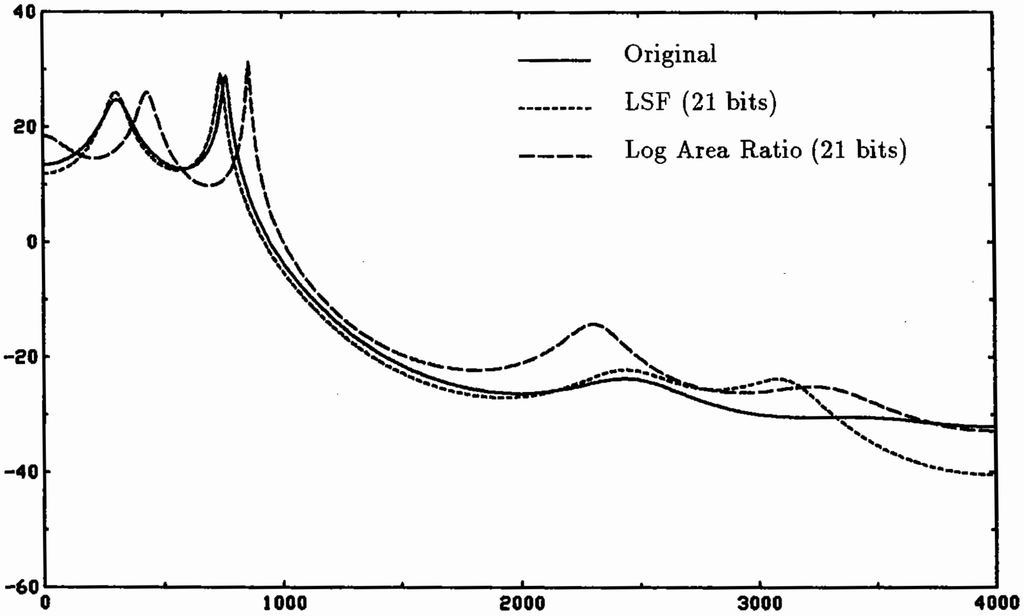


Fig. 10 LPC spectra for frame 8 of OAKF8 (160 sample frames).

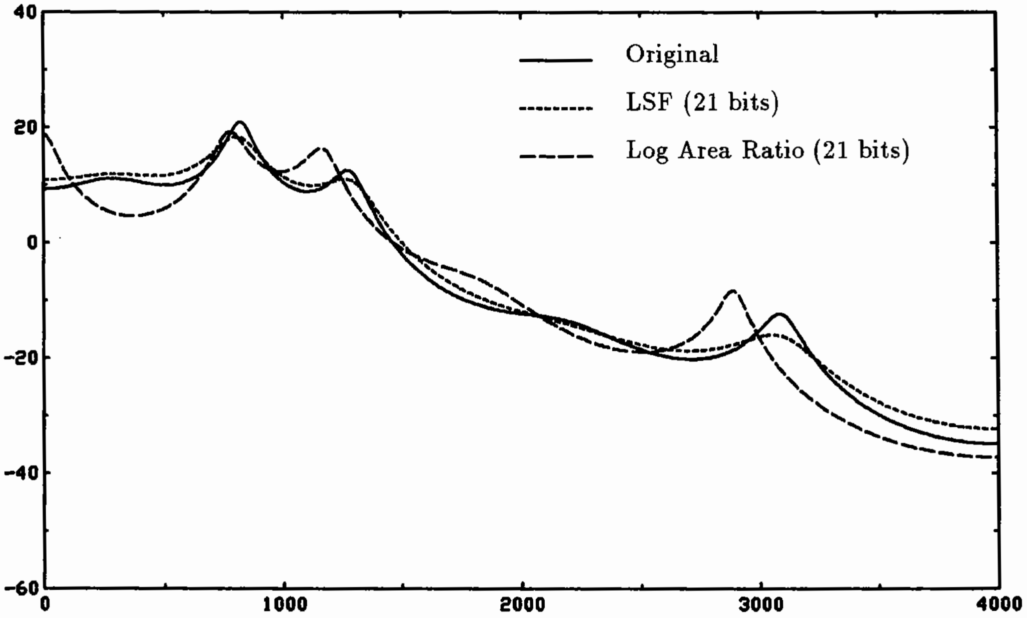


Fig. 11 LPC spectra for frame 54 of OAKF8 (160 sample frames).

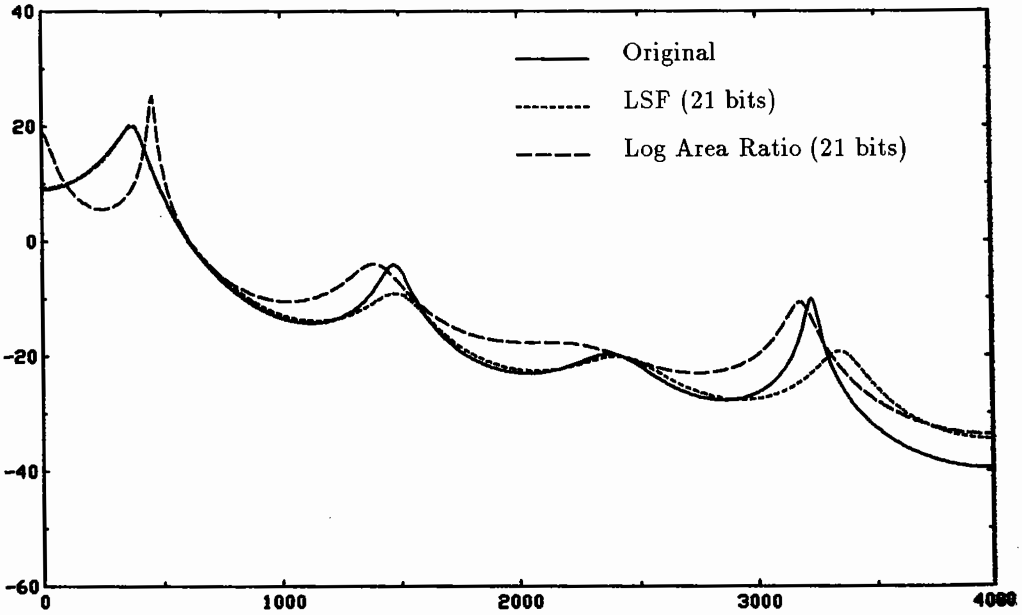


Fig. 12 LPC spectra for frame 41 of CANM8 (160 sample frames).

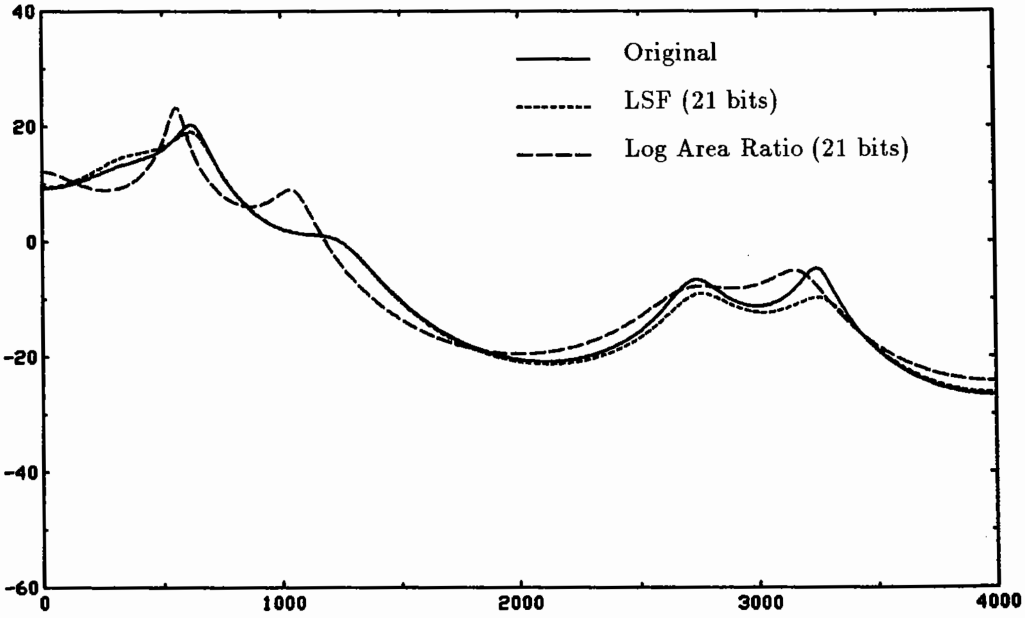


Fig. 13 LPC spectra for frame 70 of CANM8 (160 sample frames).

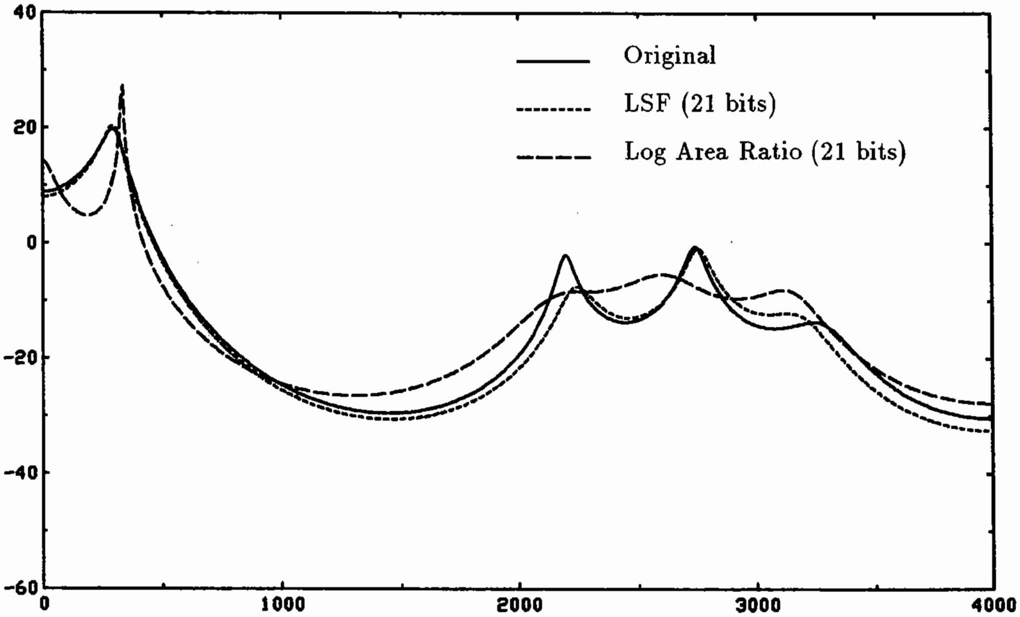


Fig. 14 LPC spectra for frame 34 of DEPM8 (160 sample frames).

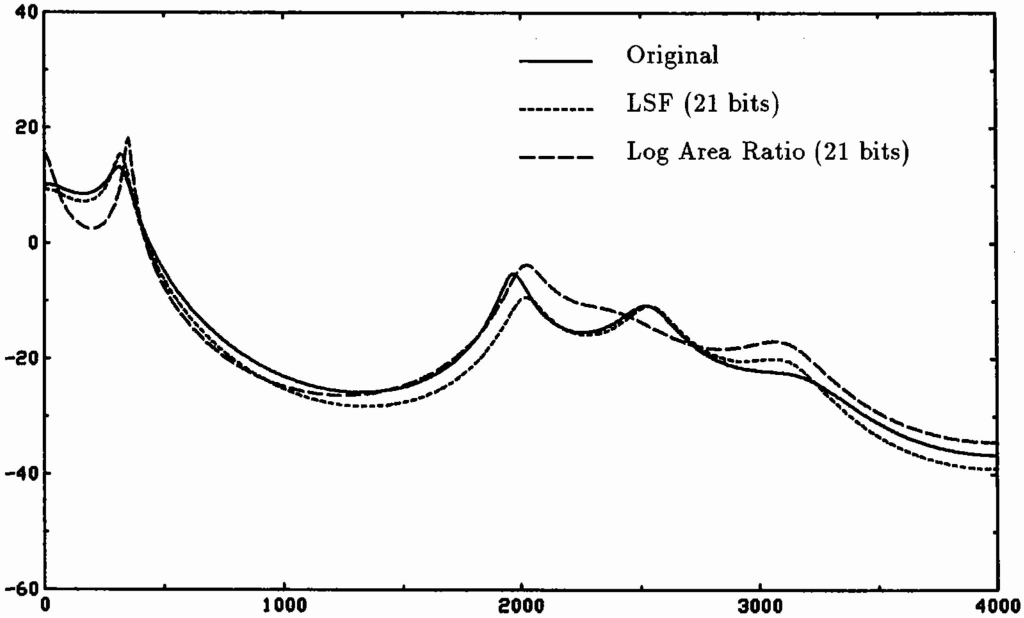


Fig. 15 LPC spectra for frame 52 of DEPM8 (160 sample frames).

4. Adaptive LSF quantizer

In the previous sections, the time invariant prototype of the line spectral frequency quantizer was designed and tested in a simplified coder with no distortion in the excitation signal. It is found that in the presence of the quantization noise in the excitation signal, the LSF quantization error becomes obvious in certain situations. Further improvement of the LSF quantizer is necessary and can be achieved with the use of adaptive DPCM (ADPCM) coder for the odd-numbered frequencies. Specifically, both the quantizers and predictors of the DPCM coders can be made time varying. In the next sections, the design of the ADPCM coders will be discussed.

4.1 Adaptive gain control

In order to increase the signal to quantization noise ratio and the dynamic range of the quantizer for non-stationary difference signals, adaptive quantizers are required. In particular, each adaptive quantizer is designed to be a combination of a time varying gain control and a time invariant non-uniform quantizer. This combination is equivalent to dynamically scaling the step sizes in the quantizer. The time varying gain factor is determined from previous quantizer output values.

The main function of the time varying gain control is to reduce possible overload and granular quantization errors. Overload occurs when the input value exceeds the range of the quantizer. Granular errors occur when the input value falls below the smallest step size. Let $G_i(n)$ be the gain factor for the input difference signal to Q_i where $i = 1, 3, \ldots, p-1$. The difference sample is normalized by dividing by $G_i(n)$. The adaptation mechanism is designed such that $G_i(n)$ is increased if previous quantizer output is overloaded and decreased if the previous quantizer output is small. Thus, the dynamic range of the effective difference samples is reduced or enlarged as appropriate. A simple but effective way to detect the presence of overload or granular error is based on the quantizer output levels. Overload error can be assumed to be present if the output level selected is either the minimum or maximum quantizer output level. Similarly granular error is present if the selected output is one of the innermost output levels. In the absence of signs of overload or granular error in previous quantization outputs, the input difference signal can be assumed to be in the correct range for the quantizer and the gain factor $G_i(n)$ remains unaltered. Experimental results show that decreases in $G_i(n)$ should take into account the presence of overload or granular error in not only the previous value and but also the one before that. The algorithm for updating to $G_i(n)$ was determined empirically and is shown in Table 5[†].

Given the bit assignments and the statistics of the scaled input difference signals for the 8 speech signals, a new set of quantizers for the even- and odd-numbered coefficients is designed accordingly.

[†] $G_0(n)$ for the 9th LSF, which is coded with only 1 bit, is always set to 1 because overload information cannot be derived from these quantizer outputs

$G_i(n)$	Conditions
$1.5\,G_i(n-1)$	Overload error at time n - 1
$G_i(n-1)/1.5$	Granular error at time n - 1
$0.5\,(G_i(n-1)+1)$	No overload or granular error at time $n-1$, but overload or granular error at time $n-2$
$G_i(n-1)$	No overload or granular error

Table 5 Adaptive quantizer gain factor.

4.2 Adaptive prediction and mean estimation

In the first version of the DPCM coder design for the odd-numbered LSF's, the predictor was implemented with a unit delay. One way to improve the coder further is to use a multiple coefficient time varying predictor.

It is noted that the line spectral frequency input to a DPCM coder has non-zero mean value. In the following we make use of an unbiased mean-estimator to help compensate for this non-zero mean. In fact, the mean estimation can be time varying and adapted from the DPCM decoded output samples. Let $\bar{L}_i(n)$ be the estimated mean value of the i^{th} odd-numbered LSF n and $\hat{l}_i(n-1)$ be the DPCM decoded output of the i^{th} input at time $n-1$. Then an unbiased mean estimate is

$$\bar{L}_i(n) = (1-\beta)\bar{L}_i(n-1) + \beta\hat{l}_i(n-1) \ , \tag{12}$$

where $0 \leq \beta \leq 1$. The mean-estimator transfer function can be expressed as

$$B(z) = \frac{\beta z^{-1}}{1-(1-\beta)z^{-1}} \ . \tag{13}$$

By choosing the value of β appropriately, the mean value can be updated at different speeds. With $\beta = 0$, the mean value is constant and with $\beta = 1$, the mean value is just the previous coder output sample. A block diagram of the system with the mean compensation is shown in Fig. 16.

We note that the effects of the one-tap predictor filter and the mean-estimator can be combined into a single predictor filter. The original one-tap predictor can be represented as

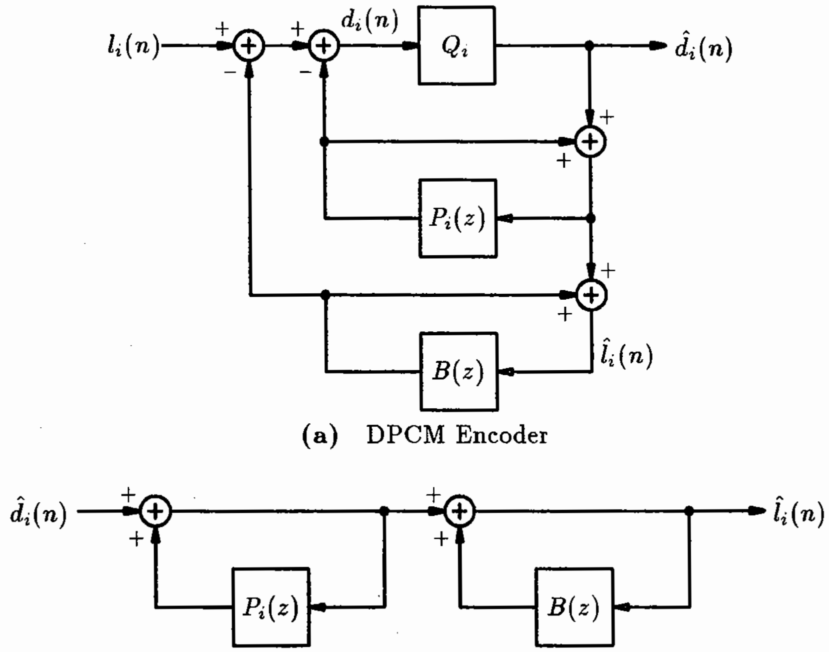
$$P_i(z) = az^{-1} \ , \tag{14}$$

where a is the predictor coefficient. The equivalent predictor formed by the combination of mean estimation and one-tap prediction can be shown to be

$$P_i'(z) = P_i(z) + B(z) - P_i(z)B(z) \ . \tag{15}$$

The corresponding synthesis filter used to reconstruct the output LSF is

$$\frac{1}{1-P'(z)} = \frac{1-(1-\beta)z^{-1}}{(1-z^{-1})(1-az^{-1})} \ . \tag{16}$$



(a) DPCM Encoder

(b) DPCM Decoder

Fig. 16 Block diagram of the DPCM system with mean compensation.

This form shows that the presence of the mean-estimator gives the system a second order denominator.

In the first test, each 1-tap time invariant predictor tap weight is assigned a value equal to the autocorrelation of the mean compensated input signal. The average autocorrelation of the input compiled with 8 speech files varies with the value of β. When $\beta \to 0$, the correlation value tends to be high and vice-versa. This is as expected because of the whitening capacity of the mean compensation. Table 6 shows this relationship.

β	Normalized autocorrelations				
	1^{st}	3^{rd}	5^{th}	7^{th}	9^{th}
0	0.967	0.974	0.994	0.996	0.997
0.25	0.543	0.530	0.685	0.647	0.678
0.50	0.364	0.333	0.510	0.486	0.645
0.75	0.196	0.137	0.318	0.316	0.605

Table 6 Normalized autocorrelation coefficient $r(1)/r(0)$ of the LSF's with mean compensation.)

Experimental results show that, in general, the system performance depends strongly on the tap weights of the predictor, with the best performance corresponding to values of a near unity.

Yet, the best value of β is found to be 0.5. Although the weights chosen do not correspond to the normalized correlations of the zero mean inputs with $\beta = 0.5$, the combination of $\beta = 0.5$ with predictor coefficients near unity gives the best performance. The predictor coefficients used were $(0.967, 0.974, 0.994, 0.996, 0.997)$.[†].

Next, a time varying predictor is considered for the DPCM coders. In the last section, the autocorrelation values were computed based on long term statistics of the line spectral frequencies. In an adaptive mode, the autocorrelation value is computed based on backward adaptation. The autocorrelations are computed over a finite window. In so doing, the decoder can be adapted in exactly the same way without any additional transmission of information. The speed of adaptation is controlled by the memory of the process (number of terms used to compute the autocorrelation). With a long memory sequence, the adaptation is slow. Note that the correlation values computed can be negative in some cases. The subjective quality of the reproduced signals is not as good as that of the signals obtained with time invariant predictor selected as given earlier. With $\beta = 0.5$, the predictor tends to converge to values which are smaller than those for best performance (see Table 6).

[†] A 2-tap time invariant predictor was also considered. No significant improvement in subjective quality was obtained.

5. Quantization of LPC parameters in CELP

In the previous sections, the design of a line spectral frequency quantizer was discussed. It is noted that the previous discussion was based on an ideal environment in which the LSF quantization error is the only kind of error in the system. Although the quantizer developed performs well in the simple configuration of Fig. 1, the final effect of the LSF quantization noise in the presence of other errors in a speech coder has to be investigated. The LSF quantizer and the speech coding system as a whole has to be efficiently integrated so as to minimize the perceptual effect of the resulting error. The attention of the following sections will be on the integration of the LSF quantizer with a CELP coder.

5.1 System Configurations

In a Code Excited Linear Predictive coder, the input speech signal first has its short time spectral characteristics removed by a short time prediction error filter (also referred to as the LPC or formant filter). The residual signal is processed by a second prediction error filter to remove the predictable information due to pitch periodicities. The final residual signal is vector quantized by selecting a waveform from a codebook of waveforms [7]. In practice, the pitch filtering and the vector quantization can either be carried out independently or, as in the approach adopted here, be jointly optimized. Given the formant filter parameters (LPC filter coefficients), the encoder searches for the jointly optimal values of pitch synthesis filter (pitch lag and filter coefficients), residual codeword, and the codeword scaling factor which minimize the energy of a frequency weighted error signal. The waveform selection procedure can be termed analysis by synthesis.

The configuration for the calculation of the frequency weighted error is shown in Fig. 17. In this figure $P(z)$ is the pitch predictor and $F(z)$ is the formant predictor. Two equivalent configurations are shown. In both cases the weighting filter depends on the formant structure of the segment of speech being analyzed. The factor γ in the weighting filter is used to control the coding noise spectrum. For the experiments, γ is set to $1/0.75$. Each codeword in the CELP dictionary is passed through the system and the resulting frequency weighted error calculated. The codeword which most closely resembles the input speech in the frequency weighted sense is chosen. The index of this codeword is sent to the decoder. The decoder then synthesizes a signal using the codeword corresponding to the index.

A set of computer simulations was carried out to evaluate the tradeoff between the use of unquantized and quantized LPC coefficients in the residual codeword selection process. As shown in Fig. 18, there are four possible configurations corresponding to the block diagram in Fig. 17(b). The filter $F(z)$ uses unquantized coefficients, while $\hat{F}(z)$ uses quantized coefficients. In the first two configurations, unquantized LPC coefficients are used to realize the formant prediction filter for

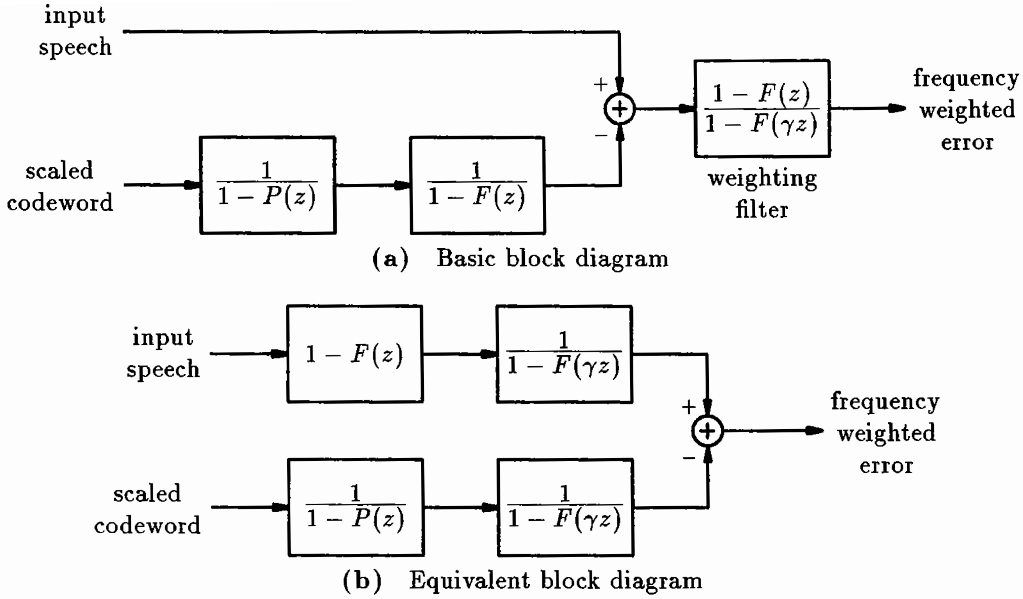
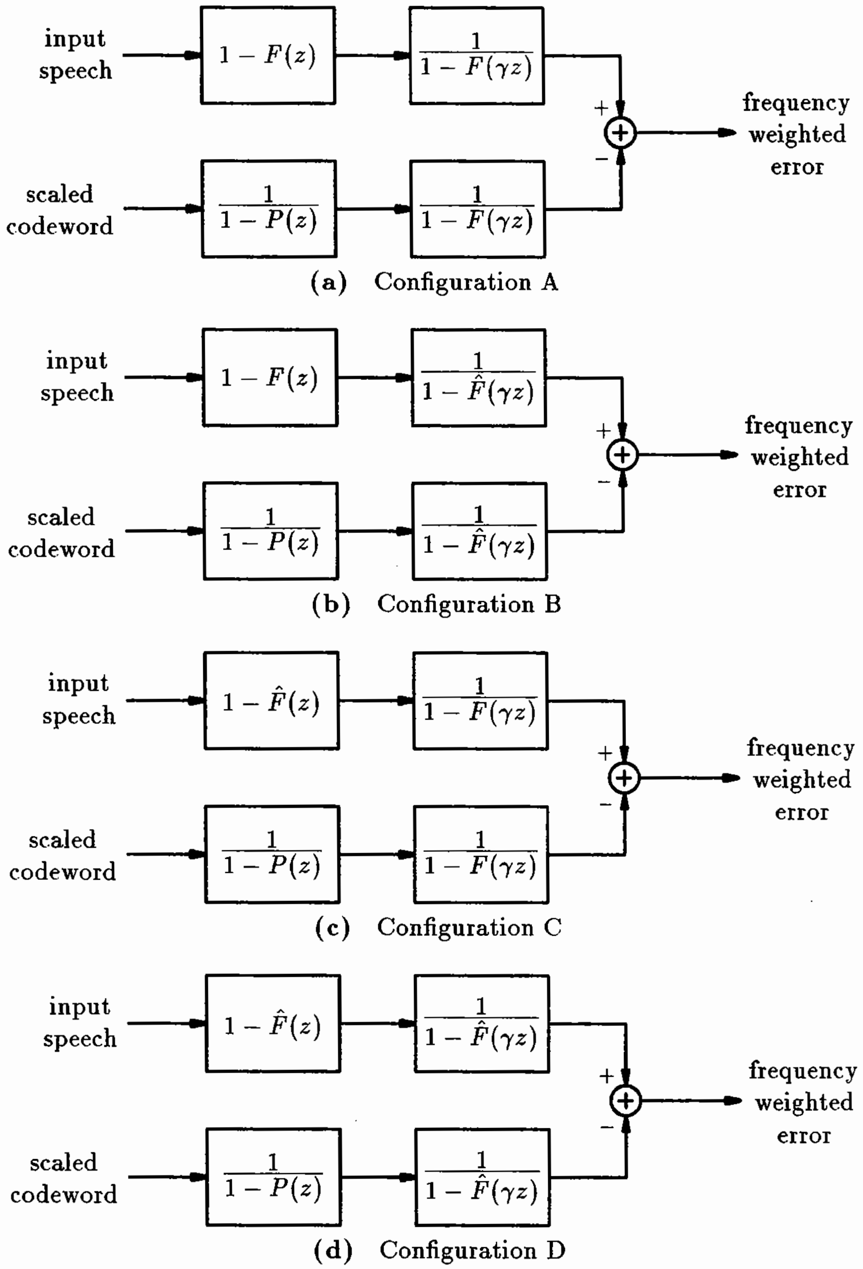


Fig. 17 Block diagram showing the calculation of the frequency weighted error

removing the predictable information, whereas the last two configurations make use of the quantized LPC coefficients for the same purpose. The number of cases is kept to four by requiring that the denominator of the frequency weighting filter (see Fig. 17(a)) be the same in both branches of the block diagrams. Note that these configurations are used only for choosing the appropriate codeword. The actual synthesis of speech in the decoder will of necessity use quantized synthesis filters coefficients.

The different configurations will behave slightly differently. One can argue that configuration A ignores the effect of quantization of the formant filter completely. The actual decoder which uses quantized coefficients for the synthesis filter may suffer as a result. On the other hand in configuration B, one can see that the speech signal used as the reference signal has the formant structure removed and a quantized formant structure reinserted. The codeword path also uses the same quantized formant structure. This means that the chosen codeword does not try to compensate for the quantization of the formant synthesis filter. This has the effect of decoupling the error due to filter coefficient quantization and the error due to coding of the excitation signal. In configuration D, the codeword does try to compensate for the quantization of the formant synthesis filter. Configuration C is included for completeness.

The first set of experiments is carried out using the 8 speech signals as listed earlier. The quantized LPC coefficients and unquantized values of the pitch lag, pitch predictor tap coefficient, and codeword gain factors are used in the decoder. The configuration used in the experiments has a single pitch coefficient, and the codeword selection process chooses one of 32 waveforms every 40



(a) Configuration A

(b) Configuration B

(c) Configuration C

(d) Configuration D

Fig. 18 Block diagrams of four different configurations, differing as to which filters have quantized coefficients.

samples.

Experimental results based on informal subjective listening tests show that the configuration can be classified into two groups: configurations A and B, and configurations C and D. In general, A and B generate output signals with lower energy content than C and D do. As a result, the outputs from A and B sound less natural and clear. All four configurations occasionally generate undesired or distorted frequency components in the forms of noise and clicks, although the problem is more annoying and severe with C and D. Between A and B, there is no consistent difference in subjective quality in the reproduced speech, although A generally outperforms B. Between C and D, the clicks from configuration D in general have less perceptual impact than those from configuration C.

In view of the experimental results, it is concluded that a better quality synthesis results if quantized LPC coefficients are used for waveform selection (configuration D). The price for the improvement is the presence of annoying and obvious clicks. Also, the experiments indicate that less distortion is possible when the formant prediction filter and the γ-weighted formant synthesis filter are realized with the same set of coefficients—quantized or unquantized.

There are several effects due to the use of quantized LPC coefficients (as in configuration D) on the performance of a CELP system. First, the formant residual signal is not as white as would have been if the prediction filter were realized with unquantized LPC coefficients. However, the encoder more accurately models the processing that will be carried out in the decoder. The encoder chooses the best excitation waveform by taking the presence of the quantized filter into account. Because there is an actual improvement in terms of naturalness in the reproduced speech. The advantage of being able to partially compensate for the filter quantization overwhelms the negative effect of using slightly more colored formant residual.

5.2 Energy dependent gain adaptation

Configuration D in the previous section gives the most natural reproduced output signals, but with the price of some annoying clicks in the signal. In this section, we consider means to mitigate these disturbances.

Subjective tests indicate that the distortion observed appears at the beginning of a words or sequence of words. The coding process is sufficiently smooth before and after the clicks. Parameter traces of the odd-numbered line spectral frequencies before and after DPCM coding confirm that there is overload of the DPCM quantizer in the positions where the clicks are detected. To eliminate the clicks, the step sizes used in the DPCM coder are adapted to the energy of the signal. With a view to constraining the bit rate required, the adaptation is of backward nature. The energy content of the previously synthesized frame of speech is used rather than that of the one being processed. No additional information is transmitted to serve this purpose.

The mechanism of the adaptation is devised so that when the energy of the previous reproduced output frame is less than a preset level, the gain factor inside a DPCM coder is assigned a large value to reduce the effect of a possible sudden change in the LSF's at the onset of speech. Inside a speech segment, the gain factor is updated as usual (see Section 4.1). In the presence of successive analysis frames with silence, the assignment of a large gain factor will cause granular noise. However, this noise does not cause significant harm since the level of the signal is low in these regions.

An energy threshold is used to control the behaviour of the gain factor. The gain factors for all DPCM coders are assigned the same large value when the energy content of a previous frame is below the reference level. The value used is a compromise between too large a value which will result in a long time to decay to normal values after the onset of speech, and too small a value which will not track the change in the LSF's at the onset of speech. The energy threshold level also has to be chosen to properly distinguish silence and speech. The final values of the energy reference level and the gain factor are obtained empirically so as to minimize the average subjective distortion for a variety of speech signals.

In effect this energy dependent gain adaptation overrides the normal gain adaptation during silence. The gain control is taken over again by the normal gain adaptation in speech segments whose energy content is sufficiently high. Experimental results indicate that it is advantageous to let the gain factor be adapted slowly back to its nominal value, immediately after the energy dependent gain adaptation is turned off.

Computer simulation results show that with a proper setting of the gain factor and energy reference level, the clicks observed in configuration D of the previous section disappear while the subjective quality of the synthesized signals remains high. It is concluded that the energy dependent gain adaptation mechanism is worthwhile.

6. Interpolation of the line spectral frequencies

In all the previous experiments, the LPC coefficients are calculated every 160 samples (20 ms). It is found that better subjective quality can be obtained if the spectral information is updated every 80 samples (10 ms). Implemented naively, this implies a doubling of the bit rate for filter coefficient transmission. In this section, an interpolation scheme for LSF's is proposed to allow the LSF's to be updated more often without the full penalty of a doubling of the bit rate. This interpolation will take advantage of the fact that linear interpolation of properly ordered LSF's gives another set of properly ordered LSF's.

6.1 Interpolation with no additional information

In a simple scheme, LSF's are generated for alternate frames, say the odd-numbered frames. The LSF's for even-numbered frames are found by interpolating between the LSF's of the adjacent odd-numbered frames. Interpolation of the filter coefficients implies that an extra delay has to be accommodated. The interpolation cannot be carried out until after both sets of LSF's for the odd-numbered frames are available.

There are two different strategies as to how the LSF's for the alternate frames are generated.

Method A

In method A, a set of LSF's is computed for every 80-sample frame. Line spectral frequencies for the odd-numbered frames are quantized as usual to get the quantized LSF's. The LSF's for the even-numbered 80-sample frames are the average of the quantized LSF's from the neighboring odd-numbered frames.

Method B

In method B, sets of LSF's are again calculated for odd-numbered frames. However, the window used to calculate the LSF's overlaps the even-numbered frames. For instance, for 80 samples frames, a window of 160 samples is used to compute the LSF's. In this way changes in the LSF's are smoothed out.

6.1.1 Comparison of method A and B

In method A, the resulting interpolated LSF's depend entirely on the spectral characteristics of the neighboring two odd-numbered frames. These interpolated LSF's contain no spectral information pertaining to the samples in the even-numbered frame. In method B, resulting interpolated LSF's for the even-numbered frames contain some spectral information pertaining to this frame. However, the spectral information is calculated with a larger window and so is not as representative for the odd-numbered frames themselves.

Computer simulations show that the resulting perceptual effect is speech/speaker dependent. In general, little improvement in subjective quality of the output of the coding system is obtained from either interpolation scheme. For instance, there may be very slight improvement observed with one utterance, while no change is found with the others. It is concluded that the computations to implement interpolation as described are not warranted.

6.2 Interpolation using additional information

In the previous section, linear interpolation of line spectral frequencies using an equal weighting of the neighbouring frames was used. Linear interpolation of the LSF's can be carried out using a variable interpolation factor as well. With zero bits for the interpolation process, method A has to discard the unquantized line spectral frequencies for every second frame. However, the spectral information in these frames in method A can be used during the interpolation process if additional bits are made available for transmission. Consider a more general interpolation formulation with an interpolation factor $\alpha_i(n)$,

$$\hat{l}_i(n) = \alpha_i(n)\,\hat{l}_i(n-1) + \big(1 - \alpha_i(n)\big)\,\hat{l}_i(n+1) \qquad i = 1, 2, \ldots, p\,. \tag{17}$$

Methods A and B both use $\alpha_i(n) = 1/2$. At the price of some additional information bits, it is possible to select a more appropriate interpolation factor from a quantized set of values. Selection of the best interpolation factor is an error minimization problem.

The problem is to choose the interpolation factor so as to minimize some error measure. If $q_i(n)$ is the quantization error for a particular LSF, the overall error can be expressed as

$$E(n) = \sum_{i=1}^{p} [w_i(n)\,q_i(n)]^2\,, \tag{18}$$

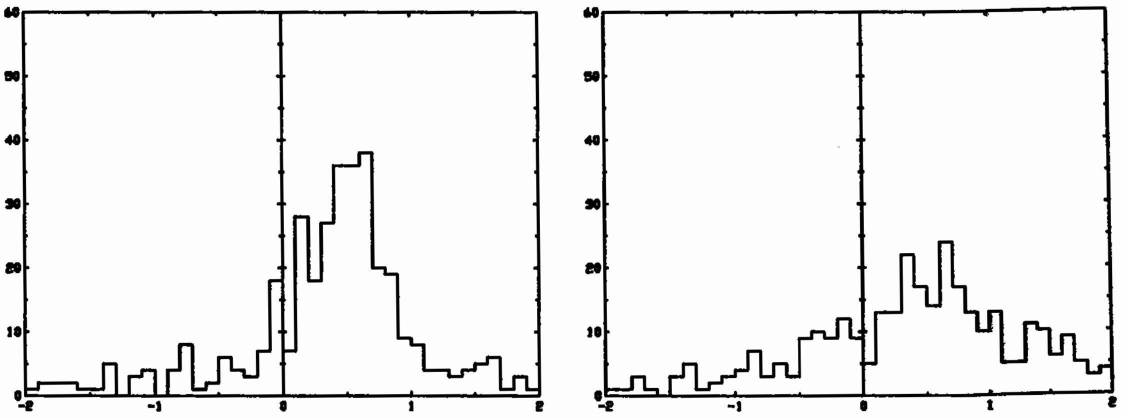
where $w_i(n)$ is a weighting factor of the i^{th} interpolation error. The optimal interpolation factor(s) are chosen to minimize the total weighted error energy $E(n)$.

If each LSF is to have its own interpolation factor, then in the absence of quantization of the interpolation factor, interpolation can produce an arbitrary set of LSF's and the interpolation error can be set to zero. The optimal interpolation factors are

$$\alpha_i^*(n) = \frac{\hat{l}_i(n+1) - l_i(n)}{\hat{l}_i(n+1) - \hat{l}_i(n-1)}\,. \tag{19}$$

Note that for this case, the individual interpolation factors do not depend on the weighting. With quantization of the interpolation factors, interpolation with separate factors for each LSF is essentially a form of DPCM coding with delayed decisions and is more properly considered a coding scheme than an interpolation scheme.

A set of histograms of the interpolation factors for each LSF coefficient was calculated. The histograms are compiled for the eight speech signals for non-silent segments. Figure 19 shows the optimal interpolation factors for the 2^{nd} LSF and the 8^{th} LSF. Both histograms show peaks near 1/2, but the histogram for $\alpha_2^*(n)$ has a smaller variance. Both histograms show that the interpolating factor goes outside the range $[0,1]$.[†]



(a) Histogram for $\alpha_2^*(n)$ (a) Histogram for $\alpha_8^*(n)$

Fig. 19 Histograms of $\alpha_i^*(n)$

A more realistic approach that does not require excessive overhead is to use a single quantized interpolation factor for all LSF coefficients. If all LSF's use the same interpolation factor, then the optimal interpolation factor is given by

$$\alpha^*(n) = \frac{\sum\limits_{i=1}^{p} w_i^2(n)\left[\hat{l}_i(n+1) - \hat{l}_i(n-1)\right]\left[\hat{l}_i(n+1) - l_i(n)\right]}{\sum\limits_{i=1}^{p} w_i^2(n)\left[\hat{l}_i(n+1) - \hat{l}_i(n-1)\right]^2} \,. \tag{20}$$

The error weighting factor $w_i(n)$ appears in this expression.

According to Kang and Fransen, the perceptual significance of error in a LSF depends on the distance of that LSF from others and on its location in the frequency domain. Specifically, it is found that errors in closely spaced LSF's are perceptually more significant, and the perceptual effect of distortion diminishes with frequency. Accordingly, the error weighting factor is defined as the product of two terms,

$$w_i(n) = \left(1 + D - \frac{d_i(n)}{d_{\max}(n)}\right)\left(1 - l_i(n)/F\right) \qquad i = 1, 2, \ldots, p \,.$$

Effectively, the first term modifies the weight based on the relative closeness of an LSF to others. The value $d_i(n)$ is the minimum absolute distance of the i^{th} LSF from its neighbours and the value

[†] An interpolation factor within the range $[0,1]$ is needed to guarantee an absence of LSF crossovers.

$d_{\max}(n)$ is the maximum absolute distance of all of the adjacent pairs of LSF's. The parameter D is chosen to be 0.01. Because $0 < d_i(n)/d_{\max}(n) \leq 1$, the first term is positive (and nonzero). The second term modifies the weight based on the frequency of the LSF. The parameter F in the second term is set to 1.6. Note that the weighting factors depend only on the (unquantized) LSF's in the frame to be interpolated.

We assign 2 bits for the transmission of the interpolation factor. The quantized interpolation factors are chosen using a histogram of the optimal interpolation factor as shown in Fig. 20. Three sets of interpolation factor quantizers were evaluated.

$$\alpha : (0.05, 0.35, 0.55, 0.80) \qquad \text{I}$$

$$\alpha : (0.00, 0.35, 0.55, 1.00) \qquad \text{II}$$

$$\alpha : (0.21, 0.37, 0.52, 0.68) \qquad \text{III}$$

The first set, I, is chosen noting the symmetry of the values in Fig. 20 about 0.4. II is chosen as in I but with the minimum and maximum values set 0 and 1. Set III is obtained by dividing the histogram in [0, 1] into equal area regions.

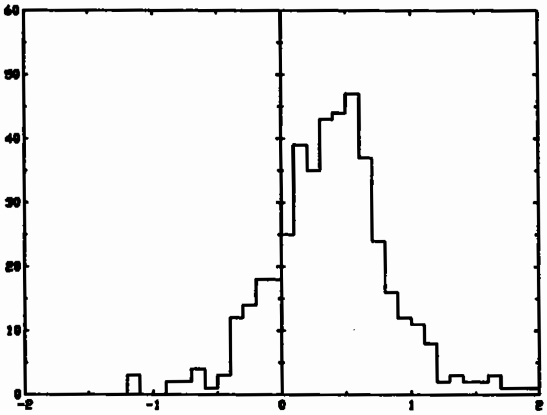


Fig. 20 Histogram of $\alpha^*(n)$

A coder using the interpolation scheme was simulated. The simulation is performed with four different speech input signals (TOMF8, CANM8, OAKF8, and PIPM8). In each frame to be interpolated, the energy of weighted error $E(n)$ is computed for each interpolation factor value. The best of these is used to interpolate the LSF's from the neighbouring frames. The quality of the synthesized signals was evaluated in informal subjective listening tests and compared with the synthesized signals with LSF quantization for every 160 samples and 80 samples.

It is found that using interpolated quantized line spectral frequencies for every second frame using an additional two bits does bring an improvement. The subjective quality of the resulting signals is consistently better than that reproduced with the synthesis filter updated based on quantized LSF's with a 160 sample frame size. In general, the output based on interpolation sound smoother and clearer while those based on the longer frames are buzzy and have ripples. In some cases, the results with interpolation are even slightly better than those with the LSF's coded every 80 samples. This means that the error due to interpolation can be less in some cases than the error due to quantization.

The three different interpolation factor quantizers give results which are signal dependent. One combination may be the best choice for one signal while it may be rated second for the others. However, on the average set I gives the best reproduced signals.

7. Summary and conclusions

Two types of LSF quantizers have been considered. In the first, the center frequency and the offset frequency of pairs of LSF's are coded. However, this method is plagued by LSF's changing order after quantization. This crossover problem manifests itself as unstable synthesis filters. The crossover problem can be artificially eliminated, but the quality of the coding is still poor (21 bits/frame).

The second type of LSF quantizer uses a combination of interframe and intraframe coding. Odd-numbered LSF's are coded differentially in time, while even-numbered LSF's are coded relative to the odd-numbered LSF's. This strategy outperforms the center/offset frequency quantizer. In addition, this form of quantizer has essentially no problem with LSF crossovers. An adaptive gain factor is introduced into the differential coder for the odd-numbered LSF's for further improvement in the quality. A modified differential coding configuration which has an additional mean value tracker was also studied.

In a CELP coder context, it was found that the codeword selection process should use the quantized formant filter parameters. In such a configuration, the effect of a noisy excitation signal introduces new degradations into the coding process. It is found that the lack of continuity of the LSF tracks at the onset of speech can lead to extraneous clicks. This is partially cured by using a gain adaptation which uses cues from the energy of the signal. Essentially the gain is increased in low level segments in anticipation of the onset of speech. This allows the LSF quantizer to quickly home in on correct values in the speech segment.

Finally several different interpolation schemes are considered. A new scheme, which transmits some additional information (2 bits) can be used to update the filter parameters twice as fast with a minimal increase in transmission overhead.

LSF's seem to offer many benefits for the representation of LPC parameters. This study has shown that LSF coding can be used to control the spectral distortion due to quantization. In a CELP coder, LSF's offer good rendition of the LPC synthesis filter at bit rates around 1100 b/s. We believe that further improvements can be made as the spectral features that are important for good coding are better understood. In addition, more sophisticated heuristics can probably improve the performance of LSF coders, especially in the critical regions corresponding to the onset of speech.

References

1. J. Makhoul, "Linear prediction: a tutorial review", *Proc. IEEE*, vol. 63, pp. 561–580, April 1975.
2. F.K. Soong and B-H. Juang, "Line spectrum pair (LSP) and speech data compression", *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, San Diego, California, pp. 1.10.1–1.10.4, March 1984.
3. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Springer-Verlag, 1976.
4. G. S. Kang and L. J. Fransen, "Low bit rate speech encoders based on line spectrum frequencies (LSFs)", *Naval Research Laboratory Report 8857*, November 1984.
5. P. Kabal and R.P. Ramachandran, "The computation of line spectral frequencies using Chebyshev polynomials", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1419–1426, Dec. 1986.
6. J.R. Crosmer and T.P. Barnwell, III, "A low bit rate segment vocoder base on line spectrum pairs", *Proc. Int. Conf. on Acoustics, Speech and Signal Processing*, Tampa, Florida, pp. 7.2.1–7.2.4, April 1985.
7. M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High-quality speech at very low bit rates", *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tampa, Florida, pp. 25.1.1–25.1.4, March 1985.