

Sine-Wave Amplitude Coding Using A Mixed LSF/PARCOR Representation

by

Robert B. Dunn

Submitted to the Department of Electrical Engineering and
Computer Science

in partial fulfillment of the requirements for the degree of

Master of Science in Bioelectrical Engineering

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

May 1995

© Robert B. Dunn, MCMXCV. All rights reserved.

The author hereby grants to MIT permission to reproduce and
distribute publicly paper and electronic copies of this thesis
document in whole or in part, and to grant others the right to do so.

Author: _____

Department of Electrical Engineering and Computer Science

May 12, 1995

Certified by.. _____

Dr. Robert J. McAulay

Senior Staff, MIT Lincoln Laboratory

Thesis Supervisor

Accepted by _____

Prof. Frederic R. Morgenthaler

Chairman, Departmental Committee on Graduate Students

MASSACHUSETTS INSTITUTE
OF TECHNOLOGY

JUL 17 1995

ARCHIVES

LIBRARIES

Sine-Wave Amplitude Coding Using A Mixed LSF/PARCOR Representation

by

Robert B. Dunn

Submitted to the Department of Electrical Engineering and Computer Science
on May 12, 1995, in partial fulfillment of the
requirements for the degree of
Master of Science in Bioelectrical Engineering

Abstract

An all-pole model of the speech spectral envelope is used to code the sine-wave amplitudes in the Sinusoidal Transform Coder [1, 2, 3]. While line spectral frequencies (LSFs) are currently used to represent this all-pole model, it is shown that a mixture of line spectral frequencies and partial correlation (PARCOR) coefficients [4] can be used to reduce complexity without a loss in quantization efficiency. The new representation is applied in the Sinusoidal Transform Coder to reduce the time required to compute all-pole model parameters by a factor of four. Objective and subjective measures of speech quality demonstrate that this does not result in reduced quality. In addition, the use of split vector quantization is shown to substantially reduce the number of bits needed to code the all-pole model.

Thesis Supervisor: Dr. Robert J. McAulay
Title: Senior Staff, MIT Lincoln Laboratory

Contents

1	Introduction	6
2	Sinusoidal Transform Coding	8
3	All-Pole Model Parameter Computation	14
3.1	All-Pole Model Representations	14
3.2	Parameter Transformations	16
3.3	Arithmetic Accuracy Requirements	19
3.4	Computational Complexity	24
3.5	Discussion	26
4	Parameter Coding	28
4.1	Scalar Quantization	29
4.2	Vector Quantization	34
4.3	Discussion	38
5	Summary	40
A	Closed Form Solution for 8th Order LSFs	41
B	Perceptually Weighted Spectral Distortion	45
C	Significance Testing	47

List of Figures

2-1	Example of interpolated spectral envelope.	10
2-2	Comparison of all-pole models.	12
2-3	Warped and unwarped fit to spline envelope.	13
3-1	System for computing predictor coefficients from mixed LSF/PARCOR representation.	16
3-2	System for computing mixed LSF/PARCOR representation using Durbin's recursion.	18
3-3	System for computing mixed LSF/PARCOR representation using the autocorrelation-lattice method.	19
3-4	System for comparing reference spectrum with computed spectrum.	21
3-5	Comparison of reference and distorted spectral envelopes.	23
4-1	Relative spectral sensitivity of PARCOR coefficients.	30
4-2	Probability density estimates for PARCOR coefficients k_1 through k_8	31
4-3	Probability density estimates for PARCOR coefficients k_9 through k_{14}	32
4-4	Correlation of parameters.	36
4-5	Correlation of the $6th$ vector element.	37
B-1	Bark scale critical band filters.	46
B-2	Perceptually based spectral weighting function.	46
C-1	Likelihood of listening test outcome.	48

List of Tables

3.1	Error computing LSFs using single precision arithmetic.	22
3.2	Error computing predictor coefficients using single precision arithmetic.	23
3.3	Error computing predictor coefficients using double length arithmetic.	24
3.4	Time needed to compute all-pole model parameters.	25
3.5	Time needed to compute predictor coefficients.	26
3.6	Time needed to compute all-pole model parameters and predictor coefficients.	27
4.1	Bit allocations for scalar quantization.	33
4.2	Distortion caused by scalar quantization.	33
4.3	Listening test comparing scalar quantization systems.	34
4.4	Bit allocations for vector quantization.	37
4.5	Distortion caused by scalar and vector quantization.	38
4.6	Listening test comparing scalar and vector quantization.	38
4.7	Distortion caused by scalar quantization of PARCOR representation.	39
B.1	Frequency bands for spectral distance measure.	45

1 Introduction

Speech coding at low data rates (4800 to 2400 bits per second) has been important primarily for secure voice transmission. Significant improvements in quality have made vocoders viable for other applications where there is an increasing demand for bandwidth such as cellular telephone, personal communication systems, and satellite communications. These applications require low power implementations using a single digital signal processor unit so the computational complexity of vocoder algorithms must be kept low. In this thesis it is demonstrated how the computational complexity of the Sinusoidal Transform Coder (STC) can be reduced by using a mixture of line spectral frequencies (LSFs) and partial correlation (PARCOR) coefficients to represent the all-pole model of the sine-wave amplitudes.

The mixed LSF/PARCOR representation was introduced in [4] to increase the quantization efficiency of 10th order linear predictive coding. In STC improvements in quality have been obtained by using model orders much higher than the basic 10th order system, but this increases the complexity of the implementation due to the difficulty in computing the LSF parameters. The mixed LSF/PARCOR representation is used here to code the higher order all-pole model of the sine-wave amplitudes in STC and decrease the computational complexity of the vocoder. It is demonstrated that the mixed LSF/PARCOR representation maintains the coding efficiency of the current LSF system.

The Sinusoidal Transform Coder is introduced in section 2 and the all-pole model for the sine-wave amplitudes is described. In section 3 the mixed LSF/PARCOR representation is presented. The arithmetic accuracy needed for computations involving LSFs is evaluated and the time required for these computations is determined. Sec-

tion 4 shows that when the parameters are quantized, using the mixed representation results in no loss in quality as compared to a purely LSF representation for coding the higher-order all-pole model of the sine-wave amplitudes at 2400 bits per second. This is demonstrated with both objective and subjective measures of speech quality. In addition, it is shown that split vector quantization can be used to reduce the number of bits needed to code the sine-wave amplitudes from 50 to 45. The extra bits can then be used to code other parameters with more fidelity, thereby increasing the overall quality of the system.

2 Sinusoidal Transform Coding

The sine-wave model for a short duration segment of speech is given by

$$s(n) = \sum_{k=1}^N A_k \cos(nk\omega_0 + \phi_k) \quad (2.1)$$

where ω_0 is the fundamental frequency, A_k are the sine-wave amplitudes and ϕ_k are the sine-wave phases [3, 5]. In sinusoidal transform coding the parameters ω_0 , A_k and ϕ_k are estimated, coded and transmitted at a frame rate of 15 to 20 milliseconds. The receiver then synthesizes each frame of speech using the model in (2.1) and successive segments are combined with an overlap-add technique. The pitch ω_0 is estimated by minimizing the mean squared error between the observed and reconstructed speech [6, 5]. The sine-wave phases ϕ_k are estimated by sampling the speech spectrum $S(\omega)$ at the pitch harmonics $k\omega_0$ where $S(\omega)$ is the Fourier transform of $s(n)$. The harmonic sine-wave amplitudes are estimated as in the Spectral Envelope Estimation Vocoder (SEEVOC) [7].

For good speech reconstruction, directly coding the pitch, the amplitudes and the phases requires at least 13000 bits per second. Encoding speech at low data rates (4800 to 1200 bits per second) requires modeling the amplitudes and the phases such that the parameters of the model can be coded more efficiently than the parameters themselves. The phase of unvoiced speech is modeled as a uniformly distributed random variable on the interval $[-\pi < \phi_k < \pi]$ where the amount of the speech spectrum that is unvoiced is determined by the degree to which the harmonic model is well-fitted to the original set of sine-waves, a parameter that can be transmitted using from 2 to 4 bits. During voiced speech the phase is modeled as a linear combination

of the phase of the vocal tract transfer function phase and the excitation phase. A synthetic excitation phase is generated by the synthesizer and a minimum phase model of the vocal tract transfer function is used, where the magnitude of the transfer function is estimated by interpolating the sine-wave amplitudes using cubic spline functions. With this model, an appropriate set of sine-wave phases can be generated at the receiver from the pitch and the cubic spline envelope of the sine-wave amplitudes so that the voiced speech phase information does not need to be coded explicitly. Using the harmonic, minimum-phase model with the harmonic samples of the cubic spline envelope allows very good quality speech to be synthesized.

Various models have been used to represent the sine-wave amplitudes including the SEEVOC [7], cepstral [3] and all-pole [1, 2] models. The latter model has proven to be the most efficient for coding speech at 2400 bits per second, and it is that model which will be the focus of this thesis. The form of the model is

$$H(z) = \frac{G}{A(z)} \quad (2.2)$$

where

$$A(z) = 1 + \sum_{i=1}^M a_i z^{-i}. \quad (2.3)$$

Using the terminology associated with linear prediction, $A(z)$ is the inverse filter, a_i are the predictor coefficients, G is the gain, and M is the model order [8, 9]. In direct linear predictive analysis of the speech waveform the predictor coefficients are computed from the autocorrelation coefficients of the time waveform. For a sampling rate of 8000 Hertz this requires that the model order be limited to about 10. If the model order is increased beyond 10, the envelope of $H(z)$ can begin to resolve the harmonics of the underlying speech spectrum during voiced speech, and this results in poor estimates of the sine-wave amplitudes, particularly when the parameters are quantized.

The problem of a limited model order is avoided in STC by using an alternate technique to estimate an all-pole model of the speech spectral envelope. The model representation can then be made arbitrarily close to the estimated spectral envelope

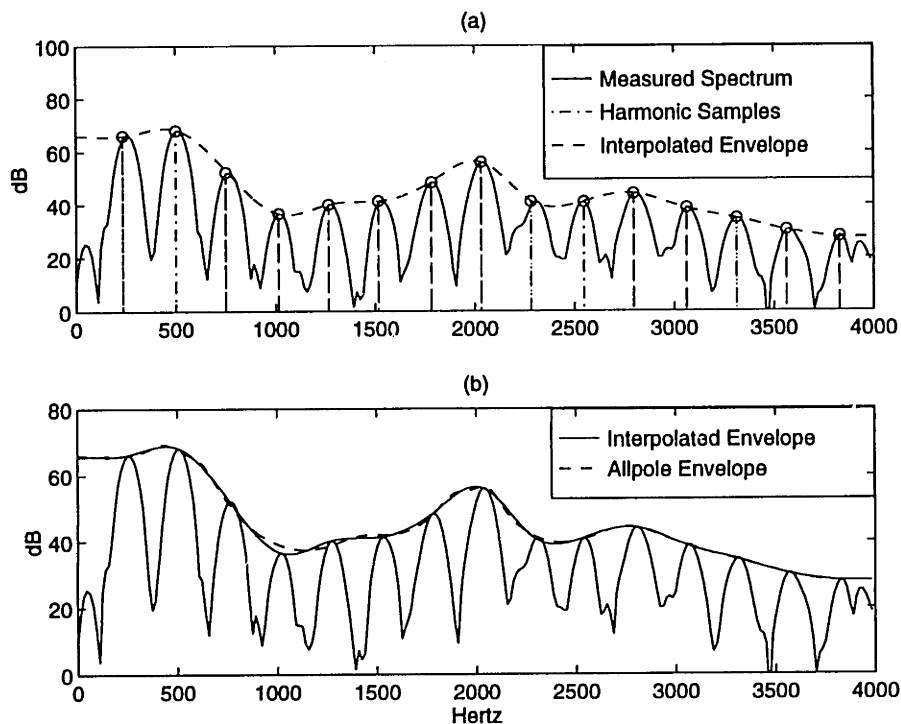


Figure 2-1: Example of (a) cubic spline interpolation of sine-wave amplitudes and (b) 22nd order all-pole model of the interpolated envelope.

by increasing the order of the model. The underlying spectral envelope is first estimated by interpolating the logarithm of the sine-wave amplitudes using cubic spline functions and the all-pole model is then fitted to this estimate of the spectral envelope [1]. In this method the autocorrelation coefficients are found using the inverse cosine transform of the interpolated sine-wave amplitudes and the predictor coefficients are then computed using Durbin's recursion. An all-pole model order of 22 has been found to be sufficient to reproduce good quality speech. Figure 2-1 shows an example of the cubic spline interpolation of the sine-wave amplitudes in comparison with a 22nd order all-pole model of the estimated envelope.

Very good quality speech can be synthesized by using the spline envelope itself to represent the sine-wave amplitudes but when an all-pole model is used to represent the amplitudes a high model order is needed to maintain this good quality. When using a model order greater than 10, good quality cannot be achieved with time domain linear predictive analysis. Figure 2-2 shows that rather than fitting the spectral envelope more closely as the model order is increased, time domain linear predictive

analysis begins to resolve the pitch harmonics and the estimated spectral envelope develops sharp peaks. This contrasts the all-pole fit to the spline interpolation of the sine-wave amplitudes which fits the underlying spectral envelope more closely as the model order is increased.

It was found that warping the interpolated spectral envelope on a perceptual scale prior to fitting the all-pole model allows a reduction in model order from 22 to 14 while maintaining good quality in the reconstructed speech [2]. The warping function used is

$$W(\omega) = \alpha \log(1 + \beta\omega) \quad (2.4)$$

where the parameters α and β allow flexibility in designing a coder at multiple bit rates. Values of $\alpha = 170$ and $\beta = 0.554$ are found to be adequate for coding at rates of 2400 to 4800 bits per second. Figure 2-3 demonstrates how the 14th order spectral fit can be improved using perceptual warping. In the low frequency region where the ear is sensitive to narrower bandwidth distortions, the fit to the estimated spectrum is better than in the high frequency region where the ear is less sensitive to a similar distortion. This contrasts the all-pole model of the unwarped spectrum in which the distortion is more uniform with respect to frequency.

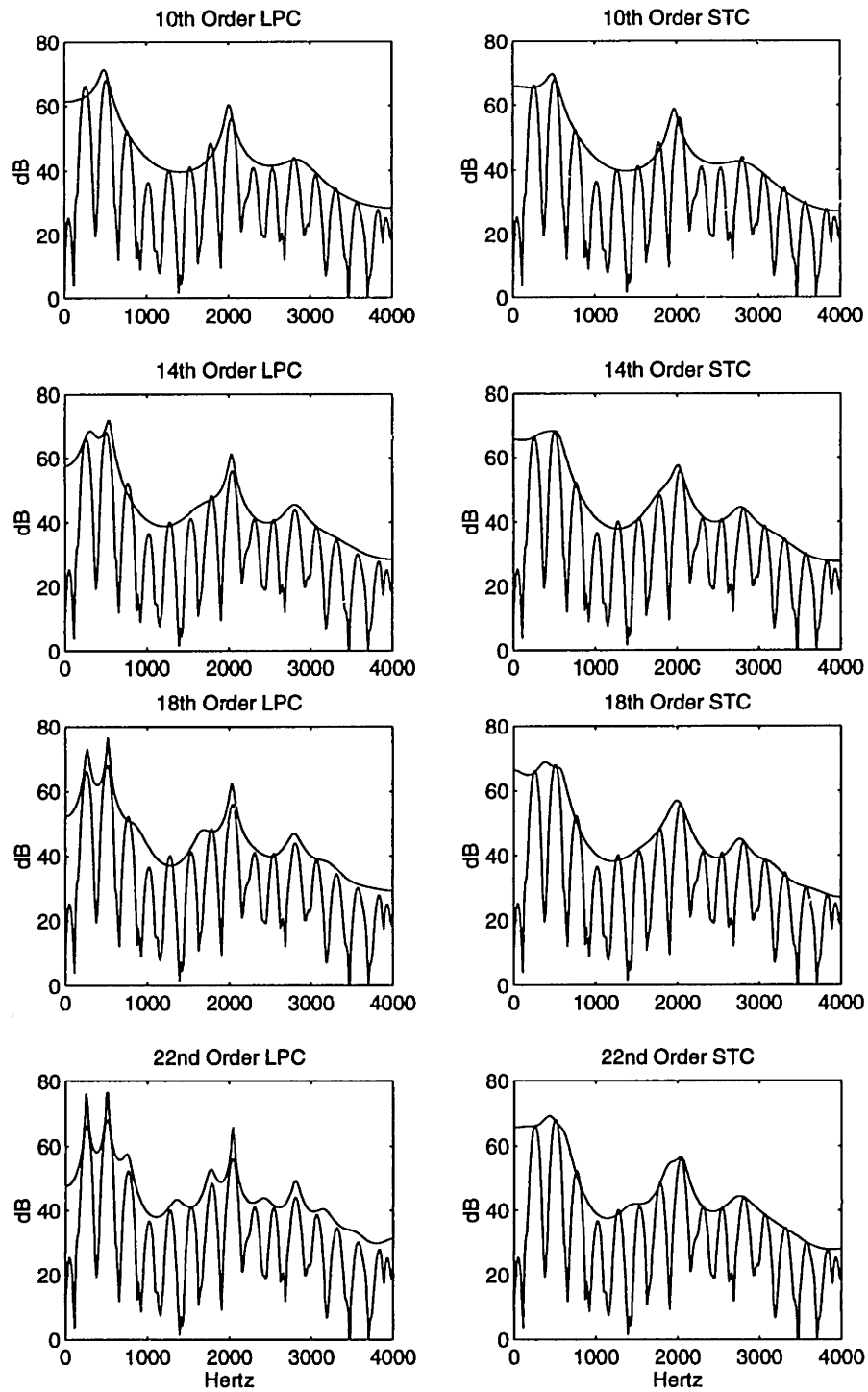


Figure 2-2: Comparison of time domain linear predictive analysis (LPC) and all-pole fit to interpolated sine-wave amplitudes (STC).

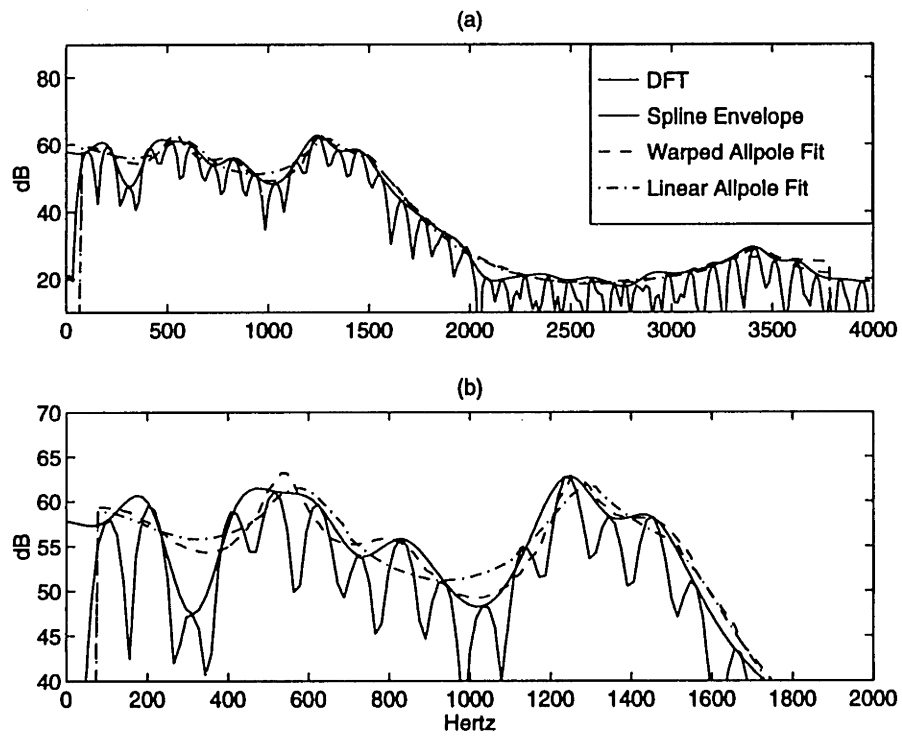


Figure 2-3: Warped and unwarped fit to spline envelope.

3 All-Pole Model Parameter Computation

The all-pole model can be characterized by the predictor coefficients, as shown in equation (2.3). However, these parameters are known to have poor quantization properties and hence are not used for speech coding. Their dynamic range is relatively large and the stability of the filter $H(z)$ cannot be guaranteed when they are quantized [10]. Instead the predictor coefficients are transformed to an alternate representation which is more suitable for coding such as partial correlation (PARCOR) coefficients or line spectral frequencies (LSFs). A third representation introduced recently in [4] uses a mixture of LSFs and PARCOR coefficients to represent the all-pole model. There are a number of methods for transforming from the predictor coefficients to these parameter sets and for transforming back to the predictor coefficients. In this thesis the numerical accuracy required for these transformations is determined as well as their computational complexity.

3.1 All-Pole Model Representations

The inverse filter $A(z)$ can be implemented as a lattice filter (assuming $H(z)$ is stable) where the partial correlation coefficients k_i in the lattice are then uniquely related to the predictor coefficients [8]. These PARCOR coefficients are very useful for coding. They have limited dynamic range ($-1 \leq k_i \leq 1$), the stability of the reconstruction filter $H(z)$ can be guaranteed when they are quantized, and they are efficiently computed from the autocorrelation coefficients using Durbin's recursion. The PARCOR coefficients can also be computed using an algorithm that minimizes the forward and

backward prediction error in the lattice filter implementation of $A(z)$ [11]. This second method has the advantage that each PARCOR coefficient k_m can be computed based on the quantized values of the previous coefficients \hat{k}_1 through \hat{k}_{m-1} (where \hat{k}_m represents the quantized value of k_m .) This possibly compensates for some of the quantization effects.

A more efficient parameter set for coding the all-pole model is the line spectral frequencies. This is the representation currently used in STC to represent the all-pole model of the sine-wave amplitudes for coding at 4800 to 2400 bits per second. The LSFs for an M th order all-pole model are defined as follows. Two artificial polynomials of order $M + 1$ are created from the M th order inverse filter $A(z)$ according to

$$P(z) = A(z) + z^{-(M+1)}A(z^{-1}) \quad (3.1)$$

$$Q(z) = A(z) - z^{-(M+1)}A(z^{-1}). \quad (3.2)$$

The line spectral frequencies f_i correspond to the roots of $P(z)$ and $Q(z)$ which are on the unit circle (i.e. at $z = e^{j2\pi f_i}$), where the trivial roots that always occur at $f = \frac{1}{2}$ and $f = 0$ are ignored. Substantial research has shown that the LSFs can be coded efficiently and the stability of the synthesis filter can be guaranteed when they are quantized. This parameter set has the advantage of better quantization and interpolation properties than the corresponding PARCOR coefficients [12]. However it has the disadvantage that solving for the roots of $P(z)$ and $Q(z)$ can be more computationally intensive than computing the PARCOR coefficients.

The all-pole model can also be represented using a mixture of LSFs and PARCOR coefficients as in [4]. In this representation the M th order all-pole model is specified by N LSFs, f_1 through f_N (where $N < M$), and by PARCOR coefficients, k_{N+1} through k_M . The predictor coefficients for this all-pole model can then be obtained as in Figure 3-1. The N th order LSFs are converted to predictor coefficients which are then converted to PARCOR coefficients k_1 through k_N . These are combined with k_{N+1} through k_M and the M th order predictor coefficients are computed from the M

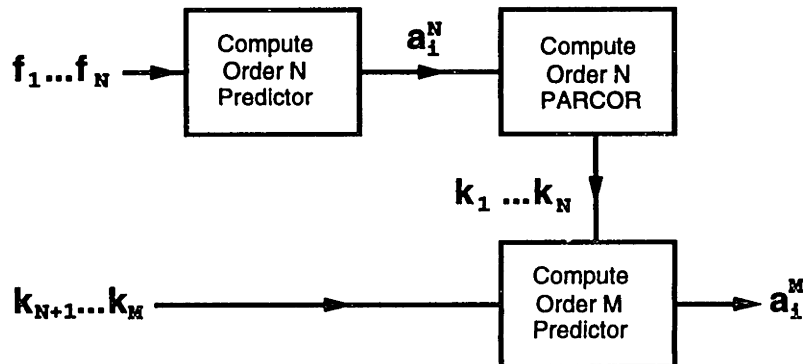


Figure 3-1: System for computing predictor coefficients from mixed LSF/PARCOR representation.

PARCOR coefficients.

3.2 Parameter Transformations

Several methods can be used to solve for the roots of $P(z)$ and $Q(z)$ to determine the LSFs. To find these roots $e^{j\omega}$ is substituted for z in (3.1) and (3.2) and the following functions with equivalent (non-trivial) roots are formed

$$P'(\omega) = \cos(L\omega) + \hat{p}_1 \cos((L-1)\omega) + \hat{p}_2 \cos((L-2)\omega) + \dots + \frac{\hat{p}_L}{2} \quad (3.3)$$

$$Q'(\omega) = \cos(L\omega) + \hat{q}_1 \cos((L-1)\omega) + \hat{q}_2 \cos((L-2)\omega) + \dots + \frac{\hat{q}_L}{2} \quad (3.4)$$

where $L = \frac{M}{2}$ and \hat{p}_i and \hat{q}_i are determined as in Appendix A [13]. Using the appropriate trigonometric substitutions, $P'(\omega)$ and $Q'(\omega)$ can be converted to L th order polynomials in $\cos(\omega)$. Since polynomials of order 4 and less can be solved in closed form, the L th order polynomial can then be used to efficiently compute the LSFs when the all-pole model order is 8 or less. When the model order is greater than 8 a closed form solution is not known as polynomials of order 5 and greater cannot be solved in closed form [14]. For these higher model orders the roots of $P'(\omega)$ and $Q'(\omega)$ are found numerically with root solving techniques. One method described in [13] directly searches for roots of (3.3) and (3.4). A second method expresses (3.3) and (3.4) as series expansions in Chebyshev polynomials and thereby significantly

reduces the computational burden for finding the roots [15].

The inverse transformation from LSFs back to predictor coefficients is computed directly from

$$A(z) = \frac{1}{2}[P(z) + Q(z)] \quad (3.5)$$

using

$$P(z) = (1 + z^{-1}) \prod_{i=1}^{M/2} [1 - 2 \cos(2\pi f_i^p) z^{-1} + z^{-2}] \quad (3.6)$$

$$Q(z) = (1 - z^{-1}) \prod_{i=1}^{M/2} [1 - 2 \cos(2\pi f_i^q) z^{-1} + z^{-2}] \quad (3.7)$$

where f_i^p are the LSFs corresponding to the roots of $P(z)$ and f_i^q are the LSFs corresponding to the roots of $Q(z)$. The computation for this transformation can be reduced by a factor of 4 by applying the method described in [15] using series expansions in Chebyshev polynomials.

An all-pole model order of at least 14 is needed to to synthesize good quality speech but the numerical root solving techniques used to find the LSFs are computationally expensive. An advantage of the mixed LSF/PARCOR representation is that the LSF order can be limited to $N = 8$ so that the efficient closed form solution can be used to compute the LSFs. Two different methods are then used to compute the mixed representation. The first, shown in Figure 3-2, uses Durbin's recursion to compute the predictor and PARCOR coefficients. This recursion is given by

$$E_0 = R(0) \quad (3.8)$$

$$k_i = - \left[R(i) + \sum_{j=1}^{i-1} a_j^{(i-1)} R(i-j) \right] / E_{i-1} \quad (3.9)$$

$$a_i^{(i)} = k_i \quad (3.10)$$

$$a_j^{(i)} = a_j^{(i-1)} + k_i a_{i-j}^{(i-1)}, \quad 1 \leq j \leq i-1 \quad (3.11)$$

$$E_i = (1 - k_i^2) E_{i-1} \quad (3.12)$$

where $R(i)$ are the autocorrelation coefficients and equations (3.9)-(3.12) are solved

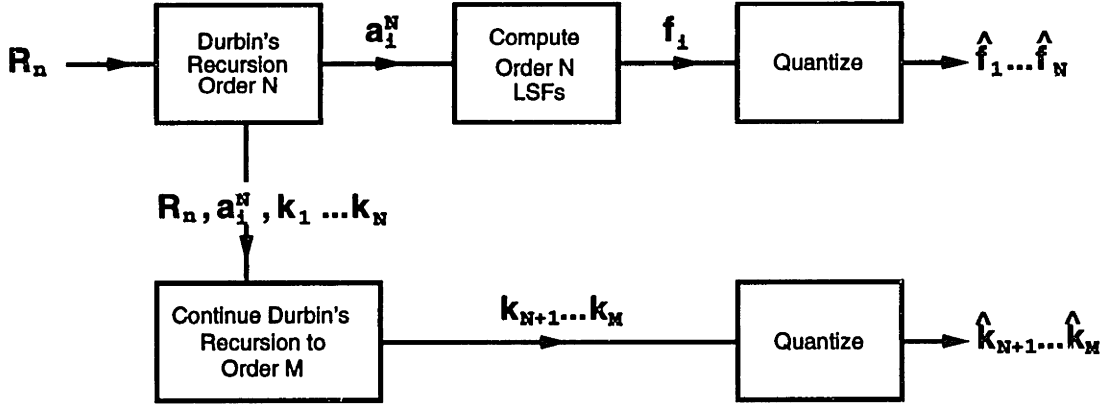


Figure 3-2: System for computing mixed LSF/PARCOR representation using Durbin's recursion.

recursively for $i = 1, 2, \dots, N$ to determine the predictor coefficients a_j^N corresponding to an N th order all-pole model. The LSFs f_1 through f_N are computed from a_j^N using the closed form solution described in Appendix A. Then PARCOR coefficients k_{N+1} through k_M are computed by continuing Durbin's recursion from the N th iteration. The LSFs and PARCOR coefficients are then quantized independently.

A second method for computing the mixed representation is shown in Figure 3-3 where the LSFs and PARCOR coefficients are not quantized independently. Instead, PARCOR coefficients k_{N+1} through k_M are computed using a lattice formulation so that the quantization of the LSFs can be taken into account [4]. To begin, the N th order predictor coefficients a_j^N are computed from Durbin's recursion and the N th order LSFs f_1 through f_N are computed from a_j^N in closed form. The LSFs are quantized and these quantized values (\hat{f}_1 through \hat{f}_N) are transformed to predictor coefficients \hat{a}_j^N . This is done using either the direct computation in equations (3.6) and (3.7) or by using the series expansion in Chebyshev polynomials. Next, PARCOR coefficients \hat{k}_1 through \hat{k}_N are computed from \hat{a}_j^N using the backward recursion

$$k_i = a_i^{(i)} \quad (3.13)$$

$$a_j^{(i-1)} = \frac{a_j^{(i)} - a_i^{(i)} a_{i-j}^{(i)}}{1 - k_i^2}, \quad 1 \leq j \leq i-1 \quad (3.14)$$

where the index i takes on the values $N, N-1, \dots, 1$ in that order. Initially $a_j^{(N)} =$

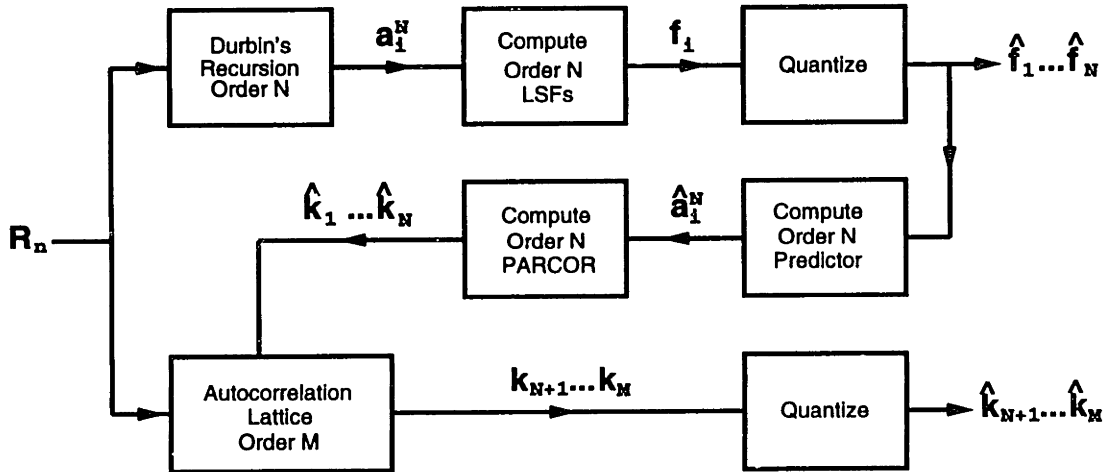


Figure 3-3: System for computing mixed LSF/PARCOR representation using the autocorrelation-lattice method.

\hat{a}_j^N , $1 \leq j \leq N$. Finally, the autocorrelation coefficients and \hat{k}_1 through \hat{k}_N are used in the *autocorrelation-lattice* method [11] to compute k_{N+1} through k_M . In [4] these last PARCOR coefficients were quantized with vector quantization after all $M - N$ of them had been computed. When scalar quantization is used these PARCOR coefficients can be quantized within the autocorrelation-lattice computation.

To convert the mixed LSF/PARCOR representation back to predictor coefficients (Figure 3-1) the quantized LSFs are first transformed to N th order predictor coefficients using either the direct form or the expansion in Chebyshev polynomials. These predictor coefficients are transformed to \hat{k}_1 through \hat{k}_N , and combined with \hat{k}_{N+1} through \hat{k}_M which have been transmitted directly to provide the PARCOR representation of the M th order system.

3.3 Arithmetic Accuracy Requirements

It has been observed that numerical accuracy problems can occur in computations involving LSFs when the model order is 14 or greater. These problems are avoided in workstation-based implementations by using double precision arithmetic, but for real-time DSP implementations such as on the TMS320C30, double precision arithmetic must be performed in software causing a substantial increase in computational

complexity. The various methods for computing LSFs from predictor coefficients and the corresponding inverse transformations are closely examined below to determine which transformations require double precision so that the use of higher precision arithmetic can be avoided when unnecessary.

A number of differences must be accounted for when comparing the accuracy of the workstation implementation with the accuracy of the DSP implementation. The numerical output of the two implementations will in general differ because even single precision arithmetic is implemented differently on the two processors. The compilers are also different and some arithmetic operations are likely to be performed in a different order causing further differences in roundoff error. Another problem is that when observing differences in the values of LSFs it is not readily apparent how much of a difference is significant. A small error in an LSF can lead to a large error in the corresponding spectral envelope when two LSFs are close together [16].

Instead of comparing LSFs directly, the LSFs are converted back to predictor coefficients and the corresponding spectral envelopes are compared. These envelopes are computed in the workstation using double precision arithmetic. The various methods for transforming to and from LSFs were first tested on the workstation using a double precision implementation. A reference spectral envelope computed from the original predictor coefficients for a given all-pole model was compared to the spectral envelope computed from the predictor coefficients which had been transformed to LSFs and back as in Figure 3-4. Over 4500 frames of speech from 6 different speakers (3 male and 3 female) were tested and the largest difference between the envelopes at any point for any frame was less than 1 thousandth of a dB regardless of the method used for the transformations. The double precision workstation implementation was judged to have negligible numerical error associated with it.

To observe the accuracy of the DSP transformation from predictor coefficients to LSFs, those transformations were performed on the DSP while the inverse transformation from LSFs back to predictor coefficients was performed on the host workstation using double precision arithmetic. The reference spectral envelope was compared to the spectral envelope corresponding to the predictor coefficients that were computed

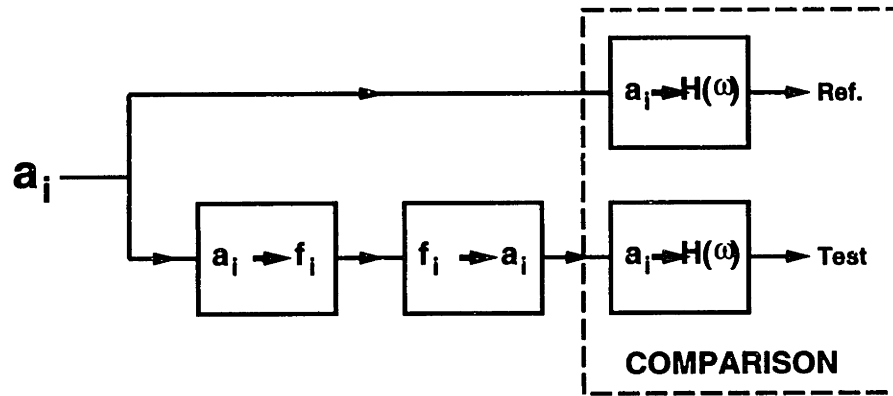


Figure 3-4: System for comparing reference spectrum with computed spectrum.

on the DSP. The inverse transformations from LSFs back to predictor coefficients were tested in a similar manner where the forward transformation was computed on the workstation using double precision arithmetic and the inverse transformation was computed on the DSP.

The workstation used was a Sun Microsystems SPARCstation 10 and the DSP was a Texas Instruments TMS320C30. To avoid a loss in accuracy when transferring data between the SPARCstation and the TMS320C30, all data was transferred in IEEE single precision binary format. Although the SPARCstation uses the IEEE format to represent floating point numbers the TMS320C30 used its own floating point representation. A format conversion between the IEEE format to the TMS320C30 floating point format is performed in software by the TMS320C30. This format conversion is exact except for “soft” zeros (denormalized members) [17].

The worst case spectral error was measured for the transformation from predictor coefficients to LSFs using three methods: the closed form solution (8th order model only), direct root solving, and by using expansions in Chebyshev polynomials. The direct root solving was implemented using code from the current real-time version of STC where the resolution of the search for roots was adjusted to provide sufficient accuracy while minimizing computation time. For the method using expansions in Chebyshev polynomials, the initial search resolution was adjusted such that the algorithm would require the same computation time as the direct root solving method. Table 3.1 shows the largest magnitude error in the spectral envelope (for the 4500

Root Solving Method	All-pole Model Order		
	8	14	22
Closed Form	0.006	—	—
Direct Root Solving	0.156	0.230	0.337
Polynomial Expansion	0.005	0.010	0.036

Table 3.1: Spectral error (in dB) when computing LSFs from predictor coefficients using *single* length arithmetic.

frames) when LSFs are computed from predictor coefficients using the DSP. This error is on the order of a few hundredths of a dB for the transformation from predictor coefficients to LSFs using the series expansion in Chebyshev polynomials. The direct root solving method has an error which is roughly an order of magnitude greater. It was also observed that significantly increasing the initial search resolution for the series expansion in Chebyshev polynomials in order to further increase accuracy would cause the algorithm to fail when using single precision arithmetic for model orders of 14 and greater.

The conversion of LSFs back to predictor coefficients requires greater arithmetic precision than the forward transformation. The worst case spectral error for both the direct computation of predictor coefficients and computation using the series expansion in Chebyshev polynomials is over 40 dB for 18th and 22th order systems. Figure 3-5 shows an example of error in the spectral envelope caused by transforming 22nd order LSFs to predictor coefficients using single precision arithmetic on the DSP. Table 3.2 shows that for systems of order 8 and 10 the worst case error is less than 1 tenth of a dB which is acceptable. For the 14th order system the worst case error shown in the table is about 1 dB. This should not seriously degrade quality but the measurement was made using a limited amount of speech (4500 frames.) Listening tests performed over a larger database revealed that using single precision arithmetic for a 14th order system occasionally results in speech frames with poorly reconstructed envelopes.

The dynamic range of floating point numbers on the DSP can be increased from about 6 decimal digits using single length arithmetic to more than 13 decimal digits

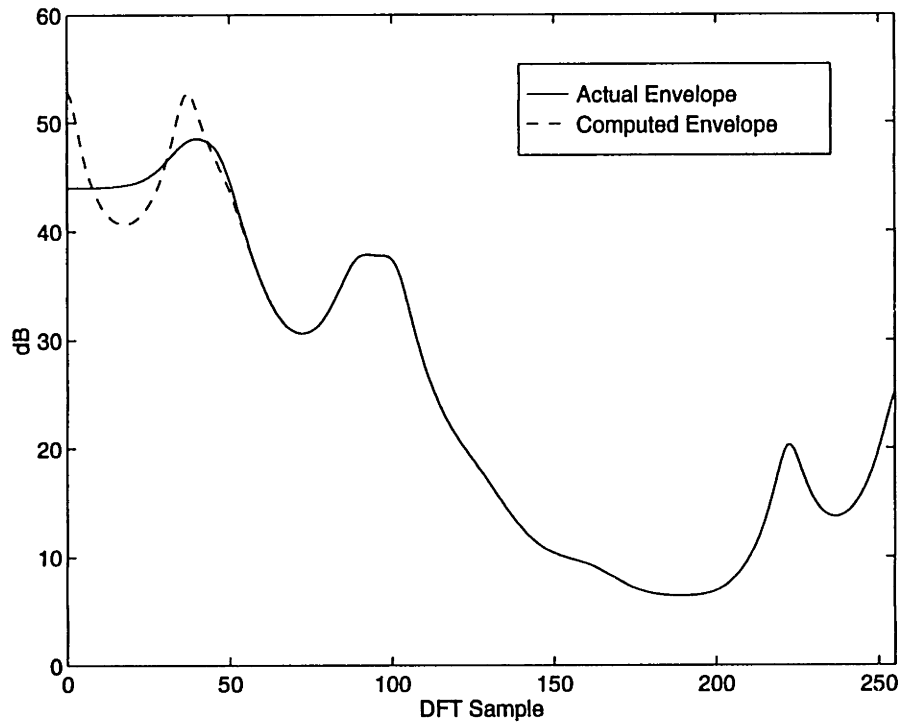


Figure 3-5: Comparison of reference spectral envelope with test spectral envelope where the test envelope has distortion caused by roundoff error in single precision arithmetic.

Transformation Method	All-pole Model Order				
	8	10	14	18	22
Direct Computation	0.023	0.092	1.108	43.63	72.58
Polynomial Expansion	0.020	0.060	0.898	55.87	81.12

Table 3.2: Spectral error (in dB) when computing predictor coefficients from LSFs using *single* length arithmetic.

Transformation Method	All-pole Model Order				
	8	10	14	18	22
Direct Computation	0.001	0.001	0.001	0.001	0.002
Polynomial Expansion	0.001	0.001	0.001	0.001	0.001

Table 3.3: Spectral error (in dB) when computing predictor coefficients from LSFs using *double* length arithmetic.

using double length arithmetic [17]. However using double length arithmetic substantially increases computation time. Single precision addition and multiplication each require only one instruction cycle on the TMS320C30 but double length addition requires 25 instruction cycles and double length multiplication requires 35 instruction cycles. A software implementation of double precision IEEE floating point requires even more computation time than double length arithmetic. Table 3.3 shows that the worst case error for the conversion from LSFs to predictor coefficients using double length arithmetic is only two thousands of a dB. But use of double length arithmetic increases the computation time for this transformation by a factor of from 5 to 10.

Single precision arithmetic is adequate for computing PARCOR coefficients but the mixed LSF/PARCOR representation requires computations involving LSFs. If the order for the LSFs is restricted to $N = 8$ then it can be seen from Tables 3.1 and 3.2 that single precision arithmetic is adequate for computing the transformations to and from the mixed representation. This allows the use of computationally expensive double length arithmetic to be avoided.

3.4 Computational Complexity

The computational complexity of various transformations was determined by measuring the time taken to perform the requisite computations on the TMS320C30. This was measured by using the internal TMS320C30 timer to periodically increment a counter and observing the average count obtained during the execution of the subroutine. The average count is multiplied by the timer period to determine the average time used by the routine. The timer period chosen must be short enough to have ad-

All-Pole Representation	Transformation Method	All-pole Model Order	Time in ms
LSF	Direct Root Solving	14	1.19
LSF	Polynomial Expansion	14	1.16
LSF/PARCOR	Autocorrelation-Lattice	14	1.13
LSF/PARCOR	Durbin's Recursion	14	0.35
LSF	Direct Root Solving	22	1.71
LSF	Polynomial Expansion	22	1.84
LSF/PARCOR	Autocorrelation-Lattice	22	2.42
LSF/PARCOR	Durbin's Recursion	22	0.47

Table 3.4: Time needed to compute all-pole model parameters.

equate resolution but not so short that the timer interrupts occur too frequently and thereby account for a significant amount of processor time. A timer counter period of 0.1 milliseconds was used and in practice it was found that this method for timing gave a resolution of about 0.01 milliseconds (it is assumed that the subroutine always begins execution with a random delay from the previous counter increment.)

The time needed to compute the all-pole representation from the autocorrelation coefficients is shown in Table 3.4. This demonstrates that for a 14th order system the mixed LSF/PARCOR representation using Durbin's recursion is about three times faster than the other three methods. Direct root solving, the polynomial expansion, and the mixed LSF/PARCOR representation using the autocorrelation-lattice method all require roughly the same amount of time. For a 22nd order system the mixed LSF/PARCOR representation using Durbin's recursion is again over three times faster than direct root solving or the polynomial expansion method, while computing the mixed LSF/PARCOR representation using the autocorrelation-lattice method requires more time than any of the other methods.

For the inverse transformation shown in Table 3.5, the time needed to compute the predictor coefficients from the mixed LSF/PARCOR representation is much less than the time needed to compute them from a 14th order LSF representation. This is in part because the transformation of 14th order LSFs to predictor coefficients must be performed with double length arithmetic while for the mixed LSF/PARCOR

All-Pole Representation	Transformation Method	All-pole Model Order	Time in ms
LSF	Direct Computation	14	3.59
LSF	Polynomial Expansion	14	1.14
LSF/PARCOR	Polynomial Expansion	14	0.21
LSF	Direct Computation	22	8.39
LSF	Polynomial Expansion	22	2.42
LSF/PARCOR	Polynomial Expansion	22	0.27

Table 3.5: Time needed to compute predictor coefficients from all-pole model parameters.

representation where the order of the LSFs is limited to 8, this transformation can be done with single precision arithmetic.

3.5 Discussion

The above analysis shows that for a 14th order LSF system, the fastest and most accurate method for computing LSFs from predictor coefficients and for the inverse transformation from LSFs to predictor coefficients is the method using a series expansion in Chebyshev polynomials. The inverse transformation was found to require double length arithmetic when the all-pole model order is 14 or greater. Table 3.6 compares the computation time for this method to the computation time for the mixed LSF/PARCOR representation using the autocorrelation-lattice method and using Durbin's recursion. The mixed representation requires substantially less computation than a 14th order LSF representation and if the autocorrelation-lattice method is not used there is an additional substantial savings in computation. The worst case error in the spectral envelope is comparable for all three methods.

All-Pole Representation	Transformation Method	All-pole Model Order	Time in ms
LSF	Polynomial Expansion	14	2.31
LSF/PARCOR	Autocorrelation-Lattice	14	1.34
LSF/PARCOR	Durbin's Recursion	14	0.56
LSF	Polynomial Expansion	22	4.26
LSF/PARCOR	Autocorrelation-Lattice	22	2.50
LSF/PARCOR	Durbin's Recursion	22	0.73

Table 3.6: Total time needed to compute all-pole model parameters and then transform coded parameters to predictor coefficients.

4 Parameter Coding

In the previous section it was shown that the mixed LSF/PARCOR representation requires substantially less computation time than a purely LSF representation. In this section the coding properties of the mixed LSF/PARCOR representation are examined using both scalar and vector quantization to determine if there is a cost in quantization efficiency associated with using the mixed LSF/PARCOR representation. While in [4] the mixed LSF/PARCOR representation was used to improve quantization efficiency for a 10th order LPC system, the mixed representation is applied here to a 14th order all-pole model of the sine-wave amplitudes in STC. The quantization efficiency of the mixed LSF/PARCOR representation is compared to that of the 14th order LSF representation currently used in STC for speech coding at 2400 bits per second. The quality of the speech synthesized from both systems is evaluated using a perceptually weighted spectral distortion measure and using blind AB listening tests.

An objective measure for speech quality which has been used extensively is the L_2 norm of the log spectral distance [18, 19, 20]. If the L_2 norm is weighted with respect to frequency, the correlation of the distortion measure with subjective measures of speech quality increases significantly [21]. The weighted spectral distortion measure is

$$SD_w(i) = \left[\frac{1}{\pi} \int_0^\pi W(\omega) [10 \log_{10} P_i(\omega) - 10 \log_{10} \hat{P}_i(\omega)]^2 d\omega \right]^{1/2} \quad (4.1)$$

where $P_i(\omega)$ is the power spectrum of the all-pole model on frame i , $\hat{P}_i(\omega)$ is the power spectrum of the quantized all-pole model on frame i and $W(\omega)$ is a perceptually based weighting function described in Appendix B. Averaging this weighted spectral distortion is a measure of speech quality. Another useful measure of quality is to

examine the percentage of frames for which the weighted distortion exceeds specified thresholds, typically 2 dB and 4 dB, because a few frames with large distortion can result in poor quality even though the average distortion is small [19, 20].

4.1 Scalar Quantization

The current version of STC codes a 14th order all-pole model of the sine-wave amplitudes with an LSF representation using 50 bits. The LSFs are differentially coded using a scalar quantizer designed as in [22]. For comparison, a 14th order mixed LSF/PARCOR representation (with $N = 8$) was also scalar quantized with 50 bits. In low-rate STC coding frame-fill techniques are used to encode the all-pole model on alternate subframes, however frame-fill techniques for the mixed LSF/PARCOR representation have not been studied. In the comparisons below, *every* subframe is quantized.

Properties of the PARCOR coefficients can give some insight into the coding properties of the mixed LSF/PARCOR representation. The spectral envelope of the all-pole model is increasingly sensitive to quantization error in any one of the PARCOR coefficients as the magnitude of that coefficient value approaches a value of 1. An estimate of the sensitivity function (after [10]) is shown in Figure 4-1. The PARCOR coefficients for a 14th order model of the sine-wave amplitudes were collected from 790,000 frames of speech in the TIMIT database [23] and histograms were generated. The histograms in Figures 4-2 and 4-3 are scaled as samples of the probability density functions of the PARCOR coefficients and are compared with Gaussian distributions having means and variances corresponding to those of the PARCOR coefficients. These figures show that k_9 through k_{14} are all concentrated between -0.4 and +0.4 where their spectral sensitivity is nearly uniform. To minimize the expected coding distortion when using scalar quantization for k_9 through k_{14} a Lloyd-Max quantizer designed for Gaussian random variables was used [24]. The 8 LSFs in the mixed representation are quantized as in [22].

The quantization efficiency of the mixed representation was evaluated using the

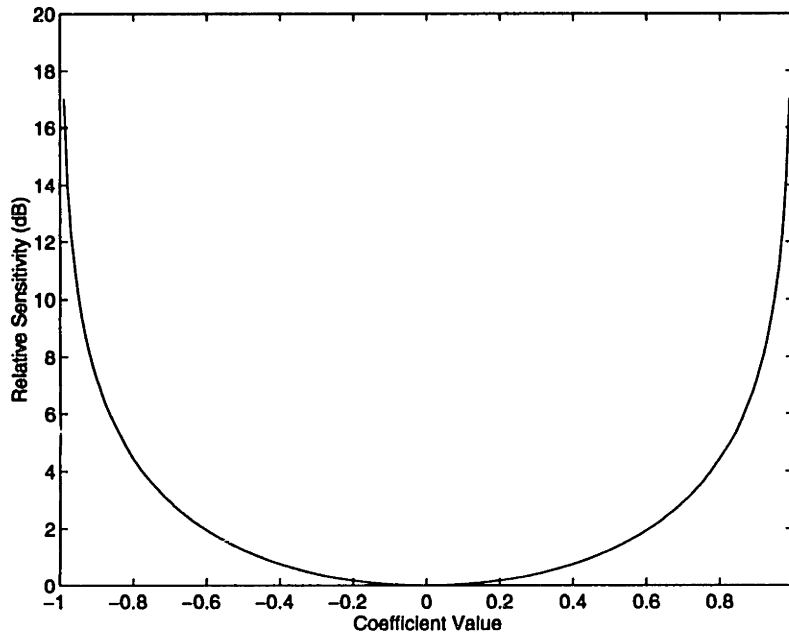


Figure 4-1: Relative spectral sensitivity of PARCOR coefficients.

autocorrelation-lattice method to compute each PARCOR coefficient based on the quantized values of the previously computed PARCOR coefficients. Quantization efficiency of the mixed representation was also tested using Durbin's recursion to compute all of the PARCOR coefficients as when the 8 LSFs and 6 PARCOR coefficients are quantized simultaneously.

The bit allocation for the 14th order LSF system was determined from extensive listening tests and is listed in Table 4.1. An appropriate bit allocation was needed for the mixed LSF/PARCOR representation. With 14 parameters and 50 bits to assign, some method was required to limit the number of combinations to be evaluated. The method used was as follows. Starting with a given bit allocation, the increase in mean distortion was observed when one of the parameters was quantized with one fewer bits. This was repeated for each of the parameters and a new bit allocation was formed by decreasing by one the number of bits assigned to the parameter which showed the least change in mean distortion and increasing by one the number of bits assigned to the parameter that showed the greatest change in mean distortion. This process was repeated until the new bit allocation no longer had a lower mean distortion than the previous allocation had. In this procedure the mean distortion

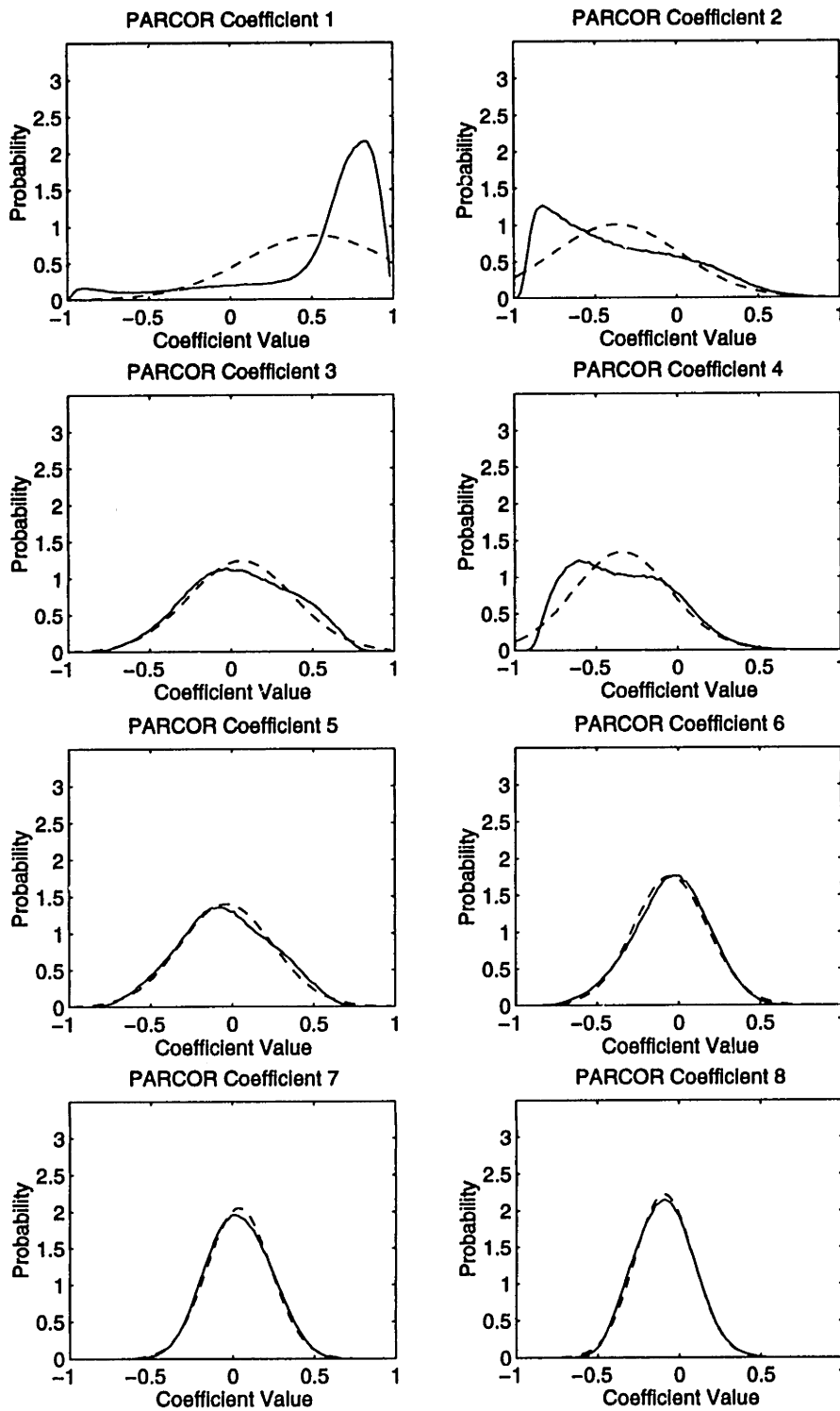


Figure 4-2: Probability density estimates for PARCOR coefficients k_1 through k_8 (solid lines) and Gaussian distributions (dashed lines).

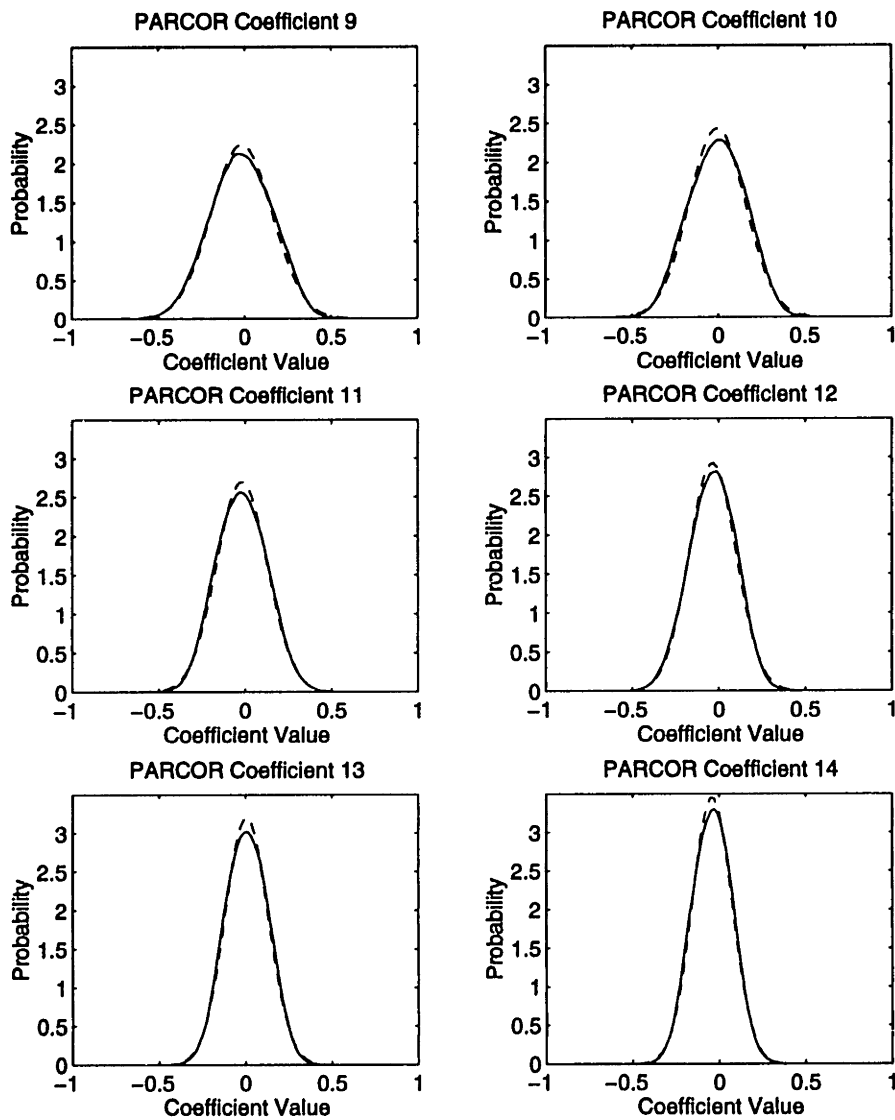


Figure 4-3: Probability density estimates for PARCOR coefficients k_9 through k_{14} (solid lines) and Gaussian distributions (dashed lines).

14th order LSF Bit Allocation														
Parameter	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	f_9	f_{10}	f_{11}	f_{12}	f_{13}	f_{14}
# of Bits	4	4	4	4	4	4	4	4	3	3	3	3	3	3

Mixed LSF/PARCOR Bit Allocation														
Parameter	f_1	f_2	f_3	f_4	f_5	f_6	f_7	f_8	k_9	k_{10}	k_{11}	k_{12}	k_{13}	k_{14}
# of Bits	5	4	4	4	5	5	4	4	3	3	3	2	2	2

Table 4.1: Bit allocations for scalar quantization of pure LSF and mixed LSF/PARCOR all-pole model parameters using 50 bits.

Coded Parameters	Mean SD_w in dB	% $SD_w > 2$ dB	% $SD_w > 4$ dB
Pure LSF	0.69	0.79	0.00
LSF/PARCOR (Lattice)	0.74	1.18	0.02
LSF/PARCOR (Durbin)	0.73	0.82	0.00

Table 4.2: Weighted spectral distortion in dB caused by scalar quantization with 50 bits.

was measured over 1000 frames of speech from 2 male and 2 female speakers. As a starting point for the process, the bit allocation in Table 4.1 for the 14th order LSF representation was used for the mixed representation where the bit allocations for f_9 through f_{14} for the 14th order LSF representation correspond to the bit allocations for k_9 through k_{14} for the mixed representation and the bit allocations for f_1 through f_8 have a direct correspondence between the two representations.

Using the bit allocations in Table 4.1 the distortion was measured for the three coding conditions: the current 14th order LSF, the mixed LSF/PARCOR representation with quantization of PARCOR coefficients within the autocorrelation-lattice computation, and the mixed LSF/PARCOR representation computed with Durbin's recursion. These measurements were made using a database of 6 speakers (3 male and 3 female) with 10 sentences per speaker. The results shown in Table 4.2 demonstrate that both forms of the mixed representation perform about as well as the 14th order LSF system. The mean distortion differs by only 0.05 dB between the three conditions tested.

Listener	Preference for Current LSF	Likelihood Assuming no Preference for Current LSF
A	15/24	0.27
B	12/24	0.73
C	12/24	0.73
D	10/24	0.92
E	10/24	0.92

Table 4.3: Listening test comparing scalar quantization of the current 14th order LSF representation with scalar quantization the mixed LSF/PARCOR representation.

A listening test was used to verify that the performance of the mixed LSF/PARCOR system using Durbin's recursion is as good as the 14th order LSF system. A two-alternative forced-choice experiment was performed using 24 sentences from the same speech database used in the distortion measurements (4 sentences each from 3 male and 3 female speakers). The experience level of the listeners ranged from none to extensive. Table 4.3 lists the number of times out of 24 that each listener preferred the current STC 14th order LSF system. The table also shows the likelihood that this score would occur if the listener could not tell the difference between the two conditions (i.e. the listener is choosing randomly). The likelihood function is described in more detail in Appendix C. These results show that there is no significant preference for the current 14th order LSF system over the mixed LSF/PARCOR system using Durbin's recursion. Most listeners reported that for a few sentences there was an apparent difference between the two systems but that for most sentences there was no noticeable difference.

4.2 Vector Quantization

Vector quantization has the potential to decrease the number of bits needed to code the all-pole model since it can take advantage of the correlation between elements in a vector. In addition, vector quantization allows for a fractional bit allocation which can lead to a more efficient allocation of bits than in scalar quantization.

The correlation between the LSFs of a 14th order all-pole model was measured over 138,000 frames of speech from the TIMIT database. Correlations were also measured between the LSFs of an 8th order all-pole model and between the PARCOR coefficients of a 14th order all-pole model. Figure 4-4 shows plots of the correlation between parameters where the horizontal axes correspond to the elements of the vector and the vertical axis is the correlation coefficient between the corresponding vector elements. In Figure 4-5 the correlation coefficient between the 6th vector element and each of the other vector elements is plotted for 14th order LSFs, 8th order LSFs, and 14th order PARCOR coefficients. The plots show that 14th order LSFs are more correlated than 8th order LSFs, and that 14th order PARCOR coefficients are less correlated still. For 14th order LSFs the correlation between adjacent LSFs is about 0.8 and for 8th order LSFs the correlation is about 0.65. For 14th order PARCOR coefficients this correlation is around 0.2. This significant amount of correlation suggests that a substantial coding gain can be achieved by using vector quantization instead of scalar quantization.

Vector quantization tables were trained using the LBG algorithm [25] with 190,000 14th order LSF vectors and 190,000 mixed LSF/PARCOR vectors. The vectors were generated from the TIMIT database and the input speech was equally divided between males and females using 163 speakers from all eight dialect regions in the database. Split vector quantization was performed [20] where the 14th order LSF vector was split into groups of 3, 3, 4 and 4, and the mixed LSF/PARCOR vector was split into groups of 4, 4, 3 and 3. Bit allocations, shown in Table 4.4, for the two representations were determined in the same manner described above for scalar quantization of the mixed LSF/PARCOR representation.

The test results are shown in Table 4.5 demonstrating that with 45 bits using vector quantization the 14th order LSF and the mixed LSF/PARCOR representations have similar performance to scalar quantization using 50 bits. A listening test was performed to compare 45 bit vector quantization of the mixed LSF/PARCOR system using Durbin's recursion to the current 14th order LSF scalar quantized system (50 bits). The results are shown in Table 4.6 and demonstrate that there is no significant

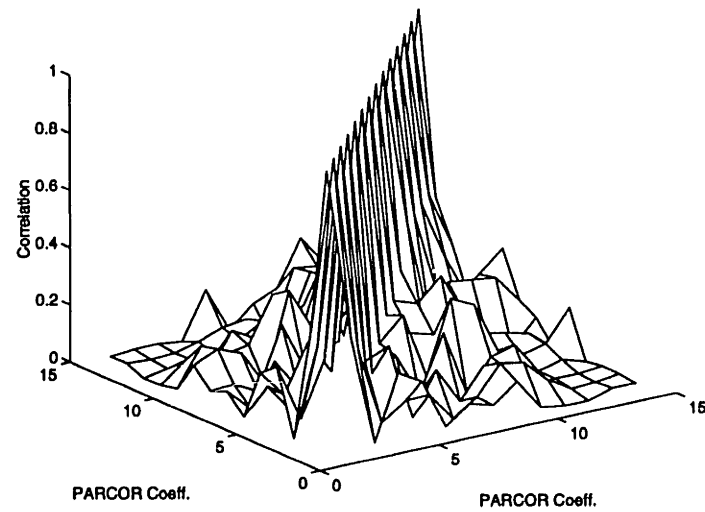
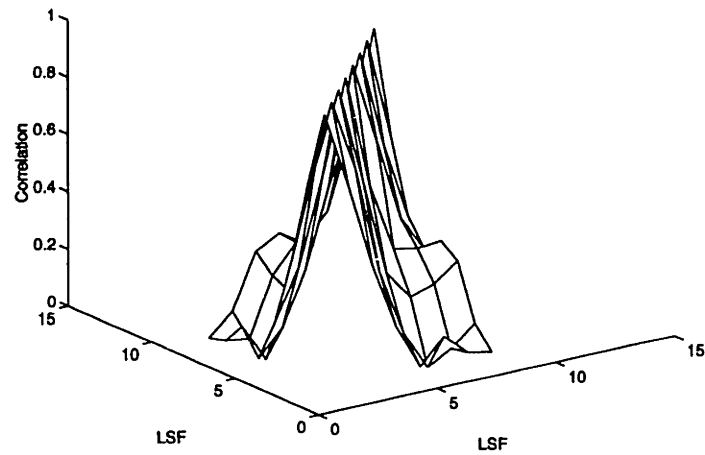
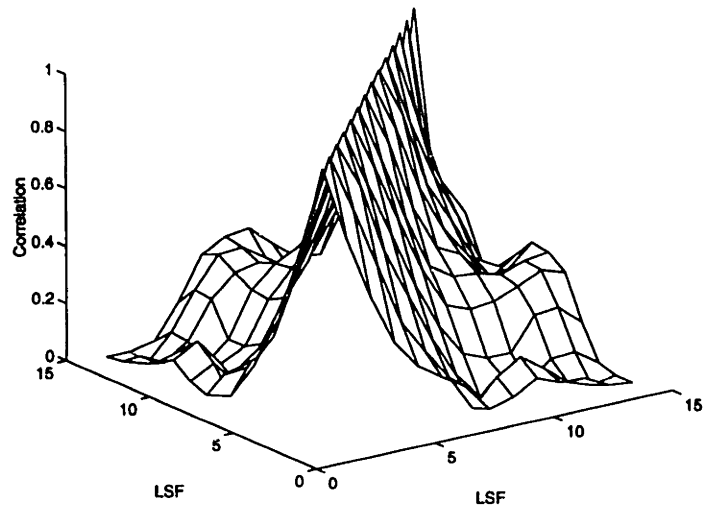


Figure 4-4: Correlation of 14th order LSFs (top), 8th order LSFs (middle) and 14th order PARCOR coefficients (bottom).

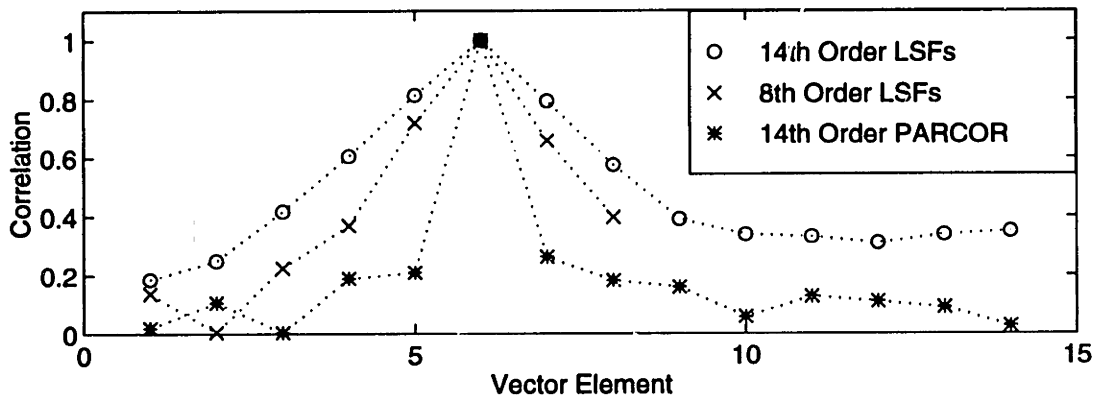


Figure 4-5: Correlation of the 6th vector element.

14th order LSF Bit Allocation				
Parameter Vector	f_1, f_2, f_3	f_4, f_5, f_6	f_7, f_8, f_9, f_{10}	$f_{11}, f_{12}, f_{13}, f_{14}$
# of Bits	11	10	12	12
Mixed LSF/PARCOR Bit Allocation				
Parameter Vector	f_1, f_2, f_3, f_4	f_5, f_6, f_7, f_8	k_9, k_{10}, k_{11}	k_{12}, k_{13}, k_{14}
# of Bits	15	15	8	7

Table 4.4: Bit allocations for vector quantization of pure LSF and mixed LSF/PARCOR all-pole model parameters using 45 bits.

Coded Parameters	Bits	Mean SD_w in dB	% $SD_w > 2$ dB	% $SD_w > 4$ dB
LSF Scalar	50	0.69	0.79	0.00
LSF Vector	45	0.76	0.62	0.00
LSF/PARCOR Vector (lattice)	45	0.80	1.68	0.05
LSF/PARCOR Vector (Durbin)	45	0.79	1.33	0.03

Table 4.5: Weighted spectral distortion caused by scalar and vector quantization.

Listener	Preference for Current LSF	Likelihood Assuming no Preference for Current LSF
A	14/24	0.41
B	10/24	0.92
C	12/24	0.72
D	12/24	0.72
E	12/24	0.72

Table 4.6: Listening test comparing the current scalar LSF quantization (50 bits) with vector quantization of the mixed LSF/PARCOR representation (45 bits).

difference between the two systems.

4.3 Discussion

The results of the above experiments show that scalar quantization of the mixed LSF/PARCOR representation performs as well as the current 14th order LSF scalar quantization. Furthermore, it has been shown that the number of bits needed to code the all-pole model of the sine-wave amplitudes can be decreased from 50 to 45 by using split vector quantization. This is true for the 14th order LSF representation as well as for the mixed LSF/PARCOR representation. It was also found that the objective evaluation of speech quality using the perceptually weighted spectral distortion measure correlates well with the results of the listening tests.

It is noteworthy that using the autocorrelation-lattice method to compute the PARCOR coefficients from quantized parameter values does not improve quantiza-

Coded Parameters	Mean SD_w in dB	% $SD_w > 2$ dB	% $SD_w > 4$ dB
PARCOR (Lattice)	0.94	4.88	1.62
PARCOR (Durbin)	1.02	5.60	2.79

Table 4.7: Weighted spectral distortion in dB caused by scalar quantization of 14th order PARCOR representation with 50 bits.

tion efficiency of the mixed representation. A likely explanation for this is that the expected spectral sensitivity to quantization of PARCOR coefficients k_9 through k_{14} is low compared to that of the first few coefficients. When coding a purely PARCOR representation there is a significant coding gain obtained by using the autocorrelation-lattice method and quantizing the PARCOR coefficients within the recursion. An objective measure of this observation is shown in Table 4.7 which demonstrates that although the mean distortion for the two coding methods is very close, there are at least 1.0% fewer outlier frames with distortion > 4 dB when quantization is done within the autocorrelation-lattice computation. An informal listening test confirmed that there is significant improvement in speech quality when quantization is done within the lattice structure.

5 Summary

The computational complexity of sine-wave amplitude coding for STC has been reduced by representing the all-pole model in terms of a mixture of LSFs and PARCOR coefficients. Using objective and subjective measures of performance, it has been shown that this was accomplished without a reduction in speech quality. In addition, it has been demonstrated that the use of split vector quantization can reduce the number of bits needed to code the spectrum from 50 to 45.

There are two reasons for the reduction in complexity associated with the mixed LSF/PARCOR representation. First, the need for computationally expensive root solving is avoided because the order of the LSFs can be limited to 8 so that a closed form solution can be used. Second, it was shown that the transformation from LSFs to predictor coefficients requires double length arithmetic, substantially increasing the computation time when a model order of 14 or higher is used. With the mixed representation, the order of the LSFs can be limited to 8, avoiding the use of double length arithmetic, while the overall model order can still be 14 or greater.

In [4] the lattice structure was used to compute the PARCOR coefficients after the LSFs had been quantized but it was found here that for the higher-order all-pole model in STC, computing the PARCOR coefficients based on the quantized LSFs does not improve quantization performance. The PARCOR coefficients can instead be computed using Durbin's recursion resulting in a substantial savings in computation time.

In future work, the frame-fill properties of the mixed LSF/PARCOR representation will be investigated as well as computationally efficient methods for vector quantization.

Appendix A

Closed Form Solution for 8th Order LSFs

The following is a closed form solution for computing the LSFs of an 8th order all-pole model from the predictor coefficients. This derivation is originally due to [13]. The predictor polynomial is

$$A(z) = 1 + \sum_{k=1}^m a_k z^{-k} \quad (\text{A.1})$$

where m is the model order and a_k are the predictor coefficients. The LSFs f_i are the frequencies corresponding to of the the roots of the polynomials

$$P(z) = A(z) + z^{-(m+1)} A(z^{-1}) \quad (\text{A.2})$$

$$Q(z) = A(z) - z^{-(m+1)} A(z^{-1}) \quad (\text{A.3})$$

which are on the unit circle (i.e. at $z = e^{j2\pi f}$.) Substituting the predictor polynomial in (A.2) and (A.3) gives

$$\begin{aligned} P(z) = & 1 + (a_1 + a_m)z^{-1} + (a_2 + a_{m-1})z^{-2} \dots \\ & + (a_{m-1} + a_2)z^{-(m-1)} + (a_m + a_1)z^{-m} + z^{-(m+1)} \end{aligned} \quad (\text{A.4})$$

$$\begin{aligned} Q(z) = & 1 + (a_1 - a_m)z^{-1} + (a_2 - a_{m-1})z^{-2} \dots \\ & + (a_{m-1} - a_2)z^{-(m-1)} + (a_m - a_1)z^{-m} + z^{-(m+1)} \end{aligned} \quad (\text{A.5})$$

which can be rewritten as

$$P(z) = 1 + p_1 z^{-1} + p_2 z^{-2} \dots + p_2 z^{-(m-1)} + p_1 z^{-m} + z^{-(m+1)} \quad (\text{A.6})$$

$$Q(z) = 1 + q_1 z^{-1} + q_2 z^{-2} \dots - q_2 z^{-(m-1)} - q_1 z^{-m} - z^{-(m+1)} \quad (\text{A.7})$$

where

$$p_k = a_k + a_{m-k+1} \quad \text{for } k = 1, 2 \dots \frac{m}{2} \quad (\text{A.8})$$

$$q_k = a_k - a_{m-k+1} \quad \text{for } k = 1, 2 \dots \frac{m}{2}. \quad (\text{A.9})$$

It is easily seen that $P(z)$ is a symmetric polynomial with a root at $z = -1$ and $Q(z)$ is an antisymmetric polynomial with a root at $z = 1$. These two roots are fixed and have no information about $A(z)$, so without loss of information the polynomials $P(z)$ and $Q(z)$ can be reduced from order $m + 1$ to order m by removing these roots.

$$\begin{aligned} \hat{P}(z) &= P(z)/(1 + z^{-1}) & (\text{A.10}) \\ &= 1 + \hat{p}_1 z^{-1} + \hat{p}_2 z^{-2} \dots \hat{p}_2 z^{-(m-2)} + \hat{p}_1 z^{-(m-1)} + z^{-m} \end{aligned}$$

$$\begin{aligned} \hat{Q}(z) &= Q(z)/(1 - z^{-1}) & (\text{A.11}) \\ &= 1 + \hat{q}_1 z^{-1} + \hat{q}_2 z^{-2} \dots \hat{q}_2 z^{-(m-2)} + \hat{q}_1 z^{-(m-1)} + z^{-m}. \end{aligned}$$

The reduced polynomials are now *both* symmetric and the coefficients are

$$\hat{p}_k = (-1)^k + \sum_{j=1}^k p_k (-1)^{n-k} \quad (\text{A.12})$$

$$\hat{q}_k = 1 + \sum_{j=1}^k p_k. \quad (\text{A.13})$$

Since the roots of $P(z)$ and $Q(z)$ lie on the unit circle, $\hat{P}(z)$ and $\hat{Q}(z)$ can be evaluated at $z = e^{j\omega}$. Making this substitution, factoring out $e^{-j\frac{m}{2}\omega}$, and grouping terms with like coefficients yields

$$\begin{aligned}
\hat{P}(z)|_{z=e^{j\omega}} &= e^{-j\frac{m}{2}\omega} \left[(e^{j\frac{m}{2}\omega} + e^{-j\frac{m}{2}\omega}) + \hat{p}_1(e^{j\frac{m-2}{2}\omega} + e^{-j\frac{m-2}{2}\omega}) + \right. \\
&\quad \left. \hat{p}_2(e^{j\frac{m-4}{2}\omega} + e^{-j\frac{m-4}{2}\omega}) + \dots + \hat{p}_{m/2} \right] \\
&= 2e^{-j\frac{m}{2}\omega} \left[\cos\left(\frac{m}{2}\omega\right) + \hat{p}_1 \cos\left(\frac{m-2}{2}\omega\right) + \right. \\
&\quad \left. \hat{p}_2 \cos\left(\frac{m-4}{2}\omega\right) + \dots + \frac{\hat{p}_{m/2}}{2} \right] \\
&= 2e^{-j\frac{m}{2}\omega} P'(\omega)
\end{aligned} \tag{A.14}$$

$$\begin{aligned}
\hat{Q}(z)|_{z=e^{j\omega}} &= e^{-j\frac{m}{2}\omega} \left[(e^{j\frac{m}{2}\omega} + e^{-j\frac{m}{2}\omega}) + \hat{q}_1(e^{j\frac{m-2}{2}\omega} + e^{-j\frac{m-2}{2}\omega}) + \right. \\
&\quad \left. \hat{q}_2(e^{j\frac{m-4}{2}\omega} + e^{-j\frac{m-4}{2}\omega}) + \dots + \hat{q}_{m/2} \right] \\
&= 2e^{-j\frac{m}{2}\omega} \left[\cos\left(\frac{m}{2}\omega\right) + \hat{q}_1 \cos\left(\frac{m-2}{2}\omega\right) + \right. \\
&\quad \left. \hat{q}_2 \cos\left(\frac{m-4}{2}\omega\right) + \dots + \frac{\hat{q}_{m/2}}{2} \right] \\
&= 2e^{-j\frac{m}{2}\omega} Q'(\omega).
\end{aligned} \tag{A.15}$$

The equations for $P'(\omega)$ and $Q'(\omega)$ can be solved in general for order m by substituting $x = \cos \omega$ and forming a Chebyshev series as in [15]. However, for $m \leq 8$ a closed form solution can be formulated. For the particular case $m = 8$

$$P'(\omega) = \cos 4\omega + \hat{p}_1 \cos 3\omega + \hat{p}_2 \cos 2\omega + \hat{p}_3 \cos \omega + \frac{\hat{p}_4}{2} \tag{A.16}$$

$$Q'(\omega) = \cos 4\omega + \hat{q}_1 \cos 3\omega + \hat{q}_2 \cos 2\omega + \hat{q}_3 \cos \omega + \frac{\hat{q}_4}{2}. \tag{A.17}$$

These equations are converted to polynomials in $\cos \omega$ as follows. The third term in (A.16) and (A.17) is reduced by applying the trigonometric identity $\cos 2\omega = 2 \cos^2 \omega - 1$, and the first term is reduced by multiple applications of the identity. The second term is reduced as

$$\begin{aligned}
\cos 3\omega &= \cos 2\omega \cos \omega - \sin 2\omega \sin \omega & (A.18) \\
&= (2 \cos^2 \omega - 1) \cos \omega - 2 \sin \omega \cos \omega \sin \omega \\
&= 2 \cos^3 \omega - \cos \omega - 2 \sin^2 \omega \cos \omega \\
&= 2 \cos^3 \omega - \cos \omega - 2(1 - \cos^2 \omega) \cos \omega \\
&= 4 \cos^3 \omega - 3 \cos \omega.
\end{aligned}$$

With $P'(\omega)$ and $Q'(\omega)$ written in terms of $\cos \omega$ and like terms collected, the resulting polynomials are

$$P'(\omega) = 8 \cos^4 \omega + 4\hat{p}_1 \cos^3 \omega + (2\hat{p}_2 - 8) \cos^2 \omega + (\hat{p}_3 - 3\hat{p}_1) \cos \omega + 1 + \frac{\hat{p}_4}{2} - \hat{p}_2 \quad (A.19)$$

$$Q'(\omega) = 8 \cos^4 \omega + 4\hat{q}_1 \cos^3 \omega + (2\hat{q}_2 - 8) \cos^2 \omega + (\hat{q}_3 - 3\hat{q}_1) \cos \omega + 1 + \frac{\hat{q}_4}{2} - \hat{q}_2. \quad (A.20)$$

The roots are found in closed form by substituting $x = \cos \omega$ and using Ferrari's method [26] to solve for x . The LSFs are

$$f_i = \frac{\omega_i}{2\pi} \quad (A.21)$$

where ω_i are the zeros of $P'(\omega)$ and $Q'(\omega)$. The coefficients \hat{p}_k are computed from (A.8) and (A.12) and the coefficients \hat{q}_k are computed from (A.9) and (A.13).

Appendix B

Perceptually Weighted Spectral Distortion

It was found in [21] that perceptually weighting the L_2 norm of the log spectral distance significantly increases the correlation of the distortion measure with subjective measures of speech quality. The weighted spectral distortion measure is

$$SD_w(i) = \left[\frac{1}{\pi} \int_0^\pi W(\omega) [10 \log_{10} P_i(\omega) - 10 \log_{10} \hat{P}_i(\omega)]^2 d\omega \right]^{1/2} \quad (\text{B.1})$$

where $P_i(\omega)$ is the power spectrum of the all-pole model on frame i , $\hat{P}_i(\omega)$ is the power spectrum of the quantized all-pole model on frame i and $W(\omega)$ is the perceptually based weighting function. In [21] the spectral envelope was divided into 6 bands ranging from 200 to 3400 Hertz using the frequencies in Table B.1. The speech bandwidth of STC is approximately 80 to 3800 Hertz and an appropriate weighting function for this bandwidth is needed.

A set of Bark scale critical band filters (after [27]) was used to divide the spectrum

Band	Bandwidth in Hertz
1	200-400
2	400-800
3	800-1300
4	1300-1900
5	1900-2600
6	2600-3400

Table B.1: Frequency bands used to perceptually weight the log spectral distance measure in [21].

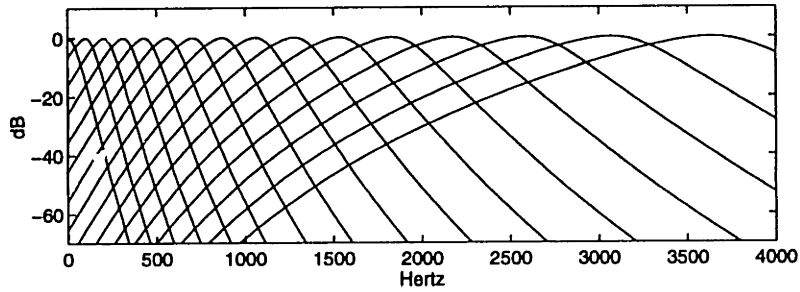


Figure B-1: Bark scale critical band filters.

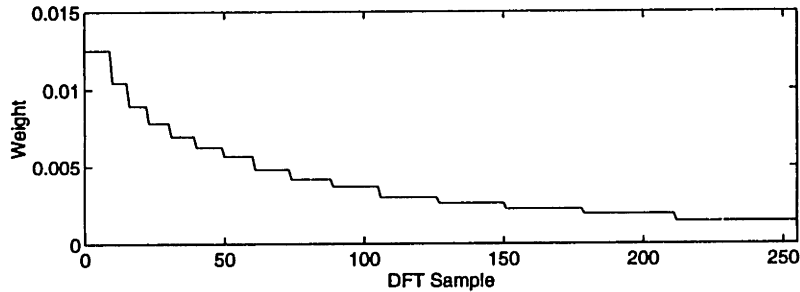


Figure B-2: Perceptually based spectral weighting function.

into 15 perceptually spaced bands. The band edge frequencies correspond to the intersections of adjacent filters which occur at their 3 dB points (Figure B-1.) The weighting function $W(\omega)$ is constant within each band (Figure B-2), and the average distortion in each band is weighted equally with respect to the other bands (i.e the area under $W(\omega)$ is the same for each band.) The exception to this is the first band which is derived from a filter centered at 0 Hz. This band has less than half the bandwidth of any other band so in this band $W(\omega)$ is arbitrarily set to be the same as in the adjacent band.

Appendix C

Significance Testing

A significance test is needed to evaluate the results of the two-alternative forced-choice listening tests. One measure of the significance of a test result is to determine the likelihood of the particular test outcome under the hypothesis H_0 that the listener cannot tell the difference between system \mathcal{A} and system \mathcal{B} (i.e. the probability of responding \mathcal{A} on any given trial is $\frac{1}{2}$.) This likelihood is the probability of the listener responding at least n times in favor of system \mathcal{A} out of N trials given that hypothesis H_0 is true [28]. The probability of exactly m responses in favor of system \mathcal{A} given H_0 is

$$p(m|H_0) = \binom{N}{m} \left(\frac{1}{2}\right)^m \left(1 - \frac{1}{2}\right)^{N-m} \quad (\text{C.1})$$

$$= \frac{N!}{m!(N-m)!} \left(\frac{1}{2}\right)^N. \quad (\text{C.2})$$

The probability of at least n responses in favor of system \mathcal{A} given H_0 is then

$$p(m \geq n|H_0) = \sum_{k=n}^N \frac{N!}{k!(N-k)!} \left(\frac{1}{2}\right)^N. \quad (\text{C.3})$$

The likelihood of n responses out of 24 in favor of system \mathcal{A} given H_0 is plotted in Figure C-1. This shows that there must be at least 18 responses in favor of system \mathcal{A} before there is less than 0.05 probability of observing that score by chance (i.e. if the listener does *not* prefer system \mathcal{A}).

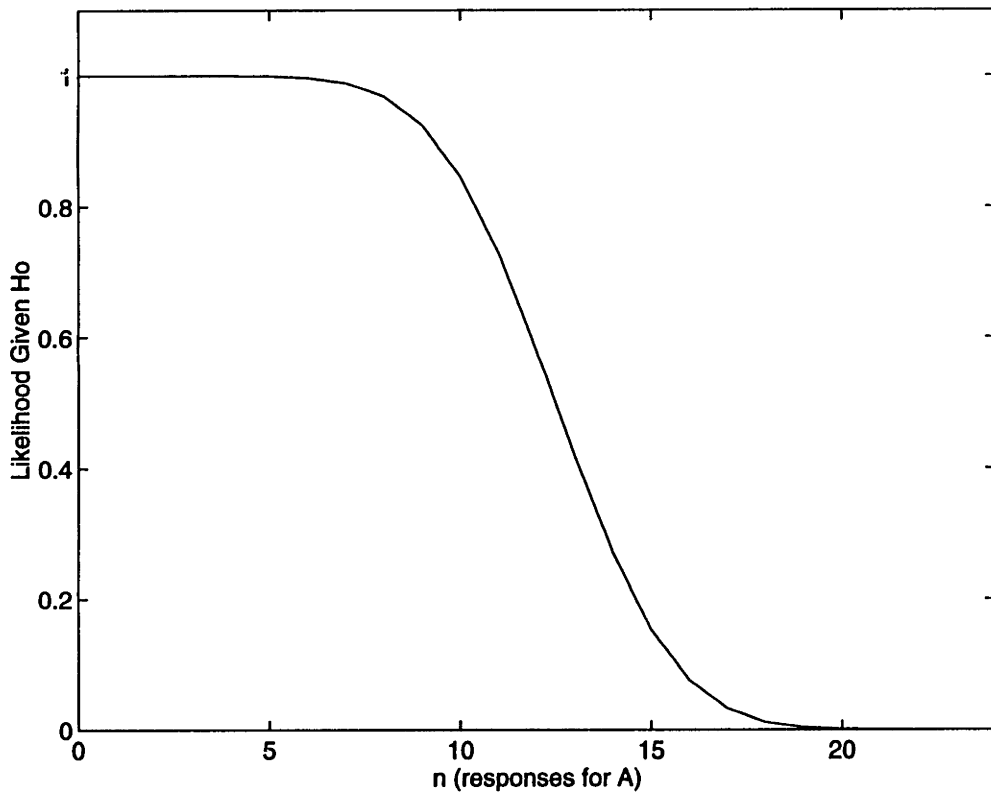


Figure C-1: Likelihood of n responses out of 24 in favor of system \mathcal{A} given H_0 .

Bibliography

- [1] Terrence G. Champion, Robert J. McAulay, and Thomas F. Quatieri. High-order allpole modelling of the spectral envelope. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages I-529-532, Adelaide, South Australia, April 1994.
- [2] Robert J. McAulay, Terrence G. Champion, and Thomas F. Quatieri. Sine-wave amplitude coding using high-order allpole models. In *Signal Processing VII: Theories and Applications*, pages 395-398, Edinburgh, Scotland, U.K., September 1994. European Association for Signal Processing.
- [3] Robert J. McAulay and Thomas F. Quatieri. *Advances in Speech Signal Processing*, chapter Low Rate Speech Coding Based on the Sinusoidal Model, pages 165-208. Marcel Dekker, Inc., New York, NY, 1992. Editors Sadaoki Furui and M. Mohan Sondhi.
- [4] Cheung-Fat Chan. Efficient quantization of LPC parameters using a mixed LSP/PARCOR representation. In *Signal Processing VII: Theories and Applications*, pages 939-942, Edinburgh, Scotland, U.K., September 1994. European Association for Signal Processing.
- [5] Robert J. McAulay and Thomas F. Quatieri. *Speech Coding and Synthesis*, chapter Sinusoidal Coding. Elsevier. Editors W. B. Kleijn and K. K. Paliwal (to be published).
- [6] Robert J. McAulay and Thomas F. Quatieri. Pitch estimation and voicing detection based on a sinusoidal speech model. In *Proc. IEEE International Conference*

- on Acoustics, Speech, and Signal Processing*, pages 249–252, Albuquerque, New Mexico, April 1990.
- [7] Douglas B. Paul. The spectral envelope estimation vocoder. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-29(4):786–794, August 1981.
- [8] John Makhoul. Linear prediction: A tutorial review. *Proceedings of the IEEE*, 63(4):561–580, April 1975.
- [9] J. D. Markel and A. H. Gray Jr. *Linear Prediction of Speech*, pages 5–15. Springer-Verlag, New York, New York, 1976.
- [10] R. Viswanathan and John Makhoul. Quantization properties of transmission parameters in linear predictive systems. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, ASSP-23(3):309–321, June 1975.
- [11] John Makhoul. Stable and efficient lattice methods for linear prediction. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 25(5):423–428, October 1977.
- [12] Noboru Sugamura and Fumitada Itakura. Speech analysis and synthesis methods developed at ECL in NTT. *Speech Communication*, 5:199–215, 1986.
- [13] Frank K. Soong and Biing-Hwang Juang. Line spectrum pair LSP and speech data compression. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 1.10.1–1.10.4, New York, NY, April 1984.
- [14] Leonard Eugene Dickson. *New First Course in the Theory of Equations*, page 42. John Wiley & Sons, Inc., New York, New York, 1939.
- [15] Peter Kabal and Ravi Prakash Ramchandran. The computation of line spectral frequencies using chebyshev polynomials. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 34(6):1419–1426, December 1986.

- [16] G. S. Kang and L. J. Fransen. Low-bit rate speech encoders based on line-spectrum frequencies (LSFs). NRL Report 8857, Naval Research Laboratory, Washington, D.C., January 1985.
- [17] Al Lovrich. *Digital Signal Processing Applications with the TMS320C30 Family: Theory, Algorithms, and Implementations*, chapter Doublelength Floating Point Arithmetic on the TMS320C30, pages 137–168. Prentice Hall, Englewood Cliffs, New Jersey, 1990.
- [18] Augustine H. Gray Jr. and John D. Markel. Distance measures for speech processing. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 24(5):380–391, October 1976.
- [19] Bishnu S. Atal, Richard V. Cox, and Peter Kroon. Spectral quantization and interpolation for CELP coders. In *Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages II-69–72, Glasgow, Scotland, U.K., May 1989.
- [20] Kuldip K. Paliwal and Bishnu S. Atal. Efficient vector quantization of LPC parameters at 24 bits/frame. *IEEE Transactions on Speech and Audio Processing*, 1(1):3–14, January 1993.
- [21] Schuyler R. Quackenbush, Thomas P. Barnwell III, and Mark A. Clements. *Objective Measures of Speech Quality*. Prentice Hall Signal Processing Series. Prentice Hall, Englewood Cliffs, New Jersey, 1988.
- [22] Frank K. Soong and Biing-Hwang Juang. Optimal quantization of LSP parameters. *IEEE Transactions on Speech and Audio Processing*, 1(1):15–24, January 1993.
- [23] William M. Fisher, George R. Doddington, and Kathleen M. Goudie-Marshall. The DARPA speech recognition research database: Specifications and status. In *Proc. DARPA Workshop on Speech Recognition*, pages 93–99, Palo Alto, California, February 1986.

- [24] Allen Gersho and Robert M. Gray. *Vector Quantization and Signal Compression*, pages 179–180. Kluwer Academic Publishers, Boston, Massachusetts, 1992.
- [25] Y. Linde, A. Buzo, and R.M. Gray. An algorithm for vector quantizer design. *IEEE Transactions on Communications*, COM-28:84–95, January 1980.
- [26] Jean-Pierre Tignol. *Galois' theory of algebraic equations*, pages 31–35. Longman Scientific & Technical, Essex, England, 1988.
- [27] Shihua Wang, Andrew Sekey, and Allen Gersho. An objective measure for predicting subjective quality of speech coders. *IEEE Journal on Selected Areas in Communications*, 10(5):819–829, June 1992.
- [28] Richard J. Larsen and Morris L. Marx. *Mathematical Statistics and Its Applications*, pages 287–293. Prentice-Hall, Englewood Cliffs, New Jersey, 1986.