SRC TR 89-1

Efficient Encoding of Speech LSP Parameters Using the Discrete Cosine Transformation

by

N. Farvardin and
R. Laroia

EFFICIENT ENCODING OF SPEECH LSP PARAMETERS USING THE DISCRETE COSINE TRANSFORMATION †

Nariman Farvardin and Rajiv Laroia
Electrical Engineering Department,
Institute for Advanced Computer Studies
and Systems Research Center
University of Maryland
College Park, Maryland 20742

Abstract

In this paper, the intraframe and interframe correlation properties are used to develop two efficient encoding algorithms for speech line spectrum pair (LSP) parameters. The first algorithm (2-D DCT), which requires relatively large coding delays, is based on two-dimensional (time and frequency) discrete cosine transform coding techniques; the second algorithm (DCT-DPCM), which does not need any coding delay, uses one-dimensional discrete cosine transform in the frequency domain and DPCM in the time domain. The performance of these systems for different bit rates and delays are studied and appropriate comparisons are made. It is shown that an average spectral distortion of approximately 1 dB^2 can be achieved with 21 and 25 bits/frame using the 2-D DCT and DCT-DPCM schemes, respectively. This is a noticeable improvement over the previously reported bit rates of 32 bits/frame and above [1], [2].

† This work was supported in part by National Science Foundation grants NSFD MIP-86-57311 and NSFD CDR-85-00108, and in part by grants from Martin Marietta Laboratories and General Electric.

1. Introduction

While various methods for speech analysis-synthesis are known [3], the line spectrum pair (LSP) method, first introduced and studied by Itakura and Sugamura [4], is generally accepted as one of the most efficient speech analysis-synthesis techniques. The LSP parameters, which represent the short-term speech spectrum, are completely equivalent, in a mathematical sense, to other linear predictive coding coefficients such as the LPC or PARCOR coefficients [3]. However, the LSP parameter have some properties which make them more attractive than the LPC and PARCOR coefficients. The increased use of LSP parameters in speech coding applications has spurred a significant research activity in developing more efficient quantization algorithms [5], [6], [1], [2].

In this paper, we will develop two systems for encoding the LSP parameters; in both systems, we attempt to utilize the correlation between the LSP parameters within a frame and between frames in an effort to reduce the average number of bits/parameter for a given level of quantization distortion. Similar ideas have been used in encoding the harmonic magnitudes in [11]. The first scheme is based on the two-dimensional discrete cosine transform coding ideas that have been used successfully in image coding applications [7]. The second system is a hybrid of one-dimensional discrete cosine transform (to reduce the intraframe correlation) and DPCM in the time domain (to reduce the interframe correlation). For both schemes, we have developed an algorithm for reordering of quantized LSP parameters which is needed to ensure the stability of the synthesis filter; we have shown that this algorithm can only reduce the mean-squared quantization error.

For the spectral distortion measure, we have studied the performance of the above schemes and have made appropriate comparisons against other quantization algorithms. Our results indicate that using the two-dimensional transform coding scheme with a delay of 100 msec, an average spectral distortion of 1 dB^2 can be obtained with only 21 bits/frame.

The rest of this paper is organized as follows. In Section 2, the LSP parameters are described and empirical evidence on the intraframe and interframe correlation of LSP parameters is provided. In Section 3 the two-dimensional discrete cosine transform coding of the LSP parameters is described, followed by Section 4 in which the a hybrid system using the discrete cosine transform and DPCM is presented. In Section 5, performance results of different systems are presented and comparisons are made. Finally, in Section 6 a summary and suggestions for further research is provided.

2. LSP Parameters

For a given order p, the linear predictive analysis of speech results in an all-pole filter $H(z) = 1/A_p(z)$, where

$$A_p(z) = 1 + \alpha_1 z^{-1} + \alpha_2 z^{-2} + \cdots + \alpha_p z^{-p} \ . \tag{1}$$

The parameters $\{\alpha_i\}_{i=1,2,\ldots,p}$, are commonly referred to as the LPC coefficients. It is easy to verify that the polynomial $A_j(z)$, associated with a jth-order LPC analysis, satisfies the following recurrence relationship [3]

$$A_j(z) = A_{j-1}(z) - k_j z^{-j} A_{j-1}(z^{-1}),\ j = 1, 2, \ldots, p, \tag{2}$$

where $A_0(z) = 1$. In (2), the parameters $\{k_j\}_{j=1,2,\ldots,p}$, are called the PARCOR coefficients. For $j = p+1$ in (2), we have

$$A_{p+1}(z) = A_p(z) - k_{p+1} z^{-(p+1)} A_p(z^{-1}) \ . \tag{3}$$

In (3) consider two extreme artificial boundary conditions: $k_{p+1} = 1$ and $k_{p+1} = -1$. These conditions correspond, respectively, to a *complete closure* and a *complete opening* at the glottis in the acoustic tube model [3]. Under these conditions, the polynomial $A_{p+1}(z)$ can be expressed as

$$P(z) = (1 - z^{-1}) \prod_{i=2,4,\ldots,p} (1 - 2z^{-1}\cos\omega_i + z^{-2}) \ , \tag{4a}$$

for $k_{p+1} = 1$, and

$$Q(z) = (1 + z^{-1}) \prod_{i=1,3,\ldots,p-1} (1 - 2z^{-1}\cos\omega_i + z^{-2}) \ , \tag{4b}$$

for $k_{p+1} = -1$, where it is assumed that $\omega_1 < \omega_3 < \cdots < \omega_{p-1}$, and $\omega_2 < \omega_4 < \cdots < \omega_p$ and that p is an even integer.

It is clear from (4) that $e^{j\omega_i}$, $i = 1, 2, \ldots,\ p$, are the roots of the polynomials $P(z)$ and $Q(z)$. The parameters $\{\omega_i\}_{i=1,2,\ldots,p}$, are defined as the *line spectrum pair* (LSP) parameters. Note that $\omega_0 = 0$ and $\omega_{p+1} = \pi$ are fixed roots of $P(z)$ and $Q(z)$, respectively. In view of the above definition, the LSP parameters can be interpreted as the resonant frequencies of the vocal tract under the two extreme artificial boundary conditions at the glottis [4]. From now on, we will use $\mathbf{\Omega}_n = [\Omega_{n,1}, \Omega_{n,2}, \ldots,\ \Omega_{n,p}]^T$ to denote the p-dimensional random vector of LSP parameters associated with the nth frame of speech.

The polynomials $P(z)$ and $Q(z)$ possess some interesting and important properties summarized in the following:

1.) All roots of $P(z)$ and $Q(z)$ lie on the unit circle.

2.) The roots of $P(z)$ and $Q(z)$ alternate on the unit circle. Specifically, the following relationship is satisfied for all n:

$$0 = \Omega_{n,0} < \Omega_{n,1} < \cdots < \Omega_{n,p-1} < \Omega_{n,p} < \Omega_{n,p+1} = \pi\ . \tag{5}$$

Property (2) above, hereafter referred to as the *ordering* property, indicates that the LSP parameters within a frame are correlated; an effective use of this correlation can improve the quantization of LSP parameters. Examples are the differential quantization scheme of [5] and the adaptive quantization of [2]. To demonstrate the *intraframe* correlation property of the LSP parameters, we have used a data base of speech samples (described in Table 1) to compute the correlation coefficient between $\Omega_{n,i}$ and $\Omega_{n,j}$, denoted by $\phi_{i,j},\ i,j\ = 1, 2, \ldots, p$. The matrix $\Phi = [\phi_{i,j}]$, presented in Table 2, indicates a relatively high correlation between neighboring LSP parameters within a frame.

Similarly, to investigate the *interframe* correlation of the LSP parameters, we have also computed the correlation coefficient between $\Omega_{n,i}$ and $\Omega_{n-k,i}$, denoted by $\psi_{i,k}$. These results are presented in Table 3 for $k = 1, 2, \ldots,\ 10$, indicating a strong correlation between the LSP parameters of the same order in neighboring frames.

The results in Tables 2 and 3 indicate that there is a strong correlation between the LSP parameters of adjacent frames and neighboring parameters in the same frame. Therefore, any compression algorithm that effectively utilizes this correlation can result in improved performance over those that do not use this correlation. One such technique that has been used successfully in the context of image coding is the two-dimensional (2-D) discrete cosine transform (DCT) coding. In what follows we will describe the 2-D DCT coding of the LSP parameters.

3. 2-D DCT Coding of LSP Frames

For notational convenience, let us consider the $p \times L$ matrix $\mathbf{X} = [X_{i,j}]$, where $X_{i,j} \triangleq \Omega_{j,i+1}$, $i \in J_0^{p-1}$, $j \in J_0^{L-1}$, where $J_i^j \triangleq \{k : i \leq k \leq j\}$. Here, $\mathbf{X}$ is a *block* of L successive frames of LSP vectors. The 2-D DCT of the matrix $\mathbf{X}$, denoted by the $p \times L$ matrix $\mathbf{Y} = [Y_{k,l}]$, is defined, for $k \in J_0^{p-1}$, $l \in J_0^{L-1}$, according to

$$Y_{k,l} = \frac{2}{\sqrt{pL}} C_k C_l \sum_{i=0}^{p-1} \sum_{j=0}^{L-1} X_{i,j} \cos\frac{(2i+1)k\pi}{2p} \cos\frac{(2j+1)l\pi}{2L}, \tag{6}$$

where $C_0 = 1/\sqrt{2}$ and $C_k = 1,\ k \in J_1^{\max(p,L)-1}$. Here, $Y_{k,l}$ is called the (k,l)th transform coefficient. Upon performing the transformation, the transform coefficients are quantized by means of scalar quantizers and encoded separately. Suppose $\hat{Y}_{k,l}$ denotes the quantized version of $Y_{k,l}$. Then, the inverse transform (2-D IDCT) of these quantized coefficients is described by

$$\hat{X}_{i,j} = \frac{2}{\sqrt{pL}} \sum_{k=0}^{p-1} \sum_{l=0}^{L-1} C_k C_l \hat{Y}_{k,l} \cos\frac{(2i+1)k\pi}{2p} \cos\frac{(2j+1)l\pi}{2L}, \tag{7}$$

for $i \in J_0^{p-1}$, $j \in J_0^{L-1}$.

The 2-D DCT, which has been used successfully in image compression applications, performs on a 2-D array of highly correlated variables (as demonstrated in Tables 2 and 3), and produces a set of nearly *uncorrelated* transform coefficients which can then be transmitted more efficiently [7].

In encoding the LSP blocks, we will confine our attention to minimizing the mean squared-error (MSE) between the block $\mathbf{X}$ and its replica $\hat{\mathbf{X}} = [\hat{X}_{i,j}]$. Since the 2-D DCT is a *unitary* transformation [7], the MSE can be expressed as

$$D = \frac{1}{pL} E(\| \mathbf{X} - \hat{\mathbf{X}} \|^2) = \frac{1}{pL} E(\| \mathbf{Y} - \hat{\mathbf{Y}} \|^2), \tag{8}$$

where $\hat{\mathbf{Y}}$ is the matrix of quantized transform coefficients. This relationship implies that to minimize the MSE between $\mathbf{X}$ and $\hat{\mathbf{X}}$, it suffices to minimize the MSE between $\mathbf{Y}$ and $\hat{\mathbf{Y}}$. The term on the right hand side of (8), in turn, can be expressed as

$$D = \frac{1}{pL} \sum_{k=0}^{p-1} \sum_{l=0}^{L-1} \lambda_{k,l} d_{k,l}(b_{k,l}), \tag{9a}$$

where $\lambda_{k,l}$ is the variance of $Y_{k,l}$ and $d_{k,l}(b_{k,l})$ is the variance-normalized minimum MSE incurred in quantizing $Y_{k,l}$ with $b_{k,l}$ bits.

Therefore, the problem of minimizing the MSE is that of determining the bit assignment map $\mathbf{b} = [b_{k,l}]$, in such a way as to minimize the expression in (9a) while satisfying the average number of bits per LSP coefficient, b_{av}, according to

$$\frac{1}{pL} \sum_{k=0}^{p-1} \sum_{l=0}^{L-1} b_{k,l} \leq b_{av}. \tag{9b}$$

This is a standard bit allocation problem discussed in [7], [9], [10]. A related issue which is of critical importance is that of the stability of the synthesis filter. Recall from (5) that the LSP parameters satisfy an ordering property which is the necessary and sufficient condition for the stability of the synthesis filter [4]. Thus, for the synthesis filter to be stable, the quantized LSP parameters $\{\hat{\Omega}_{n,i},\ i \in J_1^p\}$ must also satisfy the ordering property. Since the above quantization procedure uses scalar quantizers that operate on the transform coefficients separately, it is quite possible that the ordering property is violated by this procedure. In what follows we establish a technique that reorders the quantized LSP parameters to satisfy the ordering property without increasing the distortion.

Let the mapping $f : \mathbb{R}^p \to \mathbb{R}^p$ be defined such that for any arbitrary vector $\mathbf{u} \in \mathbb{R}^p$, $\mathbf{v} = f(\mathbf{u})$ be the sorted version of $\mathbf{u}$, such that $v_1 \leq v_2 \leq \cdots \leq v_p$.

Theorem: Let $\mathbf{u}$ be an arbitrary vector in $\mathbb{R}^p$. Let $\mathbf{w}$ be a vector in $\mathbb{R}^p$ such that $w_1 \leq w_2 \leq \cdots \leq w_p$. Then, if $\mathbf{v} = f(\mathbf{u})$, we have $\| \mathbf{w} - \mathbf{v} \| \leq \| \mathbf{w} - \mathbf{u} \|$.

Proof: The vector $\mathbf{v} = f(\mathbf{u})$ can be obtained from $\mathbf{u}$ by a finite number of switchings of two components as described in the following steps:

0. Set $k = 1$; $\mathbf{u}^{(0)} = \mathbf{u}$.
1. Identify the kth smallest component of $\mathbf{u}^{(k-1)}$, say, $u_{j_k}^{(k-1)}$, and interchange the kth and j_kth elements of $\mathbf{u}^{(k-1)}$ to generate a new vector $\mathbf{u}^{(k)}$.
2. If $k = p-1$, stop; else, set $k = k+1$ and go to (1).

From the above algorithm, it is clear that $\mathbf{v} = f(\mathbf{u}) = \mathbf{u}^{(p-1)}$. Now we will show that $\| \mathbf{w} - \mathbf{u}^{(k)} \| \leq \| \mathbf{w} - \mathbf{u}^{(k-1)} \|$, $\forall k \in J_1^{p-1}$. We have

$$A \triangleq \| \mathbf{w} - \mathbf{u}^{(k)} \|^2 - \| \mathbf{w} - \mathbf{u}^{(k-1)} \|^2$$

$$= w_k(u_k^{(k-1)} - u_k^{(k)}) + w_{j_k}(u_{j_k}^{(k-1)} - u_{j_k}^{(k)}). \tag{10}$$

But, $u_k^{(k)} = u_{j_k}^{(k-1)}$ and $u_{j_k}^{(k)} = u_k^{(k-1)}$. Therefore, (10) reduces to

$$A = (w_k - w_{j_k})(u_k^{(k-1)} - u_{j_k}^{(k-1)}). \tag{11}$$

Notice that by the definition of j_k, we always have $j_k \geq k$ which, in turn, implies $w_k \leq w_{j_k}$ and $u_k^{(k-1)} \geq u_{j_k}^{(k-1)}$; hence $A \leq 0$, which proves the theorem.

Let us suppose that $\hat{\mathbf{\Omega}}_n$, the quantized version of $\mathbf{\Omega}_n$, violates the ordering property in (5), i.e., $\hat{\mathbf{\Omega}}_n \neq f(\hat{\mathbf{\Omega}}_n)$. If $\tilde{\mathbf{\Omega}}_n = f(\hat{\mathbf{\Omega}}_n)$, then, based on the above theorem, not only does $\tilde{\mathbf{\Omega}}_n$ satisfy the ordering property, but $\| \mathbf{\Omega}_n - \tilde{\mathbf{\Omega}}_n \| \leq \| \mathbf{\Omega}_n - \hat{\mathbf{\Omega}}_n \|$, i.e., it results in a smaller Euclidean distance from $\mathbf{\Omega}_n$.

Remark 1: In quantizing the transform vectors we have not used the additional information that the $\mathbf{\Omega}_n$'s satisfy the ordering property; performing f on $\hat{\mathbf{\Omega}}_n$ enables us to utilize this knowledge to further reduce the average quantization error.

Remark 2: Using the same notation as in the Theorem, if $\mathbf{u}$ violates the ordering property, then, in general, there are uncountably many vectors which can be obtained from $\mathbf{u}$ such that they satisfy the ordering property while reducing the Euclidean distance from $\mathbf{w}$; $\mathbf{v} = f(\mathbf{u})$ is only one such vector. For example, it is possible to show that for any such $\mathbf{u}$, there exists an $\alpha_{\mathbf{u}} > 0$ such that for $0 \leq \alpha < \alpha_{\mathbf{u}}$, $\mathbf{v}_\alpha \triangleq \alpha\mathbf{u} + (1-\alpha)\mathbf{v}$ satisfies the ordering property and $\| \mathbf{v}_\alpha - \mathbf{w} \| \leq \| \mathbf{u} - \mathbf{w} \|$.

The observation made in Remark 2 suggests that for a quantized LSP vector that violates the ordering property, there is a continuum of vectors which satisfy the ordering property and which are closer to the original LSP vector. While an extra level of optimization can be used to determine the *best* choice of this vector, in this work we use $\tilde{\mathbf{\Omega}}_n = f(\hat{\mathbf{\Omega}}_n)$ as the final reconstructed vector.

The block diagram of the 2-D DCT coding of the LSP parameters is illustrated in Fig. 1. Details on the quantization and bit assignment issues are provided in Section 5.

Consider a 2-D DCT coding scheme which encodes L LSP frames at a time. This system requires an encoding delay of LT_f where T_f is the frame period (10 msec in our studies). Clearly, the system performance gets better by increasing L. While in one-way communication situations (e.g., store-and-forward) the delay associated with large values of L does not cause a major difficulty, in two-way communication situations this encoding delay is undesirable and should be minimized. As an alternative system which utilizes the intraframe and interframe correlation of LSP parameters without introducing large delays we have considered a hybrid DCT-DPCM system which we will briefly describe in the following.

4. Hybrid DCT-DPCM Coding of LSP Vectors

In the hybrid DCT-DPCM scheme, we perform a one-dimensional (1-D) DCT on the LSP parameters associated with each frame (eq. (6) with $L = 1$) in an effort to reduce the intraframe correlation. This is the followed by DPCM coding of the resultant DCT coefficients to utilize the interframe correlation between the DCT coefficients of successive frames. Specifically, a separate DPCM encoder is designed for each of the p DCT coefficients. The justification for using DPCM on the 1-D DCT coefficients resides in the strong correlation between the DCT coefficients of neighboring frames. The correlation coefficient between the ith DCT coefficients of two neighboring frames, denoted by θ_i, is presented in Table 4.

The operation of the DCT-DPCM system is very similar to that illustrated in Fig. 1 with the exception that $L = 1$ and the bank of scalar quantizers is replaced by a bank of DPCM encoders. At the very end, the mapping f is still needed to ensure the stability of the synthesis filter.

5. Performance Results

In this section we will present simulation results to demonstrate the efficacy of the schemes proposed in Sections 3 and 4. In all of these results the data base and the experimental conditions are as described in Table 1, unless stated otherwise.

While the squared-error distortion criterion is used for the design of the quantizers, we use the spectral distortion as a measure of system performance. The average spectral distortion is defined according to

$$D_s = \frac{1}{N_f}\sum_{n=1}^{N_f}\left(\frac{100}{\pi}\int_0^{\pi}(\log S_n(\omega) - \log\hat{S}_n(\omega))^2 d\omega\right), \quad (\text{dB}^2), \tag{12}$$

where $S_n(\omega)$ and $\hat{S}_n(\omega)$ are the spectra of the nth speech frame without quantization and with quantization, respectively, and N_f is the total number of frames. The spectral distortion measure is known to have a good correspondence with subjective measures [8].

A. 2-D DCT

We have obtained simulation results for the 2-D DCT coding scheme of Section 3 for bit rates of b_{av} =1.5, 2.0, 2.5, 3.0, 3.5 and 4.0 bits/parameter for $L = 1, 2, \ldots, 10$. In obtaining these results we have assumed that the DCT coefficients of the LSP parameters have a Gaussian distribution. Experimental study of the distribution of these parameters has indicated that the Gaussian assumption is reasonable. The bit assignment algorithm used is based on the steepest descent algorithm described in [9] and modified in [10]. An example of the bit allocation matrix is tabulated in Table 5 for the case with $L = 10$ and $b_{av} = 2.0$ bits/parameter. The large variation between the number of bits assigned to different DCT coefficients is again another indication of strong correlation among LSP parameters. Notice that the $b_{k,l}$'s drop more rapidly in the l direction indicating a stronger interframe correlation as compared to intraframe correlation.

The average spectral distortions for different values of L and bit rates are tabulated in Table 6. For comparison purposes, we have obtained performance results for two additional systems: (i) the system using uniform quantization of the LSP parameters (UQ) as described in [2] and the adaptive quantization in the backward direction (AQBW) also described in [2]. These results are also included in Table 6.

It can be seen from these results that the 2-D DCT coding scheme offers substantial performance improvements compared to the UQ and AQBW schemes. Generally speaking, the performance improves as L gets larger. It is interesting to note that an average spectral distortion of 1 dB^2 (generally accepted as the perceptually significant difference limen) can be obtained with $b_{av} = 2.5$ bits/parameter and $L = 4$. Further, if we are willing to tolerate larger delays, the same distortion can be achieved with $L = 10$ at $b_{av} = 2.1$ bits/parameter.

B. Hybrid DCT-DPCM

For the hybrid DCT-DPCM scheme, we have obtained performance results at the same rates as for the 2-D DCT scheme; these results are also included in Table 6. Here, the quantizers for the DPCM encoders are designed based on a training sequence of prediction errors. Specifically, if $Y_{n,i}$ is used to denote the ith DCT coefficient of the nth frame and $\bar{Y}_i$ is its mean value, then the sequence $\{E_{n,i}\}$, $n = 1, 2, \ldots, N_f$, where $E_{n,i} = [Y_{n,i} - \bar{Y}_i] - \theta_i[Y_{n-1,i} - \bar{Y}_i]$ is used for the design of the quantizer associated with the ith DPCM loop. The bit assignment among the different quantizers is based on the variances of $E_{n,i}$, $i = 1, 2, \ldots, p$. Again as an example, for the case with $b_{av} = 2.5$ bits/parameter, the bit allocation vector is $\mathbf{b} = [4, 3, 3, 3, 2, 2, 2, 2, 2, 2]$.

The performance results of the DCT-DPCM scheme presented in Table 6 indicate that 1 dB^2 average spectral distortion can be achieved with 2.5 bits/parameter. At higher bit rates, the performance of this scheme is better than that of 2-D DCT with $L = 4$, while at lower rates this behavior is reversed.

An important issue in encoding the LSP parameters is that of distribution of the spectral distortion about its mean value. This distribution is important as it provides information on the percentage of frames that suffer from spectral distortions that are significantly larger than the average value. In comparing two encoding systems with the same average spectral distortion, clearly the one that has a smaller number of frames with large distortion should be preferred. For comparison purposes the histogram of the spectral distortion for the UQ (3.8 bits/parameter), AQBW (3.5 bits/parameter), DCT-DPCM (2.5 bits/parameter) and 2-D DCT with $L = 4$ (2.5 bits/parameter) schemes are presented in Fig. 2. Note that in all these cases the average spectral distortion is approximately 1 dB^2. The results in Fig. 2 indicate that the histograms of the spectral distortion in these cases are very similar; however, it must be noted that the percentage of frames with very large spectral distortions is larger in the 2-D DCT and DCT-DPCM schemes as compared to the UQ and AQBW schemes.

Finally, we have conducted a few experiments on speech input chosen from out of the training sequence. We limited ourselves to cases with $D_s \approx 1$ dB^2. In all these cases, the D_s that we obtained on out of the training data, was less then 1.3 dB^2. We feel that increasing the size of the training sequence will decrease the sesitivity of the system performance to out of the training data.

6. Summary and Conclusions

We have described two schemes for quantization of speech LSP parameters. The basic idea in developing these schemes has been to use the intraframe and interframe correlation of LSP parameters to reduce the bit rate for a given level of fidelity. The first scheme which uses 2-D DCT coding of the LSP parameters, enables us to encode the LSP parameters with a spectral distortion of 1 dB^2 at 21 bits/frame. The second scheme, which unlike the 2-D DCT does not require large encoding delays, uses a hybrid of 1-D DCT and DPCM to use the time and frequency domain correlation of LSP parameters; this system achieves the 1 dB^2 spectral distortion at 25 bits/frame.

Replacing the 2-D DCT by the 2-D Karhunen-Loeve transformation and using entropy coding to encode the quantized DCT coefficients are among the interesting ideas for further research on this subject.

References

1. F. K. Soong and B. H. Juang, "Optimal Quantization of LSP Parameters," *Proceedings of ICASSP-88*, pp. 394-397, April 1988.
2. N. Sugamura and N. Farvardin, "Quantizer Design in LSP Speech Analysis-Synthesis," *IEEE Journal on Selected Areas in Communications*, vol. 6, pp. 432-440, Feb. 1988.
3. L.R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals,* Prentice-Hall, Englewood Cliffs, NJ, 1978.
4. F. Itakura and N. Sugamura, "LSP Speech Synthesizer, Its Principle and Implementation," *Trans. of the Committee on Speech Research*, ASJ, S79-46, November, 1979, (in Japanese).
5. F. K. Soong and B. H. Juang, " Line Spectrum Pair (LSP) and Speech data Compression," *Proceedings of ICASSP-84*, pp. 1.10.1 - 1.10.4, 1984.
6. J.R. Crosmer and T.P. Barnwell, "A Low Bit Rate Segment Vocoder Based on Line Spectrum Pairs," *Proceedings of ICASSP-85*, pp. 240-243, March 1985.
7. N. S. Jayant and P. Noll, *Digital Coding of Waveforms - Principles and Applications to Speech and Video,* Prentice-Hall, Englewood Cliffs, NJ, 1984.
8. N. Kitawaki, K. Itoh and K. Kakehi, "Speech Quality Measurement Methods for Synthesized Speech," *Review of ECL, NTT*, Japan, vol. 29, no. 9-10, pp. 895-906, 1981.
9. A. V. Trushkin, " Optimal Bit Allocation Algorithm for Quantizing a Random Vector," *Journal of Problems of Information Transmission*, pp. 156-161, Jan. 1982.
10. V. Vaishampayan and N. Farvardin, "Optimal Block Cosine Transform Image Coding for Noisy Channels," *Proceedings of the Twentieth Annual Conference on Information Sciences and Systems*, pp. 260-266, Princeton, NJ, March 1986.
11. J.C. Hardwick and J.S. Lim, "A 4.8 KBPS Multi-Band Excitation Speech Coder," *Proceedings of ICASSP-88*, pp. 374-377, April 1988.

Table 1: Experimental Conditions

Speakers	2 Male and 2 Female
Sentences	4 Sentences/Speaker
Sampling Frequency	8 KHz
Frame Period	10 msec
Window	30 msec Hamming
Analysis Order	10
Number of Frames	5,445

Table 2: Intraframe Correlation Coefficients $\phi_{i,j}$

j / i	1	2	3	4	5	6	7	8	9	10
1	1.00	0.65	-0.30	-0.35	-0.41	-0.49	-0.39	-0.40	-0.36	-0.20
2	0.65	1.00	0.28	0.11	-0.07	-0.13	-0.07	-0.05	-0.06	-0.07
3	-0.30	0.28	1.00	0.72	0.50	0.53	0.46	0.54	0.39	0.28
4	-0.35	0.11	0.72	1.00	0.72	0.62	0.46	0.42	0.45	0.21
5	-0.41	-0.07	0.50	0.72	1.00	0.79	0.52	0.47	0.34	0.26
6	-0.49	-0.13	0.53	0.62	0.79	1.00	0.71	0.61	0.49	0.28
7	-0.39	-0.07	0.46	0.46	0.52	0.71	1.00	0.73	0.58	0.41
8	-0.40	-0.05	0.54	0.42	0.47	0.61	0.73	1.00	0.58	0.46
9	-0.36	-0.06	0.39	0.45	0.34	0.49	0.58	0.58	1.00	0.41
10	-0.20	-0.07	0.28	0.21	0.26	0.28	0.41	0.46	0.41	1.00

Table 3: Interframe Correlation Coefficients $\psi_{i,k}$

k / i	1	2	3	4	5	6	7	8	9	10
1	0.93	0.84	0.76	0.68	0.61	0.55	0.50	0.45	0.41	0.36
2	0.89	0.75	0.63	0.54	0.46	0.38	0.32	0.27	0.22	0.18
3	0.92	0.80	0.70	0.60	0.51	0.43	0.36	0.30	0.24	0.20
4	0.92	0.82	0.73	0.64	0.56	0.49	0.43	0.37	0.32	0.27
5	0.95	0.88	0.81	0.74	0.67	0.61	0.54	0.48	0.43	0.37
6	0.94	0.85	0.77	0.69	0.62	0.56	0.49	0.44	0.38	0.33
7	0.93	0.83	0.75	0.66	0.58	0.50	0.43	0.37	0.31	0.26
8	0.91	0.81	0.72	0.64	0.56	0.49	0.43	0.37	0.32	0.28
9	0.87	0.73	0.64	0.55	0.48	0.42	0.37	0.33	0.29	0.25
10	0.82	0.66	0.57	0.50	0.44	0.38	0.34	0.30	0.27	0.24

Table 4: Interframe Correlation of DCT Coefficients θ_i

i	0	1	2	3	4	5	6	7	8	9
θ_i	0.95	0.90	0.95	0.90	0.90	0.84	0.83	0.85	0.87	0.79

Table 5: The Bit Allocation Matrix;

$L = 10$, $b_{av} = 2.0$ bits/parameter

l	0	1	2	3	4	5	6	7	8	9
k										
0	6	5	4	3	3	2	2	2	1	1
1	5	4	4	3	2	2	2	1	1	0
2	6	5	4	3	2	2	1	1	1	0
3	5	4	3	2	2	2	1	1	1	0
4	5	4	3	2	2	1	1	1	0	0
5	4	3	3	2	2	1	1	1	0	0
6	4	3	2	2	2	1	1	1	0	0
7	4	3	3	2	2	1	1	1	0	0
8	4	3	3	2	2	1	1	1	0	0
9	4	3	2	2	1	1	1	1	0	0

Table 6: Average Spectral Distortion (dB^2) for

2-D DCT, DCT-DPCM, UQ and AQBW Schemes

b_{av}	1.5	2.0	2.5	3.0	3.5	4.0
2-D DCT						
$L = 1$	7.09	4.62	2.66	1.62	0.86	0.47
$L = 2$	3.39	2.08	1.31	0.80	0.47	0.29
$L = 3$	2.75	1.72	1.14	0.66	0.43	0.24
$L = 4$	2.34	1.56	0.98	0.59	0.36	0.21
$L = 5$	2.23	1.44	0.90	0.54	0.31	0.18
$L = 6$	2.03	1.32	0.83	0.50	0.31	0.17
$L = 7$	2.08	1.29	0.83	0.50	0.29	0.16
$L = 8$	1.97	1.22	0.80	0.46	0.27	0.15
$L = 9$	1.91	1.20	0.76	0.45	0.26	0.14
$L = 10$	1.85	1.16	0.74	0.43	0.26	0.14
DCT-DPCM	5.34	2.43	0.98	0.55	0.27	0.15
UQ	26.54	12.94	6.22	2.99	1.57	0.73
AQBW	13.70	7.05	3.90	1.73	0.95	0.44

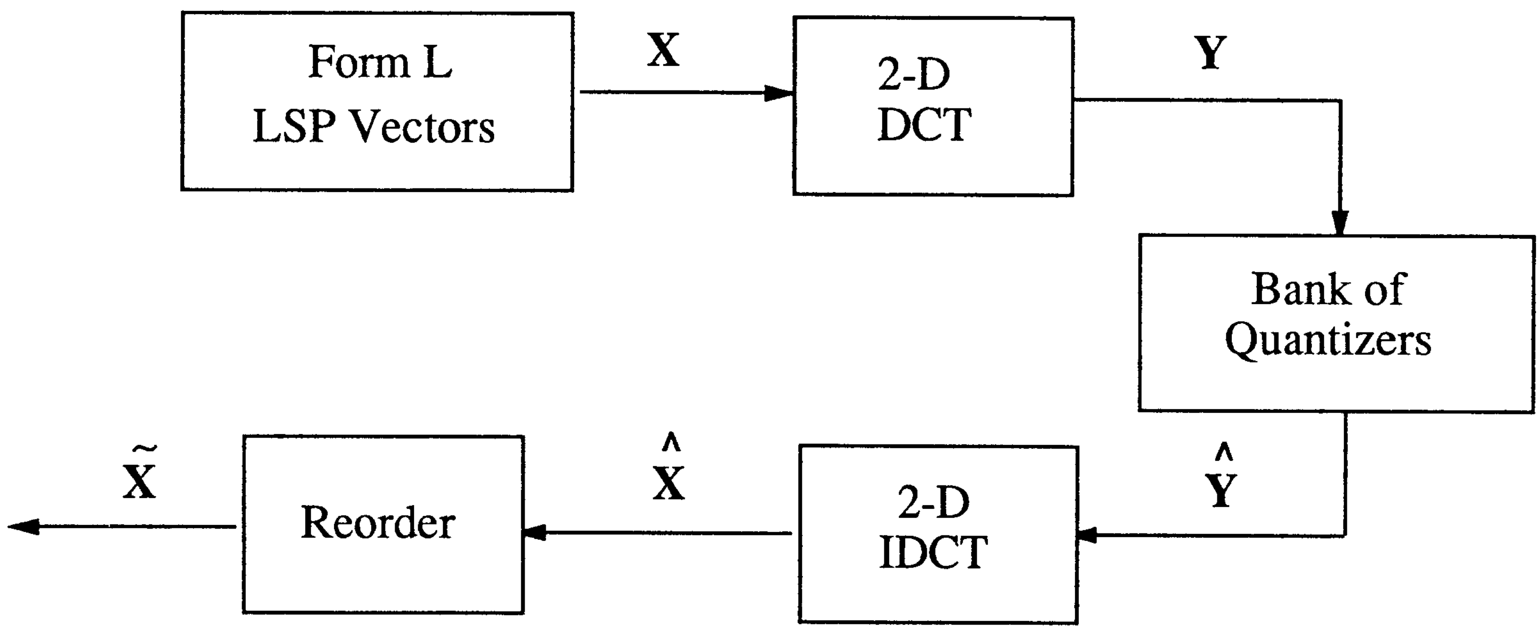

Figure 1: Block Diagram of 2-D DCT Coding Scheme

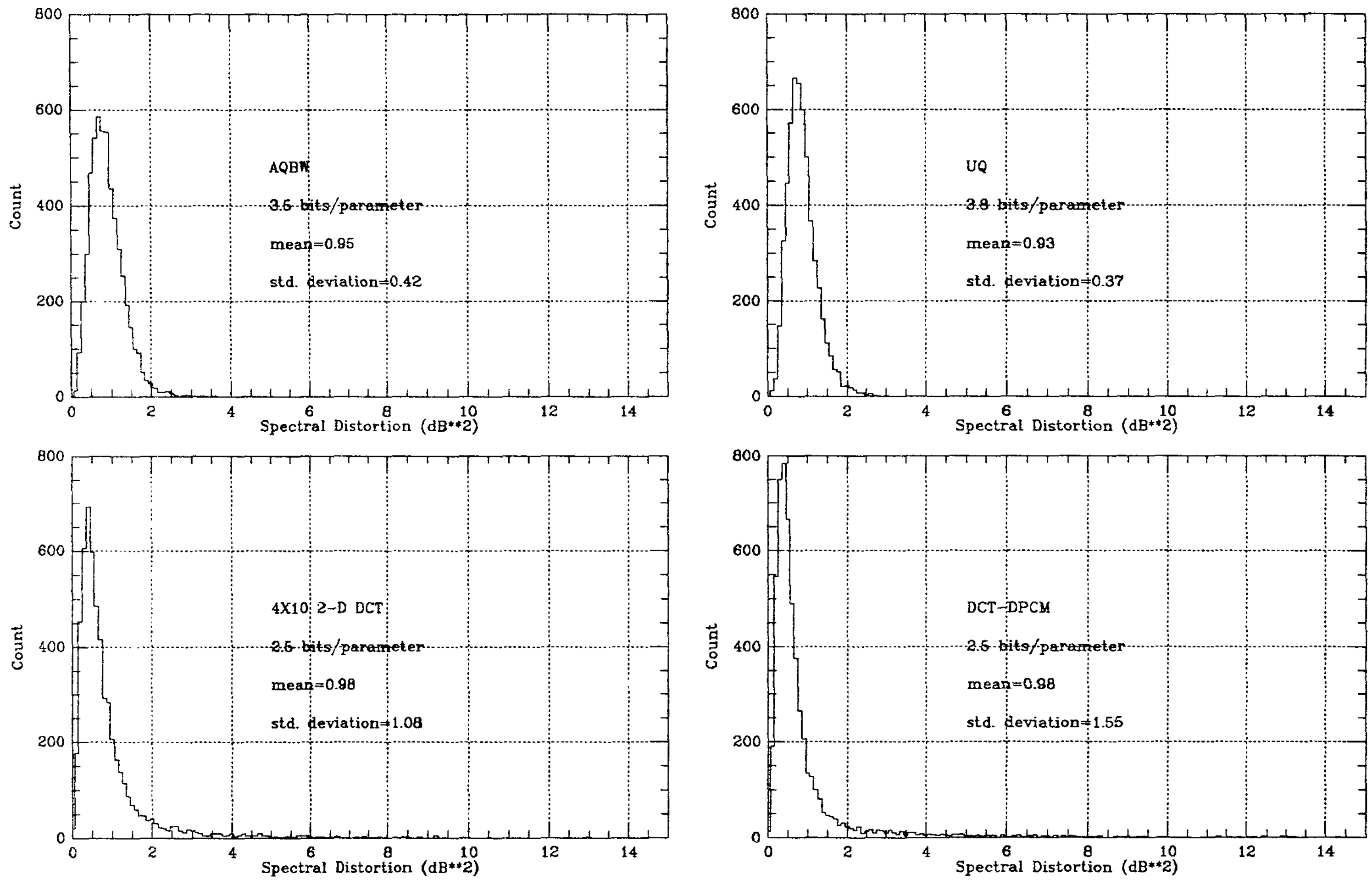

Figure 2: Histogram of Spectral Distortion for Different Schemes