SRC TR 87-113

Quantizer Design in LSP Speech Analysis-Synthesis

by

Noboru Sugamura and
Nariman Farvardin

Quantizer Design in LSP Speech Analysis-Synthesis †

by

Noboru Sugamura
Electrical Engineering Department and Systems Research Center
University of Maryland
College Park, Maryland 20742
on leave from
NTT Integrated Communications Laboratories
Nippon Telegraph and Telephone Corporation
Yokosuka Kanagawa 238-03 Japan
and
Nariman Farvardin
Electrical Engineering Department,
Institute for Advanced Computer Studies
and Systems Research Center
University of Maryland
College Park, Maryland 20742

Abstract

The LSP speech analysis-synthesis method is known as one the most efficient vocoders. An important issue in encoding of the LSP parameters is that a certain ordering relationship between the LSP parameters is required to insure the stability of the synthesis filter. This requirement has an important impact on the design of quantizers for the LSP parameters. In this paper, the performance of several algorithms for the quantization of the LSP parameters is studied. A new adaptive method which utilizes the ordering property of the LSP parameters is presented. A combination of this adaptive algorithm with non-uniform step size quantization is shown to be a very effective method for encoding the LSP parameters. The performance of the different quantization schemes is studied on a long sequence of speech samples. For the spectral distortion measure, appropriate performance comparisons between the different quantization schemes are rendered.

† This work was supported in part by a National Science Foundation grant NSFD CDR-85-00108, and in part by a grant from the Minta Martin Fund for Aeronautical Research, University of Maryland.

1. Introduction

Various speech analysis-synthesis methods for speech transmission under 9.6 kbps have been studied. Among these, the LSP (Line Spectrum Pair) analysis-synthesis method was proposed by Itakura and Sugamura in 1979 for the first time [1]. This method is known as the one of the most efficient speech analysis-synthesis techniques [2], [3]. The LSP parameters, which represent the short-time speech spectra, are completely equivalent, in a mathematical sense, to other linear predictive coding coefficients, such as the LPC or the PARCOR coefficients. However, the LSP parameters have some additional interesting properties which makes them more attractive than the LPC or PARCOR coefficients. Experimental results indicate that high-quality synthesized speech can be obtained using LSP parameters at relatively low rates. Subjective performance tests [4] also confirm the high quality of the synthesized speech.

The work in [4] is limited to uniform scalar quantization of the LSP parameters with the spectral distortion measure. In a more recent work, Soong and Juang [5] have utilized some basic properties of the LSP parameters to develop a differential quantization scheme for encoding the LSP parameters. Furthermore, vector quantization of the LSP parameters is reported in [6].

In this paper we study several quantization methods for encoding the LSP parameters. The spectral distortion measure is used for objective comparison of the different quantization schemes. Among others, an adaptive quantization scheme is developed which utilizes the so-called *ordering* property of the LSP parameters. It is shown that while this algorithm is fairly simple to implement, it results in noticeable performance improvements over all other scalar quantization schemes for the LSP parameters. The relationship between the average spectral distortion and the bit rate for the different quantization schemes is studied.

The rest of this paper is organized as follows. Section 2 includes a brief review of the LSP speech analysis-synthesis method. In this section, the relationship between the LSP parameters and the more familiar LPC and PARCOR coefficients is described and some basic properties of the LSP parameters are discussed. In Section 3 several algorithms for the quantization of the LSP parameters are presented. This is followed by Section 4 in which a performance comparison of the different quantization schemes is presented. Finally, in Section 5 a summary and conclusions is provided.

2. LSP Speech Analysis and Synthesis

In this section, we provide a brief description of the LSP speech analysis and synthesis and its implications. The interested reader is encouraged to refer to references [1]-[5] for more details.

2.1. LSP Speech Analysis and Its Interpretation

For a given order p, the linear predictive coding (LPC) analysis results in an all-pole filter $1/A_p(z)$, described by

$$H(z) \;=\; \frac{1}{A_p(z)} \;=\; \frac{1}{1+\alpha_1 z+\alpha_2 z^2+\ldots+\alpha_p z^p}, \tag{2.1}$$

where $z = e^{-j\omega}$, $-\pi < \omega < \pi$. The parameters $\{\alpha_i\}_{i=1,\ldots p}$, are well-known as the LPC coefficients [16].

It is easy to verify that the polynomial $A_n(z)$ satisfies the following relationship [1]:

$$\begin{aligned} A_n(z) &= A_{n-1}(z) - k_n B_{n-1}(z), & A_0(z) &= 1, \\ B_n(z) &= zB_{n-1}(z) - k_n A_{n-1}(z), & B_0(z) &= 1, \end{aligned} \tag{2.2}$$

where $A_n(z)$ and $B_n(z)$ are related by

$$B_n(z) \;=\; z^{n+1} A_n(1/z). \tag{2.3}$$

In (2.2), the parameters $\{k_i\}_{i=1,2,\ldots,p}$, are called the PARCOR coefficients. It is important to note that the PARCOR coefficients are completely equivalent to the LPC coefficients in a mathematical sense. In other words, the PARCOR coefficients and the LPC coefficients represent the same spectral information. In the context of speech compression, it is well-known that the LPC coefficients are inappropriate for quantization. This is primarily due to their wide dynamic range and the concomitant instability issues in the synthesis filter

[16]. For the PARCOR coefficients, however, it is known that if all $|k_i|$ are less than one, the stability of the synthesis filter is guaranteed. The PARCOR coefficients are also interpreted as the reflection coefficients at the boundary of the acoustic tube model for the vocal tract [16].

For $n = p + 1$ in (2.2), we have

$$\begin{aligned} A_{p+1}(z) &= A_p(z) - k_{p+1}B_p(z) \\ &= A_p(z) - k_{p+1}z^{p+1}A_p(1/z). \end{aligned} \tag{2.4}$$

In (2.4) consider two extreme artificial boundary conditions , $k_{p+1} = 1$ and $k_{p+1} = -1$. These conditions correspond to a *complete closure* and a *complete opening* at the glottis in the acoustic tube model, respectively. Under the conditions, $k_{p+1} = 1$ and $k_{p+1} = -1$, the polynomial $A_{p+1}(z)$ coincides with the polynomials

$$\begin{aligned} P(z) = &\; A_p(z) - z^{p+1}A_p(1/z) \\ = &\; 1 + (\alpha_1 - \alpha_p)z + \ldots + (\alpha_p - \alpha_1)z^p - z^{p+1}, \end{aligned} \tag{2.5a}$$

and

$$\begin{aligned} Q(z) = &\; A_p(z) + z^{p+1}z_p(1/z) \\ = &\; 1 + (\alpha_1 + \alpha_p)z + \ldots + (\alpha_p + \alpha_1)z^p + z^{p+1}, \end{aligned} \tag{2.5b}$$

respectively.

When p is an even integer [1], the polynomials $P(z)$ and $Q(z)$ can be expressed as [1]

$$P(z) = (1 - z) \prod_{i=2,4,\ldots,p} (1 - 2z\cos\omega_i + z^2), \tag{2.6a}$$

and

$$Q(z) = (1 + z) \prod_{i=1,3,\ldots,p-1} (1 - 2z\cos\omega_i + z^2), \tag{2.6b}$$

where it is assumed that $\omega_1 < \omega_3 < \cdots < \omega_{p-1}$ and $\omega_2 < \omega_4 < \cdots < \omega_p$.

[1] Throughout this paper, we will confine attention to even values of p.

It is clear from (2.6) that $e^{-j\omega_i}$, $i = 1, 2, \cdots, p$, are the roots of the polynomials $P(z)$ and $Q(z)$. The parameters $\{\omega_i\}_{i=1,2,\cdots,p}$ are defined as the *Line Spectrum Pair* (LSP) parameters [1]. It is important to note that $\omega_0 = 0$ and $\omega_{p+1} = \pi$ are fixed roots of $P(z)$ and $Q(z)$, respectively, and will be excluded from the LSP parameters. 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. We will elaborate on this later.

The polynomials $P(z)$ and $Q(z)$ possess some very interesting and important properties summarized in the following [1]:

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

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

$$0 = \omega_0 < \omega_1 < \omega_2 < \ldots < \omega_{p-1} < \omega_p < \omega_{p+1} = \pi. \tag{2.7}$$

¿From now on, we will refer to the above relationship as the *ordering* property of the LSP parameters.

The following derivation leads to an interpretation of the LSP spectral representation of speech. Using the LSP parameters, the power transfer function of $H(z)$ can be calculated as follows:

$$\begin{aligned} |H(e^{-j\omega})|^2 &= 1/|A_p(e^{-j\omega})|^2 \\ &= 4/|P(e^{-j\omega}) + Q(e^{-j\omega})|^2 \\ &= 2^{-p}/\{\sin^2\frac{\omega}{2}\prod_{i=2,4,\ldots,p}(\cos\omega - \cos\omega_i)^2 \\ &\qquad + \cos^2\frac{\omega}{2}\prod_{i=1,3,\ldots,p-1}(\cos\omega - \cos\omega_i)^2\}. \end{aligned} \tag{2.8}$$

Equation (2.8) implies that $|H(e^{-j\omega})|^2$ has a strong resonance at frequency ω when at least two LSP parameters are located near ω. Therefore, the LSP parameters can be interpreted as a representation of an all pole filter by means of the location density of p discrete frequencies, namely $\{\omega_1, \omega_2, \cdots, \omega_p\}$, in the frequency domain. In view of the steps used in defining the LSP parameters, they are clearly equivalent to the LPC coefficients $\{\alpha_i\}_{i=1,2,\cdots,p}$, and the PARCOR coefficients $\{k_i\}_{i=1,2,\cdots,p}$. However, it is established experimentally that the LSP parameters have better quantization and interpolation properties comparing with the LPC and PARCOR coefficients [2].

2.2. LSP Speech Synthesis

The filter $H(z)$ can be realized by a feedback loop, with a unity feedforward gain and a feedback gain of $A_p(z)-1$. The feedback gain $A_p(z)-1$, in turn, can be realized directly using the LSP parameters as follows [1]:

$$\begin{aligned} A_p(z) - 1 = & \; [(P(z)-1) + (Q(z)-1)]/2 \\ = & \; \frac{z}{2}\left[\sum_{\substack{i=2 \\ (i=\text{even})}}^{p} (a_i + z) \prod_{\substack{j=0 \\ (j=\text{even})}}^{i-2} (1 + a_j z + z^2) - \prod_{\substack{j=2 \\ (j=\text{even})}}^{p} (1 + a_j z + z^2) \right. \\ & \left. + \sum_{\substack{i=1 \\ (i=\text{odd})}}^{p-1} (a_i + z) \prod_{\substack{j=-1 \\ (j=\text{odd})}}^{i-2} (1 + a_j z + z^2) + \prod_{\substack{j=1 \\ (j=\text{odd})}}^{p-1} (1 + a_j z + z^2) \right], \quad (2.9) \end{aligned}$$

where $a_i = -2\cos\omega_i$, $i = 1, 2, \ldots, p$, and $a_{-1} = a_0 = -z$.

It is established in [1] and [9] that equation (2.7) (i.e., the ordering property of the LSP parameters), is the *necessary* and *sufficient* condition for the stability of the LSP synthesis filter. In an LSP-based speech compression system, the LSP parameters must be quantized. Therefore, in order to guarantee stability for the LSP synthesis filter, we must make certain that the quantized versions of the LSP parameters also satisfy the ordering property. This is a very important issue in the design of quantizers and dequantizers for the LSP parameters. We will elaborate on this in the following section.

3. Algorithms for Quantization of the LSP Parameters

In a speech coding situation based on LSP analysis, the LSP parameters must be quantized and encoded. In this section, several algorithms for the quantization of the LSP parameters are described. In order to reduce the average quantization distortion, some of the algorithms utilize the ordering property of the LSP parameters as described by (2.7). In this paper, only scalar quantization is considered. Also, throughout the paper, the distortion measure used for the design of the quantizers is the squared-error distortion measure.

3.1. Uniform Quantization (UQ)

Let $\omega_{i,\min}$ and $\omega_{i,\max}$ denote the minimum and the maximum values of the ith LSP parameter, respectively. Let us assume that these values are precomputed for all LSP parameters based on a long sequence of speech samples. In the uniform quantization (UQ) scheme, it is assumed that all LSP parameters are quantized by means of uniform quantizers. The step size of the quantizer for the ith LSP parameter is denoted by Δ_i and described by

$$\Delta_i = \frac{\omega_{i,\max} - \omega_{i,\min}}{2^{b_i}}, \quad i = 1, 2, \ldots p, \tag{3.1}$$

where b_i is the number of quantization bits for the ith LSP parameter.

While it might be argued that the best choice of the step size may not equal to that given by (3.1) (because the LSP parameters are not uniformly distributed), empirical evidence has shown that the values of Δ_i, $i = 1, 2, \cdots, p$, given by (3.1) yield good performance results.

Despite the fact that the LSP parameters satisfy the ordering property described by (2.7), since they are quantized independently, there is a possibility that their quantized versions do not satisfy (2.7). This phenomenon, which is merely the result of quantization noise, is more likely to happen in the low bit rate quantization of the LSP parameters (low bit rate speech coding). As mentioned before, this violation of ordering of the LSP parameters results in an instability of the LSP synthesis filter, and therefore, should be avoided. In what follows we will describe a dequantization algorithm which, in the decoder, chooses the best set of reconstruction levels in an effort to minimize the squared-error quantization distortion while maintaining stability of the synthesis filter.

Let $\hat{\omega}_{i,j}$, $i = 1,2,\ldots,p$; $j = 1,2,\ldots,2^{b_i}$, denote the jth reconstruction level of the quantizer operating on the ith LSP parameter. Then $d_{i,j}$, $i = 1,2,\ldots,p$; $j = 1,2,\ldots,2^{b_i}$, the squared-error in quantizing the ith LSP parameter, ω_i, to the jth level $\hat{\omega}_{i,j}$, is given by

$$d_{i,j} \;=\; (\omega_i \;-\; \hat{\omega}_{i,j})^2. \tag{3.2}$$

The following algorithm describes a procedure for optimum decoding of the LSP parameters such that the squared-error distortion is minimized while stability of the LSP synthesis filter is maintained.

[**Algorithm I**]

(1) For $i = 1$, calculate $d_{1,j}$ and set $G_{1,j} = d_{1,j}$ for $j = 1,2,\cdots,2^{b_1}$.

(2) For $i = 2$ to p,

  (i) calculate $d_{i,j}$ for $j = 1,2,\cdots,2^{b_i}$,

  (ii) for each $j = 1,2,\cdots,2^{b_i}$, find an index j' such that $G_{i-1,j'} + d_{i,j}$ is minimized subject to $\hat{\omega}_{i-1,j'} < \hat{\omega}_{i,j}$. Define $G_{i,j} = G_{i-1,j'} + d_{i,j}$. Here, $G_{i,j}$ denotes the *minimum accumulated quantization error* for the jth reconstruction level of the ith LSP parameter,

  (iii) for each jth reconstruction level for the ith LSP parameter, memorize the corresponding index j' obtained in (ii) and define $L_{i,j} \triangleq j'$.

(3) For the pth LSP parameter, find that index j^* which minimizes $G_{p,j}$, i.e., $G_{p,j^*} = \min\{G_{p,j}\}$, where the minimization is over all $j = 1,2,\cdots,2^{b_p}$. Denote this index by $\ell_p = j^*$.

(4) Find the set of indices of optimum reconstruction levels for the $(p-1)$st, $(p-2)$nd, $\cdots$, 1st , LSP parameters, denoted by $\{\ell_k\}_{k=p-1,p-2,\cdots,1}$, by backtracking $L_{i,j}$'s. More specifically, the index ℓ_i of the optimum reconstruction level for the ith LSP parameter is obtained recursively according to

$$\ell_i = L_{i+1,\ell_{i+1}}, \; i = p-1, p-2, \cdots, 1, \tag{3.3}$$

with $\ell_p = j^*$.

The procedure described above yields the optimum choice of reconstruction levels, which minimizes the squared quantization error while preserving the ordering property as described in (2.7). It must be noted that this algorithm could be used for the decoding of the LSP parameters regardless of the type of individual quantizers – uniform or non-uniform.

The above algorithm is a dynamic programming-based algorithm which, compared to the exhaustive search method, substantially reduces the number of computations needed to determine the optimum choice of the reconstruction levels. In fact, the computational complexity of this algorithm can be further reduced by noticing that the calculation of the minimum accumulated quantization error in Step (2) is not necessary for all possible values of the index j. For instance, let $\hat{\omega}_{i,j}$ and $\hat{\omega}_{i+1,k}$ be such that $\hat{\omega}_{i,j} \geq \hat{\omega}_{i+1,k}$. In this case only the two combinations $\hat{\omega}_{i,j-1}$ and $\hat{\omega}_{i+1,k}$ $(\hat{\omega}_{i,j-1} < \hat{\omega}_{i+1,k})$, $\hat{\omega}_{i,j}$ and $\hat{\omega}_{i+1,k+1}(\hat{\omega}_{i,j} < \hat{\omega}_{i+1,k+1})$ should be calculated.

3.2 Non-Uniform Quantization (NUQ)

It is well-known that non-uniform quantization is superior to uniform quantization in the sense that it results in a smaller average quantization error, except in the case of the uniform distribution. Let us now suppose that the number of quantization levels and the probability distribution function of all LSP parameters are known. The objective is to find the optimum quantization levels in the sense of minimizing the expected value of quantization errors. More specifically, for a generic LSP parameter, say, ω, if we denote its quantized version by $q(\omega)$, the objective in designing an N-level quantizer is to choose the optimum reconstruction levels and thresholds to minimize

$$\begin{aligned} D &= \int_{\omega_{\min}}^{\omega_{\max}} (\omega - q(\omega))^2 p(\omega) d\omega \\ &= \sum_{j=1}^{N} \int_{T_{j-1}}^{T_j} (\omega - \hat{\omega}_j)^2 p(\omega) d\omega, \end{aligned} \tag{3.4}$$

where the T_j's are the *threshold levels* ($T_0 = \omega_{\min}$ and $T_N = \omega_{\max}$), the $\hat{\omega}_j$'s are the *reconstruction levels*, $p(\omega)$ is the probability density function of the LSP parameter, and

$\omega_{\min}$ and $\omega_{\max}$ are the minimum and the maximum of the support of $p(\omega)$, respectively. It is easy to show that for the squared-error distortion measure, the optimum choice of the threshold levels satisfies $T_j = (\hat{\omega}_j + \hat{\omega}_{j+1})/2,\ j = 1, 2, \cdots, N-1$.

Algorithms for solving this problem are given by Lloyd [7] and Max [8]. However, the algorithms suggested in [7] and [8] provide only locally optimum quantizer structures. A dynamic programming-based algorithm can be used to obtain the globally optimum quantizer structure, [10], [11]. This algorithm is briefly described in the following. The interested reader should consult with [10] and [11] for details.

[**Algorithm II**]

First we discretize the interval $[\omega_{\min}, \omega_{\max}]$ into M subintervals of equal lengths. This will result in a set of discretization points described by $\Gamma = \{\gamma_1, \gamma_2, \cdots, \gamma_{M-1}\}$, where $\gamma_m = \omega_{\min} + (\omega_{\max} - \omega_{\min})m/M,\ m = 1, 2, \cdots, M-1$. The set Γ is a discrete collection of points from which we will choose the reconstruction levels. Obviously, the larger the value of M, the more precise will be the value of the reconstruction levels.

(1) For all $\hat{\omega}_{1,j} \in \Gamma$ [2], compute

$$D_{1,j} = \int_{\omega_{\min}}^{\hat{\omega}_{1,j}} (\omega - \hat{\omega}_{1,j})^2 p(\omega) d\omega. \tag{3.5}$$

(2) For each $n = 2, 3, \cdots, N-1$, and for all $\hat{\omega}_{n,j'} \in \Gamma$ compute

$$\begin{aligned} D_{n,j'} = \min \Bigg[D_{n-1,j} &+ \int_{\hat{\omega}_{n-1,j}}^{T_{j'}} (\omega - \hat{\omega}_{n-1,j})^2 p(\omega) d\omega \\ &+ \int_{T_{j'}}^{\hat{\omega}_{n,j'}} (\omega - \hat{\omega}_{n,j'})^2 p(\omega) d\omega \Bigg], \end{aligned} \tag{3.6}$$

where the minimization is with respect to the index j such that $\hat{\omega}_{n-1,j} < \hat{\omega}_{n,j'}$ and $T_{j'} = (\hat{\omega}_{n-1,j} + \hat{\omega}_{n,j'})/2$.

[2] To avoid confusion with the notation developed earlier, we emphasize that $\hat{\omega}_{n,j}$ denotes the jth choice in the set Γ which is being examined for the nth reconstruction level.

(3) Finally, to determine $\hat{\omega}_N$, compute

$$D_{N,j'} = \min\left[D_{N-1,j} + \int_{\hat{\omega}_{N-1,j}}^{T_{j'}} (\omega - \hat{\omega}_{N-1,j})^2 p(\omega) d\omega \right.$$

$$\left. + \int_{T_{j'}}^{\hat{\omega}_{N,j'}} (\omega - \hat{\omega}_{N-1,j})^2 p(\omega) d\omega + \int_{\hat{\omega}_{N,j'}}^{\omega_{\max}} (\omega - \hat{\omega}_{N,j'})^2 p(\omega) d\omega \right], \quad (3.7)$$

where the minimization is with respect to the index j such that $\hat{\omega}_{N-1,j} < \hat{\omega}_{N,j'}$ and $T_{j'} = (\hat{\omega}_{N-1,j} + \hat{\omega}_{N,j'})/2$. Then find the minimum value of $D_{N,j'}$ among all choices of j'. After finding this, the optimum reconstruction levels can be obtained by backtracking the process described in Step (2).

In the actual computation of the optimum reconstruction levels, the discrete version of the probability density function of the LSP parameters, computed from a long sequence of speech samples, is used.

It must be noted that performance improvements can be obtained by designing the quantizers based on the conditional probability density functions $p(\omega_i|\omega_{i-1})$, $i = 2, 3, \cdots, p$. This is because of the inherent dependency between the consecutive LSP parameters which is best described by the ordering property described by (2.7). In this work, however, we have only used the marginal probability density functions of the LSP parameters to avoid the additional complexity associated with the design and implementation of quantizers based on the conditional densities.

We have to mention at this point that while the use of optimum mean square error quantizers certainly reduces the average quantization error as compared to the uniform quantization case, there is still a possibility that the correct ordering of the LSP parameters is not preserved at the output of these quantizers (especially in the low bit rate region). However, Algorithm I can still be applied to obtain the best *dequantization* of the LSP parameters while preserving the correct order. Our results in the next section on NUQ are indeed obtained by combining Algorithms I and II; Algorithm II is used for the design of quantizers for individual LSP parameters while Algorithm I is used for optimal decoding.

3.3 Differential Quantization (DQ)

Due to the ordering property of the LSP parameters, it is conceivable that the frequency difference of consecutive LSP parameters possess a smaller dynamic range compared with the LSP parameters themselves. Experimental evidence indicates that this is indeed true. This has become the motivation for a method first reported by Soong and Juang [5],in which instead of quantizing the LSP parameters, the differences of consecutive LSP parameters are quantized–hence, the term differential quantization. Since we have used this algorithm for comparison purposes, in what follows we will provide a brief description of it.

[**Algorithm III**]

(1) Quantize ω_1 to $\hat{\omega}_1$ and set $i = 1$;

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

(3) Quantize $\Delta\omega_i$ to $\Delta\hat{\omega}_i$ [3];

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

(5) If $i = p$, stop; otherwise set $i = i + 1$ and go to (2).

3.4 Adaptive Quantization (AQ)

Finally, in this subsection we will present an alegant algorithm which utilizes the ordering property of the LSP parameters to reduce the quantization error. The algorithm, essentially, uses the fact the knowledge of the value of ω_i (or its quantized version) provides useful information about the range of possible values of ω_{i+1}; this information can be used to design a better quantizer for the $(i + 1)$st parameter. The algorithm is detailed in the following.

3.4.1. Forward Adaptive Quantization (AQFW)

[**Algorithm IV**]

(1) Quantize ω_1 to $\hat{\omega}_1$ with b_1 bits and set $i = 1$.

(2) Compare $\hat{\omega}_i$ and $\omega_{i+1,\min}$,

[3] In Algorithm III, $\hat{\omega}_i$ and $\Delta\hat{\omega}_i$ denote the quantized versions of ω_i and $\Delta\omega_i$, respectively.

(a) If $\hat{\omega}_i \leq \omega_{i+1,\min}$, then quantize ω_{i+1} to $\hat{\omega}_{i+1}$ with b_{i+1} bits using a uniform quantizer with step size, $\Delta_{i+1} = (\omega_{i+1,\max} - \omega_{i+1,\min})/2^{b_{i+1}}$, over the range $[\omega_{i+1,\min}, \omega_{i+1,\max}]$;

(b) If $\hat{\omega}_i > \omega_{i+1,\min}$, then quantize ω_{i+1} to $\hat{\omega}_{i+1}$ with b_{i+1} bits using a uniform quantizer with step size, $\Delta_{i+1} = (\omega_{i+1,\max} - \hat{\omega}_i)/2^{b_{i+1}}$, over the range $[\hat{\omega}_i, \omega_{i+1,\max}]$, because we know $\hat{\omega}_{i+1}$ must be greater than $\hat{\omega}_i$. (This is the case where the knowledge of ω_i becomes useful in quantizing ω_{i+1}.)

(3) If $i = p - 1$, stop; otherwise, set $i = i + 1$ and go to (2).

3.4.2 Backward Adaptive Quantization (AQBW)

The idea used in AQFW can also be applied in the backward direction. This is outlined in the following.

[**Algorithm V**]

(1) Quantize ω_p to $\hat{\omega}_p$ with b_p bits and set $i = p$.

(2) Compare $\hat{\omega}_i$ and $\omega_{i-1,\max}$,

(a) If $\hat{\omega}_i \geq \omega_{i-1,\max}$, then quantize ω_{i-1} to $\hat{\omega}_{i-1}$ with b_{i-1} bits using a uniform quantizer with step size $\Delta_{i-1} = (\omega_{i-1,\max} - \omega_{i-1,\min})/2^{b_{i-1}}$ over the range $[\omega_{i-1,\min}, \omega_{i-1,\max}]$.

(b) If $\hat{\omega}_i < \omega_{i-1,\max}$, then quantize ω_{i-1} to $\hat{\omega}_{i-1}$ with b_{i-1} bits using a uniform quantizer with step size $\Delta_{i-1} = (\hat{\omega}_i - \omega_{i-1,\min})/2^{b_{i-1}}$ over the range $[\omega_{i-1,\min}, \hat{\omega}_i]$.

(3) If $i = 2$, stop; otherwise set $i = i - 1$ and go to (2).

In AQFW and AQBW, since the ordering of the LSP parameters is preserved after quantization, parameter inversion never occurs.

Obviously, the average value of quantization error can be reduced comparing with the UQ method which always quantizes the LSP parameters over their full distribution range. Observe that in AQFW we start by quantizing the first LSP parameter while in AQBW we start by quantizing the last LSP parameter. In fact, the choice of the initial LSP parameter is rather arbitrary. If the middle-order parameter (e.g., the $(p/2)$th) is chosen as the initially quantized parameter, the quantization procedure can proceed in parallel in both directions, namely forward and backward. The advantage of such a technique is that parallel architectures can be used for the implementation of the algorithm which could result in a saving in the processing time. In this paper, to avoid confusing the issue, we merely study the AQFW and AQBW.

3.5 Adaptive Quantization with Adaptive Bit Allocation (AQ-AB)

Let us consider, for instance, the AQFW algorithm. This algorithm uses the knowledge of the value of one LSP parameter to reduce the range of possible values of the *next* LSP parameter. Thus, it is conceivable that sometimes the next LSP parameter can be quantized with a smaller number bits with nearly the same average distortion as in the nonadaptive UQ case. This has become the motivation to study the potential advantages of such adaptive bit allocations. We propose a very simple algorithm for adaptive bit allocation in AQFW and AQBW, hereafter referred to as AQFW-AB and AQBW-AB, respectively. These algorithms are described below.

In Algorithms IV or V, if the following relationship is satisfied, ω_i can be quantized with approximately the same distortion with one fewer bit.

[1] In Algorithm IV,

$$\omega_{i,\max} - \hat{\omega}_{i-1} \leq (\omega_{i,\max} - \omega_{i,\min})/2. \tag{3.8a}$$

[2] In Algorithm V,

$$\hat{\omega}_{i+1} - \omega_{i,\min} \leq (\omega_{i,\max} - \omega_{i,\min})/2. \tag{3.8b}$$

3.6 Adaptive Quantization with Non-Uniform Step Size (AQ-NU)

Obviously, combining the ideas in AQ and NUQ will result in an improved quantization scheme referred to as AQ-NU. The algorithm is as follows.

[**Algorithm VI**]

(1) Quantize ω_1 to $\hat{\omega}_1$ with b_1 bits using a non-uniform quantizer determined by the procedure mentioned in Section 3.2, and set $i = 1$.

(2) Compare $\hat{\omega}_i$ and $\omega_{i+1,\min}$,

(a) If $\hat{\omega}_i \leq \omega_{i+1,\min}$, then quantize ω_{i+1} to $\hat{\omega}_{i+1}$ with b_{i+1} bits using a non-uniform optimum quantizer over the range $[\omega_{i+1,\min}, \omega_{i+1,\max}]$. The non-uniform reconstruction levels must be precomputed using the conditional probability density function $p(\omega_{i+1}|\hat{\omega}_i)$ for each reconstruction level of the ith LSP parameter which satisfy $\hat{\omega}_i \leq \omega_{i+1,\min}$.

(b) If $\hat{\omega}_i > \omega_{i+1,\min}$, then quantize ω_{i+1} to $\hat{\omega}_{i+1}$ with b_{i+1} bits using a non-uniform optimum quantizer over the range $[\hat{\omega}_i, \omega_{i+1,\max}]$. In this case also, conditonal distribution functions are used as in Step (a).

(3) If $i = p - 1$ stop; otherwise, set $i = i + 1$ and go to (2).

Similar modifications can be made to incorporate non-uniform quantization into the AQBW scheme.

Clearly, the above adaptive quantization scheme with non-uniform thresholds and reconstruction levels is superior to both the UQ and the AQ schemes. The the AQFW-NU and AQBW-NU are superior to NUQ because they use *smaller* ranges for quantization of the LSP parameters. Also, the AQFW-NU and AQBW-NU are superior to AQ because they use *better* quantizers for individual LSP parameters.

However, to implement this quantization scheme, the conditional density function and the non-uniform reconstruction levels using the conditional densities are needed. In fact, if b_i is the number of quantization bits for encoding the ith LSP parameter, the number of conditional density functions which must be precomputed is the following:

$$1 + \sum_{i=2}^{p} 2^{\sum_{j=1}^{i-1} b_j}$$

In the case where $p = 10$, the computation of so many conditional density functions and the corresponding optimal quantizers is rather impractical.

To alleviate this difficulty, an approximation of these conditional densities can be used. Specifically, suppose the support of the density of ω_i is discretized (uniformly) into M_i subintervals, each of which is represented by one point, say $\beta_{i,1}, \beta_{i,2}, \cdots, \beta_{i,M_i}$.

Then the desired conditional density $p(\omega_{i+1}|\hat{\omega}_i)$ is approximated by one of $p(\omega_{i+1}|\beta_{i,m} \leq \hat{\omega}_i \leq \beta_{i,m+1})$, $m = 1, 2, \ldots, M_i - 1$. In this case, the total number of conditional density functions which must be computed is:

$$\sum_{i=1}^{p-1} M_i.$$

4. Performance Results

We have studied the performance of the LSP-based speech coding system at different rates using the various quantization algorithms described in Section 3 on a long sequence of speech samples. The spectral distortion measure is used for objective comparison of these quantization schemes. In what follows we will describe and discuss some of our results. Some information about the contents of speech data samples and the experimental conditions are summarized in Table 1. The experimental conditions here are nearly the same as those used in [12], except the difference of language.

4.1. LSP Parameter Distribution

The distribution range and histogram of the LSP parameters are computed for the sequence of speech samples described in Table 1. These results are illustrated in Table 2 and Fig. 1. In Table 2 and Fig. 1, ω_i is converted to f_i, using the relationship $f_i = (\omega_i/2\pi) \cdot f_s$, Hz, where f_s is the sampling frequency. These values are used as the boundary points of the quantization ranges for various quantization schemes described in Section 3. A discretized version of the probability density function of each parameter is also computed over the appropriate range with 9 bits (2^9 discretization intervals). These approximations to the density functions are used to design the optimum quantizers described in Section 3.2.

4.2. Spectral Sensitivity

To determine how the quantization bits should be allocated to the different LSP parameters, we need to have a measure of the sensitivity of the system to the effect of quantization in individual LSP parameters. Since the spectral distortion is used as a measure of the system performance, we have computed the spectral sensitivity of individual LSP parameters.

The spectral sensitivity associated with an LSP parameter, say ω_i is defined as follows. Suppose ω_i is perturbed by a small amount δ and suppose this perturbation results in an average spectral distortion $SD(\delta)$. Then, the spectral sensitivity of ω_i is defined as $SD(\delta)/\delta$. For the speech data base used in our studies, the spectral sensitivity of the different LSP parameters are computed and illustrated in Table 3. For comparison purposes, the spectral sensitivity results for Japanese speech [12] are also included in Table 3. In view of the results in Table 3, there is no significant difference between the spectral sensitivity of the LSP parameters of English and Japanese. Only ω_1 and ω_2, which are closely related to the first formant frequency, have a slightly higher sensitivity compared with other parameters.

4.3. Performance Comparison

Based on the spectral sensitivity results tabulated in Table 3, we have adopted a uniform bit allocation among all LSP parameters as they appear to be almost equally sensitive to the quantization noise. The average spectral distortion is defined as follows.

$$SD = \frac{1}{N}\sum_{n=1}^{N}\left(\frac{1}{\pi}\int_0^{\pi}(\log S_n(\omega) - \log \hat{S}_n(\omega))^2 d\omega\right), \qquad (\mathrm{dB})^2 \qquad (3.9)$$

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 is the total number of frames. The spectral distortion measure is known to have a good correspondence with subjective measures [13].

In our experiments, all speech samples which are used to extract the parameter distribution ranges, are used to compute the average spectral distortion. In AQFW-NU or AQBW-NU, to simplify the design and implementation, we have considered a rather heuristic, but efficient, modification of Algorithm VI. Specifically, let us assume that $\Delta_1, \Delta_2, \cdots, \Delta_{b^i}$ are the step sizes of the optimum quantizer for the ith LSP parameter in the non-adaptive case. Then, in the adaptive case, namely AQFW-NU and AQBW-NU, if the range of the parameter is shrunk due to adaptation, the values of the step sizes

$\Delta_1, \Delta_2, \cdots, \Delta_{b^i}$ are reduced by the same proportionality factor. This argument is based on the assumption that the shape of the conditional densities are more-or-less similar to that of the marginals and only the ranges vary.

The experimental results are summarized in Fig. 2. A few comments about these results are in order.

(1) Clearly, AQFW, AQBW, DQ and NUQ are superior to UQ.

(2) The NUQ method has a performance which is nearly the same as AQFW or AQBW with uniform quantization. This is an indication of the fact that the LSP parameters do not possess a uniform distribution and that noticeable performance improvements can be obtained through the use of optimal non-uniform quantizers. Also, our experimental results indicate that the AQBW is slightly superior to AQFW. The reason for this is that in AQBW the LSP parameters of lower order, possessing higher spectral sensitivity, are quantized with smaller step sizes than in AQFW.

(3) Adaptive bit allocation in adaptive quantization method (AQFW-AB or AQBW-AB) is not very effective in reducing quantization bits. This is because the cases where the quantized range is less than half of full range do not occur very often, and hence the performance is nearly the same as AQFW or AQBW.

(4) The adaptive and non-uniform quantization methods (AQFW-NU and AQBW-NU) have offered the best results in reducing the spectral distortion at a given bit rate. Specifically, for the same value of the spectral distortion, the adaptive non-uniform quantization scheme (AQFW-NU or AQBW-NU) requires about 6 bits/frame fewer than the uniform case. Using the AQFW-NU or AQBW-NU algorithms, a spectral distortion of 1 dB^2, which is established as the difference limen of spectral distortion [14], can be obtained with about 32 bits/frame. This corresponds to approximately a 20% bit reduction compared with the PARCOR-based speech coding system described in [15].

5. Summary and Conclusions

We have studied several issues related to quantizer design and performance in LSP speech coding. Several quantization and dequantization algorithms for LSP parameters are proposed and their performances are compared to each other. Among these, the differential quantization algorithm and the adaptive quantization algorithms utilize the *ordering* property of the LSP parameters.

The rate vs. average spectral distortion performance of the different quantization schemes are determined experimentally.

According to the experimental results, the algorithm which combines the non-uniform step size quantizers with an adaptive encoder offers the best performance. Using this method, an average spectral distortion under 1dB^2 can be achieved with only 32 bits/frame.

The performance of these algorithms on out-of-the-training-sequence-data remained to be studied. Also, a more careful study of the effect of the optimal design of the quantizers based on the actual conditional probability density functions of the LSP parameters appears to be necessary.

References

[1] F. Itakura and N. Sugamura, "LSP Speech Synthesizer," Tech. Rept. S79-46, Speech Group, Acoustical Society of Japan, Nov. 1979 (in Japanese).

[2] N. Sugamura and F. Itakura, "Speech Analysis and Synthesis Methods Developed at ECL in NTT – From LPC to LSP," Speech Communication, Vol. 5, No. 2, June 1986.

[3] H. Wakita, "Linear Prediction Voice Synthesizer: Line-Spectrum Pair (LSP) is the Newest of Several Techniques," Speech Technology, Vol. 1,pp. 17-22, Fall, 1981.

[4] N. Sugamura, S. Sagayama and F. Itakura, "A Study on Speech Quality of Synthesized Speech by LSP," Tech. Rept. S80-29, pp. 229-233, Speech Group, Acoustical Society of Japan, July, 1980 (in Japanese).

[5] F. K. Soong and B. H. Juang, "Line Spectrum Pair (LSP) and Speech Data Compression," Proc. ICASSP-84, pp. 1.10.1-1.10.4 (1984).

[6] N. Sugamura and F. Itakura, "Speech Data Compression by Spectral Pattern Matching," IECE Trans., Vol. J65-A, No. 8, pp. 834-841, Aug. 1982 (in Japanese).

[7] S. P. Lloyd, "Least Squares Quantization in PCM," IEEE Trans. Inform. Theory, IT-28, pp. 129-136, March 1982.

[8] J. Max, "Quantizing for Minimum Distortion," IRE Trans. Inform. Theory, IT-6, pp. 7-12, March 1960.

[9] S. Sagayama, "Stability Condition of LSP Speech Synthesis Digital Filter," Proc. of ASJ, pp. 153-154, March 1982 (in Japanese).

[10] D. K. Sharma, "Design of Absolutely Optimal Quantizers for a Wide Class of Distortion Measures," IEEE Trans. Inform. Theory, IT-24, pp. 693-702, June 1978.

[11] A. V. Trushkin, "On Design of Optimal Quantizer," VI International Symposium on Information Theory, Tashkand, The Soviet Union, pp. 290-292, Sept. 1984.

[12] N. Sugamura and F. Itakura, "Speech Data Compression by LSP Speech Analysis and Synthesis Technique," IECE Trans., Vol. J64-A, No. 8, pp. 599-605, Aug. 1981 (in Japanese).

[13] N. Kitawaki, K. Itoh and K. Kakehi, "Speech Quality Measurement Methods for Synthesized Speech," Review of ECL, NTT Japan, Vol. 29, Nos. 9-11, Sept.-Dec. 1981.

[14] Y. Tohkura and F. Itakura, "Spectral Sensitivity of PARCOR and Difference Limen of Spectral Distortion," Proc. IECE Spring Meeting, p. 1231, 1977 (in Japanese).

[15] Y. Tohkura and F. Itakura, "Improvement of Voice Quality in PARCOR Bandwidth Compression System," IECE Trans., Vol. J61-A, No. 3, pp. 254-261, Mar. 1978 (in Japanese).

[16] L. R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signals*,Prentice-Hall, Inc., Englewood Cliffs, NJ, 1978.

Table 1 Experimental Conditions

(Speech Data)	
Speakers Sentences	2 Male and 2 Female 4 Sentences/Speaker
Sampling Frequency	8 KHz
(LSP Analysis)	
Frame Period	10 msec
Window	30 msec Hamming Window
Analysis Order	10 $(\omega_1, \omega_2, \ldots, \omega_{10})$
Total Number of Frames	5,445

Table 2 LSP Parameter Distribution Range

[KHz]

LSP	min	min 1% tail	max 1% tail	max
f_1	0.05	0.05	0.50	0.55
f_2	0.14	0.18	0.88	1.13
f_3	0.37	0.44	1.26	1.44
f_4	0.69	0.78	1.66	1.82
f_5	0.83	0.94	2.06	2.25
f_6	1.27	1.36	2.44	2.57
f_7	1.52	1.74	2.78	2.89
f_8	2.04	2.26	3.14	3.24
f_9	2.57	2.80	3.52	3.61
f_{10}	3.06	3.30	3.80	3.86

Table 3 Spectral Sensitivity of LSP Parameters

[dB/Hz]

Language / LSP Parameter	English	Japanese[12]
ω_1	0.018	0.022
ω_2	0.016	0.020
ω_3	0.013	0.013
ω_4	0.012	0.014
ω_5	0.014	0.015
ω_6	0.012	0.014
ω_7	0.012	0.012
ω_8	0.012	0.014
ω_9	0.011	0.012
ω_{10}	0.012	0.011

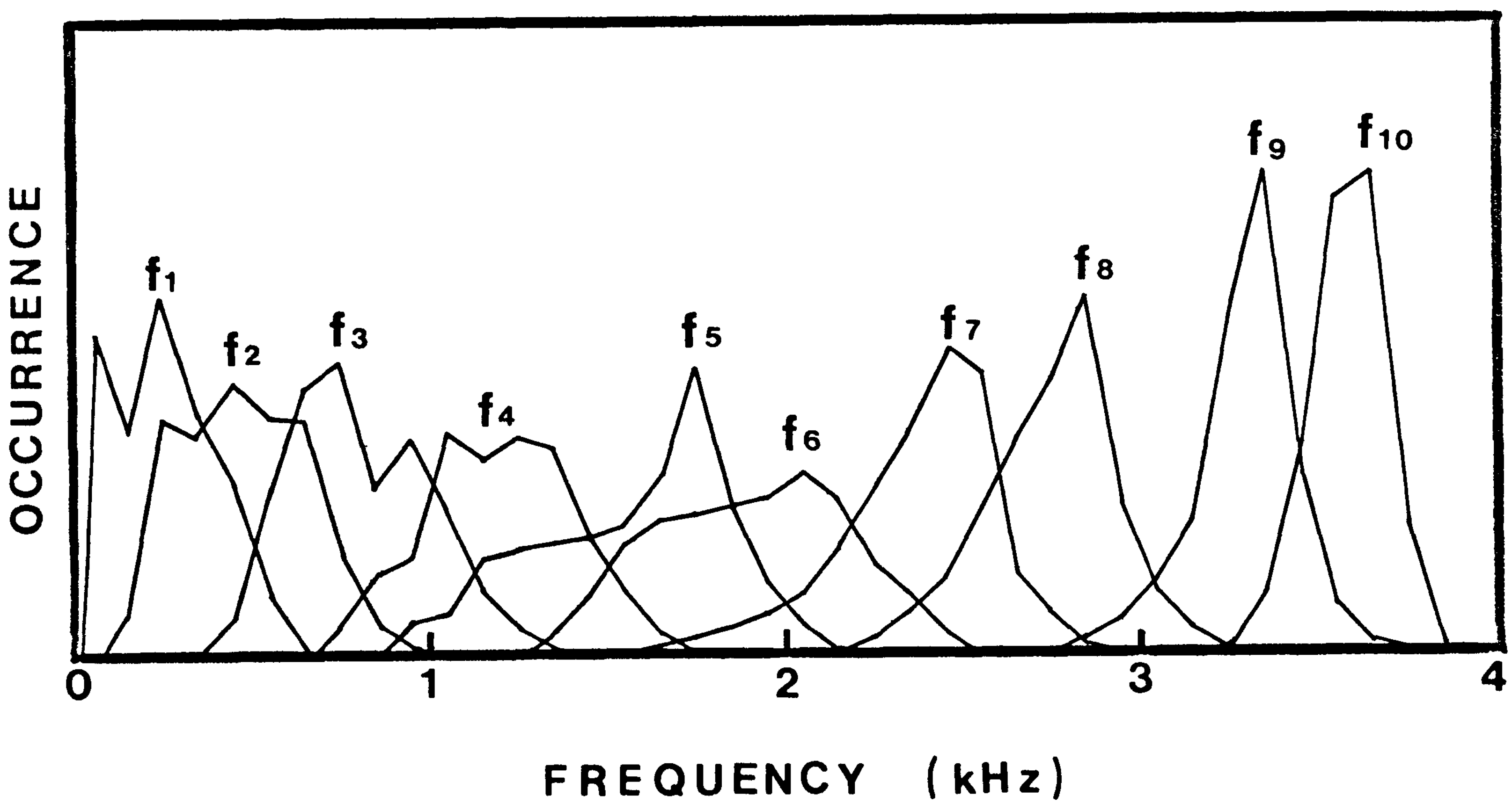

Figure 1: LSP Parameter Distribution

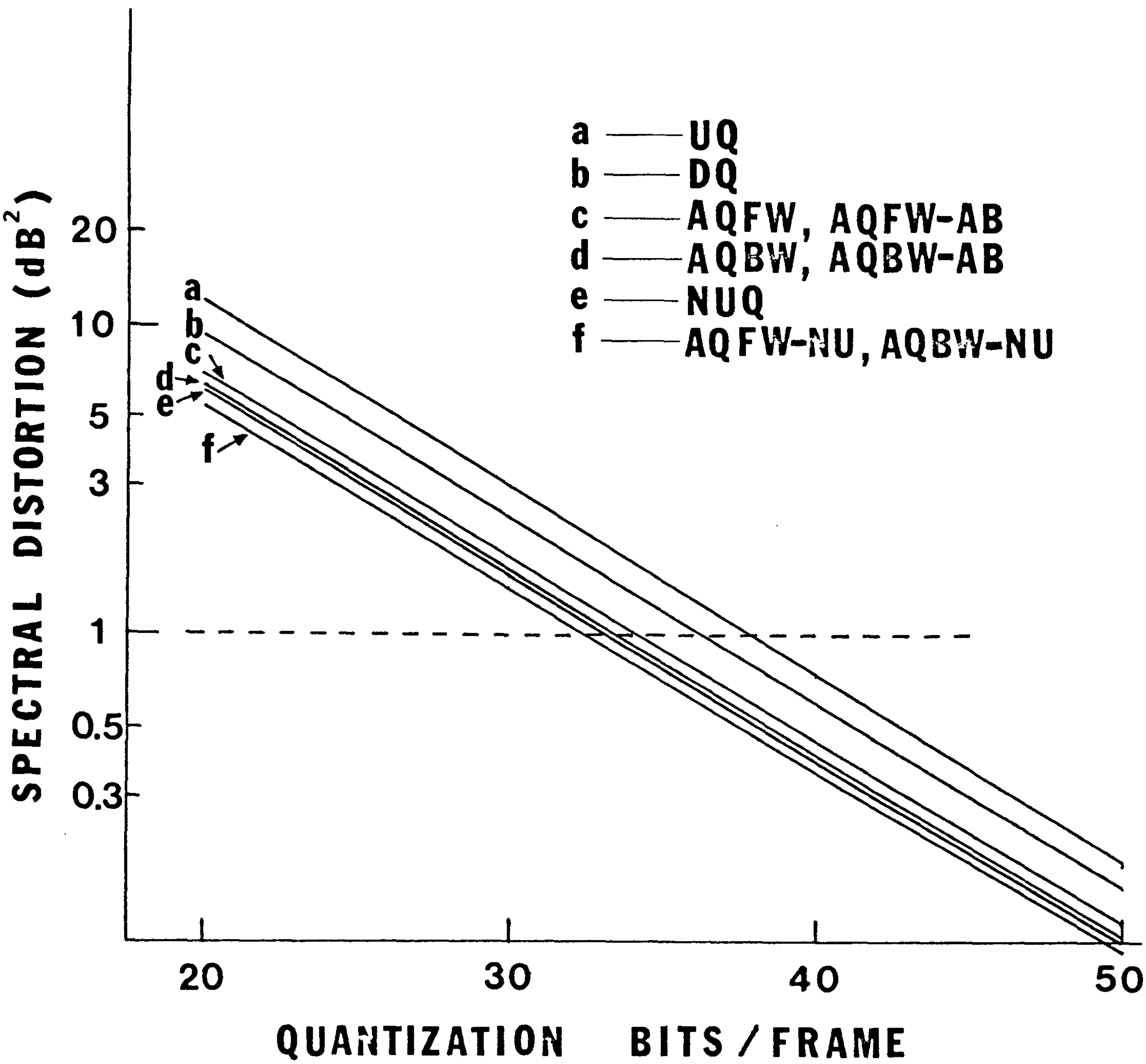

Figure 2: Relationship Between Bit Rate/Frame and Spectral Distortion in Various Quantization Methods