

# Incorporating Spectral Dynamics into LSF Vector Quantization with Error Shaping

Chin-Teng Lin,

Department of Computer Science,  
Department of Electrical and Control Engineering,  
National Chiao-Tung University,  
1001 Ta Hsueh Road, Hisnchu, Taiwan, 300, R.O.C.

Jen-Feng Chung, Der-Jenq Liu, Hsi-Wen Nein,

Department of Electrical and Control Engineering,  
National Chiao-Tung University,  
1001 Ta Hsueh Road, Hisnchu, Taiwan, 300, R.O.C.

## Abstract

This paper proposes an innovative technique to incorporate the information of spectral dynamics into the line spectrum frequency (LSF) vector quantization (VQ). This technique is based on a modified weighted log-spectral distortion (MWLSD) measure. It can be used to shape the spectral quantization distortion distribution into any frequency-dependent shaping curve, and simultaneously reduce the spectral dynamics distortion between quantized spectra and unquantized spectra. To reduce the high computational complexity of the MWLSD measure during the search procedure in the LSF VQ, a quadratically weighted distortion (QWD) measure used to approximate the MWLSD measure is derived based on the theoretical analysis of the MWLSD measure. A simplified quadratically weighted distortion (SQWD) measure is also proposed to further reduce the computational complexity of the QWD measure for practical applications, whose computational complexity is almost equal to that of weighted mean square error (WMSE) measure. Three experiments are performed to verify the proposed technique in the LSF VQ.

**Keywords:** Weighted LSD, quadratically weighted distortion measure, sensitivity matrix, human perception, speech coding, LPC quantization.

# 1 Introduction

The linear predictive coding (LPC) model for speech signals is used in many modern speech compression systems. In these systems, the LPC filter coefficients are usually transformed to line spectrum frequency (LSF) parameters which give a very effective representation for the quantization of LPC information [1], [2].

Currently, almost all LSF vector quantization (VQ) method generally uses static measure as the performance indicator of a quantizer. The static measure commonly includes log-spectral distortion (LSD) measure [3] and weighted log-spectral distortion (WLSLSD) measure [4]. It only considers the static quantization distortion error between the unquantized spectrum and quantized one at each frame. The LSD and WLSLSD measures in dB units are defined as

$$D_{\text{LSD}}(\mathbf{a}^{(t)}, \hat{\mathbf{a}}^{(t)}) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \left[ 10 \log_{10}(P^{(t)}(w)) - 10 \log_{10}(\hat{P}^{(t)}(w)) \right]^2 dw, \quad (1)$$

$$D_{\text{WLSLSD}}(\mathbf{a}^{(t)}, \hat{\mathbf{a}}^{(t)}) = \frac{1}{2\pi B_0} \int_{-\pi}^{\pi} B^2(w) \left[ 10 \log_{10}(P^{(t)}(w)) - 10 \log_{10}(\hat{P}^{(t)}(w)) \right]^2 dw, \quad (2)$$

where  $w$  is the radian frequency, and  $P^{(t)}(w)$  and  $\hat{P}^{(t)}(w)$  are the linear prediction power spectra before and after quantization at the  $t$ th frame index, respectively,

$$P^{(t)}(\omega) = \left| 1 - \sum_{k=1}^v a_k^{(t)} e^{-j\omega k} \right|^{-2}, \quad \hat{P}^{(t)}(\omega) = \left| 1 - \sum_{k=1}^v \hat{a}_k^{(t)} e^{-j\omega k} \right|^{-2}, \quad (3)$$

where  $v$  is the order of LPC filter and is constrained to be even,  $\mathbf{a}^{(t)} = [a_1^{(t)}, a_2^{(t)}, \dots, a_v^{(t)}]^T$  and  $\hat{\mathbf{a}}^{(t)} = [\hat{a}_1^{(t)}, \hat{a}_2^{(t)}, \dots, \hat{a}_v^{(t)}]^T$  are the linear prediction coefficients corresponding to the unquantized and quantized LPC parameters at the  $t$ th frame index, respectively. In Eq. (2),  $B(w)$  is called the error weighting function, and  $B_0$  is a normalization factor making  $B(w)/B_0$  to be unity root mean square (RMS). The values of  $B(w)$  are all real and positive and the function  $B(w)$  is required to be symmetric, i.e.,  $B(-w) = B(w)$ . In Eqs. 1 and 2, the spectral-dynamics distortion is not taken into account in measuring the performance of LSF VQ. However, it has been shown [5, 6] that the spectral dynamics plays an important role in human perception. In other words, the mean values of the LSD and WLSLSD are not very indicative of the perceived distortion.

There are only a few existing research works of LSF VQ, which take the spectral dynamics into account. Most of these works focus on smoothing the

LSF parameter trajectories rather than directly employing the spectral dynamics information to smooth the quantized spectra's dynamics or to reduce the spectral-dynamics distortion between the unquantized and quantized spectra. In [5, 6], the LSF parameter trajectories are subjected to constrained smoothing during encoding and decoding. One difficulty of using these methods is the need of determining constant parameters in the proposed LSF constrained smoothing procedures. Koishida et al. [7] proposed a spectral quantization scheme which uses statistics of the static and dynamic features of LSF parameters to smooth the dynamics of quantized LSF parameters. One drawback of this scheme is that the dimension of quantization space (including static and dynamic features) is twice larger than that in the conventional scheme which uses static features only. This make the spectral quantization scheme from impractical use due to higher quantization bit rates. In addition, all these previously proposed methods did not use the actual spectral dynamics information of unquantized speech spectra, nor did they take quantization noise shaping into account.

To make better use of the spectral dynamics information and take quantization noise shaping into account in LSF VQ, we now propose a modified WLSD (MWLSD) measure. The MWLSD measure uses the relative relation by spectral difference, i.e., the information of spectral dynamics, between two successive frames of speech spectra and is defined as

$$\begin{aligned}
 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda}) = & \frac{1}{2\pi B_0} \int_{-\pi}^{\pi} B^2(w) \left[ 10 \log_{10}(P^{(t)}(w)) - 10 \log_{10}(\hat{P}^{(t)}(w)) \right]^2 + \\
 & B^2(w) \left[ \left( 10 \log_{10}(P^{(t)}(w)) - 10 \log_{10}(P^{(t-1)}(w)) \right) - \right. \\
 & \left. \left( 10 \log_{10}(\hat{P}^{(t)}(w)) - 10 \log_{10}(\hat{P}^{(t-1)}(w)) \right) \right]^2 + \\
 & B^2(w) \left[ 10 \log_{10}(P^{(t-1)}(w)) - 10 \log_{10}(\hat{P}^{(t-1)}(w)) \right]^2 dw, \quad (4)
 \end{aligned}$$

where  $P^{(t)}(w)$ ,  $\hat{P}^{(t)}(w)$  and  $B(w)$  are already defined in Eq. (4), and  $P^{(t-1)}(w)$  and  $\hat{P}^{(t-1)}(w)$  are the linear prediction power spectra before and after quantization at the  $(t-1)$ th frame index, respectively. In Eq. (4),  $\Lambda$  and  $\hat{\Lambda}$  are the LPC parameters of two successive frames before and after quantization, respectively, which are defined as

$$\begin{aligned}
 \Lambda &= [\lambda_1, \lambda_2, \dots, \lambda_v, \lambda_{v+1}, \lambda_{v+2}, \dots, \lambda_{2v}]^T \\
 &= [\mathbf{a}^{(t)T}, \mathbf{a}^{(t-1)T}]^T = [a_1^{(t)}, a_2^{(t)}, \dots, a_v^{(t)}, a_1^{(t-1)}, a_2^{(t-1)}, \dots, a_v^{(t-1)}]^T, \quad (5)
 \end{aligned}$$

$$\begin{aligned}\hat{\Lambda} &= [\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_v, \hat{\lambda}_{v+1}, \hat{\lambda}_{v+2}, \dots, \hat{\lambda}_{2v}]^T \\ &= [\hat{\mathbf{a}}^{(t)T}, \hat{\mathbf{a}}^{(t-1)T}]^T = [\hat{a}_1^{(t)}, \hat{a}_2^{(t)}, \dots, \hat{a}_v^{(t)}, \hat{a}_1^{(t-1)}, \hat{a}_2^{(t-1)}, \dots, \hat{a}_v^{(t-1)}]^T,\end{aligned}\quad (6)$$

where  $\mathbf{a}^{(t-1)} = [a_1^{(t-1)}, a_2^{(t-1)}, \dots, a_v^{(t-1)}]^T$  and  $\hat{\mathbf{a}}^{(t-1)} = [\hat{a}_1^{(t-1)}, \hat{a}_2^{(t-1)}, \dots, \hat{a}_v^{(t-1)}]^T$  are the linear prediction coefficients corresponding to the unquantized and quantized LSF parameters at the  $(t-1)$ th frame index. In the whole paper, the superscripts  $t$  and  $t-1$  are used to denote the frame indices.

To design a quantizer that directly minimizes the overall MWLSD is quite difficult due to the high computational complexity of the MWLSD measure. Hence, simple measures such as quadratically weighted distortion (QWD) measure or weighted mean square error (WMSE) measure are required. In this paper, a theoretical analysis based on the MWLSD measure is performed to derive the QWD measure which theoretically approximates the MWLSD measure to any degree and has much lower computational complexity. A technique which enables the QWD measure to quantize only one LSF-parameter vector rather than two successive LSF-parameter vectors is also proposed. A simpler measure, called simplified QWD — SQWD, is also derived by simplifying the QWD measure to further reduce its computational complexity for practical applications. The computational complexity of the SQWD measure is almost equal to that of the WMSE measure. From the results in our experiments, it is demonstrated that the quantizers trained by using the weights obtained from the MWLSD measure intuitively perform better than those trained by using the weights obtained from the static measure such as the LSD measure or WLSLSD measure, since the MWLSD measure considers the spectral distortion and the spectral-dynamics distortion between unquantized spectra and quantized spectra at the same time.

This paper is organized as follows. Section 2 derives an approximate solution of the MWLSD measure for the LSF VQ schemes and results in a QWD measure and a SQWD measure in training and quantizing. In Section 3, two sets of experimental LSF VQ results, and the experimental testing results of applying the proposed technique to the LSF VQ of a CELP coder are presented. Finally, conclusions are drawn in Section 4.

## 2 Theoretical Analysis of the MWLSD Measure

In this section, the weights used in QWD measure are derived in Section 2.1 and the experimentally determined weights used in SQWD measure are given in Section 2.2. An algorithm of computing the weights used in the QWD and SQWD measures is presented in Section 2.3.

## 2.1 Replacement of the MWLSD measure by the QWD measure

In this subsection, we shall study the replacement of the MWLSD measure by a quadratically weighted distortion (QWD) measure for computational simplicity. We first introduce a theorem in [8] showing that any continuously differentiable distortion function which satisfies some properties can be approximated by a QWD function to any desired degree for short distances. The assumption of short distance is hold for the approximation of spectral dynamic distortion measures throughout the paper.

**Theorem 2.1** *Let  $\mathbf{x}$  be an  $n$  dimensional vector and assume that a vector quantizer maps  $\mathbf{x}$  to an output vector  $\hat{\mathbf{x}}$ , and  $d(\mathbf{x}, \hat{\mathbf{x}})$  is a continuously differentiable distortion function having the following properties:*

- $d(\mathbf{x}, \hat{\mathbf{x}}) \geq 0$  with equality holding only if  $\mathbf{x} = \hat{\mathbf{x}}$ .
- $d(\mathbf{x}, \hat{\mathbf{x}})$  having continuous partial derivatives of third order almost everywhere.
- The second-order derivatives of  $d(\mathbf{x}, \hat{\mathbf{x}})$ ,  $\frac{\partial^2 d(\mathbf{x}, \hat{\mathbf{x}})}{\partial x_j \partial x_k}$ , for all  $j, k$ , being positive definite almost everywhere.

Then, this  $d(\mathbf{x}, \hat{\mathbf{x}})$  measure can be approximated exactly by a QWD measure, i.e.,

$$d(\mathbf{x}, \hat{\mathbf{x}}) = \frac{1}{2}(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{D}(\mathbf{x})(\mathbf{x} - \hat{\mathbf{x}}) \quad \text{for small distances,} \quad (7)$$

where  $\mathbf{D}(\mathbf{x})$  is an  $n$  by  $n$  dimensional matrix with its  $(j, k)$ th element defined by

$$D_{j,k}(\mathbf{x}) = \left. \frac{\partial^2 d(\mathbf{x}, \hat{\mathbf{x}})}{\partial x_j \partial x_k} \right|_{\hat{\mathbf{x}} = \mathbf{x}}, \quad (8)$$

and is denoted as the “sensitivity matrix,” since its elements represent the relative sensitivity of quantizing the various parameters.

*Proof:* See Li et al. [9].

In general, the sensitivity matrix of the QWD measure defined in Eq. (7) is not diagonal. The QWD measure provides a simple method to computing the centroid of a quadratic measure, whereas it may be impossible to efficiently compute the centroid of the “true” distortion measure (such as the MWLSD or WLSLSD measure). Thus this allows the quantizer to be built using the generalized Lloyd or LBG algorithms [10].

It is readily apparent that Theorem 2.1 holds for the MWLSD measure in Eq. (4) only if the weighting function  $B(w)$  has continuous third-order derivatives, since the MWLSD measure satisfies the properties required by  $d(\mathbf{x}, \hat{\mathbf{x}})$  in Theorem 2.1. Hence, we can use the QWD measure defined in Eq. (7) to replace the MWLSD measure in designing an LSF quantizer. To design an LSF quantizer based on the QWD measure, we need to determine the sensitivity matrix in Eq. (7). This issue is what we focus on in the rest of this subsection. For analysis convenience, we rewrite Eq. (4) as follows:

$$D_{\text{MWLSD}}(\Lambda, \hat{\Lambda}) = \frac{\alpha}{2\pi B_0} \int_{-\pi}^{\pi} B^2(w) \left[ \ln(A^{(t)}(w)) - \ln(\hat{A}^{(t)}(w)) \right]^2 + B^2(w) \left[ (\ln(A^{(t)}(w)) - \ln(A^{(t-1)}(w))) - (\ln(\hat{A}^{(t)}(w)) - \ln(\hat{A}^{(t-1)}(w))) \right]^2 + B^2(w) \left[ \ln(A^{(t-1)}(w)) - \ln(\hat{A}^{(t-1)}(w)) \right]^2 dw, \quad (9)$$

where

$$\begin{aligned} \alpha &= (10/\ln(10))^2, \\ A^{(t)}(w) &= \left| 1 - \sum_{k=1}^v a_k^{(t)} e^{-jwk} \right|, \\ \hat{A}^{(t)}(w) &= \left| 1 - \sum_{k=1}^v \hat{a}_k^{(t)} e^{-jwk} \right|, \\ A^{(t-1)}(w) &= \left| 1 - \sum_{k=1}^v a_k^{(t-1)} e^{-jwk} \right|, \end{aligned}$$

and

$$\hat{A}^{(t-1)}(w) = \left| 1 - \sum_{k=1}^v \hat{a}_k^{(t-1)} e^{-jwk} \right|.$$

To derive the sensitivity matrix for LSF parameters based on the MWLSD measure, we first find out the sensitivity matrix for LPC parameters based on the MWLSD measure and then apply Jacobian matrix of transformation, which transforms LSF parameters to LPC parameters, to it to obtain the sensitivity matrix for LSF parameters based on the MWLSD measure.

### 2.1.1 Sensitivity matrix for LPC parameters based on the MWLSD measure

The following theorem gives the exact form of the sensitivity matrix for LPC parameters based on the MWLSD measure defined in Eq. (9).

**Theorem 2.2** *Let  $h^{(t)}[n]$ ,  $h^{(t-1)}[n]$ , and  $b_s[n]$  denote the impulse responses of the discrete-time filters  $1/A^{(t)}(w)$ ,  $1/A^{(t-1)}(w)$ , and  $B_s(w) \triangleq B^2(w)$  in Eq. (9), respectively. Notice that  $h^{(t)}[n]$  and  $h^{(t-1)}[n]$  are both causal signals and  $b_s[n]$  is a symmetric and noncausal signal. Then, the elements of the*

sensitivity matrix for LPC parameters based on the MWLSD measure are

$$\left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{\lambda}_m \partial \hat{\lambda}_n} \right|_{\hat{\Lambda}=\Lambda} = R_{Am}(m, n), 1 \leq m, n \leq 2v, \quad (10)$$

where

$$R_{Am}(m, n) = \begin{cases} R_{SM}^{(t)}(k, l), & k = m, l = n, 1 \leq m, n \leq v, \\ R_{SM}^{(t-1)}(k, l), & k = m - v, l = n - v, v + 1 \leq m, n \leq 2v, \\ R_{CM}^{(t, t-1)}(k, l), & k = m, l = n - v, 1 \leq m \leq v, v + 1 \leq n \leq 2v, \\ R_{DM}^{(t-1, t)}(k, l), & k = m - v, l = n, v + 1 \leq m \leq 2v, 1 \leq n \leq v, \end{cases} \quad (11)$$

where  $R_{SM}^{(t)}(k, l)$ ,  $R_{SM}^{(t-1)}(k, l)$ ,  $R_{CM}^{(t, t-1)}(k, l)$ , and  $R_{DM}^{(t-1, t)}(k, l)$  are defined as

$$R_{SM}^{(t)}(k, l) = \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda}=\Lambda} = 8\alpha R_{sm}^{(t)}(k, l), \quad 1 \leq k, l \leq v, \quad (12)$$

$$R_{SM}^{(t-1)}(k, l) = \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t-1)} \partial \hat{a}_l^{(t-1)}} \right|_{\hat{\Lambda}=\Lambda} = 8\alpha R_{sm}^{(t-1)}(k, l), \quad 1 \leq k, l \leq v, \quad (13)$$

$$R_{CM}^{(t, t-1)}(k, l) = \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t-1)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda}=\Lambda} = -4\alpha R_{cm}^{(t, t-1)}(k, l), \quad 1 \leq k, l \leq v, \quad (14)$$

$$R_{DM}^{(t-1, t)}(k, l) = \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t)} \partial \hat{a}_l^{(t-1)}} \right|_{\hat{\Lambda}=\Lambda} = -4\alpha R_{cm}^{(t, t-1)}(l, k), \quad 1 \leq k, l \leq v. \quad (15)$$

In Eq. (12),

$$R_{sm}^{(t)}(k, l) = \sum_{n=-\infty}^{\infty} b_s[n] (r_c^{(t)}[(k+l)-n] + r^{(t)}[(k-l)-n]), \quad (16)$$

where  $r^{(t)}[n]$  is the traditional autocorrelation function of the impulse response  $h^{(t)}[n]$ , i.e.,

$$r^{(t)}[n] = \sum_{m=-\infty}^{\infty} h^{(t)}[m] h^{(t)}[m+n], \quad (17)$$

and  $r_c^{(t)}[n]$  is the convolution of  $h^{(t)}[-n]$  with  $h^{(t)}[-n]$ , i.e.,

$$r_c^{(t)}[n] = \sum_{m=0}^{\infty} h^{(t)}[-m]h^{(t)}[-(n-m)] = \sum_{m=n}^0 h^{(t)}[-m]h^{(t)}[m-n], \quad n \leq 0. \quad (18)$$

In Eq. (13),  $R_{sm}^{(t-1)}(k, l)$  is exactly equal to the value of  $R_{sm}^{(t)}(k, l)$  computed at the  $(t-1)$ th frame index. In Eqs. (14) and (15),

$$R_{cm}^{(t,t-1)}(k, l) = \sum_{n=-\infty}^{\infty} b_s[n] \left( r_e^{(t,t-1)}[(k+l)-n] + r_d^{(t,t-1)}[(k-l)-n] \right), \quad (19)$$

where  $r_d^{(t,t-1)}[n]$  is the crosscorrelation function of the impulse response  $h^{(t)}[n]$  with  $h^{(t-1)}[n]$ , i.e.,

$$r_d^{(t,t-1)}[n] = \sum_{m=0}^{\infty} h^{(t)}[m]h^{(t-1)}[m-n], \quad (20)$$

and  $r_e^{(t,t-1)}[n]$  is the convolution of  $h^{(t)}[-n]$  with  $h^{(t-1)}[-n]$ , i.e.,

$$r_e^{(t,t-1)}[n] = \sum_{m=0}^{\infty} h^{(t)}[-m]h^{(t-1)}[-(n-m)] = \sum_{m=n}^0 h^{(t)}[-m]h^{(t-1)}[m-n], \quad n \leq 0. \quad (21)$$

*Proof:* See Appendix A.

From the results in Theorem 2.2, the sensitivity matrix for LPC parameters based on the MWLSD measure can be written as a  $2v$  by  $2v$  dimensional matrix  $\mathbf{D}_\lambda(\Lambda)$ , given as

$$\mathbf{D}_\lambda(\Lambda) = \left[ \begin{array}{c|c} \mathbf{R}_{sm}^{(t)} & \mathbf{R}_{cm}^{(t,t-1)} \\ \hline \mathbf{R}_{cm}^{(t,t-1)T} & \mathbf{R}_{sm}^{(t-1)} \end{array} \right] \triangleq \left[ \begin{array}{ccc|ccc} R_{sm}^{(t)}(1,1) & \cdots & R_{sm}^{(t)}(1,v) & R_{cm}^{(t,t-1)}(1,1) & \cdots & R_{cm}^{(t,t-1)}(1,v) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{sm}^{(t)}(v,1) & \cdots & R_{sm}^{(t)}(v,v) & R_{cm}^{(t,t-1)}(v,1) & \cdots & R_{cm}^{(t,t-1)}(v,v) \\ \hline R_{cm}^{(t,t-1)}(1,1) & \cdots & R_{cm}^{(t,t-1)}(v,1) & R_{sm}^{(t-1)}(1,1) & \cdots & R_{sm}^{(t-1)}(1,v) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ R_{cm}^{(t,t-1)}(1,v) & \cdots & R_{cm}^{(t,t-1)}(v,v) & R_{sm}^{(t-1)}(v,1) & \cdots & R_{sm}^{(t-1)}(v,v) \end{array} \right] \quad (22)$$



A surprising result can be obtained from Eqs. (12) and (13) in Theorem 2.2. It is

$$R_{sm}^{(t)}(i, j) = 2(\mathbf{D}_A^{(t)})_{i,j} \triangleq 2 \left. \frac{\partial^2 D_{\text{WLS D}}(\mathbf{a}^{(t)}, \hat{\mathbf{a}}^{(t)})}{\partial \hat{a}_i^{(t)} \partial \hat{a}_j^{(t)}} \right|_{\hat{\mathbf{a}}^{(t)} = \mathbf{a}^{(t)}}, \quad (23)$$

$$R_{sm}^{(t-1)}(i, j) = 2(\mathbf{D}_A^{(t-1)})_{i,j} \triangleq 2 \left. \frac{\partial^2 D_{\text{WLS D}}(\mathbf{a}^{(t-1)}, \hat{\mathbf{a}}^{(t-1)})}{\partial \hat{a}_i^{(t-1)} \partial \hat{a}_j^{(t-1)}} \right|_{\hat{\mathbf{a}}^{(t-1)} = \mathbf{a}^{(t-1)}}, \quad (24)$$

where  $D_{\text{WLS D}}(\mathbf{a}^{(t)}, \hat{\mathbf{a}}^{(t)})$  is defined in Eq. (2), and  $\mathbf{D}_A^{(t)}$  and  $\mathbf{D}_A^{(t-1)}$  are the two sensitivity matrices for the LPC parameters based on the WLS D measure defined in [4]. The matrix  $\mathbf{D}_\lambda(\Lambda)$  has two properties listed in the following:

- (a)  $\mathbf{D}_\lambda(\Lambda)$  obviously is a symmetric matrix from Eq. (22), since  $\mathbf{R}_{sm}^{(t)}$  and  $\mathbf{R}_{sm}^{(t-1)}$  are symmetric matrices shown in [4].
- (b)  $\mathbf{D}_\lambda(\Lambda)$  is positive-definite almost everywhere from Theorem 2.1.

Property (a) is useful for saving the computation time in the calculation of  $\mathbf{D}_\lambda(\Lambda)$ .

### 2.1.2 Sensitivity matrix for LSF parameters based on the MWLS D measure

In order to define LSF parameters, the inverse filter,  $A^{(t)}(z)$ , is used to construct two polynomials:

$$\bar{P}^{(t)}(z) = A^{(t)}(z) + z^{-(v+1)} A^{(t)}(z^{-1}) \quad \text{and} \quad \bar{Q}^{(t)}(z) = A^{(t)}(z) - z^{-(v+1)} A^{(t)}(z^{-1}). \quad (25)$$

The roots of these polynomials are usually called LSF parameters, denoted as  $\mathbf{w}^{(t)} = [w_1^{(t)}, w_2^{(t)}, \dots, w_v^{(t)}]^T$ . Some important properties of LSF parameters are detailed in [2, 3].

Denote the transformation of LPC vectors  $\Lambda = [\mathbf{a}^{(t)T}, \mathbf{a}^{(t-1)T}]^T$  to the LSF parameters set  $\Theta$  by the function  $\Theta(\Lambda)$ , and the reverse transformation by  $\Lambda(\Theta)$ , where  $\Theta$  represents the two successive vectors of the LSF parameters at the  $t$ th and  $(t-1)$ th frame indices, denoted as  $\Theta = [\theta_1, \dots, \theta_v, \theta_{v+1}, \dots, \theta_{2v}]^T = [\mathbf{w}^{(t)T}, \mathbf{w}^{(t-1)T}]^T = [w_1^{(t)}, \dots, w_v^{(t)}, w_1^{(t-1)}, \dots, w_v^{(t-1)}]^T$ , where  $\mathbf{w}^{(t-1)}$  is the LSF parameters corresponding to  $\mathbf{a}^{(t-1)}$  at the  $(t-1)$ th frame index. The element of the sensitivity matrix  $\mathbf{D}_\theta(\Theta)$  for the LSF parameters based on the MWLS D

measure is then given by

$$\begin{aligned}
(\mathbf{D}_\theta(\Theta))_{k,l} &= \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda(\Theta), \Lambda(\hat{\Theta}))}{\partial \hat{\theta}_k \partial \hat{\theta}_l} \right|_{\hat{\Theta} = \Theta} \\
&= \sum_{m=1}^{2v} \sum_{n=1}^{2v} \frac{\partial \lambda_m(\hat{\Theta})}{\partial \hat{\theta}_k} \frac{\partial \lambda_n(\hat{\Theta})}{\partial \hat{\theta}_l} \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{\lambda}_m \partial \hat{\lambda}_n} \right|_{\hat{\Theta} = \Theta, \hat{\Lambda} = \Lambda(\Theta)} \\
&= (\mathbf{J}_\theta^T(\Theta) \mathbf{D}_\lambda(\Lambda) \mathbf{J}_\theta(\Theta))_{k,l}, \tag{26}
\end{aligned}$$

where  $\hat{\Theta} = [\hat{\theta}_1, \dots, \hat{\theta}_v, \hat{\theta}_{v+1}, \dots, \hat{\theta}_{2v}]^T$  denotes the quantized  $\Theta$ ,  $\mathbf{D}_\lambda(\Lambda)$  is given in Eq. (22), and  $\mathbf{J}_\theta(\Theta)$  is the Jacobian matrix of the transform  $\Lambda(\Theta)$ , which has its  $(j, k)$ th element defined by

$$(\mathbf{J}_\theta(\Theta))_{j,k} = \frac{\partial \lambda_j(\Theta)}{\partial \theta_k}. \tag{27}$$

The elements  $(\mathbf{J}_\theta(\Theta))_{j,k}$  for  $1 \leq j \leq v$ ,  $v+1 \leq k \leq 2v$ , and for  $v+1 \leq j \leq 2v$ ,  $1 \leq k \leq v$  are zero since the LPC parameters  $\mathbf{a}^{(t)}$  and  $\mathbf{a}^{(t-1)}$  are only the function of the LSF parameters  $\mathbf{w}^{(t)}$  and  $\mathbf{w}^{(t-1)}$ , respectively. This implies that the Jacobian matrix  $\mathbf{J}_\theta(\Theta)$  is block diagonal, written as

$$\mathbf{J}_\theta(\Theta) = \left[ \begin{array}{c|c} \mathbf{J}^{(t)}(\mathbf{w}^{(t)}) & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{J}^{(t-1)}(\mathbf{w}^{(t-1)}) \end{array} \right], \tag{28}$$

where  $\mathbf{J}^{(t)}(\mathbf{w}^{(t)})$  denotes the Jacobian matrix of the transformation from the LSF parameters  $\mathbf{w}^{(t)}$  to the LPC parameters  $\mathbf{a}^{(t)}$  at the  $t$ th frame index,  $\mathbf{J}^{(t-1)}(\mathbf{w}^{(t-1)})$  is the matrix of  $\mathbf{J}^{(t)}(\mathbf{w}^{(t)})$  computed at the  $(t-1)$ th frame index, and  $\mathbf{O}$  is a  $v$  by  $v$  zero matrix. By substituting Eq. (22) into Eq. (26) and using Eq. (28), the sensitivity matrix  $\mathbf{D}_\theta(\Theta)$  can be rewritten as

$$\begin{aligned}
\mathbf{D}_\theta(\Theta) &= \mathbf{J}_\theta^T(\Theta) \mathbf{D}_\lambda(\Lambda) \mathbf{J}_\theta(\Theta) \\
&= \left[ \begin{array}{c|c} \mathbf{J}^{(t)T}(\mathbf{w}^{(t)}) \mathbf{R}_{sm}^{(t)} \mathbf{J}^{(t)}(\mathbf{w}^{(t)}) & \mathbf{J}^{(t)T}(\mathbf{w}^{(t)}) \mathbf{R}_{cm}^{(t,t-1)} \mathbf{J}^{(t-1)}(\mathbf{w}^{(t-1)}) \\ \hline \mathbf{J}^{(t-1)T}(\mathbf{w}^{(t-1)}) \mathbf{R}_{sm}^{(t,t-1)T} \mathbf{J}^{(t)}(\mathbf{w}^{(t)}) & \mathbf{J}^{(t-1)T}(\mathbf{w}^{(t-1)}) \mathbf{R}_{sm}^{(t-1)} \mathbf{J}^{(t-1)}(\mathbf{w}^{(t-1)}) \end{array} \right] \\
&\triangleq \left[ \begin{array}{c|c} \mathbf{K}_{sm}^{(t)} & \mathbf{K}_{cm}^{(t,t-1)} \\ \hline \mathbf{K}_{cm}^{(t,t-1)} & \mathbf{K}_{sm}^{(t-1)} \end{array} \right]. \tag{29}
\end{aligned}$$

It is interesting to note that the two matrices  $\mathbf{K}_{sm}^{(t)}$  and  $\mathbf{K}_{sm}^{(t-1)}$  in Eq. (29) have the following property:

$$\mathbf{K}_{sm}^{(t)}(i, j) = 2(\mathbf{D}_w^{(t)})_{i,j} \triangleq 2 \left. \frac{\partial^2 D_{\text{WLS}}(\mathbf{w}^{(t)}, \hat{\mathbf{w}}^{(t)})}{\partial \hat{w}_i^{(t)} \partial \hat{w}_j^{(t)}} \right|_{\hat{\mathbf{w}}^{(t)} = \mathbf{w}^{(t)}}, \tag{30}$$

$$\mathbf{K}_{sm}^{(t-1)}(i, j) = 2(\mathbf{D}_w^{(t-1)})_{i,j} \triangleq 2 \left. \frac{\partial^2 D_{\text{WLS}}(\mathbf{w}^{(t-1)}, \hat{\mathbf{w}}^{(t-1)})}{\partial \hat{w}_i^{(t-1)} \partial \hat{w}_j^{(t-1)}} \right|_{\hat{\mathbf{w}}^{(t-1)} = \mathbf{w}^{(t-1)}} \tag{31}$$

where  $\hat{\mathbf{w}}^{(t)}$  and  $\hat{\mathbf{w}}^{(t-1)}$  are the quantized LSF vectors corresponding to  $\mathbf{w}^{(t)}$  and  $\mathbf{w}^{(t-1)}$ , respectively,  $D_{\text{WLS D}}(\mathbf{w}^{(t)}, \hat{\mathbf{w}}^{(t)})$  is defined in Eq. (2), and  $\mathbf{D}_w^{(t)}$  and  $\mathbf{D}_w^{(t-1)}$  are the two sensitivity matrices for the LSF parameters based on the WLS D measure defined in [4]. This therefore results in that the sensitivity matrix  $\mathbf{D}_\theta(\Theta)$  given in Eq. (29) can be used in the QWD measure to approximate the MWLS D measure in LSF VQ. When the QWD measure with the sensitivity matrix  $\mathbf{D}_\theta(\Theta)$  is used in the LSF VQ during training and quantizing, the quantized LSF parameters of two successive frames,  $\hat{\mathbf{w}}^{(t)}$  and  $\hat{\mathbf{w}}^{(t-1)}$ , should be found out, given as

$$\begin{aligned}
 d(\Theta, \hat{\Theta}) &= d(\mathbf{w}^{(t)}, \mathbf{w}^{(t-1)}, \hat{\mathbf{w}}^{(t)}, \hat{\mathbf{w}}^{(t-1)}) & (32) \\
 &= \begin{bmatrix} \mathbf{e}^{(t)T}, \mathbf{e}^{(t-1)T} \end{bmatrix} \mathbf{D}_\theta(\Theta) \begin{bmatrix} \mathbf{e}^{(t)T}, \mathbf{e}^{(t-1)T} \end{bmatrix}^T \\
 &= \mathbf{e}^{(t)T} \mathbf{K}_{sm}^{(t)} \mathbf{e}^{(t)} + 2\mathbf{e}^{(t)T} \mathbf{K}_{cm}^{(t,t-1)} \mathbf{e}^{(t-1)} + \mathbf{e}^{(t-1)T} \mathbf{K}_{sm}^{(t-1)} \mathbf{e}^{(t-1)},
 \end{aligned}$$

where  $\mathbf{e}^{(t)}$  and  $\mathbf{e}^{(t-1)}$  are both  $v$ -dimensional error vectors, defined as  $\mathbf{e}^{(t)} = \mathbf{w}^{(t)} - \hat{\mathbf{w}}^{(t)}$  and  $\mathbf{e}^{(t-1)} = \mathbf{w}^{(t-1)} - \hat{\mathbf{w}}^{(t-1)}$ , respectively. However, on the other hand, the QWD measure can be used to find out only one quantized LSF vector  $\mathbf{w}^{(t)}$  during the search procedure in the LSF VQ, if the error vector at the  $(t-1)$ th frame index,  $\mathbf{e}^{(t-1)}$ , is assumed to be constant. This assumption is reasonable since the error vector  $\mathbf{e}^{(t-1)}$  is always determined by the known unquantized and quantized LSF parameters at the  $(t-1)$ th frame index with a zero error vector starting at the first frame, i.e.,  $\mathbf{e}^{(0)} = 0$ . That is,  $\mathbf{e}^{(t-1)}$  in Eq. (32) can be assumed to be a constant vector in quantizing the LSF parameters  $\mathbf{w}^{(t)}$ . In such a case, Eq. (32) can be rewritten as

$$d(\mathbf{w}^{(t)}, \hat{\mathbf{w}}^{(t)}) = \mathbf{e}^{(t)T} \mathbf{K}_{sm}^{(t)} \mathbf{e}^{(t)} + 2\mathbf{e}^{(t)T} \mathbf{e}_s + \mathbf{e}_u, \quad (33)$$

where  $\mathbf{e}_s = \mathbf{K}_{cm}^{(t,t-1)} \mathbf{e}^{(t-1)}$  and  $\mathbf{e}_u = \mathbf{e}^{(t-1)T} \mathbf{K}_{sm}^{(t-1)} \mathbf{e}^{(t-1)}$ .

The QWD measure defined in Eq. (33) with the constant error vector  $\mathbf{e}^{(t-1)}$  will be used in this paper. The reasons for using the QWD measure defined in Eq. (33) are given as follows. The bit rate required to quantize one LSF vector  $\mathbf{w}^{(t)}$  by using Eq. (33) is less than that required to quantize two successive LSF vectors  $\Theta$  by using Eq. (32) for achieving the same quantization performance, since the dimension of  $\Theta$  is twice larger than that of  $\mathbf{w}^{(t)}$ . In addition, the computational requirement for quantizing one LSF vector  $\mathbf{w}^{(t)}$  by using Eq. (33) is much lower than that for quantizing two successive LSF vectors  $\Theta$  by using Eq. (32), because  $\mathbf{e}_s$  and  $\mathbf{e}_u$  at the right-hand side of Eq. (33) are constant during the search procedure in the LSF VQ once  $\mathbf{e}_s$  and  $\mathbf{e}_u$  are beforehand computed.

## 2.2 Approximation of the QWD measure by the SQWD measure

Computing the full QWD measure for each vector in a codebook during quantization may be still too complex in some applications; therefore, the use of simpler distortion measure is desirable. The major part of the computational complexity in the LSF VQ by using the QWD measure defined in Eq. (33) obviously comes from the first term at the right-hand side of Eq. (33); i.e., the full matrix  $\mathbf{K}_{sm}^{(t)}$ . If the weighting matrix  $\mathbf{K}_{sm}^{(t)}$  is diagonal, then the computational complexity of the QWD measure can be greatly reduced during the search procedure in the LSF VQ. However, according to the result in [4],  $\mathbf{K}_{sm}^{(t)}$  is generally not a diagonal matrix.

The approximation to the QWD measure can be achieved by taking the main diagonal elements of the three matrices  $\mathbf{K}_{sm}^{(t)}$ ,  $\mathbf{K}_{cm}^{(t,t-1)}$ , and  $\mathbf{K}_{sm}^{(t-1)}$ , and setting the other elements of these matrices to zero. This approximate measure is named *simplified quadratically weighted distortion* (SQWD) measure and written as

$$d(\mathbf{w}^{(t)}, \hat{\mathbf{w}}^{(t)}) = \mathbf{e}^{(t)T} \mathbf{K}_{smd}^{(t)} \mathbf{e}^{(t)} + 2\mathbf{e}^{(t)T} \mathbf{e}_{sd} + \mathbf{e}_{ud}, \quad (34)$$

where  $\mathbf{K}_{smd}^{(t)}$ ,  $\mathbf{e}_{sd}$  and  $\mathbf{e}_{ud}$  are defined as

$$\mathbf{K}_{smd}^{(t)} = \text{diag}(\mathbf{K}_{sm}^{(t)}), \quad \mathbf{e}_{sd} = \mathbf{K}_{cmd}^{(t,t-1)} \mathbf{e}^{(t-1)} \quad \text{and} \quad \mathbf{e}_{ud} = \mathbf{e}^{(t-1)T} \mathbf{K}_{smd}^{(t-1)} \mathbf{e}^{(t-1)}, \quad (35)$$

where the operator  $\text{diag}(\mathbf{S})$  is to set the elements of matrix  $\mathbf{S}$  to zero except the main diagonal elements, and the two matrices  $\mathbf{K}_{smd}^{(t-1)}$  and  $\mathbf{K}_{cmd}^{(t,t-1)}$  are defined as  $\mathbf{K}_{smd}^{(t-1)} = \text{diag}(\mathbf{K}_{sm}^{(t-1)})$  and  $\mathbf{K}_{cmd}^{(t,t-1)} = \text{diag}(\mathbf{K}_{cm}^{(t,t-1)})$ . The diagonal elements of  $\mathbf{K}_{smd}^{(t)}$ ,  $\mathbf{K}_{smd}^{(t-1)}$ , and  $\mathbf{K}_{cmd}^{(t,t-1)}$  used in the SQWD measure are named *diagonal approximation* (DA) weights. The computational complexity of the SQWD measure is almost equal to that of WMSE measure since the major part of the computational complexity in LSF VQ by using the SQWD measure defined in Eq. (34) comes from  $\mathbf{e}^{(t)T} \mathbf{K}_{smd}^{(t)} \mathbf{e}^{(t)}$ , which is a form of WMSE measure, and  $\mathbf{e}_{sd}$  and  $\mathbf{e}_{ud}$  at the right-hand side of Eq. (34) are constant during the search procedure in the LSF VQ once  $\mathbf{e}_{sd}$  and  $\mathbf{e}_{ud}$  are beforehand computed.

The motivation for taking this approximation comes from the result in [8] saying that the diagonal elements of the sensitivity matrix are the optimal weights used to optimally approximate the full quadratic distortion measure by a simple WMSE measure. The QWD measure defined in Eq. (33) surely is different from the full quadratic distortion measure (since the second term and third term at the right-hand side of Eq. (33) are generally nonzero); hence, the result in [8] cannot be directly applied to this QWD measure. Even though, the SQWD measure is a reasonable distortion measure used to approximate the

QWD measure for the following two reasons. First, the information of spectral dynamics between two successive frames can be significantly incorporated into the SQWD measure by using  $\mathbf{K}_{cmd}^{(t,t-1)}$  of Eq. (35). If only the first and third terms at the right-hand side of Eq. 34 are used in the SQWD measure, then this measure will become a static measure without taking any information of spectral dynamics into account, and will be further reduced to a WMSE measure since the error vector  $\mathbf{e}^{(t-1)}$  is constant during quantization procedure. Second, the SQWD measure is always nonnegative, i.e., the following matrix is nonnegative definite:

$$\mathbf{K}_{\text{SQWD}} \triangleq \left[ \begin{array}{c|c} \mathbf{K}_{smd}^{(t)} & \mathbf{K}_{cmd}^{(t,t-1)} \\ \hline \mathbf{K}_{cmd}^{(t,t-1)T} & \mathbf{K}_{smd}^{(t-1)} \end{array} \right]. \quad (36)$$

The proof of the nonnegative-definite property of  $\mathbf{K}_{\text{SQWD}}$  is given in Appendix B.

### 2.3 Computational implementation

In order to compute the sensitivity matrix for LSF parameters based on the MWLSD measure given by Eq. (29), we only need to determine the sensitivity matrix for LPC parameters,  $\mathbf{D}_\lambda(\Lambda)$  in Eq. (22), and the Jacobian matrix of the transformation from LSF parameters to LPC parameters,  $\mathbf{J}^{(t)}(\mathbf{w}^{(t)})$  in Eq. (28), since  $\mathbf{J}^{(t)}(\mathbf{w}^{(t-1)})$  is exactly equal to the matrix of  $\mathbf{J}^{(t)}(\mathbf{w}^{(t)})$  computed at the  $(t-1)$ th frame index. The efficient computation algorithm of  $\mathbf{J}^{(t)}(\mathbf{w}^{(t)})$  can be found in [8]. Thus, what remains is the computation of  $\mathbf{D}_\lambda(\Lambda)$ .

Only the terms,  $R_{sm}^{(t)}(k, l)$  and  $R_{cm}^{(t,t-1)}(k, l)$  for  $1 \leq k, l \leq v$ , are required to calculate in computing  $\mathbf{D}_\lambda(\Lambda)$  from Theorem 2.2. The two problems existing in computing  $R_{sm}^{(t)}(k, l)$  and  $R_{cm}^{(t,t-1)}(k, l)$  are the same as those existing in computing the sensitivity matrix for LPC parameters based on the WLS measure [4]. One problem is that the impulse response  $h^{(t)}[n]$  of the LPC filter  $1/A^{(t)}(z)$  is infinite. The other is that  $b_s[n]$  may not be a finite sequence. The first problem can be solved by assuming that  $h^{(t)}[n]$  is zero after some large time index  $L$ . The other is solved by truncating  $b_s[n]$  through applying the autocorrelation sequence of a FIR window whose center locating at time index 0 if  $b_s[n]$  is an infinite sequence. The detailed description for solving these two problems can be found in [4].

With the above two problems being fixed,  $R_{sm}^{(t)}(k, l)$  and  $R_{cm}^{(t,t-1)}(k, l)$ , can be computed as follows. Without loss of generality, assume that  $h^{(t)}[n]$  and  $h^{(t-1)}[n]$  are both zero after some large time index  $L$ , and  $b_s[n]$  is zero outside the range  $-N \leq n \leq N$ . We then can rewrite Eqs. (17)–(18) and

Eqs. (20)–(21) as

$$r^{(t)}[k] = r^{(t)}[-k] = \begin{cases} \sum_{n=0}^{L-k} h^{(t)}[n]h^{(t)}[n+k], & k = 0, \dots, v \\ \sum_{n=1} a_n^{(t)} r^{(t)}[k-n], & k = v+1, \dots, N+v-1, \end{cases} \quad (37)$$

$$r_c^{(t)}[k] = \sum_{n=k}^0 h^{(t)}[-n]h^{(t)}[n-k], \quad k = 0, -1, \dots, -N+1, \quad (38)$$

and

$$r_d^{(t,t-1)}[k] = \sum_{n=0}^{L-k} h^{(t)}[n]h^{(t-1)}[n-k], \quad (39)$$

$$k = -(N+v-1), \dots, -1, 0, 1, \dots, (N+v-1),$$

$$r_e^{(t,t-1)}[k] = \sum_{n=k}^0 h^{(t)}[-n]h^{(t-1)}[n-k], \quad k = 0, -1, \dots, -N+1. \quad (40)$$

The elements of  $R_{sm}^{(t)}(k, l)$  and  $R_{cm}^{(t,t-1)}(k, l)$  can thus be easily computed by the following equations:

$$R_{sm}^{(t)}(k, l) = \sum_{n=-N}^N b_s[n]r^{(t)}[(k-l)-n] + \sum_{n=k+l}^N b_s[n]r_c^{(t)}[(k+l)-n], \quad (41)$$

$$R_{cm}^{(t,t-1)}(k, l) = \sum_{n=-N}^N b_s[n]r_d^{(t,t-1)}[(k-l)-n] + \sum_{n=k+l}^N b_s[n]r_e^{(t,t-1)}[(k+l)-n]. \quad (42)$$

This completes the computation implementation of the sensitivity matrix for LPC parameters,  $\mathbf{D}_\lambda(\Lambda)$ , based on the MWLSD measure. Finally, the sensitivity matrix for LSF parameters,  $\mathbf{D}_\theta(\Theta)$ , based on the MWLSD measure can be determined by Eq. (29), and the DA weights are the main diagonal elements of the matrices  $\mathbf{K}_{sm}^{(t)}$  and  $\mathbf{K}_{cm}^{(t,t-1)}$  given by Eq. (34).

We presently evaluate the computational complexity for calculating the DA weights as follows. The impulse response  $h^{(t)}[n]$ , for  $0 \leq n \leq L$ , requires  $vL$  multiply-adds (multiply-add meaning a multiplication operation followed by an addition operation), where  $v$  denotes the order of LPC model,  $L$  denotes the maximum time index of impulse response  $h^{(t)}[n]$ . Equations (37), (38), (39), and (40) require  $(L+1-v/2)(v+1) + v(N-1)$ ,  $N(N+1)/2$ ,  $[(L+1) - (N+v-1)/2](N+v)$ , and  $N(N+1)/2$  multiply-adds, respectively, where  $N$  denotes the length of  $b_s[n]$  for  $n > 0$ . The first term and second

term at the right-hand side of Eqs. (41) and (42) require  $2v(2N + 1)$  and  $2(N - v)(2v - 1)$  multiply-adds, respectively. The approach to computing the Jacobian matrix  $\mathbf{J}^{(t)}(\mathbf{w}^{(t)})$  by using the method described in [8] requires about  $20v$  multiply-adds and 10 sine (sine function) and 10 cosine (cosine function) of LSF parameters. However, the cosine function of the LSF parameters is typically computed by the root-search procedure in finding LSF parameters. Finally, the computation for the diagonal terms of  $\mathbf{K}_{sm}^{(t)}$  and  $\mathbf{K}_{cm}^{(t,t-1)}$  in Eq. (29) requires about  $2(v^3 + v^2)$  multiply-adds. The total computation complexity for calculating the DA weights therefore is

$$\begin{aligned} \mathbf{C}_{\text{DA}} = & \{vL + (L + 1 - v/2)(v + 1) + v(N + 1) + \\ & [(L + 1) - \frac{N+v-1}{2}] (N + v) + N(N + 1) + 2v(2N + 1) + \\ & 2(N - v)(2v - 1) + 20v + 2v^3 + 2v^2\} \text{ multiply - adds} \\ & + 10 \text{ sine.} \end{aligned} \quad (43)$$

### 3 EXPERIMENTAL RESULTS

The speech database used in our experiments consists of Chinese sentences recorded from 13 speakers including 8 men and 5 women. The database is divided into two parts. One is named training database including 8 speakers (5 men and 3 women) and the other is named testing database including 5 speakers (3 men and 2 women) independent of those used in the training database. Speech signals are low-pass filtered at 4 kHz and digitized at a sampling rate of 8 kHz. A tenth-order LPC analysis based on Levinson-Durbin algorithm is performed every 20 ms using a 25 ms-length analysis Hamming window. Moreover, a fixed 50 Hz bandwidth expansion is applied to each pole of the LPC vector, by replacing  $a_i$  by  $a_i\gamma^i$ , for  $1 \leq i \leq 10$ , where  $\gamma = 0.994$ . The LPC parameters with bandwidth expansion of each frame are transformed to LSF parameters by using Chebyshev polynomials [11]. Notice that the spectra corresponding to silence are excluded. Finally, we have 42,747 LSF vectors for training in the training database, and 29,098 LSF vectors for testing in the testing database.

The codebooks are trained according to a multi-stage VQ structure by using the joint optimization procedure with 3 or 4 stages and search depth of  $M = 8$  as described in [12]. The iterations before the error converges require about 100 time steps. The bit allocation of each stage for various bit rates is shown in Table 1.

### 3.1 Experiment 1

To examine the spectral dynamics of quantized spectra at each frequency based on the SQWD measure in LSF VQ, the weighting function in Eq. (4) is set to 1 in this experiment. That is, the weights of all frequencies are equal in this weighting function. In such a case, the matrices  $\mathbf{K}_{sm}^{(t)}$  and  $\mathbf{K}_{sm}^{(t-1)}$  in Eq. (29) are diagonal according to the results of previous section and [8]. Under this condition, the DA weights determined according to the proposed MWLSD (SQWD) measure, and the WMSE weights determined according to the WLSLSD measure in [8] are named “dynamic Gardner weights” and “static Gardner weights”, respectively. In order to compare the performance of the dynamic Gardner weights to that of the static Gardner weights, two LSF vector quantizers are trained by using these two sets of weights, respectively.

The curves in Fig. 1 illustrate that the quantized spectral dynamics based on the dynamic Gardner weights are closer to the unquantized spectral dynamics than that based on the static Gardner weights. For precise analysis, two performance measures are used. One is the average log-spectral distortion (SD) in dB, which is commonly used to measure the spectral quantization distortion of a quantizer. The average SD is defined as

$$\overline{\text{SD}} = \frac{1}{N_f} \sum_{n=1}^{N_f} \sqrt{\frac{1}{\pi} \int_0^\pi \left( 10 \log_{10}(S_n(w)) - \log_{10}(\hat{S}_n(w)) \right)^2 dw}, \quad (44)$$

where  $S_n(w)$  and  $\hat{S}_n(w)$  are the power spectra of the  $n$ th speech frame without and with quantization, respectively, and  $N_f$  is the total count of frames in the testing database. The other measure is the average log-spectral-dynamics distortion (SDD) in dB, which is used to measure the spectral-dynamics quantization distortion of a quantizer, which is defined as

$$\overline{\text{SDD}} = \frac{1}{N_f} \sum_{n=1}^{N_f} \sqrt{\frac{1}{\pi} \int_0^\pi \left( 10 \log_{10}(S_n(w)/S_{n-1}(w)) - \log_{10}(\hat{S}_n(w)/\hat{S}_{n-1}(w)) \right)^2 dw}, \quad (45)$$

where  $S_{n-1}(w)$  and  $\hat{S}_{n-1}(w)$  are the power spectra of the  $(n-1)$ th speech frame without and with quantization, respectively, and  $S_n(w)$ ,  $\hat{S}_n(w)$  and  $N_f$  are defined in Eq. (44).

The values of  $\overline{\text{SD}}$  and  $\overline{\text{SDD}}$  at different bit rates are listed in Table 2. The  $\overline{\text{SD}}$  value for the dynamic Gardner weights is larger than the  $\overline{\text{SD}}$  value for the static Gardner weights at each bit rate as listed in Table 2. Besides, Table 2 also shows that the  $\overline{\text{SDD}}$  value for the dynamic Gardner weights is smaller than the  $\overline{\text{SDD}}$  value for the static Gardner weights at each bit rate. This reveals that the quantized spectral dynamics based on the dynamic Gardner weights is significantly maintained to be closer to the original (unquantized) spectral dynamics than that based on the static Gardner weights, even though the



average spectral distortion ( $\overline{\text{SD}}$ ) for the dynamic Gardner weights is larger than that for the static Gardner weights. In other words, our proposed technique can be surely used to effectively reduce the spectral-dynamics distortion ( $\overline{\text{SDD}}$ ) at the expense of increasing the spectral distortion ( $\overline{\text{SD}}$ ) slightly. Table 2 also lists the values of  $\overline{\text{SD}}$  and  $\overline{\text{SDD}}$  of the LSF multistage vector quantizers trained by using the QWD measure. By comparing the distortion values corresponding to the QWD and SQWD measures, respectively, we can see the extent to which the performance degrades due to the approximation loss of using the SQWD measure to replace the QWD measure.

### 3.2 Experiment 2

A subjective listening test is performed in this experiment to compare the performance of LSF vector quantizers trained by using dynamic weights to that trained by using static weights, where the dynamic weights and static weights are the weights of the sensitivity matrices for LSF parameters based on the MWLSD (SQWD) measure and the WLSLSD measure, respectively. In this test, two error shaping functions related to human perception are incorporated into the LSF VQ. One is the Bark weighting function which relates to auditory perception of human ear [13], given as

$$W_B(w) = \frac{1}{25 + 75(1 + 1.4(\frac{F_s w}{2000\pi})^2)^{0.69}}, \quad (46)$$

where  $F_s$  is the sampling frequency of 8 kHz. The impulse response of the Bark weighting function,  $W_b[n]$ , can be approximately obtained by applying a finite autocorrelation sequence with length 79 to the 256-point IDFT of  $W_B(w)$  (assuming the time indices of the IDFT of  $W_B(w)$  is ranged from  $-127$  to  $128$ ). The autocorrelation sequence is the autocorrelation function of a Hamming window with length 40 centered at time index 0. The positive-definite property of the sensitivity matrices for LSF parameters based on the MWLSD (SQWD) measure and the WLSLSD measure with the Bark weighting function can thus be maintained when the finite impulse response  $W_b[n]$  is used in computing the sensitivity matrices by the methods described in the previous section and [4]. The other error shaping function is  $h_l(w)$  defined in [4], which is one of the best error weighting functions that match the criterion of transparent quantization of LPC information proposed by Paliwal and Atal [3]. The frequency response of  $h_l(w)$  is shown in Fig. 2. The impulse response of  $h_l(w)$ , denoted by  $h_l[n]$ , is designed to be a FIR filter for  $-10 \leq n \leq 10$  to avoid the computational problem in calculating the sensitivity matrices for LSF parameters based on the MWLSD (SQWD) measure and the WLSLSD measure.

As stated before, the sensitivity matrices for the LSF parameters based on the MWLSD (SQWD) measure with the above two error weighting functions

can be obtained by using the impulse responses  $W_b[n]$  and  $h_l[n]$ , and the result of previous section. Each of these two sensitivity matrices can be divided into four submatrices according to Eq. 29. The left-upper submatrix of each of these sensitivity matrices is nondiagonal, since the lengths of these two impulse responses are both larger than 3 [4]. This causes the DA weights of these two sensitivity matrices to be used according to the result of previous section. The DA weights corresponding to the impulse responses  $W_b[n]$  and  $h_l[n]$  are named “dynamic Bark weights” and “dynamic  $h_l$  weights”, respectively. The static weights without considering the information of spectral dynamics are also computed by applying  $W_B(w)$  and  $h_l(w)$  into the WLSD measure and then taking the diagonal elements of each of the sensitivity matrices for the LSF parameters based on the WLSD measure through the use of the impulse responses  $W_b[n]$  and  $h_l[n]$ , and the method described in [4]. The static weights corresponding to the impulse responses  $W_b[n]$  and  $h_l[n]$  are named “static Bark weights” and “static  $h_l$  weights”, respectively.

A diagram of the model for studying the effects of LSF VQ with the dynamic weights or static weights is shown in Fig. 3 [14]. A prediction residual is formed by filtering the speech signal using an unquantized prediction filter, and the synthetic speech is generated by exciting a quantized inverse prediction filter with the undistorted residual. In this way, the effects of LSF VQ can be studied separately from any encoding of the residual.

Three original speech patterns and the LSF quantizers trained with the dynamic weights or static weights at six different bit rates listed in Table 1 are used in this test. Seventy-two synthetic speech patterns are thus generated by utilizing the testing model shown in Fig. 3; eighteen (3 speech signals by 6 different bit rates) for the dynamic Bark weights, eighteen for the dynamic  $h_l$  weights, eighteen for the static Bark weights, and eighteen for the static  $h_l$  weights.

To evaluate the quality of the synthetic speech signals with the LSF vector quantizers trained by using these four sets of weights, the common Mean Opinion Score (MOS) test is performed. A total of eighteen judges joint the MOS test. Each judge listens to each synthetic speech signal and then gives a score. The synthetic speech signals are presented in random order. For each synthetic speech signal, the judges are asked to rate the subjective score, which is divided into 5 levels: 5 for excellent, 4 for good, 3 for fair, 2 for poor, 1 for bad. With these scores, a MOS value is calculated by taking the arithmetic mean of the scores which correspond to the synthetic speech signals generated by using the same set of weights at the same bit rate. The MOS values of this subjective test are shown in Table 3. Table 3 shows that the MOS values corresponding to dynamic Bark/ $h_l$  weights is larger than static Bark/ $h_l$  weights at each bit rate. This implies that the quality of the synthetic speech signals with the LSF vector quantizers trained by using the dynamic Bark/ $h_l$  weights

is felt to be better than that by using the static Bark/ $h_l$  weights at each bit rate. This also confirms that incorporating the spectral dynamics information into the LSF VQ results in a better improvement of the subjective quality of the synthetic speech.

In order to further compare the performance of the dynamic weights to that of the static weights in numerical analysis, two performance indicators are defined. They take auditory perception of human ear into account (i.e., incorporating the Bark error weighting function into the spectral distortion computation). These two performance indicators are similar to  $\overline{\text{SD}}$  and  $\overline{\text{SDD}}$ . One is the average weighted log-spectral distortion (WSD), given as

$$\overline{\text{WSD}} = \frac{1}{N_s} \sum_{n=1}^{N_s} \sqrt{\frac{1}{W_{B_0}\pi} \int_0^\pi W_B^2(w) \left(10 \log_{10}(S_n(w)) - 10 \log_{10}(\hat{S}_n(w))\right)^2 dw}, \quad (47)$$

where  $N_s$  is the number of frames used in computing  $\overline{\text{WSD}}$ ,  $W_B(w)$  is defined in Eq. (46),  $W_{B_0} = 2.603 \times 10^{-5}$  is the normalization factor of  $W_B(w)$ , and  $S_n(w)$  and  $\hat{S}_n(w)$  are defined in Eq. (44). The other performance indicator is the average weighted log-spectral-dynamics distortion (WSDD), defined as

$$\overline{\text{WSDD}} = \frac{1}{N_s} \sum_{n=1}^{N_s} \sqrt{\frac{1}{W_{B_0}\pi} \int_0^\pi W_B^2(w) \left[10 \log_{10}(S_n(w)/S_{n-1}(w)) - 10 \log_{10}(\hat{S}_n(w)/\hat{S}_{n-1}(w))\right]^2 dw}, \quad (48)$$

where  $N_s$ ,  $W_B(w)$ , and  $W_{B_0}$  are defined in Eq. (47), and  $S_n(w)$ ,  $\hat{S}_n(w)$ ,  $S_{n-1}(w)$ , and  $\hat{S}_{n-1}(w)$  are defined in Eq. (45).

The three original speech patterns used in the previous subjective listening test are also used in this numerical analysis. The values of  $\overline{\text{WSD}}$  and  $\overline{\text{WSDD}}$  corresponding to the Bark weights and the  $h_l$  weights at different bit rates are listed in Tables 4 and 5, respectively. Tables 4 and 5 also list the values of  $\overline{\text{WSD}}$  and  $\overline{\text{WSDD}}$  of the LSF multistage vector quantizers trained by using the QWD measure with the Bark and  $h_l(w)$  weighting functions, respectively, at different bit rates for observing the performance degradation due to the approximation loss of using the SQWD measure to replace the QWD measure. Tables 4 and 5 show that the  $\overline{\text{WSDD}}$  ( $\overline{\text{WSD}}$ ) values for the dynamic Bark weights and the dynamic  $h_l$  weights are smaller (larger) than the  $\overline{\text{WSDD}}$  ( $\overline{\text{WSD}}$ ) values for the static Bark weights and the static  $h_l$  weights at each bit rate, respectively. This indicates that a smaller  $\overline{\text{WSDD}}$  value is significantly more important than a larger  $\overline{\text{WSD}}$  value at the same bit rate and the same error weighting function from the viewpoint of speech quality according to the results of the above subjective listening tests. That is, the speech quality can be significantly improved by reducing the weighted spectral-dynamics distortion ( $\overline{\text{WSDD}}$ ) at the expense of increasing the weighted spectral distortion ( $\overline{\text{WSD}}$ ) slightly.

It is noted that  $\overline{\text{WSD}}$  is also a good performance indicator for a quantizer to indicate the approaching degree of its spectral quantization-error-distribution curve to the inverse curve of a desired weighting function [4]. Hence, Tables 4 and 5 also show that the spectral quantization error distributions for the dynamic Bark weights and the dynamic  $h_l$  weights can be effectively shaped into the inverse curves of the Bark and  $h_l$  weighting functions, respectively, since the differences in the values of  $\overline{\text{WSD}}$  between the dynamic Bark ( $h_l$ ) weights and the static Bark ( $h_l$ ) weights at each bit rate are small. This demonstrates that the proposed technique can shape the spectral quantization error distribution into the selected frequency-dependent shaping curve and simultaneously improve the speech quality of the synthetic speech by reducing the spectral-dynamics distortion between the unquantized and quantized spectra.

### 3.3 Experiment 3

The results of Experiments 1 and 2 show that the LSF VQ trained by using the proposed technique achieves the desired effect of decreasing the spectral-dynamics distortion between unquantized spectra and quantized spectra, and improving the reconstructed speech quality for the given error weighting function. In order to further test how this proposed technique affects the overall speech quality in an actual speech coding algorithm, the proposed technique is applied to the LSF quantization of a CELP coder. In this experiment, the codebooks trained at 30, 27, 24, 21, 18 and 15 bits/frame with the DA weights obtained in Experiment 2 (for the Bark weighting function and for the  $h_l$  weighting function) are used for this test. The three speech patterns used in Experiment 2 are prepared for coding; each of them is analyzed and then reconstructed by the original CELP coder and the modified CELP coders that use the LSF vector quantizers obtained in Experiment 2. The LSF quantizer used in the original CELP coder is a non-uniform scalar quantizer using 34 bits/frame.

This experiment produces thirty-nine reconstructed speech signals; three for the original CELP coder, eighteen (3 speech signals by 6 different bit rates) for the modified CELP coders corresponding to the Bark weighting function, and eighteen for the modified CELP coders corresponding to the  $h_l$  weighting function.

In order to compare the performance of the original CELP coder to that of the MCELP coder, a subjective listening test similar to the MOS test in Experiment 2 is performed except using the above reconstructed speech signals in this experiment. The MOS values for these coders are shown in Table 6. Table 6 shows that the MOS value of the MCELP coder are larger than that of the original CELP coder at each bit rate except 18 bits/frame and 15 bits/frame. This implies that the speech quality of the reconstructed speech signals us-

ing the MCELP coder is better than/[comparable to] that using the original CELP coder, even at lower bit rates such as 21 bits/frame. In other words, the results show that the approach incorporating the error shaping technique with the spectral dynamics information into the LSF quantization of CELP coder does not produce strange distortion, and moreover, results in lower bit rates for better or equivalent speech quality, even at 21 bits/frame.

## 4 CONCLUSIONS

This paper proposes an innovative technique to incorporate the spectral dynamics information of LPC spectra into LSF vector quantization (VQ) based on a modified weighted log-spectral distortion (MWLSD) measure. To achieve lower computation complexity, a quadratically weighted distortion (QWD) measure is derived to approximate the MWLSD measure. The LSF VQ trained by the QWD measure has been shown to converge to the LSF VQ trained by the MWLSD measure. The quadratic matrix used in the QWD measure, the so-called “sensitivity matrix,” is given by the second term of Taylor series expansion of the MWLSD measure and has been proven to be a nondiagonal matrix. An approximate computation algorithm for calculating this sensitivity matrix is provided. An experimentally determined matrix used to approximate the sensitivity matrix is also provided to further reduce the computational complexity of the QWD measure, and results in the simplified quadratically weighted distortion (SQWD) measure. The effect of maintaining the spectral dynamics of quantized spectra to be close to that of unquantized spectra by using the SQWD measure based on the proposed technique has been verified in Experiment 1. The subjective listening test in Experiment 2 confirms that the proposed technique can improve the quality of synthetic speech and decreasing the spectral-dynamics distortion between unquantized spectra and quantized spectra at the expense of increasing the spectral distortion slightly. Finally, the test of applying the proposed technique with Bark weighting function or  $h_l$  weighting function to the LSF quantization (34 bits/frame originally) of an actual CELP coder in Experiment 3 has shown that the proposed scheme can result in equivalent speech quality at lower bit rates (e.g., 21 bits/frame).

**ACKNOWLEDGEMENTS.** This work was supported by the Ministry of Economic Affairs, Taiwan R.O.C., under Grant 95-EC-17-A-02-S1-032.

## References

- [1] F. Itakura, Line spectrum representation of linear predictive coefficients of speech signals, *J. Acoust. Soc. Amer.*, **57** (1975), S35(A).

- [2] F.K. Soong and B.H. Juang, Line spectrum pair (LSP) and speech data compression, in *Proc. of ICASSP'84*, (1984), 1.10.1-1.10.4.
- [3] K.K. Paliwal and B.S. Atal, Efficient vector quantization of LPC parameters at 24 bits/frame, *IEEE Trans. on Speech and Audio Processing*, **1** (1993), 3-14.
- [4] H.W. Nein and C.T. Lin, Incorporating error shaping technique into LSF vector quantization, *IEEE Trans. on Speech and Audio Processing*, **9** (2001), 73-86.
- [5] W.B. Kleijn and R. Hagen, On memoryless quantization in speech coding, *IEEE Signal Processing Letters*, **3** (1996), 228-230.
- [6] H.P. Knagenhjelm and W.B. Kleijn, Spectral dynamics is more important than spectral quantization, in *Proc. of ICASSP'95*, (1995), 732-735.
- [7] K. Koishida, T. Masuko, and T. Kobayashi, Spectral quantization using statistics of static and dynamic features, *IEEE Workshop on Speech Coding for Telecommunications Proceeding*, (1997), 19-20.
- [8] W.R. Gardner and B.D. Rao, Theoretical analysis of the high-rate vector quantization of LPC parameters, *IEEE Trans. on Speech and Audio Processing*, **3** (1995), 367-381.
- [9] J. Li, N. Chaddha, and R.M. Gray, Asymptotic performance of vector quantizers with a perceptual distortion measure, *IEEE Trans. Information Theory*, **45** (1995), 1082-1091.
- [10] S.P. Lloyd, Least squares quantization in PCM, *IEEE Trans. on Information Theory*, **28** (1982), 129-137.
- [11] P. Kabal and P. Ramachandran, The computation of line spectral frequencies using Chebyshev polynomials, *IEEE Trans. on Acoust. Speech and Signal Processing*, **34** (1986), 1419-1426.
- [12] W.P. LeBlanc, B. Bhattacharya, S.A. Mahmoud, and V. Cuperman, Efficient search and design procedures for robust multi-stage VQ of LPC parameters for 4 kb/s speech coding, *IEEE Trans. on Speech and Audio Processing*, **1** (1993), 373-385.
- [13] R.P. Cohn and J.S. Collura (1997), Incorporating perception into LSF quantization - some experiments, in *Proc. of ICASSP'97*, (1997), 1347-1350.

- [14] T. Eriksson, J. Lindén, and J. Skoglund, Interframe LSF quantization for noisy channels, *IEEE Trans. on Speech and Audio Processing*, **7** (1999), 495-509.

**Received: September 21, 2005**

Table 1: Bit allocation of each stage of 3-stage/4-stage VQ for various bit rates in the experiments.

Bit rates	Stage 1 (bits)	Stage 2 (bits)	Stage 3 (bits)	Stage 4 (bits)
30 bits	8	8	7	7
27 bits	7	7	7	6
24 bits	6	6	6	6
21 bits	6	5	5	5
18 bits	6	6	6	0
15 bits	5	5	5	0

Table 2: Spectral Distortion ( $\overline{SD}$ ) values and Spectral-Dynamics Distortion ( $\overline{SDD}$ ) values of the LSF multistage vector quantizers trained by using the SQWD measure with dynamic Gardner weights and static Gardner weights at six different bit rates in Experiment 1, where  $\overline{SD}$  and  $\overline{SDD}$  values in parentheses correspond to the quantizers trained by using the QWD measure.

$\overline{SD}$ and $\overline{SDD}$ in dB				
Bit rates	dynamic Gardner weights		static Gardner weights	
	$\overline{SD}$	$\overline{SDD}$	$\overline{SD}$	$\overline{SDD}$
30	0.899 (0.911)	0.989 (0.966)	0.834	1.153
27	1.076 (1.091)	1.179 (1.146)	0.996	1.369
24	1.291 (1.312)	1.403 (1.372)	1.236	1.668
21	1.538 (1.554)	1.644 (1.607)	1.443	1.926
18	1.752 (1.781)	1.882 (1.831)	1.643	2.158
15	2.084 (2.122)	2.205 (2.148)	1.981	2.525



Table 3: Subjective test results for the LSF vector quantizers trained by using the dynamic Bark weights, dynamic  $h_l$  weights, static Bark weights, and static  $h_l$  weights in the testing model.

Bit rates	MOS Values			
	Bark weighting function		$h_l$ weighting function	
	Dynamic Bark weights	Static Bark weights	Dynamic $h_l$ weights	Dynamic $h_l$ weights
30	4.20	4.16	4.15	4.12
27	4.14	4.08	4.11	4.05
24	4.07	3.97	4.05	3.94
21	3.99	3.85	3.97	3.80
18	3.77	3.62	3.74	3.56
15	3.34	3.20	3.28	3.16

Table 4: Weighted Spectral Distortion ( $\overline{\text{WSD}}$ ) values and Weighted Spectral-Dynamics Distortion ( $\overline{\text{WSDD}}$ ) values of the LSF multistage vector quantizers trained by using the SQWD measure with dynamic Bark weights and static Bark weights at six different bit rates in Experiment 2, where  $\overline{\text{SD}}$  and  $\overline{\text{SDD}}$  values in parentheses correspond to the quantizers trained by using the QWD measure.

$\overline{\text{WSD}}$ and $\overline{\text{WSDD}}$ in dB				
Bit rates	dynamic Bark weights		static Bark weights	
	$\overline{\text{WSD}}$	$\overline{\text{WSDD}}$	$\overline{\text{WSD}}$	$\overline{\text{WSDD}}$
30	0.800 (0.804)	0.063 (0.056)	0.768	0.075
27	0.949 (0.953)	0.074 (0.068)	0.889	0.086
24	1.123 (1.126)	0.087 (0.080)	1.071	0.103
21	1.329 (1.334)	0.102 (0.093)	1.246	0.118
18	1.512 (1.516)	0.114 (0.106)	1.451	0.135
15	1.825 (1.831)	0.135 (0.124)	1.745	0.157

Table 5: Weighted Spectral Distortion ( $\overline{\text{WSD}}$ ) values and Weighted Spectral-Dynamics Distortion ( $\overline{\text{WSDD}}$ ) values of the LSF multistage vector quantizers trained by using the SQWD measure with dynamic  $h_l$  weights and static  $h_l$  weights at six different bit rates in Experiment 2, where  $\overline{\text{SD}}$  and  $\overline{\text{SDD}}$  values in parentheses correspond to the quantizers trained by using the QWD measure.

$\overline{\text{WSD}}$ and $\overline{\text{WSDD}}$ in dB				
Bit rates	dynamic $h_l$ weights		static $h_l$ weights	
	$\overline{\text{WSD}}$	$\overline{\text{WSDD}}$	$\overline{\text{WSD}}$	$\overline{\text{WSDD}}$
30	0.949 (0.952)	0.074 (0.064)	0.873	0.086
27	1.129 (1.082)	0.088 (0.077)	1.047	0.102
24	1.352 (1.357)	0.104 (0.094)	1.267	0.121
21	1.613 (1.619)	0.122 (0.109)	1.524	0.143
18	1.858 (1.862)	0.138 (0.127)	1.731	0.159
15	2.204 (2.211)	0.160 (0.145)	2.089	0.187

Table 6: Subjective test results for the original and MCELP coder with LSF vector quantizers trained by using the DA weights corresponding to Bark  $h_l$  weighting functions at six different bit rates.

Bit rates	MOS Values		
	Original CELP	MCELP	
		Bark weighting function	$h_l$ weighting function
34	3.34	-	-
30	-	3.59	3.58
27	-	3.57	3.54
24	-	3.55	3.52
21	-	3.44	3.42
18	-	3.29	3.24
15	-	2.96	2.89

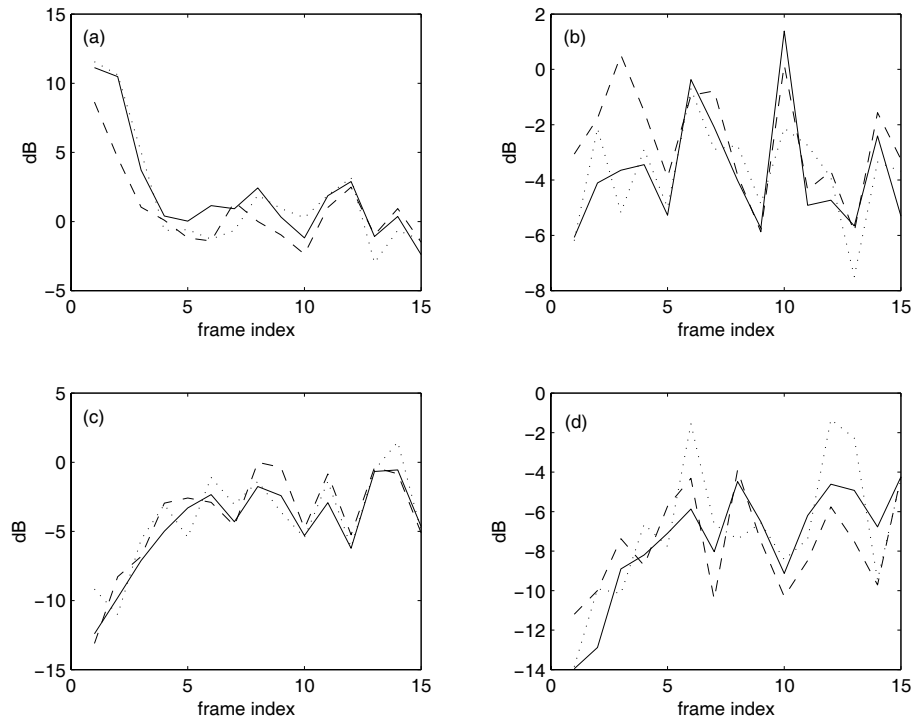


Figure 1: Spectral tracks at the frequencies of (a) 1 kHz, (b) 2 kHz, (c) 3 kHz, and (d) 4 kHz for 15-bit LSF multistage vector quantizers using two different sets of weights in Experiment 1, where dashed line corresponds to the dynamic Gardner weights, dotted line to static Gardner weights, and solid line to unquantized spectra.

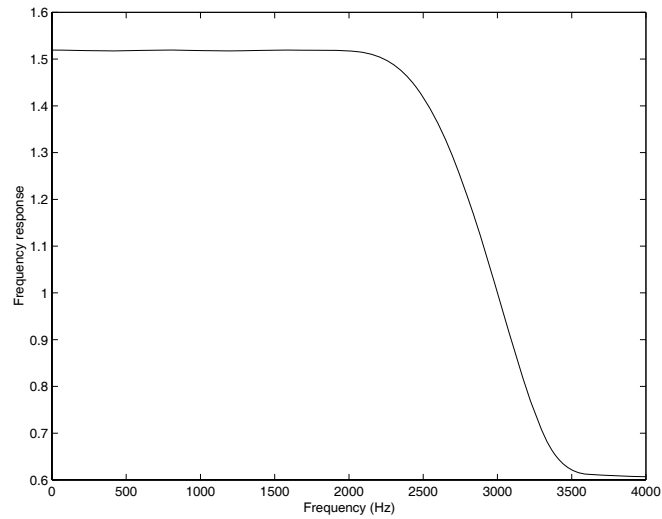


Figure 2: Frequency response of  $h_l(w)$  used in Experiment 2.

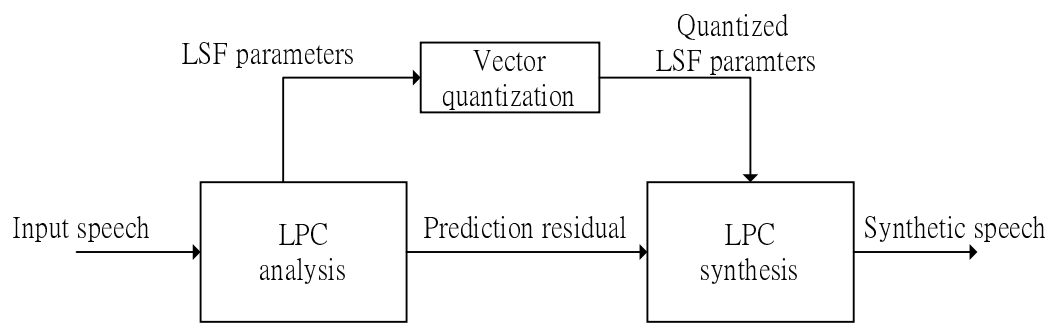


Figure 3: Synthetic speech production for the subjective listening test in Experiment 2.

## APPENDIX A

This appendix provides the proof of Theorem 2.2. The proof of Theorem 2.2 is obviously completed if Eqs. 12–15 are proved. We derive Eq. (12) firstly. Let  $a_0^{(t)} = a_0^{(t-1)} = -1$ . Then, we have  $A^{(t)}(w) = -\sum_{m=0}^v a_m^{(t)} e^{-jwm}$  and  $A^{(t-1)}(w) = -\sum_{m=0}^v a_m^{(t-1)} e^{-jwm}$ . This results in

$$|A^{(t)}(w)|^2 = \sum_{m=0}^v \sum_{n=0}^v a_m^{(t)} a_n^{(t)} \cos(w(m-n)), \quad (\text{A.1})$$

$$|A^{(t-1)}(w)|^2 = \sum_{m=0}^v \sum_{n=0}^v a_m^{(t-1)} a_n^{(t-1)} \cos(w(m-n)). \quad (\text{A.2})$$

By substituting these results into Eq. (9) and performing simple differentiation [8], we can show that

$$\begin{aligned} & \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda} = \Lambda} \\ &= 8\alpha \sum_{m=0}^v \sum_{n=0}^v a_m^{(t)} a_n^{(t)} \frac{1}{2\pi} \int_{-\pi}^{\pi} B^2(w) \frac{\cos(w(k+l-m-n)) + \cos(w(k-l-m+n))}{|A^{(t)}(w)|^4} dw \\ &= 8\alpha \sum_{m=0}^v \sum_{n=0}^v a_m^{(t)} a_n^{(t)} [f^{(t)}(k+l-m-n) + f^{(t)}(k-l-m+n)], \end{aligned} \quad (\text{A.3})$$

where

$$f^{(t)}(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} B_s(w) \frac{e^{jwx}}{|A^{(t)}(w)|^4} dw \quad (\text{A.4})$$

is the inverse discrete-time Fourier transform of  $B_s(w)/|A^{(t)}(w)|^4$  since  $B_s(w)$  is a real function. Thus we have  $F^{(t)}(w) = B_s(w)/|A^{(t)}(w)|^4$ , and  $F^{(t)}(z) = B_s(z)/(A^{(t)2}(z)A^{(t)2}(z^{-1}))$ .

Using the similar notations in [8], let

$$\begin{aligned} M^{(t)}(x) &= \sum_{m=0}^v \sum_{n=0}^v a_m^{(t)} a_n^{(t)} f^{(t)}(x-m-n) \quad \text{and} \\ N^{(t)}(x) &= \sum_{m=0}^v \sum_{n=0}^v a_m^{(t)} a_n^{(t)} f^{(t)}(x-m+n) \end{aligned} \quad (\text{A.5})$$

Then we have

$$\left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda} = \Lambda} = 8\alpha (M^{(t)}(k+l) + N^{(t)}(k-l)). \quad (\text{A.6})$$

Since

$$M^{(t)}(z) = A^{(t)2}(z)F^{(t)}(z) = \frac{B_s(z)}{A^{(t)2}(z^{-1})} \quad (\text{A.7})$$

and

$$N^{(t)}(z) = A^{(t)}(z)A^{(t)}(z^{-1})F^{(t)}(z) = \frac{B_s(z)}{A^{(t)}(z)A^{(t)}(z^{-1})}, \quad (\text{A.8})$$

we can take inverse  $Z$  transforms of  $M^{(t)}(z)$  and  $N^{(t)}(z)$ , and use the  $Z$  transform pairs of Eqs. (17) and (18) to obtain the following equality

$$\begin{aligned} \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda} = \Lambda} &= R_{SM}^{(t)}(k, l) = 8\alpha R_{sm}^{(t)}(k, l) \\ &= 8\alpha \sum_{n=-\infty}^{\infty} b_s[n] \left( r_c^{(t)}[(k+l)-n] + r^{(t)}[(k-l)-n] \right). \end{aligned} \quad (\text{A.9})$$

This completes the proof of Eq. (12).

By performing the steps taken in showing Eq. (12) except using the LPC parameters  $\mathbf{a}^{(t-1)}$  instead of  $\mathbf{a}^{(t)}$ , a similar proof can be derived for Eq. (13). In order to prove Eq. (14), a simple second-order differentiation with respect to  $\hat{a}_k^{(t-1)}$  and  $\hat{a}_l^{(t)}$  is directly performed in Eq. (9). Then we can show that

$$\begin{aligned} \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t-1)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda} = \Lambda} &= \frac{-2\alpha}{2\pi} \int_{-\pi}^{\pi} B^2(w) \left[ \frac{e^{-jwk}}{A^{(t-1)}(w)} + \frac{e^{jwk}}{A^{(t-1)*}(w)} \right] \left[ \frac{e^{-jwl}}{A^{(t)}(w)} + \frac{e^{jwl}}{A^{(t)*}(w)} \right] dw. \\ &= \frac{-2\alpha}{2\pi} \int_{-\pi}^{\pi} B^2(w) \left[ \frac{e^{jw(-k-l)}}{A^{(t)}(w)A^{(t-1)}(w)} + \frac{e^{jw(k+l)}}{A^{(t)*}(w)A^{(t-1)*}(w)} + \right. \\ &\quad \left. \frac{e^{jw(k-l)}}{A^{(t)}(w)A^{(t-1)*}(w)} + \frac{e^{jw(l-k)}}{A^{(t)*}(w)A^{(t-1)}(w)} \right] dw, \end{aligned} \quad (\text{A.10})$$

where ‘\*’ denotes the complex conjugate operator. By using the conjugate symmetric property of Fourier transform for the time-domain real sequences  $\mathbf{a}^{(t)}$  and  $\mathbf{a}^{(t-1)}$ , Eq. (A.10) becomes

$$\left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t-1)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda} = \Lambda} = \frac{-4\alpha}{2\pi} \int_{-\pi}^{\pi} B^2(w) \left[ \frac{e^{jw(k-l)}}{A^{(t-1)*}(w)A^{(t)}(w)} + \frac{e^{jw(k+l)}}{A^{(t-1)*}(w)A^{(t)*}(w)} \right] dw. \quad (\text{A.11})$$

Through the use of the inverse discrete-time Fourier transform of  $B^2(w)$  and the Fourier transform pairs of Eqs. (20) and (21), we can directly rewrite Eq. (A.11) to obtain the following equality

$$\begin{aligned} \left. \frac{\partial^2 D_{\text{MWLSD}}(\Lambda, \hat{\Lambda})}{\partial \hat{a}_k^{(t-1)} \partial \hat{a}_l^{(t)}} \right|_{\hat{\Lambda} = \Lambda} &= R_{CM}^{(t,t-1)}(k, l) = -4\alpha R_{cm}^{(t,t-1)}(k, l) \\ &= -4\alpha \sum_{n=-\infty}^{\infty} b_s[n] \left( r_e^{(t,t-1)}[(k+l)-n] + r_d^{(t,t-1)}[(k-l)-n] \right). \end{aligned} \quad (\text{A.12})$$

This completes the proof of Eq. (14). Equation (15) can also be proved by following the similar steps taken in the proof of Eq. (14). The proof of Theorem 2.2 therefore is completed since Eqs. (12)–(15) have been verified.

## APPENDIX B

This appendix provides the proof of the nonnegative-definite property of  $\mathbf{K}_{\text{SQWD}}$  in Eq. (36). First, we consider the contributions of the first term and third term at the right-hand side of Eq. (9) to the sensitivity matrix for the LSF parameters  $\Theta$ ,  $\mathbf{D}_\theta(\Theta)$ . These contributions can be denoted as two matrices  $\mathbf{D}^{(t)}$  and  $\mathbf{D}^{(t-1)}$ , respectively, which can be obviously computed by using the method described in [4], where  $\mathbf{D}^{(t)}$  and  $\mathbf{D}^{(t-1)}$  are both  $2v$  by  $2v$  dimensional matrices and have the following special form,

$$\mathbf{D}^{(t)} = \left[ \begin{array}{c|c} \mathbf{D}_u^{(t)} & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{O} \end{array} \right] \quad \text{and} \quad \mathbf{D}^{(t-1)} = \left[ \begin{array}{c|c} \mathbf{O} & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{D}_u^{(t-1)} \end{array} \right], \quad (\text{B.1})$$

where  $\mathbf{O}$  is a  $v \times v$  zero matrix,  $\mathbf{D}_u^{(t)}$  and  $\mathbf{D}_u^{(t-1)}$  are generally  $v \times v$  nonzero square matrices.

We now consider the second term at the right-hand side of Eq. (9). Let the matrix  $\mathcal{D}^{(t,t-1)}$  denote the contribution of the second term at the right-hand side of Eq. (9) to the sensitivity matrix  $\mathbf{D}_\theta(\Theta)$ . The approach to obtaining  $\mathcal{D}^{(t,t-1)}$  will be given latter. In general, the matrix  $\mathcal{D}^{(t,t-1)}$  can be divided into four  $v \times v$  square submatrices,

$$\mathcal{D}^{(t,t-1)} = \left[ \begin{array}{c|c} \mathcal{D}_{11}^{(t,t-1)} & \mathcal{D}_{12}^{(t,t-1)} \\ \hline \mathcal{D}_{21}^{(t,t-1)} & \mathcal{D}_{22}^{(t,t-1)} \end{array} \right], \quad (\text{B.2})$$

where  $\mathcal{D}_{11}^{(t,t-1)}$ ,  $\mathcal{D}_{12}^{(t,t-1)}$ ,  $\mathcal{D}_{21}^{(t,t-1)}$  and  $\mathcal{D}_{22}^{(t,t-1)}$  are generally nonzero matrices. The sensitivity matrix for LSF parameters,  $\mathbf{D}_\theta(\Theta)$ , can thus be rewritten as

$$\mathbf{D}_\theta(\Theta) = \mathbf{D}^{(t)} + \mathcal{D}^{(t,t-1)} + \mathbf{D}^{(t-1)}. \quad (\text{B.3})$$

By using Eqs. (B.1)–(B.3), we can then rewrite  $\mathbf{K}_{\text{SQWD}}$  as

$$\mathbf{K}_{\text{SQWD}} = \mathbf{U}^{(t)} + \mathcal{U}^{(t,t-1)} + \mathbf{U}^{(t-1)}, \quad (\text{B.4})$$

where  $\mathbf{U}^{(t)}$ ,  $\mathcal{U}^{(t,t-1)}$ , and  $\mathbf{U}^{(t-1)}$  are defined as

$$\mathbf{U}^{(t)} = \left[ \begin{array}{c|c} \text{diag}(\mathbf{D}_u^{(t)}) & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{O} \end{array} \right], \quad \mathbf{U}^{(t-1)} = \left[ \begin{array}{c|c} \text{diag}(\mathbf{D}_u^{(t-1)}) & \mathbf{O} \\ \hline \mathbf{O} & \mathbf{O} \end{array} \right], \quad (\text{B.5})$$

$$\mathcal{U}^{(t,t-1)} = \left[ \begin{array}{c|c} \text{diag}(\mathcal{D}_{11}^{(t,t-1)}) & \text{diag}(\mathcal{D}_{12}^{(t,t-1)}) \\ \hline \text{diag}(\mathcal{D}_{21}^{(t,t-1)}) & \text{diag}(\mathcal{D}_{22}^{(t,t-1)}) \end{array} \right], \quad (\text{B.6})$$

where the operator  $\text{diag}(\cdot)$  is defined in Eq. (35). The matrix  $\mathbf{K}_{\text{SQWD}}$  is nonnegative definite if the quadratic form  $\mathbf{e}^T \mathbf{K}_{\text{SQWD}} \mathbf{e}$  is greater than or equal to

0 for any nonzero  $\mathbf{e}$ , where  $\mathbf{e}$  is a  $2v$ -dimensional column vector. That is, if  $\mathbf{e}^T \mathbf{K}_{\text{SQWD}} \mathbf{e} \geq 0$  is proved, then the proof of the nonnegative-definite property of  $\mathbf{K}_{\text{SQWD}}$  is completed. We shall now show that  $\mathbf{e}^T \mathbf{K}_{\text{SQWD}} \mathbf{e} \geq 0$ . By using Eq. (B.4),  $\mathbf{e}^T \mathbf{K}_{\text{SQWD}} \mathbf{e}$  can be written as

$$\mathbf{e}^T \mathbf{K}_{\text{SQWD}} \mathbf{e} = \mathbf{e}^T \mathbf{U}^{(t)} \mathbf{e} + \mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e} + \mathbf{e}^T \mathbf{U}^{(t-1)} \mathbf{e}. \quad (\text{B.7})$$

The first term and third term at the right-hand of Eq. (B.7) are always non-negative due to the nonnegative-definite property of the  $\mathbf{D}^{(t)}$  and  $\mathbf{D}^{(t-1)}$  proved in [4]. Hence,  $\mathbf{K}_{\text{SQWD}}$  is nonnegative definite if and only if  $\mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e} \geq 0$ .

To show  $\mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e} \geq 0$ , the elements of  $\mathcal{U}^{(t,t-1)}$  should be derived firstly. The elements  $\mathcal{U}^{(t,t-1)}$  readily form a subset of the elements of  $\mathcal{D}^{(t,t-1)}$  from Eq. (B.6). In other words, the elements of  $\mathcal{U}^{(t,t-1)}$  can be directly obtained from Eq. (B.6) if the elements of  $\mathcal{D}^{(t,t-1)}$  are given. We now derive the elements of  $\mathcal{D}^{(t,t-1)}$  as follows. In order to calculate  $\mathcal{D}^{(t,t-1)}$ , the second term at the right-hand side of Eq. (9) is rewritten in the following:

$$\begin{aligned} \mathcal{D}_z(\Theta, \hat{\Theta}) &= \frac{\alpha}{2\pi B_0} \int_{-\pi}^{\pi} B^2(w) \left[ \left( \ln(A^{(t)}(w)) - \ln(A^{(t-1)}(w)) \right) - \right. \\ &\quad \left. \left( \ln(\hat{A}^{(t)}(w)) - \ln(\hat{A}^{(t-1)}(w)) \right) \right]^2 dw. \end{aligned} \quad (\text{B.8})$$

Thus, the elements of  $\mathcal{D}^{(t,t-1)}$  can be directly computed by performing simple second-order differentiation with respect to  $\hat{\theta}_i$  and  $\hat{\theta}_j$ , and then calculating the second derivatives at  $\hat{\Theta} = \Theta$ , given as

$$\begin{aligned} \mathcal{D}_{i,j}^{(t,t-1)} &= \frac{\partial^2 \mathcal{D}_z(\Theta, \hat{\Theta})}{\partial \hat{\theta}_i \partial \hat{\theta}_j} \Big|_{\hat{\Theta} = \Theta} \\ &= \frac{2\alpha}{2\pi} \int_{-\pi}^{\pi} B^2(w) \left\{ \left[ \left( \frac{\mathbf{J}_i^{(t)*}(w)}{A^{(t)*}(w)} + \frac{\mathbf{J}_i^{(t)}(w)}{A^{(t)}(w)} \right) - \left( \frac{\mathbf{J}_i^{(t-1)*}(w)}{A^{(t-1)*}(w)} + \frac{\mathbf{J}_i^{(t-1)}(w)}{A^{(t-1)}(w)} \right) \right] \right. \\ &\quad \left. \left[ \left( \frac{\mathbf{J}_j^{(t)*}(w)}{A^{(t)*}(w)} + \frac{\mathbf{J}_j^{(t)}(w)}{A^{(t)}(w)} \right) - \left( \frac{\mathbf{J}_j^{(t-1)*}(w)}{A^{(t-1)*}(w)} + \frac{\mathbf{J}_j^{(t-1)}(w)}{A^{(t-1)}(w)} \right) \right] \right\} dw, \end{aligned} \quad (\text{B.9})$$

where  $\mathbf{J}_i^{(t)}(w)$  and  $\mathbf{J}_i^{(t-1)}(w)$  are defined as

$$\mathbf{J}_i^{(t)}(w) = \frac{\partial A^{(t)}(w)}{\partial \theta_i} = \begin{cases} \frac{1}{2} \frac{\partial \bar{P}^{(t)}(w)}{\partial \theta_i}, & i \text{ is odd and } i \leq v \\ \frac{1}{2} \frac{\partial \bar{Q}^{(t)}(w)}{\partial \theta_i}, & i \text{ is even and } i \leq v \\ 0, & i > v \end{cases} \quad (\text{B.10})$$

$$\mathbf{J}_i^{(t-1)}(w) = \frac{\partial A^{(t-1)}(w)}{\partial \theta_i} = \begin{cases} \frac{1}{2} \frac{\partial \bar{P}^{(t-1)}(w)}{\partial \theta_i}, & i \text{ is odd and } i > v \\ \frac{1}{2} \frac{\partial \bar{Q}^{(t-1)}(w)}{\partial \theta_i}, & i \text{ is even and } i > v \\ 0, & i \leq v, \end{cases} \quad (\text{B.11})$$



where  $\bar{P}^{(t)}(w)$  and  $\bar{Q}^{(t)}(w)$  are the discrete-time Fourier transforms of  $\bar{P}^{(t)}(z)$  and  $\bar{Q}^{(t)}(z)$  defined in Eq. (25), respectively, and  $\bar{P}^{(t-1)}(w)$  and  $\bar{Q}^{(t-1)}(w)$  denote the values of  $\bar{P}^{(t)}(w)$  and  $\bar{Q}^{(t)}(w)$  at the  $(t-1)$ th frame index, respectively. Notice that the matrix  $\mathcal{D}^{(t,t-1)}$  is symmetric from Eq. (B.9). Since  $\mathcal{D}^{(t,t-1)}$  has been determined, the  $(i, j)$ th element of  $\mathcal{U}^{(t,t-1)}$  can then be obtained from Eq. (B.6), given as

$$\mathcal{U}_{i,j}^{(t,t-1)} = \begin{cases} \mathcal{D}_{i,j}^{(t,t-1)}, & i = j, \text{ or } i = j + v, \text{ or } j = i + v, \\ 0, & \text{otherwise.} \end{cases} \quad (\text{B.12})$$

After the elements of  $\mathcal{U}^{(t,t-1)}$  are obtained, the proof of  $\mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e} \geq 0$  can be proceeded as follows. By using the results of Eqs. (B.9) and (B.12), the quadratic form  $\mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e}$  can be computed as

$$\begin{aligned} \mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e} &= \sum_{i=0}^{2v} \sum_{j=0}^{2v} e_i \mathcal{U}_{i,j}^{(t,t-1)} e_j \\ &= \sum_{i=0}^v \sum_{j=0}^v e_i \mathcal{D}_{i,j}^{(t,t-1)} e_j + 2 \sum_{i=0}^v \sum_{j=0}^v e_i \mathcal{D}_{i,j+v}^{(t,t-1)} e_{j+v} + \\ &\quad \sum_{i=0}^v \sum_{j=0}^v e_{i+v} \mathcal{D}_{i+v,j+v}^{(t,t-1)} e_{j+v} \\ &= \frac{2\alpha}{2\pi} \int_{-\pi}^{\pi} B^2(w) [\sum_{i=0}^v \sum_{j=0}^v e_i \phi_i \phi_j e_j + 2 \sum_{i=0}^v \sum_{j=0}^v e_i \phi_i \phi_{j+v} e_{j+v} + \\ &\quad \sum_{i=0}^v \sum_{j=0}^v e_{i+v} \phi_{i+v} \phi_{j+v} e_{j+v}] dw \\ &= \frac{2\alpha}{2\pi} \int_{-\pi}^{\pi} B^2(w) [\sum_{i=0}^v (e_i \phi_i + e_{i+v} \phi_{i+v})]^2 dw, \end{aligned} \quad (\text{B.13})$$

where  $e_i$  is the  $i$ th element of the vector  $\mathbf{e}$ , and  $\phi_i$  is defined as

$$\phi_i = \left( \frac{\mathbf{J}_i^{(t)*}(w)}{A^{(t)*}(w)} + \frac{\mathbf{J}_i^{(t)}(w)}{A^{(t)}(w)} \right) - \left( \frac{\mathbf{J}_i^{(t-1)*}(w)}{A^{(t-1)*}(w)} + \frac{\mathbf{J}_i^{(t-1)}(w)}{A^{(t-1)}(w)} \right). \quad (\text{B.14})$$

Thus,  $\mathbf{e}^T \mathcal{U}^{(t,t-1)} \mathbf{e} \geq 0$  can be readily seen from Eq. (B.13). This implies that  $\mathbf{K}_{\text{SQWD}}$  is a nonnegative-definite matrix. This completes the proof.