

QUANTIZATION OF SEW AND REW MAGNITUDE FOR 2 KB/S WAVEFORM INTERPOLATION SPEECH CODING

Jing Li and Changchun Bao

Speech and Audio Signal Processing Lab, Beijing University of Technology, Beijing 100022
Email: lijing@emails.bjut.edu.cn, baochch@bjut.edu.cn

ABSTRACT

This paper presents the quantization schemes for the magnitude spectra of the slowly evolving waveform (SEW) and rapidly evolving waveform (REW) components in a 2 kb/s waveform interpolation (WI) coder. The SEW magnitude spectrum is quantized using a DCT-based predictive vector quantization approach. The REW magnitude spectrum is quantized using a matrix quantizer based on the combined dimension conversion method. The objective measures and subjective results indicate that the proposed quantization schemes are effective in achieving good quantization accuracy.

1. INTRODUCTION

Recently, to develop speech coding capable of achieving toll quality at 4 kb/s and below is one of major interests. Waveform interpolation [1] speech coding is one of the speech compression techniques that have the potential to provide high quality reconstructed speech at low bit rates.

In WI speech coding [1], the encoder extracts pitch-cycle waveforms from the linear prediction (LP) residual signal at regular intervals of about 2-3 ms. The extracted characteristic waveforms (CW) are aligned and normalized to form a two-dimensional representation that illustrates the characteristic waveform features (one pitch-cycle each) as a function of the phase and describes the evolution of the speech waveform shape along the time axis. At the decoder, the two-dimensional CW surfaces are converted back into the one-dimensional residual signal using a CW and a pitch length at every sample point obtained by linear interpolation.

In order to enhance the coding efficiency, the CW is low-pass and high-pass filtered respectively along the time axis to produce the slowly evolving waveform (SEW) and the rapidly evolving waveform (REW). The SEW

component represents mainly the periodic component of the speech signal while the REW component contains the aperiodic component in the speech signal. These components contain very different types of perceptual information. The SEW component requires a quite precise description with a low update rate while the REW component requires a high update rate but a rough description in its transmission.

From the viewpoint of speech quality, the SEW and REW components are essential. Since only a very limited number of bits can be used for coding the SEW and REW components, the quantization of SEW and REW components is probably one of the most difficult task in waveform interpolation. Various techniques have been proposed for quantization of SEW and REW magnitude spectra. At bit rates below 4kb/s, a common approach for quantization of SEW magnitude spectra is to employ the relation between the SEW and the REW magnitude spectra [1, 4]. And commonly in quantizing REW magnitude spectra, the correlation between successive REW magnitude spectra is exploited [1, 4, 5, 8].

This paper presents the quantization schemes for the magnitude spectra of the SEW and REW components in a 2 kb/s WI coder. The proposed quantization schemes are effective in achieving good quantization accuracy. This paper is organized as follows. In Section 2, the quantization of the SEW magnitude spectra is described. In Section 3, the quantization of REW magnitude spectra is described. In Section 4, the results of this work are summarized.

2. SEW MAGNITUDE QUANTIZATION

In waveform interpolation coders, the update rate for SEW magnitude spectra is usually one SEW every speech frame. In this paper, the interval of 25 ms is used. To quantize SEW magnitude spectrum is a complicated case. The reason is that its dimension $K(t)$ varies with the pitch period $P(t)$, and the relation between $K(t)$ and $P(t)$ is $K(t) = \lfloor P(t) / 2 \rfloor$ at the instant t . In our WI implementation, the pitch period P varies in the range 20-120 samples and thus this dimension K varies in the range 10-60 (the DC

This work was supported by National Natural Science Foundation of China under Grant 60372063 and Natural Science Foundation of Beijing under Grant 4042009.

component is usually omitted). Commonly, the most convenient way to quantize the variable-dimension spectral vectors is a variable dimension vector quantizer (VDVQ) [2].

In this scheme, the SEW magnitude spectrum is first transformed into the discrete cosine transform (DCT) domain. Then the resulting vector is quantized with an 8-bit analysis-by-synthesis (A-b-S) predictive vector quantizer using perceptual weighting.

2.1. DCT-based dimension conversion

The most significant factors affecting the objective accuracy of the quantized variable-dimension magnitude spectra are the transmission rate and the number of bits assigned for the quantization task [5]. Another significant factor is the coding technique that is used to alleviate the problem of the inconstant vector dimension [5]. In [3], the dimension conversion technique based on DCT has been proved to be a strong candidate for many applications. Therefore, in our SEW magnitude quantization scheme, the process of dimension conversion based on DCT is used as well.

The basic idea of the DCT-based dimension conversion technique is to first perform the discrete cosine transform of the variable-dimensional vector and to zero-pad or truncate the DCT-domain vector to a fixed length prior to quantization. Then at decoder the DCT-domain vector is again truncated or zero-padded to the original length and converted back into the original domain by using inverse transform [5].

Thinking of the computational complexity and memory requirements, it is necessary for a vector quantizer based on DCT to select an appropriate dimension that is as low as possible while still maintaining good quantization quality. The test was performed with twelve different codebook dimensions by using a WI coder to quantize original SEW magnitude spectra. The used measure are the weighed signal-to-noise (WSNR) and the spectral distortion (SD) caused by the quantizer. Their calculated equations are

$$WSNR = 10 \log_{10} \left(\frac{\mathbf{x}^T \mathbf{W} \mathbf{x}}{(\mathbf{x} - \hat{\mathbf{x}})^T \mathbf{W} (\mathbf{x} - \hat{\mathbf{x}})} \right) (dB) \quad (1)$$

and

$$SD = \sqrt{\frac{1}{K} \sum_{k=1}^K \left(10 \log_{10} \frac{x_k^2}{\hat{x}_k^2} \right)^2} (dB) \quad (2)$$

where, \mathbf{x} and $\hat{\mathbf{x}}$ denote the original and quantized SEW magnitude spectra vectors, x_k and \hat{x}_k denote the k th elements of \mathbf{x} and $\hat{\mathbf{x}}$. The elements of the diagonal weighting matrix \mathbf{W} in Equation (1) is given by [6]

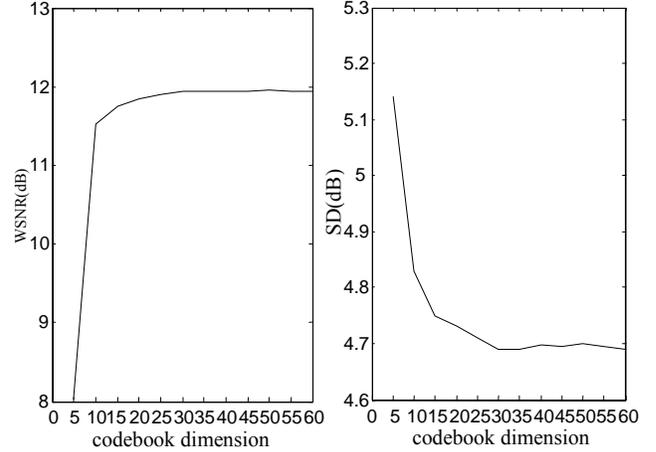


Figure 1. Quantization accuracy of an 8-bit quantizer as a function of codebook dimension

$$w_{kk} = \frac{1}{K} \left| \frac{gA(z/\gamma_1)}{A(z)A(z/\gamma_2)} \right|_{z=e^{j2\pi k/P}}^2 \quad k = 1, \dots, K \quad (3)$$

where the weighting parameters γ_1 and γ_2 were set to 0.94 and 0.6 respectively, g is the corresponding gain and $A(z)$ denotes the LPC polynomial.

The test results are depicted in Figure 1 using the 8-bit full search vector quantizer. The output WSNR and SD of the scheme are illustrated in the left and right part of Figure 1. Evidently, no significant improvement can be achieved by increasing the codebook dimension beyond half of the maximum dimension 60. Therefore, it can be concluded that the codebook dimension 30 offers a good trade-off. The conclusion is consistent with that proposed by [5].

2.2. SEW predictive A-b-S vector quantization

It has been observed that there seems to be some correlation between the successive SEW magnitude spectra and a dual predictive approach for SEW quantization has been proposed [4]. In this paper, another SEW predictive A-b-S vector quantization scheme based on DCT is proposed. Figure 2 illustrates the proposed scheme which uses the past quantized SEW magnitude to predict the current SEW magnitude. At the encoder, the SEW magnitude spectra are first down sampled and transformed into DCT domain. The resulting 30-dimension vectors are quantized by A-b-S vector quantization. At the decoder, the DCT-domain magnitude vectors are transformed back into magnitude spectra.

The WSNR and SD of the scheme are illustrated in Figure 3. The weighted signal-to-noise ratio seems to increase approximately linearly with increasing bit rate, while the spectral distortion seems to decrease approximately

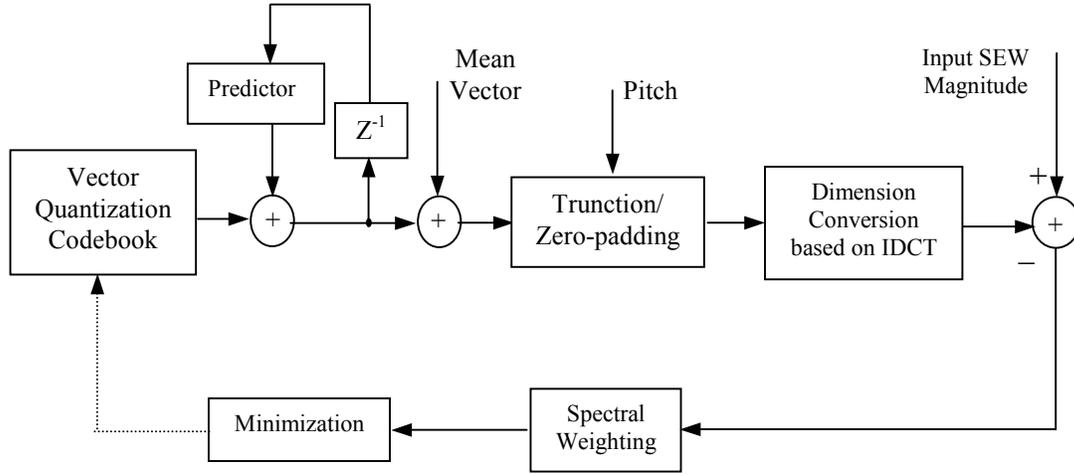


Figure 2. Structure of the predictive A-b-S SEW vector quantization

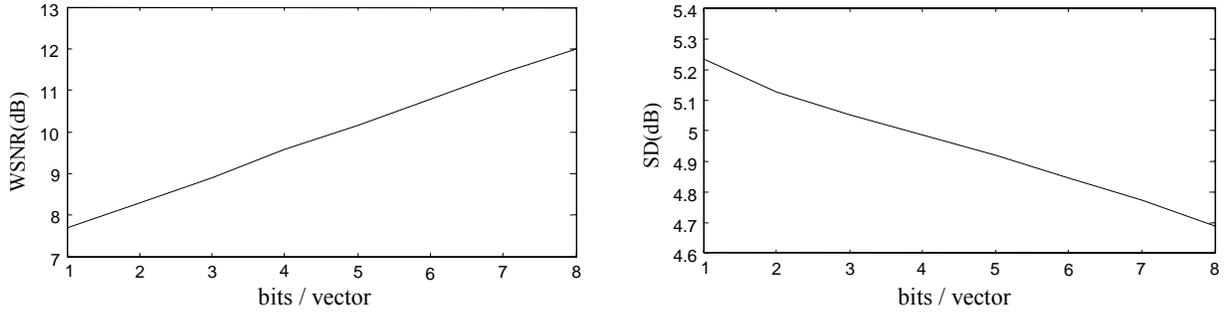


Figure 3. Quantization accuracy of the predictive A-b-S SEW vector quantization

linearly with increasing bit rate. So the quantization accuracy of the SEW magnitude quantizer can increase with increasing bit rate in an objective sense. In our WI implementation, an 8-bit quantizer is employed.

3. REW MAGNITUDE QUANTIZATION

REW magnitude spectrum is still a variable dimension vector. It has been ascertained that one of unique features of REW quantization is low-resolution quantization with a high update rate. A simple and efficient way to obtain a low-resolution representation is to represent the magnitude spectra as a linear combination of a small number of basis functions such as orthonormal polynomial [4] and DCT coefficient [5]. In this paper, the REW magnitude spectra are first converted into a fixed dimension vector, which covers the available frequency range from zero to the Nyquist frequency using the variable dimension of vector. Then the resulting magnitude vector is transformed into DCT-domain vector. Finally, a matrix quantization scheme is performed.

3.1. Combined dimension conversion method

The combined dimension conversion method includes the variable dimension of vector and the DCT-based dimension conversion technique. The variable dimension of vector is based on VDVQ method proposed by [2]. The basic idea is to represent the K -dimensional input vector \mathbf{x} using a K_{MAX} -dimensional vector \mathbf{x}'

$$x'(k) = \begin{cases} x(i), & k = \left\lceil \frac{2 \times i \times K_{MAX}}{P(t)} \right\rceil, i = 1, 2, \dots, K, \quad k = 1, 2, \dots, K_{MAX} \\ 0.0, & \text{otherwise} \end{cases} \quad (4)$$

where, $x'(k)$ denotes the k th element of \mathbf{x}' , $x(i)$ denotes the i th element of \mathbf{x} , and $\lceil \cdot \rceil$ denotes rounding towards infinity. The inverse process is to reconstruct the K -dimensional magnitude vector

$$x(i) = x'(k), \quad k = \left\lceil \frac{2 \times i \times K_{MAX}}{P(t)} \right\rceil, \quad i = 1, 2, \dots, K, \quad k = 1, 2, \dots, K_{MAX} \quad (5)$$

In order to obtain the reduced dimension magnitude vector, we transform \mathbf{x}' into DCT-domain vector using the dimension conversion technique based on DCT. Although in Section 2 the basic process of dimension conversion based on DCT has been introduced, there are some differences in their transform processes. The basic idea of the technique adopted from [7] is to first represent the K_{MAX} -dimensional vector \mathbf{x}' using a p -dimension vector \mathbf{z} that contains the DCT coefficients,

$$z(m) = \sqrt{\frac{u_m}{K_{MAX}}} \sum_{k=1}^{K_{MAX}} x'(k) \cos\left(\frac{\pi(2k-1)(m-1)}{2K_{MAX}}\right), m = 1, 2, \dots, p, \quad (6)$$

where $u_m = \begin{cases} 1 & m = 1 \\ 2 & m \neq 1 \end{cases}$, and $p \leq K_{MAX}$. Then the quantization process is performed for the vector \mathbf{z} in DCT domain. At the decoder, the elements of the K_{MAX} -dimensional magnitude vector are obtained using the inverse transform,

$$\hat{x}'(k) = \sum_{m=1}^p \sqrt{\frac{u_m}{K_{MAX}}} \hat{z}(m) \cos\left(\frac{\pi(2k-1)(m-1)}{2K_{MAX}}\right), k = 1, 2, \dots, K_{MAX} \quad (7)$$

where $\hat{z}(m)$ denotes the m th quantized DCT coefficient. Finally, the quantized magnitude vector can be reconstructed using the equation (5). In our WI implementation, p is set to 5.

3.2. REW magnitude quantization

In [5] and [8], matrix quantization of REW magnitude spectra has been proposed. In this paper, we use a matrix quantizer to quantize the DCT-domain REW magnitude spectra matrix obtained by using the combined dimension conversion method. At the encoder, the update rate for REW magnitude spectra is set to five REW each 25 ms frame and each REW magnitude spectrum is represented with five DCT coefficients. Thus, the resulting matrix consists of 25 elements.

To evaluate the performance of the proposed quantization scheme, we have conducted an informal subjective A/B listening test to compare WI reconstructed speech using either the 10-bit quantizer proposed by [8] or the 4-bit quantizer proposed by this paper. The test data sampled by 8 kHz included 16 speech sentences in Chinese, spoken by female and male speakers. Eight listeners participated in the test. The test results listed in Table 1. As can be seen from Table 1, for male speakers, female speakers, and the total speakers, the 4-bit quantizer preference is 4.2%, 7.7% and 6.0% higher than that of the 10-bit quantizer respectively. Therefore, the subjective quality of the 4-bit quantizer exceeds that of the 10-bit quantizer.

Test	10-bit quantizer	4-bit quantizer	No preference
Male	20.8%	25.0%	54.2%
Female	19.2%	26.9%	53.8%
Total	20.0%	26.0%	54.0%

Table 1. Results of the subjective A/B test

4. SUMMARY

We have presented the quantization schemes for the magnitude spectra of the SEW and REW components in a 2 kb/s WI coder. The frame length is 25 ms, and ten characteristic waveforms are extracted every frame. In this coding scheme, the DCT-domain SEW magnitude spectra vectors are quantized with an 8-bit predictive A-b-S vector quantizer presented in Section 2 and the REW magnitude spectra are quantized with a 4-bit matrix quantizer discussed in Section 3. The results of quantizing SEW and REW magnitude spectra given in this paper indicate that the dimension conversion based on DCT and predictive A-b-S quantization used for SEW magnitude spectra are effective in achieving a good quantization accuracy, and the combined dimension conversion method and matrix quantization used for REW magnitude spectra are efficient in preserving the smoothing of the REW magnitude spectra.

5. REFERENCES

- [1] W. B. Kleijn, and J. Haagen. "Waveform interpolation for coding and synthesis," in *Speech Coding and Synthesis*, W. B. Kleijn and K. K. Paliwal, eds. Amsterdam, Elsevier Science, pp. 175-207, 1995.
- [2] A. Das, A. V. Rao, and A. Gersho, "A variable-dimension vector quantization of speech spectra for low-rate vocoders," *Data Compression Conference*, pp. 420-429, 1994.
- [3] J. Nurminen, A. Heikkinen, and J. Saarinen, "Objective evaluation of methods for quantization of variable-dimension spectral vectors in WI speech coding," in *Proc. Eurospeech 2001*, Aalborg, Denmark, pp.1969-1972, 2001.
- [4] O. Gottesman, and A. Gersho, "High quality enhanced waveform interpolative coding at 2.8kbps," in *Proc. IEEE ICASSP'2000*, pp. 1363-1366, 2000.
- [5] J. Nurminen, A. Heikkinen, and J. Saarinen, "Quantization of magnitude spectra in waveform interpolation speech coding," in *Proc. Norsig 2001*, Trondheim, Norway, pp. 65-69, 2001.
- [6] O. Gottesman, and A. Gersho, "Enhanced waveform interpolative coding at 4 kbps," in *Proc. 1999 IEEE Workshop on Speech Coding*, Porvoo, Finland, pp. 90-92, 1999.
- [7] P. Lupini, and V. Cuperman. "Vector quantization of harmonic magnitudes for low-rate speech coders," in *Proc. IEEE Global Telecommunications Conference*, pp. 858-862, 1994.
- [8] J. Nurminen, A. Heikkinen and J. Saarinen. "A novel quantization scheme for the noise-like component in waveform interpolation speech coding," in *Proc. ICASSP*, pp. 649-652, 2002.