Dither-like Data Hiding in Multistage Vector Quantization of MELP and G.729 Speech Coding

Pao-Chi Chang and Hsin-Min Yu

Dept. of Electrical Eng., National Central Univ.,
Chung-Li, Taiwan 320
pcchang@ee.ncu.edu.tw

Abstract

In this paper, we present a speech data hiding technique that utilizes the characteristics of multistage vector quantization (MSVQ) and subtractive dithering to maintain high speech reconstruction quality. The last stage of MSVQ is used to store the data to be embedded. Similar to the subtractive dithering, the noise like hidden data is subtracted from the first stage of the encoder and added back to the MSVQ decoder. As a result, the degradation caused by hiding secret data is significantly reduced compared to the traditional simple substitution method.

1.Introduction

Data hiding is the art of embedding secret messages in a multimedia signal. While cryptography protects the content of messages, data hiding conceals their very existence and provides extra security. However, most current data hiding techniques developed for speech cannot defense the attack of linear predictive coding (LPC) that is widely used in speech communication systems. Therefore, a different approach that is to hide the secret message in the already compressed bit stream instead of the speech signal itself is considered. It is assumed that no further compression attacks exist before the extraction stage in this case. In this paper, we present a data hiding technique, called dither-like data hiding (DDH) method, which utilizes the characteristics of multistage vector quantization (MSVQ) and subtractive dithering. The DDH method is applied to two commonly used speech coding standards, G.729 [1] and MELP [2], which consist of MSVQ structures. The processed data stream format is still compatible with MELP or G.729 speech coding standard.

DDH method and its basic concepts are presented in the next section. Section 3 describes the technical fundamentals of the DDH method, while Section 4 presents the simulation results when the DDH method is applied to MELP and G.729. Section 5 summarizes this paper.

2. Dither-Like Data Hiding

Dithering is a technique that embeds a noise-like signal into a system. Depending on the purposes and conditions of dither systems, the dither signal can be embedded into different parts in a system. This also results in different reconstruction quality [3] [4]. Generally, the dither system can be classified as subtractive dither system and non-subtractive dither system. In a subtractive dither system the overall quantization noise is equal to the quantization error of the original quantizer. In a MSVQ, the signals in latter stages tend to be less correlated [5]. Therefore, if the index of the last stage is substituted by the data to be hidden, the last stage can be viewed as a random noise generator that generates uncorrelated data with previous stages. By subtracting this "random noise" from the input of the MSVQ encoder, and adding it back at the MSVQ decoder, which is equivalent to a subtractive dither system, the degradation caused by hiding secret data can be reduced compared to the traditional simple substitution method. This method is referred as a dither-like data hiding (DDH) method.

3. Data Hiding Algorithm in DDH

We use a two-stage VQ, illustrated in Fig. 1, as an example to describe how a MSVQ works. Fig. 1(a) and 1(b) represent the encoder and the decoder of a two-stage VQ, respectively. The quantizer Q_1 includes the pair of the encoder E_1 and the decoder D_1. The input vector X is quantized by the first stage VQ encoder E_1, which generates the first stage VQ index, denoted by i_1. The

0-7803-7576-9/02$17.00 © 2002 IEEE

quantized approximation $D_1(i_1)$ is then subtracted from X producing the error vector e_2. This error vector is then applied to a second VQ encoder E_2. The overall approximation $\hat{X}$ to the input X is formed by summing the first and the second VQ outputs, $D_1(i_1)$ and $D_2(i_2)$. The encoder of this VQ scheme simply transmits a pair of indices (i_1 and i_2) specifying the selected code vectors for each stage and the task of the decoder is to perform two table lookups to generate and then sum the two code vectors.

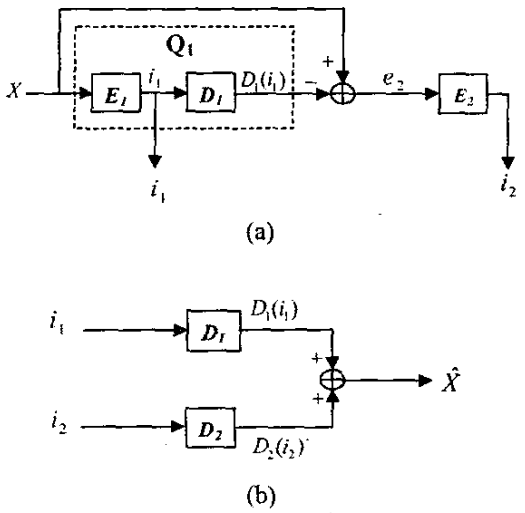

Fig. 1. A two-stage VQ, (a) Encoder (b) Decoder

Fig. 2 shows the conventional data hiding method applied to a two-stage VQ with the second stage index used to hide data, where m denotes the secret data to be hidden. The major difference between the proposed DDH method, illustrated in Fig. 3, and the conventional simple replacement method is that the noise like hidden data is subtracted from the first stage of the encoder and added back to the MSVQ decoder, which is the same as the subtractive dithering, i.e., the code vector indexed by m (denoted by $D_2(m)$) is first extracted and subtracted from X. Both indices i_1 and i_2 (that is the message m) are sent to the decoder, which performs exactly the same procedure as a MSVQ decoder to reconstruct the speech. At the same time, the secret message m is obtained as i_2.

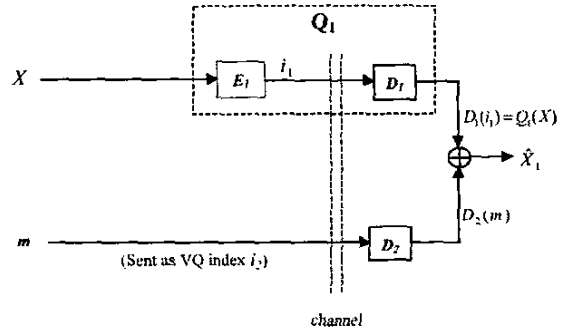

Fig. 2. Block diagram of a conventional simple replacement data hiding method applied to a two-stage VQ

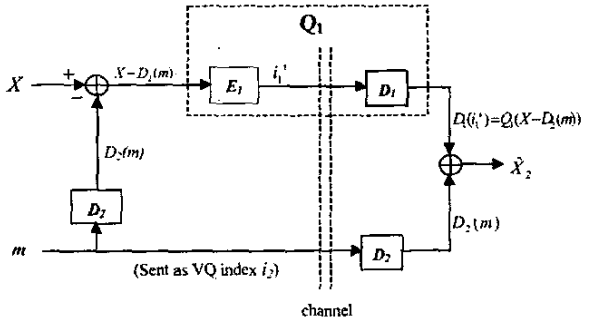

Fig. 3. Block diagram of the proposed DDH method applied to a two-stage VQ

Based on the mean squared-error (MSE) criterion, the conventional simple replacement method results in MSE_1

$$MSE_1 = E[\|X - \hat{X}_1\|^2] = E[\|X - Q_1(X) - D_2(m)\|^2] \quad (1)$$

while the DDH method yields MSE_2

$$MSE_2 = E[\|X - \hat{X}_2\|^2] = E[\|X - Q_1(X - D_2(m)) - D_2(m)\|^2]$$
$$= E[\|(X - D_2(m)) - Q_1(X - D_2(m))\|^2]. \quad (2)$$

Assume Q_1 is a high-resolution quantizer, and

$$\varepsilon_1 = X - Q_1(X) \quad (3)$$

$$\varepsilon_2 = (X - D_2(m)) - Q_1(X - D_2(m)) \quad (4)$$

representing the quantization errors of the first stage and the second stage, respectively, where m is a random number, i.e., the hidden data can be arbitrary, $D_2(m)$ and ε_1 are i.i.d. (independent, identically distributed), then (1) and (2) can be simplified by (3) and (4) to (5) and (6), respectively.

$$MSE_1 = E[\|X - Q_1(X) - D_2(m)\|^2] = E[\|\varepsilon_1 - D_2(m)\|^2]$$
$$= E[\|\varepsilon_1\|^2] - 2E[\varepsilon_1^T D_2(m)] + E[\|D_2(m)\|^2]$$
$$= E[\|\varepsilon_1\|^2] + E[\|D_2(m)\|^2] \quad (5)$$

$$MSE_2 = E[\|(X - D_2(m)) - Q_1(X - D_2(m))\|^2]$$
$$= E[\|\varepsilon_2\|^2] = E[\|\varepsilon_1\|^2] \quad (6)$$

It is observed that $MSE_1 > MSE_2$ by comparing (5) and (6). Namely, in an ideal condition, the distortion caused by the DDH method is smaller than the distortion caused by the conventional simple replacement method. Because the quantization error caused by hiding data is taken into consideration in DDH method, the overall performance represented by the average quantization error is considerably smaller than the quantization error produced by directly replacing the index with the secret data. Note that in a real situation, this conclusion may not always be true if $D_2(m)$, X, and ε_1 are not i.i.d and Q_1 is not a high resolution quantizer.

To apply the DDH method to speech coding systems, we choose the U.S. Military speech coding standard MELP and ITU-T standard G.729 as the target for study since both of them utilize MSVQ in the coding process. Fig. 4 shows the 10-dimensional two-stage VQ with a split second stage, which is used in G.729. Fig. 5 depicts the 10-dimensional four-stage VQ used in MELP. Both of them are designed to quantize the line spectral frequency (LSF). In Fig. 4 and Fig. 5, we assume that each Q denotes a pair of encoder E and decoder D, shown as in Fig. 1. In G.729, since the human ear is less sensitive to the high frequency part of speech, $Q_{2\text{-}2}$ which is designed to quantize high LSF is chosen to hide data. Fig. 6 and Fig. 7 show the two-stage VQ of G.729 and the four-stage VQ of MELP, respectively, when the proposed DDH method is applied.

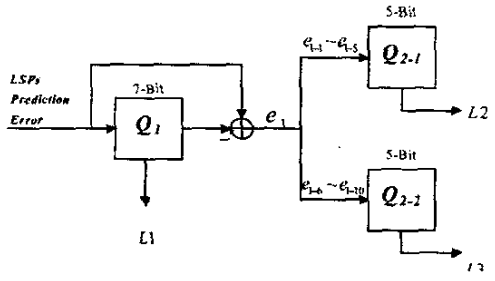

Fig. 4. Two-stage VQ in G.729

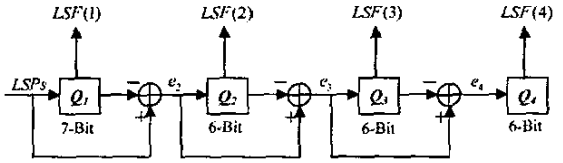

Fig. 5. Four-stage VQ in MELP

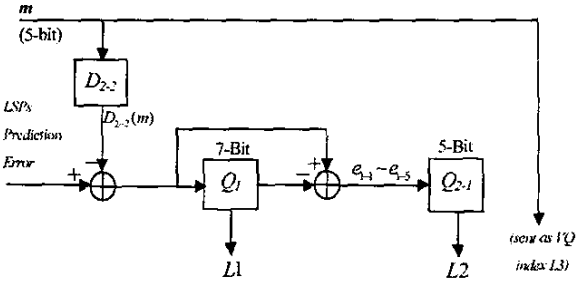

Fig. 6. Block diagram of the proposed DDH method applied to the two-stage VQ in G.729

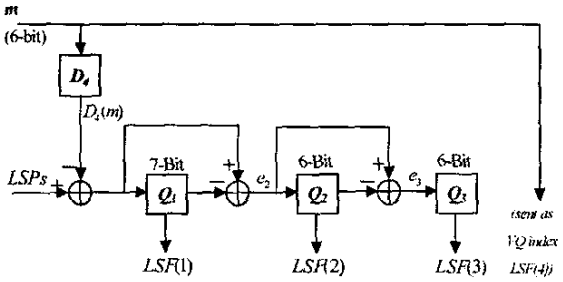

Fig. 7. Block diagram of the proposed DDH method applied to the four-stage VQ in MELP

4. Performance Evaluation

4.1. Data embedding capacity

The proposed DDH method has a fixed data hiding rate because the hidden date are placed at the same and a fixed position in a frame. MELP contains three frame types: voice, jitter voice, and unvoice. However, the data hiding positions in the proposed method are located in the parameters that all frames have. Therefore, the embedding rate is not affected by switching frame types. This method can embed 6 bits per frame to MELP. With the frame size 22.5 ms, the embedding rate for MELP is about 266.67 bps. G.729 is a fixed rate codec. Therefore, the embedding rate is also fixed. The proposed method can embed 5 bits per frame to G.729. With the frame size 10 ms, G.729 can provide 500 bps embedding rate.

4.2. Objective measurements

Since both of the two MSVQs are used to quantize the LSF, the average spectral distortion (*sd*) is chosen to measure the quantization performance. The average *sd* has been used extensively to measure LPC quantization performance. The spectral distortion for the *i*th frame, sd_i, is defined (in dB^2) as follows [6]:

$$sd_i = \frac{1}{N/2}\sum_{k=0}^{N/2-1}\left[10\log_{10}\frac{|A_i'(k)|^2}{|A_i(k)|^2}\right]^2 \qquad (7)$$

where N is the number of DFT points, $A_i'(k)$ is the k-th DFT coefficient in i-th frame of LPC all-pole filter after LSF quantization, and $A_i(k)$ is the k-th DFT coefficient of i-th frame of LPC all-pole filter before LSF quantization. The overall distortion (*SD*) can be calculated by summing up a sequence of *sd*

$$SD = \frac{1}{M}\sum_{i=1}^{M} sd_i \qquad (8)$$

where M is the total number of speech frames.

Table 1 lists the average SD of the original MSVQ and the two data hiding methods. The DFT used in the calculation is 1024-point. The data to be hidden was generated at random. As shown in Table 1, the performance of the proposed DDH method is significantly better than the direct replacement method.

Table 1. Spectral Distortion (SD) measurement

	SD (in dB^2)		
	Original MSVQ	Simple replacement method	Proposed DDH method
MELP	2.584	6.301	4.538
G.729	1.416	6.497	5.232

Note: the secret data was generated at random

4.3. Subjective measurements

The method used in the subjective listening test for MELP and G.729 is ITU-T defined Comparison Category Rating (CCR) method [7][8]. In the CCR procedure, listeners are presented with a pair of data embedded and no data embedded sentences on each trial, a short period of silence, and another pair of sentences. Listeners have to evaluate the quality of the second sentence compared to the quality of the first sentence in each pair. The order of the data embedded speech and no data embedded speech is chosen at random for each trial. The listeners rate the quality of the second sample relative to the quality of the first sample using the scale as shown in Table 2.

Table 2. The rating scale used in the CCR test

Score	Description
3	Quality of the second is much better than quality of the first.
2	Quality of the second is better than quality of the first.
1	Quality of the second is slightly better than quality of the first.
0	Quality of the second and quality of the first are about the same.
-1	Quality of the second is slightly worse than quality of the first.
-2	Quality of the second is worse than quality of the first.
-3	Quality of the second is much worse than quality of the first.

In the subjective test, the test sentences include ten Chinese sentences and three English sentences, spoken by both male and female. Twenty listeners performed the test. The overall average test results when DDH method is applied to MELP and G.729 are shown in Table 3. The data to be hidden was also generated at random.

Table 3. The results of CCR listening test

Codec	Simple replacement method	DDH method embedded
MELP	-0.45	-0.02
G.729	-0.12	-0.03

Note: the secret data was generated at random

As shown in Table 3, the performance of the proposed DDH method is better than the direct replacement method and nearly imperceptible to the original speech codecs.

5. Summary

In this paper, we have presented a speech data hiding technique that utilizes the characteristics of MSVQ and subtractive dithering to maintain high speech reconstruction quality. The proposed DDH method is also compatible with the original speech codec. Simulations show that the performance of the proposed DDH method is significantly better than the direct replacement method both objectively and subjectively. No restrictions are applied to the hidden data in this method.

6. References

[1] ITU-T Recommendation G.729, "Coding of speech at 8 kbit/s using Conjugate-Structure Algebraic-Code-Excited Linear-Prediction (CS-ACELP)," March 1996.

[2] L. M. Supplee, R. P. Cohn, J. S. Collura, A. V. McCree, "MELP : the new federal standard at 2400 bps," *IEEE ICASSP*, vol. 2, 1997, pp. 1591-1594.

[3] B. Chen, G. W. Wornell, "Quantization index modulation : a class of provably good methods for digital watermarking and information embedding," *IEEE Trans. on Information Theory*, vol. 47, no. 4, May 2001, pp. 1423-1443.

[4] R. M. Gray, T. G. Stockham, Jr., "Dithered quantizers," *IEEE Trans. on Information Theory*, vol. 39, no. 3, May 1993, pp. 805-812.

[5] A. Gersho, R. M. Gray, *Vector Quantization and Signal Compression*, Boston: Kluwer, 1992.

[6] A. M. Kondoz, *Digital Speech Coding for Low Bit Rate Communications Systems*, Wiley, 1994.

[7] ITU-T Recommendation on P.800, "Methods for subjective determination of transmission quality," Aug. 1996.

[8] S. Yeldener, J.C. de Martin, and V. Viswanathan, "A mixed sinusoidally excited linear prediction coder at 4 KB/S and below," *Proc. IEEE ICASSP*, vol. 2, 1998, pp. 589-592.