Journal of Information Hiding and Multimedia Signal Processing ©2010 ISSN 2073-4212
Ubiquitous International Volume **1**, Number **1**, January **2010**

Designing Side Information of Multiple Description Coding

Akinori Ito and Shozo Makino

Graduate School of Engineering
Tohoku University
6-6-05 Aramaki aza Aoba, Sendai, 980-8579 Japan
{aito,makino}@makino.ecei.tohoku.ac.jp

Received March 2009; revised June 2009

Abstract. *In this paper we investigated methods for designing side information of multiple description coding when transmitting two values independently. For methods that use one bit, we investigated '1-bit quantization,' 'sign correction' and 'difference quantization' methods. For those that use two bits, we investigated '2-bit quantization,' 'sign correction+difference quantization' methods. From theoretical analysis and numerical experiments, it has been found that the quantization-based method is best when correlation of the original data is weak, while 'difference quantization' or combination of sign correction is better when the original data have strong correlation. Then we applied the methods to multiple description coding of speech signals.*
Keywords: Multiple Description Coding, Packet Loss Concealment, Correlation, Side Information

1. **Introduction.** Multiple description coding (MDC) [1] is a technique to split one multimedia data into two or more 'equally important' parts. An MDC is useful for increasing quality of multimedia data transmitted over a lossy packet network [2]. Various methods for realizing MDC have been proposed so far. The correlating transform (CT) [3] is one of the most promising methods for MDC. The CT is a method to give correlation to two independent data. As the CT is a kind of linear transform, it does not use any explicit extra data to improve the quality of the recovered data (although it requires more than a half of the original bandwidth to keep the quality of the signal without packet loss equal to the original quality).

The multiple description scalar quantizer (MDSQ) [4] is a method of splitting one quantized value into two values. Using the MDSQ, we can split any scalar value into two descriptions. The drawback of the MDSQ is that the efficiency of quantization decreases if we want to increase the quality of restored signal when only one description is received.

The other way of realizing MDC is to combine fine quantization and coarse quantization. Let $Q_f(x)$ be a fine quantizer for data x and $Q_c(x)$ be a coarse quantizer for x. When sending a pair of data (x_1, x_2), we generate two descriptions

$$\begin{aligned} y_1 &= \langle Q_f(x_1), Q_c(x_2)\rangle \\ y_2 &= \langle Q_f(x_2), Q_c(x_1)\rangle \end{aligned} \quad (1)$$

and send y_1 and y_2 independently. At the receiver side, when we receive both $y_1 = \langle p_1, q_1 \rangle$ and $y_2 = \langle p_2, q_2 \rangle$, we recover the original data by

$$\begin{aligned} \hat{x}_1 &= Q_f^{-1}(p_1) \\ \hat{x}_2 &= Q_f^{-1}(p_2) \end{aligned} \tag{2}$$

and thus q_1 and q_2 are abandoned. When only y_1 is received, the original data are estimated as follows.

$$\begin{aligned} \hat{x}_1 &= Q_f^{-1}(p) \\ \hat{x}_2 &= Q_c^{-1}(q) \end{aligned} \tag{3}$$

Jiang and Ortega [5] proposed an MDC method for PCM-coded audio signal. Their method combines finely quantized even-numbered samples with coarsely quantized odd-numbered samples, and vice versa. They exploited 12bit PCM encoder as a fine quantizer and 4bit ADPCM encoder as a coarse quantizer. Let the input samples be $x_1, x_2, \ldots$, and let $Q_P(x)$ and $Q_A(x)$ be PCM and ADPCM quantizers respectively. The method by Jiang and Ortega encodes the input signal as the two descriptions,

$$\langle Q_P(x_1), Q_A(x_2) \rangle, \langle Q_P(x_3), Q_A(x_4) \rangle, \ldots$$

and

$$\langle Q_P(x_2), Q_A(x_1) \rangle, \langle Q_P(x_4), Q_A(x_3) \rangle, \ldots$$

When only the first description is received, we restore the original signal from the following code sequence.

$$Q_P(x_1), Q_A(x_2), Q_P(x_3), Q_A(x_4), \ldots$$

This method performs well, but this method has two problems. First, it is difficult to apply this method to other application such as video or high-efficiency audio codec, because this method strongly depends on PCM and ADPCM codecs. Second, it does not utilize the fact that two contiguous speech samples have strong correlation. If we use this fact, we can develop better method.

Therefore, we propose a general framework of designing side information. In our framework, a pair of values is transmitted and one or two bits of side information is appended to one value for estimating the other value of the pair. First we give a mathematical discussion for a case where two values have no correlation. Then we carry out a simulation for a case that has correlation. Finally, we apply the proposed method to a PCM-based speech transmission.

2. **Problem formulation.**

In this section we formulate the problem we want to discuss. Let X and Y be random variables to be transmitted, which obey distribution $N(0, \sigma^2)$. Let $f(X, Y)$ be a function that maps (X, Y) into a value (either a scalar or a vector) that has one- or two-bit information. Suppose that we transmit a pair (X, Y) using two independent channels. Now we transmit $X' = (X, f(X, Y))$ and $Y' = (Y, f(Y, X))$ independently. When both X' and Y' are received, we can recover the original data (X, Y) without using $f(X, Y)$ or $f(Y, X)$. When Y' is lost, we recover Y by

$$\hat{Y} = g(X, f(X, Y)). \tag{4}$$

The problem is in how to design f and g so that it maximizes correlation between Y and $\hat{Y}$.

Note that, in most cases, the problem is not to maximize correlation between Y and $\hat{Y}$ but to minimize mean squared error between Y and $\hat{Y}$, i.e. $E[(Y-\hat{Y})^2]$. As Y is assumed to obey $N(0,\sigma^2)$,

$$E[(Y-\hat{Y})^2] = E[Y^2] + E[\hat{Y}^2] - 2E[Y\hat{Y}] \tag{5}$$

$$= E[Y^2]E[\hat{Y}^2]\left(\frac{1}{E[\hat{Y}^2]} + \frac{1}{E[Y^2]} - 2\frac{E[Y\hat{Y}]}{E[Y^2]E[\hat{Y}^2]}\right) \tag{6}$$

$$= \sigma^2 E[\hat{Y}^2]\left(\frac{1}{\sigma^2} + \frac{1}{E[Y^2]} - 2\rho(Y,\hat{Y})\right) \tag{7}$$

where $\rho(Y,\hat{Y})$ is a correlation coefficient between Y and $\hat{Y}$. Therefore, assuming $E[\hat{Y}^2] \approx \sigma^2$, minimizing error is equivalent to maximizing correlation between Y and $\hat{Y}$.

3. **Uncorrelated case for 1-bit extra information.** First we investigate the case where X and Y have no correlation and $f(X,Y)$ gives only 1-bit information.

3.1. **1-bit quantization.** The simplest f is a 1-bit quantizer, i.e.

$$f(X,Y) = sign(Y) = \begin{cases} 1 & \text{if } Y \geq 0 \\ -1 & \text{otherwise} \end{cases} \tag{8}$$

$$g(X,c) = c\alpha \ (\alpha > 0) \tag{9}$$

In this case, the correlation coefficient between Y and $\hat{Y} = g(X, f(X,Y))$ is $\sqrt{2/\pi} \approx 0.7979$ regardless of the value of α. $E[(Y-\hat{Y})^2]$ can be minimized when $\alpha = \sqrt{2/\pi}\sigma$.

3.2. **Sign correction.** Next, we investigate a method based on $sign(Y)$. Sign of Y has one-bit information. Using sign information, we can improve correlation. This method corrects the sign of $\hat{Y}$ using the extra information, still the absolute value is taken from X.

$$f(X,Y) = sign(Y) \tag{10}$$

$$g(X,c) = c|X| \tag{11}$$

This method uses the absolute value of X instead of a constant value in Eq. (9). It is not a good idea to use $|X|$ when X and Y have no correlation. However, if $|X|$ and $|Y|$ are correlated, it is capable of providing a good estimation.

When X and Y are uncorrelated, the correlation coefficient of Y and $\hat{Y}$ is $2/\pi \approx 0.6366$, which is worse than 1-bit quantization.

3.3. **Quantization of difference.** Now we consider another method based on quantization of $X-Y$. When X and Y have some correlation, it becomes reasonable to quantize the difference between X and Y. This idea can be realized as follows.

$$f(X,Y) = sign(Y-X) \tag{12}$$

$$g(X,c) = X + \beta c \tag{13}$$

When X and Y are uncorrelated, the maximum correlation between Y and $\hat{Y}$ can be obtained when $\beta = \sqrt{\pi}\sigma$, and correlation coefficient is $1/\sqrt{\pi-1} \approx 0.6833$, that is better than sign correction but worse than 1-bit quantization.

4. **Uncorrelated case for 2-bit extra information.**

4.1. **2-bit quantization.** Next, let us consider when 2-bit information is available. The straightforward extension of a 1-bit case is to use 2-bit quantization. In this case, we use the following f and g:

$$q_2(x,t) = \begin{cases} -t & \text{if} \quad x < -t \\ -1 & \text{if} \quad -t \le x < 0 \\ 1 & \text{if} \quad 0 \le x < t \\ t & \text{if} \quad t \le x \end{cases} \tag{14}$$

$$f(X,Y) = q_2(Y/\alpha, k) \tag{15}$$

$$g(X,c) = c\alpha \tag{16}$$

where $\alpha = \sqrt{2/\pi}\sigma$. In this case, correlation coefficient of Y and $\hat{Y}$ is

$$\rho(Y,\hat{Y}) = \frac{2e^{-\frac{k^2}{\pi}}(e^{\frac{k^2}{\pi}} + k - 1)}{\sqrt{2\pi\left((1-k^2)\Phi\left(\frac{k}{\sqrt{\pi}}\right) + k^2\right)}} \tag{17}$$

where

$$\Phi(z) = \frac{2}{\sqrt{\pi}} \int_{-\infty}^{z} e^{-x^2} dx. \tag{18}$$

This correlation coefficient becomes maximum when $k \approx 1.8507$ and the value of the correlation coefficient is 0.8867.

4.2. **Combination of sign correction and difference quantization.** Next let us consider a method based on a combination of one-bit methods. The sign correction method and the difference quantization method can be combined, which requires 2-bit information. In this case,

$$f(X,Y) = (sign(Y), sign(Y-X)) \tag{19}$$

$$g(X,(c_1,c_2)) = c_1|X| + \beta c_2. \tag{20}$$

In this case it is difficult to calculate the correlation coefficient analytically. Therefore we carried out a numerical experiment, and we obtained the maximum correlation coefficient of 0.8562 when $\beta = 0.66$.

5. Numerical experiment for correlated case.

5.1. **Experimental conditions.** When X and Y are correlated, the correlation between Y and $\hat{Y}$ becomes larger for 'sign correction' and 'quantization of difference' method. However, it is difficult to calculate an analytical solution for such cases. Therefore, we carried out a numerical experiment using a Monte Carlo simulation for the case that X and Y have positive correlation.

First, we prepared 30000 pairs of random numbers (x_i, y_i), where x_i and y_i obeys independent Gaussian distribution $N(0,1)$. Next, we calculated (x'_i, y'_i) by

$$\begin{pmatrix} x'_i \\ y'_i \end{pmatrix} = \frac{1}{\sqrt{1+\gamma^2}} \begin{pmatrix} 1 & \gamma \\ 1 & -\gamma \end{pmatrix} \begin{pmatrix} x_i \\ y_i \end{pmatrix}. \tag{21}$$

where $0 \le \gamma \le 1$. Using this transform, correlation coefficient of x'_i and y'_i becomes $(1-\gamma^2)/(1+\gamma^2)$ while x'_i and y'_i obeys $N(0,1)$. Then the data $\{(x'_i, y'_i)\}$ are used for the simulation.

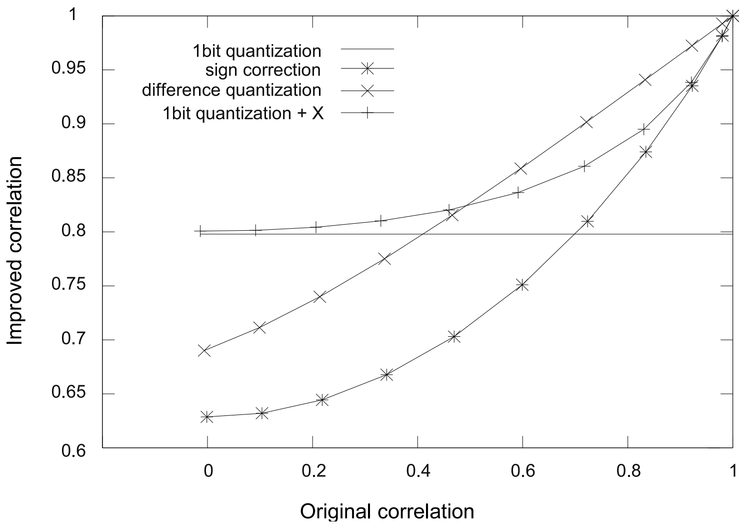

FIGURE 1. Improvement of correlation coefficient using 1 bit

5.2. **1-bit case.** Using these data, we calculated the correlation coefficient between Y and $\hat{Y}$ when using the methods described in section 3. In addition to these methods, we investigated the combination of 1-bit quantization and X. Using 1-bit quantization, we can know whether Y is positive or negative. If X and Y have some correlation, combining the quantized value of Y with X would improve the correlation between Y and $\hat{Y}$. Therefore, we examined the following method.

$$f(X, Y) = sign(Y) \tag{22}$$

$$g(X, c) = wX + (1 - w)c\sqrt{\frac{2}{\pi}}\sigma \tag{23}$$

Here, w is a weighting factor to be determined.

Figure 1 shows the experimental results. When the correlation coefficient of the original data is small, '1-bit quantization' or 'combination of 1-bit quantization and X' gave the higher correlation. When the correlation coefficient of the original data is larger than 0.5, 'difference quantization' method gave the best correlation.

Figure 2 shows the optimum parameter (β for 'difference quantization' and w for 'quantization+X', 'quantization+$|X|$). As for β, the optimum value monotonically decreases from $\sqrt{\pi}\sigma$ to nearly zero. As the change of improved correlation around the optimum β is not large, it seems reasonable for approximating the optimum value of β as

$$\beta = \sqrt{\pi}\sigma(1 - \rho(X, Y)) \tag{24}$$

where $\rho(X, Y)$ stands for the correlation coefficient of X and Y.

As for the combination weight w for 'quantization+X' method, $w = 0$ for $\rho(X, Y) = 0$ and $w = 1$ for $\rho(X, Y) = 1$. The value of w becomes 0.5 when $\rho(X, Y) = 0.82$, which means that contribution of the quantized value and X become equivalent when the correlation of the original data is comparable to the correlation of the quantized value (≈ 0.8). Here, we can approximate the optimum w as

$$w = \frac{\lambda_1}{\lambda_1 + \lambda_2} \tag{25}$$

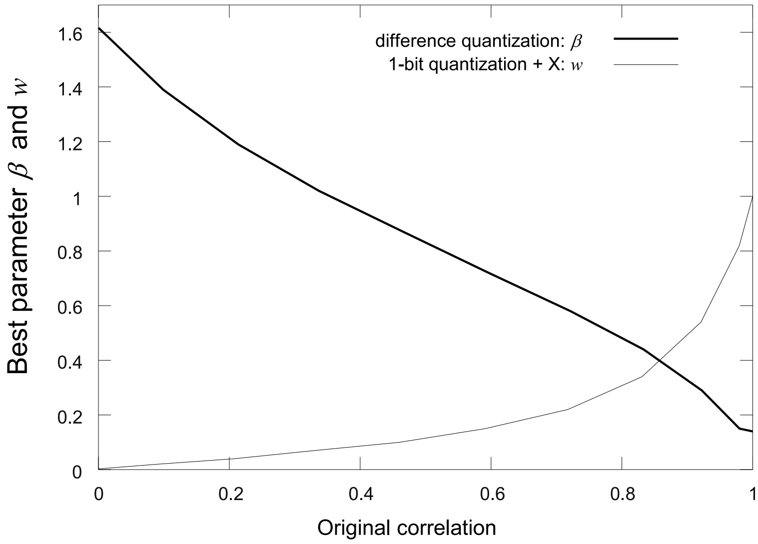

FIGURE 2. Optimum value of β and w

where

$$\lambda_1 = \frac{|\rho(X,Y)|}{1-|\rho(X,Y)|} \tag{26}$$

$$\lambda_2 = \frac{|\rho(sign(Y),Y)|}{1-|\rho(sign(Y),Y)|} \tag{27}$$

5.3. **2-bit case.** Next we investigate the case where 2 bits are available for enhancing the correlation. In addition to the methods described in section 4, we investigated the following two methods.

1. 2-bit difference quantization
   We examined 2-bit version of difference quantization.

$$f(X,Y) = q_2((Y-X)/g_1, g_2/g_1) \tag{28}$$

$$g(X,c) = X + g_1 c \tag{29}$$

   Here, g_1 and g_2 are quantization steps to be optimized.
2. Combination of 2-bit quantization and X
   The 2-bit quantization value can be combined with X when the original correlation is high.

$$f(X,Y) = q_2(Y/\alpha, k) \tag{30}$$

$$g(X,c) = wX + (1-w)c\alpha \tag{31}$$

The experimental result is shown in Figure 3. Similar to the 1-bit case, the combination of 2-bit quantization and X gave good correlation when the original correlation is under 0.5. When the original correlation was high, the 2-bit difference quantization was the best of all methods.

Figure 4 shows the optimum values of β, w, g_1 and g_2. The value of optimum β for zero correlation data was smaller than $\sqrt{2/\pi}$, because the correlation of the original data was raised first by the sign correction. Parameters g_1 and g_2 can be approximated as

$$g_i = G_i(1-\rho(X,Y))^2 \tag{32}$$

where ρ is the correlation coefficient of the original data, $G_1 = 0.85$ and $G_2 = 1.89$.

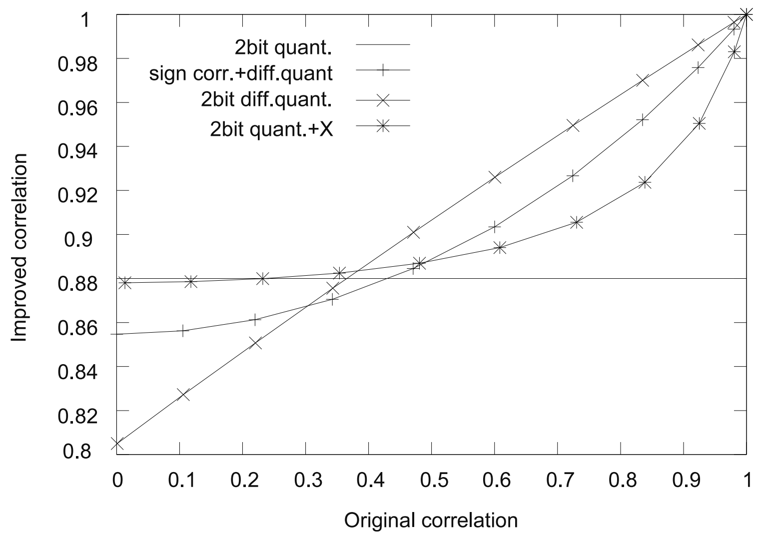

FIGURE 3. Improvement of correlation coefficient using 2 bits

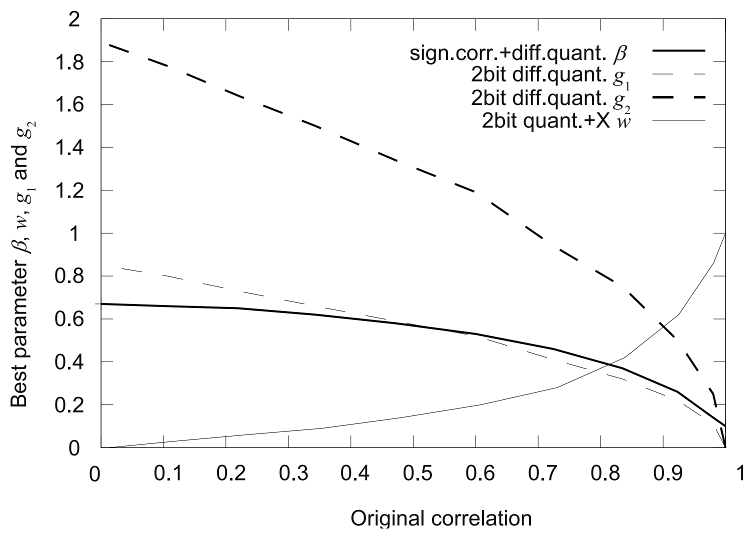

FIGURE 4. Optimum value of β, w, g_1 and g_2

5.4. **When correlation is negative.** The above discussion assumes that X and Y have positive correlation. So what happens when they have negative correlation?

The 1-bit and 2-bit quantization methods are not affected by correlation between X and Y because they only needs the sign and variance of the sample to be restored.

Consider the sign correction method. When X and Y have negative correlation, it means that $-X$ and Y have positive correlation. The sign correction method estimates Y as

$$\hat{Y} = sign(Y)|X| = sign(Y)|-X| \tag{33}$$

Therefore, the result of sign correction method under negative correlation is exactly same as that under positive correlation.

Next, consider the difference quantization method. As explained before, when X and Y have negative correlation, then $-X$ and Y have positive correlation. Therefore, if we

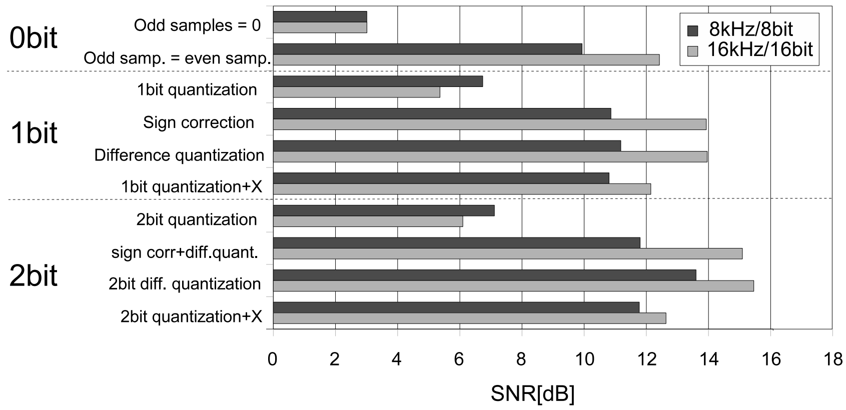

Figure 5. Experimental result

know that X and Y have negative correlation, we can use the following formulae.

$$f(X, Y) = sign(Y + X) \tag{34}$$

$$g(X, c) = -X + \beta c \tag{35}$$

However, when the correlation between X and Y changes depending on time, it will be difficult to apply this method.

6. **Application to a speech signal.** We investigated if these methods are effective for a real signal. As an example, we applied the methods to recover a speech signal split into even and odd samples[6]. We simulated a situation in which the odd samples are lost, and the lost samples were recovered as follows.

$$\hat{x}_{2i+1} = g(x_{2i}, f(x_{2i}, x_{2i+1})) \tag{36}$$

The speech samples for the experiment were taken from the ASJ continuous speech database [7]. We used 100 sentence utterances, 50 of which were uttered by a male speaker and the other 50 were uttered by a female speaker. The speech samples were encoded in two conditions: 16 kHz sampling/16 bit linear quantization and 8 kHz sampleing/8 bit linear quantization. We performed automatic voice activity detection, and the non-speech parts were excluded from computation. Even and odd samples were in strong correlation; the correlation coefficient of the 16kHz/16bit data is 0.94 and that of the 8kHz/8bit data is 0.90.

Figure 5 shows the SNR result of the methods. We tried methods that used additional 0 bit information (odd samples=0 and odd=even), 1 bit information (1bit quantization, sign correction, difference quantization and 1bit quantization+X) and 2-bit information (2bit quantization, sign correction+difference quantization, 2bit difference quantization, 2bit quantization+X).

When using 1 bit of extra information, the ‘difference quantization’ method gave the best result, while the ‘sign correction’ and ‘1bit quantization+X’ methods gave almost as good result as the difference quantizaiton method. The tendency of the qualities by all methods was almost similar for both of 16kHz/16bit and 8kHz/8bit data, although the absolute quality of the 16kHz/16bit data was better than that of 8kHz/8bit data. The quality of “difference quantization” method was not very different from other method, contrary to expectations from Figure 1. The reason why the performance of the difference quantization method was as good as the other methods seems to be a fact that the

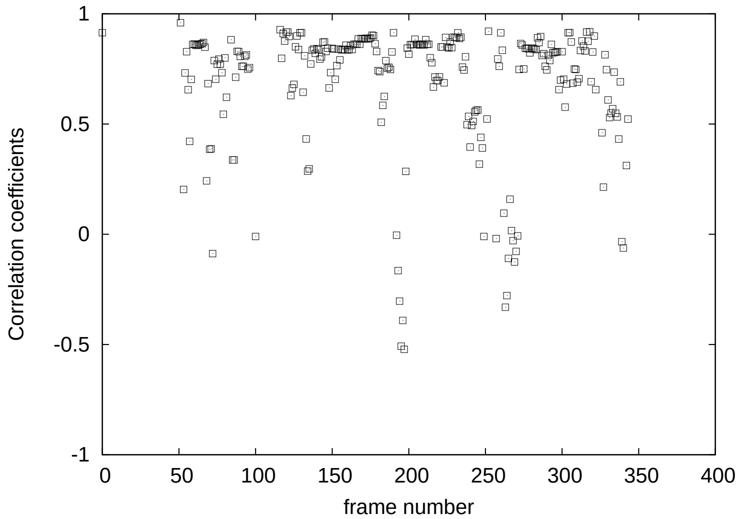

FIGURE 6. Frame number and correlation coefficients

correlation coefficient between even and odd samples depends on time. Figure 6 shows correlation coefficients of a speech signal, calculated frame by frame. In this figure, one frame includes 100 samples. From this figure, it is obvious that the local correlation become smaller or even negative in some frames, although the correlations are high in the most frames. As the difference quantization method in this experiment assumed that the correlation was uniform throughout the signal, the quality of the restored signal degrades in those parts where correlation was low.

When using 2 bits, the '2bit difference quantization' method gave the best result. This method is very similar to Jiang's method[5] where 2bit DPCM is employed as a coarse quantizer.

Figure 7 shows the rate-distortion plot of the results by all methods for 16kHz/16bit data. From this result, we can understand that the 'sign correction+difference quantization' and '2bit quantization+X' methods have little advantage over the '1bit difference quantization' method considering the increase of bitrate.

In summary, we could confirm the improvement of the proposed method when applied to speech signal encoded in PCM. Application to the proposed method to MP3-encoded music signal can be found in [8].

7. **Conclusions.** We analyzed how one or two bits of extra information can improve correlation between two values. From the results of a numerical experiment, it was found that the best method differs according to the correlation of the original data. When correlation of two values is low, the method based on quantization gives better result; when they have high correlation, quantization of difference is the best method. Then, we carried out an experiment in which we applied the method to a real speech signal.

Note that the proposed method can be used not only for two independent samples such as even and odd sample pair but also for two descriptions generated using the conventional MDC methods such as CT and MDSQ. In this case, two kinds of redundancies, R_1 and R_2, are included in a description; R_1 is caused by the first MDC, and R_2 is caused by the proposed method. Here, it is not still clear that the improvement obtained by the proposed method is bigger than the improvement of the conventional MDC when using the redundancy of $R_1 + R_2$. We need more work to clarify the performance of the combination of MDC and the proposed method.

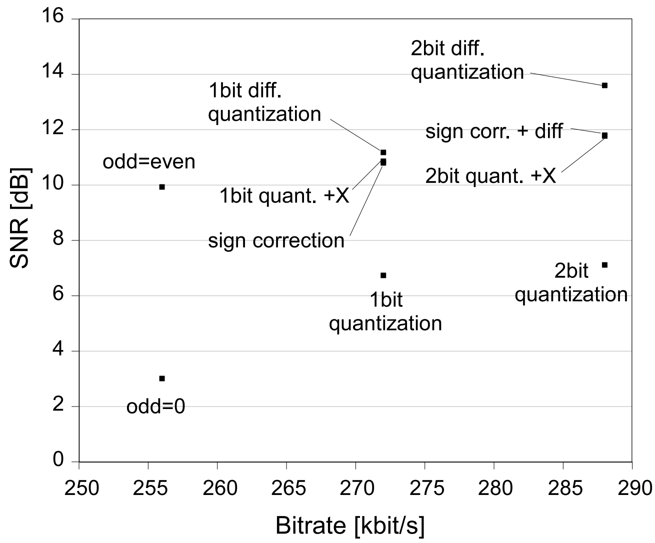

Figure 7. Rate-distortion plot for 16kHz/16bit condition

As a future work, we want to apply this framework to the real application of MD coding such as video coding [9].

Acknowledgement. This study was partly supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) No. 051302004 of the Ministry of Internal Affairs and Communications of Japan.

REFERENCES

[1] V. K Goyal, Multiple description coding: compression meets the network, *IEEE Signal Processing Magazine*, pp. 74–93, 2001.

[2] Y. Lee, K. Joohee, Y. Altunbasak and R. M. Mersereau, Layered coded vs. multiple description coded video over error-prone networks, *Signal Processing: Image Communication*, vol. 18, pp. 337-356, 2003.

[3] V. K. Goyal and J. Kovačević, Generalized multiple description coding with correlating transform, *IEEE Trans. Inf. Theory*, vol. 47, no. 6, pp. 2199-2224, 2001.

[4] V.A. Vaishampayan, Design of multiple description scalar quantizers, *IEEE Trans. Inform. Theory*, vol. 39, pp. 821–834, 1993.

[5] W. Jiang, and A. Ortega, Multiple description speech coding for robust communication over lossy packet networks, *Proc. of IEEE-ICME 2000*, vol. 1, pp. 444–447, 2000.

[6] N.S. Jayant and S.W. Christensen, Effects of packet losses in waveform coded speech and improvements due to an odd-even sample-interpolation procedure, *IEEE Trans. Commun.*, vol. 29, pp. 101-109, 1981.

[7] T. Kobayashi, S. Itahashi, S. Hayamizu and T. Takezawa, ASJ continuous speech corpus for research, *Journal of the Acoustical Society of Japan,* vol. 48, no. 12, pp. 888–893, 1992 (in Japanese).

[8] A. Ito, K. Konno, S. Makino and M. Suzuki, Packet loss concealment for MDCT-based audio codec using correlation-based side information, *Proc. of IIH-MSP*, pp. 612–615, 2008.

[9] A. Reibman, H. Jafarkhani, Yao Wang, and M. Orchard, Multiple description video using rate-distortion splitting, *Proc. of Int. Conf. on Image Processing* , vol. 1, pp. 978–981, 2001.