Journal of Information Hiding and Multimedia Signal Processing ©2010 ISSN 2073-4212
Ubiquitous International Volume **1**, Number **4**, October **2010**

Multiple Description Coding Using Time Domain Division for MP3 coded Sound Signal

Ho-seok Wey[1,2], Akinori Ito[3], Takuma Okamoto[1,3], and Yoiti Suzuki[1,3]

[1]Research Institute of Electrical Communication,Tohoku University
2-1-1, Katahira, Aoba-ku, Sendai, Miyagi 980-8577, JAPAN
[2]Graduate School of Information Sciences, Tohoku University, JAPAN
Tel.: +81-22-217-5461, Fax.: +81-22-217-5535, E-mail: {*fhswey, okamoto, yoh*}@ais.riec.tohoku.ac.jp
[3]Graduate School of Engineering, Tohoku University,
6-6-05 Aramaki aza Aoba, Aoba-ku, Sendai, Miyagi 980-8579, JAPAN
Tel.,Fax.: +81-22-795-7084, E-mail: aito@makino.ecei.tohoku.ac.jp

Received April 2010; revised May 2010

Abstract. *In audio communications over a lossy packet network, packet loss concealment techniques are needed to mitigate a user's frustration when perceiving the deterioration of the quality of the decoded signal. Multiple description coding (MDC) is a useful solution to this problem. In this paper, we describe an MDC method for concealing packet losses for wideband sound signal streaming based on the sample splitting method in the time domain and encoding by an MPEG-1 audio layer III (MP3) encoder. To enhance the quality of the restored signal, we applied a Wiener filter to the higher frequency part of the restored signal. Experiments were conducted to compare the proposed method with several conventional methods, conrming that the proposed method showed higher quality results than the conventional methods for a range of bit rates from 128 to 192 kbps when there were heavy packet losses.*

Keywords: Audio Coding, MP3, Packet loss concealment, Multiple description coding

---

1. **Introduction.** Along with the rapid advancement of computers and network technologies, numerous technologies related to multimedia information distribution via networks based on the Internet Protocol (IP) have been widely and actively studied. In streaming distribution of large and high-quality multimedia data, such as video and sound information, the Transmission Control Protocol (TCP) [1] is currently widely used. If the TCP is used for broadcasting multimedia communications, however, unicast communications are necessary between a server and the destinations, resulting in heavy trac in the network. Therefore, IP multicast-based distribution using the User Datagram Protocol (UDP) [2] or the Realtime Transport Protocol (RTP) [3] seems suitable for these kinds of applications because these protocols simply send data to many destinations simultaneously without resending the same data. In simple transmission of packet data using UDP or RTP without any Quality of Services (QoS) control, however, packet losses inevitably occur. In this case, methods are required to mask the degradation caused by the packet losses. This process is called Packet Loss Concealment (PLC). Various PLC methods have been proposed [4, 5] and can be roughly classied into two types. One is sender-based repair and the other is received-based repair. However, most of the former methods such as Forward Error Correction (FEC) [68] have a drawback, i.e., the bandwidth in the network path is increased because the FEC requires extra data in addition to the original

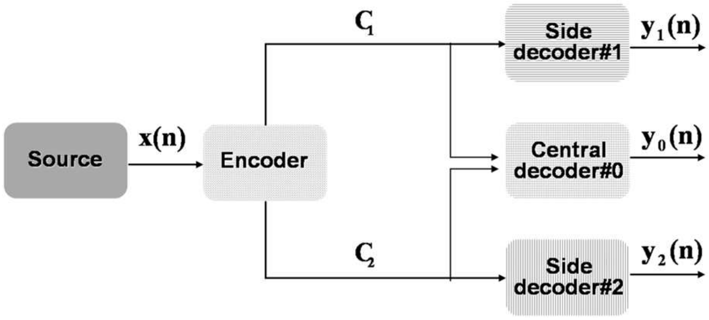


Figure 1: The fundamental structure of two-channel Multiple Description Coding (MDC)

data. On the other hand, most of the latter methods [9, 10] cannot well preserve high sound quality well because they do not use any additional information on the signal in the lost packet. Thus, all kinds of PLC methods always have some trade-off problem between the transmitted information capacity and the packet loss concealment quality. To settle this trade-off problem, the Multiple Description Coding (MDC) scheme [11] has been attracting attention. This is because the PLC method can flexibly balance the data quality with bitrate without any retransmission processing in the network path. In this case, MDC may be more useful for increasing the quality of multimedia data transmitted over a lossy packet network than layered coding when the requirement of real-time playing is severe [12].

In the research area of MDC, the target of most research studies has been images and video streaming [13, 14], and several research studies have combined the multiple description coding with data hiding such as watermarking [15, 16]. There have also been many studies on audio codecs, such as linear Pulse Code Modulation (PCM) and Differential PCM (DPCM) [17], Logarithmic PCM [18], Adaptive Differential PCM (ADPCM) [18,19], and Code Excited Linear Prediction (CELP) [20]. However, these are basically studies of MDC for narrowband sound signals such as speech signals. On the other hand, demand for streaming services for higher quality sound (i.e., wideband sound signal) is increasing recently. Nevertheless, there have been few studies of MDC for wideband sound signal codec such as MP3 [21]. Ito *et al.* have proposed an MDC approach for an MP3 codec [22]. Their method is based on a frequency-domain element division, where two descriptions are generated by dividing the frequency spectrum of the signal into one with odd-number components and the other with even-numbered components. As will be described in detail in Section 2, this method can only be applied to codec such as MP3 and Advanced Audio Coding (AAC). Moreover, using Ito's method, the bitrate increases because of the auxiliary information. On the other hand, an MDC approach based on a time-domain division has also been proposed for wideband sound signals [23]. In this method, two descriptions in the time domain are calculated by constructing one description with odd-number samples and the other with even-numbered samples. However, the sound quality of the decoded signal by this MDC method deteriorates even when no packet loss occurs because of distortions caused by the compression process of the employed codecs.

In this paper, we propose an MDC method for wideband audio signals, applicable to a streaming service that is robust against packet losses. The basic framework of our method is based on the time-domain MDC architecture for wideband sound signals [23] since this MDC method is potentially advantageous for high-quality packet loss recovery. As mentioned above, however, the original time-domain MDC method has a sound quality deterioration problem. Therefore, we herein first examine the reason for deterioration of the decoded signal introduced by the time-domain MDC, and then we propose a method for improvement of sound quality by time-domain MDC by applying a post-filter that compensates for high-frequency distortions.

This paper is organized as follows. Section 2 summarizes the concept of MDC and the MDC design problem. Section 3 proposes an improvement of the MDC by time-domain element division. The experimental results are then given in Section 4. Finally, Section 5 concludes this paper.

2. MDC for a wideband sound signal codec.

2.1. **The Concept of MDC.** MDC is a coding technique which enables us to divide a source into multiple data streams. Each of these data streams, called a description, can be decoded separately. When all of these data streams are received, these streams are decoded unitedly to recover the signal with the highest quality. Figure 1 shows the fundamental structure of signal transmission by multiple description coding using two channels. The MDC procedure is as follows [11]:

(1). The original signal $x(n)$ is divided into multiple sub-data by a multiple description (MD) encoder as illustrated in Figure 1, each of which (either C_1 or C_2) is called a description

(2). If one of the descriptions is lost, the decoded result of the single description is generated from the remaining description using a side decoder. The MD side decoder needs to be designed so that the decoded signal by this side decoder can be perceived to be similar to the original signal, though there will be a certain degree of degradation due to the lost descriptions.

(3). If all of the descriptions are received, the output signal is decoded by means of the central decoder using both descriptions. Using the central decoder, the decoded signal is perceived to be identical or very similar to the original signal.

By applying the MDC scheme, robust audio transmission can be realized. The realization methods of MDC can be divided into two categories: one is a code data division method that is applied within the audio and the other is a source data division method that is applied before using the encoder.

2.2. **Conventional MDC for wideband sound signal codecs.** Arean *et al.* proposed an MDC method for a wideband sound signal codec using a correlating transform [24]. This method, however, is insufficient because the improvement of sound quality by their method is observed only when the bitrate of an audio signal is very low, at which point the absolute quality of the signal is insufficient.

Ito *et al.* proposed an MDC method for an MP3 codec based on frequency-domain element division, where the Modified Discrete Cosine Transform (MDCT) coefficients are split into even- and odd-numbered spectral components [22]. This method improves the quality of the audio signal degraded by packet loss, but the drawback is that this method requires side information (7.6 kbit/s) in addition to the encoded signal. On the other hand, another MDC approach has been proposed based on time-domain element division in which samples of the source signal are split into even- and odd-numbered subsamples and those subsamples are independently encoded with a wideband sound signal codec such as MP3 [23]. This method has another problem, i.e., the sound quality of the decoded signal by the MD central decoder deteriorates even when no packet loss occurs.

2.3. **MDC based on time domain sample splitting.** To realize an MDC method for high-quality wideband sound signal transmission, we propose an MDC method based on the MDC structure proposed by Ito *et al.* [23]. This basic MDC structure has the following two stages: time domain sample splitting [17] and data compression using an

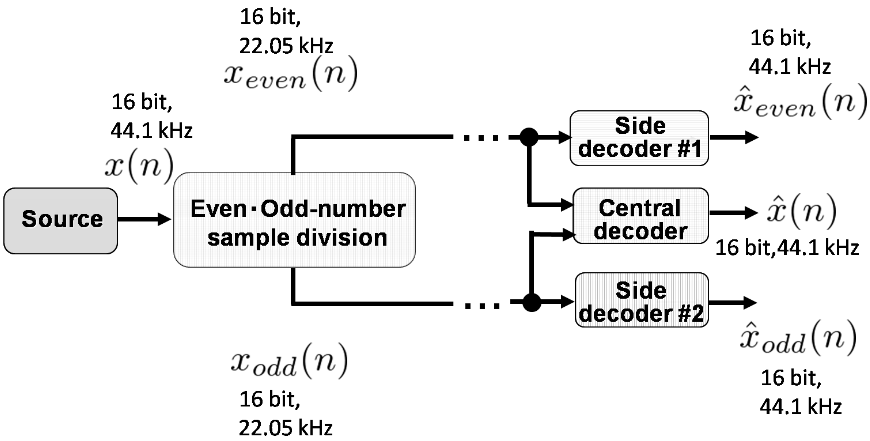


Figure 2: The time domain samples splitting MDC structure proposed by Jayant *et al.* [17].

MP3 codec [23]. The basic concept of time domain sample splitting [17] is to divide samples of the input signal into even-numbered and odd-numbered subsamples. Because two contiguous samples in a sound signal are highly correlated, one set of subsamples can be estimated from the other subsamples. If the two sets of subsamples are received correctly, the original signal can be perfectly recovered by interlocking the two subsamples.

We express $x(n)$ as a sample of the source signal in time domain. Here, $x(n)$ consists of a summation between even-numbered subsamples $x_{even}(n)$, expressed as

$$x_{even}(n) = x(2n), \qquad \text{n=0,1,2,}\cdots,\frac{N}{2}-1 \tag{1}$$

and odd-numbered subsamples $x_{odd}(n)$, expressed as

$$x_{odd}(n) = x(2n+1), \qquad \text{n=0,1,2,}\cdots,\frac{N}{2}-1 \tag{2}$$

Let W_N be

$$W_N \triangleq e^{-j\frac{2\pi}{N}}. \tag{3}$$

Then, $X(k)$, the discrete Fourier transform (DFT) result of $x(n)$, is expressed as

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}kn}, \tag{4}$$

$$= \sum_{n=0}^{N-1} x(n)W_N^{kn}, \qquad (0 \leq k < N). \tag{5}$$

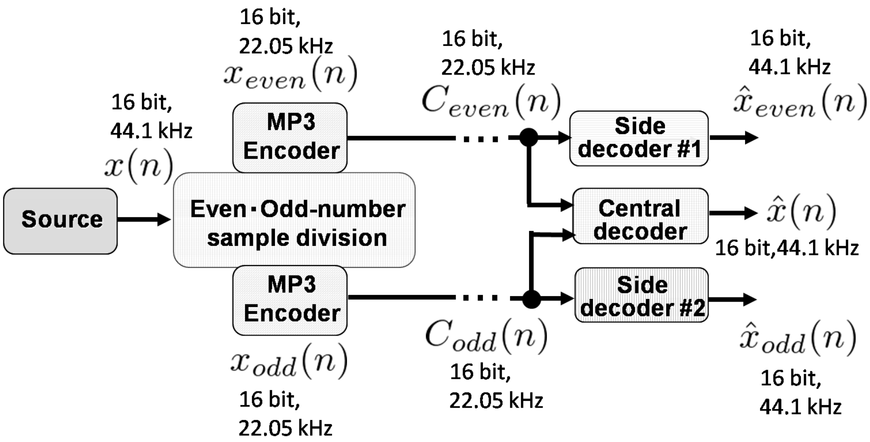


Figure 3: The basic MDC structure for wideband sound signal proposed by Ito *et al.* [23].

Here, $X(k)$ can be expressed in terms of the spectrum summation between $x_{even}(n)$ and $x_{odd}(n)$ as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n)W_N^{nk}, \tag{6}$$

$$= \sum_{m=0}^{\frac{N}{2}-1} x(2m)W_N^{2mk} + \sum_{m=0}^{\frac{N}{2}-1} x(2m+1)W_N^{(2m+1)k}, \tag{7}$$

$$= \sum_{m=0}^{\frac{N}{2}-1} x_{even}(m)W_{\frac{N}{2}}^{mk} + W_N^k \sum_{m=0}^{\frac{N}{2}-1} x_{odd}(m)W_{\frac{N}{2}}^{mk}, \tag{8}$$

$$= X_{even}(k) + W_N^k X_{odd}(k), \qquad (0 \leq k < \frac{N}{2}). \tag{9}$$

2.4. **The basic MDC structure for wideband sound signal codec.** Figure 3 shows the basic MDC framework for a wideband sound signal based on time-domain sample splitting. The procedure of this MDC is as follows:

Multiple Description (MD) Encoding: First, the source signal samples $x(n)$ are split into odd-numbered samples $x_{odd}(n)$ and even numbered samples $x_{even}(n)$ in the time domain. After that, an MP3 encoder compresses each set of subsamples into a low bit rate description ($C_{even}(n)$,$C_{odd}(n)$). These descriptions are sent using individual packets.

MD Decoding: If no packet loss occurs (both descriptions $C_{even}(n)$ and $C_{odd}(n)$ are received), these received packets are decompressed using an MP3 decoder. Then the decompressed samples are rearranged into their original sequence to decode the source signal. If packet loss occurs and one of $C_{even}(n)$ or $C_{odd}(n)$ is lost, the other description ($C_{odd}(n)$ or $C_{even}(n)$) is decompressed by the decoder. Samples of the decoded signal from the single description are then up-sampled and substituted instead of the missing packet intervals. This processing is justified by the fact that the correlation between the source signal and the up-sampled subsamples is very high [17]. If packet losses occur at both transmission paths simultaneously, zero value samples are substituted instead of the lost packet.

Table 1: ODG score description: Five-grade impairment scale.

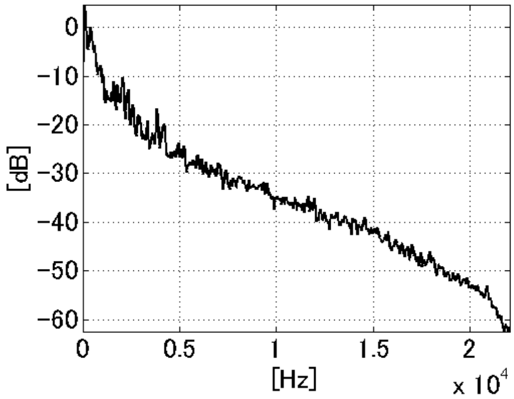
ODG score	Impairment
0.00	Imperceptible
–1.00	Perceptible but not annoying
–2.00	Slightly annoying
–3.00	Annoying
–4.00	Very annoying

2.5. **Observation of distortion introduced by the basic MDC for wideband sound signal codec.** Using this basic MDC structure with an MP3 encoder, the decoded signal by each single description can remain at a certain level of sound quality, equal to the one of a sound signal that has half of the sampling rate compared with the source signal. However, the sound quality of the decoded signal by both descriptions deteriorates even when both descriptions are received. To confirm the sound quality deterioration we measured the quality of the sound signal decoded from both descriptions. The bitrate for one description was 80 kbit/s. Then we compared the quality of the decoded signal with that of a signal from a single description. We selected 30 musical pieces from the Real World Computer Partnership (RWCP) music database [25] as test signals. These signals were selected from three different genres (Classical, jazz and pop music), there being 10 pieces in one genre. The length of a signal was about 20 seconds. We estimated each of the Objective Differential Grade (ODG) scores as a measurement of sound quality.

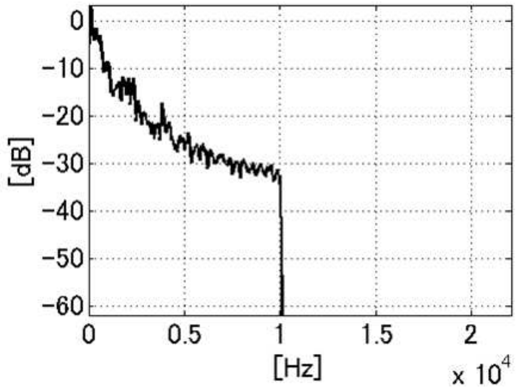
The ODG score is calculated by using Perceptual Evaluation of Audio Quality (PEAQ) [26]. The PEAQ attempts to simulate the perceptual properties of the human ear to estimate subjective quality of a test audio signal. Table 1 shows a description of the ODG score. For the calculation of each ODG score, we used the linear PCM signals not encoded by MP3 as source signals against other evaluation signals. As an experimental result, the average ODG score for the decoded signal from both descriptions was -2.89, while the average ODG for the decoded signal using the single description was -2.61. As these experimental results indicate, using both descriptions did not improve the quality of sound. To investigate reasons for the deterioration of the decoded sound, we compared the power spectrums of the source signal, the decoded signal using both descriptions and that using only one description. Figure 4 shows an example of the power spectra. From the power spectrum of the signal decoded from a single description (Fig. 4(b)), it was confirmed that this signal has only the lower half of the spectral components compared with the one of the source signal (Fig. 4(a)). In contrast, the spectrum decoded from both descriptions (Fig. 4(c)) is distorted with respect to the original signal in two aspects. One is a lack of spectral components observed around 10 to 12 kHz. The other is large spectral distortion observed at a frequency over 12 kHz.

Here, we discuss the causes of the distortions observed in the spectrum of Fig. 4(c). First, lack of spectral components around 10 to 12 kHz is caused by the MP3 codec. When encoding each of the descriptions at the half sampling rate (22.05 kHz in this example), the encoder cuts off the higher spectral components near the Nyquist frequency (over 10 kHz in this case) to reduce bitrate. Therefore, by interlocking the output signal from both descriptions, the cut-off parts of the spectral components appear at the very center of the entire frequency range of the decoded signal.

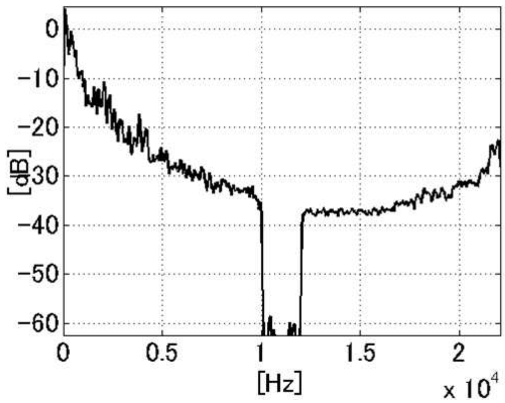
Next, the spectral distortions over 12 kHz are caused by quantization noise introduced by the MP3 encoder and its aliasing noise. Let $c(n)$ be the decoded signal by combining



(a) Source signal



(b) Decoded signal using the MD side decoder in the basic MDC



(c) Decoded signal using the MD central decoder in the basic MDC

Figure 4: Power spectrum comparison

the decoder outputs from both descriptions, $c_{even}(n)$ and $c_{odd}(n)$. Here, $c(n)$ is written as follows:

$$c(2n) = c_{even}(n), \tag{10}$$
$$c(2n+1) = c_{odd}(n), \tag{11}$$

According to the discussion in section 2.3, the spectrum of the decoded signal can be written as follows:

$$C(k) = C_{even}(k) + W_N^k C_{odd}(k), \qquad (0 \leq k < \frac{N}{2}). \tag{12}$$

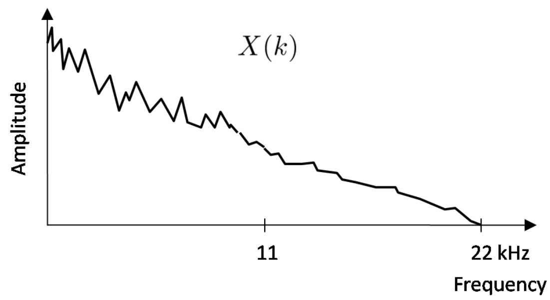
Here, as $c_{even}(n)$ is a signal generated by encoding $x_{even}(n)$ and decoding the code, the spectrum $C_{even}(k)$ can be written as $X_{even}(k)$ with quantization noise and filtering.

$$C_{even}(k) = L(k)X_{even}(k) + E_{even}(k), \tag{13}$$

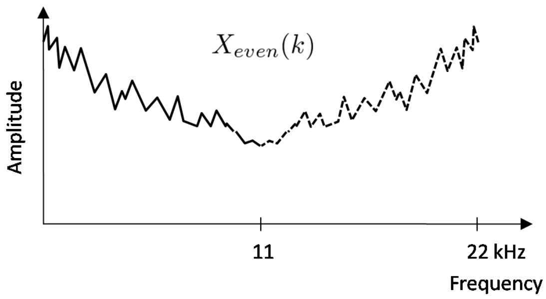
where $L(k)$ is a low pass filter introduced by the MP3 encoder and $E_{even}(k)$ is the quantization noise. Therefore, when $0 \leq k < \frac{N}{2}$, $C(k)$ can be written as follows.

$$C(k) = L(k)\{X_{even}(k) + W_N^k X_{even}(k)\} + E_{even}(k) + W_N^k E_{odd}(k), \tag{14}$$
$$= L(k)X(k) + E_{even}(k) + W_N^k E_{odd}(k), \qquad (0 \leq k < \frac{N}{2}). \tag{15}$$

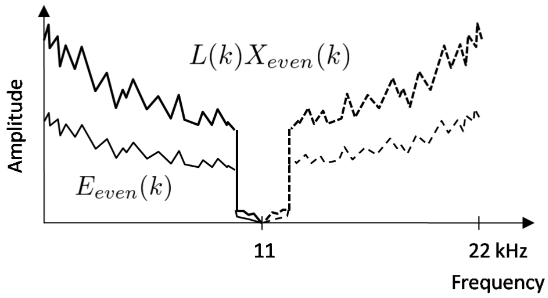
The spectral distortion introduced by the encoder is depicted in Figure 5. By calculating $x_{even}(n)$ and $x_{odd}(n)$, the spectrum is folded at $k = N/4$ (Fig. 5(b)). The low



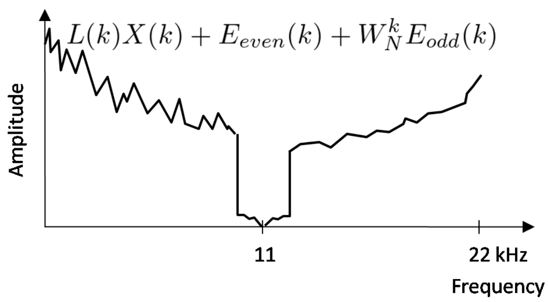
(a) Original signal



(b) Decimated signal for the single description



(c) Distorted signal by LPF $L(k)$ and the quantization noise $E_{even}(k)$



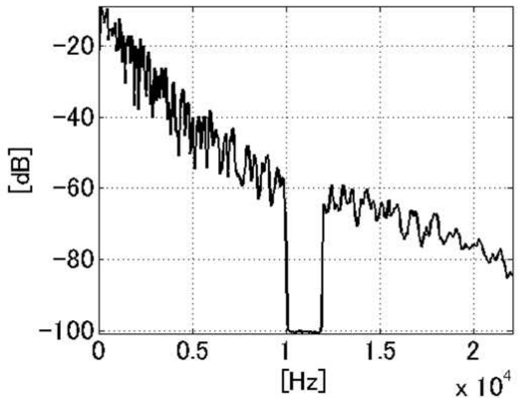
(c) Resotred signal

Figure 5: Introduction of spectral distortion by the MP3 encoder

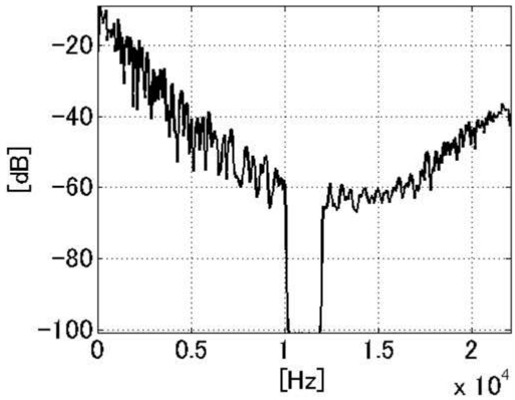
pass filter $L(k)$ cuts off the high frequency components of $X_{even}(k)$ and $X_{odd}(k)$. This filter is regarded as a band-stop filter when considering the frequency range up to $N/2$ (Fig. 5(c)). In addition, the quantization errors $E_{even}(k)$ and $E_{odd}(k)$ are added to the signal. As the MP3 codec employs frequency-by-frequency quantization based on the psychoacoustic properties, the magnitude of the quantization noise in a frequency bin is roughly in proportion to that of the input signal (Fig. 5(c)). When $c_{even}(n)$ and $c_{odd}(n)$ are combined, the higher parts of $X_{even}(k)$ and $X_{odd}(k)$ cancel each other; however, as $E_{even}(k)$ and $E_{odd}(k)$ are not cancelled, they remain in the higher frequency of the restored signal (Fig. 5(d)).

To evaluate the effect of sound quality deterioration by the spectral notch and the spectral distortion at the higher frequency individually, we compared the sound quality of the signals processed by two kinds of quality degradation processes. One was a band stop filter with cut-off frequencies from 10 to 12 kHz to simulate the filter of the MP3 codec. The other was the basic MDC under the bitrate condition of 80 kbit/s, which involved both a spectral notch and high-frequency distortion. We used the same musical pieces as in the previous experiment and calculated ODG scores. Figure 6 shows an example of the power spectrum calculated by the above-mentioned two methods. The average ODG by the band stop filter was -0.65 and that by the basic MDC was -2.89 (the same result as in the previous experiment). This result suggests that sound quality degradation introduced by the basic MDC is mainly caused by high-frequency distortion. Therefore, in the remainder of this paper, we focus on how to restore this spectral distortion and we do not treat the problem of the spectral notch in the rest of this paper.

3. **Improvement of sound quality using the Wiener filter.** On the basis of the previous result, it was proved that the decoded signals by the basic MDC required some



(a) Filtered source signal by band stop filter



(b) Decoded signal using the MD central decoder in the basic MDC for wideband audio codec

Figure 6: Power spectrum comparison

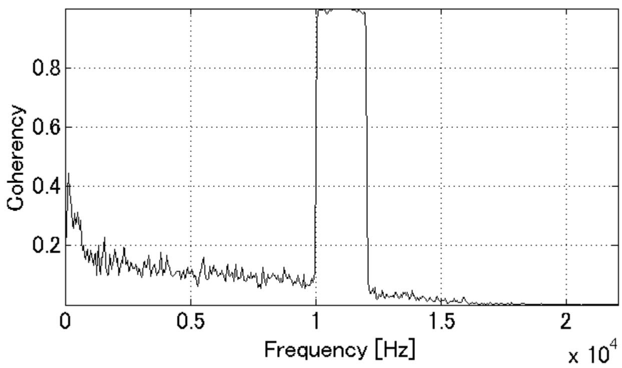


Figure 7: Example of the average spectral coherence functions between the source signal and the difference error by the basic MDC

restoration. Thus, we applied a noise reduction technique to the decoded signals for reducing the spectral distortion.

Most noise reduction techniques are designed based on the assumption that the noise is additive and has no correlation to the original signal. To confirm if this assumption is valid, we measured the spectral coherence function derived from the source signal and the difference error between source one and the decoded signal by basic MDC. Then we calculated their average coherence function (The source signals were selected from RWCP music database, i.e., three musical pieces from different genres such as classical, jazz and pop music). The spectral coherence function C_{XY} of two spectra X and Y is given by

$$C_{XY}(k) = \frac{|E_t[X(k,t)Y^*(k,t)]|^2}{E_t[|X(k,t)|^2]E_t[|Y(k,t)|^2]}, \tag{16}$$

where $X(k,t)$ and $Y(k,t)$ are spectra of two signals at frame t and frequency k, $Y^*(k,t)$ is the complex conjugate of $Y(k,t)$, and $E[\cdot]$ is temporal average. If two signals X and Y are uncorrelated, $C_{XY}(k)$ become nearly zero for all k; if any spectral components of the two signals are correlated, the coherence function of that frequency becomes larger.

The average of the coherence functions between the original signal and the residue of the decoded signal is shown in Figure 7. The coherence value in the frequency range

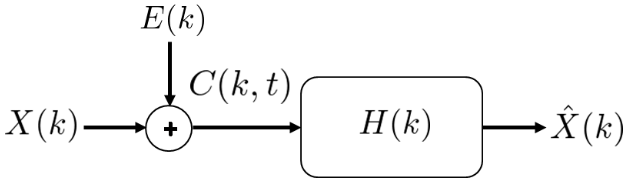


Figure 8: Basic block diagram of the Wiener filter

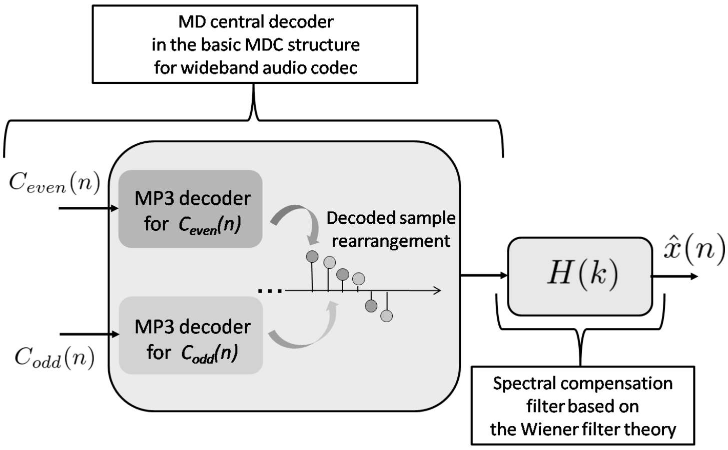


Figure 9: New central MD Decoder structure in the proposed MDC method

from 12 to 15 kHz is less than 0.1 and becomes negligible above 15 kHz. The spectral distortion above 12 kHz can, thus, be regarded as additive noise. Therefore, the use of a Wiener filter to reduce spectral distortion above 12 kHz is justified.

The Wiener filter is a basic type of optimal filter which works to reduce uncorrelated noise against the source signal [27]. A spectral subtraction filter is also a popular type of optimal filter using the estimated error characteristics [27, 28]. In general, spectral subtraction works better than the Wiener filter as long as the spectrum of noise is estimated appropriately [29]. However, since the spectral subtraction is a non-linear filter, it sometimes causes undesirable distortion such as musical noise, especially when the noise estimation is not appropriate. Consequently, we considered introducing spectral compensation filters designed on the basis of the Wiener filter theory into our proposed MDC method.

Figure 8 shows the basic block diagram of the Wiener filter. The decoded signal $C(k,t)$ consists of the source signal $X(k,t)$ disrupted by additive noise $E(k)$ and corresponds to the difference error between the source signal and the decoded signal by the basic MDC. The Wiener filter is labeled $H(k)$, and $\hat{X}(k,t)$ is the decoded signal from which the spectral distortion is reduced.

The Wiener filter $H(k)$ is given by

$$H(k) = \frac{E_t\Big[\big\{X(k,t)+E(k,t)\big\}X^*(k,t)]\Big]}{E_t\Big[X(k,t)X^*(k,t)\Big]+E_t\Big[E(k,t)E^*(k,t)\Big]}. \tag{17}$$

To apply the Wiener filter to only the higher frequency part of the signal, a high-pass filtered variation of $E(k,t)$ is used as the residual error spectrum in our proposed method. For this, the employed high-pass filter $H_{12k}(k)$ has a 12 kHz cutoff frequency and is given by

$$E(k,t) = H_{12k}(k)(C(k,t) - X(k,t)) \tag{18}$$

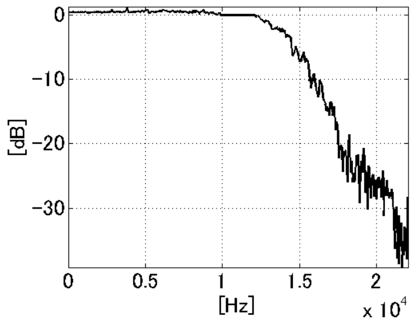
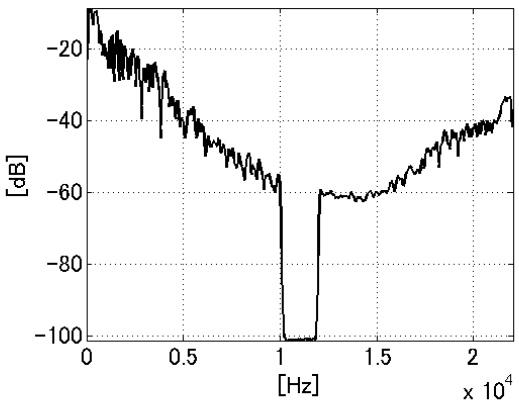
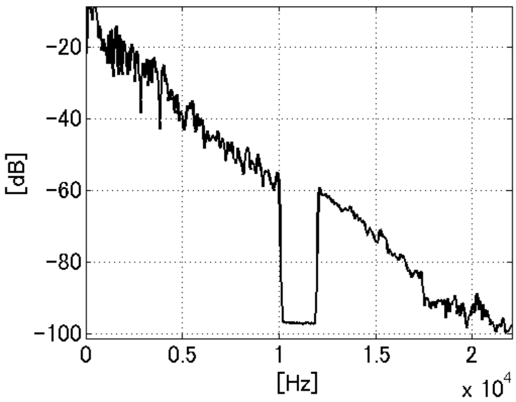


Figure 10: Example of the spectral compensation filter characteristics with full bitrate condition: 160 kbps



(a) Signal decoded by basic MDC before the spectral compensation



(b) Decoded signal after spectral compensation using the Wiener filter exhibited Fig.10

Figure 11: An example of the spectral compensation result

Figure 9 shows the inside of the central MD decoder in the proposed MDC structure. This central MD decoder in the proposed MDC works in the same way as the basic MDC with the spectral compensation filter designed by Eq. (17). An example of the filter characteristics is shown in Fig.10. Figure 11 shows an example of the spectral compensation result using the Wiener filter exhibited in Fig. 10.

4. Experiments.

4.1. **Experimental setup.** To evaluate the proposed method, we carried out several experiments simulating sound data transmission with two independent transmission paths and decoding. We set a packet size of 1152 samples per channel. We assumed only random packet losses, considering that burst packet losses can be converted into random packet losses if appropriate packet interleaving is applied.

The packet loss rates were set to 0 (No packet losses), 1, 2, 5 and 10%, respectively. The audio signals used in the experiment were same as those in the previous experiments. Table 2 and 3 show the experimental conditions. We examined two conventional methods (Single MP3 and Double MP3) and two MDC methods (The basic MDC and Proposed MDC).

Table 2: Experimental conditions: Categories of the source signals

Source signal	30 musical pieces (about 20 s) from Classical, Jazz, Popular genres, taken from the RWC music database
Sampling condition	16 bit, Stereo 44.1 kHz
Packet size	1152 samples per channel

Table 3: Experimental conditions: Categories of packet loss condition

Condition	Average packet loss rate [%]	Total bitrate [kbps]
No packet loss	0	128, 160 192, 224 256, 320
Random packet loss	1	
	2	
	5	
	10	

Single MP3: The single MP3 method means a condition of a simple one-channel transmission of a single data packet encoded with an MP3 encoder at 128, 160, 192, 224, 256, and 320 kbit/s. When using this method, if packet losses occur, the missing packet intervals are filled with zero values.

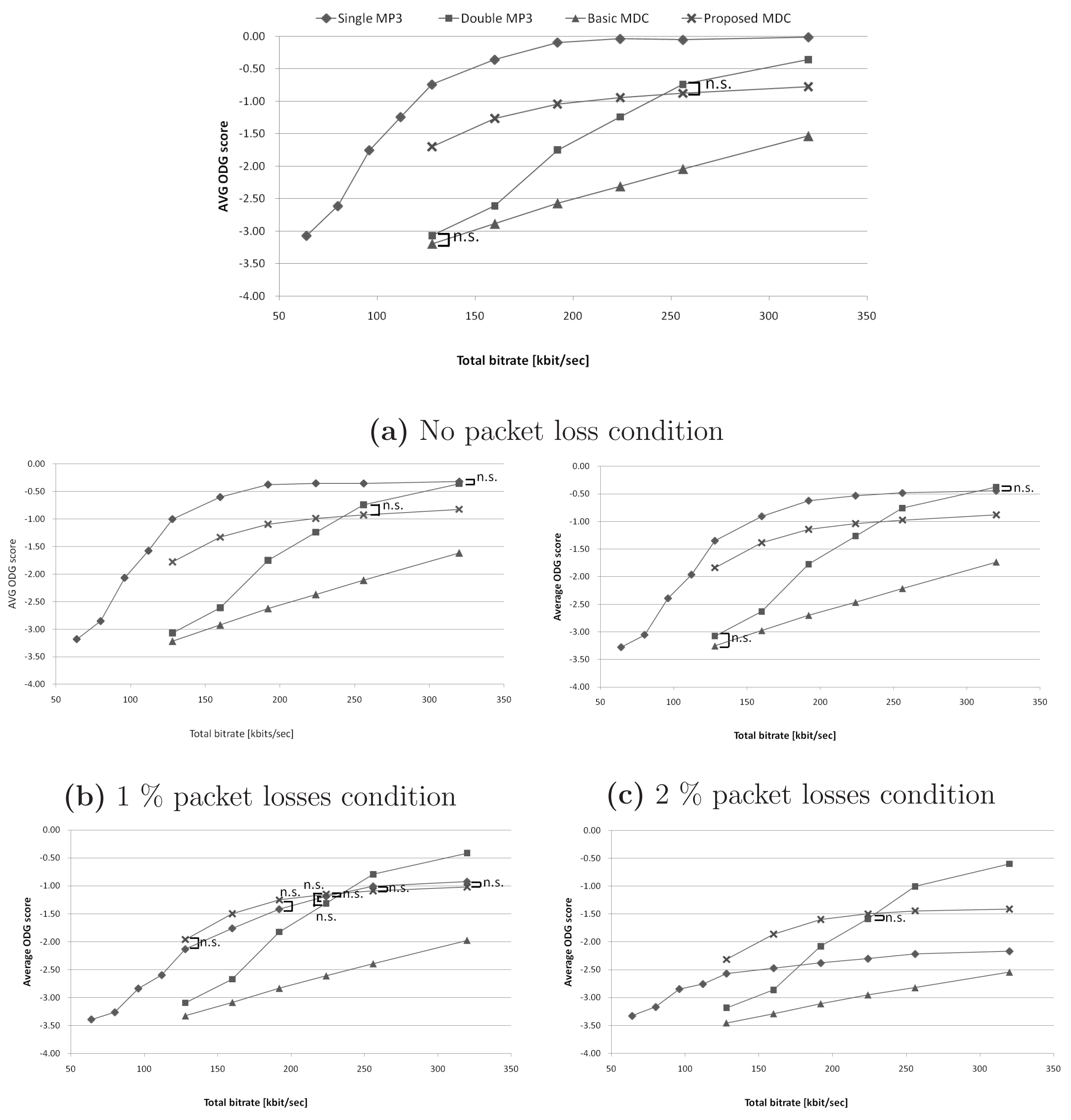
Double MP3: The Double MP3 method is a method with simple redundancy involving the transmission of two identical MP3-encoded data packets, each of which is encoded in half of the total bit rate. When packet losses occur in one channel, this method allows the recovery of the missing interval by using the packet data from the other transmission path.

Basic MDC: With this method, two different MP3-encoded half bitrate data packets using time domain sample splitting were transmitted.

Proposed MDC: The proposed MDC method is a modification of the basic MDC method with the Wiener filter as described in the previous section. When using either the basic or proposed MDC method, if packets from one channel are lost, the missing packet intervals are substituted from the decoded description delivered via the other transmission path. If packet losses occur at both transmission paths simultaneously, zero value samples are substituted instead of the missing packet intervals (This is common to the methods outlined above).

The ODG scores are calculated by Perceptual Evaluation of Audio Quality (PEAQ). Moreover, we also estimated ODG scores of the signals encoded by the Single MP3 method at 64, 80, 96, and 112 kbps, for comparison.

4.2. **Results and discussion.** Figure 12 shows the estimated ODG score as a function of the total bitrate for the four types of encoding methods. As mentioned in the previous section, they comprise three different types of PLC methods, including the proposed one, and the single MP3 encoding method. The ordinate shows ODG scores and higher ODG scores mean better sound quality. Here, the six kinds of total bit rate were applied to the three kinds of PLC methods; they range from 128 kbps (64 kbps×2) to 320 kbps (160



(a) No packet loss condition

(b) 1 % packet losses condition

(c) 2 % packet losses condition

(d) 5 % packet losses condition

(e) 10 % packet losses condition

Figure 12: Comparison of average ODG scores for the different target methods under several percent random packet loss rate condition.
(The square brackets "]" indicate that the combinations of two encoding methods for which the difference is statistically insignificant.)

kbps×2). Panel (a) of this figure shows the ODG scores of the no-packet-loss condition. Panels (b) to (e) correspond to packet loss rates of 1%, 2%, 5% and 10%, respectively.

To examine whether the differences of the ODG scores for the four types of encoding methods are statistically significant, a three-way analysis of variance with repeated-measure design was conducted by using ANOVA. Here, packet-loss rate (5 levels), bitrates (6 levels) and encoding methods (4 levels) were used as the factors, and 30 music pieces were treated as a repeated measure.

As a result, all the main effects were found to be statistically significant (packet-loss rate: $F(4, 36) = 697.42$, $p < .01$; bitrate: $F(5, 45) = 1214.61$, $p < .01$; encoding method:

$F(3,27) = 175.73$, $p < .01$). Moreover, the three-way interaction (packet-loss rate × bitrate × encoding method: $F(60,540) = 40.44$, $p < .01$) was significant and so were all the two-way interactions (packet-loss rate × bitrate; $F(20,180) = 42.923$, $p < .01$; packet-loss rate × bitrate × encoding method: $F(12,108) = 467.45$, $p < .01$; bitrate × encoding method: $F(15,135) = 164.64$, $p < .01$). Therefore, simple main effects as well as simple-simple main effects relating to the encoding method were examined and all were shown to be significant beyond $p = .01$. Then a post-hoc multiple comparison test (Ryan's method) was conducted with the significance rate of $p = 0.05$. The result showed that most of the differences between the two encoding methods were significant but that a few were not. The combinations of encoding methods for which the difference is statistically insignificant are indicated by a square bracket in Fig. 12. Insignificant combinations relating specifically to the proposed MDC are found for the following conditions: For no-packet-loss condition, double MP3 at 256 kbps; for 1% packet loss condition, double MP3 at 256 kbps; for 5% packet loss condition, single MP3 at 128, 192, 224, 256 and 320 kbps and double MP3 at 224 kbps; for 10% packet loss conditions, double MP3 at 224 kbps.

The ODG score results shown in panels of Fig. 12 (a) to (e) indicate that the proposed MDC method outperforms the basic MDC method from 1.6 points to 0.5 points. This improvement was statistically significant for all the conditions. Compared with the double MP3 method, the proposed MDC method exhibits higher ODG scores when the total bitrate is from 128 to 224 kbps. Nevertheless, the double MP3 method exhibits significantly higher ODG scores than those by the proposed MDC method for the total bitrates from 256 to 320 kbps when the packet loss is as heavy as 5 and 10%. The ODG score of the single MP3 method greatly depends on the packet loss rate. When the packet loss rate is relatively low (i.e., equal to or less than 2%) the ODG scores of the single MP3 are significantly best among all methods. However, when the packet loss rate is more than 2%, the proposed MDC method and the double MP3 method share the best ODG scores. As a result, the proposed MDC method yields the significantly highest ODG scores at a total bitrate of 160 kbps when the packet loss rate is 5%, and at a total bitrate of from 128 to 192 kbps when the packet loss rate is 10%.

Figure 13 shows the ODG scores by the four methods with respect to the packet loss rate when the total bitrate is 160 kbps. No square bracket is shown in this figure because all the differences among the four encoding methods for a packet loss rate are statistically significant at this bitrate. The result shown in this figure clearly depicts that the single MP3 method is sensitive to the packet loss rate, suggesting that the quality of audio signal under a lossy packet network will be very unstable if the single MP3 method is employed. As a result, the single MP3 method cannot achieve the highest ODG scores anymore when the packet loss rate is 5% or 10%. Instead, under the packet loss rates of 5% and 10%, the proposed MDC method exhibits the significantly highest ODG scores at this total bitrate of 160 kbps.

In summary, the proposed MDC method is robust against the change of packet loss rates, as shown in Fig.13. Additionally, the proposed MDC method significantly outperforms the other PLC methods examined in this study from the point of view of sound quality at practical total bitrates less than or equal to 192 kbps, under a high packet loss rate of 10%. These results indicate that the proposed MDC method can realize practical and stable streaming of wideband sound signals on a transmission path even with drastically changing packet loss rates.

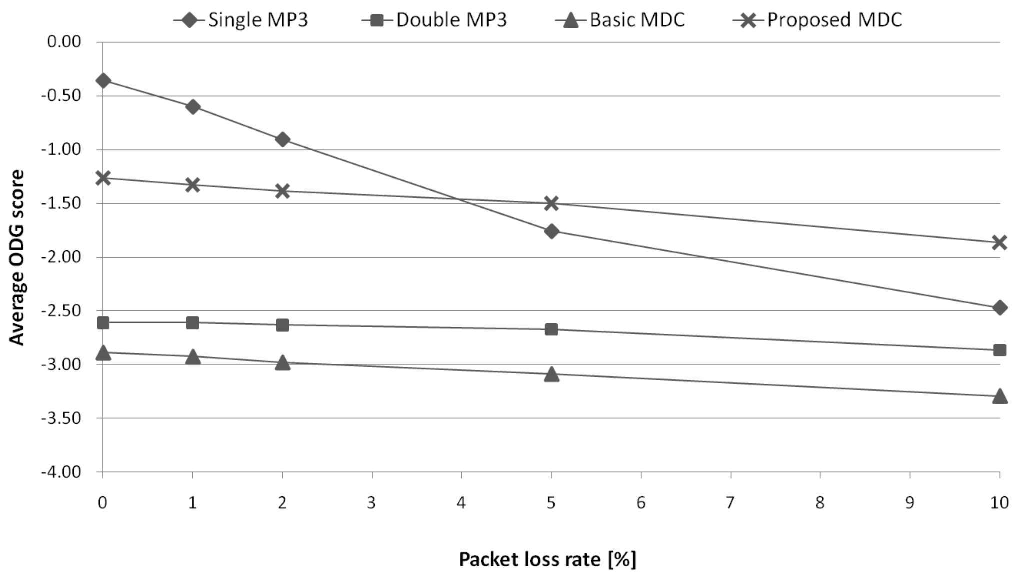


Figure 13: An example of the average ODG scores as a function of the packet loss rate for the four examined methods at a total bitrate condition at 160 kbps. (Differences of the ODG scores among the four encoding methods for a packet loss rate are all statistically significant.)

5. **Summary.** We have herein presented the need for mitigation of the playback quality deterioration of the provided data for real time-wideband digital audio broadcasting application over a lossy packet network. To realize this concept, we have proposed the MDC method in which the sample splitting process in the time domain is combined with high efficiency audio codec such as MP3 and the Wiener filter for frequency compensation. The experimental results showed that the proposed MDC method can improve the range of ODG score from about 0.2 to 1.6 compared with the conventional methods, when the range of bit rates is from 128 to 192 kbps and when the transmitted random packet loss rate is 10%. Further work is required to investigate the performance of our proposed method in comparison with another high efficiency audio codec such as AAC.

Acknowledgments. This study was partially supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) of the Ministry of Internal Affairs and Communications, by the Minisitry of Education, Culture, Sports, Science and Technology of Japan and by the Global Center of Excellence (GCOE) Program of the Ministry of Education, Culture, Sports, Science and Technology of Japan. We wish to thank Yukio Iwaya, Shuichi Sakamoto, and our colleagues of the Network Audio Media Research Group, Shunichiro Abe, Hironori Handa, for their kind advice. We deeply appreciate Maori Kobayashi for her great help with the statistical analysis. We are also grateful to the members of the Shiroishi Rotary Club for their support of one of the authors, Hoseok Wey.

REFERENCES

[1] J. Postel, Transmission control protocol, *RFC793*, 1981.

[2] J. Postel, User datagram protocol, *RFC768*, 1980.

[3] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, RTP:A transport protocol for real-time applications, *RFC1889*, 1996.

[4] C. Perkins, O. Hodson, and V. Hardman, A survey of packet loss recovery techniques for streaming audio, *Proc. of IEEE Network*, vol. 12, no. 5, pp. 40-48, 1998.

[5] B. W. Wah, X. Su, and D. Lin, A Survey of error concealment schemes for real-time audio and video transmissions over the Internet, *Proc. of Int. Symp. on Multimedia Software Engineering*, pp. 17-24, 2000.

[6] J. Rosenberg, and H. Schulzrinne, An RTP payload format for generic forward error correction, *RFC2733*, 1999.

[7] T. Niizeki, K. Kondo, K. Takano, and K. Nakagawa, *On Quality Improvement of Degraded MPEG Audio Due to Packet Loss Using Highly Compressed Overhead Data*, *Proc. of Int. Symp. on Frontiers in Speech and Hearing Research*, Tokyo, Japan, Mar. 2006

[8] Y. Wang, A. Ahmaniemi, D. Isherwood, and W. Huang, Content-based UEP: A new scheme for packetloss recovery in music streaming, *Proc. of ACM Int. Conf. on Multimedia*, pp. 412-421, 2003.

[9] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W. C. Wong, Waveform substitution techniques for recovering missing speech segments in packet voice communications, *IEEE Trans. Acoustics, Speech, and Sig. Processing*, vol. 34, no. 6, pp. 1440-1448, 1986.

[10] O. J. Wasem, D. J. Goodman, C. A. Dvorak, and H. G. Page, The effect of waveform substitution on the quality of PCM packet communication, *IEEE Trans. ASSP*, vol. 36, no. 3, pp. 342-348, 1988.

[11] V. K. Goyal, Multiple description coding: Compression meets the network, *Proc. of IEEE Signal Processing Magazine*, vol. 8, no. 5, pp. 74-93, 2004.

[12] Y. Lee, J. Kim, 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.

[13] D.Wu, Y. T. Hou, W. Zhu, Y. Zhang and J. M. Peha, Streaming video over the Internet: Approaches and directions, IEEE Trans. Circuits and Systems for Video Tech., vol. 11, no. 3, pp. 282-300, Mar. 2001.

[14] A. Vitali, *Multiple Description Coding-A New Technology for Video Streaming over the Internet*, *EBU Technical Review*, Oct. 2007.

[15] N. Conci, M. Carli, and G. Boato, Data hiding tools for multiple description video coding, *Electronic Letters*, vol. 45, no. 2, pp. 98-100, 2009.

[16] J. S. Pan, Y. C. Hsin, H. C. Huang, and K. C. Huang, Robust image watermarking based on multiple decription vector quantization, *Electronics Letters*, vol. 40, no. 22, pp. 1409-1410, 2004.

[17] N. S. Jayant, Subsampling of a DPCM speech channel to provide two self-contained half-rate channels, *Bell Sys. Tech. Jour.*, vol. 60, no. 4, pp. 501-509, 1981.

[18] H.Wey, R. Nishimura, A. Ito, M.Kobayashi and Y. Suzuki, Application of multiple description scalar quantization to speech codec, *Proc. of Japan-China Joint Conference on Acoustics*, 2007.

[19] 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.

[20] B. W. Wah, D. Lin, LSP-based multiple description coding for real-time low bit-rate voice transmissions, *Proc. of IEEE-ICME 2002*, vol. 2, pp. 597–600, 2002.

[21] ISO, Information technology coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s part 3: audio, *ISO/IEC 11172–3*, 1993.

[22] A. Ito, T. Sakai, K. Konno, S. Makino, and M. Suzuki, Packet loss concealment for MDCT-based audio codec using correlation-based side information, *Int. Journal of Innovative Computing, Information & Control*, vol. 6, no. 3, pp. 1347–1362, 2010.

[23] A. Ito, T. Sakai, M. Suzuki and S. Makino, Packet loss concealment of an audio stream by time domain and frequency domain multiple description, *Proc. of Japan-China Joint Conference on Acoustics*, 2007.

[24] R. Arean, J. Kovacevic and V.K. Goyal, Multiple description perceptual audio coding with correlating transforms, *IEEE Trans. Speech Audio Processing*, vol. 8, no. 2, pp. 140–145, Mar. 2000.

[25] M. Goto, Development of the RWC music database, *Proc. of the 18th Int. Conf. On Acoustics*, pp. 553–556, 2004.

[26] ITU, Method for objective measurements of perceived audio quality, *ITU-R. BS. 1387-1*, 11, 2001

[27] Y. Ephraim, H. Lev-Ari, and W. J. J. Roberts (eds.), A brief survey of speech enhancement, the electronic handbook, *CRC Press*, 2003

[28] S. F. Boll, Suppression of acoustic noise in speech using spectral subtraction, *IEEE Trans. Acoustics Speech and Signal Process*, vol. ASSP-27, no. 2, pp. 113–120, 1979.

[29] J. Meyer and K. U. Simmer, Multi-channel speech enhancement in a car environment using Wiener filtering and spectral subtraction, *Proc. of Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 1167–1170, 1997.