Proceedings of 20th International Congress on Acoustics, ICA 2010

23–27 August 2010, Sydney, Australia

Multiple description coding for an MP3 coded sound signal

Ho-Seok Wey(1)(2), **Akinori Ito**(3), **Takuma Okamoto**(1)(3) **and Yôiti Suzuki**(1)(2)

(1) Research Institute of Electrical Communication, Tohoku University, Sendai, Japan
(2) Graduate School of Information Sciences, Tohoku University, Sendai, Japan
(3) Graduate School of Engineering, Tohoku University, Sendai, Japan

PACS: 43.60.Dh, 43.60.Ek, 43.75.Zz

ABSTRACT

For audio communications over a lossy packet network, packet loss concealment techniques are needed. Multiple description coding (MDC) is a useful solution to this problem. This report describes an improvement of the MDC method for wideband audio streaming based on the sample splitting method in the time domain for a sound signal encoded by an MP3 encoder. We investigated the causes of deterioration associated with the conventional method. Using those findings, we propose a method to improve sound quality based on Wiener filtering and the use of a spectrum complement. Experiments were conducted to compare results obtained using the proposed method with those of the conventional method. Improvements of 0.12–1.65 in terms of Perceptual Evaluation of Audio Quality (PEAQ) were obtained over the conventional method.

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 TCP is used for broadcasting multimedia communications, however, unicast communications are necessary between a server and the destinations, resulting in heavy traffic over the network. Therefore, the IP-multicast based distribution using the User Datagram Protocol (UDP) [2] or the Realtime Transport Protocol (RTP) [3] seems suitable for applications of these kinds because these protocols simply send data to many destinations simultaneously without resending the same data. In transmitting packet data simply using UDP or RTP with no Quality of Services (QoS) control, however, packet losses invariably occur. Such cases necessitate methods to mask the degradation caused by the packet losses. This process is called Packet Loss Concealment (PLC). Various PLC methods have been proposed to date[4–7]. 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, Multiple Description Coding (MDC) scheme [8–10] is recently attracting researcher's attention. This is because the PLC method can flexible balance the data quality with bitrate without any retransmission processing in the network path.

In the research area of MDC for audio codecs, demand for the streaming services for higher quality sound (i.e. wideband sound signal) is increasing. Nevertheless, only a few studies have investigated MDC for wideband sound signal codecs such as MP3. Several MDC approaches for MP3 codec based on a frequency-domain element division using modified discrete cosine transform (MDCT) coefficients have been proposed [11, 12]. However, these methods using MDCT coefficients cannot adapt to the playback quality deterioration caused by pre-echoing and other effects because these cannot correspond to the switching of the window function for MP3. Ito *et al.* proposed another MDC approach for wideband sound signal [13]. They proposed preparation of two descriptions in the time domain by constructing a description with odd-numbered samples and another with even-numbered samples. This method is an MDC based on time-domain element division. 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 processes that the codecs use.

As described in this paper, we propose an MDC method for wideband sound signals. The method is applicable to a streaming service that is robust against packet losses. The basic framework of our method is the time-domain process based MDC architecture for wideband sound signals [13] because this MDC method offers the potential advantage of achieving good quality for packet loss recovery. As described above, however, this MDC method has a sound-quality deterioration problem. Therefore, we first investigate the reason for deterioration of the decoded signal introduced by the time-domain MDC. Furthermore, we propose a method for improvement of sound quality for time-domain MDC by application of a Wiener filter that compensates high-frequency distortions combined with spectrum interpolation processing. Experimental results demonstrate the availability of our proposed method.

2. MDC BASED ON TIME-DOMAIN SAMPLE SPLITTING

2.1 Multiple Description Coding (MDC) scheme

MDC is a coding technique enabling description of a source into multiple data streams. Each of these data streams, called descriptions, can be decoded separately. On the other hand, all of these data streams can be decoded together. Actually, MDC takes the following process steps [9]:

(1): The original signal $x(n)$ is divided into multiple subdata by a multiple description (MD) encoder. Each is called a description (C_1, C_2).

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

(3): The output signal is decoded using both descriptions using the central decoder if all the descriptions are received. Using the central decoder, the decoded signal is perceived to be identical or very similar to the original signal.

By applying the MDC scheme to sound information transmission, robust information transmission can be realized. The realization methods of MDC are divisible into two categories: a code data division method that is applied within the codec processing and a source data division method that is applied before using the codec.

2.2 Outline of the 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 introduced by Ito *et al.* [13] because this structure is designed for wideband sound signal transmission application. This basic MDC structure has the following two stages: time domain sample splitting [14] and data compression using the MP3 codec [13]. The basic concept of time domain sample splitting [14] is to divide samples of the input signal into even-numbered subsamples and odd-numbered subsamples. Two contiguous samples in a sound signal have high correlation. Therefore, one set of subsamples can be estimated from the other subsamples. The original signal can be recovered perfectly by interlocking the two subsamples if the two sets of subsamples are received correctly.

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

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

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

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

Let W_N be

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

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

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

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

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

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

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

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

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

Figure 1 portrays the basic MDC framework for a wideband sound signal based on time-domain sample splitting. This MDC takes the following process steps.

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. Subsequently, 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: These received packets are decompressed using an MP3 decoder if no packet loss occurs (both descriptions $C_{even}(n)$ and $C_{odd}(n)$ are received). Then the decompressed samples are rearranged into their original sequence to decode the source signal. The other description ($C_{odd}(n)$ or $C_{even}(n)$) is decompressed by the decoder if packet loss occurs and one of $C_{even}(n)$ or $C_{odd}(n)$ is lost.

Then, samples of the decoded signal from the single description are up-sampled and substituted for 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[14]. Zero value samples are substituted instead of the lost packet if packet losses occur at both transmission paths simultaneously.

2.3 Observation of distortion introduced by the basic MDC

Using this basic MDC structure with an MP3 encoder, the decoded signal by each single description can maintain a certain level of sound quality equal to that of the sound signal that has half the sampling rate of the source signal. However, the sound quality of the decoded signal by both descriptions deteriorates even when both descriptions are received. To investigate reasons for deterioration of the decoded sound, we compared the power spectra of the source signal, the decoded signal using both of descriptions, and that using only one description. Figures 2(a) and 2(b) respectively portray examples of the power spectra of the source signal and the decoded signal using the basic method. The spectrum decoded from both descriptions (Fig. 2(b)) is distorted with respect to the original signal in two respects. One is lack of spectral components observed around 10–12 kHz. The other is large spectral distortion observed over 12 kHz frequencies.

Here, we discuss the causes of the distortions observed in the spectrum of Fig. 2(b). First, the lack of spectral components around 10–12 kHz results from the MP3 codec. When encoding each of the descriptions at the half sampling rate (22.05 kHz in this example), the encoder cut off the higher spectral components near the Nyquist frequency (over 10 kHz in this case) for reducing the 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

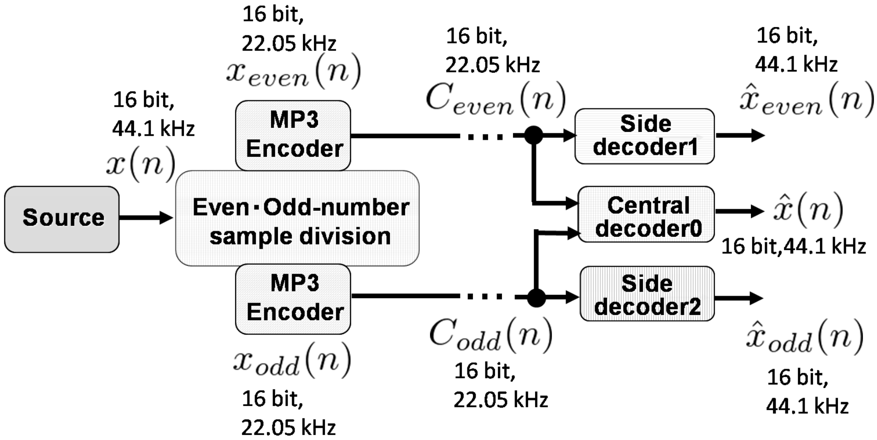


Figure 1: Basic MDC structure for the wideband sound signal proposed by Ito *et. al* [13]

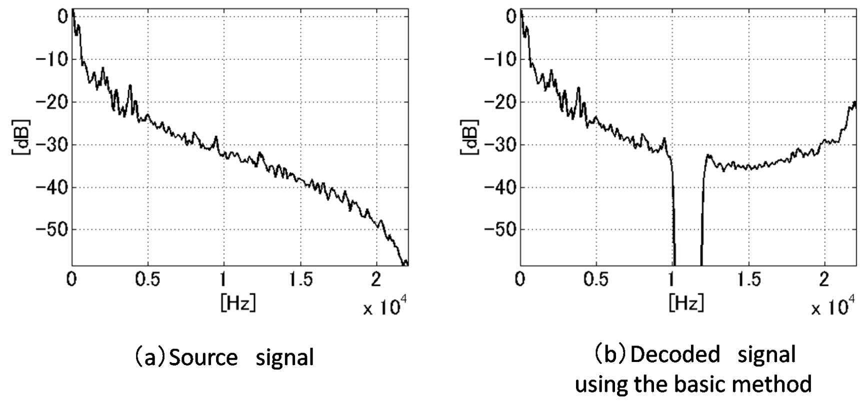


Figure 2: Power spectrum comparison between the source signal and the decoded signal using the basic MDC method

range of the decoded signal.

Next, the spectral distortions over 12 kHz results from quantization noise introduced by the MP3 encoder and its aliasing noise.

Letting $c(n)$ be the decoded signal by combining the decoder outputs from both descriptions, $c_{even}(n)$ and $c_{odd}(n)$, then $c(n)$ is written as

$$c(2n) = c_{even}(n), \qquad (10)$$

$$c(2n+1) = c_{odd}(n). \qquad (11)$$

The spectrum of the decoded signal can be written as follows.

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

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,

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

where $L(k)$ is a filter and $E_{even}(k)$ is the quantization noise. Therefore, when $0 \le k \le \frac{N}{2}$, $C(k)$ can be written as

$$\begin{aligned} C(k) &= L(k)\{X_{even}(k) + W_N^k X_{even}(k)\} \\ &\quad + E_{even}(k) + W_N^k E_{odd}(k), \qquad (14) \\ &= L(k)X(k) + E_{even}(k) + W_N^k E_{odd}(k), \\ &\qquad (0 \le k \le \frac{N}{2}). \qquad (15) \end{aligned}$$

The spectral distortion introduced by the encoder is depicted in Figs. 3(a)– 3(d). By calculating $x_{even}(n)$ and $x_{odd}(n)$, the spectrum is folded at $k = N/4$ (Fig. 3(b)). The low pass filter $L(k)$ cuts off the high frequencies of $X_{even}(k)$ and $X_{odd}(k)$; it is regarded as a band-pass filter when considering the frequency range up to $N/2$ (Fig. 3(c)). In addition, the quantization error $E_{even}(k)$ and $E_{odd}(k)$ are added to the signal. Because the MP3 codec employs frequency-by-frequency quantization based on the psychoacoustic property, magnitude of a frequency bin of the quantization noise is roughly proportional with that of the input signal (Fig. 3(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, because $E_{even}(k)$ and $E_{odd}(k)$ are not cancelled, they remain in the higher frequency of the restored signal (Fig. 3(d)).

3. IMPROVEMENT OF SOUND QUALITY USING THE WIENER FILTER AND SPECTRUM COMPLEMENT

Based on the previous result, it is proved that the decoded signals by the basic MDC require some restoration. Therefore, we apply a noise reduction technique to the decoded signals for reducing the spectral distortion over 12 kHz and a spectrum complement for alleviating the lack of spectral components of 10–12 kHz.

3.1 Spectral compensation using the Wiener filter

Most noise reduction techniques are designed based on the assumption that the noise is additive and has no correlation to the

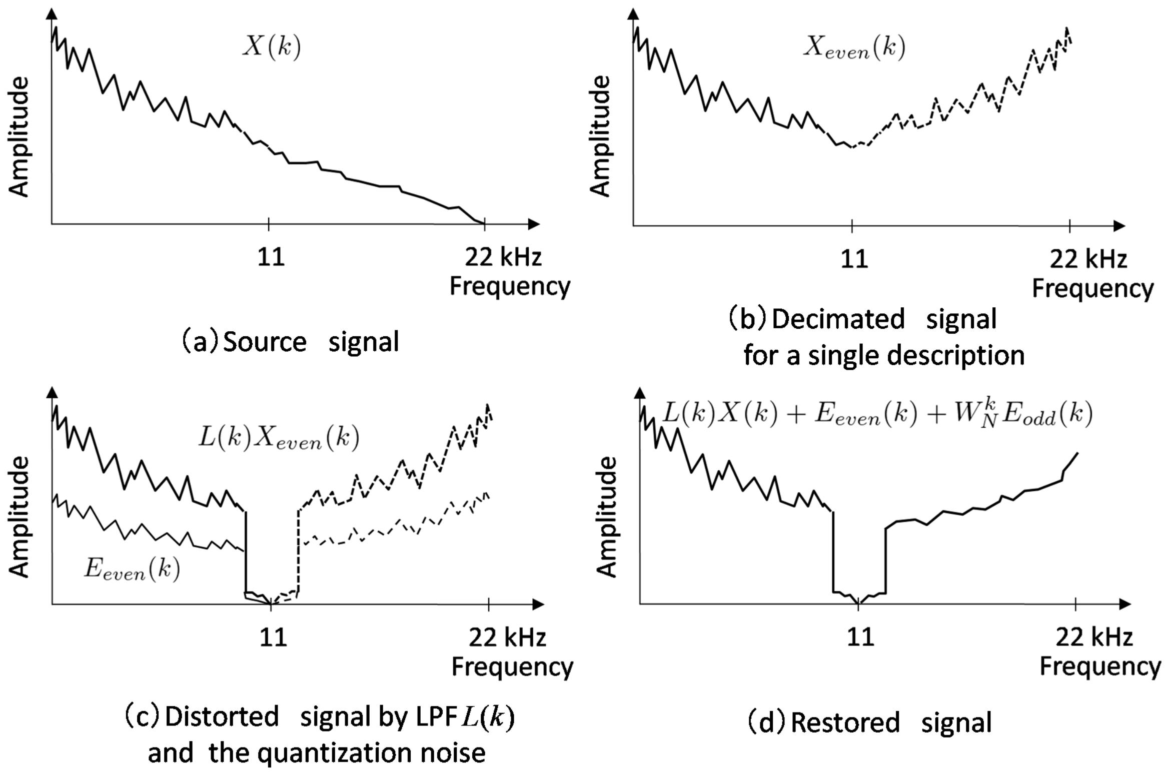


Figure 3: Introduction of spectral distortion by the MP3 encoder.

original signal. To confirm whether this assumption holds, we measured spectral coherence functions derived from the source signal and the difference error between the source one and the decoded signal by the basic MDC. We calculated their average coherence function (The source signals were selected from the Real World Computer Partnership (RWCP) music database[18], three musical pieces from different genres: classical, jazz, and pop music.). The spectral coherence function C_{XY} of two spectra X and Y are given as

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

where $X(k,t)$ and $Y(k,t)$ are spectra of two signals at frame t and where frequency k, $Y^*(k,t)$ is the complex conjugate of $Y(k,t)$, and $E[\cdot]$ is the temporal average. If two signals X and Y are uncorrelated, then $C_{XY}(k)$ becomes nearly zero for all k; if any spectral component of the two signals is correlated, then 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 Fig. 4. This result shows that coherency in the frequencies of 12–15 kHz is much less than 0.1, and that coherency greater than 15 kHz is almost zero. This result confirms that spectral distortion greater than 12 kHz in the residual difference error is qualified assuming application of several noise reduction technique and that this distortion component can be regarded as an additive noise against the source signal. Therefore, we employed the Wiener filter for reducing spectral distortion over 12 kHz.

Wiener filtering is a basic method of optimal filtering. This filter works for reducing uncorrelated noises against the source signal [15]. The spectral subtraction filter is also known as a popular method of optimal filtering using estimated error characteristics [15, 16]. In general, spectral subtraction works

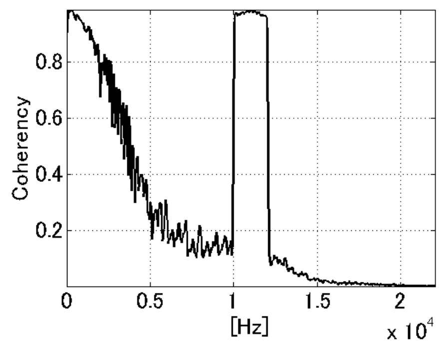


Figure 4: Example of the average spectral coherence functions between the source signal $X(k,t)$ and the difference error $E(k,t)$ using the basic MDC.

better than the Wiener filter as long as the noise spectrum is estimated appropriately [17]. However, because the spectral subtraction is a nonlinear filter, it sometimes causes undesired distortion such as musical noise, especially when the noise estimation is inappropriate. Consequently, into our proposed MDC method, we consider introducing spectral compensation filters designed based on Wiener filter theory.

The decoded signal $C(k,t)$ consists of the source signal $X(k,t)$ disrupted by additive noise $E(k,t)$, corresponding to the difference error between the source signal and the decoded signal by the basic MDC. The Wiener filter is labeled as $H(k)$; $\hat{X}(k,t)$ is the estimated input signal, corresponding to the decoded signal of which the spectral distortion is reduced.

The Wiener filter $H(k)$ is given as shown below.

$$H(k)=\frac{\mathrm{E}_t\left[\hat{X}(k,t)X^*(k,t)\right]}{\mathrm{E}_t\left[X(k,t)X^*(k,t)\right]+E_t\left[E(k,t)E^*(k,t)\right]}, \qquad (17)$$

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

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

Figure 5 shows the inside of the central MD decoder in the proposed MDC structure; this method is called the proposed method 1. This central MD decoder in the proposed MDC works the decoding function of the central MD decoder in the basic MDC in the same way and also works as an spectral compensation filter having the filter characteristics shown in Fig. 5. These spectral compensation filter characteristics are based on the difference error of the basic MDC.

3.2 Spectral complement processing to fill the spectral gap

As described previously, spectral components from around 10–12 kHz are lost in the decoded signals by the basic MDC as well as those using the proposed method 1. Some improvement in the decoded sound quality might be expected if this spectral gap could be appropriately filled. Therefore, to complement this spectral gap from around 10– 12 kHz, spectral components of the same bandwidth of 8–10 kHz that are the lower neighborhood of the gap are duplicated. The spectral complement scheme is shown in Fig. 6. Proposed method 1 combined with the spectral complement is called the proposed method 2. Detailed processes of proposed method 2 are explained below.

(1): Spectral components around 8–10 kHz are extracted using bandpass filtering (6(b)) from the decoded signal shown in Fig. 6(a).
(2): Extracted components are shifted to the band gap based on the Hilbert transform (6(c)).
(3): Band level for the frequency band of 8–10 kHz and that of 12–14 kHz are calculated respectively. Then, the straight line connecting these two levels is calculated (6(d)) and its gradient is extracted.
(4): The gradient of the spectral components of 8– 10 kHz is calculated using a straight least-squares regression line. The gradient of the shifted frequency component is corrected to be identical to the value estimated in step (3) using the differences of the gradients (6(e)).
(5): The power level of the shifted frequency band (10–12 kHz) is adjusted so that the best sound quality, on average, is obtained. The adjustment gain of −5 dB was obtained from the results of preliminary experiments.

Figure 7 presents an example of the spectral compensation result obtained using method 2.

4. EXPERIMENTS

4.1 Experimental setup

To evaluate the proposed method, we conducted several experiments simulating sound data transmission with two virtual 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 well resolved as random packet losses if appropriate packet interleaving is applied.

The packet loss rates are set, respectively, to 0 (No packet losses), 5 and 10 %. The audio signals used as test signals in the experiment were 30 musical pieces from the RWCP music database [18]. These signals were selected from three genres (classical, jazz and pop music) such that 10 pieces were of each genre. The signal length was about 20 s.

Tables 1 and 2 present the experimental conditions.

Table 1: Experimental conditions: Categories of the source signals

Source signal	30 musical pieces from Classical, Jazz and Popular genres taken from the RWC music database [18]
Sampling condition	16 bit, Stereo 44.1 kHz, 20 s
Packet size	1152 samples per channel

Two conventional methods (Single MP3 and Double MP3) and three MDC methods (basic MDC and proposed MDC methods 1 and 2) were examined.

Single MP3: The single MP3 method is 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, then 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 using the packet data from the other transmission path.

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

Proposed MDC (method 1): The proposed MDC method 1 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, then the missing packet intervals are substituted from the decoded description delivered via the other transmission path.

Proposed MDC (method 2): The proposed MDC method 2 is a modification of the proposed method combined with the spectral complement.

Zero value samples are substituted for the missing packet intervals (It is common to these methods outlined above.) if packet losses occur simultaneously at both transmission paths. The Objective Differential Grade (ODG) scores are calculated using Perceptual Evaluation of Audio Quality (PEAQ) [19]. Actually, PEAQ is designed to simulate the perceptual properties of the human ear to estimate the subjective quality of a test audio signal. Table 3 shows the ODG score explanation. Moreover, for comparison, we estimated ODG scores of the signals encoded using the Single MP3 method at 64, 80, 96, and 112 kbps bitrates.

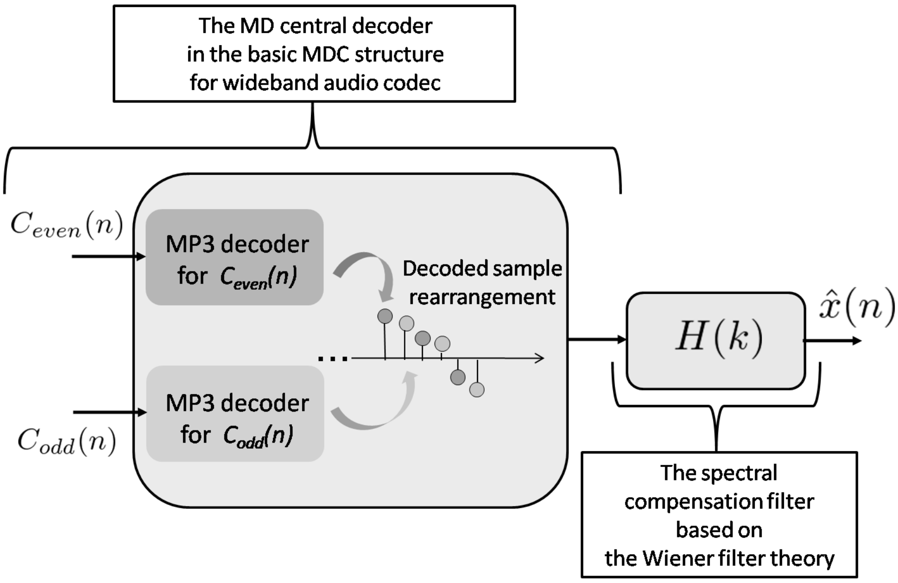


Figure 5: New central MD Decoder structure in the proposed MDC method 1.

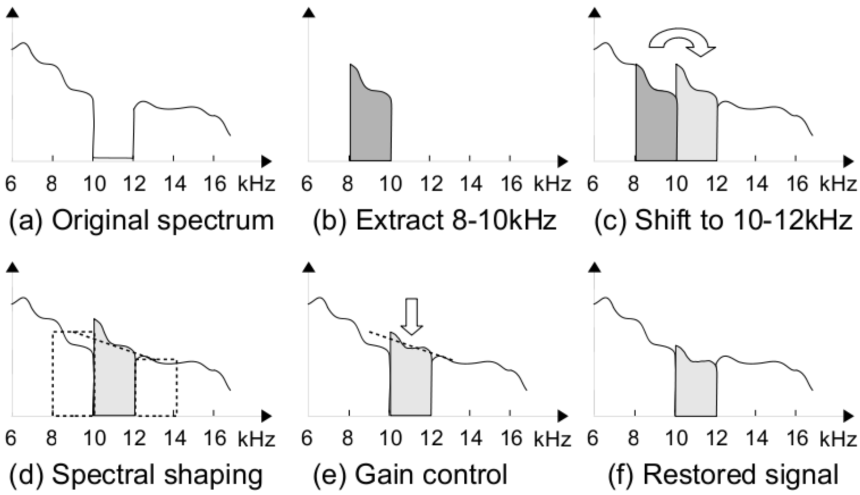


Figure 6: Spectral correction processing in the proposed MDC method 2.

Table 2: Experimental conditions: Categories of packet loss condition

Condition	Average packet loss rate [%]	Total bitrate [kbit/s]
Random packet loss	0	128,160, 192,224, 256,320
	5	
	10	

Table 3: 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

4.2 Results and discussion

First, the estimated ODG score of the proposed method 1 and 2 as a function of the total bitrate is shown in Fig. 8. These results show that the proposed method 2 is superior to the proposed method 1 in all conditions. The ODG score in method 2 is better by 0.04–0.14 than method 1. Therefore, the results of method 2 are discussed below.

Figure 9 shows the estimated ODG score as a function of the total bitrate for encoding methods of the four types. As described in the previous section, they are the three different types of PLC methods including the proposed one and the simple MP3 encoding method. Here, the six kinds of total bit rate were applied to PLC methods of the three kinds: from 128 kbps (64 kbps $\times$ 2) to 320 kbps (160 kbps $\times$ 2). The bitrate applicable to a single description of the two MDC methods is limited up to 160 kbps because the sampling frequency for the description is 22.05 kHz. Bitrates from 64 kbps to 112 kbps were also examined along with those from 128 kbps to 320 kbps for the simple MP3 encoding method. Panel (a) of this figure shows the ODG scores of no-packet loss condition. Panels (b) and (c)

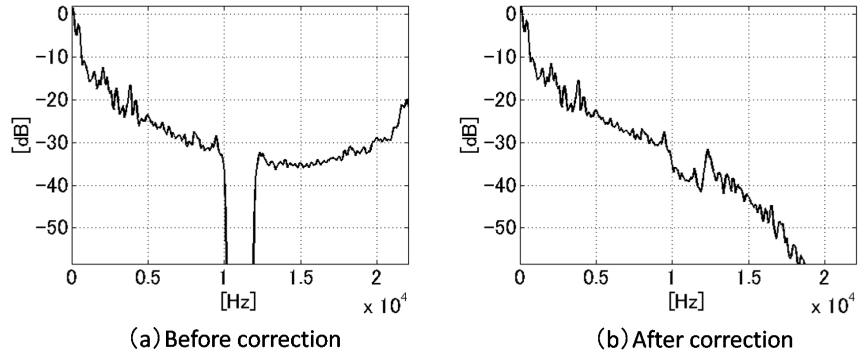


Figure 7: Example of the spectral correction result applying the proposed MDC method 2.

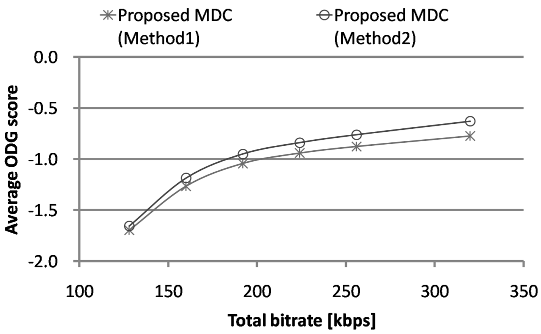


Figure 8: Comparison of the average ODG scores of proposed methods 1 and 2.

respectively correspond to the packet loss rates of 5 % and 10 %.

The ODG score results shown in Figs. 9(a) to 9(c) indicate that the proposed MDC method 2 outperforms the basic MDC method for all conditions by a rate of about 1.7 points higher at the maximum to about 0.9 points at the minimum. Moreover, these results show that the proposed MDC method exhibits higher ODG scores than those obtained using the double MP3 method with the extent of the total bitrates of 128–224 kbps, while the double MP3 method exhibits higher ODG scores than those obtained using the proposed MDC method for total bitrates of 256–320 kbps.

Next, we examine the results at every different packet loss rate in Fig. 9. In the no packet loss condition (Fig. 9(a)), it is apparent that the single MP3 method exhibits the highest ODG scores among the methods over all bitrates examined. The proposed MDC method 2, on the other hand, exhibits the second highest ODG scores for total bitrates of 128–224 kbps.

In summary, the proposed MDC method outperforms the other PLC methods examined in this study from the perspective of sound quality at practical total bitrates less than 256 kbps, under high packet loss rates such as 5 % and 10 %.These results indicate that the proposed MDC method can realize practical and stable streaming of wideband sound signals on the transmission path even with drastically changing packet loss rates.

5. CONCLUSION

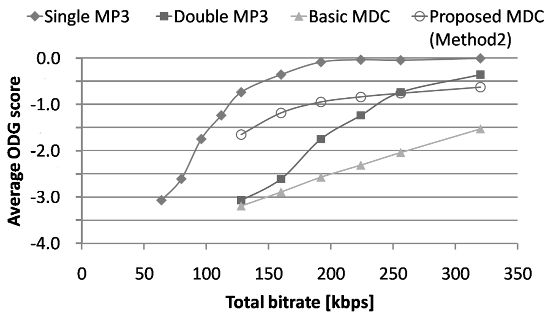
We herein presented the need for mitigating the playback quality deterioration of the provided data for a real time-wideband digital audio broadcasting application over a lossy packet network. To do so, we have proposed the MDC method of a combination of the sample splitting process in the time domain with a high-efficiency audio codec such as MP3, and the Wiener filter for frequency compensation. Moreover, spectral complement processing was combined with the method using the Wiener filter. Experimental results show that our proposed MDC method can improve ODG scores to 0.12–1.65 compared with the conventional methods in terms of the playback sound quality under conditions of bit rates of 160– 192 or 224 kbps, and of transmitted random packet loss rates from 5 or 10%. Additional work is necessary for investigation of the performance of our proposed method with another high-efficiency audio codec such as AAC.

ACKNOWLEDGMENTS

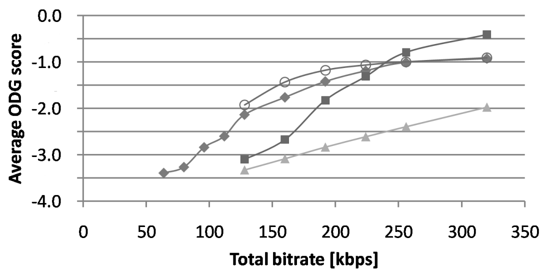
This study was partially supported by the Strategic Information and Communications R&D Promotion Programme (SCOPE) by the Ministry of Internal Affairs and Communications of Japan and the GCOE program (CERIES) of the Graduate School of Engineering, Tohoku University. We are also grateful to the members of the Shiroishi Rotary Club for their support.

REFERENCES

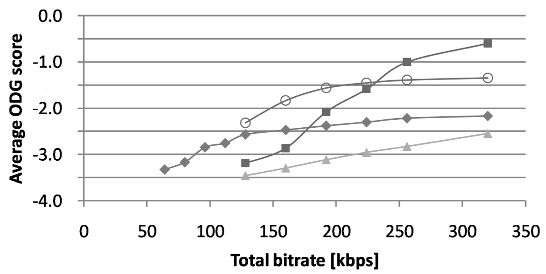
[1] J. Postel, Transmission Control Protocol, *RFC793*, 1981.
[2] J. Postel, User Datagram Protocol, *RFC768*, 1980.
[3] H. Schulzrinne, S. Casner, R. Frederick, and V. Jacobson, "RTP: A Transport protocol for real-time applications," *RFC1889*, 1996.
[4] O. J. Wasem, D. J. Goodman, C. A. Dvorak, and H. G. Page, "The effect of waveform substitution on the quality of PCM packet communication," *IEEE Trans. ASSP-36(3)*, pp. 342–348, 1988.
[5] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery technique for streaming audio," *IEEE Network*, vol. 12, no. 5, pp. 40–48, 1998.
[6] 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. Int. Symp. on Multimedia Software Engineering*, pp. 17-24, 2000.
[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. Int. Symp. on Frontiers in Speech and Hearing Research*, Tokyo, Japan, Mar. 2006.
[8] 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.
[9] V. K. Goyal, "Multiple Description Coding: Compression Meets the Network," *IEEE Signal Processing Magazine*, vol. 8, no. 5, pp. 74–93, 2004.



(a) Average packet loss rate: 0 % condition



(b) Average packet loss rate: 5 % condition



(c) Average packet loss rate: 10 % condition

Figure 9: Comparison of average ODG scores for different target methods under a several percent random packet loss rate condition.

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

[11] H. Ofir, D. Malah, and I. Cohen, "Audio packet loss concealment in a combined MDCT–MDST domain," *IEEE Signal Processing Letters*, vol. 14, no. 12, pp. 1032–1035, 2007.

[12] M. Suzuki, T. Sakai, A. Ito, and S. Makino, "Reduction method of side information for packet loss concealment based on spectrum striping coding," *Proc. 19th Int. Cong. of Acoustics*, 2007.

[13] 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. Japan–China Joint Conference on Acoustics*, 2007.

[14] 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.

[15] Y. Ephraim, H. Lev-Ari, and W. J. J. Roberts, "A brief survey of speech enhancement," *the Electronic Handbook*, CRC Press, 2003.

[16] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust. Speech Signal. Process.*, vol. ASSP-27, pp. 113–120, 1979.

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

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

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