Journal of Information Hiding and Multimedia Signal Processing
Ubiquitous International

Volume **5**, Number **2**, April **2014**

Packet Loss Concealment of Voice-over IP Packet using Redundant Parameter Transmission Under Severe Loss Conditions

Takeshi Nagano

Research Organization of Electrical Communication
Tohoku University
6-6-5 Aramaki aza Aoba, Aoba-ku, Sendai 980-8579, Japan
nagano@spcom.ecei.tohoku.ac.jp

Akinori Ito
Graduate School of Engineering
Tohoku University
6-6-5 Aramaki aza Aoba, Aoba-ku, Sendai 980-8579, Japan
aito@spcom.ecei.tohoku.ac.jp

Received November, 2013; revised November, 2013

Abstract. *This paper describes an outline of a project for developing a VoIP codec that can be used under a very severe communication environment where half of the packets drop. The codec is based on G.729 CS-ACELP, and a packet loss concealment (PLC) methods with redundant information will be used for enhancing speech quality. First, we assessed the importance of G.729 parameters, where we found that parameters related to the spectral shape and gain were relatively important. Then we evaluated speech quality when those important parameters were redundantly transmitted. Next, we developed two methods to reduce bitrate of the redundant parameters: one is to use the bit-flip function, and the other one is to use a discriminative model. From the experimental result, we found that both of the methods gave similar results, where quality improvement is almost in proportion to the redundant bitrate.*
Keywords: Voice over IP, G.729, Packet loss concealment

1. **Introduction.** The Voice over Internet Protocol (VoIP) is widely used nowadays as a method to make a phone call. The VoIP network is more exible and low cost than the conventional public switched telephone network (PSTN) because of the exibility of an IP network. Not only under an ordinary situation, IP-based telephony is known to be more robust than PSTN-based phone network under a situation of large-scale disaster [1, 2]. However, under a special situation such as a large-scale disaster, congestion of VoIP traffic is inevitable. Not only the traffic issue, the disaster such as an earthquake, flood or tsunami will destroy network facilities, which make the situation severer. Under a congested network, a real-time audio communication such as VoIP needs a packet loss concealment (PLC) technique [3]. As ratio of VoIP traffic with respect to all traffic is very small, packet losses are not considered as a big problem in a normal situation. In fact, very severe packet loss condition assumed in conventional researches was 15% to 20% loss rate [4, 5], which actually rarely occur. Network management protocol for error reduction

was also researched[6]. However, in a situation under a large-scale disaster, it will be difficult to guarantee the network quality and the network traffic becomes unpredictable.

Our goal is to develop a method for VoIP application to be used in a severe situation. Target packet loss rate of our research is around 50%, which can be occur when the network is almost down.

2. Conventional Packet Loss Recovery Techniques for VoIP.

Our target codec for VoIP is G.729 [7]. There have been a number of works for recovering packet losses of G.729. These works can be categorized into two types: the waveform-level concealment and the parameter-level concealment.

Waveform-level concealment methods can be used for not only G.729 codec but also any VoIP codecs. For example, Lee et al. proposed a waveform-based method that used waveform synthesis and overlap-add technique [8]. Although the waveform-based methods require no redundant information, performance of those methods under severe conditions will be limited.

Parameter-level concealment methods estimate parameters in a G.729 packets using interpolation, prediction or redundant information. The most straightforward method is a packet-copy, which simply uses the most recently received packet instead of the lost packet. Wang and Gibson [9] proposed a method to interpolate LSF parameters in a G.729 packet to enhance the recovered speech quality. Other methods transmit redundant information with a packet to enhance the recovered speech. Most methods use error correction code such as Reed-Solomon coding [10]. To reduce the amount of redundant bits, the unequal error protection (UEP) [11, 12] applies error-correction coding to a part of a packet or limited packets among all packets that have relatively large effect on the recovered speech quality. UEP-based error concealment methods work well, but the error protection based on the error correction code does not work at all when the packet loss rate exceeds the limit of the error correction, which causes sudden and drastic degradation of speech quality.

Thus, we propose an error concealment method under severe packet loss condition, which has "graceful degradation" property. This method is based on the design of redundant information for multiple description coding [13].

3. Importance of Parameters in a G.729 Packet.

3.1. Structure of G.729 Packet.

G.729 (CS-ACELP, 8kbps) is a low-bitrate speech codec standardized by the International Telecommunication Union (ITU) [7]. G.729 is a kind of Code-Excited Linear Prediction (CELP) codec, where the line spectrum pair (LSP) parameter is calculated by linear predictive coding (LPC) synthesis

lter, and the excitation signal is calculated from residue of LPC analysis. In the G.729 decoder, two subframes are used for encoding one packet. The extracted parameters are packed into a packet in every 10 ms, and transmitted in 8 kbit/s bitrate. Table 1 shows a list of parameters of G.729. All parameters are summarized by kinds of parameters: LSP (parameters for line spectrum pair calculation), PITCH (for pitch calculation), CODE (excitation signal calculation) and GAIN (of pitch gain and adaptive codebook gain) in Table 2.

3.2. Influence of parameter loss on speech quality.

In this section, we investigate importance of each parameter on speech quality. We first split the parameters in a packet into the above four groups (LSP, PITCH, CODE and GAIN). Then we conducted a simulation where the parameters of the selected groups are lost and that of the other

groups are preserved. We tried 2^4 - 1 = 15 combinations of the lost groups. We used MOS-LQO [14] as an evaluation metric, which was calculated based on PESQ [15]. Opticom OPERA [16] was used for calculating the MOS-LQO value.

Table 1. Parameters of G.729 codec

	symbol	role	# of bits
LSF	L0	Moving-average predictor codebook	1
	L1	1st codebook index	7
	L2	2nd codebook(lower) index	5
	L3	2nd codebook(upper) index	5
1st subframe	P1	Pitch period	8
	P0	Parity check on 1st period	1
	C1	Codebook index(position)	13
	S1	Codebook index(sign)	4
	GA1	Pitch and codebook gains(1st codebook)	3
	GB1	Pitch and codebook gains(2nd codebook)	4
2nd subframe	P2	Pitch period(relative)	5
	C2	Codebook index(position)	13
	S2	Codebook index(sign)	4
	GA2	Pitch and codebook gains(1st codebook)	3
	GB2	Pitch and codebook gains(2nd codebook)	4

Table 2. Parameters of each process

process	parameter	total [bit]
LSP	L0, L1, L2, L3	18
PITCH	P0, P1, P2	14
CODE	C1, S1, C2, S2	34
GAIN	GA1, GB1, GA2, GB2	14

Table 3. Evaluation data

Data set	ITU-T P.50 Real speech AMERICAN_ENGLISH(197 s) JAPANESE(173 s)
Contents	8 files by male speakers and 8 files by female speakers for each set

First, we selected "target groups" from the four parameter groups. In the simulation experiment, we randomly dropped packets in 50% probability. When recovering the lost packet, we recovered the correct parameters of the target group and the parameters of the other groups were copied from the previous packet. We can assess the impact on losing parameters of a certain group by comparing the quality of signals from results from different target groups. Figure 1 shows the owchart of the experiment.

Table 3 denotes about the evaluation data. The evaluation set is selected AMERICAN ENGLISH and JAPANESE from ITU-T P.50 data. Total length of set is 197 seconds and 173 seconds.

Figure 2 shows the experimental result, where the X-axis denotes bit length of the transmitted parameter, and the Y-axis denotes MOS-LQO.

In figure 2, total bit length of the target groups is denoted in the legend (number in a parenthesis). From the result, we can see that influence of the parameter on MOS-LQO varies from group to group. When only one target group is protected, LSP (18) and GAIN (14) showed relatively good quality. Among the all combination, GAIN+LSP (32) showed good quality considering the bitrate of the target group. Therefore, we consider how to protect LSP and GAIN parameters in the next section.

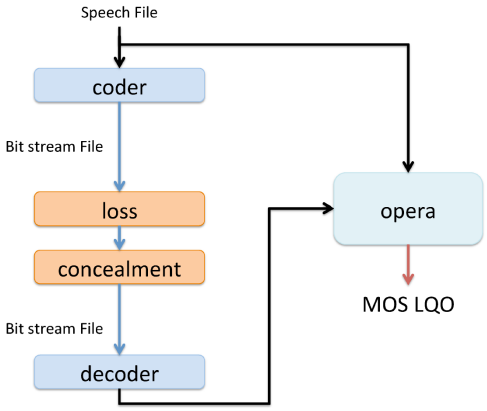

FIGURE 1. Experiment flow

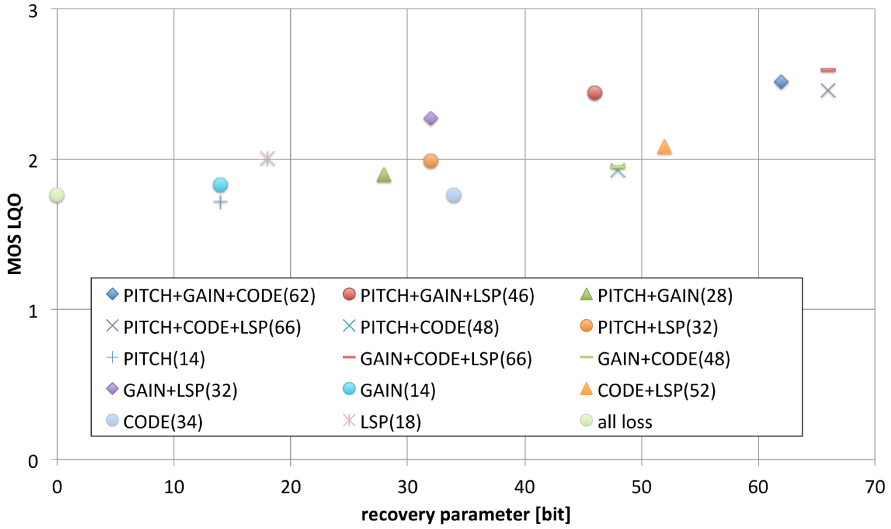

FIGURE 2. Parameter and MOS LQO.(frame loss rate=50%)

4. **Packet loss concealment by redundant transmission of GAIN and LSP parameters.**

4.1. **Redundancy of GAIN and LSP parameters.**

From the result of the previous section, we found that GAIN and LSP parameter group had large impact on speech quality. In this section, we consider protecting parameters in GAIN and LSP groups by redundantly transmitting the parameters.

Parameters in GAIN and LSP groups include two-stage vector quantization (VQ) indexes. For the GAIN group, GA1 and GA2 are the first stage indexes, GB1 and GB2 are the second stage indexes of the two-stage VQ. For the LSP group, L1 is the first stage index, L2 and L3 are the second stage indexes.

In the two-stage VQ, the input vector is quantized at the first stage, and the error is further quantized at the second stage [17]. Therefore, information of the first stage is more important than that of the second stage. Thus, we compared the two conditions on redundancy: in the first condition, all parameters in GAIN and LSP groups were redundantly transmitted; in the second condition, only the first VQ codes were redundantly transmitted.

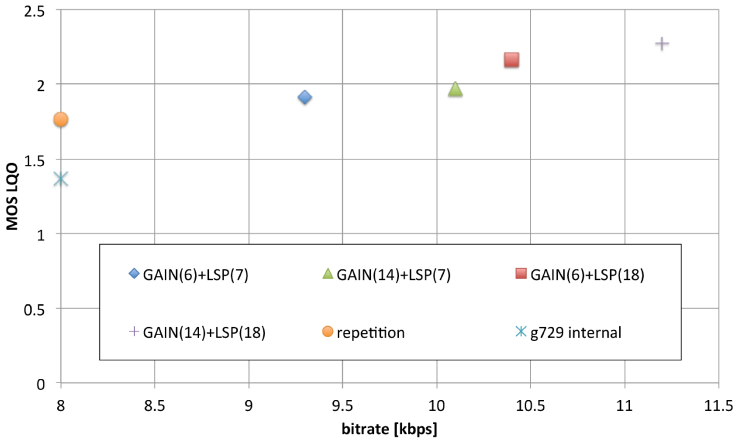

FIGURE 3. MOS LQO of GAIN and LSP parameter combination

We assume random packet loss, and the loss rate was 50%. Note that the packet loss under higher loss rate tends to be bursty. Here, we assume using interleaving technique to avoid burst packet loss [18]. The redundant information of a frame was attached to the previous frame. Thus, when a packet was lost and the previous packet was received, we used the redundant bits of the previous packet to recover the parameters. If the previous packet was also lost, the parameters were copied from the nearest packet. The parameters that were not included in the redundant bits were just copied from the nearest packet. The evaluation data were same as that in Table 3.

Figure 3 shows the experimental result. The X axis denotes bit rate (kbit/s) of the transmitted parameter including redundant bits. Here, "G.729 internal" shows the result that uses G.729 internal packet loss concealment method. "repetition" shows the result that copies the nearest previously received parameters. GAIN(6) denotes the result where the first VQ codes were used as the redundant information (number in the parentheses means the number of redundant bits), and GAIN(14) denotes the result with all GAIN bits as redundant bits. LSP(7) denotes the result where the first VQ codes were used as the redundant information, and LSP(18) denotes the result with all LSP bits as redundant bits. "+" denotes a combination of parameters.

In figure 3, we can see that improvement in speech quality is almost in proportion to the amount of redundant information. From the comparison between GAIN(6) and GAIN(14), and comparison LSP(7) and LSP(18), we can conclude that influence of GAIN and LSP was almost same.

4.2. **Bitrate reduction using bit flip.** As shown in the previous section, loss of GAIN and LSP parameter groups strongly affects the speech quality. When GAIN and LSP are transmitted as redundant information, we need considerable increase of bitrate. Thus, we consider a method to reduce the redundant information. Then we propose a method to use bit flip as redundant information. Bit flip is defined as an operation by which an arbitrary bit of a variable is reversed. Bit ip that reverse the i-th bit of variable x is

defined as follows.

$$\mathrm{flip}(x, i) = x \oplus 2^{i-1} \tag{1}$$

Here, $\oplus$ denote exclusive-or operation.

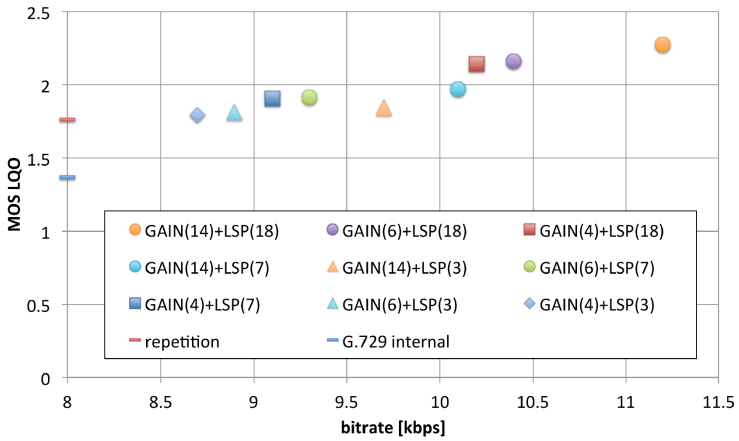

FIGURE 4. MOS LQO of GAIN and LSP parameter bit reduction using bit flip

The basic idea of bit-flip-based redundancy is as follows. Let the j-th parameter at time t be $p_{t,j}$. When $p_{t,j}$ is lost and no redundant information of $p_{t,j}$ is available, we just use $p_{t+1,j}$ instead of $p_{t,j}$. Here, we determine $i(t, j)$ where $\mathrm{flip}(p_{t+1,j}, i(t, j))$ is a better approximation of $p_{t,j}$ than $p_{t+1,j}$. As the bit length of $i(t, j)$ is approximately $\log_2 p_{t,j}$, we can reduce the bitrate by sending $i(t, j)$ compared with sending $p_{t,j}$ itself as redundant information. Note that we could use $p_{t+1,j}$ rather than $p_{t+1,j}$ for concealment; the reason why we use $p_{t+1,j}$ is that we can send redundant information without latency because $i(t, j)$ can be calculated when we have $p_{t+1,j}$. If we use $p_{t-1,j}$ for concealment, we cannot send $p_{t-1,j}$ until $i(t, j)$ is calculated using $p_{t,j}$.

Bit flip position is calculated as follows. First, distance between code vector in the codebook is defined. Let $\mathrm{x}[k] = (x_1[k], \ldots, x_M[k])$ be the k-th code vector in the codebook. Then the distance between two code vectors $d(a, b)$ is calculated as an Euclidean distance of two vectors,

$$d(a, b) = \|\mathrm{x}[a] - \mathrm{x}[b]\| . \tag{2}$$

Then we calculate the optimum flip position $i(t, j)$ as

$$i(t, j) = \arg\min_i d(\mathrm{flip}(p_{t+1,j}, i), p_{t,j}). \tag{3}$$

Then $i(t, j)$ is transmitted as redundant information instead of $p_{t,j}$.

In GAIN parameters, bit length of the codebook index GA1 and GA2 is reduced with bit flip from 6 bits to 4 bits. In LSP parameters, bit length of the codebook index L1 is reduced from 7 to 3. Because the flipped parameter $\mathrm{flip}(p_{t+1,j}, i(t, j))$ might be different from $p_{t,j}$, the quality of the restored signal will be degraded compared with the case when $p_{t,j}$ is used as redundant information.

4.3. **Experiment.** We conducted an experiment to compare the bit-flip-based and redundant information based methods. The condition of this is same as the previous experiment.

Figure 4 shows the result under the assumption of random loss. In figure 4, "GAIN(4)" shows result of GAIN with bit flip, and "LSP(3)" shows result of LSP with bit flip. From this result, we can see that all results except the G.729 internal have linear relationship,

which shows that the bitrate reduction and quality degradation using the bit flip was within the same trade-off relationship of the other conditions.

TABLE 4. Training data of the descriminative model

Corpus	ATR speech database, set B
Speaker	1 male(MHO), 1 female(FKN)
Sentences	From A1 to A10
# of frames	8914

5. A Parameter Selection Method Using Discriminative Model.

5.1. **Basic idea.** If a parameter is lost, speech quality is degraded. The impact of parameter loss on speech quality depends on characteristic of the frame and kind of the lost parameter. Thus, we tried to model the relationship between kind of the lost parameter and degradation of speech quality. Using the estimated model, we can protect only those parameters that have a larger impact on the speech quality using FEC.

5.2. **Selection of parameters using a discriminative model.** In the following equation, d denotes the impact of parameter loss on the speech quality.

$$d_{t,j} = MOS(S) - MOS(S_{loss}(t,j)) \tag{4}$$

where S and S_loss denotes G.729 coded speech without and with errors, respectively. Here, $S_{loss}(t,j)$ denotes the speech in which parameter group $j \in \{1,2,3,4\}$ (which correspond to PITCH, LSP, CODE and GAIN) in the t-th frame is lost. $MOS(S)$ denotes the MOS-LQO value with respect to the encoded speech S. If $d_{t,j}$ is large, it means that the subjective impact of losing the parameter group is large. When calculating $S_{loss}(t,j)$, we used a uniform random number to conceal the lost parameters in the group.

Then we divide $S_{loss}(t,j)$ for all combination of t and j into either "severe parameter loss" set L_0 and not-so-severe parameter loss" set L_1 using a threshold θ, as follows:

$$\begin{array}{l}(t,j) \in L_0 \text{ if } d_{t,j} > \theta \\ (t,j) \in L_1 \text{ otherwise}\end{array} \tag{5}$$

If we use small θ, most losses are judged as "sever" losses, while using larger θ makes the judgement less strict.

Once all parameter losses are categorized into the two sets, we can train a classifier to discriminate a speci

c parameter loss into one of the "severe" and "not-so-severe" classes using a feature vector t_t. Let $C_{j,\theta}(\mathrm{f}_t)$ be a classifier that returns either 0 or 1, where 0 means that the degradation of $S_{loss}(t,j)$ is severe and 1 means it isn't. We can train various $C_{j,\theta}$ by changing the threshold θ. We used the parameter values from $(t-n)$-th to $(t+n)$-th frame:

$$\mathbf{f}_t = (\mathrm{p}_{t-n}, \mathrm{p}_{t-n+1}, \ldots, \mathrm{p}_{t+n}) \tag{6}$$

where p_t denotes the vector of the parameter values in the t-th frame. Dimension of the vector was 15, as shown in Table 1.

When sending t-th frame of a speech signal S, we extract feature vectors f_t, and evaluates the impact of parameter loss $C_{j,\theta}(\mathrm{f}_t)$. If $C_{j,\theta}(\mathrm{f}_t) = 0$; the j-th parameter is transmitted redundantly. Here, we can control the tradeoff between the bitrate and speech quality by changing the threshold θ.

5.3. **Evaluation.** We employed a Support Vector Machine (SVM) [19] as discriminative model. The SVM was trained using the training data described in Table 4. As described above, we trained one SVM for each combination of a threshold value and one of the four parameter groups. We used RBF (Radial Basis Function) for the kernel function of the SVM. We used libsvm [20] for training and using SVM. Tab. 5 describes the evaluation data.

TABLE 5. Evaluation data

Data set	ITU-T P.50 Real speech, AMERICAN ENGLISH (197sec.) and JAPANESE (173sec.)
Contents	8 files by male speakers and 8 files by female speakers for each set

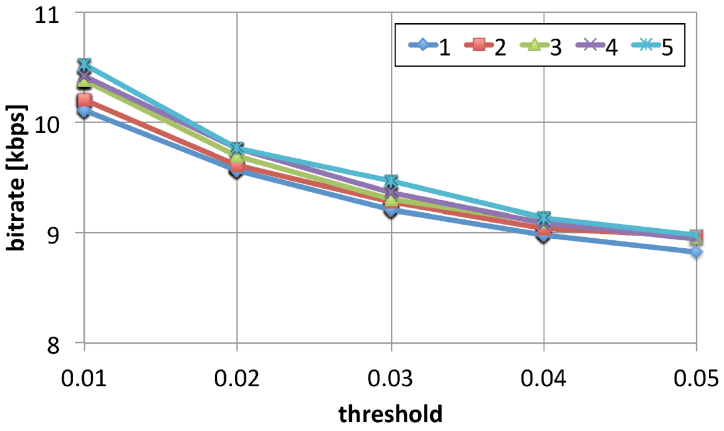

FIGURE 5. Threshold θ and bitrate

In the experiment, parameters of each frame were packed into a packet, and the redundant parameters of the previous frame were packed together, too. In addition, a 4-bit header was appended to each of the packet to indicate which parameter is appended redundantly. When simulating the packet losses, we used the Gilbert model as the packet loss model. The packet loss rate was varied from 10% to 50%, and the average burst size was set to 2.0.

When a packet is lost, the redundant parameters were recovered correctly, and the other parameters were copied from the previous packet. Figure 5 shows the experimental result, where the X-axis denotes the threshold and the Y-axis denotes the bitrate, and each legend denotes number of previous/following frames. For example, the line with legend "1" is the result where three frames (one current frame, one previous and one following frame) was used. In figure 5, we can see that the bitrate can be controlled using the threshold. Figure 6 shows the experimental result of quality of the recovered sound when the average burst size was 2.0.The X-axis denotes the packet loss rate and the Y-axis denotes MOS LQO. In the legend, "K-M" means that K is the number of feature frame and M is the threshold. Here, "G.729 internal" shows the result that uses G.729 internal packet loss concealment method, and "repetition" shows the result when the nearest previously received parameters were copied. In Figure 6, each vertical bar denotes MOS LQO value, where the lowest point denotes the 50% loss case, and highest

point denotes the 10% loss case. From this result, we can see that the quality of the speech linearly improved with respect to the bitrate.

6. **Discussion.** Proposed method can be integrated with layered coding. Layered coding has two layers. One layer is base layer that have high priority in the network. Another layer is enhancement layer that have lower priority than base layer. If base layer is received, data can be decoded with acceptable quality. Furthermore if enhancement layer is received, quality of data can be improved.

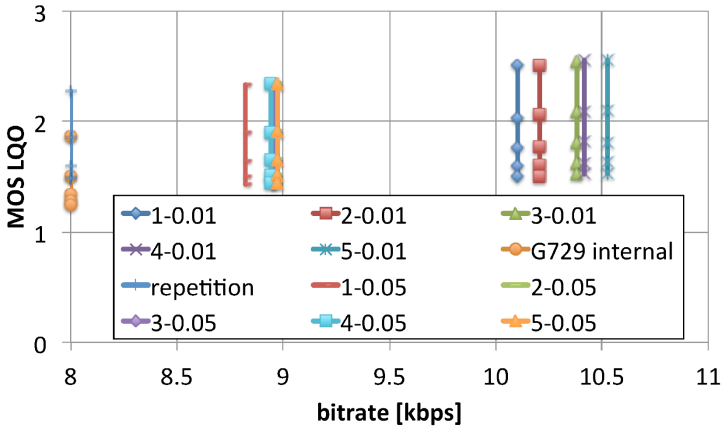

FIGURE 6. Bitrate and MOS-LQO

In proposed method, it is considerable that G.729 paramters are assigned to base layer and PLC parameters such as bit-flip position or redudant parameters are assigned to enhancement layer. Under severe packet loss condition, decoded speech quality can be improved when enhancement layer is received. Concepts of layered coding are also used such as scalable coding[21] and watermarking[22] and so on.

7. **Conclusion.** We investigated packet loss concealment methods in severe packet loss condition. First, we investigated impacts of parameter loss on the speech quality, and found that GAIN and LSP parameter groups were more important than the other two groups. Then we evaluated the effect of redundant parameter transmission, and confirmed improvement of 1.5 MOS-LQO. Then we developed two methods to control trade-off between bitrate and speech quality. The first one is to use bit flip, and the other one is to use discriminative models. From the experimental results, we found that both methods could control the bitrate but the speech quality was almost linear to the redundant bitrate. Considering the simplicity and effectiveness, the method based on the bit flip can be a better choice for this task.

In these works, we assumed that the packet loss pattern is random. In the actual situation, packet loss pattern will be often bursty. Therefore, we need to use interleaving technique, which increases the latency of the speech. Then we will investigate the trade-off between latency and speech quality through subjective evaluation.

Acknowledgment. A part of results presented in this paper was achieved by carrying out an MIC program "Research and development of technologies for realizing disaster-resilient networks" (the no.3 supplementary budget in 2011 general account).

REFERENCES

[1] T. Ichiguchi, Robust and usable media for communication in a disaster, *Science & Technology Trends: Quarterly Review*, no. 41, pp. 44-55, 2011.

[2] K. Cho, C. Pelsser, R. Bush, and Y. Won, The Japan earthquake: the impact on traffic and routing observed by a local ISP, *Proc. of the Special Workshop on Internet and Disasters*, 2011.

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

[4] J. Turunen, P. Loula, and T. Lipping, Assessment of objective voice quality over best-effort networks, *Journal of Computer Communications*, vol. 28, no. 5, pp. 582-588, 2005.

[5] L. Deng and, R. A. Goubran, Assessment of effects of packet loss on speech quality in VoIP, *Proc. of The 2nd IEEE Internatioal Workshop on Haptic, Audio and Visual Environments and Their Applications*, pp. 49-54, 2003.

[6] T. H. Liu, S. C. Yi, and X. W. Wang, A fault management protocol for low-energy and efficient wireless sensor networks, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 4, no. 1, pp. 34-45, 2013.

[7] G. 729: Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear-prediction (CS-ACELP), *ITU-T Recommendations*, 2007.

[8] M. K. Lee, S. K. Jung, H. G. Kang, Y. C. Park, and D. H. Youn, A packet loss concealment algorithm based on time-scale modification for CELP-type speech coders, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 116-119, 2003.

[9] J. Wang, and J. D. Gibson, Parameter interpolation to enhance the frame erasure robustness of CELP coders in packet networks, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 745-748, 2001.

[10] W. Jiang, and H. Schulzrinne, Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss, *Proc. of the 12th international workshop on Network and operating systems support for digital audio and video*, pp. 73-81, 2002.

[11] U. Horn, K. Stuhlmüler, M. Link, and B. Girod, Robust internet video transmission based on scalable coding and unequal error protection, *Journal of Signal Processing: Image Communication*, vol. 15, no. 1-2, pp. 77-94, 1999.

[12] H. Sanneck, and N. T. L. Le, Speech property-based FEC for internet telephony applications, *Proc. of Multimedia Computing and Networking*, SPIE 3969, pp. 38-51, 2000.

[13] A. Ito, and S. Makino, Designing Side Information for Multiple Description Coding, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 1, no. 1, pp. 10-19, 2010.

[14] P.862.1: Mapping function for transforming of P.862 to MOS-LQO, *ITU-T Recommendations*, 2003.

[15] P.862: Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs, *ITU-T Recommendations*, 2001.

[16] Opticom GmbH, *OPERA voice/audio quality analyzer*, available at http://www.opticom.de/products/opera.html.

[17] B. H. Juang, and A. Gray Jr., Multiple stage vector quantization for speech coding, *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, pp. 597-600, 1982.

[18] T. K. Chua, and D. C. Pheanis, QoS evaluation of sender-based loss-recovery techniques for VoIP, *Journal of IEEE Network*, vol. 20, no. 6, pp. 14-22, 2006.

[19] C. J. C. Burges, A tutorial on support vector machines for pattern recognition, *Journal of Data Mining and Knowledge Discovery*, vol. 2, no. 2, pp. 121-167, 1998.

[20] C. C. Chang, and C. J. Lin, LIBSVM : a library for support vector machines, *ACM Trans. Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1-26, 2011.

[21] M. H. Taieb, J. Y. Chouinard, D. Wang, K. Loukhaoukha, and G. Huchet, Progressive coding and side information updating for distributed video coding, *Journal of Information Hiding and Multimedia Signal Processing*, vol. 3, no. 1, pp. 1-11, 2012.

[22] F. C. Chang, H. C. Huang, and H. M. Hang, Layered access control schemes on watermarked scalable media, *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology*, vol. 49, no. 3, pp. 443-455, 2007.