A Packet Loss Recovery of G.729 Speech Under Severe Packet Loss Condition

Takeshi NAGANO* and Akinori ITO*
* Tohoku University, Sendai, Japan
E-mail: nagano@spcom.ecei.tohoku.ac.jp Tel/Fax: +81-22-795-7084

***Abstract*—In a VoIP application, packet losses degrade speech quality. Especially, IP network under a large-scale disaster should cause severe packet losses. We investigate influence of parameter loss to speech quality using G.729. And we investigated an effect of packet loss concealment method using redundant G.729 parameters. As compared with "repetition" method, the proposed method could improve speech quality. We also propose a bitrate reduction method by sending bit flip position instead of codebook index.**

I. Introduction

The Voice over Internet Protocol (VoIP) is widely used nowadays as a method to make a phone call. Compared with the conventional public switched telephone network (PSTN), VoIP network is more flexible and low cost. Not only under an ordinary situation, IP-based telephony is known to be more robust compared with 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, typhoon, tsunami or tornado 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]. However, as ratio of VoIP traffic with respect to all traffic is very small [4], packet losses are not big problem in a normal situation. In fact, very severe packet loss condition assumed in conventional researches was around 15% loss rate [5], which actually rarely occur. However, in a situation under a large-scale disaster, it will be difficult to guarantee the network quality and the network traffic become 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.

Our target codec of VoIP is G.729 [6], which is a kind of code-excited linear prediction (CELP) codec. In a CELP-based codec, the input speech is split into frames, parameters are extracted frame-by-frame, and then the quantized parameters are packed into a packet. When concealing a lost packet, there could be two possibility for estimating the signal in the lost packet. The first kind of methods estimate the lost parameters from the previous (and occasionally the next) packets. After estimating the parameters, waveform is generated from the recovered parameters. The second kind of methods directly generate waveform of the lost part from the previous waveforms.

Examples of waveform level concealment methods are WSOLA [8] and AR-model-based waveform synthesis [9]. Parameter level concealment methods are divided into methods that estimate the parameters of the lost frame using the previous frames and methods that exploit redundant information attached to the standard packets. The most straightforward way of the former method is to estimate the lost parameters by copying the parameters from the previous packet. Yang et al. [10] proposed a method to estimate the lost frames by interpolation considering packet loss state. The latter method uses redundant information such as parity bits to recover the lost packet. This kind of technology is called "forward error correction (FEC)." Jiang and Schulzrinne [11] investigated relationship between the signal quality and loss pattern with parity-based FEC. When protecting all bits in a packet is too costly, only the important part in a packet can be rigidly protected using FEC and the other part is estimated from the previous packet (unequal error protection [12]).

In this paper, we investigated methods with parameter-level FEC. First, we investigate impact of losing a certain parameter in a frame on the quality of speech. Next, we investigate how the quality can be improved using unequal error protection, protecting only those parameters that have large impact on the speech quality. Finally, we proposed a method using bit flip to improve the speech quality with less redundant bits.

II. Overview of G.729

G.729(CS-ACELP, 8kbps) is a low-bitrate speech codec standardized by the ITU [6]. G.729 is a kind of a CELP codec, where the line spectrum pair (LSP) parameter is calculated by linear predictive coding (LPC) synthesis filter, 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 I 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 II.

III. 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

TABLE I
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 II
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

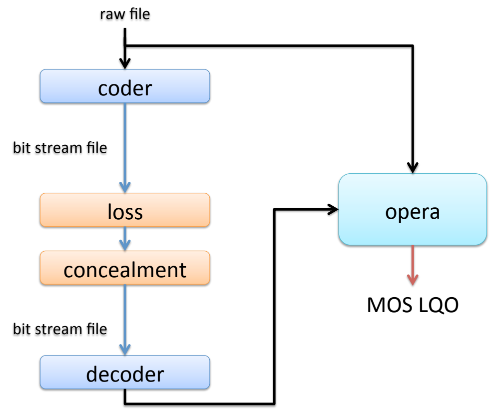

Fig. 1. Experiment flow

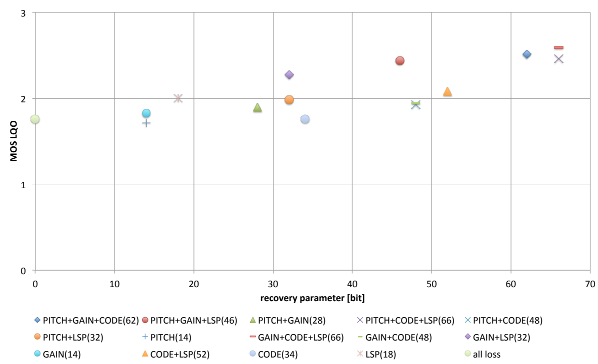

Fig. 2. Parameter and MOS LQO.(frame loss rate=0.5)

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 [13] as an evaluation metric, which was calculated based on PESQ [14]. Opticom OPERA [15] was used for calculating the MOS-LQO value.

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 the lost parameters of a certain group by comparing the quality of signals from results from different target groups. Figure 1 shows the flowchart of the experiment.

Table III denotes the evaluation data. The evaluation set is selected AMERICAN_ENGLISH and JAPANESE set of ITU-T P.50 data. Total length of the speech in the two evaluation sets were 197 seconds and 173 seconds, respectively.

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 parentheses). From the result, we can see that the influence of parameters on MOS-LQO varies from group to group.

TABLE III
EVALUATION DATA

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

When the 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.

IV. 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) index. For the GAIN group, GA1 and GA2 are the first stage index, GB1 and GB2 are the second stage index 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 [16]. 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

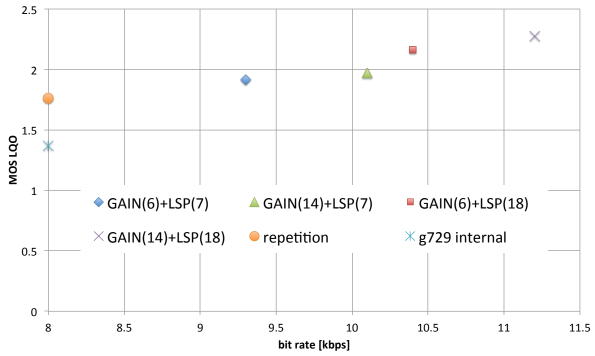

Fig. 3. MOS LQO of GAIN and LSP parameter combination.(average burst size = 1.0)

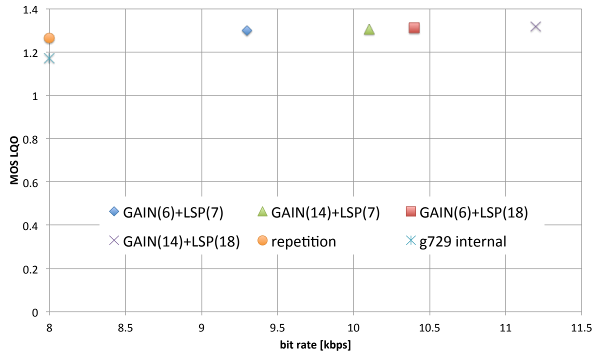

Fig. 4. MOS LQO of GAIN and LSP parameter combination.(average burst size = 5.0)

second condition, only the first VQ codes were redundantly transmitted.

We used the Gilbert model as the packet loss model. The packet loss rate was set to 50%, and the average burst size was set to 1.0 or 5.0. 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 was same as that in Table III.

Figure 3 shows the experimental result when the average burst size was 1.0. 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, 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 combination of parameters.

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

Figure 4 shows the result when average burst size was 5.0. Although the proposed method only exploited one past frame for redundant information, it showed better quality than the"repetition" condition. The advantage of the proposed method was smaller when the average burst size was longer, because the redundant information tended to be lost when the long burst loss occurred.

V. Bit length reduction of codebook index

In this section, we consider reducing bit length of redundant information. 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.

A. Bitrate reduction using bit flip

We propose a method to use bit flip as redundant information. Bit flip is defined as an operation by which arbitrary bit of a variable is reversed. Bit flip 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.

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

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

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

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

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

Then $i(t)$ is transmitted as redundant information instead of p_t.

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

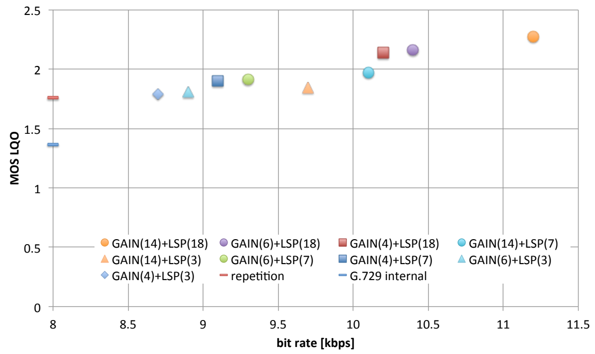

Fig. 5. MOS LQO of GAIN and LSP parameter bit reduction using bit flip(average burst size = 1.0).

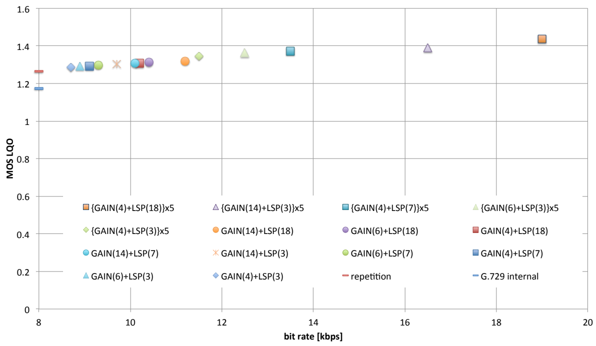

Fig. 6. MOS LQO of GAIN and LSP parameter bit reduction using bit flip(average burst size = 5.0).

from 7 to 3. Because the flipped parameter flip$(p_{t+1}, i(t))$ might be different from p_t, the quality of the restored signal will be degraded compared with the case when p_t is used as redundant information.

B. 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 5 shows the result when average burst size was 1.0. In Figure 5, "GAIN(4)" shows result of GAIN with bit flip, "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.

Figure 6 shows the result when the average burst rate was 5.0. In Figure 6, "GAIN(4)" shows the result of GAIN with bit flip, and "LSP(3)" shows that with bit flip. $\times N$ in the legend means that the redundant information was attached to the past N frames. We can see from this result that the speech quality improved when more redundant information was used. However, the quality looked to converge around MOS-LQO 1.4.

VI. Conclusions

We investigated influence of loss of parameters of G.729 for speech quality. As a result, GAIN and LSP parameter group had greater influence to the speech quality. Next, we investigated a method using GAIN and LSP parameter groups as redundant information. Finally, we investigate a method to reduce bitrate using bit flip. As result, the bit flip method could not outperform the ordinary bitrate-quality trade-off.

In the future work, we investigate possibility to exploit other redundant information for improving speech quality with smaller redundant information.

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," In *Proc. Special Workshop in Internet and Disasters*, 2011.
[3] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, no. 5, pp. 40–48, 1998.
[4] C. Labovitz, S. Iekel-Johnson, D. McPherson, J. Oberheide and F. Jahanian, "Internet Inter-Domain Traffic," in *Proc. SIGCOMM*, pp. 75–86, 2010.
[5] J. Turunen, P. Loula and T. Lipping, "Assessment of objective voice quality over best-effort networks," *Computer Communications*. vol. 28, no. 5, pp. 582–588, 2005.
[6] International Telecommunication Union, "G. 729: Coding of speech at 8 kbit/s using conjugate structure algebraic-code-excited linear-prediction (CS-ACELP)," ITU-T Recommendation, 2007.
[7] J. Rosenberg, "G. 729 error recovery for internet telephony",Project report, Columbia University,1997.
[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", In *Proc. ICASSP*, vol.1, pp. 116–119, 2003.
[9] G. Zhang and W. B. Kleijn,"Autoregressive model-based speech packet-loss concealment", In *Proc. ICASSP*, pp. 4797–4800, 2008.
[10] J. Yang, S. S. Yu, J. Zhou and Y. Gao, "A new error concealment method for consecutive frame loss based on CELP speech", *Computers & Electrical Engineering*, vol. 36, no. 5, pp. 1014–1020, 2010.
[11] W. Jiang and H. Schulzrinne, "Comparison and optimization of packet loss repair methods on VoIP perceived quality under bursty loss", In *Proc. Int. Workshop on Network and Operating Systems Support for Digital Audio and Video*, pp.73–81, 2002.
[12] U. Horn, K. Stuhlmüller, M. Link and B. Girod, "Robust Internet video transmission based on scalable coding and unequal error protection," *Signal Processing: Image Communication,* vol. 15, no. 1–2, pp. 77–94, 1999.
[13] International Telecommunication Union, "P.862.1: Mapping function for transforming of P.862 to MOS-LQO," 2003.
[14] International Telecommunication Union, "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," 2001.
[15] Opticom GmbH, "OPERA voice/audio quality analyzer.", http://www.opticom.de/products/opera.html.
[16] B.-H. Juang and A. Gray Jr., "Multiple stage vector quantization for speech coding," In *Proc. ICASSP*, vol. 7, pp. 597–600, 1982.