VQ Codebook Index Interpolation Method for Frame Erasure Recovery of CELP Coders in VoIP

Jeongseok Lim*,**, Hae Yong Yang*, Kyung Hoon Lee*, Sang Kyu Park** ***Regular Members***

ABSTRACT

Various frame recovery algorithms have been suggested to overcome the communication quality degradation problem due to Internet-typical impairments on Voice over IP(VoIP) communications. In this paper, we propose a new receiver-based recovery method which is able to enhance recovered speech quality with almost free computational cost and without an additional increment of delay and bandwidth consumption. Most conventional recovery algorithms try to recover the lost or erroneous speech frames by reconstructing missing coefficients or speech signal during speech decoding process. Thus they eventually need to modify the decoder software.The proposed frame recovery algorithm tries to reconstruct the missing frame itself, and does not require the computational burden of modifying the decoder. In the proposed scheme, the Vector Quantization(VQ) codebook indices of the erased frame are directly estimated by referring the pre-computed VQ Codebook Index Interpolation Tables(VCIIT) using the VQ indices from the adjacent(previous and next) frames. We applied the proposed scheme to the ITU-T G.723.1 speech coder and found that it improved reconstructed speech quality and outperforms conventional G.723.1 loss recovery algorithm. Moreover, the suggested simple scheme can be easily applicable to practical VoIP systems because it requires a very small amount of additional computational cost and memory space.

Key Words : Code-Excited Linear Prediction(CELP), Frame Erasure Recovery, G.723.1, Voice over IP(VoIP), VQ Codebook Index Interpolation Tables(VCIIT).

Ⅰ. Introduction

Since VoIP technology can reduce communication cost considerably and provide more efficient manageable aspect of the converged multimedia network, Internet telephony service is expected to grow rapidly and replace entire Plain Old Telephone Service(POTS) eventually. However, the best-effort service model on today's Internet inevitably causes various defects in telephony service such as packet loss, delay, jitter, and so on. This erroneous environment seriously degrades speech quality in VoIP[1], and these vulnerable aspects are impeding the widespread use of Internet telephony service at present.

Handling the Quality of Service(QoS) is regarded as an important area of VoIP technology, and various techniques and algorithms to provide or enhance it have been introduced. Conventional VoIP QoS management techniques can be categorized into two groups: one is the network-based approach which addresses its attention to develop new network protocols such as Differential Services (DiffServ) and Resource Reservation Protocol (RSVP). The other is trying to recover the missing speech packets at the terminal. Although the former is a more suitable approach to solve VoIP QoS problems, it takes too long for such protocols to diffuse throughout all network nodes and clients. Therefore, the latter, i.e., sender/receiver

* Electronics and Telecommunications Research Institute(limjs@etri.re.kr), ** Hanyang University(skpark@hanyang.ac.kr)
논문번호 : KICS2005-07-309, 접수일자 : 2005년 9월 28일

terminal-based loss recovery scheme, is more applicable to the current network environment.

Despite of its applicability to real-world problems compared to the network-based approach, most of the existing works on this area still have some practical defects such as additional bandwidth consumption, increment of delay or computational complexity. For example, the Media-specific Forward Error Correction(FEC) technique[2,9], the representative sender-based erasure recovery scheme, increases both the packet processing delay and requires more communication bandwidth. As a result, the effect of speech quality enhancement is considerably canceled out by these negative factors. Even worse, it requires much implementation cost since both the primary and secondary coders should be implemented. Similarly, much of receiver-based techniques including insertion, interpolation and regeneration, etc. also need a large amount of additional computational cost to enhance recovered speech quality[3,4].

In this paper, we suggest a novel receiver-based recovery scheme which enhances recovered speech quality with almost free computational cost and without an increase in of delay and bandwidth. Most of conventional recovery algorithms try to recover the lost speech frames by reconstructing missing coefficients or speech signal during speech decoding process, and thus they eventually need to modify the decoder software. The proposed frame recovery algorithm tries to reconstruct the missing frame itself, and does not require modifying the decoder software. In the proposed scheme, the VQ codebook indices of the erased frame are directly estimated by referring the pre-computed VCIIT using the VQ indices from the adjacent (previous and next) frames. In the pre-computing process, an element in the ith row and jth column of the VCIIT is determined by finding the optimal index which minimizes the difference between the interpolated value of the ith row and jth column element and the value of the candidate VQ codebook index.

We applied the proposed recovery scheme to the ITU-T G.723.1 [5] speech coder and found that it improved reconstructed speech quality, and outperforms the conventional G.723.1 loss recovery algorithm. Moreover, the suggested simple scheme can be easily applicable to practical VoIP systems because it requires far smaller amounts of additional computational cost and memory space.

Ⅱ. G.723.1 ERASURE RECOVERY SCHEME

2.1 Operation

ITU-T G.723.1 is the narrow band Code-Excited Linear Prediction(CELP) speech coder and consists of 5.3/6.3 Kbps dual rates. Two coders use Line Spectral Pair(LSP), adaptive codebook and fixed codebook coefficients to compress and decompress speech signals. The only difference of two coders is the fixed codebook coefficients extraction scheme with MP-MLQ for 6.3 Kbps and ACELP for 5.3 Kbps respectively. To provide robustness for indicated frame erasure, G.723.1 is equipped with receiver-based built-in error concealment strategy. If a frame erasure occurs, two decoders are switched from regular decoding routine to common erasure concealment block as shown in Fig. 1.

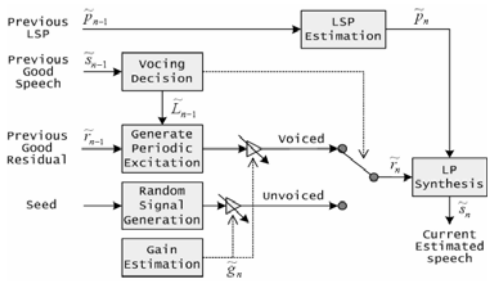

Fig. 1. Block diagram of the G.723.1 erasure recovery scheme

LSP coefficients of current frame are calculated in the decoder with (1) when speech frame is received normally. Because the encoder uses a first order differential coding method with previous LSP, the corresponding decoder reconstructs LSP by adding previous($\tilde{p}_{n-1}$) and current residual($\tilde{e}_n$) coefficients. However, in case of frame erasure, the current frame residual LSP vector($\tilde{e}_n$) is set to zero and LSP predictor coefficient b is larger than normal case as shown in (2).

$$\tilde{p}_n = b_1[\tilde{p}_{n-1} - p_{DC}] + p_{DC} + \tilde{e}_n \qquad (1)$$

$$\tilde{p}_n = b_2[\tilde{p}_{n-1} - p_{DC}] + p_{DC} \qquad (2)$$

where

$\tilde{p}_n$ quantized LSP vector of the nth frame;

p_{DC} long term DC component;

$\tilde{e}_n$ quantized residual LSP vector of the nth frame;

b first order fixed predictor, $b_1 = 12/32$ for good frames and $b_2 = 12/32$ for erased frames

The prediction of excitation signal($\tilde{r}_n$), which is reconstructed with adaptive and fixed codebook parameters in normal case, is performed in two different ways depending on the last previous good frame prior to the erased frame. The frame is checked with a voiced/unvoiced classifier based on a cross-correlation maximization function. The last 120 samples of the frame are cross-correlated with $L_2 \pm 3$ and the index which results in the maximum correlation value is chosen as the interpolation index candidate. If the prediction gain of the best vector is more than 0.58dB, the frame is declared as voiced, otherwise the frame is declared as unvoiced. The classifier returns 0 for the unvoiced case and the estimated pitch value($\tilde{L}_{n-1}$) for the voiced case. If the current frame is marked as erased, and the previous frame was classified as unvoiced, the current frame excitation($\tilde{r}_n$) is generated using an uniform random number generator. In the voiced case, the current frame is regenerated with periodic excitation having a period equal to the value provided by the classifier($\tilde{L}_{n-1}$). Finally, gain estimator attenuates 2.5dB for each frame and after erasure state continues 3 frames, the output is muted completely. The estimated excitation signal($\tilde{r}_n$) is synthesized using the LSP coefficient($\tilde{p}_n$) which is estimated with (2).

2.2 Performance Analysis

Fig. 2 shows speech quality degradation effect of G.723.1 low rate coder in an erasure environment. Fig. 2(a), 2(b) shows decoded speech signal in clean and 10% erroneous situation respectively and indicates frame error locations. The average spectral distortion(SD), which is usually used as an indicator of the perceptual quality of the LSP parameter distortion, of each case is plotted in Fig 2(c). The spectral distortion for a subframe n is defined as follows [6].

$$SD_n = \sqrt{\frac{1}{F_s}\int_0^{F_s} [10\log_{10}(P_n(f)) - 10\log_{10}(\hat{P}_n(f))]^2 df} \qquad (3)$$

Where $P_n(f)$ is LSP frequency spectrum prior to quantization in encoder side, while $\hat{P}_n(f)$ is trans-

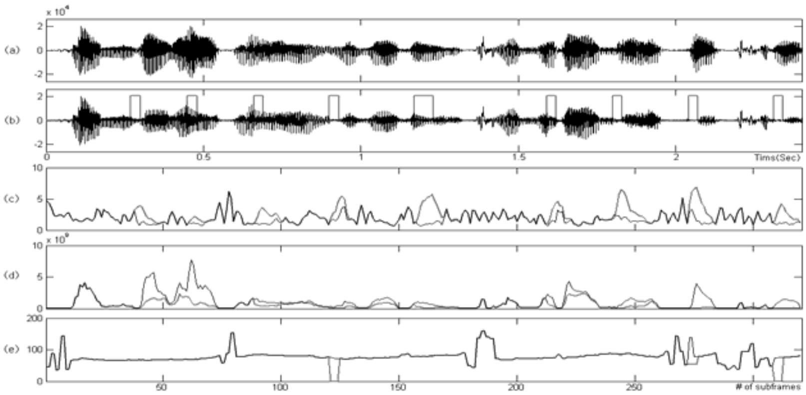

Fig. 2. Example and analysis of speech quality degradation effect due to the frame erasure (10% erasure rate).
(a) decoded signal in a clean environment
(b) decoded signal and frame error locations marked with rectangles in an erroneous environment
(c) spectral distortion contours in a clean (black line) and an erroneous (gray line) environment
(d) energy contours in a clean (black line) and an erroneous (gray line) environment
(e) pitch contours in a clean (black line) and an erroneous (gray line) environment

ferred(in the normal case) or estimated(in the erasure case) LSP spectrum in decoder side. According to Fig. 2 and (1), (2), if single LSP information is missed, LSP coefficient of current frame is naturally distorted and even affects several consecutive frames as shown in Fig. 2(c). This phenomenon results from differential coding strategy of LSP vector with the previous frame. As is the case for the LSP, energy attenuation is severe and also affects following several frames as shown in Fig. 2(d). It is caused by infection that adaptive codebook residual signal of current frame is affected by status of previous frame. Finally, Fig. 2(e) shows pitch contour of this environment. When the previous good frame is classified as voiced, new pitch lag which is equal to the previous good frame is replicated.

From the above analysis, we can recognize that degradation effect of single frame erasure in G.723.1 speech coder is very serious and propagates along several successive good frames. Moreover, if two or more frames are erased successively, reconstructed speech is rapidly attenuated and muted finally. As a necessary consequence, we can conclude that conventional G.723.1 error recovery scheme, which is a single side parameter repetition-based method, has a limitation on the recovered speech quality in a practical erroneous situation.

Ⅲ. PROPOSED ALGORITHM

3.1 Assumption

Proposed error recovery algorithm starts with following two realistic and reasonable assumptions.

A. Both of the previous and next frame packets are available in frame erasure environment.

B. VoIP system is equipped with a general purpose microprocessor possessing sufficient memory space available.

The first assumption is derived from the fact that at least one or more future frames stored in the so called jitter buffer is available at the decoder in practical VoIP systems. Such buffer is introduced to eliminate the harmful effects of delay jitter, packet reordering. However, because making jitter buffer deeper will increase a delay, selection of an adequate sized buffer is required to enhance overall VoIP communication quality. The 30~100ms jitter buffer length is typically used to absorb the effect of delay jitter in real VoIP system. In this environment, we can obtain future information and use it together with the previous frame to recover missed current frame.

The second assumption is based on the characteristics of general implementation architecture of VoIP systems, generally comprised network interface, microprocessor, Digital Signal Processor (DSP), A/D, D/A converter, and other analog/digital logic components. Because speech coder algorithm like G.723.1 consists of massive math functions, DSP is generally used as coprocessor and microprocessor performs overall system control functions such as VoIP signaling, DSP control, system management and other device control and so on. While DSP does not have sufficient memory space to execute real-time speech coder application on internal memory, the processor has large external execution memory space, at least several mega bytes, to control entire system.

3.2 Key Idea

The key ideas of the proposed algorithm with the actually acceptable assumptions described above are illustrated Fig. 3. We prepare the VCIIT in advance, and generate erased frame packet by simply referring the tables using previous and next information. These series of operations are carried out on a microprocessor without modifying speech decoder. Fig. 3 illustrates the situation where frame 3 is missed, and the microprocessor restores that erased frame from both side's good frame information available in the jitter buffer. Because microprocessor requires merely simple operations such as packet parsing and table referencing, it involves little additional complexity in computation.

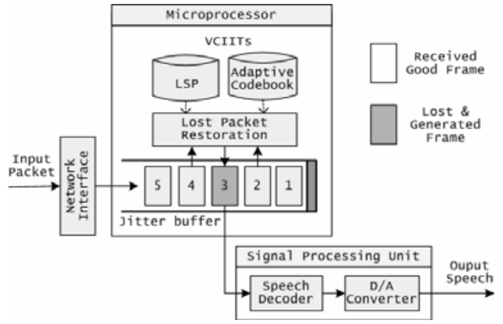

Fig. 3. Block diagram of the proposed loss recovery scheme

And necessary memory space to store VCIIT is negligible compared with the memory size of a microprocessor.

To realize the suggested ideas as described above, we should prepare two kinds of VCIIT in advance. Residual LSP vectors and adaptive codebook gain vectors are quantized with VQ codebook tables and applied following different algorithms to generate the VCIIT.

3.3 LSP VCIIT Generation

In the decodere, the residual LSP vector($\tilde{e}_n$) shown in (1) is generated by referring received LSP codebook indices to VQ codebook tables which consists of sub-vectors divided with dimension 3, 3, and 4 respectively. Each *m*th sub-vector is a vector quantized using an 8 bit VQ codebook. The *l*th entry of the *m*th split residual LSP codebook($\tilde{e}_n$) is expressed with follow:

$$\tilde{e}_{l,m} = [\tilde{e}_{1,l,m}\ \tilde{e}_{2,l,m} \cdots \tilde{e}_{K,l,m}],\ \begin{matrix} 0 \le m \le 2 \\ 1 \le l \le 256 \end{matrix},\ K = \begin{cases} 3, m=0 \\ 3, m=1 \\ 4, m=2 \end{cases} \qquad (4)$$

To simply illustrate the generation algorithm of the LSP VCIIT, we will look at the example for *m*=2. In this case, an element in the ith row and jth column of the VCIIT is determined by finding the optimal index *k* which minimizes the following error criterion $E_{k,i,j}$ defined as (7).

$$\tilde{e}_l = [\tilde{e}_{1,i}\ \ \tilde{e}_{2,i},\ \ \tilde{e}_{2,i},\ \ \tilde{e}_{2,i}],\ 1 \le i \le 256 \qquad (5)$$

$$r_{i,j} = \left(\frac{\tilde{e}_i + \tilde{e}_j}{2}\right) \qquad (6)$$

$$E_{i,j,k} = (r_{i,j} - \tilde{e}_k)\, W_{i,j} (r_{i,j} - \tilde{e}_k)^T \qquad (7)$$

where $r_{i,j}$ is a linear interpolated residual LSP vector and $W_{i,j}$ is a weighting factor for generating LSP VCIIT. The initial value of $W_{i,j}$ is 1 and if $(r_{q,i,j} - r_{(q-1),i,j}), 2 \le q \le 4$ value is larger than weighting constant wc, $W_{q,i,j}$ and $W_{(q-1),i,j}$ are multiplied by wc and if the value is less than or equal to wc, $W_{q,i,j}$and $W_{(q-1),i,j}$ are divided by wc. According to the simulation result with various values, it is concluded that the best optimal value for the weighting constant wc is 5. The generation scheme of another two VCIIT are the same except only that the range of q is different.

The weighting factor $W_{i,j}$ plays an important role in accurate estimation of interpolated codebook index based on inherent characteristics of LSP frequencies. Fig. 4 shows the effect of LSP spectrum changes according to LSP frequency movement to illustrate the effect of weighting factor. LSP frequency has the property which is gathered together around frequency hill and scattered around the valley as shown in Fig 4. Because human hearing organ is more sensitive in frequency hills than in valleys, consequently, to predict the overall interpolated LSP frequency more precisely, prediction of adjacent frequency that become narrow (in other words, $r_{q,i,j} - r_{(q-1),i,j}$ is positive) is more important.

In case of applying the interpolation tables which are made by the algorithm described above, the spectrum restoration performance of frequency

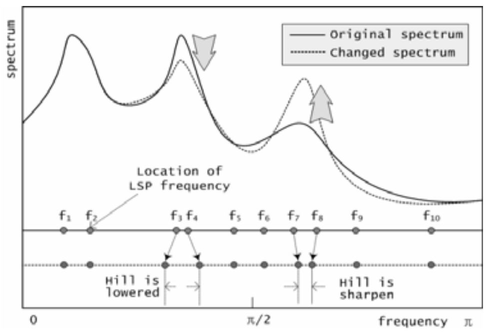

Fig. 4. The effect of LSP frequency movement

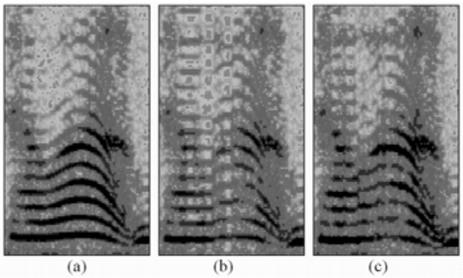

Fig. 5. The spectrogram showing performance of suggested LSP interpolation scheme
(a) decoded speech with clean channel
(b) decoded speech using G.723.1 error recovery scheme with three alternative errors
(c) decoded speech using proposed error recovery scheme with the same impairments

spectrum is shown in Fig. 5. The spectrogram of conventional scheme Fig. 5(b) has many discontinuity points and shows large amount of loss of frequency information. On the other hand, proposed scheme Fig. 5(c) can restore much more features of the original spectrum than the conventional case.

3.4 Adaptive Codebook Gain VCIIT Generation

Another two VQ codebook tables are used in adaptive codebook gain quantization. For the high rate coder, in the case that pitch lag for the subframe is smaller than 58, a gain vector codebook with 85 entries is used. The other case involving low rate coder, the index is one of 170 entries composing a gain vector codebook. And these two codebook tables identically have 20 dimension vectors. Although the encoder uses 20 dimensional gain vectors, decoder uses only the first five vectors ($\beta_{i,j}$) to decode vector gain as shown in (8) and we also use first five elements to make interpolation table.

$$u(n) = \sum_{j=0}^{j=4} \beta_{i,j} e'[n+j], 0 \leqq n \leqq 59 \quad (8)$$

where

$u(n)$ adaptive codebook excitation vector;

$\beta_{i,j}$ pitch predictor gain vector;

$e'[n]$ decoded combined excitation vector considered with current pitch lag of the subframe

In case of the 170 entry codebook table, an element in the ith row and jth column of the VCIIT is determined by finding the optimal index k which minimizes the following energy difference ($gE_{k,i,j}$) defined by the equation (11). The different point with LSP interpolation is that energy difference is used as a decision criterion because the vector is related with signal gain. The other 85 entry codebook is applied with the same criteria.

$$\boldsymbol{gp}_i = [gp_{1,i}\ gp_{2,i}\ gp_{3,i}\ gp_{4,i}\ gp_{5,i}], 1 \leqq i \leqq 170 \quad (9)$$

$$\boldsymbol{gr}_{i,j} = \left(\frac{\boldsymbol{gp}_i + \boldsymbol{gp}_j}{2}\right), 1 \leqq i,j \leqq 170 \quad (10)$$

$$gE_{k,i,j} = \sum_{1}^{5} [(\boldsymbol{gr}_{i,j})^T(\boldsymbol{gr}_{i,j}) - (\boldsymbol{gp}_k)^T(\boldsymbol{gp}_k)], \quad 1 \leqq i,j \leqq 170 \quad (11)$$

3.5 Memory Usage

The interpolation tables which are created with above method are consists of three 256 by 256 matrices for LSP, two 170 by 170 and 85 by 85 matrices for adaptive codebook gain. Owing to the fact that these tables are symmetric, microprocessor will consume about 116Kbyte memory space to store the VCIIT. While this amount of memory space is a burden to DSP, it does not cause any memory deficiency to the microprocessor because it possesses several mega bytes memory available.

3.6 Regeneration Procedure

Now we explain about regeneration process of erased packet in microprocessor side with the precomputed interpolation tables. If one frame is missed and the next frame is stored in jitter buffer, we execute the following sequence to restore the missed packet. Summary of parameter specific interpolation method is shown in Table 1.

A. Extract G.723.1 parameters of both side packets corresponding to Table 1.

B. LSP codebook indices, vector quantized parameters of a missed packet, are predicted by finding the content of the VCIIT corresponding to the row and column.

Table 1. Parameter Regeneration Method

Parameters	Data type	Quantization type	Applied Method
LSP	Indices	VQ	Referring LSP the VCIIT
Adaptive cb. lag	Value	-	Linearly interpolating with both side values
Adaptive cb. gain	Indices	VQ	Referring adaptive codebook the VCIIT
Fixed cv. gain	Indices	Log SQ	Linearly interpolating with both side indices
Fixed cb. others	Different for data type		Replicating the previous good frame

C. The last value of the previous frame and the first value of the next frame are used to linearly interpolate adaptive codebook lag because the values exist twice for every frame. After the interpolation, interpolated value continues during missed packet.

D. As is the LSP case, adaptive codebook gain index is obtained by using the VCIIT. However, in case of high rate coder using two types of VQ tables, if tables of both sides are not equal, adaptive codebook interpolation can not be performed and indices of previous frame is replicated.

E. Fixed codebook gain coefficient is quantized by a logarithmic scalar quantizer and we use linear interpolation method of indices for every subframes. The remainder resulting from integer division operation effects overall gain attenuation.

F. Other fixed codebook parameters such as pulse positions, signs, and grid index are replicated with previous good frame parameter using the characteristics of speech signal which is stationary for short range.

G. Predicted indices and values gathered by the above processes are packetized and requested to decode in DSP.

Ⅳ. Performance Evaluation

So far, we explained the proposed VQ codebook index interpolation method. In this section, we will introduce simulation environments which have been executed and evaluate the performance of the proposed method using the simulation results.

The proposed scheme is implemented in standalone with fixed-point C source code with UNIX environment separated with G.723.1 decoder and consume 116 Kbytes and some running memory spaces. The input to this program is bit-stream output from the encoder with erroneous environment and generates new corrected bit-stream.

4.1 Performance Analysis

Fig. 6 shows an example and analysis of speech quality under the identical situations as Fig. 2. Compared to conventional recovered speech signal, a large amount of spectral distortion, energy attenuation and consecutive degradation effect are now reduced by applying proposed method. Especially, energy contour shows good performance in tracking the original energy contour. This beneficial tracking effect should intuitively produce better concealment performance than simple G.723.1 coder-based scheme.

4.2 Computational Complexity

With the reconstructed speech quality enhancement, it is confirmed that this scheme nearly does not increase the computational complexity in microprocessor. Although complexity comparison between G.723.1 and the proposed algorithm is inconsistent, because proposed scheme is executed not in DSP but in microprocessor, we can guess the complexity dimension of new schemethrough comparing with speech coder. The computational complexity check is performed in Sun Blade 100 workstation with single CPU,2 Gbytes main memory and GNU compiler using the default optimization option. Table 2 shows that additional complexity increment of the suggested novel method is negligible(under 0.1%).

4.3 Simulation Environment

To verify and evaluate speech quality enhancement effect statistically, we performed the simulation under the following environments:

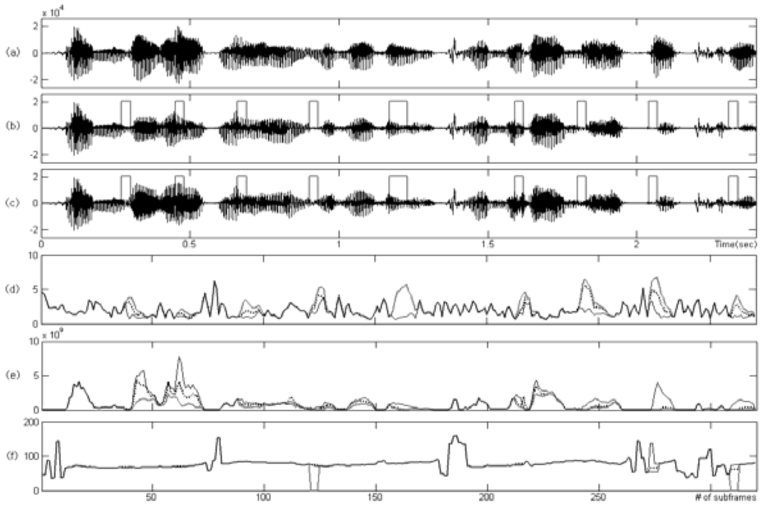

Fig. 6. Example and analysis of speech quality degradation and correction effect due to frame erasure (10% erasure rate).
(a) decoded signal in a clean environment
(b) decoded signal and error locations marked with rectangles in an erroneous environment
(c) decoded signal with proposed scheme and erroneous locations marked with rectangles in a loss environment
(d) spectral distortion contours in a clean(black line), a simple erroneous(gray line), and a recovered erroneous(dotted line) environment
(e) energy contours in clean (black line), a simple erroneous (gray line), and a recovered erroneous (dotted line) environment
(f) pitch contours in a clean (black line), a simple erroneous (gray line), and a recovered erroneous (dotted line) environment

Table 2. Comparison of computational complexity

Coder rate		5.3Kbps	6.3Kbps
G.723.1	Encoder	36.80us	47.76us
	Decoder	3.18us	3.20us
	Total	39.98us	50.96us
Proposed methods		2.09ns	1.64us
Complexity comparison ratio		0.05%	0.03%

A. Coder - ITU-T G.723.1 6.3Kbps MP-MLQ and 5.3Kbps ACELP coder which is widely used in VoIP is applied. The coder has its own error concealment mechanism as described in Section 2.

B. Performance comparison built-in receiver-based error recovery scheme of the coder and proposed VQ codebook index interpolation method are compared.

C. Packet loss model the 2nd order Gilbert model which approximates most closely to the real Internet erroneous environment [7] with each loss percentage step is used.

D. Speech quality measurement - to measure reconstructed speech quality, subjective Mean Opinion Score (MOS) is generally used. However, to reduce evaluation time and cost, alternate objective speech quality measure method is chosen. In this simulation, we use ITU-T Rec. P.862 Perceptual Evaluation of Speech Quality (PESQ) [8] algorithm whose correlation with MOS is over 90 %.

E. Test vectors - speech file database recorded with real communication environment is used. The database is comprised of Korean 32 male and 28 female speakers sampled with 16 bit, mono, 8Khz sampling rate and lasted during about 3~7 seconds. The contents are radio traffic information, weather forecasting, news and cell phone conversation etc. The test vectors have noisy characteristics to reflect real communication environment.

F. Computer simulation is performed in Sun Blade 100 workstation with single CPU, 2 Gbytes main memory and GNU compiler using default optimization option.

G. Iteration - to reduce random nature of the simulation result, we performed simulation iteratively 100 times and it takes about two weeks.

4.4 Performance Evaluation

Under the described simulation environments, we obtain comparison results as shown in Fig. 7,

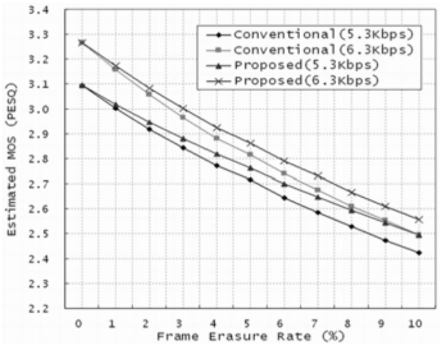

Fig. 7. Estimated MOS vs. packet erasure rate for the conventional and the proposed scheme in G.723.1 speech coder.

Table 3. Comparison of estimated MOS values

FER(%)	Conventional		Proposed	
	5.3Kbps	6.3Kbps	5.3Kbps	6.3Kbps
0	3.094	3.267	3.094	3.267
3	2.843	2.966	2.880	3.001
5	2.715	2.814	2.764	2.862
7	2.584	2.675	2.646	2.731
10	2.424	2.495	2.495	2.556

which tells us the decoded speech quality degradation effect on each case corresponding to each frame erasure rate. It shows that both two coders have tendency of rapid degradation in reconstructed speech quality according to the growth of erasure rate. However, the proposed method performs an important improvement role to mitigate the degradation rate at each coder. Even speech quality of low rate coder with proposed scheme is almost same with conventional high rate coder in case of 10% frame erasure rate. To provide specific value, Table 3 indicates estimated MOS values at typical Frame Erasure Rate (FER).

The final simulation results, Fig. 8 shows the amount of enhancement effect for each of the error rates. From the figure, we can observe following two phenomena: one is that enhancement effect becomes saturated as error rate increases. The other is that applying the proposed scheme to a low rate coder has a better enhancement effect than to high rate. The former can be explained with the fact that the probability of continuous loss becomes higher as the loss ratio increases and the suggested scheme can not meet against the situation. Because the high rate coder uses two kinds of adaptive gain VQ codebook tables,

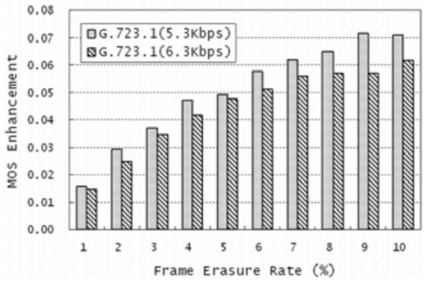

Fig. 8. Speech quality enhancement effect of each coder for each frame erasure rate.

it sometimes occur an unhappy case in which the codebook tables of both sides do not match for the latter case.

Putting all results together, we can obtain a large amount of speech quality enhancement at the expense of negligible additional complexity and memory space increment. Even more, the complexity and the memory problem can be ignored in practical VoIP systems because proposed technique is executed on the microprocessor.

Ⅴ. CONCLUSIONS

In this paper, we suggested a new receiver-based frame erasure recovery scheme, i.e., the VQ codebook index interpolation method, for erroneous VoIP applications. This novel scheme uses both previous and next information and generates erased frame by simply referring the precomputed VCIIT in microprocessor-side without modifying the speech coder. The simulation results demonstrate that the proposed scheme improves a large amount of reconstructed speech quality without additional bandwidth and delay and with negligible increment of computational complexity and memory space. Because this scheme has excellent merits such as enough enhancement effect of speech quality and easy applicability to real system, it can be applied to other multi-media communication applications built in various data networks.

REFERENCES

[1] B. Duysburgh, S. Vanhastel, B. De Vreese, C. Petrisor, P. Demeester, "On the influence

of best-effort network conditions on the perceived speech quality of VoIP connections," *Proceedings of IEEE International Conference on Computer Communications and Networks*, pp. 334-339, 2001

[2] Hardman et al., "Reliable audio for use over the Internet," *Proceedings on INET'95*, 1995.

[3] C. Perkins, O. Hodson, V. Hardman, "A Survey of packet loss recovery techniques for Streaming Audio," *IEEE Network*, Sep./Oct., 1998.

[4] B.W. Wah, Lin Dong, "A Survey of error-concealment schemes for real-time audio and video transmissions over the Internet," *Proceedings of IEEE International Symposium on Multimedia Software Engineering*, p. 1724, Dec, 2000

[5] "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kbit/s," *ITU-T Recommendation* G.723.1, 1996.

[6] W. B. Kleijn and K. K. Paliwal, "*Speech coding and synthesis*," Elsvier, Amsterdam, 1995.

[7] Dan Rubenstein, Jim Kurose, Don Towsley, "Real-time reliable multicast using proactive forward error correction," *UMASS CMPSCI Technical Report* p. 9819, 1998.

[8] "Perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs," *ITU-T Recommendation P.862*, Jan., 2002

[9] Jin-soo. Hwang, S. Choi, S. Ahn, Y. Cho, "Interleaving FEC Mechanism for Packet Loss Recovery of the Internet Phone Service," *Journal of the Korea Institute of Communication Sciences*, pp.1869~1878, Vol.25, No.12A, 2000

Jeongseok Lim **_Regular member_**

Feb. 1987 B.S. degree of Electronic Communication Engineering from Hanyang University

Feb. 1989 M.S. degree of Electronic Communication Engineering from Hanyang University

Feb. 2002 Completion of Ph.D. degree of Electronics and Computer Engineering

1989~2000 Agency for Defense Development

2000~current Senior Researcher of ETRI

<Research interests> Error correction code, Digital communications, Digital signal processing

Haeyong Yang **_Regular member_**

Feb. 1996 B.S. degree of Electrical Engineering from Pusan National University

Feb. 1998 M.S. degree of Electrical Engineering from Pusan National University

1998~2001 Hyundai Electronics

2001~current Senior Researcher of ETRI

<Research interests> Digital signal processing, VoIP.

Kyunghoon Lee **_Regular member_**

Feb. 1992 B.S. degree of Electronics Engineering from Korea University

Aug. 1994 M.S. degree of Electronics Engineering from Korea University

Aug. 1998 Ph.D. degree of Electronics Engineering from Korea University

1998~2000 Senior Researcher of ETRI

2000~2001 Senior Researcher of Tellion

2001~current Senior Researcher of ETRI

<Research interests> DSP, Image and signal processing, DRM.

Sang Kyu Park **_Regular member_**

Feb. 1974 B.S. degree of Electrical Engineering from Seoul National University

May 1980 M.S. degree of Electrical Engineering from Duke University

Jan 1987 Ph.D. degree of Electrical Engineering from the University of Michigan

1976~1978 Agency for Defense Development

Aug. 1990~Aug. 1991 Visiting scholar at University of Southern California

Mar. 1987~current Professor at Hanyang University

<Research interests> Digital communications, Spread spectrum Communications, Coding theory, PCS, IMT-2000