VoIP 망에서의 프레임손실은닉을 위한 비선형 회귀분석 기법

A Nonlinear Regression Analysis Method for Frame Erasure Concealment in VoIP Networks

최승호*, 성호상**

Seung-Ho Choi, Ho-Sang Sung

요 약 프레임 손실은 VoIP 망에서의 음질 저하의 주요 원인이다. 본 논문에서는 VoIP 망에서 주로 사용되는 CELP 기반 음성부호화기의 음질 저하를 최소화하기 위해 비선형 회귀분석 기반의 프레임손실은닉 알고리즘을 제안한다. 제안된 기법은 ITU-T G.729 표준 코덱에 적용되었으며, 기존 방법들에 비해 향상된 PESQ 성능을 보였다.

Abstract Frame erasure is one of the most difficult problems in voice over IP (VoIP) networks and is a major source of speech quality degradation. In this paper, a frame erasure concealment algorithm based on nonlinear regression analysis is presented to minimize speech quality deterioration in code-excited linear prediction (CELP) based coders. We applied the proposed scheme to the ITU-T G.729 standard and obtained improved perceptual evaluation of speech quality (PESQ) scores compared to the conventional methods.

Key Words : frame erasure concealment, nonlinear regression analysis, CELP, G.729, VoIP

I. Introduction

Frame erasure is one of the most difficult problems in voice over IP (VoIP) networks and is a major source of speech quality degradation. The lost frames have to be regenerated at the receiver using frame erasure concealment (FEC) techniques. Most VoIP systems are based on RTP/UDP/IP, which does not have a QoS control mechanism [1]. Also, code-excited linear prediction (CELP) based speech coders are known to be sensitive to both bit errors and frame erasures [2]. The FEC algorithms can be classified into the sender-based and receiver-based algorithms with regard to the place where the concealment algorithm works. The receiver-based algorithms include the repetition based forward concealment [3] and the interpolative concealment [4]. The concealment method used in the ITU-T G.729 standard speech coder that is widely used for VoIP belongs to the forward concealment [3].

Recently, a linear regression-based gain estimation method was proposed for forward FEC in CELP-type speech coders [5]. In this paper we propose a nonlinear regression analysis method for receiver-based FEC in order to more accurately estimate the lost frames. We applied the proposed method to the G.729 coder and the performance of the proposed method was evaluated under random and burst frame erasure conditions.

The remainder of this paper is organized as follows. The proposed nonlinear regression analysis techniques are described in section II. The experimental results are shown in section III. Finally, we conclude in section IV.

*준회원, 서울산업대학교 전자정보공학과
**준회원, 삼성전자 종합기술원
접수일자 2009.9.7, 수정일자 2009.10.8

Ⅱ. Frame erasure concealment by nonlinear regression analysis

This section covers the details regarding the conventional and the proposed frame erasure concealment methods.

1. FEC of G.729

In the ITU-T G.729 standard decoder, first the parameters are decoded (spectral information, adaptive-codebook vector, fixed-codebook vector and gains). The concealment strategy has to reconstruct the current lost frames, based on the previously received information. The method repeats the spectral information and replaces the missing excitation signal with one of similar characteristics, while gradually decaying its energy. The fixed-codebook gain is based on an attenuated version of the previous fixed-codebook gain and is given by

$$g_c^{(m)} = 0.98 \, g_c^{(m-1)} \tag{1}$$

where m is the frame index. The adaptive-codebook gain is based on an attenuated version of the previous adaptive codebook gain and is given by

$$g_p^m = 0.9 \, g_p^{(m-1)}, \ g_p^{(m)} < 0.9. \tag{2}$$
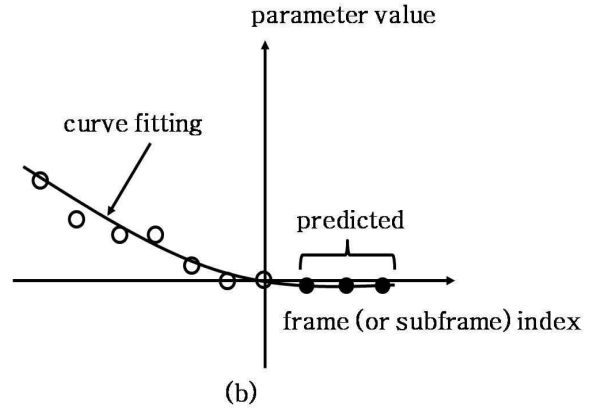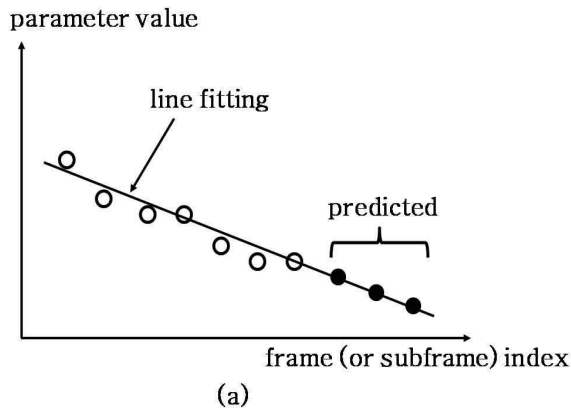
parameter value

(a)

Fig. 1. Prediction of lost frames using (a) linear and (b) nonlinear regression analysis techniques.

2. Linear regression analysis

The FEC using linear regression analysis originates from the idea that the parameters of lost frames could be more similarly recovered if the codec parameters of several preceding good frames are utilized for predicting those of lost frames. The linear regression equation is the form of (3) and the coefficients β_0 and β_1 can be determined using the ordinary least squares technique of (4) and (5).

$$y = \beta_0 + \beta_1 x \tag{3}$$

$$\beta_0 = \overline{y} - \beta_1 \overline{x} \tag{4}$$

$$\beta_1 = \frac{\sum_i x_i y_i - \overline{y} \sum_i x_i}{\sum_i x_i^2 - \overline{x} \sum_i x_i} \tag{5}$$

In (4) and (5), $\overline{x} = \frac{1}{M} \sum_i x_i$ and $\overline{y} = \frac{1}{M} \sum_i y_i$, where M is the number of the preceding good frames. Fig. 1 (a) demonstrates the basic idea of the FEC using the linear regression analysis method.

3. Nonlinear regression analysis

Instead of (3), nonlinear equations such as power,

exponential, and polynomial can be adopted to implement the proposed nonlinear regression analysis method. In this research work, we used the power equation in (6). If we remove the offset α as in Fig. 1(b), the equation can be transformed into a linear equation in (7) by taking the logarithm. Then we can determine the coefficients simply using linear regression analysis for $\log x$ and $\log y$ pairs.

$$y = \alpha + \beta_0 x^{\beta_1} \qquad (6)$$
$$\log y = \log \beta_0 + \beta_1 \log x \qquad (7)$$

III. Experimental results

To evaluate the performance of the FEC methods, we prepared 8 Korean sentences spoken by 8 speakers (4 males and 4 females) from the NTT-AT database [6]. Each sentence was 8 seconds long and was sampled at 16 kHz, followed by down-sampled to 8 kHz using the ITU-T G.191 software tool [7]. ITU-T P.862 (PESQ) was used as an objective speech quality measure for each FEC algorithm, which is known to be highly correlated with the perceptual quality measured by mean-opinion score (MOS) [8]. Fig. 2 shows an example of the waveforms reconstructed by the FEC method of G.729, linear regression, and nonlinear regression when the burst length is 3. We can see the waveform obtained by the nonlinear regression analysis is the most similar to the original waveform compared to the other methods.

The experimental results are given in Table 1 when frame erasure rate (FER) is 3% and the number of preceding good frames is 7. We can notice that the PESQ scores of the proposed methods are much higher than the conventional ones, especially when burst errors occur.
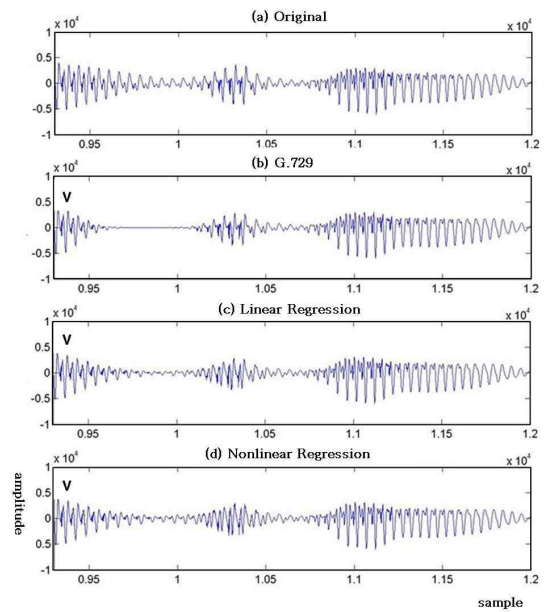

Fig. 2. Waveform comparison example of (a) original speech signal and reconstructed speech signals using the frame erasure concealment methods of (b) G.729, (c) Linear regression, and (d) Nonlinear regression.

Table 1. PESQ scores for each method at FER 3%

Method	Burst length		
	1	3	6
ITU-T G.729	3.462	3.290	2.982
Linear regression	3.483	3.394	3.079
Nonlinear regression	3.489	3.429	3.129

IV. Conclusion

In this paper, we proposed a frame erasure concealment algorithm based on nonlinear regression analysis for a CELP-type speech coder in VoIP networks. Performance evaluation was conducted based on G.729 standard speech coder under random and burst frame erasure conditions. From the PESQ measure and waveform comparison, the proposed frame erasure concealment algorithm showed much improved speech quality than the conventional techniques, especially under burst loss conditions.

References

[1] O. Hersent, D. Gurle, and J.-P. Petit, IP telephony: Packet-based multimedia communications systems, Addison Wesley, 2000.

[2] H. Sanneck and N. T. L. Le, "Speech property-based FEC for Internet telephony applications," Proc. SPIE/ACM SIGMM Multimedia Computing and Networking Conference 2000 (MMCN 2000), San Jose, CA, Jan. 2000, pp. 38-51.

[3] ITU-T Recommendation G.729, Coding of speech at 8 kbit/s using conjugate-structure algebraic-code-excited linear-prediction (CS-ACELP), ITU, Mar. 1996.

[4] J. C. de Martin, T. Unno, and V. Viswanathan, "Improved frame erasure concealment for CELP-based coders," Proc. ICASSP, Istanbul, Turkey, vol. 3, June 2000, pp. 1483-1486.

[5] C. S. Cho, N. I. Park, and H. K. Kim, "A packet loss concealment algorithm robust to burst packet loss for CELP-type speech coders," International Technical Conference on Circuits/Systems, Computers and Communications (ITC-CSCC 2008), pp. 941-944, 2008.

[6] NTT-AT, Multi-lingual speech database for telephonometry, 1994.

[7] ITU-T Recommendation G.191, Software tools for speech and audio coding standardization, ITU, Nov. 2000.

[8] ITU-T Recommendation P.862, perceptual evaluation of speech quality (PESQ), an objective method for end-to-end speech quality assessment of narrowband telephone networks and speech codecs, ITU, Feb. 2001.
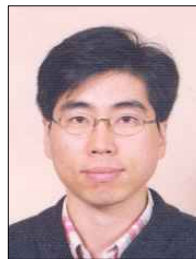
저자 소개

Seung Ho Choi(준회원)

- 1991 한양대학교 전자공학과 학사
- 1993 KAIST 전기및전자공학과 석사
- 1999 KAIST 전기및전자공학과 박사
- 1996-2002 삼성종합기술원 전문연구원
- 2008-2009 University of Florida, Visiting Scholar
- 2002-현재 서울산업대학교 전자정보공학과 부교수

<주관심분야: speech and audio coding, speech recognition, digital communications>

Hosang Sung(준회원)

- 1996 한양대학교 제어계측공학과 학사
- 1999 한양대학교 제어계측공학과 석사
- 1999-2003 한국전자통신연구원 연구원
- 2003- 삼성전자 종합기술원 전문연구원

<주관심분야: speech coding, audio coding, mobile telecommunications, and end-to-end QoE (Quality of Experience) in packet networks>