

# 연속적인 프레임 손실에 강인한 G.729 프레임 손실 은닉 알고리즘

조충상, 이영한, 김홍국  
광주과학기술원 정보통신공학과

## A Packet Loss Concealment Algorithm Robust to Burst Packet Losses for G.729

Choong Sang Cho, Young Han Lee, and Hong Kook Kim  
Department of Information and Communications  
Gwangju Institute of Science and Technology (GIST)  
E-mail: {cscho, cpumaker, hongkook}@gist.ac.kr

### Abstract

In this paper, a packet loss concealment (PLC) algorithm for CELP-type speech coders is proposed to improve the quality of decoded speech under a burst packet loss condition. The proposed algorithm is based on the recovery of voiced excitation using an estimate of the voicing probability and the generation of random excitation by permutating the previously decoded excitation. The voicing probability is estimated from the correlation using the previous correctly decoded excitation and pitch. The proposed algorithm is implemented as a PLC algorithm for G.729 and its performance is compared with PLC employed in G.729 by means of perceptual evaluation of speech quality (PESQ) and an A-B preference test under the random and burst packet losses with rates of 3% and 5%. It is shown that the proposed algorithm provides better speech quality than the PLC of G.729, especially under burst pack losses.

### I. 서론

VoIP나 VoWiFi와 같이 IP 망을 통한 음성전송에 대한 수요가 날로 증대되어 가고 있다. IP 망에서는 packet의 jitter 등에 따른 패킷손실이 발생하며 이는 음질열화를 가져온다 [1][2]. 이러한 환경에서 패킷 손실에 의한 음질열화를 최소화하는 패킷 손실 은닉 (PLC) 방법으로는 송신단에서 프레임 손실을 은닉하는 방법과 수신단에서 프레임 손실을 은닉하는 두 가지 방식으로 나눌 수 있다 [1]. 송신단 기반의 대표적

인 프레임 손실 은닉 방법으로 forward error correction, interleaving, 재전송 방법 등이 있고, 수신단 기반의 손실 은닉 방법으로는 삽입, 보간, 모델 기반의 복원 방법 등이 있다. 송신단 기반의 손실 은닉 방법은 프레임 손실이 발생하였을 경우 프레임 손실을 은닉하기 위한 추가적인 정보가 요구되기 때문에 추가적인 정보를 전송하기 위한 전송비트가 추가적으로 요구되는 단점이 있다. 하지만 높은 프레임 손실률에서도 급격한 음질열화가 발생하지 않는다는 장점이 있다 [1]. 반면에 수신단 기반의 손실 은닉 방법은 전송률이 증가되진 않지만 프레임 손실률이 높아짐에 따라 급격한 음질열화가 발생하는 단점이 있다 [1][3].

본 논문에서는 VoIP 등에 널리 이용되는 CELP 부호화기의 수신단 기반 프레임 손실 은닉 알고리즘을 제안한다. 기존의 방법 중의 하나로, [2]에서는 하나의 프레임의 파라미터가 손실된 경우 손실없이 복원된 바로 이전의 프레임과 이후 프레임의 파라미터를 선형적으로 보간하여 현재 손실된 파라미터를 복원한다. 또 다른 방법으로는 손실된 프레임과 이웃하는 프레임의 경계지점에서 신호의 크기가 급격히 변하여 발생하는 문제와 연속적인 프레임 손실이 발생할 때 상관도가 감소하는 문제를 해결하고자 하는 방법이 있다 [4][5]. 경계지점에서의 문제를 해결하기 위해서는 여기신호를 sine파형으로 표현하며, 손실된 여기신호를 extrapolation 방법으로 복원해 낸다 [4]. 이렇게 복원된 여기신호는 경계지점에서 급격히 변하지 않는 반면, 연속적인 프레임 손실이 발생할 경우에는 복원된 음성과 실제 프레임 손실이 없이 복원된 음성 사이의 상관도가 감소한다. 이러한 문제를 해결하기 위하여 연속하여 손실된 프레임 수가 증가할수록 여기신호의 크기를 감소시키는 것이 일반적이다. 또한 가장 최근

손실없이 복원된 여기신호와 동일한 에너지를 갖는 백색 잡음신호를 생성하여 여기신호에 더해주는 방법을 사용한다.

따라서 본 논문에서는 CELP 음성부호화기에서 연속적인 프레임 손실이 발생하는 환경에서 개선된 수신단 기반 프레임 손실 은닉 방법을 제안하고 이를 G.729의 프레임 손실 은닉 방법으로 구현한다. 그리고 구현된 프레임 손실 은닉 방법의 성능은 G.729에 표준으로 구현되어져 있는 프레임 손실 은닉 방법과 PESQ 및 음질비교를 통해 그 성능을 상호 비교한다.

본 논문의 구성은 다음과 같다. 서론에 이어, 제 II 장에서 G.729 표준 프레임 손실 은닉 방법을 기술하기로 한다. 제 III 장에서는 본 논문에서 제안한 프레임 손실 은닉 방법에 대해 설명하고, 제 IV 장에서 제안된 방법의 성능을 G.729 표준 프레임 손실 은닉 방법과 비교한다. 마지막으로 제 V 장에서 결론을 맺는다.

## II. G.729 표준 프레임 손실 은닉 방법

ITU-T 표준인 G.729 는 코드 여기 선형예측 (code-excited linear prediction: CELP) 기반의 음성부호화기로 8 kHz 샘플링 주파수에서 한 프레임당 10 ms (80 샘플)단위로 음성을 분석하여 8 kbit/s의 전송률로 음성 파라미터를 전송한다. G.729의 프레임 손실 은닉 방법은 수신단 기반의 방법으로 가장 최근에 손실없이 복원된 정보를 바탕으로 프레임 손실 은닉 방법을 수행한다 [6]. 우선, 프레임 손실이 발생한 프레임의 선형예측 계수는 가장 최근에 손실없이 복호화된 프레임의 선형예측 계수를 사용한다. 다음 프레임에 손실없는 프레임이 수신되었을 경우를 대비하여 선형예측 계수를 구하기 위하여 moving average (MA) 계수를 갱신하는 작업을 수행한다.

다음으로는 여기신호를 복원하기 위해 사용되는 고정 코드북 이득과 적응 코드북 이득을 연속적인 프레임 손실에 따라 순차적으로 감쇄시킨다. 여기신호를 복원하기 위해서 가장 최근 손실없이 복원된 프레임이 유성음인지 무성음인지 구분하는 과정을 수행한다. 유성음일 경우에는 이전 프레임의 피치값을 바탕으로 적응 코드북과 적응 코드북 이득을 이용하여 여기신호를 복원하고, 무성음일 경우에는 무작위로 선택된 고정 코드북의 펄스 위치 및 펄스의 부호를 이용하여 여기신호를 복원해 낸다.

## III. 유성음 확률 및 잡음 여기신호 기반의 프레임 손실 은닉 방법

본 논문에서 제안한 프레임 손실 은닉 방법은 손실

된 프레임의 여기신호를 그림 1과 같은 과정으로 복원한다. 손실된 여기신호를 복원하기 위하여 가장 최근에 손실이 없는 프레임에서 복원된 여기신호와 피치값

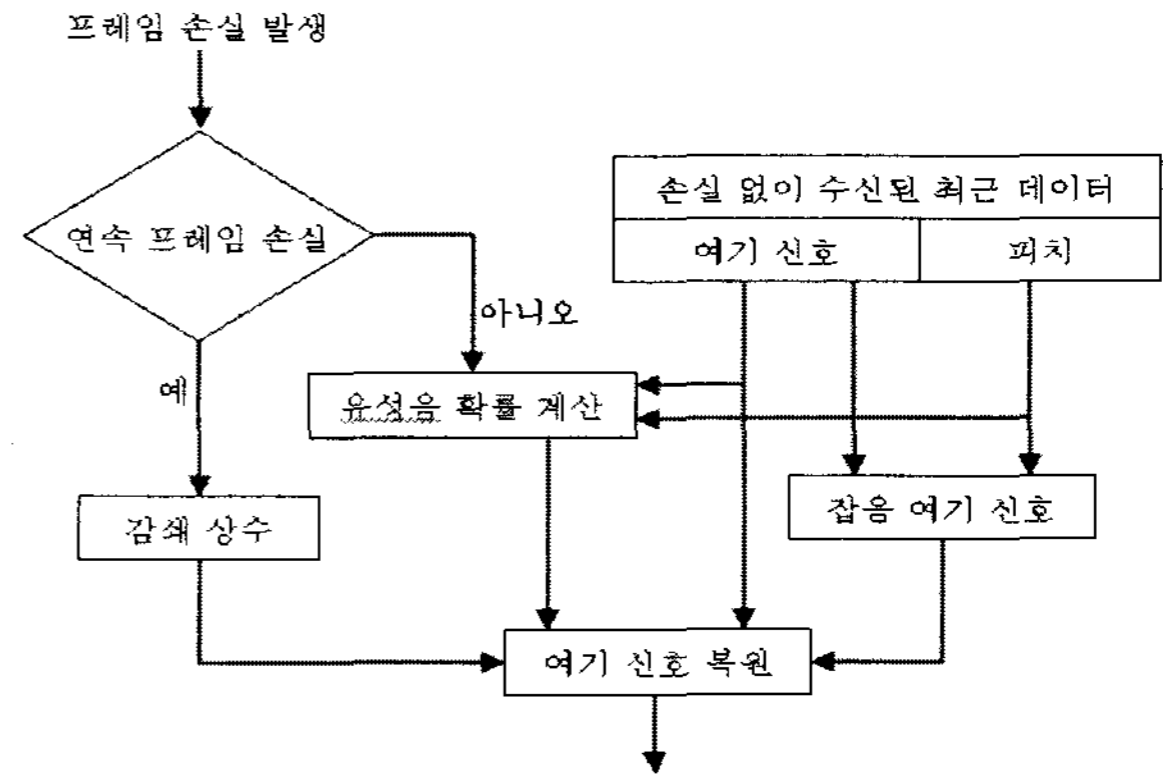


그림 1. 제안된 유성음 확률 기반 여기신호 복원 구조.

으로 부터 피치값을 기준으로 여기신호의 상관계수를 식 (1)과 같이 구한다.

$$r = \frac{|\sum_{i=0}^k x(i)x(i+P)|}{\sqrt{\sum_{i=0}^k x^2(i)}\sqrt{\sum_{i=0}^k x^2(i+P)}} \quad (1)$$

여기서  $x(n)$ 은 가장 최근에 손실없이 복원된 프레임의 여기신호이고,  $P$ 는 피치값,  $r$ 은 상관계수이다. 측정된 상관계수를 바탕으로 식 (2)와 (3)과 같이 신호가 유성음일 확률 (voicing probability)을 구한다.

$$voicing\ factor = \sqrt{r} \quad (2)$$

$$voicing\ probability = \begin{cases} 1, & voicing\ factor > 0.7 \\ \frac{voicing\ factor - 0.3}{0.4}, & 0.3 \leq voicing\ factor \leq 0.7 \\ 0, & voicing\ factor < 0.3 \end{cases} \quad (3)$$

한편, 가장 최근에 손실이 없는 프레임에서 복호화된 여기신호를 무작위로 배치 (permutation)하여 만든 잡음 여기신호와 피치를 반복하여 만든 주기적인 여기

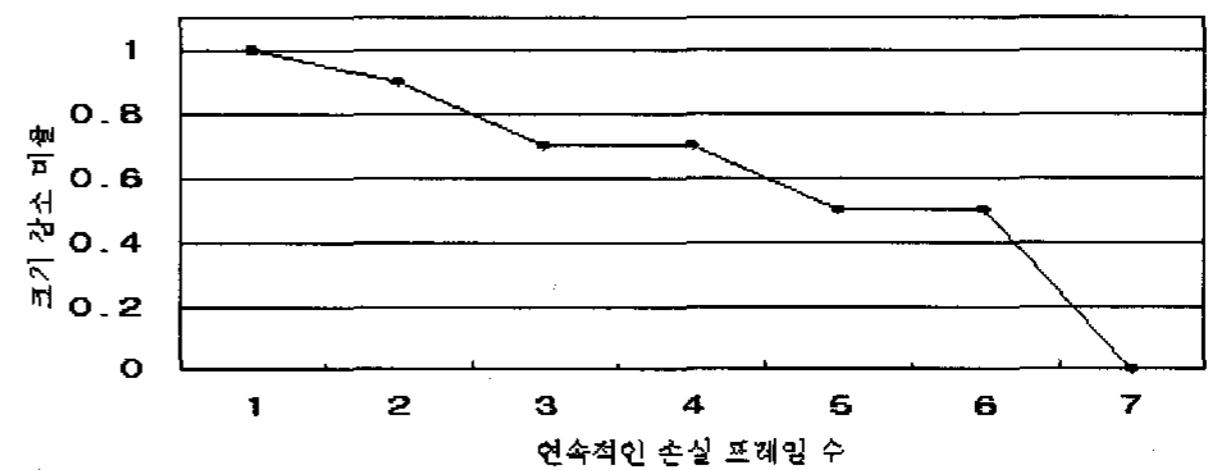


그림 2. 연속적인 프레임 손실에 따른 여기신호 크기 감소 비율.

신호 (본 논문에서는 피치 여기신호라 칭한다.)를 각각 생성한다. 생성된 두 가지의 여기신호 중 피치 여기신호에 대해 유성음일 확률로 가중치를 부여하고, 잡음

여기신호에 대해서는 (1-voicing probability)로 정의된 무성음일 확률로 가중치를 부여한다. 가중치가 적용된 두 여기신호를 합하여 첫 번째 손실된 프레임의 여기신호를 구한다.

연속적으로 프레임 손실이 발생하였을 경우 첫 번째 손실된 프레임에서와 같은 방법으로 바로 이전에 복원한 여기신호와 피치값으로 부터 피치 여기신호와 잡음 여기신호를 만들고, 첫 번째 손실된 프레임에서 구해진 유성음일 확률과 유성음이 아닐 확률을 피치 여기신호와 잡음 여기신호에 가중치로 부여하고 합하여 새로운 여기신호를 구한다. 연속적으로 손실된 프레임 수에 따라 그림 2와 같은 크기 감소 비율을 새롭게 구해진 여기신호에 곱하여 새로운 여기신호를 복원한다.

다음 프레임에 손실이 없을 경우 신호를 복원하기 위하여 프레임 손실 은닉 방법으로 복원한 여기신호 80 샘플로 적응 코드북 생성에 사용되는 메모리를 갱신하는 과정을 수행한다. 또한 고정 코드북 이득은 G.729 표준 프레임 손실 은닉방법과 같은 방법으로 갱신하는 작업을 수행한다.

선형예측계수에 대한 복원은 식 (4)와 같이 가장 최근에 손실없이 복원된 프레임의 선형예측계수의 크기를 감소시켜서 합성 필터링에 적용한다.

$$a_i^{(m)} = 0.99^i a_i^{(m-1)}, i=1, \dots, 10 \quad (4)$$

여기서  $m$  은 프레임 번호이고,  $a_i^{(m)}$  는  $m$  번째 프레임에서  $i$  번째 선형예측계수이다. 선형예측계수의 크기를 감소시키므로서 합성필터의 포먼트 밴드폭이 확장되어 결과적으로 스펙트럼이 smoothing되게 된다 [7]. 연속적인 프레임 손실이 발생할 경우는 이전 프레임에서 크기가 감소된 선형예측계수의 크기를 더욱 감소시켜 필터링에 사용한다. 또한 손실없는 프레임이 입력되었을 경우 선형예측계수를 구하기 위하여 G.729 프레임 손실 은닉 방법과 같이 moving average (MA) 계수 갱신한다.

#### IV. 실험 및 결과

본 논문에서 제안된 프레임 손실 은닉 방법의 성능을 평가하기 위하여 NTT-AT 데이터베이스[8]에 있는 8초 길이의 한국인 남성, 여성음성 각각 48 개씩을 테스트 데이터로 선정하였다. 16 kHz로 저장되어 있는 각각의 음성신호에 modified IRS 필터를 적용한 후 이를 8 kHz로 다운 샘플링하여 G.729의 입력신호로 사용하였다. 프레임 손실 환경을 위해 ITU-T 표준 G.191에 정의된 Gilbert-Elliott 모델을 사용하였다 [9]. 이러한 프레임 손실 모델을 통하여 프레임 손실률이 각각 3%와 5%인 손실패턴을 발생시켰으며 각각의 경

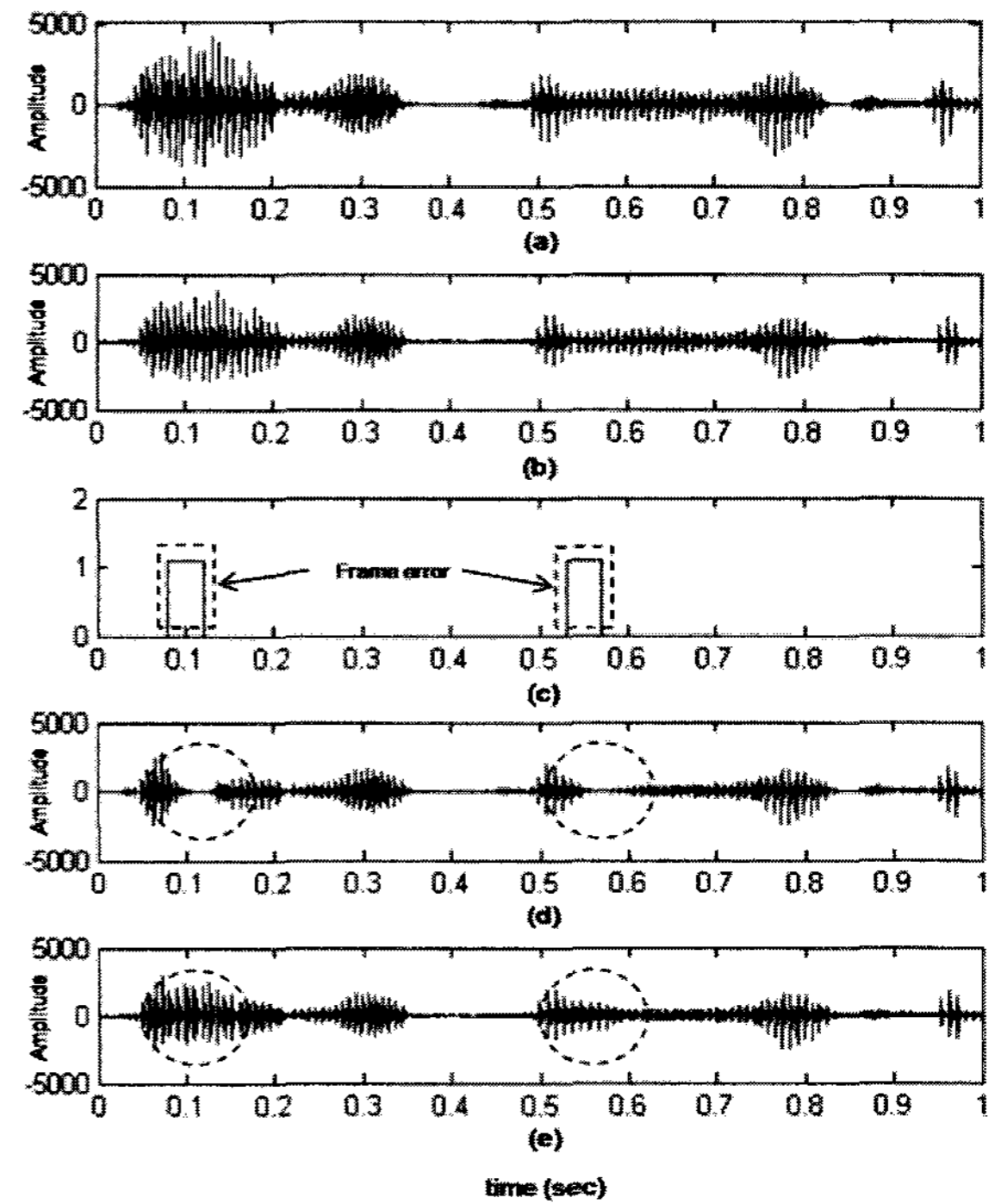


그림 3. 프레임 손실 은닉 알고리즘에 의해 복원된 파형 비교. (a) 원본 음성 파형; (b) G.729 으로 복호화된 파형; (c) 프레임 손실 패턴; (d) G.729 프레임 은닉 방법이 적용된 파형; (e) 제안된 프레임 은닉 방법이 적용된 파형.

우에 대해 연속적으로 손실된 프레임의 수가 각각 2, 3, 4, 5, 6이 되도록 수작업으로 손실패턴을 수정하였다. 성능평가 방법으로 ITU-T에서 제공하는 객관적 음질평가 방법인 PESQ [10]와 주관적 음질평가를 사용하여 G.729에 구현된 프레임 손실 은닉 방법과 제안된 프레임 손실 은닉 방법과의 성능을 비교하였다.

그림 3은 연속적인 프레임 손실에서 제안된 프레임 손실 은닉 방법과 G.729 프레임 손실 은닉 방법으로 복원된 파형을 보여주고 있다. 실험 결과, 그림 3.(a)의 원음을 G.729 부호기로 부호화하여 생성된 비트스트림이 손실없이 복호화되어 그림 3.(b)와 같은 파형으로 복호화되었다. 또한 그림 3.(c)와 같은 연속적인 프레임 손실이 발생하였을 경우 G.729 프레임 손실 은닉 방법에 의해 그림 3.(d)와 같은 파형으로 복원되었고, 제안된 방법에 의해 그림 3.(e)와 같은 파형으로 복원되었다. 그림 3.(d)의 점선 부분과 같이 G.729 프레임 손실 은닉 방법은 연속적인 프레임 손실이 발생하였을 경우 파형의 크기가 급격히 감소하지만 본 논문에서 제안된 프레임 손실 은닉 방법은 그림 3.(e)의 점선 부분과 같이 연속적인 프레임 손실에서도 급격한 손실없이 원음과 유사하게 복원할 수 있음을 볼 수 있다.

G.729 프레임 손실 은닉 방법과 본 논문에서 제안된 프레임 손실 은닉 방법을 PESQ를 통하여 비교하였다.

표 1. G.729 표준 프레임 손실은닉 방법과 제안된 프레임 손실 은닉 방법의 PESQ (MOS) 평가 비교.

연속 손실 프레임 수	3% 손실률		5% 손실률	
	G.729 손실 은닉 방법	제안된 손실 은닉 방법	G.729 손실 은닉 방법	제안된 손실 은닉 방법
1	3.40	3.40	3.25	3.25
2	3.31	3.33	3.12	3.15
3	3.26	3.32	3.04	3.10
4	3.20	3.30	2.93	3.08
5	3.21	3.30	2.94	3.10
6	3.20	3.34	2.96	3.12

표 1은 연속적인 프레임 손실에서 제안된 프레임 손실은닉 방법의 성능을 평가하기 위하여 연속적인 손실 프레임 수가 2, 3, 4, 5, 6경우에 대해 PESQ를 측정된 결과이다. 표 1에서와 같이 burstiness가 0인 경우, 즉 Gilbert-Elliott 모델에서 연속될 확률이 최소일 경우, 프레임 손실률 3%와 5%의 경우에 대해서는 유사한 성능을 보였다. 하지만, 연속적인 프레임 손실의 경우, 즉 Gilbert-Elliott 모델에서 연속될 확률이 최대일 경우, 연속적으로 손실된 프레임 개수에 따라 0.02에서 0.16의 MOS 향상을 보였다.

표 2. G.729 표준 프레임 손실 은닉 방법과 제안된 프레임 손실 은닉 방법의 A-B 선호도(%) 평가 비교.

Burstiness/ 손실율		선호도		
		G.729 손실 은닉 방법	유사함	제안된 손실 은닉 방법
$\gamma = 0$	3%	6	75	19
	5%	3	90	7
$\gamma = 1$	3%	3	39	58
	5%	27	37	37
평균		9.75	60.25	30.25

또한 본 논문에서 제안된 프레임 손실은닉 방법에 대해 주관적 음질평가를 위하여 일반인 6명을 대상으로 선호도를 실험하였다. 표 2에 보이는 바와 같이 Gilbert-Elliott 모델에서  $\gamma$ 가 0경우와 1인 경우에 대해서 실시하였다. 본 논문에서 제안된 프레임 손실 은닉 방법의 선호도가 평균 30.25%이고 G.729 프레임 손실 은닉 방법의 선호도가 9.75%로 본 논문에서 제안된 프레임 손실은닉 방법이 상대적으로 20.5% 높은 선호도를 보였다.

## V. 결론

본 논문에서는 ITU-T 표준 G.729의 프레임 손실 은닉 방법을 기술하였고, G.729에서 연속적인 프레임 손실 환경에서 음질을 향상시키기 위한 수신단 기반 프레임 손실 은닉 방법을 제안하였다. 제안된 프레임 손실 은닉 방법에서는 가장 최근 손실없이 복원된 신

호의 유성음 확률을 구하고, 가장 최근 손실없이 복원된 여기신호를 기반으로 잡음 여기신호와 피치 여기신호를 구하였다. 결과적으로 두 여기신호에 유성음일 확률과 무성음일 확률을 가중치로 부여하여 여기신호를 복원했다. 선형예측계수는 가장 최근 손실없이 복원된 프레임의 선형예측계수의 크기를 감쇄시키는 방법으로 선형예측계수를 복원했다. 본 논문에서 제안한 프레임 손실 은닉 방법은 손실률 3%와 5%에서 연속적인 프레임 손실이 발생하였을 경우 G.729 표준 프레임 손실 은닉 방법보다 PESQ의 MOS와 A-B 선호도 관점으로 더 우수한 성능을 보임을 알 수 있었다.

## 참고문헌

- [1] C. Perkins, O. Hodson, and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," *IEEE Network*, vol. 12, pp. 40-48, Sept./Oct. 1998.
- [2] C. Montminy and T. Aboulnasr, "Improving the performance of ITU-T G.729A for VoIP," in *Proc. IEEE International Conference on Multimedia and Exposition*, pp. 433-436, July 2000.
- [3] D. J. Goodman, G. B. Lockhart, O. J. Wasem, and W. C. Wong, "Waveform substitution techniques for recovering missing speech segments in packet voice communications," *IEEE Trans. Acoust. Speech Signal Process.* vol. 34, no. 6, pp. 1440-1448, Dec. 1986.
- [4] J. Lindblom, et al., "Packet loss concealment based on sinusoidal extrapolation," in *Proc. ICASSP*, pp. 173-176, May 2002.
- [5] A. Husain and V. Cuperman, "Reconstruction of missing packets for CELP-based speech coders," in *Proc. ICASSP*, pp. 241-244, May 1995.
- [6] ITU-T Recommendation G.729, *Coding of speech at 8 kbit/s using conjugate-structure code-excited linear prediction (CS-ACELP)*, Feb. 1996.
- [7] J. Lindblom, et al., "Packet loss concealment based on sinusoidal modeling," *Proc. IEEE speech coding workshop*, pp. 65-67, Oct. 2002.
- [8] NTT-AT, *Multi-lingual speech database for telephony*, 1994.
- [9] ITU-T Recommendation G.191, *Software tools for speech and audio coding standardization*, Nov. 2000.
- [10] 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 coders*, Feb. 2001.