

손실이후 프레임 정보에 의한 패킷손실은닉 알고리즘 개선

김재현(ICU), 한민수(ICU)

<차 례>

- | | |
|------------------------|------------------------------|
| 1. 서 론 | 3.4. 음성신호 감쇠와 OLA |
| 2. G.711 패킷손실은닉 알고리즘 | 4. 손실이후 프레임 정보를 이용한 패킷 손실 보상 |
| 3. 선형예측계수를 이용한 패킷손실 보상 | 5. 알고리즘 평가 및 결과분석 |
| 3.1. 음성신호의 선형예측분석 | 6. 결 론 |
| 3.2. 선형예측필터의 대역폭확장 | |
| 3.3. 잔차신호 생성 | |

<Abstract>

Improvement of Packet Loss Concealment Algorithm by Utilizing Next Good Frame Info.

Jae-Hyun Kim, Min-Soo Hahn

In real time packetized voice application, missing packets are major source of voice quality degradation. Thus packet loss concealment (PLC) algorithms are needed to guarantee QoS of VoIP. In this paper, we describe packet loss concealment scheme utilizing the next good frame which follows loss packets. When this scheme is combined with other PLC algorithms, such as G.711 pitch waveform replication recommended by ITU-T LP based PLC algorithm, additional voice quality improvement is obtained for consecutive packet loss larger than 60 msec.

* 주제어: 인터넷전화(VoIP), 통화품질(QoS), 패킷손실은닉(packet loss concealment(PLC))

1. 서 론

오늘날 인터넷 전화(VoIP: Voice over IP)가 많은 관심의 대상이 되고 있음에도 불구하고 현실적으로 IP(Internet Protocol) 네트워크는 음성과 같은 실시간 매체를 전송하는데 그리 적합하지 않다. 이것은 본질적으로 IP 네트워크의 끊어짐(connectionless) 현상과 이로 인한 낮은 서비스 품질(QoS) 때문이다. IP 네트워크를 통한 실시간 보이스 트래픽 전송을 위해서는 다음과 같은 네 가지 주요 성능 파라미터들을 고려해야 한다; 단-대-단(end-to-end) 지연, 패킷지연 변화량(delay jitter), 패킷손실(packet loss), 순서가 뒤바뀐 패킷전송(out-of-order packet delivery)[1].

이중에서 인터넷 전화의 패킷손실(packet loss)은 음성신호를 나르는 IP 패킷이 제 시간 안에 수신 측에 도착하지 않았을 때 발생한다. 그 이유로 네트워크 상에서 전송 중에 손상되는 경우와, 네트워크 기능장애(network malfunction)나, 혹은 수신 측에 너무 늦게 도착하여 음성 재생에 포함되지 못하고 제외된 경우를 들 수 있다[2].

패킷손실은 하나의 프레임으로 발생할 수도 있고, 연속된 프레임의 블록으로 발생할 수도 있다. 인터넷상에서의 패킷손실은 상호의존적으로 발생한다. 만약 n 번째 보이스 프레임이 손실되었다면 $n+1$ 번째 보이스 프레임이 손실될 확률 또한 상당히 높다[3][4].

패킷손실을 보상하기 위해서 흔히 쓰이는 기술로 FEC(Forward Error Correction)와 패킷손실은닉(PLC: Packet Loss Concealment)을 들 수 있다. FEC는 송신 측과 수신 측에서 소스에 대한 추가적인 처리가 요구되며, 그 효율성은 손실률과 손실 분포에 따라 달라진다. 또한 FEC 알고리즘은 넓은 대역폭과 부가적인 처리에 따른 지연을 필요로 한다[5]. 반면, 패킷손실은닉 알고리즘의 목적은 수신된 비트 스트림에 대해서 네트워크 상에서 사라진 데이터를 대신할 수 있는 합성음성신호를 만들어 내는 것이다. 음성신호는 대개 부분적으로 안정화 되어있기 때문에 이전의 음성신호로부터 손실된 패킷의 음성신호를 추정하는 것이 가능하다[6]. 대체로 손실은닉 알고리즘은 단일 프레임(4 - 40msec)에 대한 낮은 패킷손실률(<15%)에 대해서 효과적으로 적용될 수 있다[7].

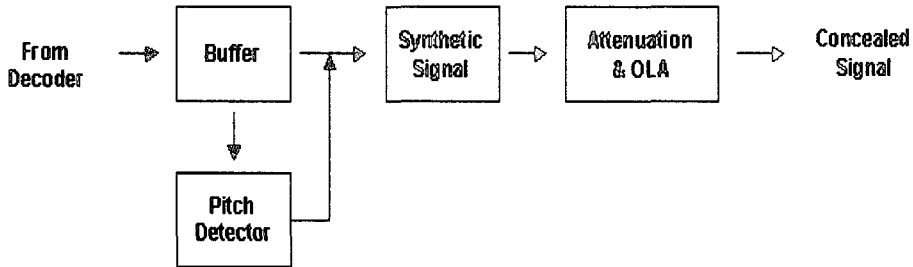
본 논문에서는 G.711 PCM 음성코더표준[8]에 제안된 PLC 알고리즘[6]과 선형예측(LP)계수를 이용한 PLC 알고리즘을 구현하였으며, 패킷손실에 대한 알고리즘의 성능평가를 위해서 주관적 평가지표(MOS)를 비교 하였다.

또한 연속적인 패킷손실에 대해서도 개별적으로 주관적 평가를 실시하여 비교 하였고, 손실구간이 상대적으로 긴 구간(>60msec)에 대해서는 본 논문에서 제안한 손실 후 프레임의 음성정보를 이용한 알고리즘을 적용하였다. 여기서 적용된 PLC 알고리즘은 PCM(Pulse Code Modulation) 음성신호에 대해서 처리되며, 다양한 음성 코딩 기술과 독립적으로 수신 측에서 적용될 수 있다.

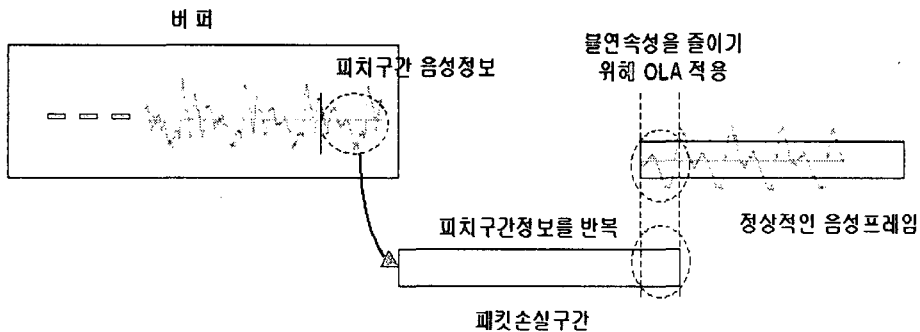
2. G.711 패킷손실은닉 알고리즘

CELP기반 코더를 비롯한 여러 음성코더(G.723.1, G.728, G.729 등)들은 자체적인 패킷손실은닉 알고리즘을 각각의 표준안에 기술하고 있으며, G.711 음성코더에도 PCM 음성 신호에 대한 PLC 알고리즘이 제안되었다[6].

이 알고리즘은 패킷손실이 발생했을 때 히스토리 버퍼에서 그 이전 프레임의 피치정보를 손실된 구간에 반복, 삽입해 줌으로써 손실을 보상한다. 위의 그림 1은 G.711 PLC 알고리즘의 전체적인 블록도를 나타내며 이에 대한 간략한 과정은 다음과 같다.



<그림 1> G.711 PLC 알고리즘 블록도



<그림 2> G.711 PLC 알고리즘 손실구간 보상

1. 정상적으로 음성 패키지가 전송되었을 때는 수신된 패키지를 디코딩하게 된다. 이때, 출력을 오디오 단자로 내보내며 피치검출을 위해서 히스토리 버

퍼에 저장한다.

2. 패킷손실이 발생하게 되면 히스토리 버퍼에 저장된 이전의 음성신호에 대해서 정규화된 교차상관도(normalized cross-correlation)의 피크점을 찾아 피치를 예측한다.
3. 손실된 구간에 대해서 히스토리 버퍼에서 가장 최근의 음성신호의 피치길 이만큼의 신호를 반복해 보상한다.
4. 패킷손실 이후의 음성 프레임과 불연속성을 줄이기 위해서 OLA(Overlap Add)를 적용하고, 연속된 패킷손실에 대해서 10ms당 20%씩 신호를 점차 적으로 감쇠시킨다.

위의 그림 2는 G.711 PLC 알고리즘이 어떻게 손실구간을 보상하는지 간단히 그림으로 나타낸 것이다.

3. 선형예측계수를 이용한 패킷손실 보상

음성신호는 시간축 상에서 상관도가 높기 때문에 일정 부분의 패킷손실에 대해서 손실이전 프레임의 선형예측(LP: Linear Prediction)계수와 잔차신호(residual signal)를 이용하여 보상을 할 수 있다[9][10]. 본 논문에서는 [10]에서 제안한 잔차신호를 이용해 LP계수기반 PLC 알고리즘을 구현하였다.

3.1. 음성신호의 선형예측분석

시간 t에서의 음성 샘플을 x_t 라고 하면, 선형예측분석[11]에 의해 현재의 음성 샘플을 이전 p개의 샘플로부터 예측할 수 있다. 이때, 예측값과 실제값간의 차이를 e_t 라고 하면 식(1)이 성립한다.

(1)

$$x_t = -\sum_{i=1}^p a_i x_{t-i} + e_t$$

LP계수를 이용한 PLC알고리즘은 손실이전에 수신한 음성 프레임의 성도정보(vocal tract info.)와 여기정보(excitation info.)를 바탕으로 패킷손실을 보상한다. LP계수를 이용한 음성합성을 통해서 음소(phoneme)단위가 변하지 않을 정도(보통 4-40msec)의 패킷손실에 대해서 음성품질 향상이 가능하다는 것을 예상할 수 있다.

3.2. 선형예측필터의 대역폭 확장

LP필터의 대역폭을 확장[9]한다는 것은 필터의 피크를 약간씩 줄이고 대역폭을 조금 늘리는 것을 의미한다. 식 (2)를 적용함으로써 이전 프레임으로부터 구해진 LP필터는 조금 더 부드러운 형태가 된다.

(2)

$$a'_i = (0.98)^i a_i \quad i = 1 \dots p$$

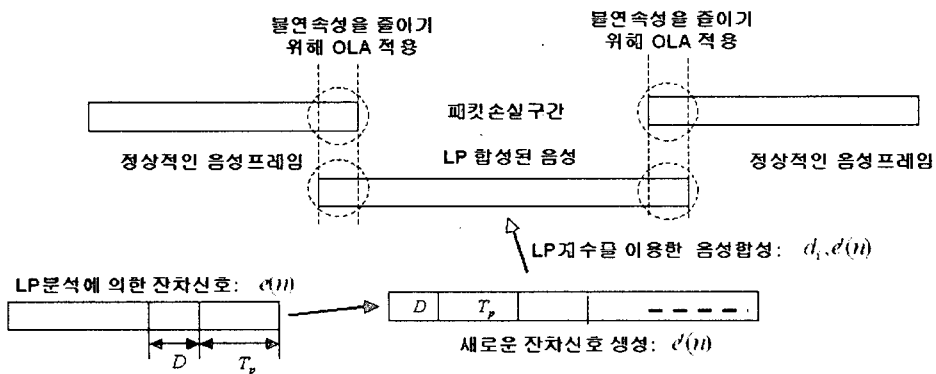
이로 인해 천천히 변화하는 음성신호의 스펙트럼 특성을 반영할 수 있다[12]. 패킷손실이 발생하는 때 프레임마다 LP필터의 대역폭을 확장하여 필터의 스펙트럼을 평활화 시키고, 여기서의 LP계수를 이용한 음성합성으로 보다 부드러운 음성변화구간(transition)을 얻을 수 있다.

3.3. 잔차신호 생성

[10]에서 적용한 잔차신호를 구하기 위해서는 이전프레임의 피치정보가 필요하다. 피치를 구하기 위해 식(3)의 AMDF(Average Magnitude Difference Function)를 사용하였다.

(3)

$$E(\tau) = \frac{1}{N} \sum_{n=0}^{N-1} |x(n) - x(n+\tau)|, \quad \tau = \tau_{\min}, \tau_{\min} + 1, \dots, \tau_{\max}$$



<그림 3> LP계수를 이용한 PLC 알고리즘

여기서 $x(n)$ 은 버퍼의 음성샘플을 나타내며 τ_{min} 과 τ_{mx} 는 각각 최소 피치와 최대 피치를 의미한다. 식(3)을 통해 구한 $E(\tau)$ 을 식(4)에 적용함으로써 해당 구간의 피치 T_p 를 구할 수 있다.

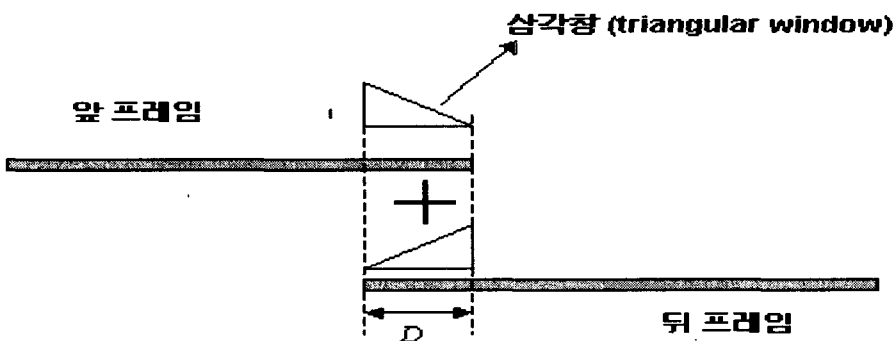
(4)

$$T_p = \arg \underset{\tau}{\text{MIN}} [E(\tau)]$$

앞 절 3.1의 LP분석에서 잔차신호 $e(n)$ 과 피치 T_p 를 이용해 새로운 잔차신호 $e'(n)$ 을 생성한다. 패킷손실 손실구간에 대해서 LP계수를 이용한 음성합성신호는 손실구간 전/후 신호와의 경계에서 불연속성("click")이 발생한다. 이를 완화해 주기 위해서 OLA를 적용하게 되는데. 이때 추가적인 중복구간(D)에 대한 음성신호 생성이 필요하며 위의 그림 3과 같이 나타낼 수 있다.

3.4 음성신호 감쇠와 OLA(Overlap Add)

앞서 잠시 언급한 것처럼 음성손실구간 전후에 대한 불연속성 때문에 OLA를 적용하며, 이에 대한 구체적인 방법은 G.711 PLC알고리즘과 동일하게 중복구간에 대해서 삼각 창(triangular window)을 사용하였다[6].



<그림 4> OLA (Overlap Add)

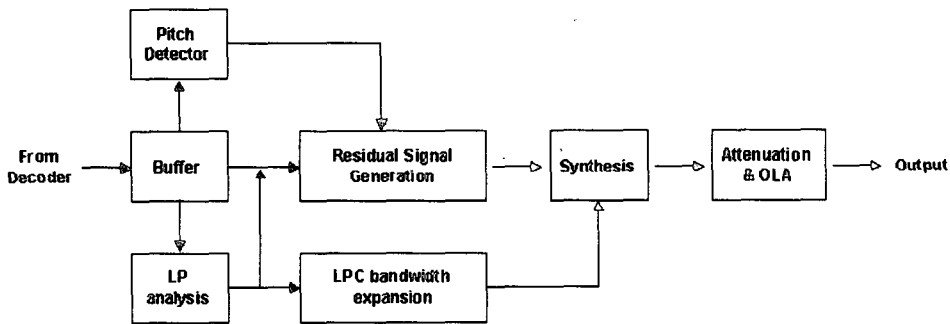
위의 그림 4는 삼각 창을 적용하여 두 프레임을 OLA하는 것을 도시화한 것이다.

손실구간이 길어질수록 복원된 음성신호는 부자연스러워진다. 그러나 점차적으로 신호를 감쇠(attenuation) 시킴으로써 음성품질을 향상시킬 수 있다[6].

아래의 그림 5는 지금까지의 LP계수를 이용한 PLC 알고리즘 블록도를 나타낸 것이다.

4. 손실이후 프레임 정보를 이용한 패킷손실 보상

사람이 발생하는 음소의 평균 길이는 대략 80msec 정도이다[4]. 따라서 패킷손실 길이가 그 이하라면 앞서 언급한 방법으로 손실보상이 가능하다. 하지만 실제로 IP 네트워크에서 발생하는 패킷손실은 연속적으로 일어나는 경우가 많으며 2개에서 10개 정도가 제일 빈번하다[4].



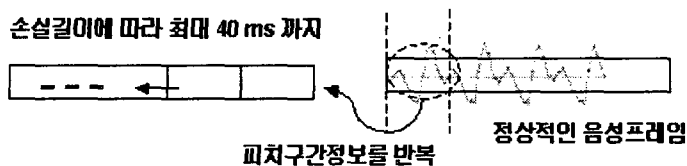
<그림 5> LP계수를 이용한 PLC 알고리즘 블록도

그러므로 한 음소길이보다 길게 음성신호의 손실이 일어날 수도 있다. 또한 음소가 변화하는 구간(transition region)에서 손실이 발생했을 경우에도 문제가 될 수 있으므로, 이에 대해 손실이후 프레임의 음성정보를 이용해 손실된 음성신호의 뒷부분을 보상해 주어야 한다. 여기서는 60msec을 넘는 음성손실에 대해서 최대 40msec을 뒤 프레임의 음소정보를 적용해 손실구간을 복원했다. 아래의 그림 6은 뒤 프레임으로부터 패킷손실을 보상하는 것을 보여준다.

AMDF를 적용하여 피치를 검출하고 피치정보를 손실구간 뒷부분에 삽입해 줌으로써 보다 자연스러운 복원음을 만들 수 있다. 예를 들어 한 패킷의 길이를 20msec라고 했을 때, 연속된 패킷손실에 대해서 이전 프레임 정보(피치길이만큼의 정보)를 너무 오랫동안 반복해 손실을 보상한다면, 인공음(artifact)에 의해 복원된 음성은 오히려 더 부자연스러워진다. 때문에 대부분의 PLC 알고리즘은 신호 감쇠를 적용하여 60msec 이후에는 묵음으로 대체하는 방법을 사용한다. 하지만 코더에서 음성을 재생하는 동안에도 다음 음성 패킷들은 버퍼에 저장되고 있으므로,

손실구간 뒷부분은 손실 이후 프레임 정보로 보상할 수 있다. 실험에서는 손실 길이에 따라 최대 40msec 까지 손실구간 뒤 부분을 보상해 주었다.

손실이후 프레임으로부터 보상을 할 경우에 추가적인 지연 문제가 발생할 수도 있다. 하지만 단-대-단(end-to-end) 통신지연의 합이 150msec 이하일 때는 대부분의 사용자는 대화에 불편을 느끼지 않으며[2], 최근에는 네트워크의 QoS(Quality of Service)를 보장하기 위해 많은 연구가 진행되고 있어 네트워크상에서의 지연문제는 어느 정도 해결되고 있다. 또한 수신 측에서 일정량의 재생 데이터를 버퍼에 저장하고 있기 때문에, 시스템 처리와 음성코더가 비동기적으로 동작되므로, 재생 스케줄 조정에 의해서 지연시간은 음성데이터 재생에 영향을 미치지 않을 수 있다[12].



<그림 6> 손실이후 프레임으로부터 보상

<표 1> 실험에 사용된 파라미터

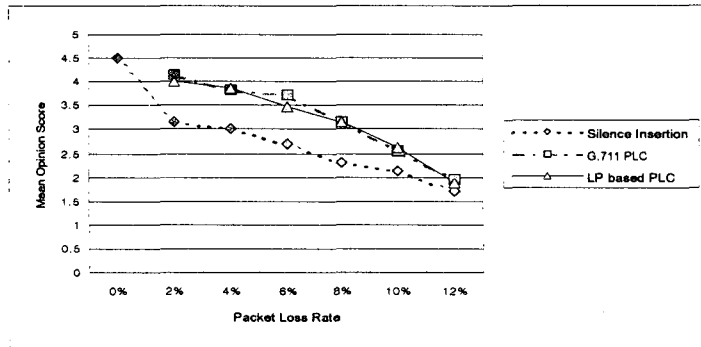
패킷크기	160 샘플 (20 msec)
표본화율	8 kHz
LPC 차수	12
τ_{min}	20 샘플
τ_{mx}	120 샘플

위의 표 1은 실험에 사용된 파라미터들을 간단히 정리한 것이다.

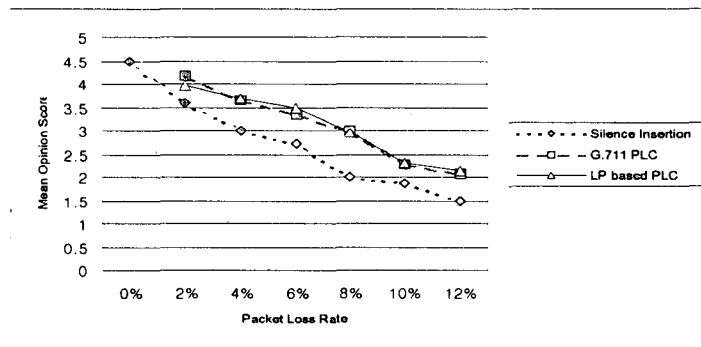
5. 알고리즘 평가 및 결과분석

G.711 PLC 알고리즘과 LP기반 PLC 알고리즘을 적용하여 패킷손실에 의해 열화된 음성신호의 품질이 얼마나 향상될 수 있는지 알아보고, 연속된 패킷손실에 대해서 손실이후 프레임으로부터 보상한 PLC 알고리즘을 기존의 알고리즘들에 추가하여 실험하였다.

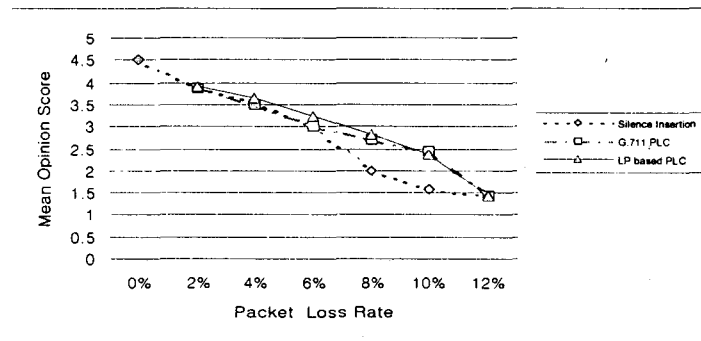
주관적평가(MOS: Mean Opinion Score)를 통해서 각각의 PLC 알고리즘에 대해서 점수를 산출했다. 20명의 피험자가 실험에 참여하였으며, 실험 데이터는 2명의 남녀가 발생한 5초-6초 분량의 나레이션 데이터 2개와 단문 데이터 2개를 사용하였다.



<그림 7> 하나씩 패킷손실이 발생했을 경우의 MOS



<그림 8> 연속으로 3개의 패킷손실이 발생했을 경우의 MOS



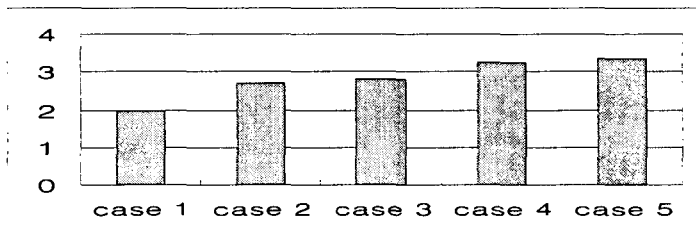
<그림 9> 연속으로 5개의 패킷손실이 발생했을 경우의 MOS

손실된 음성 데이터를 구성하기 위해서 패킷손실은 패킷마다 서로 독립적으로 발생한다고 가정하였다(iid packet loss model). 하지만 실제로 패킷손실은 패킷간에 상호의존적(inter-dependent)으로 발생한다. 이에 대한 영향을 고려하여 손실이 발생할 때 연속으로 손실되는 패킷을 3개와 5개로 늘려 MOS 테스트를 추가로 실시했다.

그림 7, 8, 9는 패킷손실률을 2% - 12%까지 변화하는 패킷손실에 따른 각 PLC 알고리즘의 주관적평가 결과이다. G.711 PLC 알고리즘과 LP기반 PLC 알고리즘은 알고리즘의 유사성(피치정보에 의해 손실보상)때문에 엇비슷한 결과를 보인다. 하지만 연속된 패킷손실이 증가 할수록 LP기반 PLC 알고리즘이 약간 더 나은 성능을 나타냈다. 보통 낮은 패킷손실률에 대해서는 같은 손실률일 경우에 연속적인 패킷손실의 MOS 결과가 더 높게 나온다. 그 이유는 전체 음성 중에서 패킷손실이 발생한 부분이 특정 부분에 몰려 있어, 그 부분은 제외하면 고르게 손실이 분포한 경우보다 훨씬 깨끗한 음질을 나타내기 때문이다.

하지만 앞의 그림 9에서 알 수 있듯이 패킷손실이 연속으로 5개 이상 발생하게 되었을 때는 앞서의 PLC 알고리즘에 의한 음질의 개선정도가 작아진다. 그 이유는 패킷손실구간이 길어지면서 음소의 변화구간에서 손실이 많이 발생하거나 하나의 음소정보가 전부 사라져 버리는 경우가 생기기 때문이다.

이 경우에 손실구간의 뒷부분을 손실이후 프레임으로부터 보상하게 되면 추가적인 음질향상을 기대할 수 있다. 여기서는 연속적인 패킷손실이 5일 때 이 알고리즘을 적용하여 MOS 테스트를 수행하였다. 아래 그림 10은 5개의 연속된 패킷손실에 대해서 8% 손실을 적용했을 때의 주관적 평가결과로써 G.711 PLC 알고리즘과 LP기반 PLC 알고리즘 모두에서 추가적인 음질향상이 있음을 확인할 수 있다.



Case 1	Silence insertion
Case 2	G.711 PLC
Case 3	LP기반 PLC
Case 4	G.711 PLC + 후 프레임 보상
Case 5	LP기반 PLC + 후 프레임 보상

<그림 10> 5연속 패킷손실, 손실률 8%일때의 MOS 비교

대략 15% 이내의 패킷손실률에 대해서 60ms 이하의 손실구간을 앞 프레임 음성정보를 이용한 PLC 알고리즘으로 패킷손실에 대해서 음질 개선이 가능하고, 평균음소 길이(80ms)보다 긴 손실구간에 대해서는 추가적으로 손실구간 이후의 프레임 음성정보를 이용해 음질을 향상시킬 수 있다.

6. 결 론

음소단위의 음성신호가 전부 손실되지 않을 정도의 패킷손실에 대해서는 G.711 PLC 알고리즘이나 LP계수를 이용한 PLC 알고리즘 모두에 대해서 음질개선이 가능하지만, 실험적으로 60msec 이상 되는 손실구간에 대해서는 앞 프레임의 음성신호 정보만으로 손실 보상을 하기에는 다소 무리가 있다. 본 논문에서 제안된 손실 후 프레임 정보를 이용해 보상할 경우에는 앞 프레임 음성정보의 과도한 반복으로 인해 발생하는 부자연스러운 인공음(artifact)을 방지할 수 있으며, 긴 구간에 대한 패킷손실 발생시 추가적인 음질향상을 가져온다. 하지만 보다 정확한 결과를 얻기 위해서는 실험에 참여하는 피험자의 수를 더 많이 늘릴 필요가 있고 평가데이터의 종류도 다양화시켜야 한다. 또한 손실 후 프레임을 이용하기 위해서 음성신호의 재생 스케줄을 세밀히 조정하여 재생시 일어날 수 있는 지연을 피할 수 있도록 고려해야 한다.

참 고 문 헌

- [1] Kos, A., B. Klepec and S. Tomazic (2002), Techniques for performance improvement of VoIP applications, *MELECON 2002*. 11th Mediterranean, pp.250~254.
- [2] Janssen, J., D. De Vleeschouwer, M. Buchli and G. H. Petit (2002), Assessing voice quality in packet-based telephony, *IEEE Internet Computing* Vol.6., May-June, pp.48~56.
- [3] Arai, M., A. Chiba and K. Iwasaki (1999), Measurement and modeling of burst packet losses in Internet end-to-end communications, *1999 Pacific Rim International Symposium on*, pp.260~267.
- [4] Kim, Chin Suk and R. Braun (2001), A survey of UDP packet loss characteristics, *Signals, Systems and Computers*, pp.200~204.
- [5] Perkins, C., O. Hodson and V. Hardman (1998), A survey of packet loss recovery techniques for streaming audio, *IEEE Network* Vol.12., Sept.-Oct., pp.40~48.
- [6] Appendix I, A High (1999), Quality Low-Complexity Algorithm for Packet Loss Concealment with G.711, *ITU-T Recommend. G.711*, Sept., pp.1~7.
- [7] 김무중, 권철홍(2002), 인터넷 전화에서 손실 패킷 복원을 위한 동적인 부가 정보 전송 기법, 「한국음향학회지」 제21권 제4호, pp.349~360.

- [8] Pulse Code Modulation (PCM) of Voice Frequencies, *ITU-T Recommend. G.711*, Nov. 1998, pp.7~10.
- [9] Lindblom, J. and P. Hedelin (2002), Packet loss concealment based on sinusoidal extrapolation, *Acoustics, Speech, and Signal Processing, IEEE International Conference on Vol.1*, pp.173~176.
- [10] Gunduzhan, E. and K. Momtahan (2001), Linear prediction based packet loss concealment algorithm for PCM coded speech, *Speech and Audio Processing, IEEE Transactions on Vol.9*, Nov., pp.778~785.
- [11] Oppenheim, A. V. and Schafer, R. W. (1975), *Digital Signal Processing*, Prentice-Hall, New Jersey, pp.815~825.
- [12] de Martin, J. C., T. Unno and V. Viswanathan (2000), Improved frame erasure concealment for CELP-based coders, *IEEE International Conference on Vol. 3.*, pp.1483~1486.
- [13] Cox, R. V. (1997), Three new speech coders from the ITU cover a range of applications, *IEEE Communications Magazine Vol.35*, Sept., pp.40~47.
- [14] Bi, Jingping, Qi Wu and Zhongcheng Li (2002), Packet delay and packet loss in the internet, *Computers and Communications*, pp.3~8.
- [15] Wasem, O. J., D. J. Goodman, C. A. Dvorak and H. G. Page (1988), The effect of waveform substitution on the quality of PCM packet communications, *IEEE Transactions on Vol.36*, Mar., pp.342~348.

접수일자: 2002년 5월 2일

게재결정: 2002년 5월 24일

▶ 김재현(Jae-Hyun Kim)

주소: 305-732 대전광역시 유성구 화암동 58-4번지 한국정보통신대학원대학교

소속: 한국정보통신대학원대학교(ICU) 음성/음향 정보 연구실

전화: 042) 866-6196

E-mail: hijhkim@icu.ac.kr

▶ 한민수(Min-Soo Hahn)

주소: 305-732 대전광역시 유성구 화암동 58-4번지 한국정보통신대학원대학교

소속: 한국정보통신대학원대학교(ICU) 음성/음향 정보 연구실

전화: 042) 866-6123

E-mail: mshahn@icu.ac.kr