

# 피치 하모닉 움직임 예측과 적응적 신호 크기 예측을 이용한 패킷 손실 은닉 알고리즘

김태하, 이인성\*

## Packet Loss Concealment Algorithm Using Pitch Harmonic Motion Estimation and Adaptive Signal Scale Estimation

Tae-Ha Kim, In-Sung Lee\*

**요약** 본 논문에서는 피치 하모닉 움직임 예측과 적응적 신호 크기 예측을 이용한 패킷 손실 은닉 알고리즘을 제안한다. 스펙트럼 움직임 예측 방법은 사용 가능한 이전 패킷의 스펙트럼 상의 움직임을 일정한 부대역으로 나누어 손실된 신호의 움직임을 예측하여 복원한다. 제안하는 알고리즘에서는 음성신호를 유성음과 무성음으로 구분하여 유성음의 경우 피치 주파수를 활용하여 피치 하모닉으로 나누어 손실된 신호의 피치 하모닉 움직임을 예측하여 복원하고 무성음의 경우 스펙트럼 움직임 예측 방법을 사용하여 신호를 복원한다. 음성 프레임의 연속 손실이 발생한 경우 LMS(Least Mean Square) 예측기를 사용하여 이전 프레임의 이득 정보를 활용하여 신호 크기를 예측하여 출력 신호의 이득을 조절하는 방법을 제안한다. 객관적 평가방법인 PESQ (Perceptual Evaluation of Speech Quality) 시험을 통해 제안된 알고리즘의 성능을 평가하였고 기존의 방법보다 MOS 0.1의 성능 개선을 보였다.

**Abstract** In this paper, we propose a packet loss concealment (PLC) algorithm using pitch harmonic motion prediction and adaptive signal amplitude prediction and. The spectral motion prediction method divides the spectral motion of the previous usable frame into predetermined sub-bands to predict and restore the motion of the lost signal. In the proposed algorithm, the speech signal is classified into voiced and unvoiced sounds. In the case of voiced sounds, it is further divided into pitch harmonics using the pitch frequency to predict and restore the pitch harmonic motion of the lost frame, and for the unvoiced sound, the lost frame is restored using the spectral motion prediction method. When the continuous loss of speech frames occurs, a method of adjusting the gain using the least mean square (LMS) predictor is proposed. The performance of the proposed algorithm was evaluated through the objective evaluation method, PESQ (Perceptual Evaluation of Speech Quality) and was showed MOS 0.1 improvement over the conventional method.

**Key words:** Packet Loss Concealment, Speech Coding, Voice over IP, Pitch Harmonic, LMS Predictor, PESQ, Gain Control

### 1. 서 론

이동통신망에서 음성 트래픽 채널을 이용하여 음성을 전달하는 방법 뿐만 아니라 데이터 채널을 이용하여

음성 통신하는 방법이 많이 사용되고 있다. 국내외에서 고속의 데이터 서비스가 가능한 시스템이 상용화되었으며 VoIP나 VoLTE와 같은 패킷망을 이용한 음성 서비스

\*School of Information Communication Eng., Research Institute for Computer and Information Communication, Chungbuk National University, Chungdaero 1, Seowongu, Cheongju, Republic of Korea (Tel: 82-43-261-3223)

\*Corresponding author: In-Sung Lee, School of Information Communication Eng., Research Institute for Computer and Information Communication, Chungbuk National University(inslee@chungbuk.ac.kr)

Received July 30, 2021

Revised August 02, 2021

Accepted August 10, 2021

가 증가하고 있다[1-3]. 이러한 VoIP 서비스의 확대 공급에 따라 음성 서비스의 품질 관리가 중요한 요소가 되고 있다[4]. 특히 패킷 손실 시 음질 저하를 막는 방법은 VoIP서비스의 중요한 경쟁력을 유지하는데 중요한 핵심 기술이 되고 있다. IP 네트워크에서 패킷을 기반으로 음성통신을 수행하는 VoIP 서비스인 경우 전송지연(Delay), 지터(Jitter), 패킷 손실(Packet loss) 등으로 인하여 QoS(Quality of Service)를 보장받지 못하고 음성 품질의 저하를 발생하게 된다[5].

네트워크 망에서 사용하는 표준 음성 부호화 방식(G.711, G.723.1, G.729 등)들은 손실된 패킷을 복구하여 손실의 영향을 최소화하게 하는 패킷 손실 은닉 Packet Loss Concealment(PLC) 방법을 디코더에 적용하고 있다[6]. 패킷 손실 은닉 방법에는 크게 두가 방식으로 나누어진다. 먼저 송신자측이 음성 데이터에 미리 부가 정보를 포함하여 전송한 후 수신자 측에서 그 부가 정보를 받아 손실된 패킷을 복구하는 송신자 및 수신자 기반의 방법이 있다. 또 다른 방법으로 수신자 측에서 손실을 확인 후 자체적으로 음성 프레임으로부터 손실 프레임을 복구하는 수신자 기반의 방법이 있다[7].

스펙트럼 움직임을 이용한 패킷 손실 은닉 알고리즘은 수신자 기반의 방법 중 하나로 프레임 손실이 발생했을 경우 사용 가능한 이전 프레임의 음성 스펙트럼을 이용하여 손실된 패킷의 스펙트럼 예측한 후 손실된 패킷을 복원한다. 스펙트럼 예측을 통한 손실된 프레임을 복구하는 방법 외에 출력신호의 이득을 조절하는 방법을 사용한다. 연속 프레임 손실의 경우 일정하게 출력 음성의 이득을 감소하여 손실된 음성이 자연스럽게 다음 프레임에 연결하게 한다[8].

본 논문에서는 스펙트럼 움직임 벡터를 구할 때 균일한 대역폭을 갖는 부대역으로 나누지 않고 VAD(Voice Activity Detection)를 이용하여 무성음과 유성음을 구분하고 유성음의 경우 피치 파라미터를 활용해 피치 하모닉의 기본 주파수에 따라 변화하며 부대역으로 나누어 스펙트럼 움직임 벡터를 추출하는 방법을 제안한다. 또 다른 제안하는 패킷 은닉 방법으로 패킷 손실시 사용하는 이득 감소율을 고정적으로 사용하는 것이 아니라 연속 손실의 경우 손실된 프레임 기준 양쪽 프레임의 최대값을 활용하여 신호

의 크기 변화를 적응적으로 예측하여 이득을 조절하는 방법을 제안하였다. 제안된 패킷 손실 은닉 알고리즘의 성능평가는 ITU-T 객관적 음질 평가 방법인 PESQ 시험으로 각각의 성능을 평가 비교하였다.

## 2. 스펙트럼 움직임 예측을 이용한 패킷 손실 은닉 알고리즘

스펙트럼상의 움직임을 이용한 패킷 손실 은닉 알고리즘은 손실된 패킷의 스펙트럼 에너지의 움직임을 이전 신호들의 움직임으로부터 예측하고 예측된 신호의 움직임을 통하여 손실된 패킷을 복원하는 알고리즘이다[8]. PLC 알고리즘은 그림1과 같이 Windowing & Overlap, 스펙트럼 분석, 스펙트럼 움직임 예측, 선형 외삽법과 보간법, IDFT or IDCT & Overlap-add 의 5단계로 구성되어진다.

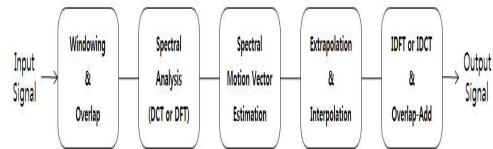


그림 1. 스펙트럼 움직임 예측을 이용한 PLC 알고리즘  
Fig. 1. PLC Algorithm using spectral motion estimation

첫 번째 단계인 Windowing & Overlap과 스펙트럼 분석에서는 입력신호를 일정한 윈도우를 씌워 시간 영역의 신호를 주파수 영역의 신호로 바꿔준다. Hamming Window를 20ms의 길이로 적용시켰고 75%의 Overlap 비율을 사용하였다.

Windowing & Overlap을 통하여 나온 신호는 DFT나 DCT를 적용하여 주파수 영역의 신호로 변환된다. Eq.(1)과 Eq.(2)는 각각 DFT와 DCT를 나타낸다.

$$X_{DFT}(k,l) = \sum_{m=0}^{N-1} x[l \cdot s + m]e^{-j2\pi mk/N} \tag{1}$$

$$X_{DCT}(k,l) = w(k) \sum_{m=0}^{N-1} x(l \cdot s + m) \cos[\pi(m + 0.5)k/N] \quad (2)$$

여기서  $N$ 은 연속적으로 분리된 신호의 길이,  $k$ 는  $0 \sim N-1$ 까지의 주파수 성분,  $s$ 는 Overlap의 비율을 나타낸다. Eq.(2)의 DCT에서  $w[n]$ 은 가중치 함수로서 Eq.(3)으로 나타낼 수 있다.

$$w[k] = \begin{cases} \frac{1}{\sqrt{N}}, & k=0 \\ \sqrt{\frac{2}{N}}, & k=1, \dots, N-1 \end{cases} \quad (3)$$

세 번째 단계인 스펙트럼 움직임 예측은 첫 번째와 두 번째 단계를 통하여 얻은 주파수 영역의 스펙트럼  $X(k,l)$ 을 Overlapping된 부대역의 갯수  $B$ 개로 나누어  $X_{SB}(k,l)$ ,  $SB = 1, \dots, B$ 로 표현하게 된다. 이렇게 각각의 부대역으로 나누어진 스펙트럼 신호는 Eq.(4)를 통해서 스펙트럼 움직임 벡터를 구할 수가 있다.

$$M_{k,l} = \arg \max_i \left( \sum_{i \in SB} f_{l,l-1}(X_{SB}(i,l), X_{SB}(i,l-1)) \right) \quad (4)$$

$$f_{l,l-1}(X_{SB}(i,l), X_{SB}(i,l-1)) = \sum_{m=1}^{SB_{-} \leq n-i} X_{SB}(m,l) X_{SB}(m-i,l-1) \quad (5)$$

여기서  $i$ 는  $k$ 번째 주파수가 존재하는 부대역의 위치, 함수  $f$ 는 교차 상관 함수를 의미하며 그 정의는 Eq.(5)와 같이 주어진다. 결국  $k$ 번째 주파수가 존재하는  $i$ 번째 부대역의 교차 상관 함수의 최대값을 갖는 지점이  $k$ 번째 주파수의 스펙트럼 움직임 벡터가 된다.

네 번째 단계인 선형 외삽법과 보간법 부분에서는 세 번째 단계에서 구한 스펙트럼 움직임 벡터를 이용하여 실제 손실된 프레임과 손상된 프레임을 복원하게 된다. 먼저 선형 외삽법을 통해 손실된 프레임을 직접 예측하여 복원하게 된다. 이전 세 개의 프레임으로부터

얻은 움직임 벡터를 이용하여 이전 세 개의 프레임을 이동한 후 그림2에서 같이 최소자승법[9]을 이용하여 손실된 신호를 찾게 된다. DFT 또는 DCT를 통하여 시간영역에서 주파수 영역으로 변환하는 과정에서 Windowing & Overlap 때문에 손실된 패킷 주변으로 손상된 패킷이 발생하게 된다. 이 손상된 패킷은 손실된 패킷을 복원한 후 Eq.(6)에 의해서 선형 보간법으로 복원하게 된다.[10]

$$I_L(x_1, x_2, T_G, l) = \frac{(l x_2 + T_G - l + 1) x_1}{T_G + 1} \quad (6)$$

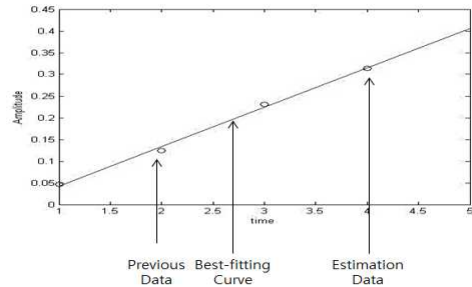


그림 2. 최소자승 예측 방법  
Fig. 2. Least square estimation method

여기서  $x_1$ 은 이전 사용가능한 프레임의 첫 번째 값,  $x_2$ 는 이후 사용가능한 프레임의 첫 번째 값,  $l$ 은 프레임 인덱스,  $T_G$ 는 손상된 프레임의 개수를 나타내며 손상된 프레임은 하나씩 복원됨으로 1로 고정된다.

### 3. 피치 하모닉 움직임 예측과 적응적 신호 크기 예측을 이용한 패킷 손실 은닉

음성은 유성음과 무성음에 스펙트럼의 모양이 다르게 나타나며 유성음에서는 피치 주파수의 배수에서 스펙트럼 에너지가 크게 나타나는 하모닉 스펙트럼 특성을 갖는다. 음성의 특성을 VAD를 통해 유, 무성음으로 구별한 후 무성음인 경우 일정한 크기의 부대역으로 나누어 움직임 벡터를 예측하여 스펙트럼을 예측하고 유성음인 경우 신호의 주기인 피치를 이용하여 피치 주기에 따라 변하는 피치 하모닉 성분으로 부대역을 나누어 움직임 벡터를 추출한다.

또다른 PLC 방법으로 출력신호의 이득을 적응적으로 조절하는 방법을 제안한다. 연속 손실의 경우 LMS 예측기를 통하여 구한 각 프레임의 최대값을 활용하여 손실된 구간에 신호의 변화를 예측해 적응적으로 이득을 조절하는 방법을 제안한다. 그림3은 제안한 피치 하모닉 움직임 예측 및 적응 이득조정 방법을 이용한 PLC 알고리즘을 나타내고 있다.

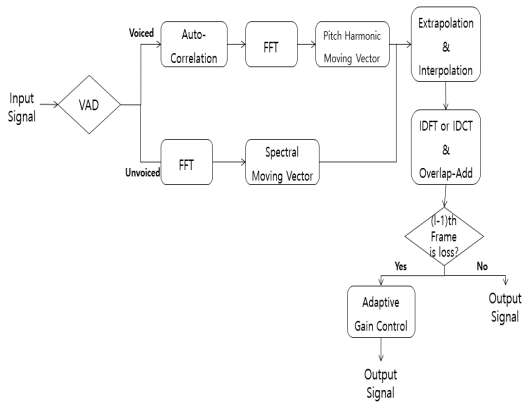


그림 3. 제안된 PLC 알고리즘  
Fig. 3. Proposed PLC Algorithm

### 3.1 음성/비음성 분리(VAD)

음성 프레임의 특성에 따라 스펙트럼의 부대역의 크기를 다르게 하기 위해 VAD를 사용한다. 본 연구에서는 계산량이 적게 요구되는 프레임의 에너지를 이용하여 음성 구간과 비음성 구간을 결정하는 방법을 사용하였다. 프레임의 에너지는 Eq.(7)을 통해 구하였다.

$$E_l = \frac{1}{L} \sum_{m=0}^{L-1} |x_l(m)| \quad (7)$$

여기서  $L$ 은 한 프레임의 길이를 나타낸다.

비 음성 신호의 에너지 값이 음성 신호의 에너지 값보다 현저히 낮기 때문에 일정 기준보다 높으면 음성 구간, 낮으면 비 음성 구간으로 Eq.(8)과 같이 결정한다.

$$VAD\_Lag = \begin{cases} 0, & E_l < \delta \\ 1, & E_l \geq \delta \end{cases} \quad (8)$$

### 3.2 피치 하모닉 움직임 벡터 추출

음성 프레임의 종류에 따라 부대역의 크기를 다르게 나누어 스펙트럼 움직임 벡터를 예측한다. Windowing & Overlap과 DFT나 DCT를 적용하기 전 자기 상관 함수[11]를 이용하여 피치 주기를 구한다. 유성음과 유성음으로 Overlap된 신호에 대해 피치를 이용한 피치 하모닉 움직임 벡터를 추출하게 되고 유성음과 무성음, 또는 무성음과 무성음으로 Overlap된 신호에서는 일정한 크기의 부대역으로 나누어 움직임 벡터를 추출하게 된다.

피치 주기를 구하는 방식으로 시간 영역에서의 방법은 자기 상관 함수를 이용한다. 자기 상관 함수는 Eq.(9)를 통하여 구하여진다.

$$R[k] = \sum_{m=0}^{L-k-1} x[m]x[m-k], k = 0, \dots, L-1 \quad (9)$$

자기상관함수  $R[k]$ 에서  $R[0]$ 는 첫 번째 최대값 지점이 되고 그 다음 두 번째 최대값을 갖는  $k$ 의 값이 시간 영역에서의 피치 주기가 된다. 피치 하모닉 부대역의 크기 결정은 먼저 피치를 시간 영역에서 구하였기 때문에 Eq.(10)을 활용하여 피치 주파수를 계산한다.

$$F0 = \frac{1}{T \cdot \frac{1}{f_s}} \quad (10)$$

여기서  $T$ 는 시간 영역에서의 피치 주기이며  $f_s$ 는 샘플링 주파수(Sampling Rate)를 말하고  $F0$ 는 주파수 영역으로 변환한 피치 주파수를 말한다. 피치 하모닉을 위한 부대역의 크기는 피치 주파수의 3배를 사용하였다. 이렇게 구한 피치 하모닉 성분으로 Eq.(11)을 통해 피치 하모닉 부대역의 크기를 결정한다.

$$PH\_SB = \frac{3F0}{f_s / DFTLength} \quad (11)$$

여기서  $DFTLength$ 는 DFT 길이를 말한다. Eq.(11)로 구한 피치 하모닉 부대역을 이용하여 음

직임 벡터를 추출하게 되는데 먼저 VAD를 이용하여 손실된 프레임이 음성 구간인지 확인한다. 손실된 프레임을 기준으로 이전 프레임이 유성음과 무성음이 섞여 있다면 피치가 정확하게 계산되지 않기 때문에 일정한 크기의 부대역로 나누어 움직임 벡터를 추출하며 유성음으로만 이루어져 있다면 추출된 피치를 활용하여 피치 하모닉 부대역로 나누어 움직임 벡터를 추출한다. 피치 하모닉 움직임 벡터는 Eq.(12)와 Eq.(13)을 통하여 계산한다.

$$P_{M_{k,l}} = \underset{i}{\operatorname{argmax}} \left( \sum_{i \in PH\_SB} f_{l,l-1}(X_{PH\_SB}(i,l), X_{PH\_SB}(i,l-1)) \right) \quad (12)$$

$$f_{l,l-1}(X_{PH\_SB}(i,l), X_{PH\_SB}(i,l-1)) = \sum_{m=1}^{N-1} X_{PH\_SB}(m,l) X_{PH\_SB}(m-i,l-1) \quad (13)$$

Eq.(11)에 의해서 만들어진 피치 하모닉 부대역은 DFT나 DCT 길이에 맞추어 떨어지지 않기 때문에 남은 밴드가 발생하게 되는데 남은 밴드는 Eq.(14)를 이용하여 구하고 Eq.(12)와 Eq.(13)을 이용하여 움직임 벡터를 구해준다.

$$Rest\_band = DFTLength - \operatorname{mod}(DFTLength, PH\_SB/2) \quad (14)$$

여기서 mod는 DFT의 길이를 PH\_SB/2의 길이로 나눈 나머지를 의미한다.

선형 외삽법 및 보간법은 최소 자승법과 Eq.(6)을 활용하여 구현하여 손실되고 손상된 신호를 복원하였다. 그림4, 그림5, 그림6은 각각 원 신호, 손실되고 손상을 입은 신호, 그리고 PLC에 의해 복원된 신호의 스펙트럼을 나타내었고 그림7은 스펙트로그램 상에서의 원 신호와 손실된 신호, 원 신호와 선형 외삽법과 보간법이 모두 적용된 신호를 나타내었다.

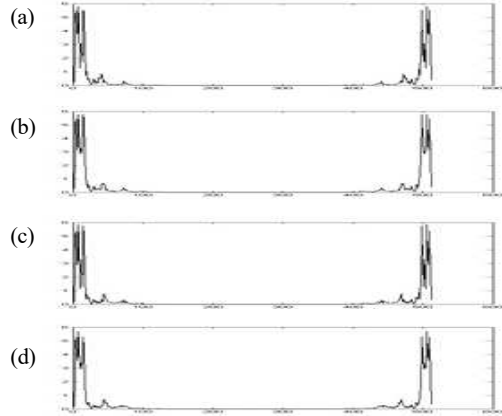


그림 4. 원신호의 스펙트럼

(a) l번째 프레임, (b) (l-1)번째 프레임  
(c) (l-2)번째 프레임(d) (l-3)번째 프레임

Fig. 4. Spectra of Original Signal

(a) lth frame (b) (l-1)th frame  
(c) (l-2)th frame (d) (l-3)th frame

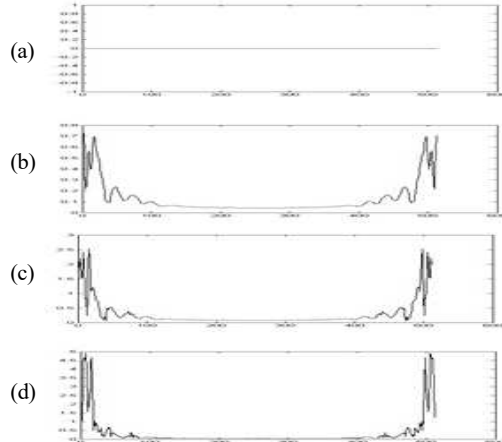


그림 5. 손실되고 손상된 프레임의 스펙트럼

(a) l번째 프레임, (b) (l-1)번째 프레임  
(c) (l-2)번째 프레임(d) (l-3)번째 프레임

Fig. 5. Damaged signal spectra due to frame loss

(a) lth frame (b) (l-1)th frame  
(c) (l-2)th frame (d) (l-3)th frame

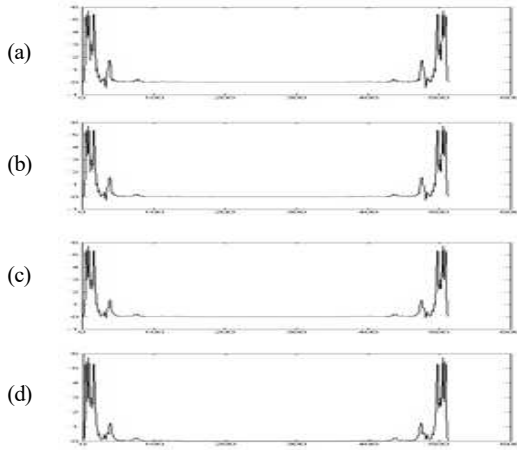


그림 6. 복원된 신호의 스펙트럼  
 (a)  $l$ 번째 프레임, (b)  $(l-1)$ 번째 프레임  
 (c)  $(l-2)$ 번째 프레임(d)  $(l-3)$ 번째 프레임  
 Fig. 6. Spectra of Reconstructed Signal  
 (a)  $l$ th frame (b)  $(l-1)$ th frame  
 (c)  $(l-2)$ th frame (d)  $(l-3)$ th frame

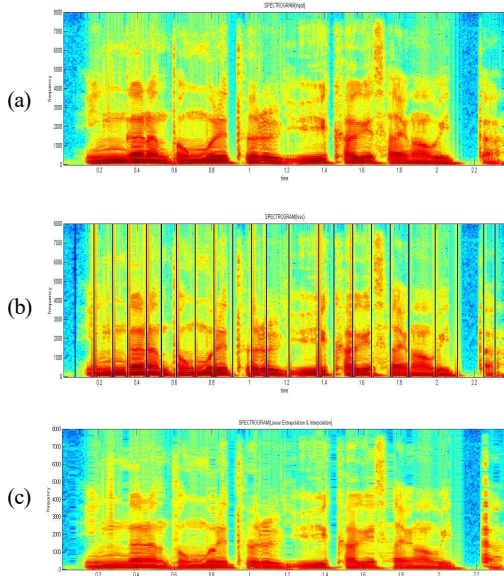


그림 7. 원신호와 복원된 신호의 스펙트로그램  
 (a) 원신호 (b) 손실된 신호 (c) PLC를 사용하여 복원된 신호  
 Fig. 7. Spectrograms of (a) the clean signal, (b) the signal with frame loss, and (c) reconstructed Signal using PLC algorithm

### 3.3 적응적 신호 크기 예측을 통한 이득 조절

음성의 경우 시간에 대해서 천천히 변화하는 특성을 가지고 있다. 손실된 음성 프레임은 가장 가까운 구간의 음성 프레임과 가장 유사한 특성을 가지게 된다. 제안하는 방법에서 LMS 예측기를 활용하여 손실된 구간의 이전과 이후의 최대값을 예측하여 적응적으로 이득을 조절한다.

이전 프레임의 최대값, 스텝 크기, 예측오차를 이용하여 필터 계수를 갱신하고 갱신된 필터계수와 이전 프레임의 최대값 곱함으로 현재 프레임의 최대값을 예측한다[13]. Eq.(15)를 이용하여 필터계수가 갱신되고 Eq.(16)을 이용하여 예측 오차를 구할 수 있으며 최종적으로 Eq.(17)을 이용하여 손실된 프레임의 최대값을 예측할 수 있다.

$$H_{l+1} = H_l + \mu \cdot E_l \cdot M_{l-1} \quad (15)$$

$$E_l = M_l - \hat{M}_{l-1} \quad (16)$$

$$\hat{M}_l = H_l \cdot M_{l-1} \quad (17)$$

여기서  $\mu$ 는 스텝 크기,  $E_l$ 는 예측오차,  $M_{l-1}$ 은 이전 프레임의 최대값을 나타내고 Fig.10에서 원 신호의 프레임 별 최대값과 예측된 최대값을 보여주고 있다.

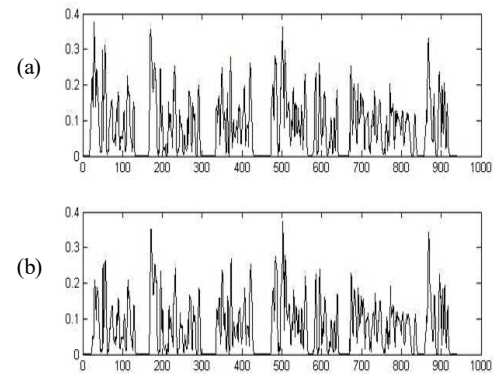


그림 8. LMS예측기를 이용한 최대값 예측  
 (a) 원 신호의 최대값 (b) 예측된 최대값  
 Fig. 8. Max value prediction using LMS predictor  
 (a) Max value of original signal (b) predicted max value

손실된 프레임은 기준으로 양쪽 프레임을 확인 한 후 모두 손실이 되지 않은 경우에는 양쪽 프레임의 최댓값을 활용하여 이득을 조절하고 양쪽 프레임 중 하나라도 손실이 생겼다면 LMS 예측기를 사용하여 예측한 최댓값을 활용하여 이득을 조절하게 된다.

증가하는 신호에 대해서는  $W_i$ , 감소하는 신호에 대해서는  $W_d$ 를 적용하여 이득을 조절하는데  $W_i$ 와  $W_d$ 는 Eq.(18)으로 구하게 되고 그림11에 나타내었다.

$$W = \begin{cases} W_i = \frac{\hat{M}_{l+1}}{\hat{M}_{l-1}} & (if \hat{M}_{l+1} > \hat{M}_{l-1}) \\ W_d = \frac{\hat{M}_{l+1}}{\hat{M}_{l-1}} & (if \hat{M}_{l+1} < \hat{M}_{l-1}) \end{cases} \quad (18)$$

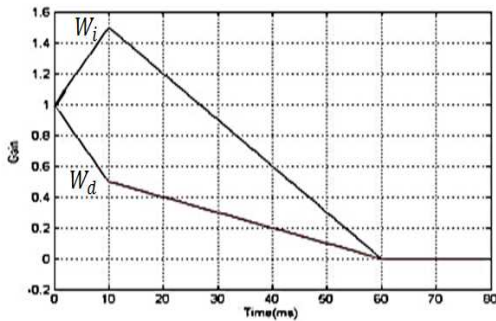


그림 9. 적응 신호 스케일 예측을 이용한 이득 조절  
Fig. 9. Gain control using adaptive signal scale estimation

### 4. 성능평가

본 논문에서 제안한 패킷손실 은닉 알고리즘을 성능 평가하기 위해서 ITU-T 객관적인 음질 평가 방법인 PESQ 평가 방법을 사용하였다[14]. Gilbert-Elliot 모델을 사용하여 IP 네트워크 상에서의 프레임 손실을 모델링하였다[15]. Gilbert-Elliot 모델의 경우 그림12와 같이 Markov Chain으로 표현되어 있다. 평균 손실률과 연속 손실 확률은 Eq.(19)와 Eq.(20)의 식으로 계산할 수 있다.

$$L_{avg} = \frac{P1}{P1 + P2} \quad (19)$$

$$L_{n-continuous} = (1 - P2) \cdot P2^{n-1} \quad (20)$$

5%, 10%, 15%, 20%의 평균 패킷 손실률 환경에서 성능 시험하였고 연속적으로 패킷 손실될 확률은 20%로 적용하여 시험하였다. 평균 손실률에 따른 P1와 P2의 값은 표1에 표기하였다.

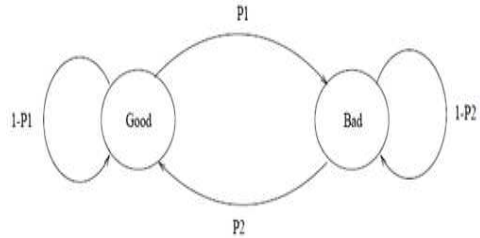


그림 10. Gilbert-Elliot 모델[15]  
Fig. 10. Gilbert-Elliot model[15]

표 1. 에러 모델의 확률  
Table 1. Probability of error model

Average Loss Rate	P1	P2
5%	0.04208	0.8
10%	0.08888	0.8
15%	0.14117	0.8
20%	0.2	0.8

표 2. 시험에 사용된 음성 신호  
Table 2. Speech signal used in test

File Name	Sampling Rate	Total Length	Frame Length
Korean_male1.wav	16kHz	28.76sec	20ms
Korean_male2.wav	16kHz	24.12sec	20ms
Korean_female1.wav	16kHz	28.98sec	20ms
Korean_female2.wav	16kHz	25.02sec	20ms

실험에 사용한 음성신호는 표2에 정의하였다. 표2의 음성 데이터를 ITU-T G.711 코덱[16]의 인코딩, 디코딩에 적용하여 전송 손실이 없을시 출력신호

의 PESQ 값을 측정하였다. G.711 코덱을 거쳐서 나온 PESQ 값은 한국남성1 음성의 경우 4.26, 한국남성2인 경우 4.20의 값을 보였고 한국여성음성1의 경우 3.95, 한국여성음2의 경우 4.21의 값을 보였다. Gilbert-Elliott Error Model을 적용하여 패킷 손실을 만든 후 제안한 알고리즘의 PESQ 성능을 측정하였다.

표3과 표4에서는 제안한 알고리즘을 적용하여 얻은 음성 신호의 PESQ 값을 보여주고 있다. 표3과 표4에서 PH-SB 방법은 피치하모닉 움직임 예측 방법만 사용한 경우이고 PH-SB+AGC는 피치하모닉 움직임 예측 방법과 적응 이득조절 방법 모두 사용한 결과이다. 스펙트럼 움직임을 이용한 패킷 손실 알고리즘의 경우 부대역의 개수 별로 증가함에 따라 음질 개선을 이룰 수 있었으며 제안한 피치 하모닉 움직임 벡터와 적응적 이득 조절을 이용한 패킷 손실 은닉 알고리즘의 성능이 기존의 8개 부대역 스펙트럼 예측 방법보다 남성 음성에서 5% 프레임 손실시 MOS 0.22, 10% 프레임 손실시 MOS 0.1, 20% 프레임 손실시 MOS 0.14의 성능 개선을 보였고 여성 음성에서는 5% 프레임 손실시 MOS 0.17, 10% 프레임 손실시 MOS 0.1, 20% 프레임 손실시 MOS 0.14의 성능 향상을 보였다. 새롭게 제안한 PH-SB+AGC 방법은 기존의 32개 부대역 스펙트럼 움직임 예측 방법보다 남성,여성 음성에서 20% 프레임 손실시 MOS 0.1의 성능 개선을 보였다.

표 3. 남성 음성 PESQ 결과 (Korean\_male.wav)  
Table 3. PESQ results (Korean\_male.wav)

Loss Rate	8 Sub-band	16 Sub-band	32 Sub-band	PH-SB	PH-SB+AGC
5%	3.52	3.54	3.74	3.74	3.74
10%	3.35	3.39	3.41	3.43	3.45
15%	3.00	3.01	3.04	3.07	3.10
20%	2.81	2.83	2.85	2.91	2.95

표 4. 여성 음성 PESQ 결과 (Korean\_female.wav)  
Table 4. PESQ results (Korean female.wav)

Loss Rate	8 Sub-band	16 Sub-band	32 Sub-band	PH-SB	PH-SB+AGC
5%	3.34	3.43	3.49	3.50	3.51
10%	3.25	3.28	3.30	3.32	3.35
15%	2.93	2.94	2.95	3.01	3.03
20%	2.71	2.73	2.75	2.80	2.85

### 5. 결 론

본 논문에서는 피치 하모닉 움직임 예측과 적응적 신호 크기 예측을 이용한 패킷 손실 은닉 알고리즘을 제안하였다. 음성 프레임의 특성에 따라 유성음과 무성음 분리하고 무성음인 경우 스펙트럼 움직임예측을 이용한 패킷 손실 은닉 알고리즘을 적용하였으며 유성음인 경우 피치 주기 값을 계산하여 피치 기본 주파수에 따라 부대역 크기를 가변적으로 나누어 피치 하모닉 움직임 예측 패킷 손실 은닉 알고리즘을 적용하였다. 또 다른 패킷 손실 은닉 방법으로 연속 패킷 손실 시 적응적으로 출력 음성 이득 조절 방법을 제안하였다. 연속 손실 상황에서 기존의 알고리즘의 경우 G.711 코덱의 패킷 손실 은닉 알고리즘의 연속 손실 모델인 일정한 감소율을 사용하여 이득을 조절하는 부분을 LMS 예측기를 사용하여 최대값을 예측 한 후 신호 변화에 따라서 적응적으로 이득을 조절하는 방법을 제안하였다.

제안한 피치 하모닉 움직임 예측과 적응적 신호 크기 예측을 이용한 패킷 손실 은닉 알고리즘의 성능을 평가하기 위해 객관적인 성능 평가 방법인 ITU-T 표준 방법인 PESQ 시험을 사용하여 성능 시험하였다. PESQ 평가 결과 각 음성 특성에 맞게 가변적으로 부대역을 조절하고 연속 손실에 대해서 적응적으로 이득을 조절 방법이 일정한 부대역으로 나누어 스펙트럼 움직임 예측하고 연속 손실에 대해 일정하게 이득을 조절한 방법에 비해 남성음성의 경우 최대 MOS 0.22, 여성음성의 경우 최대 MOS 0.17의 향상된 성능을 볼 수 있었다. 패킷 손실시 PLC 알고리즘에 의해 복원된 음성 스펙트럼의 모습이 원래의 음성의 스펙트럼을 잘 살려내고 있음을 확인할 수 있었다.



## REFERENCES

- [1] S. Y. Jo, "Trends on Standardization of Voice Service Support in LTE Networks"(in Korean), TTA Journal 139, pp.100-103, 2012.
- [2] M. S. Lee, D. Y. Kim, and B. S. Lee, "Trends of Codec Technology for 4G Mobile Enhanced Voice Service "(in Korean), Electronics and Telecommunications Trends 25, pp. 29-37, 2010.
- [3] M. S. Lee, D. Y. Kim, and B. S. Lee, "Trends on Speech Codec for Voice Communication Service"(in Korean), Electronics and Telecommunications Trends 16, pp. 46-58, 2010.
- [4] B. Y. Chang, D. W. Seo, and B. J. Park, "Study on VoIP Service Quality Management"(in Korean), J. the Institute of Webcasting, Internet and Telecommunication 11, pp.245-252, 2011.
- [5] S. H. Han, J. S. Kim, H. W. Lee, W. Ryu and M. S. Hahn, "Performance Improvement of Packet Loss Concealment Algorithm in G.711 Using Speech Characteristics"(in Korean), Phonetics and Speech Sciences, 57, pp.175-198, 2006.
- [6] ITU-T Recommendation, Appendix I: A high quality low-complexity algorithm for packet loss concealment with G.711, 1999.
- [7] V. P. Bhute and U. N. Shrawankar, "Speech Packet Concealment Techniques Based on Time-Scale Modification for VoIP", IEEE International Conference on Computer Science and Information Technology(ICCSIT), pp.825-828, 2008.
- [8] S. K. Pedram, S. Vaseghi, and B. Langari, "Audio packet loss concealment using spectral motion", IEEE International Conference on Acoustic, Speech and Signal Processing(ICASSP), pp. 6707-6710, 2014.
- [9] S. J. Miller, "The Method of Least Squares", Mathematics Department Brown University, 2006
- [10] E. Zavarehei, and S. Vaseghi, "Interpolation of Lost Speech Segments Using LP-HNM Model With Codebook Post-Processing", IEEE Transactions on Multimedia 10, pp. 493-502, 2008.
- [11] X. Mei, J. Pan and S. Sun, "Efficient algorithms for speech pitch estimation", IEEE International Symposium on Intelligent Multimedia, Video and Speech Processing(ISIMP), pp.421-424, 2001.
- [12] K. C. Kim, S. J. Park, S. P. Lee, and M. Y. Kim, "Pitch Estimation Method in an Integrated Time and Frequency Domain by Applying Linear Interpolation", The Institute of Electronics and Information Engineers 47, pp.100-108, 2010.
- [13] Zeidler, J.R., "performance analysis of LMS adaptive prediction filter", Proceedings of the IEEE 78, pp.1781-1806, 1990.
- [14] ITU-T Recommendation, 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] M. Mushkin and I. Bar-david, "Capacity and coding for the Gilbert-Elliot channels", IEEE Transactions on Information Theory 35, pp.1277-1290, 1989.
- [16] ITU-T Recommendation, Amendment 2: New Appendix III - Audio quality enhancement toolbox, 2009.

---

## 저자약력

---

김 태 하 (Tae-Ha Kim)

[일반회원]



〈관심분야〉

- 2013년 : 충북대학교 정보통신과 (학사)
- 2014년 ~ 2016년: 충북대학교 전파통신공학과(석사)

음성/오디오 신호처리

이 인 성 (In-Sung Lee)

[정회원]



<관심분야>

- 1983년 : 연세대학교 전자공학과 (학사)
  - 1985년 : 연세대학교 전자공학과 (석사)
  - 1992년 : Dept. of Electrical Eng. Texas A&M University (박사)
  - 1993년 ~ 1995년 : 한국전자 통신연구원 이동통신기술연구단 선임연구원
  - 1995년 ~ 현재 : 충북대학교 정보통신공학부 교수
- 음성/오디오 부호화, 지능신호처리, 통신신호처리