

음성신호의 특성을 고려한 패킷 손실 은닉 알고리즘

정회원 윤성완*, 강홍구*, 종신회원 윤대희*

Packet Loss Concealment Algorithm Based on Speech Characteristics

Sung-Wan Youn*, Hong-Goo Kang* *Regular Members*, Dae-Hee Youn* *Lifelong Member*

요약

VoIP(Voice over Internet Protocol)와 같은 IP 네트워크망에서는 패킷 지연, 지터, 패킷 손실 등의 이유로 QoS(Quality of Service)를 보장받지 못하기 때문에, 패킷 손실을 은닉하는 방법에 대한 연구는 필수적이다. IP망에서 사용되는 대부분의 저전송률 음성부호화기는 자체적으로 패킷 손실 은닉(PLC: Packet Loss Concealment) 알고리즘을 사용하고 있지만, 예측 기법에 기반한 양자화 특성상 패킷 손실 이후에도 에러가 전파되는 문제가 있다. 또한, 손실된 패킷의 음성신호 특성을 고려하지 않고 과거 파라미터값을 반복시키는 기존 PLC 방법은 그 구현은 쉽지만 천이구간에서의 합성신호의 음질이 심각히 저하된다. 본 논문에서는 패킷 손실 환경에서 음성신호 특성에 따른 에러전파 영향을 정량적으로 분석하고 그 결과를 토대로 보간법 기반의 새로운 PLC 알고리즘을 제안한다. 제안한 알고리즘은 파라미터별로 음성신호의 특성을 고려해 선택적으로 보간법을 적용하고, 예측 필터의 메모리를 효과적으로 갱신한다. 성능평가 결과, 제안한 알고리즘은 VoIP에서 널리 사용되는 G.729의 기존 PLC 알고리즘에 비해 다양한 FER 환경에서 성능이 향상되었다.

Key Words : Packet Loss Concealment, Frame Erasure, Speech Coding, VoIP, Speech Class

ABSTRACT

Despite of the in-depth effort to control the variability in IP networks, quality of service (QoS) is still not guaranteed in the IP networks. Thus, it is necessary to deal with the audible artifacts caused by packet losses. To overcome the packet loss problem, most speech coding standard have their own embedded packet loss concealment (PLC) algorithms which adapt extrapolation methods utilizing the dependency on adjacent frames. Since many low bit rate CELP coders use predictive schemes for increasing coding efficiency, however, error propagation occurs even if single packet is lost. In this paper, we propose an efficient PLC algorithm with consideration about the speech characteristics of lost frames. To design an efficient PLC algorithm, we perform several experiments on investigating the error propagation effect of lost frames of a predictive coder. And then, we summarize the impact of packet loss to the speech characteristics and analyze the importance of the encoded parameters depending on each speech classes. From the result of the experiments, we propose a new PLC algorithm that mainly focuses on reducing the error propagation time. Experimental results show that the performance is much higher than conventional extrapolation methods over various frame erasure rate (FER) conditions. Especially the difference is remarkable in high FER condition.

I. 서론

VoIP같은 패킷 네트워크망에서는 지연 (delay),

지터 (jitter), 패킷 손실 등으로 인해 QoS를 보장받지 못하는 문제가 있다. 급격한 음질저하를 유발하는 패킷 손실 문제를 해결하기 위해, 패킷 네트워크

* 연세대학교 전기전자공학과 디지털신호처리 연구실(yocello@dcsp.yonsei.ac.kr)

논문번호 : KICS2006-05-001, 접수일자 : 2006년 5월 10일, 최종논문접수일자 : 2006년 7월 3일

망에 사용되는 대부분의 표준 음성부호화기는 자체적인 PLC 알고리즘을 사용하고 있다. PLC 알고리즘은 크게 송신단 기반의 방법과 수신단 기반의 방법으로 분류할 수 있다¹¹. 송신단 기반의 방법은 패킷 손실이 발생할 경우를 대비해, 추가 비트 할당과 지연시간을 감수하여 현재 패킷의 정보를 다음 패킷과 함께 전송시키는 방법이며, FEC(Forward Error Correction)이 대표적이다^{12, 13}. 이에 비해 수신단 기반의 방법은 송신단에서 손실된 패킷에 대한 부가 정보를 보내지 않기 때문에, 이전 또는 이후에 수신되는 패킷의 정보를 이용하여 손실패킷을 복원하며 CELP(Code-Excited-Linear-Prediction)와 같은 저전송률 음성부호화기에서 주로 사용된다¹⁴. 하지만 예측기법을 사용한 부호화과정으로 인해 패킷 손실 이후, 후속프레임들에 에러가 전파(error propagation)되는 문제점이 있다. 또한, 동일한 프레임 에러 비율(FER: Frame Erasure Rate) 환경에서도 손실 패킷의 음성신호 특성에 따라 복구된 음성신호의 음질은 달라질 수 있다¹⁶. 예를 들면, 묵음 또는 무성음 구간의 패킷 손실은 음질의 저하가 크지 않지만, 무성음에서 유성음으로 변하는 천이구간이 손실됐을 경우, 심각한 음질의 저하가 야기된다.

본 논문에서는 음성신호의 특성을 고려한 PLC 알고리즘을 제안한다. 이를 위해 우선 LPC (Linear Prediction Coefficient), 피치, 이득값, 고정 코드북 등 CELP 기반 음성 부호화기의 파라미터들이 패킷 손실 환경에서 각 음성신호의 특성에 따라 전체 음질 저하에 미치는 영향과 중요도를 분석한다. 이 결과를 토대로 음성신호의 특성별로 선택적 보간법 및 후속프레임을 위한 효율적인 메모리 업데이트를 적용한 PLC 알고리즘을 제안하고 ITU-T G.729¹⁷ 음성부호화기에 구현한다. 제안된 PLC 알고리즘의 성능 평가를 위해 반복법 기반의 G.729 PLC 알고리즘과 함께 스펙트럼 왜곡(SD: Spectral Distortion)과 PESQ¹⁸를 측정한다.

본 논문의 구성은 다음과 같다. II장에서는 기존 G.729 음성부호화기의 PLC 알고리즘에서 발생하는 에러 전파를 관찰하고, III장에서는 음성신호의 특성에 따른 패킷 손실 영향을 분석한다. IV장에서는 음성신호의 특성을 고려한 보간법 기반의 패킷 손실 알고리즘을 제안하고, V장에서는 기존 G.729 패킷 손실 은닉 알고리즘과 비교 평가한다. 마지막으로 VI장에서는 결론과 추후 연구과제에 대해 기술한다.

II. 예측기법 음성부호화기의 에러 전파 영향

대부분의 저전송률 음성부호화기는 음성신호의 성도(vocal tract), 즉 스펙트럼 정보를 AR(Auto Regressive) 모델링하는 선형예측 기법을 사용한다. 그리고, 적응 코드북 및 고정 코드북으로 구성되는 여기(excitation) 신호 벡터를 선형 예측 계수로 구성된 합성필터에 컨볼루션(convolution) 시켜 음성신호를 재생한다. 또한, 인접한 프레임 또는 서브프레임의 정보를 이용하여 LSF(Line Spectral Frequency)와 고정 코드북 이득값과 같은 파라미터의 양자화 과정에 필요한 비트를 감소시켜 부호화 효율을 높인다. 그러나 이와 같이 프레임간의 상호의존성을 이용한 부호화 방법은 패킷 손실 환경에서 부/복호화기의 내부 메모리가 서로 달라지기 때문에 후속 프레임에 에러의 영향이 전파되고 부/복호화기의 내부메모리가 서로 동기화되기까지 상당한 시간이 소요된다. 본 장에서는 VoIP환경에서 널리 사용되는 ITU-T 8 kbit/s CS-ACELP G.729 음성부호화기를 이용해 패킷 손실 환경에서의 에러 전파 영향에 관해 분석하고 패킷 손실 은닉 알고리즘 제안시 고려해야할 문제점들을 정리한다.

프레임 에러가 발생했을 경우, G.729 음성부호화기는 복호화에 필요한 파라미터를 이전 프레임의 정보를 반복하거나 감쇠시켜 사용하는 에러 은닉(error concealment) 알고리즘을 통해 복구한다. 하지만 예측기법 부호화과정으로 인해 에러 발생 후 부/복호화기간의 필터 메모리가 불일치되고, 이 영향은 LSF와 고정 코드북 이득값의 양자화 과정에 사용되는 4차 MA 예측 필터, 과거 여기 신호 버퍼, 합성 필터 메모리에 남게 되어 합성신호의 음질을 저하시킨다¹⁹. 예를 들면, 프레임 에러가 발생했을 경우 G.729는 이전 프레임의 LSF를 손실된 프레임의 LSF로 사용하지만, 현재 프레임의 4차 MA 필터의 코드정보가 없기 때문에 식 (1)과 같이 복구된 LSF를 이용해 4차 MA 필터의 메모리를 갱신한다.

$$\hat{l}_i = \left[\hat{w}_i^{(m)} - \sum_{k=1}^4 \hat{p}_{i,k} \hat{l}_i^{(m-k)} \right] \left(1 - \sum_{k=1}^4 \hat{p}_{i,k} \right), \quad i = 1, \dots, 10. \quad (1)$$

여기서 \hat{l}_i , \hat{w}_i , $\hat{p}_{i,k}$ 는 각각 코드정보, 복구된 LSF 값, 이전 프레임의 예측 필터의 계수이다.

그림 1부터 2는 한 개의 프레임 에러가 발생된 후, MA 필터의 메모리 불일치로 인해 LSF 궤적 및 합성신호에 에러가 전파되는 영향을 보여준다.

그림 1은 한 개의 프레임 에러 발생시, 부/복호화기의 4차 MA 필터메모리가 서로 다르기 때문에 에러 영향이 지속되는 것을 보여준다. 결국, 그림 2와 같이 최종 합성된 음성 신호에서도 에러 전파가 지속됨을 알 수 있다. G.729에서 사용하는 PLC 방법은 과거 프레임의 정보를 이용하기 때문에 그 과정이 간단하다. 하지만 부/복호화기의 내부 메모리 불일치와 적응 코드북 또는 고정 코드북 중 한가지 성분으로만 구성되는 단순한 여기신호는 후속 프레임의 파라미터에 영향을 주고, 결국 합성신호의 음질을 저하시키는 요인이 된다. 이런 문제점을 최소화하기 위해 손실프레임은 물론 인접프레임의 음성신호 특성을 고려해야 한다. III 장에서는 음성 신호 특성에 따라 패킷 손실의 영향을 분석한다.

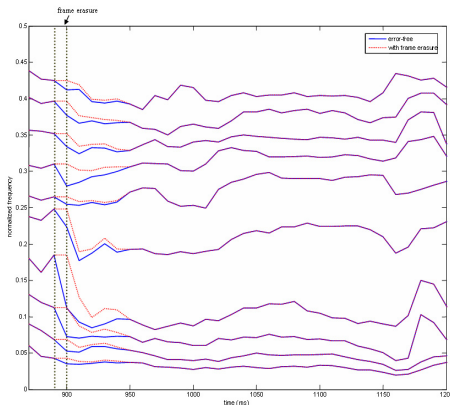


그림 1. 프레임 에러 발생시 LSF 궤적

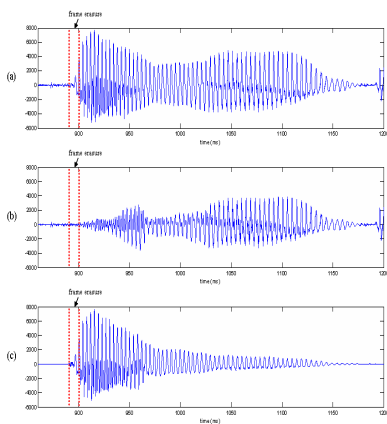


그림 2. 프레임 에러 발생시 합성신호:

- (a) 합성신호(에러 미발생),
- (b) 합성신호(에러 발생),
- (c) 합성신호의 오차

표 1. 음성 클래스 분류

class number	classes
0	silence
1	noise-likely unvoiced
2	unvoiced
3	voiced to unvoiced transition
4	onset (unvoiced to voiced transition)
5	non-stationary voiced
6	stationary voiced

III. 음성신호의 특성에 따른 패킷 손실 영향

음성신호의 특성을 고려하지 않고 손실 패킷을 복구하기 위해 과거 프레임의 정보를 사용하는 경우, 합성된 신호의 음질에 심각한 영향을 미칠 수 있다. 예를 들면, 무성음에서 유성음 천이되는 구간이 손실됐을 경우, 기존의 PLC 방법은 과거 프레임인 무성음 정보를 이용한다. 결국, 파라미터의 불일치로 인한 에러가 발생하고 손실된 프레임 이후에도 그 영향이 전파되어 음질이 심각하게 저하된다. 본 장에서는 표 1에 정리된 음성 분류 (speech classification) 방법^[6]과 NTT 한국어 음성 데이터베이스 (DB)를 사용해 스펙트럼 왜곡, 재동기화 시간 (Resynchronization time)^[9], PESQ 등으로 음성 특성에 따른 패킷 손실 영향을 분석한다.

3.1 Spectral Distortion(SD)

그림 3은 식 (2)와 같이 정의된 스펙트럼 왜곡 (SD)을 LPC, 여기신호, 그리고 합성 음성신호별로 각 음성클래스에 따라 보여준다. 손실된 프레임의 LSF를 과거 프레임의 LSF로 대체한 경우, 다음 프레임의 LSF를 예측기법으로 복호화하는 과정에서 4차 MA 예측 필터는 손실된 패킷의 LSF에 가장 높은 가중치를 적용하게 된다. 그렇기 때문에 그림 3의 LPC SD 궤적에서 볼 수 있듯이, 손실 프레임의 SD보다 이후 프레임의 SD가 높은 것을 알 수 있다. 또한, 안정적인 유성음구간 보다 잡음 성분이 많은 무성음, 유무성음의 천이구간이 다른 음성클래스보다 SD가 많음을 알 수 있다. 전체적으로는 무성음에서 유성음으로 천이되는 구간의 프레임 손실 영향이 가장 심각함을 볼 수 있다. 그리고 MA 예측 필터의 메모리 에러는 안정적인 유성음 구간에서조차 그 영향이 확연하다.

$$D_i = \frac{1}{F_s} \int_0^{F_s} [10 \cdot \log_{10}(P_i(f)) - 10 \cdot \log_{10}(\hat{P}_i(f))]^2 df, \quad (2)$$

여기서 F_s 는 8 kHz의 샘플링 주파수이고, $P_i(f)$ 와 $\hat{P}_i(f)$ 는 i 번째 프레임의 파워 스펙트럼이다.

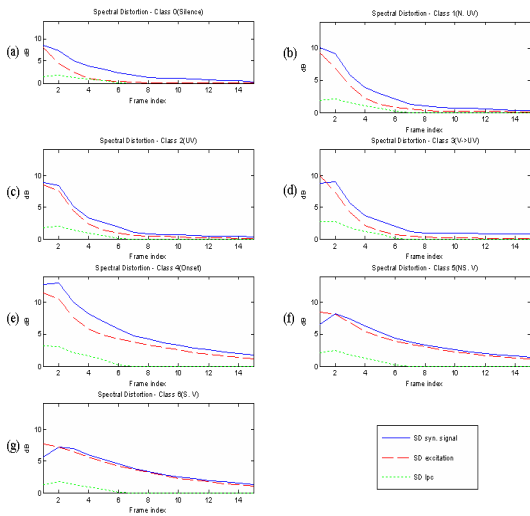


그림 3. 음성 클래스에 따른 스펙트럼 왜곡.
(a) class 0, (b) class 1, (c) class 2, (d) class 3, (e) class 4, (f) class 5, (g) class 6

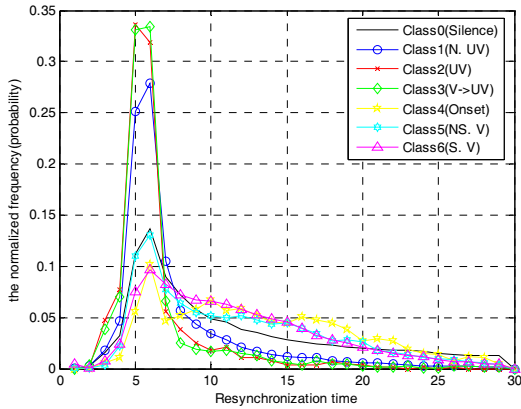


그림 4. 음성 클래스에 따른 재동기화 시간 분포

3.2 Resynchronization time

[9]에서 Rosenberg는 재동기화 시간(resynchronization time)을 이용하여 프레임 손실 후 에러가 지속되는 시간에 대해 연구했다. 프레임 손실에 대해 에러가 수렴하는 정도를 명확히 정의하기 어렵지만, 본 논문에서는 [9]에서 사용된 에너지차이를 정량화한 방법을 사용했다. 재동기화 시간은 식 (3)과 같이 손실된 프레임부터 segSNR(segmental Signal to Noise Ratio)의 추이를 관찰하여 segSNR값이 특정 문턱값 (threshold)을 넘는 순간의 프레임으로 정의

된다. 본 논문에서는 [9]에서처럼 20 dB를 문턱값으로 사용했다. 20 dB는 프레임 손실이 발생하지 않은 경우의 합성신호를 기준으로 프레임 손실의 오차신호 에너지가 기준의 1% 이하로 떨어지는 것을 의미한다.

$$segSNR = 10 \cdot \log_{10} \left\{ \frac{\sum_n x(n)^2}{\sum_n (x(n) - \hat{x}(n))^2} \right\} dB \quad n \in [1, F] \quad (3)$$

여기서 F 는 프레임 길이, $\hat{x}(n)$, $x(n)$ 는 각각 에러 유무에 따라 합성된 신호이다.

그림 4는 각 음성 클래스별로 측정된 재동기화 시간의 분포도이다. 무성음과 관련된 구간 (class 0~3)의 재동기화 시간은 주로 10번째 프레임 이내에 분포돼 있음을 볼 수 있다. 그러나 유성음 관련 구간, 특히 천이구간 (class 4)의 재동기화 시간은 손실 프레임 이후의 10번째 프레임 이상까지 상당 부분 분포되어 있다. 그림 4의 결과로부터 과거 프레임의 정보를 이용해 손실 프레임을 은닉했을 때, 무성음보다 유성음 구간에서 부/복호화기의 재동기화에 필요한 시간이 많이 소요됨을 알 수 있다. 이는 프레임 손실로 인한 에러가 무성음보다 유성음에서 오래 전파되는 것으로 해석할 수 있다.

3.3 PESQ

표 2는 음성 클래스의 분포와 프레임 손실 환경의 음질저하를 측정된 PESQ결과이다. 무성음 구간의 손실은 음질저하가 작지만, 유성음, 특히 천이구간에서 급격한 음질 저하가 일어남을 알 수 있다.

표 2의 결과와 SD, 재동기화 시간의 결과를 종합하여 패킷 손실 환경에서 음성의 특성에 따라 음질이 저하되는 경향을 파악할 수 있다. 일반적으로 음성신호의 특성을 고려하지 않고 과거 프레임의 정보만을 이용해 프레임 손실을 은닉했을 경우, 그 영향은 이후 프레임에 한동안 남아있게 된다. 본 논문에서는 보충실험을 통해, 패킷 손실 환경에서 LSF, 피치 지연, 고정 코드북 펄스, 이득값 등이 각 음성 클래스에 미치는 중요도를 표 3에 정리하였다. 표 3의 결과로 무성음에선 고정 코드북 펄스와 이득값이, 유성음에선 피치값의 영향이 상대적으로 많음을 알 수 있다. LPC 정보는 상대적인 차이는 있지만 모든 클래스에 대해 비교적 중요한 역할을 한다. IV장에서는 이 결과를 바탕으로 음성신호의 특성을 고려한 프레임 손실 은닉 알고리즘을 제안한다.

표 2. 각 음성 클래스의 분포 및 프레임 손실에 따른 PESQ

class number	Portion (%)	PESQ
error-free	X	3.654
0	33.3	3.652
1	16.2	3.638
2	0.80	3.640
3	3.00	3.632
4	3.40	3.584
5	12.0	3.624
6	31.3	3.626

표 3. 각 음성 클래스의 파라미터 중요도

class	Parameter contribution order
0	pitch < fixed-codebook < LSF < gain
1	pitch < fixed-codebook < gain < LSF
2	pitch < gain < fixed-codebook < LSF
3	pitch < fixed-codebook < LSF < gain
4	gain < LSF < fixed-codebook < pitch
5	gain < pitch < fixed-codebook < LSF
6	gain < LSF < fixed-codebook < pitch

IV. 보간법 기반의 패킷 손실 은닉 알고리즘

본 장에서는 III 장의 실험결과를 바탕으로 보간법 기반의 PLC 방법을 제안한다. 본 논문에서 제안하는 PLC 방법은 VoIP 환경을 고려해 수신자 기반의 방법으로 설정하고 송신단에서 추가 비트할당을 통해 음성 클래스 정보를 전송한다고 가정한다. 제안한 PLC 알고리즘은 VoIP에서 널리 사용되는 G.729 음성부호화기에 구현한다.

일반적으로 후속 프레임의 정보를 이용해 프레임 손실을 은닉하는 보간법 기반의 PLC 알고리즘은 G.729에서 쓰이는 반복법 기반의 프레임 손실 은닉법보다 성능이 좋다고 알려져 있다^[10]. 보간법에 이용되는 후속 프레임 정보를 얻기 위해선 추가 지연 시간이 필요하다는 단점이 있다. 하지만, VoIP와 같은 패킷 네트워크 시스템의 수신단에는 프레임 지연과 지터(jitter) 현상을 대비해 플레이아웃(playout) 버퍼가 필수적으로 존재하기 때문에 추가 지연 시간 없이 플레이아웃 버퍼에 저장된 후속 프레임을 이용할 수 있음을 가정하였다.

G.729와 같은 예측기법 음성부호화기는 LSP와 고정 코드북 이득값의 예측 오차값을 양자화하여 전송하기 때문에 파라미터값에 보간법을 직접 적용하기 어렵다. 또한, 손실 패킷의 후속 프레임들의

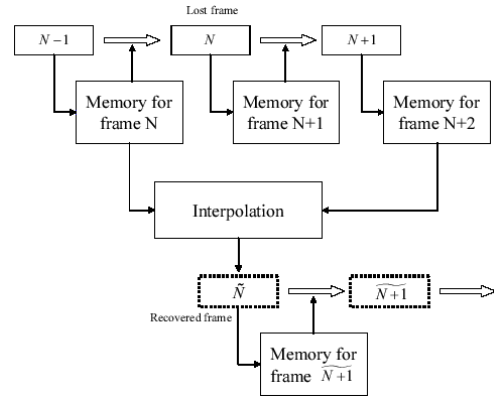


그림 5. 제안한 보간법 기반의 프레임 손실 은닉법

복호화를 위한 필터 메모리를 갱신하며 발생하는 부/복호화기간의 메모리 불일치 문제도 기존 보간법 기반 PLC 방법의 문제점이다. 본 논문에서는 손실 프레임 및 인접 프레임의 음성신호 특성을 이용한 선택적 보간법을 적용해 손실된 파라미터를 복구하고, 후속 프레임 복호화를 고려하여 효율적으로 메모리를 갱신한다. 그림 5는 본 논문에서 제안한 보간법 기반의 프레임 손실 은닉방법의 개념도이다. 복호화에 이용되는 파라미터 중 과거 프레임과의 상관관계가 없는 고정 코드북 펄스 정보는 보간법 적용이 어렵기 때문에, 기존 G.729와 마찬가지로 랜덤 펄스로 재생한다. 4.1절부터 4.3절까지는 보간법을 사용하는 각 파라미터의 손실 은닉법에 대해 기술한다.

4.1 LPC 파라미터 복구

G.729는 4차 MA 예측 필터를 사용해서 LPC 정보를 양자화하고 전송된 코드북 정보를 이용해 해당 프레임 및 후속 프레임의 LPC값을 복원한다. 부호화기의 양자화 과정에서는 2개의 MA 예측 필터 중, WMSE(weighted mean square error)를 최소화하는 MA 예측 필터를 사용하고, 사용된 필터의 인덱스를 복호화기에 전송한다.

기존 반복법 기반의 프레임 손실 은닉법은 후속 프레임 복호화를 위한 필터 메모리를 갱신하기 위해 이전 프레임에서 사용된 예측 필터를 사용한다. 하지만, 무성음 구간의 LPC 정보의 변화 추이는 유성음의 경우보다 심하기 때문에 프레임이 손실됐을 경우, 부호화기에서 사용된 예측 필터와 다른 예측 필터로 메모리 갱신을 하게 된다. G.729에서 사용하는 2개의 예측 필터는 서로 다른 계수값을 갖고 있기 때문에 부호화기와 다른 예측 필터를 사용해

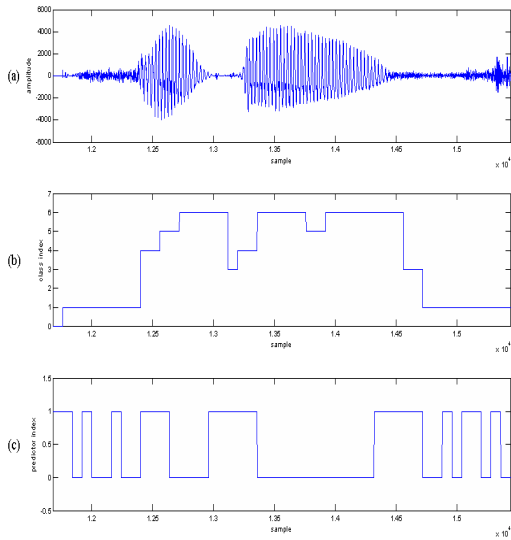


그림 6. MA 예측 필터의 변화 추이:
 (a) 음성신호,
 (b) 클래스 인덱스,
 (c) 예측 필터 인덱스

메모리를 업데이트할 경우, 후속 프레임의 LPC 값은 크게 달라지게 된다. 그림 6은 LSF를 예측하는 필터의 종류가 해당 음성 클래스의 변화 추이에 따라 빈번하게 변화하고 있음을 보여준다. 본 논문에서는 부호화단에서 해당 프레임의 LSF 예측 필터 인덱스 정보를 표현하는 1 비트를 추가로 전송해 복호화기에서 올바른 예측 필터 메모리 갱신이 이루어질 수 있도록 한다. 그리고, 현재 손실 프레임의 LSP값은 과거 프레임과 후속 프레임의 LSP값을 선형 보간해 복구한다.

4.2 피치 지연값 복구

유성음 구간의 피치 지연값은 단조로운 변이를 보이기 때문에 인접 프레임간 유사성이 높지만, 무성음 구간에서는 그 변화가 심하다. 따라서 G.729와 같이 손실된 프레임의 피치값을 과거 피치 지연값으로 대체할 경우, 무성음에서 유성음으로 천이되는 구간은 과거 무성음 프레임의 피치값을 유성음 성분이 포함된 손실 프레임의 피치로 사용하게 되어 여기 신호에 심각한 에러 전파를 야기시킨다. 또한, 과거 및 후속 프레임의 피치값을 보간하여 손실 프레임의 피치값을 복구하더라도, 보간에 사용되는 피치값들의 차이가 클 경우, 그 성능을 보장할 수 없다. 따라서, 본 논문에서는 손실 프레임 및 인접 프레임의 음성 클래스 특성을 고려하여 보간법을 선택적으로 적용하여 피치값을 복구한다.

표 4. 음성 클래스 정보에 따른 피치 지연 보간의 가중치

(N-1)th	N-th	(N+1)th	Weighting
Voiced	Voiced	Voiced	0.5, 0.5
Onset	Onset	Voiced	0.25, 0.75
Unvoiced	Onset	Onset	
Onset	Voiced	Voiced	

우선, 선택적 보간법을 효율적으로 적용하기 위해 III장에서 사용한 음성클래스를 무성음 (class 0~3), 천이구간 (class 4), 유성음 (class5~6)과 같이 3개로 단순화한다. 해당 프레임의 음성 클래스 정보는 4.1절에서처럼 후속 프레임과 함께 전송되어지는 것으로 가정한다. 본 논문에서는 표 4에 정리된 것과

같이 연속된 프레임들의 음성 클래스를 고려하여 선택적인 보간법을 적용한다. 식 (4)와 같이 유성음 구간이 연속되는 경우는 과거 및 후속 프레임에 동일한 가중치를, 무성음에서 유성음으로의 천이구간에는 후속 프레임들이 한동안 유성음으로 지속되기 때문에 과거 프레임보다 후속 프레임 피치값에 높은 가중치를 적용한다. 표 4에 정리되지 않은 무성음 구간의 피치값 복구는 기존 G.729의 방법을 사용한다.

$$T0_{cur} = \begin{cases} 0.5 \times T0_{prev} + 0.5 \times T0_{next}, & \text{when voiced} \\ 0.25 \times T0_{prev} + 0.75 \times T0_{next}, & \text{when transition} \end{cases} \quad (4)$$

여기서, $T0_{cur}$, $T0_{prev}$, $T0_{next}$ 는 각각 현재, 과거, 후속 프레임의 피치값이다.

4.3 이득값 복구

G.729는 켈레 구조(conjugate structure) 코드북을 이용해 피치 이득값과 고정 코드북 이득값을 양자화한다. 프레임이 손실 됐을 경우, 과거 프레임의 이득값을 감쇠시켜 현재 프레임의 이득값으로 사용하는데, 피치 이득값은 0.9, 고정 코드북 이득값은 0.98의 비율로 감쇠한다. 그리고, 후속 프레임의 고정 코드북 이득값 복호화를 위해 식 (5)와 같이 현재 손실 프레임의 4차 예측 필터의 평균 에너지를 4 dB 감쇠시켜 메모리를 갱신한다. 이때, 4차 예측 필터를 사용하는 LSF의 경우와 마찬가지로 고정 코드북 이득값의 예측 필터 메모리에도 에러의 영향이 남게 된다. 또한, 이 과정에서 랜덤으로 발생된 고정 코드북 펄스의 영향은 전혀 고려되지 않는다.

$$\gamma^{(m)} = \frac{G_c^{(m)}}{10^{(\bar{E}^{(m)} + \bar{E} - E^{(m)})/20}} \quad (5)$$

여기서 $\gamma^{(m)}$, $G_c^{(m)}$, $\bar{E}^{(m)}$, \bar{E} , $E^{(m)}$ 은 각각 m 번째 프레임의 고정 코드북 이득값의 코드, 고정 코드북 이득값, 예측 오류 에너지의 평균값, 평균 에너지, 여기신호 에너지이다.

본 논문에서는 이득값을 과거 및 후속 프레임의 이득값을 이용해 선형 보간한 후, 고정 코드북 이득값의 경우는 랜덤 발생된 코드북 펄스 에너지의 영향을 반영하여, 새롭게 γ 를 구하고 이를 이용해 4차 예측 필터의 메모리를 효과적으로 갱신한다. 이 과정의 개념도는 그림 7에 나타나 있다.

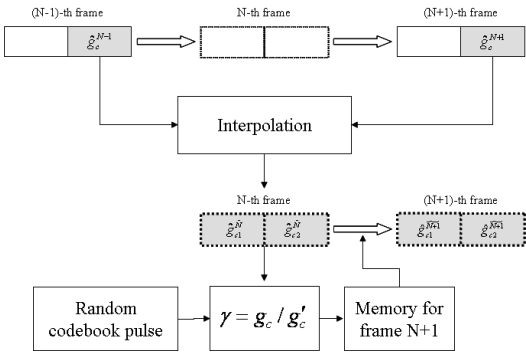


그림 7. 고정 코드북 이득값의 선형 보간 및 메모리 갱신

V. 성능 평가

본 장에서는 제안된 보간법 기반 PLC 알고리즘의 성능을 NTT 한국어 음성 DB를 사용해 반복법 기반의 G.729 프레임 손실 은닉 방법과 비교한다. NTT DB는 조용한 환경에서 8초 길이의 문장 12개를 남녀 각각 4명이 발음한 96개의 음성 샘플로 구성되었다. 본 실험에서는 1 프레임/패킷을 가정했고, 성능 평가의 척도로는 스펙트럼 왜곡과 PESQ를 사용했으며 프레임 에러 발생비율 (FER: Frame Erasure Rate)은 3~30%까지로 설정했다.

그림 8과 표 5는 기존 G.729 PLC 방법과 본 논문에서 제안한 이득값의 선형 보간법 및 예측 필터 메모리 갱신 방법에 대한 성능을 보여준다. 이 실험에서는 이득값을 제외한 나머지 파라미터는 기존 G.729 PLC방법으로 복구하였다. 선형 보간으로 이득값을 복구하는 방법은 기존 PLC의 과거 이득값의 감쇠법보다 SD 및 PESQ에서 좋은 성능을 보인다. 또한, 선형 보간된 이득값으로 후속 프레임의 이득값 복호화를 위한 예측 필터 메모리를 갱신하기 위해 랜덤 코드북 펄스 에너지를 고려한 본 논문의 방법은 기존 G.729의 메모리 갱신 방법보다

표 5. 제안한 패킷 손실 은닉 알고리즘의 PESQ 값

FER (%)	G.729 PLC	메모리갱신 (G.729)	메모리갱신 (Proposed)	Proposed PLC
0	3.654	3.654	3.654	3.654
3	3.328	3.351	3.365	3.386
5	3.167	3.199	3.218	3.244
8	2.982	3.019	3.043	3.073
10	2.880	2.919	2.946	2.975
20	2.528	2.579	2.617	2.643
30	2.292	2.351	2.396	2.412

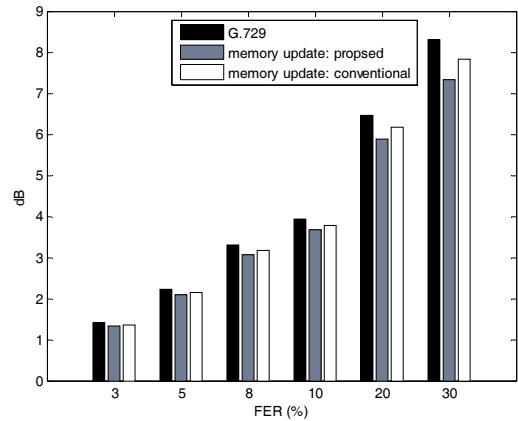


그림 8. 이득값 보간의 스펙트럼 왜곡 (SD)

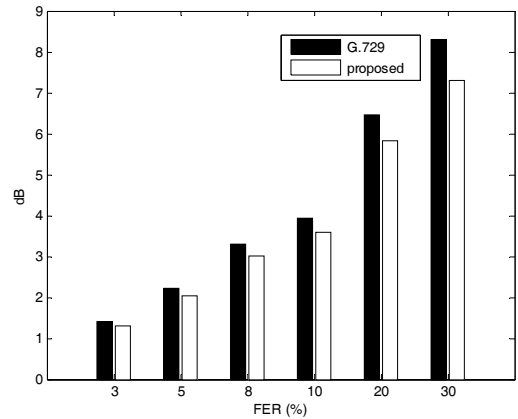


그림 9. 제안한 PLC알고리즘의 SD

좋은 성능을 보임을 알 수 있다. 성능향상의 원인은 크게 두 가지로 요약할 수 있다. 첫째로는 과거 및 후속 프레임의 이득값을 이용한 선형보간이 과거값만을 이용하는 반복법보다 원래의 이득값에 가깝게 복구하기 때문이다. 둘째로는 후속 프레임 이득값의 복호화를 위한 예측 필터의 메모리를 현재 프레임의 코드북 에너지를 고려하여 효과적으로 갱신하

기 때문이다. 이러한 성능향상은 FER이 높아질수록 확연해진다. 또한, LPC 정보 및 피치값의 선형 보간 역시 기존 G.729 PLC 방법보다 좋은 성능을 보였다.

그림 9는 선형 보간법 기반으로 복구된 파라미터를 모두 종합하여 제안한 PLC 알고리즘의 스펙트럼 왜곡이다. 이득값의 경우와 마찬가지로 기존 G.729 PLC 방법에 비해 제안한 PLC 방법의 성능이 향상됨을 알 수 있다. 표 5의 PESQ 값 역시 스펙트럼 왜곡의 성능과 일치하는 결과를 보인다. 제안한 알고리즘은 프레임 손실비율이 높아질수록 성능향상이 두드러진다.

VI. 결론

예측기법을 사용하는 대부분의 저전송률 음성부호화기는 패킷 손실이 발생했을 경우 복호화단에서 과거 정보를 이용하여 현재 손실 패킷을 은닉한다. 이 과정에서 부/복호화기간의 예측 필터의 메모리가 불일치하는 문제가 발생하고, 이로 인한 에러 영향은 오랜 시간동안 후속 패킷에 전파된다. 본 논문에서는 손실 패킷의 음성신호 특성을 고려한 PLC 알고리즘을 제안하였다. 이를 위해 음성 신호의 특성에 따라 패킷 손실 환경에서 에러가 전파되는 영향과 파라미터 중요도를 정리하였다. 이를 기반으로 본 논문에서 제안한 PLC 알고리즘은 음성 신호의 특성을 고려해 선형 보간법으로 파라미터를 복구하고, 후속 프레임들의 복호화를 위해 예측 필터의 메모리를 효과적으로 갱신하는 방법을 사용했다. 기존 G.729 PLC 방법과의 성능평가를 통해, 본 논문에서 제안한 PLC 알고리즘이 다양한 FER 환경에서 성능을 향상시킴을 확인했다.

제안된 알고리즘은 기존 방법에 비해 성능을 향상시켰지만, 파라미터의 보간법 적용시 세밀한 튜닝과 burst packet loss와 같은 다양한 환경에서의 성능평가 등 추후 연구가 필요하다. 그리고, 예측 기법을 사용하는 다양한 음성 부호화기에 대해 제안한 PLC 알고리즘을 적용하는 실험 역시 필요하다. 본 논문에서는 G.729를 선택하여 제안한 알고리즘의 성능을 비교했으나, ITU-T G.723.1 또는 AMR-NB와 같은 음성부호화기에도 본 알고리즘을 적용하여 그 성능을 검증할 필요가 있다.

참 고 문 헌

[1] B. W. Wah, et al, "LSP-based multi-

ple-description coding for real-time low bit-rate voice over IP," *IEEE Transaction on Multimedia*, vol. 7, no. 1, pp. 167-178, Feb. 2005.

[2] T. J. Kostas, et al, "Real-time voice over packet-switched networks," *IEEE Network*, vol. 12, no.1, pp. 18-27, Jan.-Feb. 1998

[3] J. C. Bolot, et al, "Adaptive FEC-Based Error Control for Internet Telephony," *Proceeding INFOCOM 99*, vol. 3, pp. 21-25, March 1999.

[4] K. Cluver, et al, "Reconstruction of missing speech frames using sub-band excitation," in *Proc. IEEE Int. Symp. Time-Frequency and Time-Scale Analysis*, pp. 277-280, June 1996.

[5] F. Mertz, et al, "Voicing-controlled frame loss concealment for adaptive multi-rate (AMR) speech frames voice-over-IP," in *Proc. Eurospeech 2003*, pp. 1077-1080, Sep., 2003.

[6] M. K. Lee, et al, "A Quality Assessment Method of Voice Communication using Packet Loss Information," *IEEE Transactions on Speech and Audio Processing*, submitted Feb., 2006.

[7] ITU-T Rec. G.729, "Coding of Speech at 8 kbit/s CS-ACELP Speech Coder," 1996.

[8] ITU-T Rec. P.862, "Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs," Feb. 2001.

[9] J. Rosenberg, "G.729 Error Recovery for Internet Telephony," Project Report, Columbia University, May 1997.

[10] J. Wang, et al, "Parameter interpolation to enhance the frame erasure robustness of CELP coders in packet networks," in *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, vol.2, pp. 745-748, May, 2001.

-
- 윤 성 완 (Sung-Wan Youn) 정회원
한국통신학회 논문지 제26권 1B호 참고
- 강 흥 구 (Hong-Goo Kang) 정회원
한국통신학회 논문지 제30권 1C호 참고
- 윤 대 희 (Dae-Hee Youn) 종신회원
한국통신학회 논문지 제26권 1B호 참고