

VoIP용 음성부호화기를 위한 패킷 손실 은닉 알고리즘에 대한 연구

이승원, 김시호, 유승형, 배건성
경북대학교 전자·전기 공학부

A Study on the Packet Loss Concealment Algorithm for Speech Coders in VoIP System

Seung Won Lee, Si Ho Kim, Seung Hyung Yu, Keun Sung Bae
School of Electronics & Electrical Engineering, Kyungpook National University, Korea
lsw@mir.knu.ac.kr, ksbae@mir.knu.ac.kr

요 약

본 연구에서는 VoIP와 같은 패킷망에서 G.729 CS-ACELP 음성부호화기에 패킷 손실 은닉 알고리즘을 적용하여, 패킷 손실로 인한 음질 저하의 완화에 관한 실험을 수행하였다. 패킷 손실 은닉은 수신된 패킷으로부터 복호된 파형을 저장해두었다가, 손실이 발생하면 피치 동기가 맞도록 선택한 파형을 손실된 패킷 자리에 대체하는 방법과 연속적인 손실 이후에 음성부호화기의 메모리를 초기화하는 방법에 기반하고 있다. 실제 VoIP 통화 실험에서 측정된 패킷 손실 분포에 대해 패킷 손실로 인한 음질 저하를 완화할 수 있음을 확인하였다.

I. 서 론

패킷망을 통해 실시간으로 음성신호를 전송하는 VoIP(Voice over Internet Protocol) 시스템에서 패킷 손실이 발생할 경우 이는 곧바로 음질 저하의 원인이 된다. 패킷의 재전송을 통한 패킷 복구 방법은 지나치게 긴 지연을 초래하여 실시간의 통신의 의미를 잃게 되며, 손실된 패킷을 0의 값을 삽입하여 대체 할 경우에는 손실되는 패킷의 수가 증가할수록 수신단에서 듣게되는 음질은 급격하게 저하된다. 따라서 패킷의 손실이 필연적으로 따르는 VoIP 시스템에서 원활한 음성통신을 위해서는 손실된 패킷에 대한 수신단에서의 적절한 대응이 필요하다. 본 연구에서는 이와 같은 문제에 대해 연구의 필요성을 느끼고, 실제 VoIP 통화 환경에서 발생하는 패킷 손실로 인한 음질 저하를 완화하기 위해 손실 은닉 알고리즘을 갖는 음성부호화기에 대해 실험하고, 통화 실험의 결과를 제시하였다. 실험에 사용한 음성부호화기는 ITU-T(International Telecommuni-

cations Union-Telecommunications standardization sector)의 G.729[1] 방식이며, 손실 은닉 방식은 이전의 성공적으로 도착한 패킷이 갖는 정보를 이용해서 복호화한 파형을 손실된 패킷에 대한 합성신호를 대체할 때 사용하는 방법에 기반하고 있다. 또한 CELP(Code Excited Linear Prediction)[2] 계열의 음성부호화기가 같은 이전 프레임에의 분석 의존성이 패킷 손실로 인한 오차 전파를 유발하는 것을 방지하기 위해 연속된 패킷 손실이 발생할 경우 음성부호화기의 메모리를 초기화하는 방법을 적용하였다. 적용된 손실 은닉 방식은 수신단 기반의 손실 은닉 기법으로 음성부호화기의 디코더 부분에만 적용되기 때문에 기존의 음성부호화기에 비해 연산량이나 지연을 크게 증가시키지 않는다.

본 논문의 구성은 다음과 같다. 2장에서는 VoIP 시스템에 사용될 수 있는 G.729 음성부호화 알고리즘에 대한 간단한 설명과 함께, 하나의 손실된 패킷이 이후에 발생하는 패킷에 어떤 영향을 주는가에 대해 살펴본다. 그리고 3장에서는 본 연구에 적용된 손실 은닉 방법과 실험을 통해 관찰한 손실 분포에 대해 설명하고, 실험 결과 및 분석을 4장에서 제시한다. 마지막으로 5장에서는 결론을 맺고 향후 연구방향을 제시한다.

II. G.729 음성부호화기

2.1 G.729 음성부호화 알고리즘

G.729 음성부호화기는 CS-ACELP(Conjugate Structure - Algebraic Code Excited Linear Prediction)[1]로 알려진 ITU-T에 의해 제정된 권고안으로, 이동통신용으로 고안된 8 kbps의 전송율을 갖는 음성부호화기이다. G.729-음성부호화기는 그림 1과 같이 음성신호를 적용 코드북 벡터(adaptive codebook vector)와 고정

코드북 벡터(fixed or algebraic codebook vector)의 합으로 구성되는 여기신호가 사람의 성도특성을 모델링하는 10차의 디지털 all-pole 필터를 통과한 출력으로 합성하는 방식이다. 이때 여기신호는 합성에 의한 분석 방법(Analysis by synthesis)을 통해 청각적으로 입력 음성신호와 가장 유사한 특성을 갖는 신호를 합성하기 위한 것을 선택하고, 합성필터의 계수는 10차의 선형 예측 분석을 통해 추정한다. 합성 필터의 출력은 좀 더 자연스러운 음질을 만들기 위해 후처리 필터를 통과하게 된다. 이들 연산은 5 msec 길이의 서브프레임과 10 msec 길이의 프레임 단위로 수행되며, VoIP 시스템에 적용하기 위해 10 msec에 해당하는 부호화된 비트열을 하나의 패킷으로 만들어 1초당 100개의 패킷을 발생하게 된다.

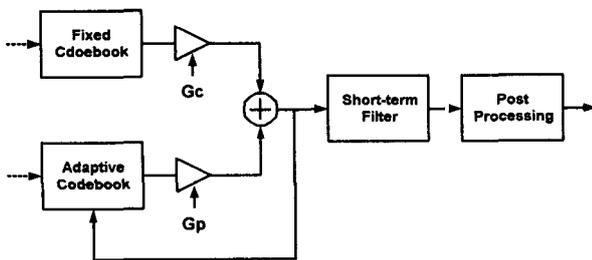


그림 1. G.729 음성부호화기의 음성합성 모델

G.729 음성부호화기는 선택 사항으로 Annex B[3]에서 정의하고 있는 음성활성도 검출기(Voice Activity Detector)와 잡음 생성(Comfort Noise Generation) 방식을 이용한 묵음 압축(Silence Compression) 기능을 가질 수 있다. 이 중에서 음성활성도 검출기(VAD)는 스펙트럼 왜곡, 전대역 에너지, 저대역 에너지, 영교차율과 이들의 평균값들을 사용하여 음성 구간 및 비음성 구간을 검출하여 음성 압축 효율을 높일 수 있을 뿐만 아니라, 패킷 손실이 발생한 경우 은닉 알고리즘의 적용에 참조될 수 있다.

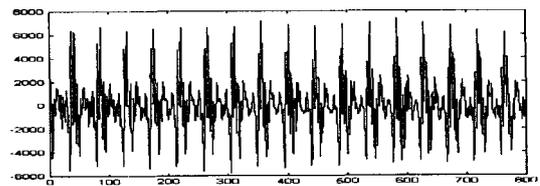
2.2 패킷 손실의 영향

G.729 음성부호화기를 적용한 VoIP 실험에서 하나의 패킷 손실이 발생하면, 그 영향은 손실된 패킷뿐만 아니라 뒤이어 생성되는 패킷에도 영향을 미치게 된다 [4]. 이는 CELP 방식에서 분석하는 특징 파라미터들이 현재 프레임의 분석 결과뿐만 아니라 이전 프레임에서 분석한 정보도 고려하기 때문이다. G.729 음성부호화 알고리즘에서 사용하는 이전 프레임의 정보들은 다음과 같다[1].

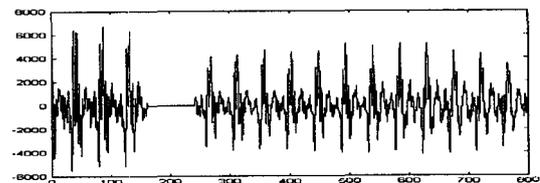
- 선형 예측(Linear Prediction) 분석 파라미터
- 고정 코드북 및 적응 코드북의 이득
- 이전의 고정 코드북 벡터의 에너지를 사용하는 이득 예측기 메모리
- LSF(Line Spectrum Frequency) 분석을 위한 4차의 MA(Moving Average) predictor 메모리

G.729 방식으로 부호화된 음성패킷의 손실이 발생할 경우 손실된 패킷이 다음 프레임에 어떤 영향을 미치는가를 알아보기 위해, 임의의 프레임을 손실로 처리하고 뒤이어 나타나는 프레임의 복호된 신호를 손실이 없는 경우와 비교해 보았다. 그림 2는 하나의 패킷 손실이 일어난 경우에 이를 '0' 삽입 방법을 통해 합성음을 복원한 경우와((b)), G.729 알고리즘에 내장된 프레임 손실 처리 알고리즘에 의한 은닉 손실 은닉 결과((c))를 손실이 없는 경우((a))와 비교한 것이다. 그림 2의 (a)에서 패킷 손실이 전혀 없는 경우에 G.729 음성부호화기를 이용한 합성음의 파형은 전반적으로 평탄한 포락선을 가지는 것을 알 수 있다. 그림 2의 (b)는 161에서 240까지의 샘플에 해당하는 한 프레임의 손실이 일어난 경우, 이를 0으로 대체한 그림을 나타내고 있다. (b)와 (c)에서 241 이후의 샘플은 정상적으로 수신된 패킷을 사용하여 복호한 신호이지만 (a)의 손실이 없는 경우와 비교해 볼 때, 신호의 전체 포락선에서 차이가 있음을 알 수 있고, 실제로 샘플단위로 비교해 볼 때도 값이 차이가 난다. 이는 앞서 설명한 것과 같이 현재 프레임에서 분석된 특징 파라미터 및 음성부호화 알고리즘의 내부 메모리가 이후 프레임 분석에도 영향을 미치기 때문이다.

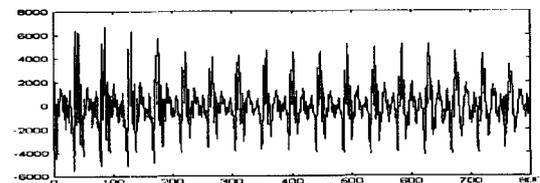
그림 3은 패킷 손실이 없는 경우의 합성음과 패킷 손실이 있는 경우의 합성음을 이용해서 구한 평균 자승 오차(Mean Square Error)를 나타내고 있다. 손실된 패킷으로 인한 오차를 직접적인 오차(A 영역: 160에서 240까지의 0 대체 값과의 오차)와 음성부호화기의 잘못된 메모리 값으로 인한 간접적인 오차(B 영역)로 나누어 비교해 볼 때, 직접적인 오차뿐만 아니라 오차의 전파로 인한 간접적인 오차 또한 상당히 큰 값을 가지는 사실을 알 수 있다. 이같은 두 가지 형태의 오차는 합성음의 음질을 저하시키며, VoIP 시스템에서 해결해야 할 문제이다.



(a) 패킷 손실이 없는 경우



(b) 손실된 패킷을 0으로 대체한 경우



(c) G.729에 내장된 은닉 알고리즘을 적용한 경우
그림 2. G.729 음성부호화기의 에러 전파

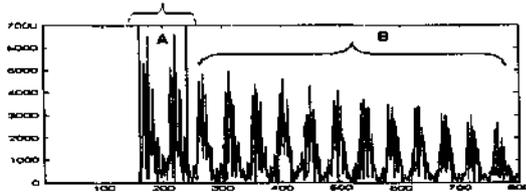


그림 3. 패킷 손실로 인한 오차(A) 및 오차의 전파(B)

III. 적용된 손실 은닉 기법

본 연구에서 실험한 손실 은닉 알고리즘은 CELP 계열의 음성부호화기가 갖는 패킷 손실의 영향으로 인한 에러의 전파를 감소시키는 것이다. 정상적으로 패킷이 수신되다가 손실되는 경우에는 이전의 복호된 파형을 이용하여 불연속점이 발생하지 않도록 동기를 맞추어서 대입하는 방법을 사용한다[5]. 이때 피치 동기가 맞도록 파형을 대입할 때 필요한 피치 정보는 마지막으로 수신된 패킷에서 추출된 피치 정보를 이용한다. 그리고 연속적인 패킷 손실 이후 정상적인 패킷이 수신되면 음성부호화기 내부에 사용되는 파라미터의 메모리 값을 초기화함으로써 이전의 손실된 패킷들에 대한 은닉 기법이 초래하는 오차의 전파를 최소화한다. 이 과정을 정리하면 다음과 같다.

(1) 정상적으로 패킷이 수신된 경우: 수신단은 복호된 음성 파형을 일정 길이의 순환 버퍼에 저장하게 된다. 패킷이 계속해서 수신될 경우 이 버퍼의 기록값은 갱신되고, 패킷 손실이 발생할 경우 버퍼에 저장된 값을 사용하게 된다. 이때 버퍼의 길이는 G.729에서 개루프 피치 탐색 범위의 최대값과 오버랩 합성시의 중복구간 길이를 고려하여 새롭게 정의한 최대 허용 피치(140 샘플)의 2.25배 길이에 해당하는 315 샘플로 설정하였다. 복호된 파형 이외에 피치 검색의 중복성을 제거하기 위해 현재 프레임에서 분석한 피치 값 또한 저장한다.

(2) 패킷이 수신되지 않은 경우: 버퍼에 저장된 파형을 이용하여 피치 동기에 맞도록 파형을 추출하여 손실된 패킷을 대체한다. 이때, 버퍼에 저장된 파형은 단순히 동기에 맞추어 반복되는 것이 아니라, 이후의 연속적인 패킷 손실이 발생할 경우를 고려하여, 감쇠 계수를 적용하여 차츰 크기를 줄이고, 단순한 반복으로 인한 buzz를 방지하기 위해 랜덤한 잡음을 적당하게 첨가한다.

(3) 패킷 손실 이후 정상적인 패킷 수신인 경우: 이전 패킷이 손실되고 현재 패킷이 정상적으로 수신된 경우에는 수신된 패킷에서 복호화한 음성신호를 바로 실제 합성음으로 사용하지는 않는다. 이는 복호된 신호가 이전의 은닉 알고리즘을 통해 합성된 합성음과 자연스럽게 연결되지 못하는 경우가 발생할 수가 있고, 이로 인한 불연속적인 재생음이 귀에 거슬리는 소리를 낼 수 있기 때문이다. 이러한 문제를 해결하기 위해 간단한 overlap-and-add 방법으로 이전의 합성된 신호와 현재

패킷에서 복호된 신호를 자연스럽게 연결시킬 수 있게 된다. 또한 음성부호화기 내부의 파라미터 메모리를 초기화함으로써, 손실 은닉 기법이 초래하는 오차가 전파되는 것을 최소화한다.

IV. 실험 결과 및 분석

4.1 VoIP 통화시의 패킷 손실 분포 경향

실험에 사용한 패킷 손실 경향은 실제 VoIP 통화 실험을 통해 측정된 데이터를 사용하였으며, VoIP 통화 실험은 2002년 4월에 대구와 제주간에 OpenPhone 프로그램[6]을 이용하여 수행하였다. 음성부호화기는 64 kbps의 전송속도를 갖는 PCM(Pulse Coded Modulation) 방식을 사용하였으며, 20 msec의 프레임을 하나의 패킷으로 만들어 초당 50개의 패킷을 발생시키고, 약 20여분의 음성데이터에 대해 통화 실험을 수행하면서 얻은 패킷 손실 경향을 측정하였다. 실험결과 약 6%의 평균 패킷 손실율을 가졌으며, 대부분의 패킷 손실이 하나의 패킷에 대해 발생하였고, 연속적인 패킷 손실은 전체 손실 중 일부에 지나지 않았다. PCM음성부호화기에서 측정된 패킷 손실 경향은 표 1과 같다.

표 1. 패킷 손실 경향 분석

연속 손실 패킷 수	백분율
1	72.95 %
2	23.60 %
3	3.44 %
4	0.01 %
5 이상	0.00 %
계	100.0 %

4.2 실험 결과 분석

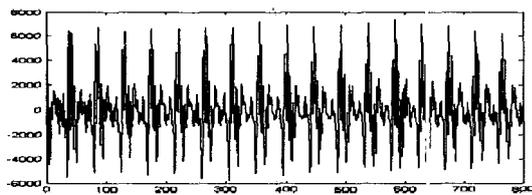
본 실험에서는 표 1과 같이 구한 실제 VoIP 통화 환경에서의 패킷 손실 분포경향에 대해, G.729 음성부호화기를 사용하는 경우, 패킷 손실 은닉 알고리즘이 적용된 경우와 그렇지 않은 경우에 대해 실험하였다. 패킷 손실 은닉 알고리즘은 G.729 음성부호화기에 내장된 방식과 본 논문의 3.1 절에 설명한 방법을 각각 적용하여 실험하였다.

먼저 패킷 손실 은닉 기법을 적용하지 않은 경우에는 손실이 일어난 패킷에 대해 0을 삽입하기 때문에 듣기에 상당히 거북한 느낌을 받게 된다. 따라서 실제 VoIP 시스템에서는 반드시 손실된 패킷에 대한 처리가 필요하다는 것을 알 수 있었다.

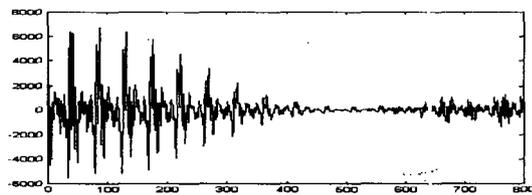
G.729에 내장된 손실 은닉 알고리즘을 적용한 경우에는 표 1의 패킷 손실 분포에서는 패킷의 손실이 그다지 많은 편은 아니어서 손실이 없는 경우와 거의 차이를 느끼지 못할 정도의 합성음을 만들어 내는 것을 확인할 수 있었다. 하지만 인위적으로 만든 연속적(4

개 이상의 연속된 패킷 손실)인 패킷 손실에 대해서는 손실의 수가 증가할수록 음질의 저하가 급격하게 일어남을 확인 할 수 있다. 이는 연속적인 손실로 인한 부적당한 값을 갖는 메모리의 누적이 상당 시간 지속되어 나타나는 현상임을 2.2절에서의 분석을 통해 예상할 수 있었다.

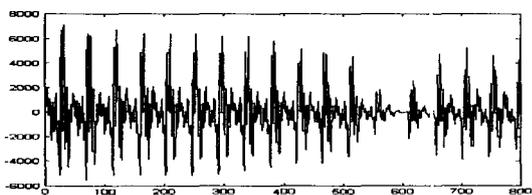
마지막으로 3.1 절에 설명된 방법으로 손실된 패킷에 대한 은닉 처리를 적용한 경우, 표 1의 손실 조건에서는 G.729에 내장된 처리 방법과 마찬가지로 손실이 없는 경우와 거의 차이를 느끼지 못할 정도의 음질을 나타내는 것을 확인하였다. 어느 정도의 패킷 손실 환경에서는 간단한 손실 은닉 알고리즘의 적용으로 합성음의 음질 열화를 상당히 완화시킬 수 있음을 확인 할 수 있었다. 그리고, 인위적으로 만든 연속적인 패킷 손실 환경에서는 논문에서 실험한 방법이 G.729에 내장된 은닉 방법보다 우수한 처리 결과를 보임을 확인하였다. 그림 3은 인위적으로 발생시킨 연속적인 패킷 손실에 대해 (a)손실이 없는 경우의 합성음과 (b)G.729에 내장된 기법을 적용한 결과, 그리고 (c)과형 대체 기법 및 음성부호화기의 메모리를 초기화하는 은닉 기법을 적용한 결과를 나타내고 있다. 그림 (b)와 (c)를 비교할 때, (b)는 잘못된 메모리 값이 지속되고, 이득을 지나치게 감소시켜서 연속된 패킷 손실이 발생하는 경우에는 과형이 급격하게 왜곡되어 원신호와와는 상당한 차이를 나타내지만, (c)의 경우에는 적당한 과형 대체 및 성공적인 수신 이후의 메모리 초기화로 보다 원음성에 가까운 합성음을 얻을 수 있음을 알 수 있었다. 하지만 안정된 유성음이 연속적으로 나타나는 구간에 대해 실험 결과가 분석된 것이어서 G.729에 내장된 손실 은닉 처리 기법과의 우수성을 성급히 판단할 수는 없다.



(a) 패킷 손실이 없는 경우



(b) G.729에 내장된 손실 은닉 처리 결과



(c) 과형 대체 및 메모리 초기화 기법 처리결과
그림 4. 연속된 패킷 손실에 대한 은닉 처리 결과

IV. 결론 및 향후 연구 방향

본 연구에서는 VoIP에 사용될 수 있는 G.729 CS-ACELP 음성부호화기를 이용한 패킷 손실 은닉 실험에 대한 결과와 실제 패킷 손실 분포를 실험적으로 측정하여 제시하였다. 패킷 손실로 인한 오차의 영향을 최소화 할 수 있도록 수신된 패킷의 복호 과형을 적당히 대체하고, 손실 이후 수신된 패킷에 대해서 음성부호화기의 메모리를 초기화하는 방법을 써서, 연속적으로 발생하는 패킷 손실에 대해 음질 저하를 감소시키는 방법을 고려했던 실험 결과를 제시하였다. 실험 결과 소수의 패킷 손실에 대해서 손실이 없는 합성음의 음질과 거의 차이가 없음을 확인하고, 연속적인 다수의 패킷 손실이 발생한 상황에서도 비교적 만족할 만한 은닉 결과를 얻을 수 있었다. 앞으로는 패킷 손실로 인한 음성부호화기의 내부 메모리 값의 빠른 정정과 다양한 패킷 손실 패턴에 대해 적응적으로 대처할 수 있는 은닉 방법에 대한 연구를 계획이다.

본 과제는 정보통신부의 정보통신학술기초연구지원 사업으로 수행한 연구결과입니다.

참 고 문 헌

- [1] ITU-T G.729: Coding of speech at 8 kbit/s using conjugate-structure algebraic code excited linear prediction(CS-ACELP), International Telecommunications Union, Mar. 1996
- [2] Kondoz A. M., *Digital Speech: Coding for Low Bit Rate Communications System*, John Wiley & Sons, 1994
- [3] ITU-T G.729 Annex B: A silence compression scheme for G.729 optimizes for terminals conforming to recommendation V.70, International Telecommunications Union, Nov. 1996
- [4] C. Montminy, T. Aboulnasr, "Improving the performance of ITU-T G.729A for VoIP", Multimedia and Expo, 2000, ICME 2000, vol.1, pp. 433-436
- [5] ITU-T G.711 Annex I: A high quality low-complexity algorithm for packet loss concealment with G.711, International Telecommunications Union, 1999
- [6] OpenH323 Project web site
<http://www.openh323.org/>