

論文2003-40TC-11-8

FEC기능을 추가한 AMR-WB 음성 부호화기를 이용한 음성 패킷 전송

(Speech Packet Transmission Using the AMR-WB Coder with FEC)

黃 晷 俊 * , 李 寅 誠 **

(Jeong-Joon Hwang and In-Sung Lee)

요 약

본 논문에서는 인터넷 환경에서 실시간 음성 통신을 가능하게 하기 위해 부가 정보를 이용한 손실 패킷 복구 방법이 첨가된 전송 방법을 제시한다. 3GPP에서 기본적으로 이동 통신 환경에서의 사용을 위해 표준화 되었고, 인터넷 환경에서의 사용을 위해 최근에 ITU-T에서 개선된 AMR-WB 음성 부호화기를 사용하였다. 인터넷과 같은 패킷 교환망 서비스에서의 패킷손실은 음질 저하를 유발하고 실시간 통신이 불가능하도록 한다. 따라서 본 논문에서는 단일 손실 발생시에 FEC(Forward Error Correction) 방법을 적용하였고 연속 손실의 경우에는 오류 은닉을 하였다. 또한 손실율에 따라 AMR-WB(Adaptive Multi-Rate Wideband) 부호화기의 특성을 이용하여 여러 모드로 동작 하는 방법을 제시한다. 인터넷 환경의 실험을 위해 길버트 모델을 이용하였다. 손실율을 변화시키며 AMR-WB 23.05 kbit/s 모드로 전송하는 방법과 SNR(Signal to Noise Rate)과 MOS(Mean Opinion Score) 측정을 통해 비교하였다. 실험한 결과 손실율이 30% 에서도 SNR은 9.8dB MOS 값은 3.0정도의 통신 가능한 높은 음질을 보였다.

Abstract

This paper suggests the packet loss recovery method to communicate in real time in the Internet. To reduce the effects of packet loss, Forward Error Correction (FEC) that adds redundant information to voice packets can be used. Adaptive Multi Rate Wideband(AMR-WB) codec which is recently selected by the Third Generation Partnership Project(3GPP) for GSM and the third generation mobile communication WCDMA system and has also been standardized in ITU-T for providing wideband speech services is used. The major cause for speech quality degradation in IP-networks is packet loss. So, We recovered single lossy packet by using FEC method and concealed continued errors. The proposed scheme is evaluated in the Gilbert Internet channel model. The high quality of audio maintained up to 30% packet loss.

Keywords : FEC, 에러은닉, AMR-WB

* 學生會員, 忠北大學校 電波工學科

(Dept. of Radio Science & Engineering Chungbuk
National University)

** 正會員, 忠北大學校 電氣電子컴퓨터工學部

(School of Electrical & Computer Engineering

Chungbuk National University)

※ 본 연구는 2002년 충북대 발전기금 재단 연구년 교수 지원으로 수행되었음.

接受日字:2003年2月5日, 수정완료일:2003年11月1日

I. 서론

인터넷과 같은 패킷 통신 망에 대해 email 혹은 FTP (File Transfer Protocol) 과 같은 과거의 데이터 서비스로부터 많은 실시간 통신 프로토콜이 생겨나고 있으며 앞으로 인터넷 망을 통한 실시간 음성 통신은 보다 중요하게 될 것이다.

AMR-WB(Adaptive Multi Rate Wideband) 부호화기는 기본적으로 GSM 과 3세대 이동통신 WCDMA 시스템을 위해 3GPP(Third Generation Partnership Project)에 사용을 위해 표준화 되었고, 최근에 IETE에서 인터넷 프로토콜에 대해서 표준화 되었다. AMR-WB 부호화기는 또한 2001년 7월에 ITU-T에 의해 16kbit/s 정도의 광대역 음성 부호화기의 표준화에 선택되었다^[1].

패킷 전송 방식의 네트워크에서 음질을 저하시키는 가장 주된 원인은 패킷 손실이다. 패킷 손실로 인해 복호화 음성이 끊어지거나 일그러지고 심지어는 통신 장애가 생기기 까지 한다^[2-4]. VoIP(Voice over IP), 화상회의, 원격의약처방 등과 같은 실시간 프로토콜의 경우 특히 지연에 민감하다.

본 논문은 AMR-WB 부호화기를 이용하여 패킷통신 망에서의 실시간 음성 통신을 가능하도록 하기 위해 추가적인 지연을 최소화 하면서 손실 패킷을 복구하는 방법을 제안한다. AMR-WB 부호화기의 다 전송률 특징을 이용하여 패킷 망의 상태에 따라 저 전송률의 부가 정보를 삽입하는 FEC(Forward Error Correction) 방법을 적용한다. 본 논문에서 제안한 시스템은 단일 패킷 손실의 경우 부가 정보를 통해 패킷을 복구하고 연속 패킷 손실에 대해서는 은닉 방법을 사용한다.

II. AMR-WB 음성 부호화기

현재 2세대 3세대 이동통신 시스템은 대역폭이 200~3400 Hz 정도인 협대역에서 동작한다. 최근 무선통신에서 멀티미디어 서비스와 통신 속도가 가능해 짐에 따라 음성을 명료하고 원 음성에 가깝게 하기 위해 50~7000 Hz 인 광대역이 제공되어진다. 1999년 이후로 3GPP 와 ETSI 는 WCDMA와 모든 비트율의 GSM 시스템을 위한 광대역 음성 부호화기의 개발과 표준화 작업을 진행했다. 그 결과 2001년 AMR-WB(adaptive multi-rate wideband) 부호화기가 표준화 되었다. AMR-WB 부호

화기는 또한 약 16 kbit/s의 광대역 음성 부호화 에 대한 표준화 작업에 참여하여 2001년 7월에 표준화 되었다.

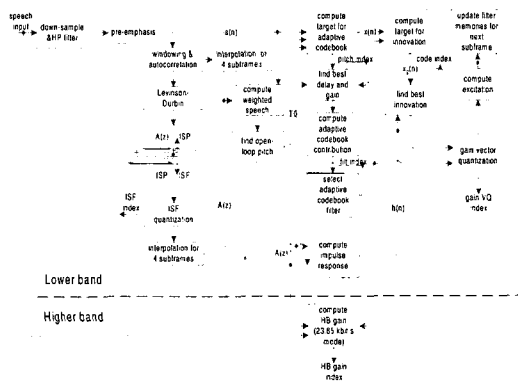


그림 1. ACELP 부호화기의 간략화된 블록도
Fig. 1. Simplified block diagram of ACELP coder.

AMR-WB는 ACELP(Algebraic Code Excited Linear Prediction coder)를 기반으로 다중 모드 ACELP를 AMR-WB라고 부른다. 6.60 kbit/s 에서 23.85 kbit/s 까지 9 개의 압축률을 가진다. VoIP와 인터넷 어플리케이션, 이동 통신, PSTN 어플리케이션, ISDN 광대역 전화, ISDN 비디오폰, 비디오 회의 등의 시스템에서 활용할 수 있다.

1. AMR-WB 음성 부호화기

AMR-WB 음성 부호화기는 비트율이 6.6, 8.58, 12.65, 14.25, 15.85, 18.25, 19.85, 23.05, 23.85 kbps의 9개의 부호화 모드로 이루어져 있다. 각 코덱은 CELP (Code-Excited Linear Predictive) 구조에 기반을 두고 있다.

12.65 kbit/s 이상의 모드는 높은 음질의 광대역 음성을 제공한다. 가장 낮은 두개의 모드 8.85 와 6.60 kbit/s 는 손실이 심한 무선 채널이나 혹은 네트워크의 폭주일 때 단지, 임시적으로 사용 되도록 제공 되어 진다.

AMR-WB 부호화기는 16 kHz 샘플링 주파수에서 동작하며, 부호화는 20ms의 블록 단위로 이루어진다. 두개의 주파수 대역, 50~6400 Hz 와 6400~7000 Hz 는 복잡도를 줄이고, 가장 중요한 주파수 영역에 더 많은 비트를 할당하기 위하여 각각 따로 부호화 된다.

부호화기의 블록도는 <그림 1>과 같다. 먼저 입력 신호는 12.8 kHz로 샘플링 주파수를 낮추고 high pass 필터와 pre-emphasis 필터가 수행된다. 필터를 통과한 신호는 LP analysis 가 수행된다. 이때 구해진 LP 파라미

터들은 스펙트럼(Immittance Spectrum Pairs)으로 변환되고 S-MSVQ(Split-Multistage Vector Quantization)을 사용하여 부호화 된다. 음성

표 1. 20ms 프레임에 대한 AMR-WB 부호화 알고리즘의 비트 할당

Table 1. Bit allocation of the AMR-WB coding algorithm for 20 ms frame.

Mode	bit rate	total number of bits per frame
8	23.85 kbit/s	477
7	23.05 kbit/s	461
6	19.85 kbit/s	397
5	18.25 kbit/s	365
4	15.85 kbit/s	317
3	14.25 kbit/s	285
2	12.65 kbit/s	253
1	8.85 kbit/s	177
0	6.60 kbit/s	132

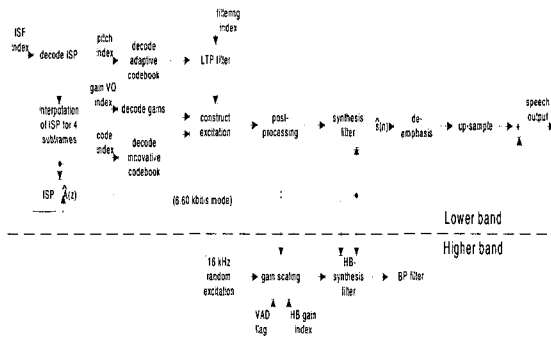


그림 2. ACELP 복호화기의 간략화된 블록도
Fig. 2. Simplified block diagram of ACELP decoder.

프레임은 각각 5 ms 씩 4 개의 하위 프레임으로 나누어진다. LP 파라미터의 양자화 또는 인터플레이션은 하위 프레임에 의해 사용된다. 전송률 모드에 따른 각 하위 프레임당 비트 할당율은 <표 1>과 같다.

2. AMR-WB 음성 복호화기

복호화기에서의 신호의 흐름도는 <그림 2>와 같다. 복호화기에서는 전송된 인덱스들은 수신된 비트스트림으로부터 추출된다. 이러한 인덱스들은 각각의 전송된 프레임에서 코더의 파라미터들을 얻기 위하여 복호화된다. 파라미터들에는 ISP 벡터, 4개의 분할된 피치 지연, 4개의 LTP 필터링된 파라미터, 4개의 혁신적인 코드벡

터, 4개의 벡터 양자화된 피치와 혁신적 이득의 셀이 있다. 23.85 kbit/s 모드에서는 또한 고대역 이득 인덱스가 복호화된다. ISP 벡터는 LP 필터 계수로 변환되며, 각 부프레임에서 LP 필터들을 얻기 위해 상호 보간된다.

그 다음 각 64샘플의 부프레임에서 여기신호는 기대되는 이득들에 의하여 크기가 변한 적응 및 고정 코드 벡터가 합해짐으로써 구해지고, 12.8 kHz의 음성은 LP 합성 필터를 통과한 여기신호를 필터링함으로써 재구성되며, 재구성된 음성은 역강조 되어진다.

마지막으로 재구성된 음성은 16 kHz로 업샘플링되고, 고대역 음성은 6 kHz로부터 7 kHz까지 주파수 대역이 더해진다.

III. 손실 복구 및 오류 은닉 방법

1. 부호화 모드 선택과 패킷 구성시스템

AMR-WB 의 다전송률을 이용해서 통신망의 상태에 따라 부가정보를 사용할지 여부를 결정하고, 기본 부호화 모드와 부가정보로 쓰일 부호화 모드를 선택한다. AMR-WB 부호화기의 경우 한 패킷당 4비트의 모드 정보를 사용한다. 즉, 16개의 모드를 사용할 수 있으므로 <표 1> 의 9개 모드 외에 7개의 모드를 추가할 수 있다. 본 논문에서는 <표 2>와 같이 모드를 추가하였다. 모드 9 - 15는 부가 정보를 추가로 사용하는 모드이다.

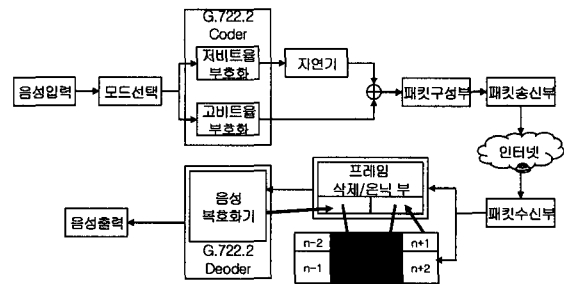


그림 3. 제안된 시스템 구조
Fig. 3. Block diagram of the suggested system.

표 2. 부가정보 삽입을 위한 추가된 모드
Table 2. appended mode for additional data.

모드	9	10	11	12	13	14	15
비트율(kbps)	8.85	12.65	14.25	15.85	18.25	19.85	23.05

손실 율에 따라 적절한 모드를 선택함으로써 대역폭의 증가 없이 부가 정보를 삽입하여 손실 패킷을 복구할 수 있다. 본 논문에서는 무손실시에 23.05 kbps로 부호화 하고 낮은 손실율에서는 15.85 kbps 부호화 데이터에 6.60 kbps 부가정보를 삽입하여 패킷을 구성하고 높은 손실율에서는 14.25 kbps 부호화 데이터에 8.85 kbps 부가 정보를 삽입하여 손실 율에 상관없이 약 23.00 kbps 정도를 유지하였다.

2. 부가 정보를 이용한 패킷 손실 복구

실시간 통신 시스템은 손실을 복구하기 위해 재전송이나 추가적인 디코딩 지연이 발생하는 인터리빙 방법은 제한된다. Mbone(Multicast Backbone) 상에서 재전송 없이 손실 복구를 위해 FEC 방법이 효과적임이 실험되었다^[4]. 이 방법의 간단한 예가 <그림 4>에 나와 있다. i 번째 부호화 패킷은 입력 신호와 1패킷 이전의 입력신호를 부가정보로 가지고 있다. 이모델은 단일 패킷의 손실을 복구 할 수 있다. $\bar{y}[i+2]$ 패킷이 손실 되었을 경우 $\bar{y}[i+3]$ 패킷을 입력받아 부가정보를 사용해서 $\bar{x}[i+2]$ 를 복구하고 또 기본 데이터를 이용해서 $\bar{x}[i+3]$ 를 복호화 할 수 있다. FEC 방법은 추가적인 지연을 거의 발생시키지 않고 원 음성에 가깝게 복구할 수 있으나 대역폭이 증가하며, 또 단기간 손실 복구만 가능하다는 단점이 있다. 하지만 인터넷에서 패킷 손실은 무작위로 발생하는 경향이 있으며 단일 손실뿐 아니라 연속 손실이 발생한다^[5]. 따라서, 최근에 인터넷 환경에 FEC 방법을 적용하기 위한 연구가 이루어지고 있다^[6,7]. 이 방법들은 패킷망의 상태에 따라 부가정보 추가 방식을 달리 하여 가능한 대역폭을 늘리지 않는 방법을 사용한다.

<그림 5>와 같이 부가 정보의 지연을 달리 하였을 때, $\bar{y}[i+1], \bar{y}[i+2], \bar{y}[i+3]$ 3개의 패킷이 연속 손실된 경우 이 후의 세 개의 패킷의 부가 정보를 이용해 $\bar{x}[i+1], \bar{x}[i+2], \bar{x}[i+3]$,을 복원 할 수 있다. 이 때, $\bar{x}[i+1]$ 은 3패킷 이후에 복원 가능하기 때문에 손실 발생시 지연은 3패킷의 지연이 발생한다. 뿐만 아니라 <그림 5>와 같이 패킷을 구성한 경우 $\bar{y}[i+1]$ 패킷 하나만 손실 되었다 하더라도 역시 3 패킷 이후에 $\bar{x}[i+1]$ 패킷이 손실 된다면 이전에 손실 되었던 $\bar{x}[i+1]$ 패킷을 복원 하기 위해 부가 정보를 사용 할 수가 없게 된다. 따라서 단일 패킷 손실 발생의 경우에도 여전히 3

패킷의 지연이 발생하게 되며, 3패킷 이후의 패킷이 손실되었을 경우에는 단일 패킷 손실 조차 복원을 할 수 없게 된다. 따라서 이와 같이 지연을 달리 하는 경우는 지연이 크게 발생하게 되므로 실시간 전송에 부적합 할 뿐만 아니라, 네트워크의 순간적인 과부하로 인해 지연이 크게 발생하는 경우에는 효과적일 수 있지만, 거의 단일 패킷이 발생하는 일반적인 네트워크의 경우에는 오히려 부가정보를 한 패킷만 지연 시킨 경우보다 음질을 더욱

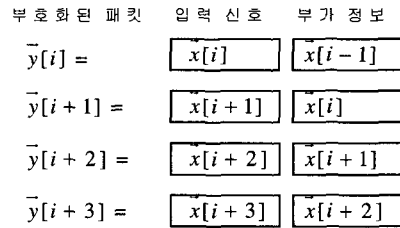


그림 4. FEC 방법의 간단한 예
Fig. 4. An example of FEC method.



그림 5. 3 패킷 이전의 부가정보를 사용한 경우
Fig. 5. The case using additional data from 3 packets ago.

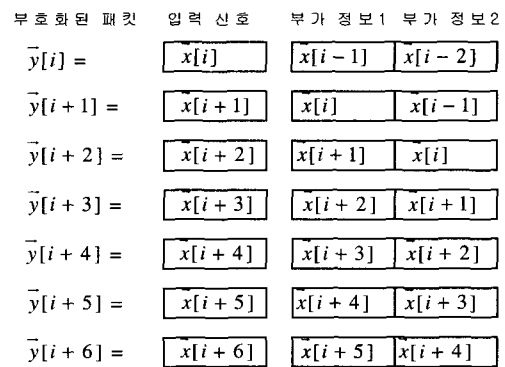


그림 6. 2개의 부가 정보를 사용한 경우
Fig. 6. The case using two additional data.

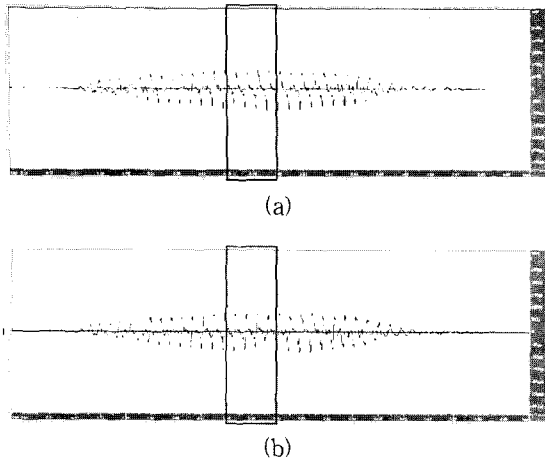


그림 7. 단일 패킷 손실을 복구한 결과 (a) 손실 발생하지 않았을 경우의 음성 출력 (b) 손실된 패킷을 복구한 후의 음성 출력

Fig. 7. Reconstructed signal after the single packet loss recovery. (a) output signal without packet loss (b) recovered output signal with additional data

떨어뜨리게 된다.

<그림 6>과 같이 패킷을 구성하는 방법은 단일 패킷 손실을 복원하기 위해서는 추가적인 복호화 지연은 발생하지 않지만 연속 손실을 복원하기 위해서는 한 패킷을 더 기다려야 하므로 복원시에 한 프레임의 지연이 추가로 발생할 수 있다. 즉, $\bar{y}[i+1]$ 과 $\bar{y}[i+2]$ 패킷이 손실 되었을 경우 $\bar{x}[i+1]$ 을 복원하기 위해 두 패킷을 기다렸다가 $\bar{y}[i+3]$ 을 수신하고서야 복원 할 수 있다. 뿐만 아니라 이 방법의 단점은 부가 정보를 위해 대역폭이 더욱 증가하게 된다는 것이다. 부가 정보를 위해 대역폭을 늘려 사용한 경우 그 효율을 비교해 보기 위해 4장에서 한패킷 전체 대역폭을 비슷하게 하고 하나의 부가 정보를 사용한 경우와 두개의 부가정보를 사용한 경우를 실험해 보았다. 그 결과 두개의 부가 정보를 사용하여 연속 손실을 복원한 경우 그 효율이 떨어짐을 확인 할 수 있었다. 이렇듯 연속 패킷 손실에 대해 부가 정보를 사용하여 손실을 복구할 경우 계속해서 부가정보를 사용하는 것은 대역폭의 증가 혹은 부호화 지연이 추가로 발생하게 되며 이것은 실시간 통신의 장애 요인이 된다. 따라서 본 논문에서는 하나의 부가 정보를 한 패킷만 지연 시켜 단일 패킷 손실에 대해서 <그림 7>의 경우처럼 복원한다.

<그림 7(a)>는 손실이 없을 경우의 출력 파형을 나타

낸 것이고 <그림 7(b)>는 박스 부분의 패킷을 손실시킨 다음에 부가 정보를 사용하여 손실된 부분을 복원한 파형이다. 연속적인 손실에 대해서는 여러 은닉을 통하여 대역폭의 증가 혹은 추가적인 지연 없이 실시간 통신에 적합하도록 하였다. 제안한 손실패킷 복구 및 오류 은닉방법이 첨가된 시스템의 구조는 <그림 3>과 같다.

3. 오류 은닉

3.1. 연속 패킷 손실 복구

부가 오디오 정보를 이용한 손실 패킷 복원방법을 사용하였을 경우 부가 오디오 정보로도 복원할 수 없는 연속 패킷 손실이 <그림 8>에서와 같이 발생하게 되었을 경우 기존의 오류 은닉 방법으로는 손실된 프레임에 침묵이나 백색잡음으로 대체하거나, 이전에 수신된 패킷을 복사하는 방법이 사용되었다.

음성 부호화기 중 가장 효과적인 손실 프레임 복원방법은 이전에 올바르게 수신한 프레임의 파라미터를 감쇄시켜서 손실된 프레임에서 사용하는 방법이다. 이 방법은 부가적인 지연을 필요로 하지 않는 장점이 있기 때문에 최근의 음성 부호화기 표준에 사용되었다. 그러나 <그림 9(b)>와 같이 일률적으로 감쇄된 값을 감쇄시키기 때문에 박스 안의 손실 부분의 음질 왜곡이 심하게 발생될 뿐만 손실 이후의 정상적으로 수신된 박스 뒤의 부분 역시 손실의 영향을 받아 파라미터 감쇄에 의한 음질 왜곡이 발생한다.

본 논문에서는 연속적으로 패킷이 손실 되었을 경우, 복원 불가능한 패킷을 재생산하고 이후 프레임의 오류를 은닉하였다. <그림 8>의 경우 가장 마지막에 손실된 n+1 번째 패킷은 n+2 번째 패킷의 부가정보를 통해 복원을 한다. 이전의 연속적인 손실 패킷에 대해서는 가장 최근에 정상적으로 복호화된 n 번째 패킷의 스트림 정보를 이용하여 복원한다. 복원 불가능한 패킷을 기준으로 가장 최근에 정상적으로 복호화된 n 번째 패킷은 과거 패킷이 되고 부가 정보로써 복원된 n+2 번째 패킷은 미래 패킷이 된다.

AMR-WB 부호화기에서 보간법이 적용될 파라미터는 LSF parameter, pitch lag, gain factor 이다. LSF의 보간법은 [9]에서 연구 되었다. 하지만 본 논문에서는 이러한 은닉을 통한 개선이 가능함을 보이는 것으로써 보간법을 사용하지 않고 과거 패킷을 그대로 사용하였다. 만약 3개 이상의 연속 패킷 손실이 발생한다면 연속적 패킷 손실이 발생하기 이전 마지막으로 올바르게 수

신한 패킷을 단순히 가장 먼저 손실이 발생한 패킷의 대체 패킷으로 복사하여 사용하는 오류 은닉 방법을 사용하였다.

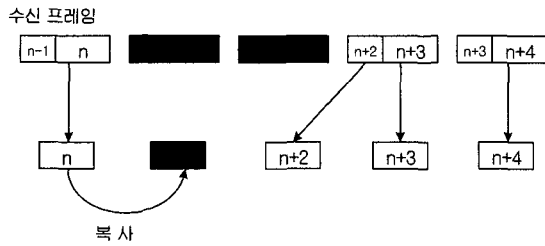


그림 8. 연속 패킷 손실인 경우 복구방법
Fig. 8. Recovery method in continual packet loss.

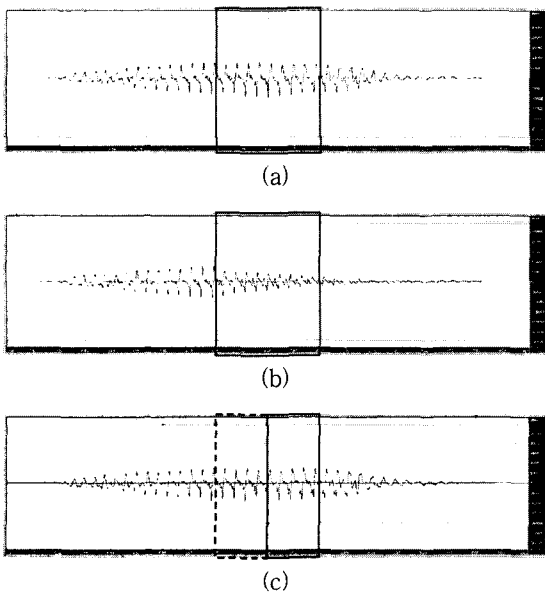


그림 9. 연속 패킷 손실을 복구한 후의 음성 파형 (a) 손실이 발생하지 않았을 경우 출력 (b) 23.05 kbit/s amr-wb 자체 에러 은닉 방법 사용 (c) 제안된 방법을 사용한 손실 복구

Fig. 9. Reconstructed signal after the continual packet loss recovery. (a) output signal in case that no packet loss occurs (b) signal which the method on 23.05 kbit/s amr-wb outputs (c) signal which the method suggested outputs

3.2. 군집 오류 은닉

CELP 기반 부호화기는 예측 부호화기 이므로 각 프레임은 현재 프레임에 의해 생성된 정보 비트뿐만 아니라 이전 프레임으로부터 영향을 받는 예측 상태를 기반으로 복호한다. 그러므로 부정확한 예측 상태는 손실된

프레임 이후의 신호를 복호하는데 음질 저하를 유발시키므로 즉 상태를 정확히 갱신하지 않고 음성신호를 복원해내서는 안 된다. 본 논문에서 사용하는 오류 은닉 방법은 부호화기에서 복원하지 못하는 프레임에 대해 과거 여기 신호를 '0'으로 초기화 하는 방법을 사용하여 상태 오류를 줄였다.

<그림 9>는 두개의 패킷을 연속 손실 시켰을때 복원한 파형이다. (a)는 손실을 발생시키지 않고 정상적으로 복호화 시킨 파형이고 (b)는 23.05kbit/s 자체 은닉 방법을 사용함으로써 파라미터 감쇄에 의하여 손실을 은닉한 파형이다. 그리고 (c)는 위에서 제안한 방법을 사용하여 손실을 복구하고 복원이 불가능한 부분에 대해서는 은닉 방법을 사용한 출력 파형이다. (b)와는 달리 (c)의 경우 마지막 박스 안의 부분은 그 다음 패킷의 부가 정보를 사용하여 복원을 하였고 점선 안의 부분은 이전에 정상적으로 수신된 패킷을 사용하여 재 구성하여 손실 부분을 복구 하였다. 박스 이후 부분은 손실 부분을 부가 정보로 복원한뒤 과거 여기 신호를 '0'으로 초기화 하여 이후 정상적으로 수신된 패킷은 파라미터 감쇄에 의한 손실이 발생하지 않고 정상적으로 복호화 하여 오류를 줄였다.

IV. 성능 평가

인터넷 환경에서의 패킷 손실과정의 상관구조를 길버트 모델과 같은 저 차의 마르코프 체인으로 나타낼 수 있다는 연구결과가 나왔다^[7, 8]. <그림 10>은 길버트 모델의 천이도이다. 손실은 B 상태에서만 일어나며 이전 프레임이 G 상태일때 손실이 발생할 확률이 p 이고 이전 프레임이 B 상태일때 손실이 발생할 확률은 q 이다. 길버트 모델을 적용하였을 경우 평균 손실 확률은 B 상태에 머무르는 평균시간을 B 상태에 머무르는 평균시간과 G 상태에 머무르는 평균시간의 합으로 나누어 식 (1)과 같이 구할 수 있다.

본 논문에서 제안한 방법의 성능을 평가하기 위해 객관적 음질 평가 방법인 SNR과 주관적 음질 평가를 위해 MOS 테스트를 실시하였으며, 비교되는 부호화기는 부호화기 자체에 내장된 오류 은닉 방법을 사용한 AMR-WB 부호화기의 23.05 Kbit/s 모드를 사용한다. 평균 패킷 손실률은 0 %, 1 %, 5 %, 10 %, 20 % 그리고 30 % 를 사용하였다. 패킷의 크기는 부호화기 자체의 프레임 크기로 하였고, q값은 0.2를 사용하였다.

$$L_{avg} = \frac{P}{P + p} \tag{1}$$

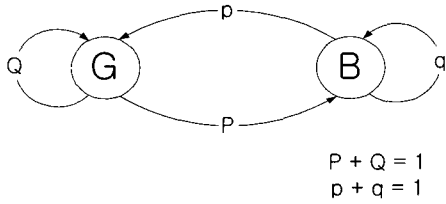


그림 10. 길버트 모델의 상태 천이도
 Fig. 10. State diagram of the Gilbert model.

객관적 성능 평가를 위해 SNR을 사용하였다. SNR은 식 (2)와 같이 계산되어진다.

$$SNR (dB) = 10 \log \left[\frac{\sum_{n=1}^{N_s} s^2(n)}{\sum_{n=1}^{N_s} (s(n) - \hat{s}(n))^2} \right] \tag{2}$$

여기서, $s(n)$ 은 원 음성 신호이고 $\hat{s}(n)$ 은 복원 음성 신호이다. 그리고 주관적 성능 평가를 위해 PESQ 프로그램을 사용하여 입력 음성과 출력 음성 간의 주관적 MOS 값을 측정하였다.

<그림 11>에서 점선으로 표시된 부분은 연속으로 패킷이 손실되는 부분을 나타낸다. (a)는 원본 신호, (b)는 AMR-WB 23.05Kbit/s 모드의 오류 은닉방법, (c)와 (d)는 제안된 손실 패킷 복구 및 오류 은닉방법을 적용하였을 경우 복원된 음성신호의 스펙트로그램들이다. 점선으로 표시된 부분을 보면 알 수 있듯이 제안된 방법이 부호화기 자체의 오류 은닉방법보다 더 많은 음성을 복구해 낼 수 있다.

평가는 남녀 각각 4문장으로 총 음성길이는 800개의 프레임(1프레임 = 20ms)으로 이루어진 입력 파일들을 사용했으며 AMR-WB 23.05 kbit/s 모드와 제안한 방법으로 (15.85+6.60) kbit/s 모드와 (14.25+8.85) kbit/s 모드를 사용하여 0% 에서 30%까지 평균 손실율에 따른 SNR(Signal to Noise Rate)과 MOS(Mean Opinion Score)측정값이 <표 3>에 나타나 있다. 무 손실인 경우를 제외하고는 본 논문에서 제안한 방법이 패킷 손실이 있다고 가정할 경우에 더 성능이 좋을 수 있다.

<표 3>을 보면 손실이 있을때 SNR값이 약1.4-7.1

dB 더 좋은 성능을 보이는 것을 알 수 있다. <표 3>과 <그림 12>을 보면 MOS 결과 값은 오류가 없는 상황에서 AMR-WB 23.05 Kbit/s 모드, 그리고 제안한 방법 모두 MOS 4.0이상의 음질을 유지 한다는 것을 알 수 있다. 즉 모두 유선 전화 품질 이상의 음질을 가지고 있다는 것이다. 그러나 손실이 발생하는 상황에서는 역시 제안된 방법이 비교 대상 부호화기보다 더 높음을 알 수 있다. 20 %와 30 %의 패킷 손실에서 제안된 방법이

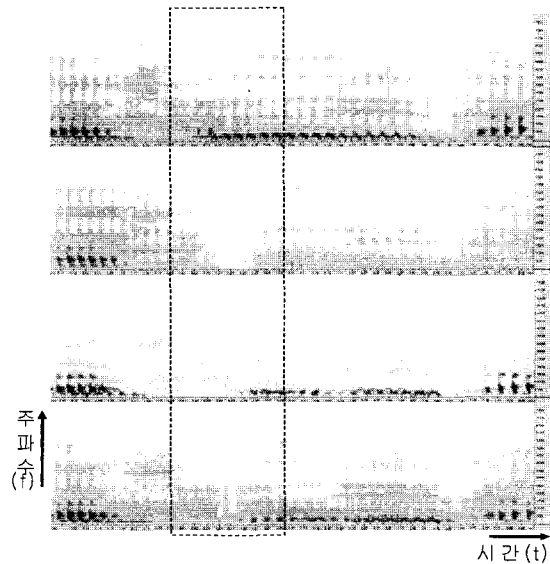


그림 11. 30 % 평균 패킷 손실 시 음성 부호화기의 출력 음성 (a) 원본 신호 (b) AMR-WB 23.05 Kbit/s 부호화기 (c) 제안된 방법(15.85 Kbit/s + 6.60 Kbit/s) (d) 제안된 방법(14.25 Kbit/s + 8.85 Kbit/s)

Fig. 11. Output signal of speech coder at 30% average packet loss rate. (a) Original signal (b) Using AMR-WB 23.05 Kbit/s (c) Suggested method(15.85 + 6.60 Kbit/s) (d) Suggested method(14.25 + 8.85 Kbit/s)

표 3. SNR 값과 MOS 값 측정 결과
 Table 3. Results of SNR and MOS tests.

평균 손실율		0%	1%	5%	10%	20%	30%
AMR-WB 23.05 (kbit/s)	MOS	4.260	3.928	3.472	3.112	2.495	2.205
	SNR(dB)	35.431	23.312	17.549	10.983	5.829	3.735
제안된방법 (15.85+6.60) 22.45 (kbit/s)	MOS	4.179	4.067	3.870	3.650	3.267	3.036
	SNR(dB)	31.051	25.660	21.493	16.175	11.245	8.583
제안된방법 (14.25+8.85) 23.10 (kbit/s)	MOS	4.145	4.047	3.871	3.673	3.338	3.106
	SNR(dB)	29.808	25.500	22.093	17.781	13.454	10.834

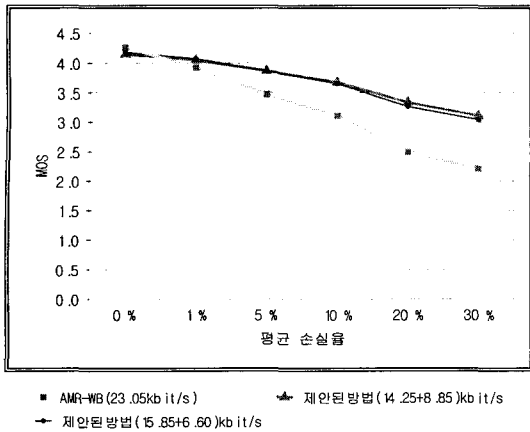


그림 12. MOS 실험 결과
Fig. 12. Results of MOS test.

표 4. 부가 정보의 수를 달리하여 실험한 결과
Table 4. Results of simulation about the different number of additional data.

평균 손실율		0%	3%	5%	10%	20%	30%
제안된 방법 (18.25+8.25) 27.10 (kbit/s)	MOS	4.245	4.202	4.034	3.884	3.595	3.142
	SNR(dB)	32.311	29.298	21.734	19.720	14.183	7.865
(14.25+6.60 +6.60) 27.45 (kbit/s)	MOS	4.161	4.090	3.991	3.796	3.540	3.139
	SNR(dB)	28.921	25.602	22.212	19.834	15.295	9.088

AMR-WB 23.05 kbit/s 보다 약 0.77-0.84 높고 MOS 3.0 이상의 음질을 유지하는 것을 확인 할 수 있다.

부가 정보의 수를 달리하여 적용 하였을 경우 손실율에 따른 성능을 알아보기 위해 <표 4>와 같이 연속 손실율은 30%로 고정하고 평균 손실율을 변화 하면서 실험 하였다. 전송률을 비슷하게 하기 위해 본 논문에서 제시한 방법은 기본 전송률은 18.25kbps 로 하고 부가 정보의 전송률은 8.85 kbps로 하여 전체 27.10 kbps로 하였고 부가 정보를 두개 사용한 경우는 기본 전송률은 14.25kbps로 하고 두개의 부가 정보를 각 6.60kbps 로 하여 전체 27.45kbps로 하였다. <표 4>에서 연속 손실율을 30%로 고정 하였으므로 평균 손실율이 적을 수록 단일 패킷 손실보다는 연속 손실의 비중이 높아진다. 연속 패킷의 비중이 높은 경우 역시 전체 전송률을 비슷하게 한 경우 하나의 부가 정보를 사용한 경우 보다 높은 음질을 유지 하였다. 따라서 연속 손실을 복원을 하기위해 부가 정보를 추가로 사용하는 것 보다는 은닉 알고리즘을 적용하는 것이 보다 효율적임을 확인 할 수 있다.

V. 결론

본 논문에서는 AMR-WB 부호화기를 이용하여 인터넷을 통한 음성 전송 및 오디오 전송에 사용하기 위한 방법을 제안하였다. 인터넷을 기반으로 한 음성 데이터 전송에서 고려해야할 단일 손실 패킷에 대한 복구 방법으로 순방향 오류 정정 방법(FEC)을 사용하였는데, AMR-WB 부호화기의 특징인 다중 전송률을 이용하여 각각 주된 오디오 정보에 부가 오디오 정보를 AMR-WB 부호화기의 저 전송률 모드를 사용하였다. 인터넷에서의 패킷 전송 모델은 Gilbert Model을 사용하였다.

평균 패킷 손실율을 바꿔 가면서 실험한 결과 SNR값 및 MOS 테스트 결과 더 좋은 성능을 보였으며, 20%의 패킷 손실에서 MOS값은 3.0 이상 SNR 값은 13 dB 이고 30%에서 MOS값이 3.0 정도 SNR값은 9.8 dB로 통신 가능한 음질 이상을 보였다.

제안된 방법은 IMT-2000 시스템과 인터넷망과의 음성 통신을 위해 AMR-WB 코더의 전송 손실 보완 방법으로 사용 가능하다.

참고 문헌

- [1] R. Salami, et al., "The Adaptive Multi-Rate Wideband codec: History and Performance," IEEE Speech Coding Workshop, pp. 144-146, October 2002.
- [2] J. Bolot and A. Vega_Garcia, "Control mechanisms for packet audio in the Internet," in Proceedings of IEEE INFOCOM, pp. 232-239, March 1996.
- [3] C. Perkins, O. Hodsom and V. Hardman, "A survey of packet loss recovery techniques for streaming audio," IEEE Network, Vol. 12, no. 5, pp. 40-48, September 1998.
- [4] C. Romer, M. Podolsky and S. McCanne, "Simulation of FEC-based error control for packet audio on the Internet," in Proceedings of IEEE INFOCOM, pp. 505-515, April 1998.
- [5] Vern Paxson, "End-to-end internet packet dynamics," IEEE Transactions On Networking, Vol. 7, no. 3, pp. 277-292, June 1999.

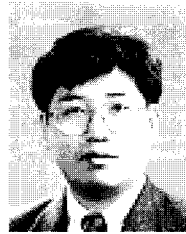
- [6] Toru Morinaga, Kazunori Mano, et al., "The Forward-Backward Recovery Sub-Codec (FB-RSC) Method : A Robust Form of Packet-Loss Concealment for Use in Broadband IP Networks," IEEE Speech Coding Workshop, pp. 62-64, October 2002.
- [7] E. Martinian and C. -E. W. Sundberg, "Low Delay Burst Erasure Correction Codes," ICC 2002. IEEE International Conference on, Vol. 3, pp. 144-146, 2002.
- [8] D. Veitch, "Understanding end-to-end Internet traffic dynamics", IEEE Globecom 98, November 1998.
- [9] J. Wang and J. D. Gibson, "Parameter interpolation to enhance the frame erasure to CELP coders in packet networks," in Proceedings of ICASSP 2001, pp. 745-748, May 2001.

저 자 소 개



黄 晟 俊(學生會員)

2002년 2월 : 충북대학교 전기전자공학부 학사. 2002년 3월~현재 : 충북대학교 전파공학과 석사과정.
<주관심분야 : 디지털 신호처리, 음성 신호처리, VoIP>



李 寅 誠(正會員)

1983년 2월 : 연세대학교 전자공학과 학사. 1985년 2월 : 연세대학교 전자공학과 석사. 1992년 12월 : Texas A&M University 전기공학과 박사. 1986년 5월~1987년 7월 : 한국통신 연구개발단 전임연구원. 1993년 2월~1995년 9월 : 한국전자통신연구원 이동통신 기술연구단 선임연구원. 1995년~현재 : 충북대 전기전자공학부 부교수. <주관심분야 : 음성 및 영상신호압축, 이동통신, 적응필터>