

음성 신호 전송에 있어서의 에러 은닉 기법

정희원 안 용 기*, 장 원 영**, 전 준 현**, 정 제 창*

Error Concealment Techniques in Audio Signal Transmission

Yong Ki Ahn*, Won Young Chang**, Joon Hyeon Jeon**, Jechang Jeong* *Regular Members*

요 약

디지털 오디오 신호는 전송중의 에러에 의해 음질이 많이 저하된다. 이 문제를 해결하기 위해 지금까지 다수의 정교한 에러 은닉 기법이 제안되었다. 본 논문에서는 에러에 의해 손실된 프레임을 복원하기 위해서 삼각 변환 커널(Sinusoidal Transform Kernel)과 패턴 매칭(Pattern Matching)을 이용하는 두 가지의 에러 은닉 기법을 제안한다. 삼각 변환 커널을 이용한 에러 은닉 기법에서는 손실된 프레임의 앞 뒤 프레임의 삼각 변환 계수를 이용한다. 앞 뒤 프레임의 삼각 변환 계수들의 에너지 집중도를 조사하여 DFT, π -ODFT, DCT, DST, DHT중에서 하나의 삼각 변환을 선택한다. 선택된 삼각 변환 계수들을 보간하여 손실된 프레임의 변환 계수를 예측한 후, 예측된 손실 프레임의 변환 계수들을 합성하여 시간 영역에서 프레임을 복원한다. 패턴 매칭을 이용한 에러 은닉 기법에서는 손실된 프레임의 앞 뒤 프레임을 기준 벡터로 이용하여 탐색영역안에서 상관 계수가 가장 높은 부분을 구한다. 구해진 부분을 이용하여 손실된 프레임을 대체한 후, 복원된 프레임이 앞 뒤 프레임과 연속성을 가질 수 있도록 보상한다. 그리고, 세 가지 종류의 오디오 신호에 대해 제안된 에러 은닉 기법을 적용한 결과 기존의 에러 은닉 기법과 비교하여 우수한 주관적, 객관적 음질을 얻었다.

ABSTRACT

Digital audio signals are severely degraded due to uncorrectable transmission errors. In order to overcome this problem, a number of elaborate error concealment algorithms have been proposed so far. In this paper, two error concealment methods, a error concealment method using sinusoidal transform kernels and a error concealment method using pattern matching, are presented in order to reconstruct the error embedded frame. In the error concealment method using sinusoidal transform kernels, we use sinusoidal transform coefficients of the previous and the next frame of the lost frame. We choose the one of transform among DFT, π -ODFT, DCT, DST and DHT, using the energy compaction of transform coefficients of the previous and the next frame. After predicting transform coefficients of the lost frame by interpolating the chosen sinusoidal transform coefficients, we reconstruct the lost frame in the time domain by synthesizing the predicted lost frame transform coefficients. And in the error concealment method using pattern matching, we search the part which has the maximum correlation in the search range, using the previous and the next frame of the lost frame as reference vector. After substituting the lost frame for the found part, we compensate the reconstructed frame to have continuity with the previous and the next frame. When applied to three kinds of audio signals, the proposed methods show superior subjective and objective quality to conventional error concealment methods.

* 한양대학교 전자통신공학과(yongsmile@lge.co.kr) 정희원, ** 한국통신 멀티미디어연구소: 정희원
논문번호: 97215-0624, 접수일자: 1997년 6월 24일

I. 서 론

오디오와 비디오 데이터는 아날로그 신호일 때보다 디지털로 부호화될 때 보다 나은 품질을 얻을 수 있다. 그러나, 디지털로 부호화될 때 데이터의 발생량이 많아지므로 압축기법을 이용하여 데이터의 발생량을 감소시켜야 한다. 디지털 오디오 신호는 MPEG 오디오 또는 AC-3 등 규정된 압축 표준에 따라 부호화되어 프레임단위로 전송되므로, 프레임 헤더내의 단 하나의 에러에 의해 프레임 전체가 손실되는 등 프레임내의 에러에 매우 민감하다. 그러므로, 에러로 인한 문제를 극복하기 위해서 에러 정정 및 검출부호가 사용된다.

MPEG-2 오디오나 AC-3 오디오의 프레임에서는 에러 검출부호인 CRC(Cyclic Redundancy Code)를 사용하고 있다. 그러나, 에러 정정 또는 검출부호는 실제 데이터 뒤에 에러를 정정하거나 검출하기 위하여 일정한 양의 부가적인 데이터를 덧붙여 전송하므로 부가적인 데이터로 인해 전체적인 데이터 양의 증가를 가져온다. 그리고, 채널상의 에러 특성을 알 수 없기 때문에 과도하게 지나친 에러 정정 및 검출 능력을 가진 부호를 사용하여 오히려 압축효율을 저하시키며, 에러 정정 및 검출 과정 중에 문제가 발생하기도 한다.

디지털 오디오 신호의 전송 에러를 극복하기 위한 또 다른 방법으로 채널을 통해 전송할 때 에러가 발생하면 복호화기 다음의 후처리 과정(Post Processing)을 통해서 발생한 에러를 인간의 청각으로 느낄 수 없도록 은닉하는 에러 은닉 기법이 있다[1-4].

본 논문에서는 제안하는 에러 은닉 기법으로서 손실된 프레임의 앞 뒤 프레임의 삼각 변환계수들을 구하고 에너지 집중도를 고려하여 하나의 삼각 변환계수를 선택한 후 그 변환계수를 보간하여 손실된 프레임의 변환계수들을 예측하고 예측된 변환계수들을 합성하여 시간 영역의 손실된 프레임의 오디오 신호를 복원하였다. 그리고, 또 다른 에러 은닉 기법으로서 손실된 프레임의 앞 뒤 프레임을 이용하여 기준이 되는 벡터를 구성하고 탐색영역 안에서 기준이 되는 벡터와 상관계수가 가장 높은 벡터를 찾아서 이를 이용하여 손실된 오디오 신호를 복원하였다.

본 논문의 구성은 다음과 같다. II장에서는 에러로 인해 손실된 프레임에 대한 기존의 에러 은닉 기법들을 살펴보고, III장에서는 제안된 삼각 변환 커널(Sinusoidal Transform Kernel)을 이용한 에러 은닉 기법과 패턴 매칭(Pattern Matching)을 이용한 에러 은닉

기법에 대하여 살펴본다. 그리고 IV장에서는 제안된 에러 은닉 기법을 세 가지의 오디오 신호에 대해 적용한 결과를 보이고 있다.

II. 기존의 에러 은닉 기법

에러가 발생하여 그 프레임을 사용하지 못하는 경우 인간이 에러를 느끼지 못하도록 에러 은닉기법을 사용하여야 한다. 여러 가지의 에러 은닉 기법 중에서 가장 간단한 기법인 에러가 발생한 프레임의 묵음으로 만드는 묵음화 기법보다 나은 음질의 에러 은닉 기법을 선택하여야 한다. 기존의 제안된 에러 은닉 기법은 다음과 같다.

1. 묵음화(Mute)

에러 은닉의 가장 간단한 기법은 복호할 수 없는 프레임을 묵음화하는 것이다. 프레임을 자주 묵음화하는 것은 주관적인 음질을 매우 떨어뜨리게 된다. 그리고 오디오 신호와 묵음 구간의 '클릭'하는 소리를 피하기 위해 'Fade-in' 또는 'Fade-out' 방법이 첨가될 수 있다. 하지만 프레임을 묵음화하는 것은 매우 귀에 거슬리며, 묵음화된 프레임의 정보는 손실된다[5-8].

2. 반복(Repetition)

복호화할 수 없는 프레임들을 묵음화하는 대신에 앞에서 올바르게 복호된 프레임으로 대체할 수 있다. 올바르게 복호된 프레임을 가지고 에러에 의해 손실된 프레임을 대체하는 방법은 오디오 신호 그 자체에 매우 민감하다. 하지만 대부분의 경우 단순히 묵음화시키는 것보다는 좋은 음질을 들려 준다. 그러나, 묵음화하는 것과는 달리 최소한 한 프레임 정도의 데이터를 저장할 수 있는 메모리가 필요하다. 프레임들이 많이 깨어진 경우에는 한 프레임을 여러 번 반복하거나 손실된 프레임의 수만큼 한 번 반복하여 에러 은닉을 할 수도 있다. 이러한 에러 은닉 기법은 순수한 사인파나 백색잡음의 경우에는 잘 적용되지만 대부분의 오디오 신호에는 그렇지 못하다[5-8].

3. 주파수 영역에서의 에러 은닉

시간 영역에서의 에러 은닉 방법에 대해서는 잘 알려져 왔지만 주파수 영역에서의 접근은 많지 않았다. 디지털 오디오 신호의 전송은 그림 1과 같이 모델링 된다. 시간 영역에서 주파수 영역으로의 변환은 MDCT(Modified Discrete Cosine Transform)이나 다위상 필터뱅크 또는 이들의 하이브리드 방법을 사용한다.

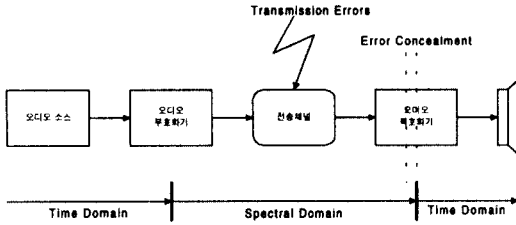


그림 1. 주파수 영역에서의 오디오 신호 전송의 모델링
Fig. 1 The modeling of audio signal transmission in frequency domain

만약 전송 중에 에러정정코드로 복구할 수 없는 에러가 발생하면 주파수 영역의 에러는 시간 영역에서는 균집에러를 유발하게 된다.

에러 은닉 방법은 Analysis 필터 बैं크의 출력을 시간과 주파수 및 진폭의 3차원 공간을 구성한 후, 문턱치(Threshold)를 주어 그 이하로 클리핑하는 방법이다. 하지만 이러한 방법은 에러에 대한 정보를 가지고 있지 않기 때문에, 특별한 종류의 오디오 신호에 대해 음질을 떨어뜨릴 가능성을 가지고 있다[9].

4. AR(Autoregressive) 모델링을 이용한 에러 은닉
손실된 프레임의 신호를 AR 프로세스로써 모델링하여 손실된 프레임을 보간하는 기법이다. 기존의 AR 모델링과는 달리, 손실된 프레임과 그 주위의 신호에 대해 1개의 AR모델을 사용하지 않고 앞 뒤 프레임을 이용하여 2개의 서로 다른 AR 모델링을 사용한다. 각각의 AR 모델링에 의해 구해진 신호에 대해 프레임의 연속성을 고려해서 Raised Cosine형태의 가중치를 부과한 후 두 신호를 합하여 손실된 프레임의 신호를 복원한다[10].

5. LSB(Least Significant Bit) Dropping을 이용한 에러 은닉

패킷 통신에 있어서, 전송측에서는 패킷의 수가 전송능력을 초과하여 발생하면 초과 발생된 패킷을 버리게 되고 수신측에서는 버퍼메모리의 지연시간내에 패킷이 도착하지 않으면 그 패킷을 잃어버리게 된다. 또한 전송선상에 혼잡도를 제어하기 위해 패킷을 버리기도 한다.

LSB는 복원된 신호의 음질에 영향을 적게 주기 때문에 LSB Dropping기법은 전송되는 패킷을 MSB(Most Significant Bit)패킷과 LSB패킷으로 만들어서, 전송 도중 LSB패킷을 손실하여도 복원된 신호의 음질에 영

향을 적게 주고자 하는 기법이다. LSB Dropping기법이 ADPCM에 적용된다고 가정하자. 이 때 LSB패킷에 손실이 발생하면 부호기의 예측기 입력과 복호기의 예측기 입력이 달라져서 복호기쪽의 양자화 스텝 크기나 예측변수 등이 바뀌어서 예측기의 출력값이 부호기의 값과 달라지게 된다. 이 문제를 해결하기 위해 그림 2와 같이 R비트의 부호어 중에서 LSB를 포함한 m비트를 제외($R' = R - m$)하여 부호화기와 복호화기의 예측기 입력으로 보내 예측신호를 만들어 낸다. ADPCM의 부호어를 R' 비트의 MSB와 m비트 LSB패킷으로 나누어 전송하면 LSB Dropping기법을 ADPCM에 적용할 때 패킷 손실에 의한 음질저하를 최소화할 수 있다[11].

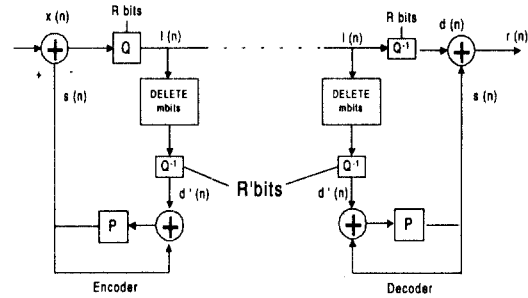


그림 2. ADPCM에 적용된 LSB Dropping시스템
Fig. 2 LSB dropping system for ADPCM

III. 제안하는 에러 은닉 기법

프레임 안에 에러가 있는가를 체크하기 위해 각 프레임마다 CRC(Cyclic Redundancy Code)가 삽입된다. 이 CRC는 비트스트림내의 가장 중요한 정보인 헤더를 검사하고 있다. 전송 중에 비트스트림 내에 에러가 발생하면 복호화기는 CRC에 의해서 에러를 발견할 수 있다. 그러나, 복호화기는 헤더정보, 비트할당 정보, 척도계수선택정보 등의 중요한 제어 정보에 에러가 있으면 그 프레임을 복호화할 수 없게 된다. 그래서 복호화기는 이러한 정보에 에러가 있는 것을 발견하면 그 프레임을 버리고 에러가 있는 프레임을 복원하기 위해 에러 은닉 기법을 사용하게 된다.

MPEG-2 오디오(ISO/IEC 13818-3)나 AC-3 오디오에 의해 부호화된 프레임이 채널상의 에러 또는 복호 과정에서 발생한 에러로 인해 복호되지 못할 경우에 앞 뒤의 올바르게 복호된 프레임을 사용하여 그 손실

된 프레임을 은닉하고자 한다.(실제 MPEG-2 오디오 계층 2에서는 1152 샘플을 한 프레임으로 그리고 AC-3 오디오에서는 1536 샘플을 한 프레임으로 사용하지만, 실험에서는 음성 신호일 경우는 64 샘플을, 오디오 신호일 경우에는 256 샘플을 한 프레임이라고 가정한다.)

1. 삼각 변환 커널(Sinusoidal transform kernel)을 이용한 에러 은닉

MPEG-2 오디오나 AC-3 오디오에서는 프레임마다 CRC를 사용하므로 어떤 프레임이 에러로 인해 손실되었는지 복호화기에서 알 수 있다. 오디오 신호의 한 프레임에서 에러가 발생하여 그 프레임이 복호되지 못하고 손실 되면 그림 3과 같이 앞 뒤의 올바르게 복호된 프레임을 사용하여 손실된 프레임을 은닉한다. 삼각 변환 커널을 이용한 에러 은닉 기법의 전체적인 구성도를 그림 4에 나타내었다.

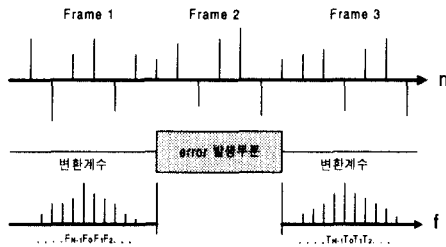


그림 3. 앞 뒤 프레임으로부터의 에러 프레임 복구
Fig. 3 Reconstruction of a error frame using the previous and the next frame

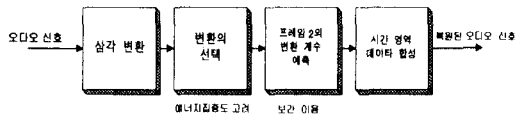


그림 4. 삼각 변환 커널을 이용한 에러 은닉 기법의 구성도
Fig. 4 The Diagram of the sinusoidal transform kernel method

먼저, 프레임 1과 프레임 3의 오디오 신호들을 각각 DFT(Discrete Fourier Transform), π -ODFT (π -Offset Discrete Fourier Transform), DCT(Discrete Cosine Transform), DST(Discrete Sine Transform), DHT (Discrete Hartley Transform) 등의 삼각 변환을 이용하여 5쌍의 변환 계수들을 구한다.

구해진 5쌍의 변환 계수들의 에너지 집중도를 조사하여 프레임 1과 프레임 3의 오디오 신호에 대해 5개의 삼각 변환중에서 에너지 집중도가 가장 높은 하나의 변환을 선택할 수 있다. 이렇게 선택된 변환을 손실된 프레임의 오디오 신호를 복원하는데 사용한다. 손실된 프레임의 앞 뒤 프레임의 변환 계수들을 보간하여 손실된 프레임의 변환 계수들을 예측하고 예측된 변환 계수들을 합성하여 손실된 프레임의 오디오 신호를 복원한다.

에너지 집중도를 이용하여 5개의 삼각 변환중에서 하나의 변환을 선택한다는 것은 프레임 1과 프레임 3의 오디오 신호의 성분을 가장 잘 특징짓는 변환 계수 성분들을 찾아낸다는 것이며, 오디오 신호가 짧은 시간동안에 정상신호(Stationary signal)라는 가정하에서 이 변환 계수들은 프레임 2에서 급격히 변화할 수 없다. 그러므로 에너지 집중도가 가장 높은 변환을 사용하여 프레임 1과 프레임 3의 오디오 신호의 특징을 가장 잘 특징짓는 변환 계수들을 구한다.

에너지 집중도를 고려하여 5개의 삼각 변환 중에서 선택된 변환을 사용하여 구해진 프레임 1의 변환 계수를 F_k , 프레임 3의 변환 계수를 T_k 로 표시하고 F_k 와 T_k 로부터 예측되는 프레임 2의 변환 계수를 S_k 로 표시한다(그림 5 참조).

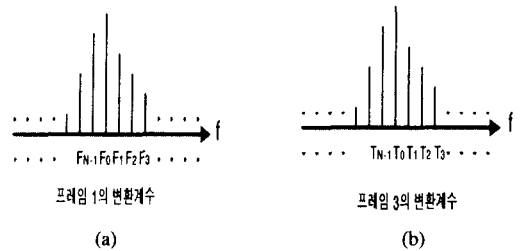


그림 5. 앞 뒤 프레임의 오디오 신호의 변환 계수 (a) 프레임 1의 변환 계수 (b) 프레임 3의 변환 계수

Fig. 5 Transform coefficients of audio signal of the previous and the next frame (a) Transform coefficients of 1st frame (b) Transform coefficients of 3rd frame

프레임 1의 오디오 데이터 변환계수	$F_0 F_1 \dots F_{N-1}$
프레임 2의 오디오 데이터 변환계수	$S_0 S_1 \dots S_{N-1}$
프레임 3의 오디오 데이터 변환계수	$T_0 T_1 \dots T_{N-1}$

다음, k번째 변환 계수인 F_k 와 T_k 를 이용하여

S_k 를 예측한다. F_k 와 T_k 의 차이를 N 등분으로 선형 보간하여 프레임 2의 n 번째 오디오 신호의 변환 계수의 예측을 식 (1)과 같이 앞 뒤 프레임의 변환 계수로부터 할 수 있다. 이 때 n 은 시간 영역에서 프레임 내의 오디오 데이터 순서를 나타낸다.

$$S_k(n) = \left[\frac{T_k - F_k}{N+1} \times (n+1) + F_k \right], \quad (1)$$

$$0 \leq k \leq N-1, \quad 0 \leq n \leq N-1$$

그림 6은 F_k 와 T_k 의 차이를 N 등분으로 선형 보간하는 과정을 보이고 있다. 실제로 N 개의 변환 계수에 대해 모두 선형 보간을 하지만 여기서는 저주파 2

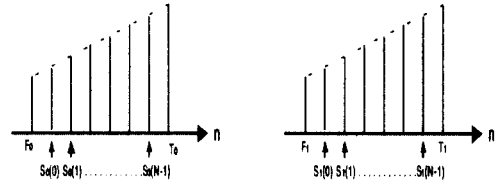


그림 6. 프레임 2 오디오 신호의 변환 계수 예측(저주파 2개)
Fig. 6 The prediction of transform coefficients of audio signal of the 2nd frame(Only 2 low frequency)

개만 보였다. 그러므로, 식 (1)을 이용하여 프레임 2의 n 번째 오디오 신호에 대한 변환 계수 $S_k(n)$, $k = 0, \dots, N-1$ 을 예측한다.

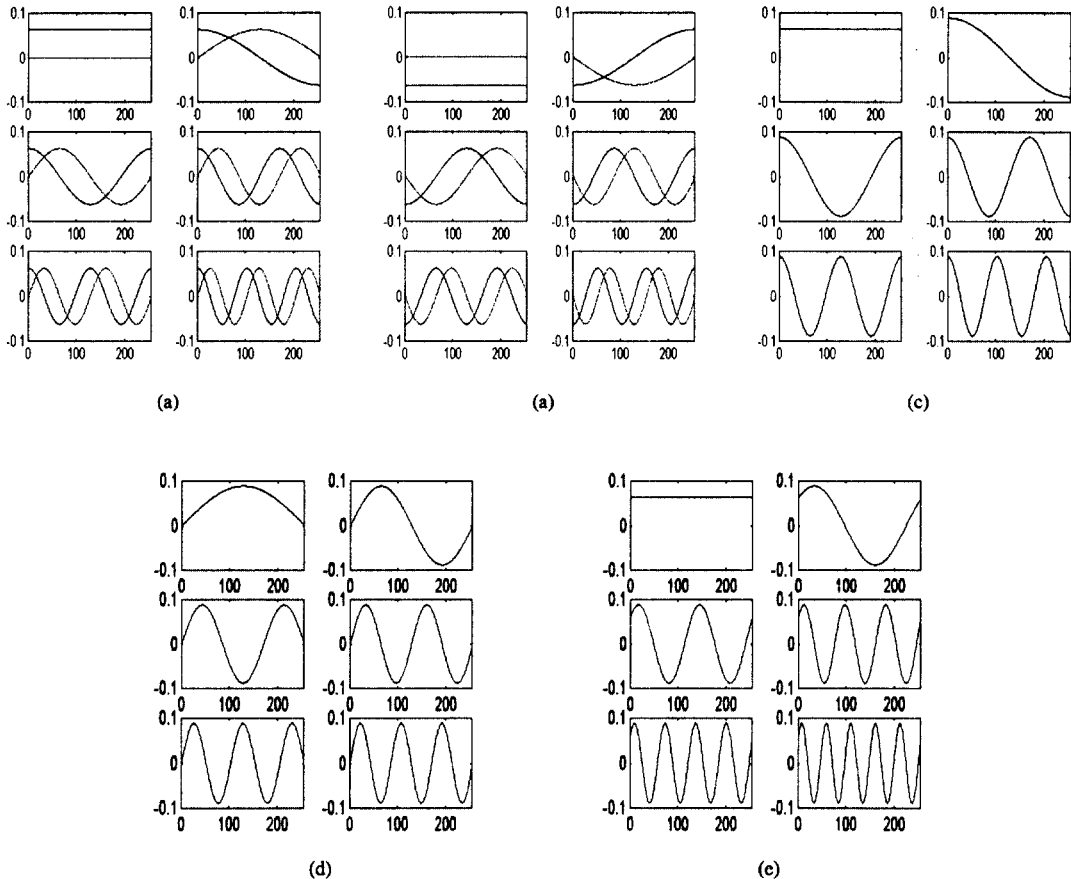


그림 7. 합성에 사용된 변환 커널(저주파 6개) (a) DFT의 변환 커널 (b) π -ODFT의 변환 커널 (c) DCT의 변환 커널 (d) DST의 변환 커널 (e) DHT의 변환 커널

Fig. 7 Transform kernels used for synthesis (Only 6 low frequency kernels) (a) Transform kernels of DFT (b) Transform kernels of π -ODFT (c) Transform kernels of DCT (d) Transform kernels of DST (e) Transform kernels of DHT

그 다음 예측된 프레임 2의 n번째 오디오 신호에 대한 변환 계수 $S_k(n)$, $k = 0, \dots, N-1$ 을 합성하여 시간 영역에서의 오디오 신호를 만들어 낸다. 이 합성의 과정에 사용되는 변환 커널을 그림 7에 나타내었다. 실제로는 N개의 커널이 모두 사용되지만 여기서는 저주파 6개만 보였다. 그리고, 복원된 프레임 2의 오디오 신호들이 프레임 1과 프레임 3과 연속성을 가지기 위해서는 오디오 신호를 합성할 때 사용되는 변환 커널들이 프레임간에 서로 연속되어야 한다. 각각의 변환에 대한 시간영역에 대한 합성식은 표 1에 나타나 있으며, 각각의 변환 커널들은 프레임간의 연속성을 지니고 있다.

$\hat{s}(n)$ 은 시간 영역에서 복구된 프레임 2의 오디오 신호이다. 이렇게 표 1에서의 합성식을 사용하여 합성한 오디오 신호는 앞 뒤의 프레임과 연속성을 갖게 된다. 그림 8은 에러에 의해 손실된 프레임 2가 에러 은닉 기법에 의해 복구된 모습을 보여준다.

2. 패턴 매칭(Pattern Matching)을 이용한 에러 은닉

지금까지 제안한 에러 은닉 방법들은 변환 영역을 기본으로 하여 시간 영역의 오디오 신호를 복원하는 방법이었다. 하지만 패턴 매칭을 이용한 에러 은닉 기법은 시간 영역을 기본으로 하는 방법이다. 에러로 인해 손실된 프레임의 앞 뒤의 오디오 신호를 이용하여 시간 영역에서 파형의 패턴 매칭을 통해서 가장 유사한

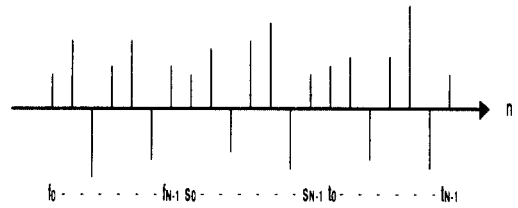


그림 8. 프레임 2의 오디오 신호의 복원
Fig. 8 The reconstruction of audio signal in the 2nd frame

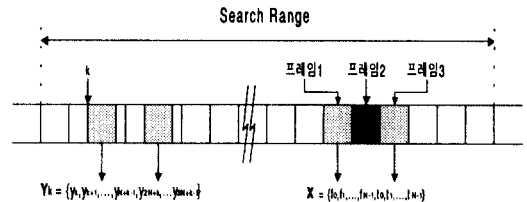


그림 9. 앞 뒤 프레임을 이용한 패턴 매칭
Fig. 9 Pattern matching using the previous and the next frame

부분으로 손실된 프레임을 대체 시키고자 하는 방법이다[13].

프레임 2가 에러로 인해 손실되었다면 그림 9와 같이 앞 뒤의 프레임 1과 프레임 3의 오디오 신호로부터 파형의 패턴 매칭의 기준이 되는 벡터 X를 만들고,

표 1. 오디오 신호의 복원 시 사용되는 합성식
Table 1. Synthesis formulas used for audio signal reconstruction

변환	오디오 신호의 복원 시 사용되는 합성식
DFT	$\hat{s}(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k(n) \exp[j(\frac{2\pi kn}{N})], \quad 0 \leq n \leq N-1$ $S_k(n) : \text{복소수}$
π -ODFT	$\hat{s}(n) = \frac{1}{\sqrt{N}} \sum_{k=0}^{N-1} S_k(n) \exp[j(\frac{2\pi kn}{N} + \pi)], \quad 0 \leq n \leq N-1$ $S_k(n) : \text{복소수}$
DCT	$\hat{s}(n) = \sum_{k=0}^{N-1} (-1)^k \alpha(k) S_k(n) \cos[\frac{\pi(2n+1)k}{2N}], \quad 0 \leq n \leq N-1$ $\alpha(0) \equiv \sqrt{\frac{1}{N}}, \quad \alpha(k) \equiv \sqrt{\frac{2}{N}} \text{ for } 1 \leq k \leq N-1$
DST	$\hat{s}(n) = \sum_{k=0}^{N-1} (-1)^{k+1} \alpha(k) S_k(n) \sin[\frac{\pi(n+\frac{1}{2})(k+1)}{N}], \quad 0 \leq n \leq N-1$ $\alpha(N-1) \equiv \sqrt{\frac{1}{N}}, \quad \alpha(k) \equiv \sqrt{\frac{2}{N}} \text{ for } 0 \leq k \leq N-2$
DHT	$\hat{s}(n) = \frac{2}{\sqrt{N}} \sum_{k=0}^{N-1} S_k(n) \sin[(\frac{2\pi kn}{N} + \frac{\pi}{4})], \quad 0 \leq n \leq N-1$

탐색 영역(Search Range)안에 있는 오디오 신호로부터 파형의 패턴 매칭의 대상이 되는 벡터 Y_k ($k \in$ Search Range)를 구한다. 프레임 1의 오디오 신호를 f_k , 프레임 3의 오디오 신호를 t_k , 탐색 영역내의 오디오 신호를 y_k 로 표시하면 벡터 X 와 벡터 Y_k 는 다음과 같다.

$$X = \{f_0, f_1, \dots, f_{N-1}, t_0, t_1, \dots, t_{N-1}\}$$

$$Y_k = \{y_k, y_{k+1}, \dots, y_{N+k-1}, y_{2N+k}, y_{2N+k+1}, \dots, y_{3N+k-1}\}$$

다음, 벡터 Y_k 와 벡터 X 와의 상관계수는 식(2)를 이용해서 구할 수 있으며, 탐색 영역내에서 벡터 Y_k 의 시작점 k 를 이동시키면서 벡터 Y_k 와 벡터 X 와의 상관계수가 가장 큰 위치 $k = K$ 를 찾는다(식(3)참조).

$$\text{상관계수(Correlation)} = \frac{\langle X \cdot Y_k \rangle}{\|X\| \cdot \|Y_k\|} \quad (2)$$

여기서,

$$\langle X \cdot Y_k \rangle = \sum_{i=0}^{N-1} f_i \cdot y_{k+i} + \sum_{i=0}^{N-1} t_i \cdot y_{2N+k+i}$$

$$\|X\| \cdot \|Y_k\| = \left\{ \sqrt{\sum_{i=0}^{N-1} f_i^2 + \sum_{i=0}^{N-1} t_i^2} \right\} \cdot \left\{ \sqrt{\sum_{i=0}^{N-1} y_{k+i}^2 + \sum_{i=0}^{N-1} y_{2N+k+i}^2} \right\}$$

$$\text{시작점 } K: \text{MAX} \left\{ \frac{\langle X \cdot Y_k \rangle}{\|X\| \cdot \|Y_k\|} \right\} \text{인 } k = K \text{ 지점, } \forall k \text{ in Search Range} \quad (3)$$

이러한 방법으로 찾은 K 를 시작점으로 하는 벡터 Y_K 를 이용하여 손실된 프레임 2의 오디오 신호를 벡터 Y_K 사이 에 있는 데이터 $\{y_{N+K}, y_{N+K+1}, \dots, y_{2N+K-1}\}$ 로써 대체할 수 있다. 그림 10은 대체시키는 과정을 나타낸다.

손실된 프레임 2의 오디오 신호에 대해 대체할 부분을 찾는 다음의 문제는 앞 뒤의 프레임과 대체된 부분이 연속성을 갖는 것이다. 이는 프레임 1의 마지막 데이터(f_{N-1})와 대체될 부분의 첫 데이터(y_{N+K})의 차이와 대체될 부분의 마지막 데이터(y_{2N+K-1})와 프레임 3의 첫 데이터(t_0)의 차이를 고려해서 식 (4)

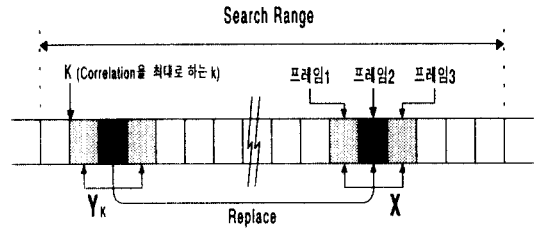


그림 10. 손실된 프레임 2의 오디오 신호의 복구
Fig. 10 Audio signal reconstruction of the lost 2nd frame

에서 구해진 보상항을 대체될 부분에 전체적으로 증가시키거나 감소시킴으로써 해결할 수 있다.

$$\text{Compensate}(n) = \begin{cases} \frac{\{(f_{N-1} - y_{N+K}) - (y_{2N+K-1} - t_0)\}}{N+1} \\ (n+1) + (f_{N-1} - y_{N+K}), & 0 \leq n \leq N-1 \\ 0, & \text{otherwise.} \end{cases} \quad (4)$$

이러한 과정을 통해서 에러에 의해 손실된 프레임 2는 탐색 영역(Search Range)안에서 상관계수가 가장 높은 파형으로 대체되고, 앞 뒤 프레임의 오디오 신호와 연속성을 가지도록 복원된다.

IV. 실험 결과 및 고찰

MPEG-2 오디오(ISO/IEC 13818-3)나 AC-3 오디오에 의해 부호화된 한 프레임이 채널에서의 에러에 의해 손실되었고, 앞 뒤의 올바르게 복호된 프레임을 사용하여 손실된 프레임을 은닉하고자 한다는 것을 가정하고 있다. 실제 MPEG-2 오디오 계층 2에서는 1152 샘플이, 그리고 AC-3 오디오에서는 1536 샘플이 한 프레임이지만 실험에서는 음성신호일 경우는 64샘플, 오디오 신호일 경우에는 256샘플을 한 프레임이라고 가정하고 있다. 선택된 3가지(표 2 참조)의 오디오 신호에 대해 목음화하는 방법과 제안된 에러 은닉 기법을 각각 사용하여 그 음질과 파형을 비교하고 있다.

표 2. 실험에 사용된 오디오 신호
Table 2. Audio signal used for simulation

	3가지 오디오 신호	샘플링 주파수	샘플당 비트수	프레임당 샘플수
1	음성 신호	8 kHz	16 bit	64 샘플
2	Mozart Magic Flute	44.1 kHz	16 bit	256 샘플
3	Kenny G 연주곡	44.1 kHz	16 bit	256 샘플

제안된 에러 은닉 기법의 성능을 평가하기 위해서 객관적인 평가는 식(5)와 같이 Relative Squared Error를 이용한 SNR(Signal-to-Noise Ratio)을 사용하고, 주관적인 평가는 복구된 파형을 직접 청취하는 방법을 사용하였다[10].

$$e^2(l) = \frac{\sum_{n=0}^{N-1} [x(l+n) - \hat{x}(l+n)]^2}{\sum_{n=0}^{N-1} x^2(l+n)}$$

$$SNR(l) = 10 \cdot \log_{10} \frac{1}{e^2(l)} \quad (5)$$

$x(n)$: 원래의 파형, $\hat{x}(n)$: 복구된 파형

음성 신호는 아주 짧은 시간 영역의 구간에서는 정상신호이고, 대부분의 에너지는 모음에 많이 몰려 있지만 음성의 인식력 관점에서는 백색잡음과 비슷한 특성을 가진 자음에 영향을 많이 받는 특징을 가지고 있다. 그래서 음성 신호의 경우에는 에러의 발생 위치에 따라 에러 은닉의 성능이 달라진다. 자음의 에러에 대한 에러 은닉 성능은 삼각 변환 커널을 이용하거나 패턴 매칭을 이용하더라도 모음의 에러에 대한 성능보다 낮음을 알 수 있다.

에러가 발생한 프레임은 묵음으로 처리한 파형과 각 에러 은닉 기법으로 복구된 프레임의 파형을 그림 11에 나타내었고 식 (5)를 이용하여 SNR을 구하였다.

한 프레임은 길이는 8ms이고 64샘플로 구성되어 있다. 또한 프레임당 샘플의 수가 패턴 매칭을 이용한 에러 은닉 기법에서는 음질에 큰 영향을 준다는 것을 알 수 있다. 즉 프레임당 샘플수가 많을 때에는 복구된 파형이 실제의 파형과 상당한 차이가 있으며, 음질도 크게 떨어짐을 알 수 있다.

Mozart Magic Flute곡과 Kenny G의 연주곡은 44.1 kHz로 샘플링되고 샘플당 16비트가 할당되었다. 한 프레임의 길이는 약 5ms이며 256샘플로 이루어져 있다. 에러가 발생한 프레임은 묵음으로 처리한 파형과 각 에러 은닉 기법으로 복구된 프레임의 파형을 그림 12와 그림 13에 각각 나타내었다.

손실된 프레임에 대해 묵음으로 처리하는 기존의 에러 은닉 기법에 대한 파형을 살펴보고, 예상한 대로 청취를 하면 묵음으로 처리된 프레임에서는 귀에 거슬리는 잡음이 발생한다.

삼각 변환 커널을 이용한 에러 은닉 기법에서는 앞 뒤 프레임과 복구된 프레임간의 불연속적인 현상을 해결하기 위해 합성할 때에 커널의 연속성을 고려하였다. 그 결과 복구된 파형을 살펴보면 불연속적인 현상은 해결되었지만 복구된 파형이 전체적으로 좌우로 뒤집혀 있음을 알 수 있다. 이러한 현상 때문에 청취시에는 그 인식력이 떨어짐을 알 수 있다. 하지만, 그 음질은 묵음으로 처리된 것보다는 나음을 알 수 있다. 손실된 프레임을 0으로 대체하여 묵음화 하는

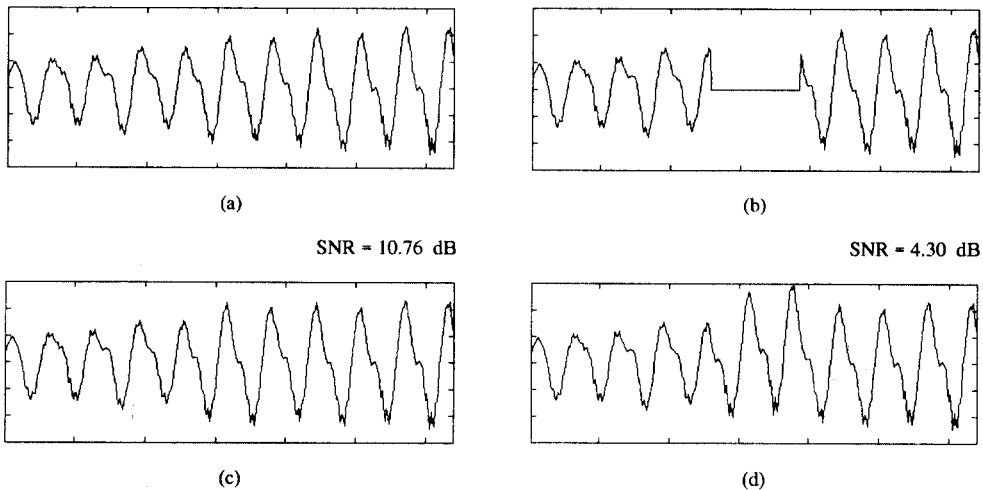


그림 11. 음성 신호의 에러 은닉 (a) 실제 프레임 2의 파형 (b) 묵음화된 프레임 2의 파형 (c) 삼각 변환 커널을 이용한 에러 은닉 기법으로 복구된 프레임 2의 파형 (d) 패턴 매칭을 이용한 에러 은닉 기법으로 복구된 프레임 2의 파형

Fig. 11 Error concealment of the speech signal (a)Original signal of the 2nd frame (b) Muted signal of the 2nd frame (c) Reconstructed signal of the 2nd frame using the sinusoidal transform kernel method (d) Reconstructed signal of the 2nd frame using the pattern matching method

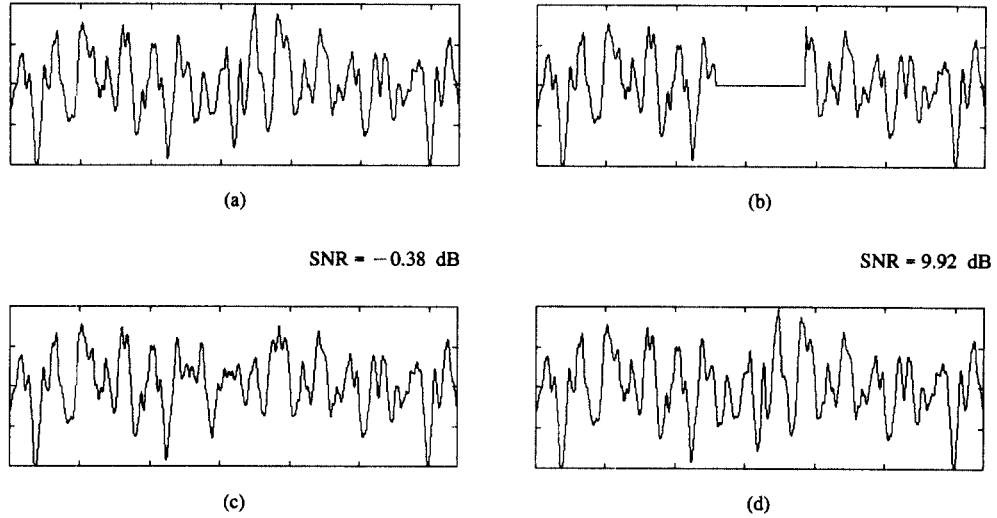


그림 12. Mozart Magic Flute곡에 대한 에러 은닉 (a) 실제 프레임 2의 파형 (b) 묵음화된 프레임 2의 파형 (c) 삼각 변환 커널을 이용한 에러 은닉 기법으로 복원된 프레임 2의 파형 (d) 패턴 매칭을 이용한 에러 은닉 기법으로 복원된 프레임 2의 파형

Fig. 12 Error concealment of the Mozart Magic Flute (a) Original signal of the 2nd frame (b) Muted signal of the 2nd frame (c) Reconstructed signal of the 2nd frame using the sinusoidal transform kernel method (d) Reconstructed signal of the 2nd frame using the pattern matching method

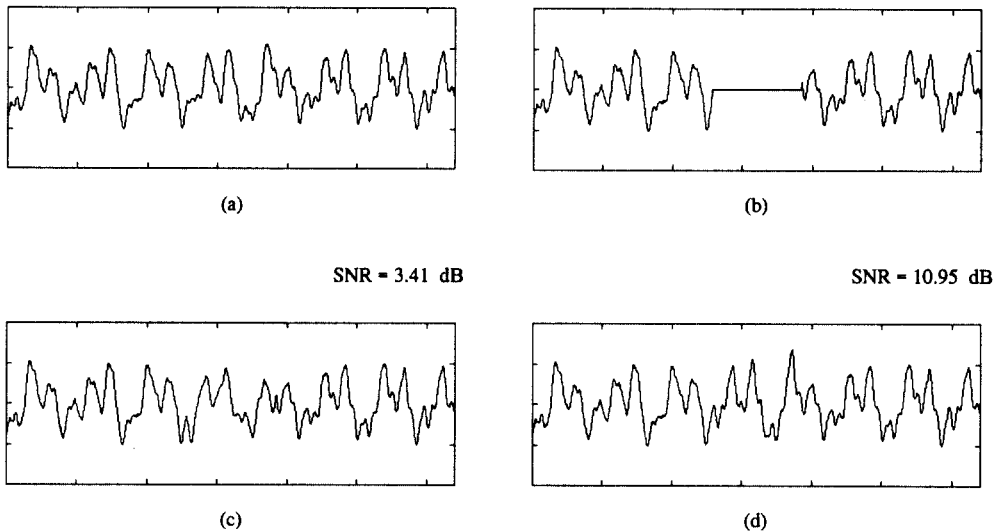


그림 13. Kenny G 연주곡에 대한 에러 은닉 (a) 실제 프레임 2의 파형 (b) 묵음화된 프레임 2의 파형 (c) 삼각 변환 커널을 이용한 에러 은닉 기법으로 복원된 프레임 2의 파형 (d) 패턴 매칭을 이용한 에러 은닉 기법으로 복원된 프레임 2의 파형

Fig. 13 Error concealment of the Kenny G's music (a) Original signal of the 2nd frame (b) Muted signal of the 2nd frame (c) Reconstructed signal of the 2nd frame using the sinusoidal transform kernel method (d) Reconstructed signal of the 2nd frame using the pattern matching method

기법의 SNR은 항상 0 dB를 나타낸다. 삼각 변환 커널을 이용한 에러 은닉 기법에서는 복원된 신호의 SNR이 0 dB보다 낮게 나타나기도 한다. 하지만 복원된 음질은 목음화 기법보다 나은 음질을 들려 준다.

이러한 주파수 영역에서의 접근과는 달리 시간 영역에서 파형의 패턴 매칭을 이용한 에러 은닉 기법의 경우에는 복구된 파형이 다른 에러 은닉 기법보다 원 오디오 신호의 파형에 가까움을 볼 수 있다. 그리고 실제 청취시에서도 가장 음질이 좋게 나왔다. 그러나, 이 기법은 탐색 영역(Search range)안에 있는 오디오 신호의 정보를 이용하므로 삼각 변환 커널을 이용한 에러 은닉 기법보다 많은 오디오 신호의 정보를 이용하고 있으며 여러 프레임을 저장할 수 있는 버퍼가 필요하다.

위에서 제안한 에러 은닉 기법을 사용하여 복원한 파형에 대한 SNR를 표 3에 다시 나타내었다. 그리고 복원된 신호를 청취하여 그 음질을 서로 비교해 보면 패턴 매칭을 이용한 에러 은닉 기법이 가장 좋았으며, 삼각 변환 커널을 이용한 에러 은닉 기법, 목음화의 순으로 음질이 낮아짐을 알 수 있다. 하지만 사람의 음성, 특히 짧은 구간동안 정상신호인 모음의 경우에는 삼각 변환 커널을 이용한 에러 은닉 기법이 가장 좋은 성능을 나타내고 있음을 알 수 있다.

표 3. 복원된 오디오 신호들의 SNR
Table 3. SNR of the reconstructed audio signals

	3 가지 오디오 신호	삼각 변환 커널을 이용한 에러 은닉	패턴 매칭을 이용한 에러 은닉
1	음성 신호	10.79 dB	4.30 dB
2	Mozart Magic Flute	-0.38 dB	9.92 dB
3	Kenny G 연주곡	3.41 dB	10.95 dB

V. 결 론

디지털 오디오 신호는 전송상의 에러에 많은 영향을 받는다. 이러한 문제를 해결하기 위해 정교한 에러 검출과 정정 기법들이 제안되었다. 그러나, 기존의 에러 정정 기법은 검출 및 정정 능력 이내의 에러들에 대해서는 기능을 잘 발휘하지만, 에러들이 더 발생하여 정정 능력을 넘어서면 에러 은닉 기법이 적용되어야 한다. 본 논문에서는 목음화, 반복, 주파수 영역에서의 클리핑 등 기존의 에러 은닉 기법에 대해 살펴보고, 삼각 변환 커널을 이용한 에러 은닉 기법과 패턴 매칭을 이용한 에러 은닉 기법을 제안하였

다. 또한 선택된 세 가지의 오디오 신호에 대해 제안된 에러 은닉 기법을 적용하여 실험한 결과 목음화나 삼각 변환 커널을 이용한 에러 은닉 기법보다 패턴 매칭을 이용한 에러 은닉 기법이 더 좋은 음질을 나타내었다.

향후 연구과제로는, 기존의 에러 은닉 기법이 주파수 영역과 시간 영역으로 따로 접근을 하고 있지만, 변환 영역과 시간 영역의 접근을 함께 사용하는 하이브리드 방법 등이 보다 연구되어야 할 분야이다.

참 고 문 헌

1. ISO CD 11172-3 "Coding of moving picture and associated audio for digital storage media up to about 1.5Mbit/s, Part 3 Audio," 1993.
2. ISO/IEC 13818-3 "Information Technology - Generic Coding of Moving Pictures and Associated Audio : Audio", 1994.
3. M. Barberis and E. F. Schroeder, "Burst Error Concealment for Digital Audio Tape and Broadcast Application," 90th AES Convention, preprint 3012, Paris 1991.
4. ATSC Document A/52 "Digital Audio Compression (AC-3) Standard," 1995.
5. H. Cha and S. Lee, "Error Concealment Method for Digital Audio Coding", International workshop on HDTV and the evolution of Television, 8B8. 34-40, Taipei, Nov. 1995.
6. D. Wiese, "Error Concealment Strategies for Digital Audio Broadcasting," 92nd AES Convention, preprint 3264, Vienna 1992.
7. D. Wiese, "Optimization of Error Detection and Concealment for ISO/MPEG/AUDIO CODECS Layer-I and II," 93rd AES Convention, preprint 3368, San Francisco 1992.
8. G. Plenge, C. Weck, D. Wiese, "Combined Channel Coding and Concealment", IEE Colloquium on 'Terrestrial DAB - Where is it going?', pp. 78, Feb. 1993.
9. J. Herre and E. Eberlein, "Error Concealment in the Spectral Domain".
10. W. Etter, "Restoration of a Discrete-Time Signal Segment by Interpolation Based on the Left-sided and Right-Sided Autoregressive Parameters," IEEE Trans. Signal Process., vol. 44, no. 5, pp. 1124-

