

일반논문-08-13-5-18

## 대역 확장을 통한 MP3 오디오의 음질 향상

허 소 영<sup>a)</sup>, 김 인 철<sup>a)‡</sup>

## A Quality Improvement of MP3-Coded Audios Using Bandwidth Extension

So Young Heo<sup>a)</sup>, and Rin Chul Kim<sup>a)‡</sup>

## 요 약

본 논문에서는 대역 확장 기법을 이용한 MP3 오디오 신호의 음질 향상 방법에 대해 살펴본다. Liu가 제안한 고주파 복원 기법을 근간으로 하여, 제안한 음질 향상 기법에서는 복원 시작 지점을 결정하는 방법과 개선된 선형 예측 방법을 제시하였다. 또한 고주파 대역 성분 생성에 저주파 성분을 복사해 이용하는 방법과 함께 AWGN을 이용하는 방법을 제시한다. 제안한 방법은 주관적 음질 평가를 통하여 MP3 오디오의 품질을 개선시킬 수 있음을 확인할 수 있었다.

## Abstract

In this paper, we investigate methods to enhance the perceptual quality of MP3-coded audios. Based on the high frequency reconstruction method by Liu, in the proposed method, we determine adaptively the starting point of high frequency reconstruction. We also present an improved linear estimation method. For high frequency component generation, we compare two methods. One is a replication of low-frequency components and the other is an insertion of additive white Gaussian noise signals. Through subjective tests, we shall show that the proposed method can improve the perceptual quality of MP3-coded audio.

Keywords : MP3, bandwidth extension, high frequency reconstruction, band-limited audio

## 1. 서 론

MPEG-1 Audio LayerⅢ<sup>[1]</sup>, 즉 MP3는 디지털 음성신호를 고효율로 압축, 부호화하기 위한 규격으로 오늘날 가장 대중적인 오디오 표현 형식이다. MP3 오디오는 원본 CD 오디오와 유사한 음질을 유지하면서 그 용량이 원본 CD 오디오의 약 1/11 정도로 매우 작아서 음악 데이터의 배포에 적합하다. 또한 다양한 MP3 부호화 S/W가 제공되고 있기 때문에 원본 CD 오디오에서 MP3 형식으로의 변환이

아주 용이해 누구나 쉽게 이용할 수 있다.

MP3는 주파수 영역에서 사람의 청각 특성을 나타내는 심리음향 모델(psychoacoustic model)을 이용하여 부호화하는 청각 오디오 부호화 기법의 대표적인 예이다. 일반적으로 MP3는 매우 높은 압축률을 성취하되 귀에 거슬리지 않는 음질을 제공하기 위해, 사람 귀에 잘 들리지 않는 약 14~15kHz 이상의 고주파 성분을 부호화하지 않는다. 따라서 이러한 MP3 오디오의 단점을 해결하기 위한 노력으로 MP3 오디오의 고주파 성분을 효과적으로 추정해 복원하고자 하는 연구가 많이 진행되었다.<sup>[2-5]</sup>

본 논문은 MP3 오디오 형식을 그대로 유지하면서 고주파 성분을 복원하여 MP3 오디오의 음질을 향상시키는 방

a) 서울시립대학교 전자전기컴퓨터공학부

Dept. of Electrical & Computer Eng., University of Seoul

‡ 교신저자 : 김인철 (rin@uos.ac.kr)

법에 대하여 살펴본다. 고주파 성분을 복원하는 대표적인 방법으로 Liu가 제시한 고주파 복원 알고리즘<sup>[2]</sup>이 있다. Liu의 고주파 복원 알고리즘은 고주파 대역의 포락선을 추정하고, 추정된 포락선에 맞추어 부호화되지 않은 고주파 대역에 저주파 성분을 복사하여 고주파 성분을 복원한다. 이 방법은 MP3 복호화의 후처리 과정에서 고주파 성분을 복원하는 방법으로, 고주파 성분을 표현하기 위해 별도의 부호화 정보를 전송하는 spectral band replication (SBR)<sup>[4]</sup> 등과는 차이가 있다.

본 논문에서는 Liu 알고리즘을 개선하기 위해 MP3 부호화 S/W마다 부호화 되지 않는 고주파 대역이 다를 수 있다는 점을 고려하여 복원이 시작되는 주파수 대역을 추정을 통해 결정한다. 또한 대역 신호의 에너지 분포를 이용하여 포락선을 추정하며, 고주파 복원을 위해 저주파 복사뿐만 아니라 AWGN(additive white Gaussian noise)을 이용하는 방법도 고려하였다. 제안하는 기법은 Lame MP3 복호기에 구현하였으며 주관적 음질 평가를 통해 그 성능을 평가하였다.

논문의 구성은 다음과 같다. 제 2장에서 Liu가 제안한 고주파 복원 기법을 근간으로 고주파 복원 과정에 대하여 살펴본 후, 제 3장에서는 제안하는 알고리즘을 Liu의 고주파 복원 기법과 비교하여 설명하겠다. 제 4장에서 실험 결과를 보이고, 제 5장에서 결론을 맺겠다.

## II. 고주파 복원

본 절에서는 Liu의 고주파 복원 기법<sup>[2]</sup>을 중심으로 고주파 복원 과정에 대하여 간략하게 살펴보겠다. Liu의 기법은 MP3 복호기의 MDCT 영역에서 고주파 복원을 수행하므로 별도의 주파수 변환 과정을 필요로 하지 않는다. 일반적으로 MP3로 압축된 오디오 신호는 고주파 성분이 제거된 대역 제한 신호로 표현되므로, MDCT (modified discrete cosine transform) 계수 또한 약 15KHz 이하의 계수들만 전송된다. 이때, 오디오 신호는 하모닉 성분 (harmonic component)들로 구성된다는 가정 하에, Liu의 기법에서는 저주파 영역의 신호들을 고주파 영역으로 복사함으로써 고주파 복

원을 수행한다.

이를 위해 먼저 최소 자승 기법 (least square method)으로 대역 제한된 MDCT 계수를 이용하여 MDCT 계수의 로그 크기 포락선에 대한 선형 모델을 추정한다. 그런 다음, 저주파 대역의 일정 구간을 고주파 대역으로 반복 복사하여 고주파 대역 신호를 합성한다. 이때 합성된 고주파 신호들의 에너지를 앞서 예측한 포락선 크기로 맞춘다. 마지막으로 복원된 고주파 대역 신호를 입력된 저주파 신호에 합하여 전 대역으로 확장된 오디오 신호로 출력된다.

### 1. 포락선 추정

$X[k]$ 를 임의의 시간 프레임에서의  $k$ 번째 MDCT 계수라고 하고, MP3로 부호화 될 때  $k < k_c$  범위의 계수만 전송된다고 하자. 그리고 집합  $M$ 은 다음과 같이 차단주파수  $k_c - 1$ 까지  $N$ 개의 로그 크기 계수로 구성된다고 하자.

$$M = \{\ln|X[k_c - N]|, \ln|k_c - N + 1|, \dots, \ln|k_c - 1|\} \quad (1)$$

포락선 추정을 위해서  $\ln|X[k]|$ 를 다음과 같이 선형 모델로 근사화한다.

$$\ln|X[k]| = a_{opt} \cdot k + b_{opt} \quad (2)$$

이때,  $a_{opt}$  과  $b_{opt}$  은 집합  $M$ 에 속한  $N$ 개의 값을 이용하여 최소자승 기법 (least squared method)으로 추정할 수 있다. 효과적인 선형모델 추정을 위해 Liu는 다음과 같은 제한 조건을 두었다. 만약  $X[k]$ 가 0에 가까운 값을 가지면 집합  $M$ 에 속한 값들의 평균 로그 크기의 0.001배 되는 값으로  $\ln|X[k]|$ 을 대체하였다. 그리고 일반적으로 주파수가 높아짐에 따라 포락선은 감소하는 경향을 가지므로  $a_{opt}$ 의 크기는 0 보다 커지지 않도록 조정하였다.

### 2. 고주파 대역 생성

차단 주파수  $k_c$ 보다 높은 영역의 고주파 성분을 생성하는 방법은 다음과 같다. 먼저, 차단 주파수 이하에 위치하는

$U$ 개의 스펙트럼 성분으로 구성된 복원단위를 정하고, 이 복원 단위를 반복적으로 복사하여 고주파 대역을 생성한다. Liu는 전송률(bit rate)이 128kbps인 MP3 오디오에 대해서는  $U$  값을 1혹은 2kHz에 대응되는 값으로 설정하고, 전송률이 96kbps일 때는 4-5kHz에 대응되는 값으로 설정하였다.

그런 다음, 복사된 고주파 영역의 에너지가 식 (2)에서 추정된 포락선과 일치하도록 크기를 조정한다. 이때 복원 단위가 임펄스 등과 같이 비정상적인 성분이 포함되어 있다면 정상적인 고주파 복원이 되지 않는다. Liu는 추가적으로 검출비율 (detection ratio)을 정의하여 복원 단위가 정상적인 성분으로 구성되어 있을 때에만 고주파 복원이 수행되도록 제한하였다.

### III. 제안한 기법

제안한 고주파 복원 기법의 블록선도는 그림 1에 제시한 바와 같다. 제안한 기법은 Liu의 고주파 복원 기법을 두지만, 다음과 같은 차이점을 가지고 있다. 먼저 Liu의 고주파 복원 기법은 MDCT 영역에서 고주파 복원을 수행하였으나 제안한 기법은 IMDCT와 합성 필터 बैं크 (synthesis filter bank) 사이의 대역 신호를 이용하여 고주파 복원을 수행한다. 두 번째로, Liu의 고주파 복원 기법은 차단 주파수  $k_c$ 를 알고 있다는 전제하에 복원 알고리즘을 적용한다. 그러나 제안하는 기법에서 차단 주파수는 추정을 통해 가변적으로 설정된다. 마지막으로 제안한 기법은 Liu의 고주파 복원 기법과 같이 저주파 대역을 고주파 대역으로 복사함으로써 고주파 성분을 생성하는 방법과 함께

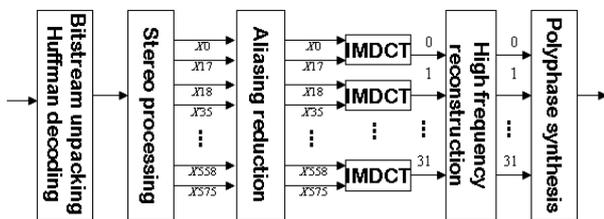


그림 1. 제안한 기법의 블록선도

Fig. 1. Block diagram of the proposed method

AWGN (additive white Gaussian noise)을 이용하여 고주파 성분을 생성하는 방법도 고려하여 그 결과를 비교한다.

계속하여 복원 시작 대역 결정, 선형 모델 계수 추정, 그리고 고주파 대역 생성에 대해 구체적으로 살펴보겠다.

#### 1. 복원 시작 대역 결정

MP3로 압축된 대부분의 오디오는 차단 주파수 14~16kHz 이상의 주파수 성분은 부호화하지 않는데, MP3 오디오의 차단 주파수는 압축할 때 사용하는 MP3 부호화 S/W에 따라 달라진다.

Liu의 고주파 복원 기법은 차단 주파수를 미리 알고 있다고 가정하고, 동일한 차단 주파수에 대하여 알고리즘을 적용한다. 만약 차단 주파수와 복원 주파수가 일치하지 않는다면, 전송된 주파수 성분의 일부가 손상되거나 복원되어야 할 고주파 대역 중 일부가 복원되지 않는 현상이 발생한다. 차단 주파수가 항상 일정하더라도 부호화 과정에서 차단 주파수 근방의 주파수 성분이 차폐되어 데이터를 상실할 수도 있다. 복원 주파수가 고정되어 있는 조건에서 이러한 상황이 발생한다면 복원 주파수 이하의 성분으로 구성되는 복원 단위가 0에 가까운 값들을 포함하게 된다. 이로 인해 복원된 고주파 대역에 주파수 공백(spectral hole)이 발생하게 되고 고주파 복원 성능이 현저하게 저하된다. 이러한 단점들을 보완하기 위해 제안한 기법에서는 MP3 오디오를 부호화 한 다음, 그 신호에 적합한 복원 주파수를 결정한다.

제안한 기법은 IMDCT 블록과 합성 필터 बैं크 사이의 32 대역 신호들을 이용한다. 복원 주파수를 결정하기 위해 먼저 차단 주파수로 예상되는 대역을 검색 주파수 범위로 설정하고, 그 주파수 범위 내의 대역별 에너지를 계산한다. 만약 한 대역의 에너지가 이전 대역의 에너지에 비해 매우 작고 그 에너지가 계속 유지되면, 그 대역을 복원 시작 대역  $n_c$ 로 결정한다. 본 논문에서는 차단 주파수보다 낮은 대역 간 에너지 차이는 5dB 이내이고 차단 주파수 전후의 에너지 차이는 매우 크다는 점을 고려하여, 대역 간 에너지가 10dB 이상 차이 나는 대역을 복원 시작 대역으로 설정한다. 검색 주파수 범위는 13.0kHz에서 16.5kHz 사이로 설정하

여, 14~16kHz 범위 내에 존재하는 차단 주파수를 검색할 수 있게 하였다. 또한 주파수 공백 등 복원 단위의 결손에 의한 영향을 최소화할 수 있도록 하였다.

2. 선형 예측을 이용한 포락선 추정

본 절에서는 선형 예측 기법을 이용하여 저주파 대역 데이터로부터 고주파 대역의 포락선을 예측하는 방법을 설명한다. Liu의 기법에서  $N$ 개의 MDCT 계수들을 이용하여 포락선을 추정하는 것에 반해, 제안하는 알고리즘에서는 대역 신호를 이용하여 포락선을 추정한다.

그림 2에서는 MP3 복호화 과정에서 각 단계별 스펙트럼을 도시하였다. 그림 2(a)는 엘리이싱 감축 이전의 로그 크기 스펙트럼이고, (b)는 IMDCT 이후의 32대역에 대한 에너지 스펙트럼이다. 그림 2(a)에 보인 것이 Liu 기법에서 사용한 데이터로서 비교적 넓은 범위에 값들이 분포하는 것을 알 수 있다. 또한, -100dB, 즉 0 근방의 계수 값들이 다수 존재하여 선형 추정의 정확도를 저하시킬 수 있다는 단점이 있다. 반면에 그림 2(b)에 보여준 것은 크기 스펙트럼이 아니라 대역별 에너지이고, 0 근방의 값이 없어 선형 추정의 정확도를 높일 수 있다는 장점이 있다.

제안한 알고리즘의 포락선 추정 방법을 구체적으로 설명하면 다음과 같다. 먼저, 10KHz 이하의 주파수 성분은 에너지가 비교적 크고, 복원하고자 하는 고주파 성분과의 연

관성이 낮으므로 10KHz 이상의 주파수 성분을 이용하여 에너지 포락선에 대한 선형 모델을 추정한다. 따라서 IMDCT 블록에서 산출되는 대역 신호 중에서 10kHz를 포함하는 14번 대역부터 앞 절에서 구한 복원 시작 대역 직전인  $n_c - 1$ 번째 대역까지의 각 대역별 에너지의 로그 값,  $P_n = 10\log_{10}|E_n|$ 을 구한다. 그런 다음,  $P_n$ 에 대한 선형 모델을 다음과 같이 설정한다.

$$P_n = 10\log_{10}|E_n| = a_0 + a_1n \tag{3}$$

여기서, 계수  $a_0$ 와  $a_1$ 은 다음과 같이 최소 자승법을 이용하여 얻을 수 있다.

$$a_0 = \left\{ \frac{1}{N} \sum_n P_n \right\} - a_1 \left\{ \frac{1}{N} \sum_n n \right\} \tag{4}$$

$$a_1 = \frac{\frac{1}{N} \sum_n n \cdot P_n - \left\{ \frac{1}{N} \sum_n n \right\} \left\{ \frac{1}{N} \sum_n P_n \right\}}{\frac{1}{N} \sum_n n^2 - \left\{ \frac{1}{N} \sum_n n \right\}^2} \tag{5}$$

여기서 모든 합은  $14 \leq n \leq n_c - 1$  범위에서 행해지고,  $N = n_c - 14$ 이다. 일반적으로 주파수 영역에서 대역 에너지는 주파수가 높아짐에 따라 감소하므로, 만약 식 (5)를 통해 추정된  $a_1$ 이 0보다 크다면 이러한 경우  $a_1 = 0$ 으로 설정한다. 식 (4) 및 (5)로 얻어진 계수들을 식 (3)에 대입하여  $n_c$  이상 대역의 에너지 추정 값으로 사용한다.

3. 고주파 대역의 복원

본 절에서는 부호화되지 않은 고주파 영역인  $n_c$  이상의 대역 신호를 복원하는 방법을 설명하겠다. 고주파 대역 복원을 위해 먼저, III.2 절에서 구한 선형 모델로 고주파 대역의 포락선을 추정한다. 그림 3은 복원하고자 하는 구간을 나타내었다. Part 1은  $n_c$ 번째 대역부터 28번째 대역까지로 구성되는데, 이 구간에서는 선형 모델로 포락선을 추정한다. 그러나 29번째 대역부터 31번째 대역까지로 구성되는 part 2 구간은 별도로 처리한다. 원본 CD 오디오를 살펴보

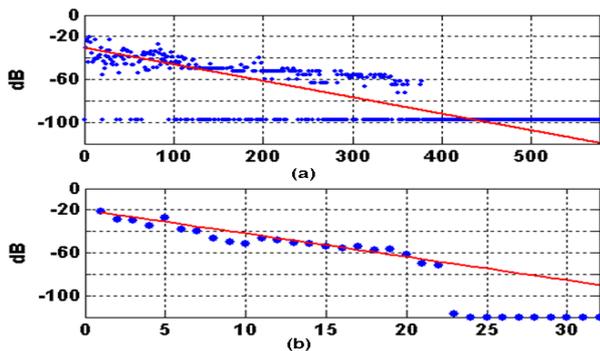


그림 2. 복호기에서 Aliasing 감축 이전의 로그 크기 스펙트럼 (a)과 IMDCT 이후의 에너지 스펙트럼(b)의 예  
Fig. 2. Example of log-magnitude spectrum before the aliasing reduction (a) and energy spectrum after IMDCT at the decoder

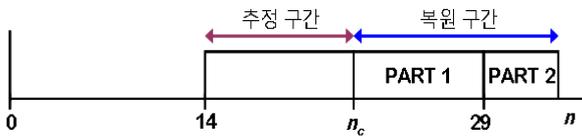


그림 3. 주파수 구간 정의

Fig. 3. Definition of frequency intervals

면 20kHz 이상의 주파수 성분은 표본화(sampling)할 때 앨리어싱을 방지하기 위한 저대역 통과 필터에 의해 감쇠가 이루어진다. 이러한 점을 고려하여 part 2 구간은 선형 모델에 의해 얻어진 포락선 추정치를 나이퀴스트(Nyquist) 주파수까지 추가적으로 감쇠시켰다. Part 2 내의 대역 신호의 포락선이 단조 감소하게 만들되, 나이퀴스트 주파수를 포함하는 마지막 대역에서 약 -30dB 만큼 추가적으로 감쇠가 일어나도록 설정하였다.

고주파 대역의 포락선을 추정한 다음, 고주파 대역 신호를 생성한다. 고주파 대역을 생성하는 방법으로 저주파 대역 신호를 복사하여 고주파 대역으로 옮기는 방법과 AWGN 신호를 삽입하는 방법, 두 가지를 이용해 그 결과를 비교해보았다. 먼저, 저주파 대역 신호를 복사하는 방법은 Liu의 기법과 유사하다. 이 방법에서는 약 4kHz 대역부터  $n_c - 1$ 번째 대역까지를 복사할 영역으로 설정한다.  $n_c$  이상인  $n$ 번째 대역 신호는  $n - 13$ 번째 대역 신호를 복사한 후, 선형 모델로 추정된 에너지가 되도록 대역 신호의 에너지를 수정한다. 이 방법은 복원이 시작되는 복원 대역  $n_c$ 의 변동에 관계없이 항상 동일한 주파수 쌍을 이루도록 되어 있다. 추후에 신호의 특성에 따라 복사할 신호를 달리 하도록 개선하면 보다 향상된 품질을 얻을 수 있을 것으로 기대된다.

다음으로, 복원할 대역의 생성을 위해 AWGN을 삽입할 수 있다. 이를 위해 평균이 0이고 분산이 1인 가우시안 난수(Gaussian random number)를 발생시킨다<sup>[6]</sup>. 그런 다음, 이 난수의 분산을 선형모델로 추정된 포락선에 맞추어  $n_c$  이상의 대역에 삽입한다. 이 방법은 부호화 되지 않은 고주파 대역은 그 크기가 비교적 작고, 부호화하기에 복잡한 신호로 되어 있다는 가정과 일반적으로 양자화 잡음은 백색 잡음으로 모델링된다는 가정에 부합되는 것이다<sup>[7]</sup>. 다음 절

에서 고주파 대역을 생성할 수 있는 두 가지 방법의 성능을 실험을 통해 비교하겠다.

#### IV. 실험 및 결과

본 절에서는 실험을 통해 고주파 복원을 이용한 MP3 음악의 음질 향상 방법의 성능을 평가하겠다.

MP3 복원 품질을 비교 평가하기 위해서는 MP3 오디오 뿐만 아니라 원본 오디오를 함께 비교하여야 한다. 이를 위해 실험에 사용되는 오디오 파일을 다음의 과정을 통해 얻었다. 먼저, CD에 수록된 오디오를 Adobe Audition 1.5를 이용하여 원본 wav로 변환한다. 이 때 wav 오디오의 표본화 주파수(sampling frequency)는 44.1kHz이고, 샘플 당 16bits로 표현된 스테레오 신호이다. 이 wav 오디오는 다시 Adobe Audition 1.5를 이용하여 MP3 파일로 변환된다. 변환된 MP3 파일은 표본화 주파수 44.1kHz, 전송률(bit rate) 128kbps, 그리고 차단 주파수 15.8kHz로 설정되었다. 제안한 음질 향상 알고리즘은 PC의 MSVC 6.0 환경에서 Lame 복호기 버전 3.97을 수정하여 구현하였다. 그리고 스펙트럼은 Blackman-Harris 윈도우를 사용한 512-pt FFT를 통해 얻었다.

성능 평가를 위해 다양한 곡에 대해 실험을 수행하여, 원본 오디오와 음질 차이를 비교적 크게 감지할 수 있는 곡을 선정하였다. 실험에 사용한 곡은 그룹 Queen의 "Don't stop me now" (일부, 40초)와 이문세의 "조조할인" (3분 58초), Westlife의 "Swear it again" (4분 7초) 등 3곡이다.

##### 1. 객관적 음질 평가

본 절에서는 본 논문에서 제시된 복원 시작 대역 결정 방법과 고주파 복원 기법의 성능을 주파수 영역에서 살펴 보겠다.

본 논문에서 복원 시작 대역  $n_c$ 는 입력 오디오 신호에 따라 가변적으로 결정된다. 그림 4에서는 첫 번째 실험곡인 Queen의 20초 위치에서 스펙트럼을 도시하였다. 가로축은 0부터 22KHz까지의 주파수를 나타내고, 세로축은 dB 단위

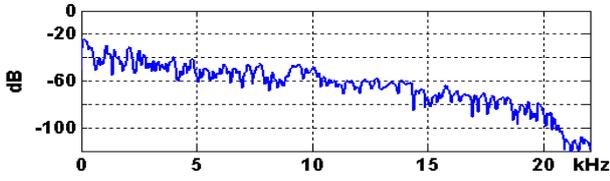


그림 4. Queen의 20초에서의 스펙트럼  
Fig. 4. Magnitude spectrum of the Queen at 20 second

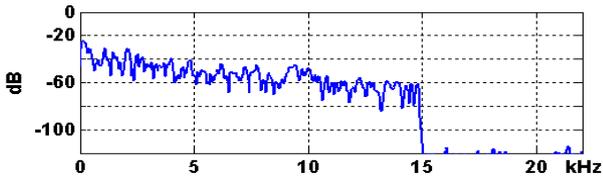


그림 5. MP3로 부호화된 Queen 20초에서의 스펙트럼  
Fig. 5. Magnitude spectrum of the MP3-coded Queen at 20 second

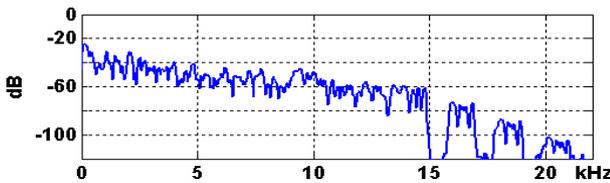


그림 6. 고정된 시작대역을 이용하여 복원된 Queen의 크기 스펙트럼  
Fig. 6. Magnitude spectrum of the decoded Queen using fixed starting band of reconstruction

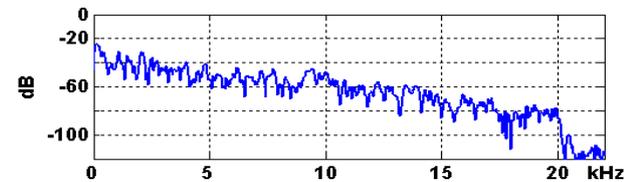


그림 7. 저주파 복사 방법을 사용하여 복원된 Queen의 크기 스펙트럼  
Fig. 7. Magnitude spectrum of the decoded Queen using the low-band replication method

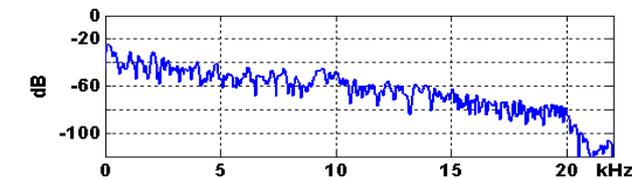


그림 8. 백색잡음 삽입으로 복원된 Queen의 크기 스펙트럼  
Fig. 8. Magnitude spectrum of the decoded Queen using an insertion of AWGN

의 크기를 나타낸다. 이 곡을 차단 주파수 15kHz인 MP3로 부호화하였을 때 동일한 시각에서의 스펙트럼을 그림 5에 도시하였다. 이와 같이 부호화된 곡을 차단 주파수 16kHz로 설정된 Liu 기법으로 복원하였을 때의 스펙트럼을 그림 6에 도시하였다. 그림 6을 참조하면, Liu의 기법에서는 복원 시작 대역이 차단 주파수와 일치하지 않아 복원되지 않은 주파수 영역 즉, 15kHz에서 16kHz 사이가 밸리(valley)와 같이 도시됨을 확인할 수 있다. 또한 복원 시작 지점 이전의 저주파 대역을 이용하여 고주파 대역을 생성하므로 차단 주파수와 복원 주파수 사이의 공백이 복원된 고주파 대역에 반복적으로 영향을 미치는 것을 확인할 수 있다. 이러한 문제는 복원 시작 대역을 가변적으로 추정함으로써 해결할 수 있다.

그림 7과 8에서는 각각 저주파 복사 및 AWGN 삽입의 방법으로 제안한 고주파 복원 알고리즘을 적용하였을 때 그림 4와 동일한 시각에서의 스펙트럼을 도시하였다. 그림 7 및 8에서는 복원 주파수를 자동적으로 설정함으로써 그림 6에서와 같은 주파수 공백(spectral hole)이 발생하지 않고 자연스럽게 스펙트럼이 복원됨을 알 수 있다. 또한, 그림 7 및 8의 주파수 포락선은 전 대역에서 그림 4의 원본 오디오와 유사하게 유지됨을 알 수 있다. 한편, 그림 7에 제시한 저주파 복사를 이용한 고주파 복원 기법은 저주파 영역에서의 패턴이 고주파 영역에 어느 정도 반복되고 있는 것으로 나타났다. 반면에 그림 8에 제시한 AWGN을 이용한 고주파 복원에서는 스펙트럼이 균일한 특성을 보임을 알 수 있다.

## 2. 주관적 음질 평가

본 논문에서 제안하는 고주파 복원 기법의 성능을 평가하기 위해 주관적 음질 평가를 수행하였다. 주관적 음질 평가는 평가자에 따라 매우 상이한 결과를 보일 수 있다는 단점이 있다. 특히, 초보자와 전문가의 품질 평가 결과는 매우 다르게 나올 수 있다고 알려져 있다. 본 장에서는 음질 평가에 익숙하지 않는 초보자들을 대상으로 주관적 평가를 시행하였다. 초보자의 경우 MP3 오디오와 원본 CD 오디오를 비교하면서 듣는 기회가 거의 없고, 대부분 보편적인 오

디오 형식인 MP3 오디오에 익숙해져 있다는 특징이 있다. 본 음질 평가에 앞서, 평가자가 원본 CD 오디오와 MP3 오디오의 차이를 인지할 수 있도록 한다. 그런 다음 오디오 형식을 알려주지 않고, 원본 CD 오디오, MP3 오디오, 저주파 복사를 이용해 고주파가 복원된 MP3 오디오, 그리고 AWGN을 이용해 고주파가 복원된 MP3 오디오를 들려주어 그 품질을 평가하도록 하였다. 각 실험 오디오는 후렴으로 들어가기 전부터 후렴부분을 포함하는 약 15~20초 구간을 들음으로서 고주파 대역의 유무에 따른 음질의 차이를 비교할 수 있도록 했다. 음질은 5단계로 평가하도록 했고, 평가자는 17명의 20대 학생들이다.

표 1은 17명의 평가자들이 3 곡에 대해 평가하여, 각 방법에 대해 총 51개의 5단계 평가값 분포와 함께 평가값의 평균과 표준편차를 제시하였다. 표 1을 참조하면, MP3 오디오(1. MP3로 표시)가 주관적 평가값이 약 3.92로 가장 낮은 값을 보였다. 저주파 복사를 이용하여 고주파 대역을 복원한 오디오(2. dec1으로 표시)는 MP3 오디오에 비해 약간 향상된 음질을 보여주고 있다. 반면에 AWGN을 이용하여 고주파 대역을 복원한 오디오(3. dec2로 표시)는 주관적 평가값이 4.27로 크게 향상된 음질을 보여주고 있다. 특히 평가값 2 (poor) 이하인 평가자가 없는 것으로 보아 항상 기준 이상의 음질을 보장할 수 있음을 알 수 있다. dec1은 MP3에 비해 평가값 5 (Excellent)로 평가한 사람이 감소한 것으로 나타났는데, 이는 저주파 복사를 이용하여 고주파를 복원할 때 평가자에 따라 오히려 주관적으로 느끼는 음질이 낮아질 수 있음을 보여준다. 그러나 dec1에 대해 평가값 3 혹은 2로 평가한 사람이 MP3에 비해 줄어들었다는 것은 주관적 음질을 전반적으로 향상시켜주는 특성이 있음을 알 수 있다. 한편, 표 1의 wav의 평가결과가 4.20인데

반해, dec2의 결과가 wav의 결과보다 오히려 높은 4.27로 나타났다. 이는 평가자들이 MP3 오디오에 익숙해져 있어 일부 평가자들은 MP3 오디오의 음질이 원본 CD 오디오보다 우수하다고 느끼기 때문인 것으로 생각된다. 또한 dec2에서 MP3 오디오의 단점을 고주파 복원을 통해 해소함에 따라 평가 결과가 향상된다고 판단할 수 있다. 그러나 wav와 dec2의 평균과 함께 표준편차를 참조하면, 95% 신뢰구간이 각각 [4.01, 4.38]과 [4.08, 4.47]로 측정되었다. 따라서 두 방법에 대한 신뢰 구간이 상당히 겹치게 되므로, 두 방법이 거의 동일한 음질을 제공하는 것으로 간주할 수 있다.

이상의 결과를 종합하면, MP3 오디오의 주관적 음질은 고주파 대역 복원을 통해 향상된다고 할 수 있다. 특히, AWGN을 이용하여 고주파 대역을 복원한 오디오는 원본 CD 오디오와 동등한 주관적 음질을 보인다고 할 수 있다.

## V. 결론

본 논문에서는 Liu의 고주파 복원 알고리즘에 기초하여 MP3 오디오의 형식을 그대로 유지하면서 고주파 대역을 복원할 수 있는 새로운 음질 향상 방법을 제안하였다.

제안한 알고리즘은 기존의 기법과 비교하여 다음과 같은 차이점이 있다. 먼저, MP3 부호화 S/W에 따라 차단 주파수가 다를 수 있다는 것을 고려하여, 입력된 오디오의 차단 주파수를 검출하여 추정된 차단 주파수를 복원 시작 대역으로 결정하였다. 그 결과, 부호화된 오디오 정보를 충분히 이용하면서, 제거된 고주파 대역을 효과적으로 복원할 수 있었다. 다음으로 32개 대역 신호의 에너지에 기초하여 주파수 포락선을 추정함으로써 고주파 대역의 포락선 복원이 보다 견실하고 정확해졌다. 고주파 대역 생성을 위해 저주파 성분과 AWGN 신호를 사용해본 고주파 대역 생성에 이용되는 성분의 영향도 고찰해보았다. 그 결과 AWGN을 이용한 고주파 대역 복원은 주관적 음질 평가 결과가 원본 CD 오디오와 동등한 것으로 나타나, MP3 오디오의 음질을 크게 개선시킬 수 있음을 확인하였다.

본 논문에서 제안한 알고리즘이 원본 CD 오디오의 음질을 완벽하게 복원하지는 못하지만, 고주파 대역의 복원을

표 1. 주관적 음질 평가 결과

Table. 1. Result of the subjective quality assessment

형식	5	4	3	2	1	평균	표준 편차
	excellent	good	fair	poor	bad		
1. MP3	12	25	12	2	0	3.92	0.80
2. dec1	8	36	6	1	0	4.00	0.60
3. dec2	21	23	7	0	0	4.27	0.70
4. wav	17	27	7	0	0	4.20	0.66

통해 보다 공간감 있고, 풍부한 오디오로 재생될 수 있음을 보였다. 제안한 알고리즘에 고조파 성분을 효과적으로 복원하는 방법, 차폐된 신호의 복원 등을 고려한다면 한층 더 원본 CD 오디오에 근접한 음질로 재생될 수 있을 것으로 기대된다.

참 고 문 헌

[1] ISO/IEC 11172-3, Information technology coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s Part 3: Audio, 1993.  
 [2] C. M. Liu, W. C. Lee, and H. W. Hsu, "High frequency re-

construction for band-limited audio signals", Proc. of 6th Int'l Conf. on Digital Audio Effects, pp. 1-6, London, UK, Sep. 2003.  
 [3] E. R. Larsen, R. M. Aarts, Audio Bandwidth Extension, John Wiley & Sons, Ltd, 2007.  
 [4] M. Dietz, L. Liljeryd, K. Kjolring, and O. Kunz, "Spectral band replication, a novel approach in audio coding" 112th AES Convention, Munich, Germany, May 2002.  
 [5] 윤원중, MP3 오디오 음악의 음질 향상에 관한 연구, 단국대학교 컴퓨터 과학 및 통계학과, 석사학위논문, Feb 2005.  
 [6] S.S Rao, Applied Numerical Methods For Engineers And Scientists, Prentice Hall, 2001.  
 [7] J. Herre, D. Schulz, "Extending the MPEG-4 AAC codec by perceptual noise substitution," 104th AES Convention, Amsterdam, Preprint 4720, May 1998.

저 자 소 개



허 소 영

- 2006년 : 서울시립대학교 전자전기컴퓨터공학부 학사 졸업
- 2008년 : 서울시립대학교 전자전기컴퓨터공학부 석사 졸업
- 주관심분야 : 신호처리, 오디오 코덱



김 인 철

- 1985년 : 서울대학교 제어계측공학과 학사 졸업
- 1987년 : 서울대학교 제어계측공학과 석사 졸업
- 1992년 : 서울대학교 제어계측공학과 박사 졸업
- 1992년 ~ 1994년 : 대우전자(주) 영상연구소 선임연구원
- 1994년 ~ 1999년 : 한성대학교 정보공학과 조교수
- 1999년 ~ 현재 : 서울시립대학교 전자전기컴퓨터공학과 교수
- 주관심분야 : 신호처리, 음향처리, 실시간 신호처리, 오디오/비디오 코덱