

MP3의 음질 향상을 위한 고주파 복원

허소영, 김인철

서울시립대학교 전자전기컴퓨터공학부

E-mail : rin@uos.ac.kr

High Frequency Reconstruction For MP3 Quality Improvement

So-Young Heo, Rin-Chul Kim

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

요약

본 논문에서는 일반적인 MP3 오디오에서 손실된 고주파 성분을 복원하여 향상된 품질의 오디오를 제공하는 방법을 고찰한다. 고주파 복원을 위해 먼저 MP3 오디오에 따라 손실된 고주파 영역이 차이남을 참조하여 고주파 복원 시작 대역을 결정한다. 그런 다음, IMDCT 이후 32개 대역으로 분할된 대역 신호의 에너지를 이용하여 고주파 대역의 에너지 포락선을 추정한다. 추정된 에너지 포락선을 참조하여 저주파 대역 신호를 복사하거나 AWGN 신호를 삽입함으로써 손실된 고주파 신호를 복원한다. 본 논문에서는 주관적 음질 평가 및 스펙트럼 분석을 통해 제안한 방법의 성능을 평가한다.

1. 서론

MPEG-1 오디오 LayerIII[1] 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 작다. 일반적으로 MP3는 매우 높은 압축률을 성취하되 귀에 거슬리지 않는 음질을 제공하기 위해, 약 14~16KHz 이상의 고주파 성분을 부호화하지 않는다. 이러한 MP3 오디오의 손실된 고주파 성분을 효과적으로 복원해 음질 향상을 얻고자 하는 연구가 많이 진행되었다[2-7].

본 논문에서는 MP3 오디오 형식을 그대로 유지하면서 고주파 성분을 복원하여 MP3 오디오의 음질을 향상시키는 방법으로 Liu가 제시한 고주파 복원 알고리즘[6]을 기초로 해서 MP3 오디오의 음질 향상을 얻을 수 있는 개선 방법을 고찰한다. 그리고 주관적 음질 평가를 통해 그 성능을 평가한다.

2. MP3의 고주파수 성분 복원 방법

일반적으로 MP3는 사람 귀에 잘 인식되지 않는 14~16KHz 이상의 고주파 성분을 부호화하지 않음으로써 사람 귀에 친숙한 낮은 주파수 성분을 보다 충실히 표현한다. 14~16KHz 이상의 고주파 성분이 제거되면, 비교적 낮은 주파수로 나타내어지는 보컬 (vocal) 성분은 증강된 크기로 재생되어 보컬 성분의 명료성 (clarity)이 약간 향상된다. 그렇지만 공간감 (spatiality)이 줄어들고 음향이 건조 (dull)해진다는 단점이 발생한다. 만약 MP3 형식의 음악에서 제거된 고주파 성분을 복원할 수 있다면 원래의 CD 음악과 유사하게 공간감이 있고 풍부한 (rich) 음악으로 재생할 수 있을 것이다.

현재까지 소개된 오디오 고주파 성분 복원 방법들은 표현 형식에

부가 정보를 추가하여야 하는 지 여부에 따라 non-blind 방법과 blind 방법으로 구별할 수 있다. Non-blind 방법의 대표적 예로서는 Coding Technology사의 SBR (spectral band replication) [2],[3]과 VLSI사의 PLUS-V 알고리즘[4]이 있다. Non-blind 방법은 비교적 충실히 고주파 성분을 복원할 수 있지만, 복원하고자 하는 고주파 성분에 대한 정보를 추가적으로 표현하기 위해 별도의 오디오 표현 형식이 필요하다.

그에 비해 blind 방법의 고주파수 복원 알고리즘의 대표적인 예는 wavelet 분석/합성 방식[5]과 Liu가 제시한 고주파 복원 알고리즘[6]이 있다. Blind 방법은 기존의 오디오 표현 형식을 그대로 사용한다는 장점이 있지만, non-blind 방법에 비해 오디오의 음질이 떨어진다는 단점이 있다.

본 논문에서는 blind 방법 중 하나인 Liu의 알고리즘을 기본으로 했다. Liu 알고리즘에서는 IMDCT 도메인에서 고주파 영역의 주파수 포락선을 추정하고, 추정된 주파수 포락선에 맞추어 부호화되지 않은 고주파 영역에 저주파 성분을 복사하여 고주파 성분을 복원한다.

본 논문에서는 Liu 알고리즘을 개선하기 위해 복원이 시작되는 주파수 영역을 추정해 고주파 대역을 복원한다. 또한, IMDCT 변환 이후의 32개 대역 신호의 에너지 분포를 이용하여 주파수 포락선을 추정하며, 고주파 복원에 저주파 복사뿐만 아니라 AWGN 을 이용한 방법도 고려하였다.

2.1 복원 시작 대역 결정

MP3로 압축된 오디오의 정확한 차단 주파수는 압축할 때 사용한 MP3 부호화 S/W에 따라 달라진다. MP3의 고주파 성분의 복원 과정

본 논문은 부분적으로 (주)문도시스템 의 연구비 지원으로 이루어졌습니다.

에서 만약 차단 주파수와 복원을 시작하는 주파수가 일치하지 않다면, 전송된 주파수 성분의 일부가 손상되거나, 복원 되어야 할 주파수 성분 중 일부가 복원되지 않는 현상이 발생한다. 그리고 차단 주파수가 항상 일정하다고 하더라도 부호화 과정에서 차단 주파수 근방의 주파수 성분이 차폐되어 부호화되지 않을 수도 있다. 이런 경우 복원 주파수를 고정시켜놓는다면 복원 주파수 이하의 주파수 성분으로 구성되는 복원 단위가 0에 가까운 값을 가질 수 있어, 고주파 복원 성능이 현저히 저하된다.

본 논문에서는 이러한 점을 고려하여 고주파수 복원 이전에 차단 주파수를 추정을 통해 가변적으로 설정하도록 했다. 먼저 차단 주파수 근방의 일정한 주파수 범위를 설정하고, 그 주파수 범위 내에서 대역별 에너지를 비교한다. 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 주파수 전후의 에너지 차이는 매우 크다는 점을 참조하여 대역 간 에너지차가 10dB 이상일 때 복원 시작 대역으로 설정하였다. 또한, 복원 시작 대역을 탐색하는 주파수 범위는 13.0~16.5KHz로 설정하여 14~16KHz 범위의 차단 주파수 검색 및 차폐에 의해 손실된 복원 단위에 의한 영향을 최소화할 수 있도록 하였다.

2.2 선형 예측을 이용한 포락선 추정

본 논문에서는 Liu 알고리즘과 마찬가지로 별다른 변환이 필요하지 않도록 부호화 과정 내 주파수 도메인의 신호를 이용한다. 다만 본 논문에서는 32개 대역으로 분해된 대역 신호를 이용한다. 로그 크기 스펙트럼이 개별적인 값으로 나타나는데 반해 대역 신호들은 대역당 18개의 샘플을 갖는다. 그리고 로그 스펙트럼에 비해 에너지 스펙트럼이 통계학적으로 견실한 특성을 보이므로 대역별 에너지를 이용하여 선형 모델 계수를 추정한다. 또한, 로그 스펙트럼은 주파수에 대해 지수 함수적으로 감소하므로, 에너지의 로그 값은 주파수 영역에서 선형적으로 감소하는 함수로 나타내어져 계산상 더 편리하다.

본 논문에서는 복원 구간과 선형 예측 구간의 연관성을 위해 10KHz 이상인 대역 신호들을 이용하는데, 10KHz를 포함하는 14번째 대역부터 앞 절에서 구한 복원 시작 대역 직전 $n_c - 1$ 번째 대역까지의 구간을 선형 예측 구간으로 이용하였다. 이 구간 내 n 번째 대역에 대한 에너지의 로그값 $\log_{10}|E_n|$ 을 얻는다. 그런 다음, 다음과 같이 $\log_{10}|E_n|$ 에 대한 선형 모델을 설정하고,

$$\log_{10}|E_n| = a_0 + a_1 n \quad (1)$$

그 계수 a_0 와 a_1 을 다음과 같이 최소 자승 접근법을 이용해 얻는다.

$$a_1 = \frac{\frac{1}{N} \sum_n n \cdot \log_{10}|E_n| - \left\{ \frac{1}{N} \sum_n n \right\} \left\{ \frac{1}{N} \sum_n \log_{10}|E_n| \right\}}{\frac{1}{N} \sum_n n^2 - \left\{ \frac{1}{N} \sum_n n \right\}^2} \quad (2)$$

$$a_0 = \left\{ \frac{1}{N} \sum_n \log_{10}|E_n| \right\} - a_1 \left\{ \frac{1}{N} \sum_n n \right\}$$

여기서 모든 합은 $14 \leq n \leq n_c - 1$ 번째 대역 범위에서 더해지고, $N = n_c - 14$ 이다. 일반적으로 주파수 영역에서 로그 스펙트럼

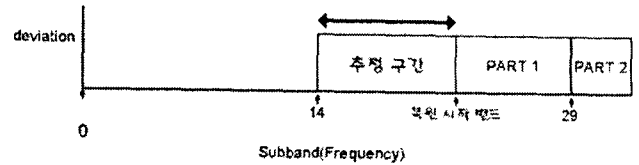


그림 1. 선형 모델을 적용하는 구간.

은 주파수가 높아짐에 따라 감소하므로, 대역 에너지 또한 대역이 높아짐에 따라 감소할 것이다. 만약 식 (2)를 통해 추정된 a_1 이 0보다 크다면 잘못 추정되었을 가능성이 높으므로, 이러한 경우 $a_1 = 0$ 으로 설정한다. 식 (2)와 같이 얻어진 계수들을 이용한 식 (1)을 통해 복원 시작 대역 n_c 이상 대역 신호들의 에너지 포락선을 추정한다.

2.3 고주파 대역의 복원

앞 절에서 구한 선형 모델 추정에 의한 고주파 대역 신호의 포락선을 이용해 고주파 성분을 복원하기 위해 본 논문에서는 복원 구간을 그림 1과 같이 두 구간으로 나누었다.

Part 1 구간은 n_c 번째 대역부터 28번째 대역까지로 구성되는데, 이 구간에서는 선형 모델로 추정된 포락선을 통해 대역별 에너지를 구하여 복원하고, part 2 구간(20KHz 이상 대역)에서는 선형 모델에 의해 얻어진 포락선 추정치를 Nyquist 주파수까지 추가적으로 감쇠가 일어나도록 설정하였다. 감쇠는 part 2 내의 대역 신호의 포락선이 단조 감소를 하되, Nyquist 주파수를 포함하는 마지막 대역에서 약 -30dB 만큼 추가적으로 감쇠가 일어나도록 설정하였다.

그리고 고주파 대역 신호에 상세 정보를 채워 넣는 방법으로 본 논문에서는 저주파 대역 신호를 복사하는 방법과 AWGN 신호를 삽입하는 방법 두 가지를 이용했다. 저주파 대역 신호를 복사하는 방법은 Liu 알고리즘과 유사하다. 복사에 이용할 저주파 대역 구간을 정한 후, 그 대역의 에너지를 선형 모델로 추정된 에너지가 되도록 대역 신호의 크기를 수정해 이용한다.

다음으로, 복원할 대역의 상세 정보로 AWGN 데이터를 삽입할 수 있다. 이를 위해 평균이 0이고, 분산이 1인 Gaussian 난수(random number)를 발생시킨다. 그런 다음, 이 난수들의 분산을 선형 모델로 추정된 에너지 포락선에 맞추어 n_c 이상의 대역에 삽입한다. 이 방법은 부호화 되지 않은 고주파 성분은 그 크기가 비교적 작고, 부호화하기에 복잡한 신호로 되어 있다는 가정과 일반적으로 양자화 잡음은 확률 함수로 나타내어질 수 있다는 가정에 부합되는 것이다.

그림 2는 본 논문에서 고주파 성분을 복원하는 지점을 MP3 복호기

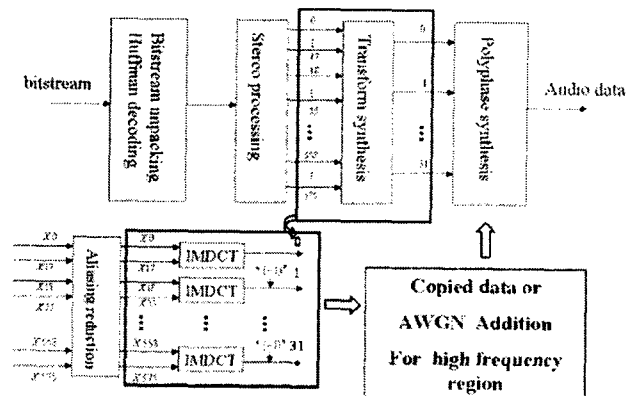


그림 2. MP3 복호기에서 고주파 복원 위치.

내에 도시한 것이다. 그림 2에서 보인 바와 같이 IMDCT가 수행된 다음, 대역 신호에 대해 고주파 복원이 수행된다.

3. 실험 및 결과

3.1 실험 방법

본 논문의 실험에 사용되는 오디오 파일은 다음의 과정을 통해 얻었다. 우선 CD에 수록된 오디오를 Adobe Audition 1.5를 이용해 표본화 주파수 44.1KHz, 샘플 당 16bits로 표현된 2 채널 스테레오의 원본 WAVE파일을 얻었다. 이 WAVE파일을 Adobe Audition 1.5를 이용해 표본화 주파수 44.1KHz, 전송률 128Kbps, 차단 주파수 15.804KHz의 MP3 파일을 얻었다. 그리고 본 논문에서 수행한 고주파 복원을 통한 음질 향상 알고리즘은 Window XP 환경의 PC에서 Microsoft Visual Studio 6.0과 C 언어를 이용해 버전 3.97의 Lame 복호기를 수정하여 구현하였다.

본 논문에서 고찰한 고주파 복원 기법의 성능을 평가하기 위해 주관적 음질 평가를 수행하였다. 본 논문에서는 음질 평가에 익숙하지 않은 초보자들을 대상으로 주관적 평가를 시행했다. 초보자의 경우 MP3 오디오와 원본 CD 오디오를 비교하면서 듣는 기회가 거의 없고, 대부분 가장 보편적인 오디오 형식인 MP3 오디오에 익숙해져 있다는 특징이 있다.

주관적 음질 평가의 순서:

1. MP3와 wav파일에 관한 간단한 특성을 알려준다. 그 후에 오디오 형식을 알려주지 않고 원본 CD 오디오와 MP3 오디오를 들려준다. 그리고 청취한 두 오디오를 구별하게 한다.
2. 오디오 형식을 알려주고, 두 오디오를 다시 들려줌으로써 MP3 오디오와 원본 CD 오디오의 특성을 유의해서 들을 수 있게 하고 두 오디오 형식의 차이를 알게 한다.
3. 오디오 형식을 알려주지 않고, 동일한 음악에 대해 4가지 다른 표현 형식의 오디오를 들려준다.
4. 서로 다른 4가지 표현 형식의 오디오 음질을 개별적으로 평가한다.

평가에 사용된 실험 오디오는 그룹 QUEEN의 "don't stop me now"(일부, 40초)과 이문세의 "조조할인"(3분 58초), WESTLIFE의 "swear it again"(4분07초) 등 3곡이다. 3개의 실험 오디오에 대해 위와 같은 방식으로 각각 서로 다른 4가지 형식의 오디오의 품질을 평가하였다. 각 실험 오디오는 후렴으로 들어가기 전부터 후렴부분을 포함하는 약 15~20초 구간을 선택해서 고주파 대역에 대한 느낌을 비교할 수 있도록 했다. 음질 평가는 표 1과 같이 5단계로 품질을 평가하였다. 그리고 평가자는 17명의 20대 학생들이다.

3.2 실험 결과

그림 3 ~ 7은 복원 시작 대역의 추정 효과의 효과를 보여주는 자료이다. 이 그림들은 첫 번째 실험곡 "don't stop me now"(queen)의 차단 주파수 15KHz의 MP3 오디오의 20초 일 때의 프레임에 대한 복원 결과를 각 처리 방법에 따라 보여준다.

복원 주파수가 16KHz로 설정된 Liu 알고리즘을 이용할 경우 차단

표 1. 주관적 음질 평가 기준

Excellent	음질이 매우 자연스럽다. (5점)
Good	음질이 약간 떨어지지만 그런대로 괜찮다. (4점)
Fair	음질이 많이 떨어지며 약간 거슬린다. (3점)
Poor	음질이 많이 왜곡되었지만 거북스러운 정도는 아니다. (2점)
bad	듣기 힘들 정도로 거북하게 들린다. (1점)

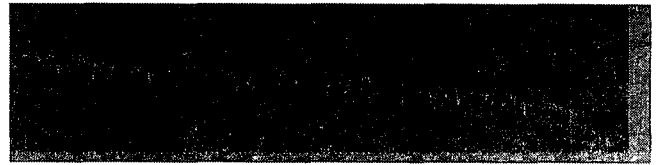


그림 3. 원본 wav 파일 (queen, 20초)

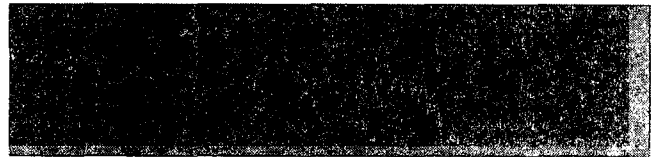


그림 4. 차단주파수 15KHz의 MP3 파일

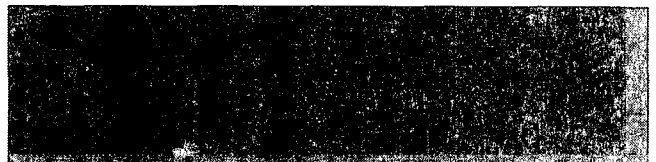


그림 5. Liu에 의한 고주파 대역 복원



그림 6. 저주파 복사에 의한 복원



그림 7. AWGN을 이용한 복원

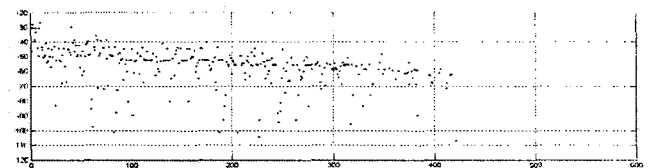


그림 8. IMDCT 변환 이전의 데이터 분포

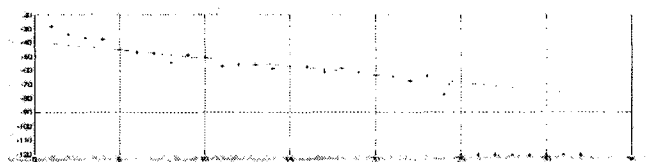


그림 9. IMDCT 이후 32 밴드 대역 에너지를 이용한 고주파 대역 추정

주파수가 그 이전인 15KHz의 MP3 파일을 복원할 때, 그림 5와 같은 문제가 발생한다. 복원 주파수 이전 대역이 복원되지 않은 문제점과 함께 고주파 복원을 위한 저주파 복사 구간에 15KHz에서 16KHz 사이의 공백이 포함됨으로써 복원된 고주파 영역에도 연속적으로 데이터 공백이 발생한다. 반면, 복원 시작점을 추정해 복원한 경우엔 저주파 복사와 AWGN을 이용한 복원 모든 경우에서 공백 없이 연속적으로 고

표 2. 주관적 음질 평가 요약.

형식	5	4	3	2	1	평균
	excellent	good	fair	poor	bad	
1. MP3	12	25	12	2	0	3.92
2. dec1	8	36	6	1	0	4.00
3. wave	17	27	7	0	0	4.20
4. dec2	21	23	7	0	0	4.27

주파 대역이 복원됨을 확인할 수 있다.

그림 8과 9를 통해 고주파 대역 복원을 위한 선형 추정을 위해서 IMDCT 변환 이후의 밴드 대역의 에너지를 이용하면, IMDCT 변환 이전의 분산된 값들을 이용하는 것보다 각 대역의 대푯값을 이용해서 조금 더 정확한 추정이 가능함을 알 수 있다.

그리고 AWGN을 이용해서 복원을 하면 고주파 대역 복원을 위해 복사에 이용할 저주파 대역 구간(단위 복사 구간)에 관한 제한을 신경 쓰지 않아도 된다. 단위 복사 구간 설정의 어려움으로는 먼저, 단위 복사 구간으로 이용할 적절한 영역을 구하는 것이 어렵고, 적절한 영역의 단위 복사 구간을 설정한다 하더라도, 모든 프레임에 대해 단위 복사 구간이 완벽하지 않음도 고려해야 한다. 앞에서 보인 예와 유사하게 단위 복사 구간에 데이터의 공백 영역이나 예상치 못한 피크가 포함되는 등의 경우 고주파 대역 복원에 영향을 미치게 된다.

다음으로 주관적 음질 평가에 대한 결과이다.

표 2를 보면 세 곡에 대한 평가의 결과를 조합하여, 5단계 평가값의 분포와 함께 평가값의 평균값을 제시하였다. (1. MP3)는 MP3로 부호화된 오디오, (2. dec1)은 저주파 복사를 이용하여 고주파 성분을 복사한 MP3 오디오를 나타낸다. 그리고 (3.wav)는 원본 CD 오디오, (4. dec2)는 AWGN을 이용하여 고주파 성분을 복사한 MP3 오디오를 나타낸다.

표 2를 참조하면, MP3 오디오가 가장 낮은 주관적 평가값을 보였다. 저주파 복사를 이용하여 고주파 성분을 복원한 오디오는 MP3에 비해 약간 향상된 성능을 보여주고 있는 반면, AWGN을 이용해 고주파 성분을 복원한 오디오는 비교적 크게 향상된 성능을 보여주고 있다. 특히 평가값 2(poor) 이하인 평가자가 없는 것으로 나타나 항상 기준 이상의 품질을 보장할 수 있음을 보였다. Dec1은 MP3에 비해 평가값 5(Excellent)로 평가한 사람이 감소한 것으로 나타났는데, 이는 저주파 복사를 이용한 고주파 복원이 평가자에 따라 오히려 주관적 품질이 감소될 수 있음을 보여준다. 그러나 평가값 3 혹은 2로 평가한 사람이 MP3에 비해 줄어들었다는 것은 주관적 품질을 전반적으로 향상시켜 주는 특성이 있음을 알 수 있다. 한편, 표 2에서는 wave의 평가값에 비해, dec2의 평가값은 wave의 평가값보다 오히려 높게 나타났다. 이는 평가자들이 MP3 오디오에 익숙해져 있어 일부 평가자들은 MP3 오디오의 품질이 원본 CD 오디오보다 우수하다고 판단했기 때문이다. 또한, dec2에서 고주파 복원을 통해 MP3의 단점을 해소함에 따라 평가값이 향상된 것으로 판단된다. 그러나, wave와 dec2의 평가값 차이는 오차 한계 이내로 볼 수 있으므로, 본 연구에서의 평가 자료에 의하면, 두 방법이 거의 동일한 음질을 제공하는 것으로 간주할 수 있다.

4. 결론

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

먼저, MP3 압축 S/W에 따라 차단 주파수가 다를 수 있다는 것을 고려하여, 입력된 오디오로부터 차단 주파수를 추정하여 추정된 차단 주파수를 복원시작 대역으로 결정하였다. 다음으로 선형 모델로 주파수 포락선을 추정할 때, IMDCT 변환 이후의 32개 대역으로 분리된 대역 신호의 에너지에 기초하여 주파수 포락선을 추정함으로써 보다 견실하고 정확한 주파수 포락선 추정이 가능해졌다. 마지막으로 복원 시작 대역부터 추정된 주파수 포락선에 맞게 고주파 신호를 채워 넣되, 저주파 성분을 고주파 영역에 복사하여 채워 넣는 방법과 함께 AWGN 신호를 채워 넣는 방법을 함께 고려했다.

그리고 본 논문에서는 3가지 실험 오디오를 이용하여 17명의 평가자를 대상으로 주관적 음질 평가를 시행한 결과, 저주파 복사를 이용하여 고주파 성분을 복원하는 방법과 AWGN을 이용하여 고주파 성분을 복원하는 방법 모두 MP3 오디오의 주관적 품질을 개선함을 알 수 있었다. 특히, AWGN을 이용하여 고주파 성분을 복원하는 방법은 주관적 음질 평가 결과가 원본 CD 오디오와 동등한 것으로 나타나, MP3 오디오의 음질을 크게 개선시킬 수 있음을 알 수 있었다.

본 논문을 통해 구현한 방법은 실질적으로 원본 CD 오디오의 품질을 완벽하게 복원하지는 못하지만, 고주파 성분의 복원을 통해 보다 공간감 있고, 풍부한 오디오를 재생시킬 수 있음을 보였다. 본 연구에서 고찰한 기법에 고조파 성분을 효과적으로 복원하는 방법, 스테레오 특성을 복원할 수 있는 방법, 차폐된 신호의 복원 등을 고려한다면, MP3 오디오를 한층 더 원본 CD 오디오에 근접한 음질로 재생할 수 있을 것으로 기대한다.

참고문헌

- [1] ISO/IEC 11172-3, *Information technology coding of moving pictures and associated audio for digital storage media at up to about 1.5Mbit/s Part3:Audio*, 1993.
- [2] Martin Dietz, Lars Liljeryd, Kristofer Kjolring, Oliver Kunz, "Spectral band replication, a novel approach in audio coding" at the *112th AES Convention*, Munich, Germany, May 10 - May 13, 2002.
- [3] Per Ekstrand, "Bandwidth extension of audio signals by spectral band replication", *Proc. of 1st IEEE Benelux Workshop in Model based Processing and Coding of Audio(MPCA-2002)*, pp.53-58, Leuven, Belgium, November 15, 2002.
- [4] "PLUS-V Specification: Public Document," VLSI Solution, Finland, Dec 17, 2001.
- [5] Corey, C., *Wavelet signal processing of digital audio with applications in electro-acoustic music*, Master's Thesis, Dartmouth College, Hanover, New Hampshire: 1996.
- [6] Chi-Min Liu, et. al., "High frequency reconstruction for band-limited audio signals", *Proc. of 6th Int'l Conf. on Digital Audio Effects*, pp. DAFX-1-DAFX-6, London, UK, Sep. 8-11, 2003.
- [7] 윤원중, "MP3 오디오 음악의 음질 향상에 관한 연구", 단국대학교 컴퓨터 과학 및 통계학과, 2004년.