

linbits 특성을 이용하여 MP3 비트스트림에 부가적인 정보를 삽입하는 방법에 관한 연구

Additive Data Insertion into MP3 Bitstream Using *linbits* Characteristics

김도형*, 양승진*, 정재호*
(Do-Hyoung Kim*, Seung-Jin Yang*, Jae-Ho Chung*)

*인하대학교 전자공학과 DSP 연구실

(접수일자: 2003년 6월 9일; 수정일자: 2003년 8월 20일; 채택일자: 2003년 9월 4일)

MP3의 사용이 늘어나면서 저작권 보호나 음악 콘텐츠 정보의 삽입을 위한 방법에 대한 요구가 증가하고 있으며 관련된 연구가 활발히 진행 중에 있다. 이 때 부가적인 정보가 삽입되어도 MP3의 비트스트림 구조가 크게 변형되어 음질의 열화나 파일 크기의 변화가 발생하여서는 안 된다. 본 연구에서는 위의 조건을 만족할 수 있도록 MP3 부호화 과정 중에서 양자화된 정수 형태의 계수가 허프만 부호화될 때, 값이 큰 샘플들의 특징을 이용하여 *linbits* 중 몇 개의 비트를 변형하여 부가적인 정보를 삽입하고, 그 음질의 열화 정도를 테스트하였다. 일련의 실험 과정을 통하여 평균 60 bytes/sec의 정보 삽입률에서 MOS 4.6 정도의 음질을 얻을 수 있음을 확인하였다. 제안된 방법을 이용하면 저작권에 대한 정보 또는 미디어 자체에 대한 정보들을 효과적으로 삽입할 수 있어 음악 데이터베이스의 운용과 같은 다양한 분야에 적용될 수 있다.

핵심용어: MP3, 디지털 워터마킹, 15-초과 양자화 계수, *linbits*, MOS 테스트

투고분야: 음향 신호처리 분야 (1,2)

As the use of MP3 audio compression increased, the demand for the insertion of additive data about copyright or information on music contents has been grown and the related research has been progressed actively. When an additive data is inserted into MP3 bitstream, it should not to happen any distortion of music quality or the change of file size, due to the modification of MP3 bitstream structure. In our study, to make these conditions satisfied, we inserted some additive data to bitstream by modifying some bits of *linbits* among the quantized integer coefficients having big values. At this time, we consider the characteristics of *linbits* and their distributions. As a result of subjective sound quality test through MOS test, we confirmed that the quality of MOS 4.6 can be achieved at the data insertion rate of 60 bytes/sec. Using the proposed method, it is possible to effectively insert an additive data such as copyright information or information about media itself, so that various applications like audio database management can be realized.

Keywords: MP3, Digital watermarking, Excess-15 quantized coefficients, *Linbits*, MOS test

ASK subject classification: Acoustic signal processing (1,2)

1. 서론

지난 수년 동안 인터넷의 보급과 압축 기술의 발달로 인하여 많은 이들이 손쉽게 원하는 정보들을 공유하고 취득할 수 있게 되었다. 여러 면에서 이러한 발전은 긍정적이고 바람직하지만, 인터넷에서 취득한 정보나 지식에 대한 재산권이나 소유권에 대한 분쟁과 갈등들이 사회적

인 문제로 대두되는 등의 부작용을 만들어내기도 하였다. 특히 MP3를 비롯한 고음질 오디오 신호 압축 기술이 발달하면서 대중들은 음질 상의 큰 차이를 느끼지 않으면서 값싸고 손쉽게 음악 콘텐츠들을 얻을 수 있게 되어 이러한 문제를 더욱 가중시키게 되었다. 근래에는 일방적인 단속과 제재만으로는 이러한 문제를 해결할 수 없다는 의견이 설득력을 얻으면서 오히려 인터넷을 통하여 압축된 형태의 음악 콘텐츠를 대중이 비교적 쉽고 저렴하게 취득할 수 있는 방안들이 마련되고 있다. 이러한 양성화 과정에서 많은 사람들이 관심을 두고 있는 내용 중

책임저자: 김도형 (dokim92@chol.com)
402-751 인천시 남구 용현동 253
인하대학교 전자공학과 DSP 연구실
(전화: 032-860-7420; 팩스: 032-863-3654)

하나의 음악 콘텐츠 자체에 제작자나 배포자에 대한 정보를 삽입하고 필요할 때 소유권을 주장할 수 있게 하는 방법에 관한 것이다.

최근에 디지털 워터마크 기술이 발전하면서 디지털 형태의 음악이나 영상에 다양한 응용을 위한 워터마크 또는 부가정보의 삽입에 대한 관심이 증가하고 있다. 이러한 워터마크 기술은 미디어를 보호하기 위하여 여러 가지 형태로 운영이 가능하다. 예를 들어 저작권 보호(copyright protection), 핑거프린팅(fingerprinting), 복사금지(copy protection), 방송 모니터링(broadcast monitoring), 데이터 인증(data authentication), 인덱싱(indexing) 등을 생각할 수 있는데[1,2], 그 중에서도 저작권 보호나 핑거프린팅 기술은 제작자가 필요에 따라 저작권 정보나 핑거프린트를 콘텐츠에 삽입하여 복사 경로를 추적하거나 소유권을 주장할 수 있는 근거를 마련하는 바람직한 기술이다. 이를 위해서는 음악 콘텐츠 파일에 필요한 정보를 부가하되 실제 사용자가 부가된 정보로 인하여 음질의 열화나 파일 크기의 변화 등을 인지할 수 없도록 하여야 한다.

근래에 개발된 몇 가지의 워터마킹 기술은 크게 시간영역에서 워터마크를 삽입하는 기술[3-5]과 주파수 영역[6] 또는 웨스트림 영역에서 워터마크를 삽입하는 기술[7]로 나눌 수 있다. 어느 경우에서나 장단점이 있을 수 있지만, 공통되는 요구 조건 몇 가지는 다음과 같이 요약된다.

- 인지불가능성: 삽입된 정보에 의하여 원래 음악 신호의 음질이 열화되어서는 안 된다.

- 강인성: 임의의 침입자에 의하여 제거되지 않아야 한다.
- 복원용이성: 미디어에 대한 권리를 가진 사람은 언제든지 필요에 의하여 삽입된 정보를 쉽고 빠르게 복원할 수 있어야 한다.

본 연구에서는 위의 조건들을 만족시킬 수 있는 새로운 정보 삽입 기법 또는 워터마킹 기법을 제안하고 그 성능을 평가하였다. 제안하는 알고리즘은 현재 활용도나 대중성에서 단연 독보적인 압축 방법인 MP3 (MPEG-1 Layer III)의 비트스트림에 부가적으로 정보를 삽입할 수 있게 한다. 이 방법은 양자화된 정수 형태의 계수의 변형을 통하여 정보를 삽입하기 때문에 비트스트림 영역에서의 정보 삽입이라고 할 수 있으며, 삽입된 정보를 비교적 쉽게 추출할 수 있기 때문에 저작권 보호나 핑거프린팅과 같은 응용에 사용될 수 있다.

2장에서는 기본 개념으로서 MP3의 부호화 과정 중 양자화와 허프만 부호화에 관련된 내용을 서술하고 3장에서는 MP3 부호화를 할 때 정보를 삽입하는 방법에 대하여 기술한다. 4장에서 정보가 부가된 MP3 비트스트림의 주관 음질 평가 및 데이터 삽입률에 관한 결과를 보일 것이고 끝으로 5장에서 결론을 맺겠다.

II. MP3 부호화와 linbits의 특징

그림 1에는 MP3 부호화의 대략적인 블록도를 보였다. 한 프레임당 1152개의 샘플을 갖는 입력 PCM 신호는 서

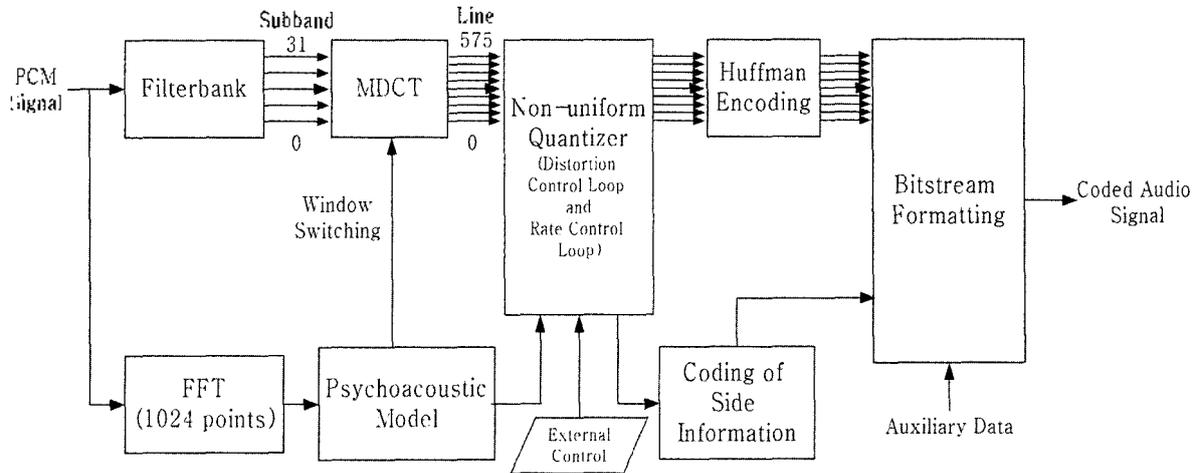


그림 1. MP3 부호화 과정을 보여주는 간단한 블록도
Fig. 1. Simple block diagram of MP3 encoding procedure.

브랜드 필터링과 MDCT (Modified Discrete Cosine Transform)를 거쳐 576개의 계수로 변환된다. 이때 같은 입력에 대하여 심리음향모델이 양자화기의 왜곡 제어 루프에서 사용할 수 있도록 각 대역에 대한 SMR (Signal-to-Masking Ratio)를 계산한다. 양자화기에서는 왜곡을 제어하기 위한 외부 순환 루프 (outer iteration loop)와 양자화에 사용되는 비트의 수를 조정하는 내부 순환 루프 (inner iteration loop)의 상호작용을 통하여 최종적인 양자화 수가 결정된다. 양자화 비트 수가 결정되면 정해진 규칙에 따라 부호화를 하게 되는데, 이 때 일반적인 MDCT 계수들에는 허프만 테이블을 이용한 부호화가 적용된다[8-10].

본 논문이 구체적인 MP3 알고리즘을 다루지는 않지만, 부가적인 정보를 삽입하는 단계인 허프만 부호화 및 비트 스트림 형성 과정의 이해를 위해 MP3의 양자화와 허프만 부호화, 그리고 *linbits*의 의미를 설명하도록 하겠다.

2.1. 양자화

양자화 과정을 간략하게 보여주는 개념도를 그림 2에 나타내었다. 실제 양자화는 레이트 제어 (rate control)를 하는 내부 순환 루프에서 수행된다. 내부 순환 루프는 왜곡을 제어 (distortion control)하는 외부 순환 루프에서 가장 먼저 수행되는 작업이다. MDCT를 통하여 변환된 계수는 부동소수점 형태의 값을 갖는데, 아래 식 (1)에 보인 비균일 양자화기를 거쳐 정수의 형태로 변환된다.

$$ix(i) = \text{nint} \left(\left(\frac{|xr(i)|}{\sqrt{2} \cdot \text{quant} + \text{quantif}} \right)^{0.75} - 0.0946 \right) \quad (1)$$

위에서 $xr(i)$ 는 576개의 MDCT 변환 계수 중에서 i 번째에 해당하는 값이고, $\text{quant} + \text{quantif}$ 는 양자화 스텝 사이즈 (quantization step size)를, $\text{nint}()$ 는 가장

가까운 정수 (nearest integer)를 말하고, 마지막으로 $ix(i)$ 는 i 번째 계수가 양자화된 정수 값을 의미한다. 이 값이 실제로 가질 수 있는 범위는 특별히 정해져 있지는 않지만 실제로는 수십 내지 수백 정도의 크지 않은 값에서 최대값을 가지게 된다.

양자화 스텝 사이즈는 내부 순환 루프에서 허용 가능한 비트의 수를 고려하여 결정된다. 허용 가능한 비트의 수를 고려한다는 것은 주어진 양자화 스텝 사이즈를 사용하였을 경우 허프만 부호화 등의 과정을 통해 총 몇 개의 비트를 사용하여야 주어진 그레놀 (granule)을 표현할 수 있는지를 계산하는 과정을 말한다.

현재의 양자화 스텝 사이즈를 사용하였을 경우 그레놀에서 사용할 수 있는 최대의 비트의 수를 넘지 않게 되면 외부 순환 루프가 동작된다. 외부 순환 루프에서는 내부 순환 루프에서 정해진 양자화 스텝 사이즈를 이용하여 양자화할 경우, 발생하는 왜곡을 계산하고 각 임계 대역 (critical band)을 증폭하여 왜곡을 제어한다.

2.2. 허프만 부호화와 *linbits*

양자화 과정의 결과로 얻어진 그레놀마다의 576개 정수 값들 중 뒷부분의 값들은 대부분 0의 값을 갖게 된다. 이것은 일반적인 오디오 신호가 고주파 영역에서 작은 에너지를 갖기 때문에 발생하는 현상인데, 0이 연속되는 부분은 연속되는 0의 개수만을 짝수로 기록하여 부호화한다. 연속된 0이 끝나는 지점에서부터 앞부분으로의 값 역시 비교적 작은 값을 갖게 되는데, -1, 0, 1의 값으로만 구성된 부분은 네 개씩 짝을 이루어 쿼드러플 (quadruple)을 위한 허프만 테이블을 이용하여 부호화한다. 이렇게 연속된 0의 짝들 (*rzero*)과 절대값이 1 이하인 쿼드러플들 (*count1*)을 제외한 나머지 영역은 두 개의 값을 짝지어서 32개의 허프만 코드 테이블을 이용하여 부호화하게 된다. *bigvalues*라고 불리는 이 영역은 다시 세 개의

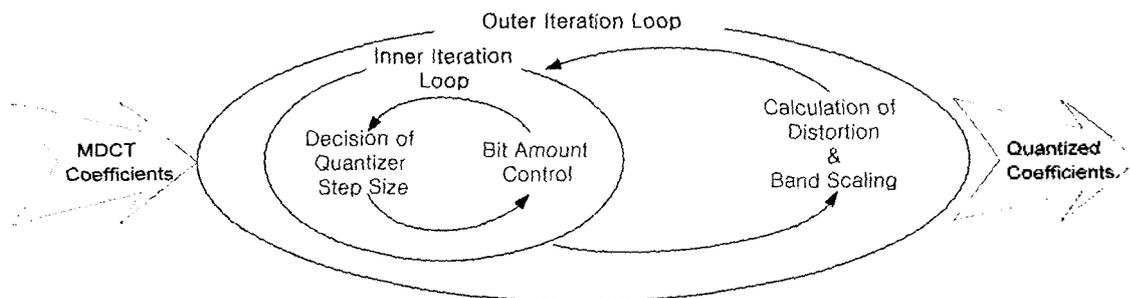


그림 2. 양자화 과정을 수행하는 순환 루프
Fig. 2. Iteration loop performing quantization.

subregion으로 분할되어 각각 서로 다른 허프만 테이블을 사용하게 된다. 어떤 허프만 테이블을 사용할 것인가는 각 subregion에 존재하는 최대값의 크기를 고려하여 결정된다.

허프만 테이블들이 부호화할 수 있는 최대값은 15이며, 주어진 영역에서의 최대값이 15보다 큰 경우 (15-초과 양자화 계수)에는 ESC (escaped) 모드가 동작한다. ESC 모드에서는 부호화하려는 값이 15보다 크면 그 값과 15의 차이는 선형적으로 이진 부호 (binary code)를 사용하여 부호화하고, 15는 해당 허프만 테이블에 있는 코드를 사용하여 부호화한다. 이 때 15를 초과하는 값을 선형적으로 부호화하기 위해 필요한 최대 비트의 수를 linbits라 부른다. 16번째 허프만 테이블에서부터 23번째 허프만 테이블까지는 테이블에 포함된 코드들은 모두 같지만, 15-초과 양자화 계수를 부호화하기 위하여 사용되는 linbits의 값이 각각 다르다. 24번째에서부터 31번째까지의 허프만 테이블도 같은 코드를 가지지만 linbits가 서로 다른 경우이다. 표 1에 32개의 허프만 코드 테이블의 특징을 보였다.

예를 들어 현재 부호화하려고 하는 subregion에서의 최대값이 25라 하자. 최대 25를 표현하기 위해서는 25와 15의 차이에 해당하는 10의 값을 선형 부호화할 수 있도록 부가적인 비트를 사용하여야 한다. 이 경우에는 linbits가 4비트이어야 한다. 해당 subregion에서 현재 부호화하려는 값이 21이라면 선택된 허프만 테이블 (테이블 번호 19 또는 24)에서 15의 값을 부호화하고, 이 값과의 차이값 (21-15=6)은 linbits를 이용하여 0110₍₂₎의 형태로

부호화되어 비트스트림에 삽입되게 된다.

III. 부가적인 정보의 삽입

3.1. ESC 모드에서의 정보의 삽입

위에서 언급한 것과 같이 MP3 부호화를 위한 허프만 코드 테이블은 양자화에 의하여 발생된 정수 형태의 값 중 최대 15의 값까지 표시할 수 있도록 설계되어 있다. 이 값보다 큰 값이 발생되면 ESC 모드가 동작하기 시작한다. ESC 모드에서는 테이블 번호 16번부터 31번까지의 ESC 테이블이 사용되는데 15의 값은 테이블에서 해당 코드를 찾아 부호화하고 주어진 값과 15의 차이에 해당하는 값은 비트 형태 그대로 비트스트림에 삽입된다.

이 때 linbits에 해당하는 비트들 중에서 LSB들을 변형하여 데이터를 삽입하여 보았다. linbits에 해당하는 모든 비트를 변형하는 것은 너무 큰 왜곡을 초래하므로 그 중 일부의 비트들을 변형하는 것에 대하여 고려하였으며, 그 경우 음질에 열화가 얼마나 발생하는지를 평가하기 위하여 MOS 테스트를 실시하였다.

ESC 모드가 동작되는 16번에서부터 31까지의 허프만 테이블에서의 최소 linbits는 1비트이고 최대 linbits는 13비트이다. 실험을 간단화하기 위하여 표 2와 같이 변형되는 비트의 수를 정하였다. 이 표에서 볼 수 있는 바와 같이 linbits가 1비트인 경우 1비트를 변형하였고 linbits가 2비트 이상이면 LSB 2비트만을 변형하였다.

표 1. 허프만 코드 테이블의 특징
Table 1. Properties of huffman code table.

| Table No. | Max Code | ESC On/Off | linbits | Table No. | Max Code | ESC On/Off | linbits |
|-----------|----------|------------|---------|-----------|----------|------------|---------|
| 0 | 0 | off | 0 | 19 | 15 | on | 4 |
| 1 | 1 | off | 0 | 20 | 15 | on | 6 |
| 2~3 | 2 | off | 0 | 21 | 15 | on | 8 |
| 4 | Not Used | off | 0 | 22 | 15 | on | 10 |
| 5~6 | 3 | off | 0 | 23 | 15 | on | 13 |
| 7~9 | 5 | off | 0 | 24 | 15 | on | 4 |
| 10~12 | 7 | off | 0 | 25 | 15 | on | 5 |
| 13 | 15 | off | 0 | 26 | 15 | on | 6 |
| 14 | Not Used | off | 0 | 27 | 15 | on | 7 |
| 15 | 15 | off | 0 | 28 | 15 | on | 8 |
| 16 | 15 | on | 1 | 29 | 15 | on | 9 |
| 17 | 15 | on | 2 | 30 | 15 | on | 11 |
| 18 | 15 | on | 3 | 31 | 15 | on | 13 |

표 2. *linbits* 중 변형되는 비트의 수
Table 2. Number of modified bits in *linbits*.

| Table No. | linbits | Number of Modified Bits |
|-----------|---------|-------------------------|
| 15 | 1 | 1 |
| 16~31 | 2~13 | 2 |

3.2. *bigvalues* 영역을 이용한 정보 삽입들의 운용

MP3 부호화되는 음악 신호의 특징에 따라 다르긴 하지만 ESC 모드는 경우에 따라 매우 빈번히 발생하게 된다. 만약 아무런 제약없이 위의 절에서 언급한 것처럼 일괄적으로 비트의 변형을 가할 경우 음질 상의 심각한 왜곡을 초래할 수밖에 없다. 이 경우 변형되는 비트의 발생 빈도를 타당하게 조절할 수 있어야 하며, 그로 인하여 데이터 삽입들의 감소를 감수하고서라도 음질의 열화를 방지할 수 있어야 한다.

본 연구에서는 어떤 특징의 음악에서 ESC 모드가 집중적으로 발생하는지를 조사하여 보았다. 이를 위해 사용된 샘플들과 그에 따른 15-초과 양자화 계수의 발생 빈도를 표 3에 나타내었다.

이 표에서 볼 수 있는 것과 같이 15를 넘는 값은 샘플의 종류에 따라 편차가 큰데 고주파 신호가 강한 신호일수록 빈도가 적고 저주파 영역에 집중된 에너지를 갖는 샘플의 경우 빈도가 높은 것임을 알 수 있었다. 이 두 가지 경우에 대한 극단적인 예들의 스펙트럼들을 그림 3과 그림 4에서 보여주고 있다. 그림 3은 위의 샘플 중에서 "andrea2"의 스펙트럼을, 그림 4는 "radio"의 스펙트럼을 보여준다.

15를 초과하는 발생 빈도가 높은 경우는 정해진 비트의 수를 이용하여 저주파 영역에 집중된 에너지를 표현할 때, 할당된 비트의 양이 저주파 대역에 집중되면서 발생하는 현상이며, 발생 빈도가 적은 경우는 역시 정해진 비

트를 고주파 대역에까지 나누어 사용하기 때문에 발생하는 현상이라고 할 수 있다. 따라서 에너지가 얼마나 분산되어 있는가는 *linbits*의 양을 결정하는 요소로 작용한다고 할 수 있다.

에너지의 분산, 즉 스펙트럼의 분포 범위를 나타낼 수 있는 값은 주파수 대역별 에너지의 비율 등 여러 형태가 사용 가능하지만, MP3 알고리즘에서 부가적인 계산없이 사용할 수 있는 지표값으로 *bigvalues*를 생각할 수 있다. 이 값은 576개의 양자화된 값 중에서 고주파 영역의 연속되는 0 (*rzero*)과 -1, 0, 1의 값이 연속되는 영역 (*count1*)을 제외하고 비교적 값이 큰 영역이 최대 어디까지이며, 어느 양자화 계수부터 실제 허프만 코드 테이블을 사용하여야 하는지를 알려 주는 값이다. 이해를 돕기 위해 그림 5에 576개의 양자화된 계수의 배열을 도식화하였다. 이 그림에서 총 576개의 양자화 계수들 중에서 *bigvalues*는 M개, *count1*은 N개, 그리고 *rzero*는 576-(M+N)개가 된다.

*bigvalues*의 값이 크다는 것은 유효한 양자화 계수의 값이 고주파 영역에 걸쳐 넓게 분포되어 있다는 것을 나타내며 반면 그 값이 작다는 것은 에너지가 저주파 영역에 집중되어 있다는 것을 나타낸다. 따라서 이 값을 살펴보면 주파수 대역에 따른 에너지 분포의 차이로 인하여 발생하는 오디오 신호별 15-초과 양자화 계수의 발생 편차를 추적할 수 있다.

이와 같은 사실을 바탕으로 하여 본 연구에서는 다음과 같이 *bigvalues*의 범위에 따라 정보 삽입을 위하여 사용되는 15-초과 양자화 계수의 변형에 제한을 두었다. *bigvalues*의 값에 따른 변형의 제한은 표 4와 같은 기준을 바탕으로 하였다.

위와 같이 제한의 기준을 정한 이유는 *bigvalues*가 작은 경우, 즉 에너지가 저주파 영역에 집중되어서 15-초과

표 3. 사용된 샘플 신호별 15-초과 양자화 계수의 발생 횟수
Table 3. Number of occurrence of excess-15 quantized values as samples.

| Sample Name | Length | Number of Occurrence | Occ./Sec. | Sample Name | 샘플 길이 | Number of Occurrence | Occ./Sec. |
|-------------|--------|----------------------|-----------|-------------|--------|----------------------|-----------|
| piano | 23.401 | 14871 | 635.486 | beetho2 | 13.572 | 3833 | 282.420 |
| oboe | 14.548 | 2666 | 183.255 | big | 11.439 | 652 | 56.998 |
| brass | 6.058 | 3590 | 592.605 | char11 | 11.584 | 4610 | 397.963 |
| pipeorgan | 6.149 | 3373 | 548.544 | john2 | 17.214 | 7319 | 425.177 |
| saxophone | 7.455 | 188 | 25.21 | kpop | 10.517 | 1439 | 136.826 |
| wash1 | 12.006 | 4396 | 366.150 | newday | 10.455 | 541 | 51.746 |
| andre1 | 14.391 | 8204 | 570.079 | radio | 11.833 | 196 | 16.564 |
| andrea2 | 27.221 | 26564 | 975.864 | secret2 | 16.440 | 6937 | 421.959 |
| andrea3 | 20.943 | 3562 | 170.081 | stan1 | 14.758 | 1951 | 132.199 |
| beetho1 | 10.792 | 2401 | 222.480 | chorus | 15.387 | 1538 | 99.955 |

적인 데이터를 삽입할 때 후보가 되는 15-초과 양자화 계수가 순차적으로 발생하지 않도록 여러 스케일팩터로 다시 분산을 시켜 음질의 향상을 꾀하였다. 예를 들어 *bigvalues*의 값이 250이라 가정하자 (표 4 참조). 15-초과 양자화 계수가 첫 번째에서부터 다섯 번째까지 연속적으로 발생하였다면, 그대로 5개의 계수값을 변형한다면 저주파 영역에 오차가 집중되어 왜곡을 초래할 것이다. 따라서 이러한 경우라도 각 스케일팩터에서의 15-초과 양자화 계수가 고르게 변형되도록 분산시켰다. 표 4에서 *bigvalues*가 400 이상인 경우에는 이 방법에 제한을 받지 않는다. 이렇게 분산을 시키면 정보 삽입으로 인한 음질의 왜곡 현상이 특정한 주파수 대역에 편중되지 않기 때문에 효과적이라고 할 수 있다.

IV. 음질 평가 테스트 및 결과

4.1. MOS 테스트를 이용한 음질 주관 평가

본 절에서는 제3장에서 설명한 MP3 비트스트림에 정보를 삽입하는 기법의 성능 중 음질에 관한 평가를 위하여 수행한 MOS 테스트에 관하여 설명한다.

4.1.1. 피험자

MOS 테스트를 위해 20대의 남녀 대학생이 참여하였다. 전체 테스트 인원은 남자 16명, 여자 4명으로 총 20명이었으며, 전문적인 훈련을 받지는 않았지만 음악을 즐

겨 듣는 사람들을 대상으로 하였다.

4.1.2. 사용된 샘플

사용된 샘플의 이름과 길이는 이미 표 3에서 보였다. 이 샘플들은 모두 CD에서 손실없이 디지털 형태로 채집한 것들이어서 모두 44.1 kHz의 샘플링 주파수와 16비트 해상도를 갖는 스테레오 신호들이다. 각 장르를 잘 대변할 수 있도록 가급적 다양한 음악들을 선택하였으며 임의의 장르별 샘플 종류는 표 5에서 볼 수 있다. 같은 음악에서도 서로 다른 분위기를 대변할 수 있는 부분들을 추출하였다.

4.1.3. MOS 테스트 방법

피험자에게 각 실험 샘플별로 두 가지의 자료를 제시하였다. 먼저 부가적인 정보를 삽입하지 않은 128 kbps MP3 음악 (A)을 들려주었고, 곧바로 본 논문에서 언급한 방법을 이용하여 부가 정보를 삽입한 128 kbps MP3 음악 (B)을 들려주었다. A 음악이 열화가 없다는 가정 하에 B 음악의 음질이 얼마나 열화되었는지를 다음 표 6과 같은 지표를 이용하여 평가하도록 하였다[11]. 이 때 피험자가 원할 경우 같은 음악을 반복적으로 들어볼 수 있도록 하였다.

일반 PC에 고급 D/A 변환 능력을 갖는 사운드 어댑터 (Sound Adapter)를 장착하여 상용 MP3 플레이어에서 재생하였고, 피험자는 고성능 모니터용 헤드폰을 이용하여 음질을 평가하도록 하였다.

음질이 가장 나쁠 수 있는 경우를 측정하기 위하여 삽

표 4. *bigvalues*에 따른 변형의 제한
Table 4. Constraint of modification as *bigvalues*.

| Range of <i>bigvalues</i> | Max Number of Modified Coefficient in <i>subregion</i> |
|------------------------------|--|
| <i>bigvalues</i> < 200 | 3 |
| 200 ≤ <i>bigvalues</i> < 300 | 5 |
| 300 ≤ <i>bigvalues</i> < 400 | 7 |
| 400 ≤ <i>bigvalues</i> | No Limit |

표 5. 실험에 사용된 샘플들의 장르별 분류
Table 5. Genres of music samples used in experiments.

| Genre | Sample Name |
|--|--|
| Classic, Semi-Classic, Classic Instrumentals | beetho1, beetho2, andre1, secret2 |
| Ballad Mood Pop | kpop, big |
| Rock Mood Pop | newday, radio, |
| Jazz | char1f, john2, stan1, wash1 |
| Solo Instrumentals | piano, brass, saxophone, pipeorgan, oboe |
| Voice, Chorus | andrea2, andrea3, chorus |

입하는 정보의 형태는 다음 표 7에서 보인 것과 같이 원래의 값과 거리가 가장 멀어지도록 변형하였다.

표 6. MOS 테스트를 위한 평가 척도
Table 6. Quality evaluation index for MOS test.

| Score | Index |
|-------|--------------------------------------|
| 5 | No audible distortion |
| 4 | Audible distortion, but not annoying |
| 3 | Slightly annoying |
| 2 | Annoying |
| 1 | Very annoying |

표 7. 최대 음질 열화를 만들기 위한 비트 변형의 조건
Table 7. Bit modification condition for making the worst sound quality.

| linbits | Original LSB Code | Modified LSB Code |
|---------|-------------------|-------------------|
| 1 bit | 0 | 1 |
| | 1 | 0 |
| | 00 | 11 |
| 2 bits | 01 | 11 |
| | 10 | 00 |
| | 11 | 00 |

표 8. 실험 결과: 샘플별 MOS 값과 정보 삽입률
Table 8. Experiment result: MOS value and data insertion rate.

| Sample Name | MOS | Variance | Data Insertion Rate |
|-------------|--------|----------|---------------------|
| pipeorgan | 4.85 | 0.134211 | 131 bytes/sec |
| piano | 4.45 | 0.260526 | 114 bytes/sec |
| brass | 4.45 | 0.365789 | 106 bytes/sec |
| oboe | 4.5 | 0.368421 | 44 bytes/sec |
| saxophone | 4.15 | 0.660526 | 5 bytes/sec |
| andre1 | 4.7 | 0.221053 | 126 bytes/sec |
| beetho1 | 4.7 | 0.221053 | 45 bytes/sec |
| stan1 | 4.5 | 0.368421 | 28 bytes/sec |
| radio | 4.15 | 0.660526 | 3 bytes/sec |
| secret2 | 4.6 | 0.252632 | 86 bytes/sec |
| beetho2 | 4.6 | 0.252632 | 58 bytes/sec |
| john2 | 4.7 | 0.326316 | 96 bytes/sec |
| andrea2 | 4.75 | 0.197368 | 126 bytes/sec |
| newday | 4.5 | 0.578947 | 12 bytes/sec |
| charl1 | 4.6 | 0.252632 | 97 bytes/sec |
| andrea3 | 4.8 | 0.168421 | 36 bytes/sec |
| big | 4.8 | 0.168421 | 13 bytes/sec |
| wash1 | 4.75 | 0.197368 | 72 bytes/sec |
| kpop | 4.55 | 0.260526 | 32 bytes/sec |
| chorus | 4.65 | 0.344737 | 20 bytes/sec |
| Average | 4.5875 | 0.313026 | 63 bytes/sec |

4.2. MOS 결과 및 각 샘플별 정보 삽입률

다음 표 8에 각 샘플에 대하여 20명의 피험자에 대한 MOS 테스트의 평균값과 분산, 그리고 초당 정보의 삽입률을 보였다. MOS 테스트의 평균값은 약 4.6으로 나타났으며, 초당 삽입할 수 있는 정보의 양은 약 63 바이트 (500여 비트) 정도로 나타났다. 초당 60여 바이트라면 1초에 약 60개의 ASCII, 또는 약 30개의 조합형 한글 문자를 삽입할 수 있다는 것이다. 이 정보량은 음악의 저작권 등을 위한 정보 외에도 가수명, 음악의 제목 등의 음악 콘텐츠 자체에 대한 정보와 대중 음악이라면 가사까지도 실을 수 있는 양이라고 할 수 있다.

그런데 표 8을 살펴보면 “saxophone”과 “radio” 샘플의 평균이 매우 작게 나온 사실을 살펴볼 수 있다. 이것은 피험자 중 두 명이 두 샘플에 대하여 각각 MOS 2점 정도의 매우 낮은 점수를 부여하였기 때문이다. 두 샘플의 분산이 다른 것들에 비하여 매우 높게 나온 것이 이를 설명한다. “newday” 샘플의 경우에도 위 두 명 중 한 명의 피험자가 MOS 3점의 매우 낮은 점수를 부과하여 평균이 떨어졌으며 역시 분산의 값이 비교적 크다. 이러한 현상은 문제의 샘플들에 실제 삽입된 정보의 양이 최대 십여 바이트로 다른 샘플보다 비트스트림의 변형이 매우 적었

음에도 발생한 것이어서 쉽게 이해가 되지 않는 결과로 여겨졌다. 문제의 피험자에게 이와 같은 현상에 대하여 언급한 결과, 피험자들이 그 음악들에 대한 선입견을 가지고 있었음을 확인하였다. 한 명의 피험자는 “radio”나 “newday” 샘플에 대하여 매우 익숙해 있어서 평소에 들었던 환경과의 차이로 인한 음질상의 선입견을 가지고 있었으며, 또 다른 한 명의 피험자는 고주파 성분이나 노이즈가 강한 락 분위기의 음악에 지나치게 거부감을 가지고 있었음을 알 수 있었다. “saxophone” 샘플은 테너 색소폰이 매우 다이나믹하게 연주된 샘플이어서 매우 조용하고 평범한 음악에만 노출되었던 사람들이 쉽게 거부감을 가질 수 있는 샘플이다. 따라서 음질의 열화 정도를 측정하지 못하고 오히려 선호도에 따른 평가를 하었다고 생각할 수 있다.

위의 표에서 또하나의 문제로 지적될 수 있는 사항은 정보 삽입률이 샘플에 따라 큰 편차를 갖는다는 것이다. 그러나 이것은 다양한 음악들의 특징을 잘 표현하는 일부 구간에서 채집된 샘플에 관한 데이터이기 때문에 전체 음악을 대상으로 할 경우에는 정보 삽입률이 평균화되어 지나치게 적은 정보 삽입이 문제가 되지는 않는다고 할 수 있다. 일례로 위의 표에서 정보의 삽입률이 가장 적은 “radio” 샘플 (3 bytes/sec)과 삽입률이 높은 그룹 중 하나인 “andrei” 샘플 (126 bytes/sec)에 대하여 전곡을 모두 압축하면서 정보의 삽입률을 살펴보니 각각 40 bytes/sec와 88 bytes/sec로 나타났다. 즉, 정보 삽입률에 대한 편차 문제는 음악의 전체 구간에 대하여 고려할 때는 실제로 심각하지 않은 것이라고 할 수 있다.

V. 결론

본 연구에서는 MP3 부호화 과정 중 양자화를 거쳐 발생된 정수 형태의 계수 중에서 그 값이 15를 초과하는 계수들이 비트스트림에 삽입될 때, *linbits* 중 LSB를 변형하여 부가적인 정보를 삽입하고 그 성능을 평가하였다. *bigvalues*의 값을 이용한 입력 신호의 주파수 대역 분포를 감안하고 *linbits*의 특성을 고려하여 변형되는 계수가 특정 대역으로 집중하는 현상을 막아 음질의 왜곡을 최소화하였다. 일련의 실험을 통하여 제안된 기법을 이용하여 부가적인 정보를 삽입할 경우 초당 63 바이트의 정보를 삽입할 수 있음을 확인하였고, 이 경우의 음질은 20명의 비전문가를 대상으로 한 MOS 테스트 결과 평균

4.6 정도인 것을 알 수 있었다. 이 방법을 이용하면 MP3 비트스트림에 음악 콘텐츠에 대한 다양한 정보를 삽입할 수 있어 저작권 보호 및 부가정보 서비스에 대한 다양한 응용 분야에 사용될 수 있을 것이라 생각된다.

본 연구는 현재 고정된 정보 삽입률을 보장할 수 없는 문제를 해결하고 보다 안전한 저작권 보호 기법을 위한 정보 삽입의 강인성에 초점을 맞추어 계속 진행할 것이다. 또한 전용 부호화기와 전용 플레이어의 개발을 통하여 실제 현장에서 쉽게 사용될 수 있도록 개발할 것이다.

참고 문헌

1. G. C. Lanelaar, J. Setyawan and L. Legendijk, “Water-marking digital image and video data,” *IEEE Signal Processing Magazine*, 17 (5), 20-46, 9, 2000.
2. M. Arnold, “Audio watermarking: features, applications and algorithms,” *Proc. of IEEE International Conference on Multimedia and Expo*, 2, 1013-1016, 2000.
3. P. Bassia and I. Pitas, “Robust audio watermarking in the time domain,” *Proc. of Ninth European Signal Processing conference*, 25-28, 1998.
4. M. F. Mansour and A. H. Tewlik, “Audio watermarking by time-scale modification,” *Proc. of IEEE International Conference on Acoustics, Speech, and Signal Processing*, 3, 1353-1356, 2001.
5. L. Boney, A. H. Tewlik, and K. N. Hamdy, “Digital watermarks for audio signals,” *Proceedings of the Third IEEE International Conference on Multimedia Computing and Systems*, 3, 473-480, 1996.
6. J. Herre, C. Neubauer, F. Siebenhaar, and R. Kulesa, “New results on combined audio compression/watermarking,” *Proc. of AES 111th Convention*, 2000.
7. S.-K. Lee and Y.-S. Ho, “Digital audio watermarking in the cepstrum domain,” *IEEE Trans. on Consumer Electronics*, 46 (3), 744-750, 8, 2000.
8. ISO/IEC 11172-3, “Information Technology - Coding of moving pictures and associated audio for digital storage media at up to 1.5 Mb/s, Part 3,” Audio, 1993.
9. K. C. Pohlmann, *Principles of Digital Audio*, 4th Ed., McGraw-Hill, 2000.
10. S. Shlien, “Guide to MPEG-1 Audio Standard,” *IEEE Trans. on Broadcasting*, 40 (4), Dec., 1994.
11. 강성훈, 방송음향, 기전연구사, 61-65, 1997.

저자 약력

● 김도형 (Do-Hyoung Kim)

1996년 2월: 인하대학교 전자공학과 (공학사)
 1998년 8월: 인하대학교 전자공학과 (공학석사)
 1998년 9월~현재: 인하대학교 전자공학과 박사과정 재학중
 1996년 5월~1997년 12월: 한국전자통신연구원 실감음향통신연구실 위촉연구원
 * 주관심분야: 오디오 신호처리, 오디오 코딩, 심리음향, 오디오 모델링, 음악음향

● 양 승 진 (Seung-Jin Yang)

2002년 2월: 인하대학교 전자공학과 (공학사)
2002년 3월~현재: 인하대학교 전자공학과 석사과정
※ 주관심분야: 디지털 워터마킹, 오디오 코딩

● 정 재 호 (Jae-Ho Chung)

1982년: University of Maryland (BSEE)
1984년: University of Maryland (MSEE)
1990년: Georgia Institute of Technology (Ph. D.)
1984년~1985년: 미 국방성 산하 해군 연구소, 신호처리실, 연구원
1991년~1992년: AT&T Bell Laboratories, 음성신호처리 연구실, 연구원 (MTS)
1992년~현재: 인하대학교 공과대학 전자공학과 교수