

지각적 특성을 이용한 저 비트오율 압축 오디오 음질개선

차혁근, 채병국, 차형태
송실대학교 전자공학과

Audio Quality Enhancement using Perceptual Property at a Low-bitrate Compression

Hyuk-Geun Cha, Byoung-Koog Chae, Hyung-Tai Cha
Dept. of Electronic Engineering, Soongsil University

hgcha@mms.ssu.ac.kr, ekstns@mms.ssu.ac.kr, hcha@ssu.ac.kr

요약

본 논문에서는 저 비트오율 압축 시 발생하는 신호 왜곡을 인간의 지각적 특성을 이용하여 음질을 개선하는 알고리즘을 제안한다. 저 비트오율 압축 과정에서 손실된 고주파 영역의 신호를 부가 정보를 사용하지 않고 손실 되지 않은 영역의 정보를 사용하여 고주파 영역의 신호를 첨가함으로써 음질을 개선하였다. 비 손실 영역의 순음 및 비 순음 성분을 검출하여 손실영역에 해당 하모닉 성분을 청각 자극 에너지로 스케일 하여 새로운 신호를 첨가한다. 원 신호와 저 비트오율 압축으로 인해 왜곡된 신호, 그리고 본 논문의 알고리즘을 이용하여 개선된 신호를 신호 대 잡음 비율 측정하고 청감 테스트를 통해 음질 개선 효과를 확인하였다.

1. 서론

오디오 압축 기술은 통신 기술의 발전으로 인하여 새로운 변화를 필요로 하게 된다. 기존의 오디오 데이터를 매체에 저장하기 위한 압축 방식에서 이동 망이나 인터넷을 통한 빠른 전송과 실시간 서비스와 같은 제한된 대역폭에서의 고 음질, 고효율의 압축 방식을 요구하게 된다. 이러한 변화에 따라 통신환경에 적합한 압축 표준안이 제정되었고 새로운 환경에서의 다양한 압축 알고리즘이 제안되고 있다.[3][4] 그러나 새로 제안

된 표준이나 압축 알고리즘은 현재 사용되고 있는 기존의 오디오 데이터[1][2]의 재활용이 불가능하고 새로운 부호화 과정이 필요로 하는 문제점을 가지고 있다.

본 논문에서는 새로운 부호화 과정 없이 기존의 오디오 데이터를 이용하여 음질을 개선하는 방법을 제안한다. 저 비트오율 압축 오디오 신호에서 비 손실 영역의 순음성분(Tonal component) 및 비 순음성분(Nontonal component)과 청각 자극에너지(Excitation energy)를 얻어 손실영역의 해당 하모닉 성분(Harmonic component)을 사용하여 손실영역에 새로운 신호를 첨가하여 음질을 개선하고자 하였다.

2. 제안된 알고리즘

2.1 순음 및 비 순음 성분 검출

순음성분의 검출은 임의의 주파수 인덱스(Frequency index) k 에 대해서 다음 식 (1)의 조건을 만족하는 성분으로 결정한다. 즉, $P(k) > P(k \pm 1)$ 인 조건을 만족하는 $P(k)$ 를 Local maximum값으로 결정하고, 이렇게 결정된 Local maximum값들 중 Δ , 로 구분되는 구간 중 최소 7dB 이상 차이 나는 주파수 성분들을 순음 성분으로 결정한다.[5]

$$S_i = \begin{cases} P(k) & | P(k) > P(k \pm 1) \\ P(k) & \geq P(k \pm \Delta_i) + 7dB \end{cases} \quad (1)$$

여기서 Δ_i 는 임계 대역(Critical band)을 기반으로 한 주파수 범위로서 다음과 같이 나타낼 수 있다.[5]

$$\Delta_i = \begin{cases} 2 & (2 < k < 63) \\ [2, 3] & (63 \leq k < 127) \\ [2, 6] & (127 \leq k < 255) \\ [2, 12] & (255 \leq k < 500) \end{cases} \quad (2)$$

비 순음 성분은 각각의 임계대역에 대해서 순음성분이 검출된 주파수 인덱스 구간 $\pm \Delta_i$ 를 제외한 주파수 성분들을 합산하여 비 순음 마스커(Masker)의 음악 레벨로 계산한다.[5]

$$P_{nm}(\bar{k}) = 10 \log_{10} \frac{r(\omega)}{10} \text{ dB} \quad (3)$$

where, $\forall P(j) \in \{P_{nm}(k, k \pm 1, k \pm \Delta_i)\}$

2.2 청각 자극에너지 계산

인간의 청각 시스템의 주파수 응답 특성은 임계 대역의 전체 신호 에너지에 대한 마스킹 특성 및 지각적 에너지 확산 현상을 나타내는 청각적 자극 에너지로 해석할 수 있다. 주파수 변별력을 나타내는 임계 대역은 바크 울(Bark ratio)로 표현하며, 청각적 자극 에너지는 이러한 바크 스케일을 기준으로 이루어지게 된다. 임의의 주파수 f 에 대해서 바크 울은 다음과 같이 표현된다.[2][5]

$$z / \text{Bark} = 13 \arctan(0.76 f / \text{kHz}) + 3.5 \arctan(f / 7.5 \text{kHz})^2 \quad (4)$$

여기서 z 는 신호의 샘플링 주파수에 의해 결정되는 전체 임계 대역의 수를 나타내고 임계대역의 대역폭은 다음과 같이 표현된다.[2][5]

$$\Delta \omega_c / \text{Hz} = 25 + 75[1 + 1.4(f / \text{Hz})^2]^{0.69} \quad (5)$$

임의의 임계 대역에서의 전력 스펙트럼 $X(\omega, i)$ 에 대한 임계 대역 에너지는 다음과 같이 계산한다.[2][5]

$$X_c(z, i) = a_c(z) \sum_{n=\alpha_c}^{\beta_c} X(\omega, i) \quad (0 \leq z \leq Z-1) \quad (6)$$

여기서 $a_c(z)$ 는 외이(外耳)에서부터 중이(中耳)까지

의 다양한 전달 특성 요소들에 의한 주파수 의존 감쇠 특성을 나타내는 감쇠 성분이다. 이러한 임계 대역 에너지를 인접 대역과 다른 대역에 영향을 주게 되는데 이를 고려해준 청각 자극 에너지는 다음과 같다.[2][5]

$$X_c(z, i) = \sum_{v=0}^{z-1} [SF(v, z) X_s(v, i)] \quad (0 \leq z \leq Z-1) \quad (7)$$

여기서 $SF(v, z)$ 는 다음과 같이 표현될 수 있다.[5]

$$SF(\Delta z) = 15.81 + 7.5(\Delta z + 0.474) - 17.5 \sqrt{1 + (\Delta z + 0.474)^2} \quad (8)$$

2.3 하모닉 성분 검출

다양한 종류의 신호는 정형파의 합 s_n 과 잡음성분의 q_n 으로 표현이 가능하다. 그 표현 수식은 다음과 같다.[7][8]

$$x_n = s_n + q_n \quad n = 0, 1, \dots, N-1 \quad (9)$$

$$s_n = \sum_{m=1}^M A_m \cos(\omega_m n + \phi_m)$$

여기서 A_m 과 ϕ_m 은 주파수가 ω_m 인 정형파 성분에 대한 크기와 위상을 나타내며, M 은 정형파의 개수를 나타낸다. 신호를 이루고 있는 정형파 성분은 기본주파수(Fundamental frequency)와 하모닉(Harmonic) 관계를 통해서 나타낼 수 있으며 기본주파수와 하모닉 관계는 다음과 같다.[6][9]

$$F_n = \begin{cases} F_0 \times 2^k & (\text{octave}) \\ F_0 \times (3/2)^k & (\text{perfect fifth}) \\ F_0 \times (4/3)^k & (\text{perfect fourth}) \\ F_0 \times (5/4)^k & (\text{major third}) \\ F_0 \times (6/5)^k & (\text{minor third}) \\ F_0 \times (9/8)^k & (\text{major second}) \end{cases} \quad (10)$$

(where, $k = 1, 2, \dots, 10$)

2.4 청각 신호의 크기 조정

식(7)을 사용하여 계산된 청각 자극에너지에서 비 순음 영역의 기본 주파수 성분과 손실영역의 하모닉 성분의 해당 청각 자극 에너지 비를 계산하여 손실영역의

첨가될 신호의 스케일 벡터(Scale factor)로 사용하며, 다음과 같이 나타낼 수 있다.

$$S_H = S_F \frac{X_H + \alpha}{X_F} \quad (11)$$

여기서 X_F 는 비 손실영역의 기본주파수의 해당하는 청각 자극에너지를 나타내며, X_H 는 손실영역의 첨가할 하모닉 성분의 청각 자극에너지를 나타낸다. α 는 첨가된 신호가 마스킹 되지 않게 하기 위한 상수 값이다.

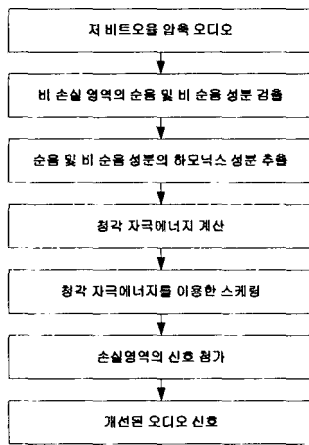


그림 1). 알고리즘 개념도

그림 1과 같이 본 논문에서 제안된 알고리즘은 별도의 부호화 과정이 없으므로 복호화 과정 후 독립적으로 사용되며, 비 손실영역에 순음 성분 및 비 순음 성분 검출과 손실영역의 해당하는 하모닉 성분을 얻어 청각 자극에너지를 이용한 첨가 신호의 스케일 과정으로 구성되어있다.

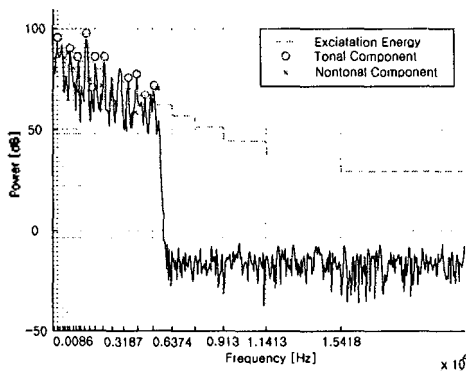


그림 2). 순음 및 비 순음검출과 청각 자극에너지

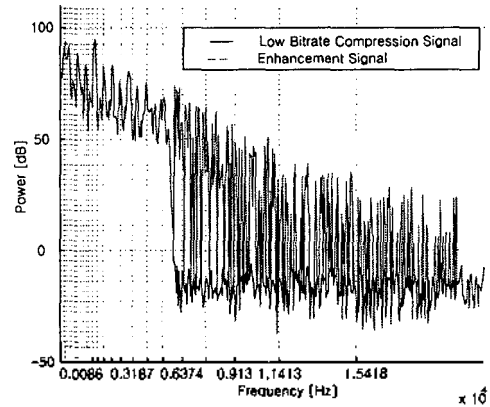


그림 3). 손실영역의 신호 첨가 결과

그림 2는 비 손실의 순음성분 및 비 순음 성분의 검출과 청각 자극에너지를 계산하는 과정을 보여주고 있으며, 그림 3에서는 저 비트오율 압축으로 인한 손실된 고주파영역에 순음성분 및 비 순음 성분의 하모닉 성분을 청각 자극에너지 값으로 스케일 하여 첨가한 신호를 보여주고 있다.

3. 실험 결과

제안된 알고리즘을 검증하기 위해 다양한 종류의 오디오 신호를 사용하였으며, 44.1kHz 샘플링 율과 16 bit/sample의 정밀도를 갖는 오디오 신호를 64kbps 비트 오율로 압축하여 실험을 수행하였다. 먼저 64kbps 비트 오율로 압축한 오디오 신호와 제안된 알고리즘을 사용하여 개선한 신호를 다양한 장르의 오디오 샘플을 이용하여 각각 객관적 측정방법인 신호 대 잡음 비(Signal-to-Noise Ratio, SSNR)를 측정하여 결과를 비교한 결과 신호차원의 좋은 개선 효과를 얻지 못하였다.[10]

Music Sample	SNR	
	64kbps	Enhancement
Symphony	25.5	24.7
Jazz	24.8	23.5
Pop	25.1	24.9
Piano	23.1	22.4

표 1). 신호 대 잡음 비 결과

객관적인 음질 향상 척도로서 사람이 직접 듣는 결과

와는 다를 수 있다. 따라서 주관적인 듣기평가 방법인 청감 테스트(Mean Opinion Score, MOS)를 실시하였으며, 결과는 표 2와 같다. 테스트는 20세 이상 청각적 장애가 없는 일반인 남녀 30명을 대상으로 객관적 평가에 사용된 동일한 오디오 샘플을 이용하여 1점에서 5점까지의 점수로 음질을 평가하였다.

Music Sample	MOS	
	64kbps	Enhancement
Symphony	2.1	4.01
Jazz	2.2	3.9
Pop	1.9	3.1
Piano	2.7	3.8

표 2). 청감 테스트 결과

객관적 음질 측정 결과와 같이 후처리 알고리즘 수행 후에도 1dB미만의 적은 신호의 왜곡이 발생됨을 확인할 수 있으며, 다양한 장르에 대한 청감 테스트 결과에서 알 수 있듯이 저 비트율로 압축된 오디오 신호에 비해서 우수한 음질 개선 효과를 확인하였다.

4. 결론

본문에서는 저 비트율 압축으로 인하여 왜곡된 오디오 신호를 추가 정보 없이 인간의 청각적 특성을 이용하여 개선하는 방법을 제안하였다. 실험 결과 지각적 음질 개선 효과를 평가하는 청감 테스트를 통해 음질 개선 효과를 확인할 수 있었다. 향후 과제로 객관적 음질 측정 개선효과 즉 신호차원의 개선을 위한 연구가 수행될 예정이다.

참고문헌

1. Davis Pan A Tutorial on MPEG/Audio Compression IEEE, 1995
2. Nikil JaYant, James Johnson and Robert Safranek "Signal Compression Based on Models of Human Perception", Proceedings of the IEEE, Vol 81, No. 10, October 1993
3. Yin H Lam and Robert W Stewart, "Perceptual Suppression of Quantization Noise in Low Bitrate Audio Coding", IEEE,

- 1998
4. 서정일, 서진수, 홍진우, 강경욱, "저비트율로 압축된 오디오의 음질 개선 방법", The Journal of the Acoustical society of Korea, 2002
5. E.Zwicker and H. Fastil, "Psychoacoustics, Facts and Models", Springer, 1999.
6. Max E. Valentinuzzi and Nestor E. Arias "Human Psychophysiological Perception of Musical Scales and Nontraditional Music" IEEE Engineering in Medicine and Bilogy, March/April, 1999
7. Boris Doval and Xavier Rodet "Estimation of Fundamental Frequency", IEEE, 1991
8. K.W.Chan and H.C.So, "Accurate Frequency Estimation for Real Harmonic Sinusoids", IEEE, 2004
9. Anssi P. Klapuri, "Multiple Fundamental Frequency Estimation Based on Harmonicity and Spectral Smoothness" IEEE, 2003
10. S. R. Quackenbush, T. P. Barnwell, III, and M. A. Clements, "Objective Measures of Speech Quality. Englewood Cliffs, 92nd AES Conv, Mar. 1992