통합 음성/오디오 부호화기의 Noise Filling 알고리즘에 대한 연구

*송정욱 **강홍구

연세대학교

*jeongook@dsp.yonsei.ac.kr

Study on Noise Filling algorithm of Unified Speech and Audio Coding

*Jeongook Song **Hong-Goo Kang

Yonsei University

요약

본 논문에서는 Unified Speech and Audio Coding (USAC)에 적용된 Noise Filling의 부호화 과정에서 음질 왜곡 정도에 따라 Noise level을 설정하는 방법을 제안한다. USAC는 Moving Picture Experts Group (MPEG)에서 표준화한 최신의 음성/오디오 통합 코덱으로 현존하는 코덱 중에 최고의 성능을 가지고 있다. 하지만, 복호화기 기술만 표준화 하여, 인코더를 설계하는 방법에 따라 음질의 차이가 존재한다. 현재 오픈 소스 기반으로 진행되고 있는 프로젝트 JAME에서는 이러한 음질 차이를 극복하고, USAC에 적용된 핵심 인코더 기술의 성능을 최대화 할 수 있는 여러 가지 방법을 포함하고 있다. 그 중 Noise Filling은 저 전송률 부호화 과정에서 양자화 되지 않는 스펙트럼에 대하여 일정한 noise level을 넣어 인지적으로 음질을 향상시키는 방법이다. 제안된 Noise Filling 부호화 방법은 현재 프레임의 음질 왜곡 정도를 반영하여, noise-like 신호 성분을 더욱 정교하게 부호화 할 수 있게 하였다.

1. 서론

전통적으로 오디오 코덱은 인간의 청각적 모델을 바탕으로 설계되어 발전하여 왔다. 이러한 청각적 모델을 심리 음향 모델(Psychoacoustic Model)이라고 부르며, 이는 인간의 가청 주파수 내에서 일반적인 소리의 인지 정도를 나타낸다. 또한 마스킹 효과(Masking effects)를 심리 음향 모델에 이용하여 차등적으로 비트를 할당하도록 설계하여 양자화 효율을 올렸다. 이러한 오디오 부호화의 기술들은 현재 대표적인 오디오 코덱인 AAC [1](Advanced Audio Coding)에 핵심 기술이 되었다.

하나의 단말기에서 통신과 방송의 융합된 컨텐츠를 구분 없이 사용할 수 있게 되자, 기존 오디오 코덱 또한 보다 다양한 컨텐츠를 효율적으로 양자화할 수 있는 새로운 구조로 다시 설계될 필요가 있었다. 이에 따라 MPEG에서는 대표적인 음성 코덱인 AMR-WB+[2](Adaptive Multi-Rate Wide-Band plus)를 AAC와 선택적으로 사용할 수 있는 USAC(Unified Speech and Audio Coding)[3]를 표준화 하였다.

Noise Filling은 USAC의 주파수 영역 부호화기에 적용된 알고리즘이다. 본 논문에서는 USAC에 적용된 Noise filling 알고리즘의 한계점을 지적하고, 매 프레임 양자화 에러에 따라 가변적으로 noise level을 설계하는 새로운 noise filling 알고리즘을 제안한다.

2. Noise Filling 알고리즘

USAC에 적용된 Noise Filling 알고리즘은 마스킹 효과에 의하여 양자화 되지 않는 스펙트럼에 대하여, 원 신호의 평균 에너지의 일정 크기만큼의 노이즈를 채워주는 방법이다. 총 8Bit로 양자화 되며, 상위 3bit는 0으로 양자화된 스펙트럼의 크기를 생성하는 Noise level을 만들며, 하위 5bit는 0으로 양자화된 주파수 밴드(sfb, scale factor band)의 크기를 생성하는 Noise offset을 만든다.

Noise level을 l이라고 하면 양자화된 n번째 스펙트럼은 다음과 같다.

$$\widehat{x_n} = \begin{cases} \pm s|q[n]|2^{0.25scf[sfb]-25} & , q[n] \neq 0 \\ \pm s2^{\frac{l-14}{3}}2^{0.25scf[sfb]-25} & , q[n] = 0 \end{cases} \quad (1)$$

여기서 $q[n]$은 스펙트럼 양자화 값을, s는 부호함수를, sfb는 scale factor band index를, $scf[sfb]$는 서브밴드 $scf[sfb]$의 scale factor값을 나타낸다. $q[n]=0$인 스펙트럼의 에너지를 $\widehat{E_{on}}$이라고 하면, 수식(1)을 이용하여 다음과 같이 전개가 가능하다.

$$\begin{aligned} \widehat{E_{on}} &= \sum_{sfb} \sum_{n \in sfb, q[n]=0} \widehat{x_n^2} \quad (2) \\ &= \sum_{sfb} \sum_{n \in sfb, q[n]=0} (\pm s2^{\frac{l-14}{3}}2^{0.25scf[sfb]-25})^2 \\ &= \sum_{sfb} 2^{0.5scf[sfb]-50} \sum_{n \in sfb, q[n]=0} 2^{\frac{2l-28}{3}} \\ &= 2^{\frac{2l-28}{3}} \sum_{sfb} 2^{0.5scf[sfb]-50} N_0[sfb] \end{aligned}$$

여기서 $N_0[sfb] = \sum_{n \in sfb, q[n]=0} 1$ 으로 scale factor 밴드에서 양자화 스펙트럼 값이 0인 개수를 말한다. 양자화된 신호의 에너지가 원신호의 에너지에 대하여 α크기만큼 비례한다고 하면, $\widehat{E_{on}}$은 다음과 같이 표현할 수 있다.

$$\widehat{E_{on}} = \sum_{n, q[n]=0} \hat{x}_n^2 = \alpha \sum_{n, q[n]=0} x_n^2 \quad (3)$$

수식 (2)과 수식 (3)을 이용하여 noise level값을 구하면, 다음과 같다.

$$l = 1.5\left[\log_2\left(\sum_{n, q[n]=0} x_n^2\right) - \log_2\left(\sum_{sfb} 2^{0.5scf[sfb]-50} N_0[sfb]\right)\right] + 14 \quad (4)$$

여기서 noise level은 [0, 7]범위에서 가장 가까운 정수로 양자화하여 복호화기에 전송된다.

Scale factor 밴드의 모든 스펙트럼이 0인 경우 noise offset, f을 이용하여 복원하며, 수식은 다음과 같다.

$$\hat{x}_n = \pm s 2^{\frac{l-14}{3}} 2^{0.25(scf[sfb]+f-16)-25}, q[n]=0 \quad (5)$$

위의 noise level 값을 얻는 방식과 같이 noise offset을 사용하여 복원된 scale factor 밴드 에너지가 원신호의 에너지에 대하여, α만큼 비례한다고 가정하면, noise offset, f은 다음 수식과 같다.

$$f = 2\log_2\left(\sum_{n, q[n]=0} x_n^2\right) - 2\log_2\left(\sum_{sfb} 2^{0.5scf[sfb]-58} N_0[sfb]\right) + \frac{56-4l}{3} \quad (6)$$

3. 제안된 Noise Filling 알고리즘

현재 USAC - JAME [4]에 적용된 Noise Filling 알고리즘의 에너지 비례 상수, 수식 (3)의 α는 0.5로 고정된 값을 가진다. 현재 프레임의 양자화 잡음의 크기와 상관없이 noise filling알고리즘을 적용할 경우, 인위적인 잡음이 삽입된 것처럼 들릴 수 있다. 따라서 에너지 비례 상수, α는 현재 프레임의 양자화 에러를 반영하여 다음과 같이 정의할 수 있다.

$$\alpha = \gamma\left(1 - \frac{\sum_{n, q[n]\neq 0} (x_n - \hat{x}_n)^2}{\sum_{n, q[n]\neq 0} x_n^2}\right) \quad (8)$$

4. 성능 비교

그림 1은 음악 샘플에 대한 기존 Noise filling과 제안된 Noise filling 방법을 비교 분석을 위하여 원신호의 스펙트로그램에 해당하는 프레임 별 스펙트럼의 에러 비율을 그린 그림이다. 여기서 스펙트럼의 에러 비율은 현재 프레임의 전체 양자화 에러와 Noise filling이 적용된 스펙트럼의 양자화 에러 비율을 말한다. 그림에서 보여주는 것과 같이 제안된 방법에서 기존의 방법과 달리 에러 비율이 일정하게 나타남을 알 수 있다.

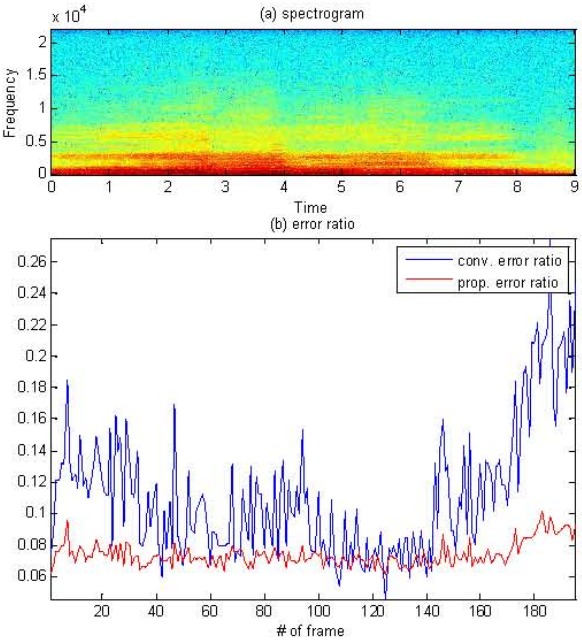

그림 1 (a) 원신호의 스펙트로그램 (b) 기존 noise filling이 적용된 에러 비율과 제안된 noise filling이 적용된 스펙트럼의 에러 비율

5. 결론

본 논문에서는 USAC에 적용된 개선된 Noise filling 방법을 제안하였다. 제안된 방법은 기존 방식 보다 매 프레임 일관되게 전체 양자화 에러와 Noise filling이 적용된 스펙트럼의 에러 비율을 유지하여, noise-like 신호 성분을 더욱 정교하게 부호화하였다.

6. 참고 문헌

[1] ISO/IEC 14496-3:2009, "Coding of Audio Visual Objects, Part 3: Audio," 2009.

[2] 3GPP, "Audio codec processing functions; Extended Adaptive Multi-Rate – Wideband(AMR-WB+) codec; Transcoding functions," 3GPP TS 26.290, 2004.

[3] M. Neuendorf, et al., "A novel scheme for low bitrate unified speech and audio coding-MPEG RM0," in Proceedings of the *126th AES Convention*, Munich, Germany, May 2009

[4] J. Song, H. O. Oh, and H.G. Kang, "Enhanced long-term predictor for Unified Speech and Audio Coding," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Prague, Czech, May 2011.