다중밴드 양자화를 적용한 USAC 부호화 기술

백승권, 임우택, 이태진
한국전자통신연구원 미디어부호화연구실
{skbeack, wtlim, tjlee}@etri.re.kr

Seungkwon Beack, Wootaek Lim, Taejin Lee

Media Coding Research Section
Electronics and Telecommunications Research Institute (ETRI)

요 약

본 논문은 USAC(Unified Speech and Audio Coding) 오디오 부호화 기술의 성능 개선에 관련한 것이다. USAC 은 FD(Frequency domain) 양자화 모듈과 LPD(Linear prediction domain) 양자화 모듈을 탑재하고 있다. 본 논문에서는 LPD 모드로부터 생성되는 잔차신호에 대하여 주파수 영역에서 다중밴드로 분할하고 각 밴드 별 양자화를 독립적으로 수행함으로써 USAC 의 LPD 모드의 양자화 효율을 개선하였다. 그 결과 동일 조건에서 제안 방법이 기존의 LPD 모드의 성능을 음질 측면에서 향상시킴을 확인할 수 있었다.

1. 서론

음성 및 오디오 부호화 기술은 지속적으로 고효율 압축과 높은 음질을 달성하기 위하여 압축 알고리즘이 개발되어 왔다. 특히 MPEG 국제 표준화 기관을 통해 표준 기술로 개발된 오디오 코덱 기술은 단방향 스트림 오디오 서비스 시장에서 성공적으로 널리 사용되어 왔다. 현재 MPEG 을 통해 표준화가 완료된 코덱으로 가장 높은 성능을 나타내고 있는 코덱은 USAC 으로 2012 년도에 표준이 제정되었으며, 현재까지도 지속적인 검증 작업이 표준화 기관을 통해 수행되고 있다[1]. USAC 은 음악신호 뿐만 아니라 음성에 대해서도 높은 성능을 나타내는 오디오 코덱으로, 기존의 오디오 코덱이 갖고 있는 음성에 대한 음질 열화 문제를 극복하였고, 기존의 오디오 코덱의 성능 또한 향상시킴으로써 차세대 오디오 코덱의 근간을 마련할 수 있었다[2]. USAC 의 부호화 방식은 hybrid 방식을 채택하였으며, 음성과 음악신호 특성에 따라 부호화 모듈을 달리한다. 주로 음악적 신호가 강한 입력 신호에 대해서는 기존의 오디오 코딩 방식인 FD 모드에서 부호화를 수행하며, 음성적 신호가 강할 경우에는 LPC(Linear prediction coefficient)활용한 LPD 코딩 방식을 적용하여 부호화 한다. 그러나 실제로는 음악 신호라도 부호화 효율이 있다면 LPD 모드로 충분히 부호화가 가능하며, 마찬가지로 음성에 가까운 신호라도 FD 모드로 부호화를 진행할 수 있다. 따라서 USAC 은 이러한 상이한 부호화 모듈 간의 스위칭이 발생했을 때 음질 왜곡을 최소화할 수 있는 평탄화 기술을 탑재하고 있으나, 여전히 입력 신호에 따라 상이한 모듈로 양자화를 수행함으로써 구조적 복잡도와 예상치 못한 스위칭으로 인하여 간헐적인 부호화 왜곡을 완벽하게 제거할 수는 없는 문제점을 안고 있다.

본 논문은 USAC 의 LPD 모드의 성능 개선에 관련한 방법을 제안하고 있다. 궁극적으로는 LPD 모드의 지속적인 성능 개선을 통해 음성/음악에 대하여 통합적인 양자화 효율을 제공할 수 있는 수준까지 개발하고자 한다. 그 일환으로, LPD 의 TCX(Transform Coded Excitation) 모드의 성능 개선을 위하여 다중 밴드 양자화 수행 전략을 반영하여 그 효과를 검증하고자 하였다.

2. TCX 부호화 기술

USAC 의 LPD 모드는 ACELP(Algebraic Code-Excited Linear Prediction)와 TCX 모드로 구성되어 있다. ACELP 는 주로 음성 신호를 부호화 하기 위하여 활용되며, TCX 는 음성의 정적인 구간 혹은 음성과 음악이 혼재되어 있는 신호의 부호화를 위하여 선택적으로 활용된다. TCX 는 LPC 필터링 이후의 잔차신호에 대하여 단구간 프레임을 취한 후, 주파수 영역으로 변환하여 양자화를 수행하는 방식이다. TCX 부호화 모드는 단구간 프레임의 크기에 따라 3 가지로 정의되어 있으며, 12.8 kHz

표본화 주파수를 기준으로 TCX20(20msec TCX), TCX40(40msec TCX), TCX80(80msec TCX)로 구성되어 있다. 본 논문에서는 제안 방식의 유효성을 검증하기 위하여 TCX80 으로 부호화 모드를 고정하고 동일 조건에서 양자화를 수행한 후 복원된 신호의 성능을 평가하고자 한다.

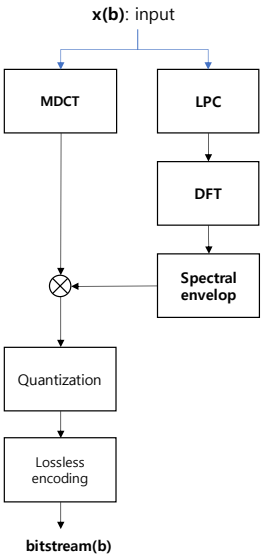

그림 1. TCX 부호화기 구조도

그림 1은 USAC의 TCX 기본 구조도를 간략하게 나타내었다. 입력 프레임 신호 블록 x(b)는 MDCT(Modified Discrete Cosine Transform)를 수행하여 주파수 계수로 변환된다. 또한 입력 블록에 대한 LPC 계수를 구한 뒤 DFT (Discrete Fourier Transform)를 거쳐 복소 주파수 계수로 변환된다. 변환 된 복소 주파수 계수에 절대값을 취하여 LPC 계수의 특성인 주파수 영역의 주파수 포락선(spectral envelop)을 얻게 되며, 포락선 정보의 역인 계수를 MDCT 계수에 곱하여 MDCT 계수를 평탄화 시킨다. 최종적으로 평탄화 된 MDCT 계수는 양자화 및 무 손실 부호화 과정을 통해 비트스트림으로 변환된다.

3. 다중밴드 양자화 TCX 부호화 기술

기존 USAC 의 TCX 는 주파수 영역에서 양자화를 수행하기 위해서 단일 스케일팩터(scale-factor)를 적용하여 양자화를 수행한다. 그러나 본 논문에서 제안하는 방식은 단일 스케일팩터 기반의 양자화 구조를 다중밴드 양자화 구조로 변환하여 양자화를 수행하였다. 이를 위해 먼저 주파수 영역의 잔차 신호를 다음과 같이 정의한다.

$$\mathbf{res}(k) = \left[res(B(k-1)), res(B(k)+1), ..., res(B(k+1)-1) \right]^T \quad (1)$$

수식 (1)로부터 양자화를 위한 잔차신호 다중밴드 수는 B 개로 정의됨을 알 수 있으며, 각 서브밴드의 잔차신호 계수 벡터열은 $\mathbf{res}(k)$ 로 표현하였다. 마찬가지로 서브밴드 경계정보 *B(B) = N/2* 이고, N 은 MDCT 변환 크기이며, *B(0) = 0* 으로 정의된다. 이렇게 정의된 각 서브밴드마다 하나의 스케일팩터 *SF(k)*를 추정한다. *SF(k)*를 추정하는 방식은 $\mathbf{res}(k)$ 를 대표하는 중간 값 에너지로부터 환산될 수 있으며, 양자화에 활용 가능한 비트 수를 참조하여 이를 추정할 수도 있다. 각 서브밴드 별로 *SF(k)*를 구하는 과정은 다음과 같은 과정을 통해 수행된다.

- 해당 서브밴드의 가용 비트 수를 할당한다. (X bits)
- 총 가용 비트로부터 표현 가능한 dB 를 계산한다. 예를 들어 1 bit 가 표현할 수 있는 dB 는 하나의 주파수 샘플에 대해서 6dB 를 표현할 수 있다. 즉, 각 서브밴드별로 dB 에너지를 구하고 가용 비트로 표현 가능한지 파악한다. (서브밴드 전체 에너지 / 6dB = 최대 필요 비트 수)
- 만약 비트가 부족하다면 12 dB (6 x 2) 로 나누어 bit 수 계산하고 가용 비트내에서 만족하는지를 파악한다.
- 비트수가 남는다면 바이너리 방식으로 6 dB 와 12 dB 를 탐색하여 가장 적절한 단위의 dB 스텝을 SF 의 크기 dB 로 찾는다.

각 서브밴드의 SF 를 추정하였다면 양자화는 다음과 같이 수행한다. 먼저 주파수 계수의 크기 정보에 대하여 양자화를 수행하며, 각 서브밴드별로 구해진 SF 값을 적용한다.

$$abs\left(\mathbf{resQ}\left(B(k):B(k+1)-1\right)\right) = \left\lfloor \left| 20\log_{10}\left(abs\left(\mathbf{res}_f\left(B(k):B(k+1)-1\right)\right)\right) - SF(k) \right\rfloor, \quad 0 \le k \le B-1 \quad (2)$$

식 (2)에서 $\mathbf{resQ}(k)$ 는 잔차 신호의 양자화된 계수이며 $\mathbf{res}_f(k)$ 는 MDCT 잔차신호 계수를 나타낸다. 또한 위상 정보의 복원을 위하여 수식 (3)과 같이 위상 정보를 취한다.

$$angle(\mathbf{resQ}(B(k):B(k+1))) = angle(\mathbf{res}_f\left[B(k):B(k+1)\right]), \quad 0 \le k \le B-1 \quad (3)$$

위상정보는 변환된 주파수 영역이 실수부만 존재한다면 각 계수의 부호(Sign)정보만이 위상정보로 취해진다. 만일에 주파수 영역이 복소수 영역이라면 arctangent 연산을 통하여 실수부와 허수부의 정보로부터 위상 정보를 얻는다. 본 논문에서는 MDCT 를 주파수 변환 방식으로 적용하였으므로, 부호 정보를 위상정보로 전송한다.

양자화된 주파수 성분의 복원 방법은 수식 (4)와 같이 나타낼 수 있다.

$$\mathbf{resQ}(B(k):B(k+1)) = abs\left(\mathbf{res}_f\left(B(k):B(k+1)\right)\right) \cdot \exp\left(j \times angle\left(\mathbf{res}_f(B(k):B(k+1))\right)\right) \quad (4)$$

4. 실험 결과

본 논문에서는 제안된 방법의 유효성을 검증하기 위하여 주관적 음질 평가를 수행하였다. 평가 환경은 표 1 과 같이 구성하였다. 먼저 주관적 음질 평가 방법은 MUSHRA 방식을 따랐다[3]. MUSHRA 방식은 블라인드 방식의 주관적 성능평가 방법으로 hidden reference 와 anchor 시스템을 포함하여 청취테스트를 수행한다.

표 1. 주관적 성능측정 평가 환경

평가 환경	채택 항목
평가 방법	MUSHRA
피험자	6 명
평가 아이템/비트율	12 개(10 초 이상)/13kbps
표본화 주파수	12.8 kHz
평가 시스템	org : Hidden reference
	lp3.5 : Anchor 3.5kHz
	System A : USAC TCX80
	System B: AMR-WB+
	System C: 제안 시스템 TCX80

본 논문의 제안 방법은 음성/오디오 코어대역에 주로 적용되는 기술로, 테스트 아이템의 표본화 주파수는 12.8 kHz 로 설정하였으며, 부호화 대역은 6.4 kHz 의 대역폭을 갖는다. 참조 비교 시스템의 선정은 음성 코덱으로 최고의 성능을 나타내는 AMR-WB+와, 오디오 코덱으로 최고 성능을 나타내는 USAC 을 참조 시스템으로 선정하였다[2, 4]. 제안 방식의 유효성을 검증하기 위하여 코딩 모드는 모두 80msec 의 TCX80 모드로 고정하여 부/복호화를 수행하였다. 비트율(Bitrate)은 13kbps 로 저 비트율을 선정하여 테스트를 수행하였다. 이는 부호화 효율 차이를 극명하게 확인하기 위해 양자화 왜곡이 심한 저 비트율에서 음질 비교를 수행하였기 때문이다. 테스트 아이템은 총 8 개로 음성 아이템 3 개, 음악 아이템 3 개, 음악과 음성이 동시에 포함된 2 개의 아이템으로 구성하였으며, 모두 MPEG 오디오 표준화 단체에서 선정한 테스트 아이템이다.

그림 2 와 3 은 청취 평가를 수행한 결과이다. 그림 2 는 청취 평가를 수행한 절대 점수에 대한 평균치를 기준으로 95% 신뢰구간을 나타낼 수 있도록 표기하였다. 신뢰구간이 겹치지 않을 경우, 비교 시스템 간에 통계적으로 의미 있는 차이가 있다고 판단할 수 있다. 절대 점수에 대한 분석결과, 아이템 별로 신뢰 구간이 겹치고 있으나 전체 평균에 대한 제안 시스템의 성능은 비교 시스템들과 신뢰구간에서 겹치지 않는 성능을 나타내고 있다. 즉, 제안된 다중 밴드 스케일팩터 기반 양자화 방식은 기존의 최신 코덱 기술과 비교하여 주목할 만한 개선된 성능을 나타내고 있음을 관측할 수 있다.

아이템별로 보다 분명한 차이를 보기 위하여 절대 평가 점수의 시스템 간의 차이를 통계적으로 분석하여 그림 3 과 같이 나타내었다. 시스템 간의 차이는 제안 시스템을 기준으로 차이 점수를 산출하였다. 차분 점수의 통계치가 평균값을 기준으로 95% 신뢰구간이 0 점에 걸치지 않는다면 두 시스템 간의 차이가 통계적으로 유 의미 하다고 판단할 수 있다. 제안시스템을 기준으로 차이 점수에 대한 통계치를 그림 3 로부터 살펴보면 3 개의 아이템에 대해서는 비교 시스템보다 우세함을 관측할 수 있으며, 마찬가지로 전체 차이 점수의 평균은 비교시스템보다 통계적으로 우세한 결과를 얻을 수 있었다.

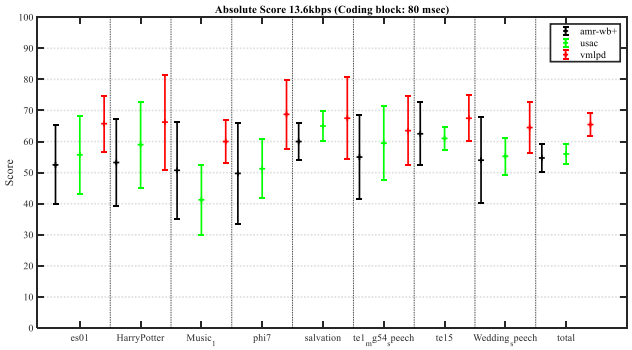

그림 2. 주관적 성능평가 절대점수 평균치 비교

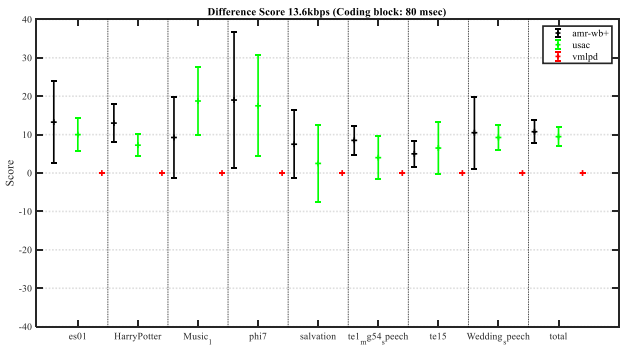

그림 3. 주관적 성능평가 절대점수차 평균치 비교 (system C 기준)

5. 결론

본 논문에서는 다중밴드 양자화를 적용한 개선된 USAC 부호화 기술을 제안하였다. 제안 방법은 기존 최고 음성 코덱인 AMR-WB+와 기존 오디오 최신 코덱인 USAC 과 성능과 비교했을 때, 통계적으로 유효한 성능 향상을 검증하였다. 본 제안 기술은 주로 장구간 선형 예측 부호화 기술에 적용하였을 때 성능 개선을 예상할 수 있다. 향후 USAC 의 단구간 선형예측 부호화 기술을 개선하거나 대치 할 수 있는 기술 개발이 필요하며, 최종적으로 다양한 부호화 모드를 활용하는 USAC 의 최적의 부호화 성능을 개선할 수 있는 기술개발 연구가 진행되어야 할 것이다.

감사의 글

이 논문은 2020 년도 정부(과학기술정보통신부)의 재원으로 정보통신기술진흥센터의 정보통신기획평가원의 지원을 받아 수행된 연구임 (No. 2017-0-00072, 초실감 테라미디어를 위한 AV 부호화 및 LF 미디어 원천기술 개발)

참고문헌

[1] ISO/IEC JTC1/SC29/WG11, “Unified Speech and Audio Coding Verification Test Report,” Torino, Italy, July 2011, MPEG2011/N12232.

[2] M. Neuendorf, et al, “The ISO/MPEG Unified Speech and Audio Coding Standard: Consistent high quality for all content types and at all bit rates,” Journal of the AES, vol. 61, no. 12, pp. 956-977, Dec. 2013.

[3] International Telecommunication Union, “Method for the subjective assessment of intermediate sound quality (MUSHRA),” 2001, ITU-R, Recommendation BS, 1543-1, Geneva, Switzerland.

[4] 3GPP, “Adaptive Multi-Rate - Wideband (AMRWB) Speech Codec; General Description,” 2002, 3GPP TS 26.171.