

MPEG-D USAC : LPD 복호화기의 고정 소수점 알고리즘 구현

*송은우 **송정욱 ***강홍구

연세대학교

*sewplay@dsp.yonsei.ac.kr

Fixed-point Implementation of LPD Decoder in MPEG-D USAC

*Eunwoo Song **Jeongook Song ***Hong-goo Kang

Yonsei University

요약

본 논문에서는 MPEG-D 오디오 서브그룹에서 진행 중인 Unified Speech and Audio Coding (USAC) 표준의 Linear Prediction Domain (LPD) 복호화기 모듈을 고정소수점 알고리즘으로 제안한다. USAC 부호화기는 두 개의 최신 음성-오디오 부호화기가 융합된 형태로, 음성 및 오디오 신호에 대하여 우수한 성능을 갖는 부호화기이다. USAC의 표준 완료와 본격적인 서비스화에 앞서서 USAC LPD 복호화기의 구조적인 특성을 분석하고, Digital Signal Processor (DSP) 구현을 위한 LPD 복호화기의 고정소수점 알고리즘을 구축하는 동시에 모듈의 복잡도를 측정하고자 한다. 또한 고정소수점 알고리즘으로 구현된 LPD 복호화기와 기존의 부동소수점 복호화기의 성능을 비교하고, LPD 복호화기의 두 가지 부호화 모드에 따른 복잡도 이슈를 다루도록 한다.

1. 서론

기존의 음성, 오디오 부호화기는 서로 다른 응용분야에서 서로 다른 방향으로 발전되어 왔기 때문에, 음성 부호화기는 음성 신호에, 오디오 부호화기는 음성신호에 성능이 떨어진다 단점이 있다. 하지만 모바일, 디지털 TV, 디지털 라디오, 오디오 북 등의 다양한 멀티미디어 기기가 보급되고 이에 따른 서비스 융합의 수요가 증가하면서, 음성과 음악 신호에 우수한 음질을 제공할 수 있는 부호화 기술이 한 가지 이슈로 떠오르게 되었다.

MPEG-D 오디오 서브그룹에서는 최신 음성 부호화기인 AMR-WB+와 최신 오디오 부호화기인 HE-AAC v2가 융합된 USAC 부호화기의 표준화 작업을 진행하였다 [1][2][3]. USAC 부호화기는 선처리 단의 신호 분류 (Signal Classifier) 모듈에서 입력신호의 특성을 파악하고, 이에 따라 두 부호화기를 선택적으로 동작시킨다. USAC의 Frequency Domain (FD) 부호화기인 HE-AAC v2는 음악 특성이 강한 신호를, Linear Prediction Domain (LPD) 부호화기인 AMR-WB+는 음성 특성이 강한 신호를 부호화한다. 97차 MPEG 회의에서 보고된 USAC 부호화기의 성능 평가 결과, 모든 비트레이트 (8kbps mono - 96kbps stereo)에서 3가지 콘텐츠 (음성, 음악, 음성과 음악이 섞인 신호)에 대하여 기존 부호화기에 비해 우수한 성능이 평가되었다 [4].

하지만 MPEG-D USAC의 표준 완료와 본격적인 상업화에 앞서 기존의 부동소수점 부호화기의 고정소수점 알고리즘은 아직 해결되지 않은 과제로 남아있다. 또한 구조적인 측면에서 USAC은 2개의 부호화기가 결합되어 있을 뿐만 아니라 Context Adaptive Arithmetic Coder (CAAC), Weighted Linear Prediction Transform (wLPT) 등, 표준화 과정에서 여러 틀들이 새롭게 도입되었기 때문에 모듈의 복잡

도가 조심스레 검토되어야 할 필요가 있다. 따라서 본 논문은 USAC 부호화기 중 LPD 모듈을 고정 소수점 알고리즘으로 구현하고, 그 성능과 복잡도를 분석하고자 한다. 고정소수점 알고리즘은 USAC의 참조 모델인 JAME 부호화기에 적용하였으며 16비트, 32비트 연산이 가능한 ITU-T 표준 고정소수점 라이브러리를 사용하였다 [5][6].

본 논문의 구성은 다음과 같다. 2장에서는 USAC 부호화기의 전반적인 구조를 설명하고, 3장에서는 USAC LPD 모듈의 부호화과정을 다루도록 한다. 4장은 실험 환경과 성능 평가로 구성되어 있으며, 5장에서는 고정소수점으로 구현한 USAC LPD 부호화기를 정리하고자 한다.

2. USAC의 구조

MPEG-D USAC의 구조는 크게 2개의 코어 대역 부호화기 모듈과 대역폭 확장 모듈로 이루어진다.

코어 대역 부호화기는 입력신호의 통계적 특성에 따라 FD 모듈 혹은 LPD 모듈이 선택적으로 동작한다. FD 모듈은 음악 특성이 강한 신호를 위한 AAC 기반의 부호화 모듈이고, LPD 모듈은 음성 특성이 강한 신호를 위한 Algebraic Code-Excited Linear Prediction (ACELP) 혹은 wLPT의 모듈이다. 프레임과 프레임 사이에서 FD 모듈과 LPD 모듈이 바뀔 때는 부드러운 전이를 위한 모드 변환 틀이 추가적으로 동작한다.

대역폭 확장 모듈은 표준 Spectral Band Replication (SBR) 방법을 기반으로 한다 [7]. SBR은 저주파 신호를 이용하여 고주파 까지 대역폭을 확장하는 방식으로 USAC에서는 밴드폭 확장 방법이 개선된 Enhanced SBR을 사용한다.

3. LPD 모듈의 복호화 과정

그림 1은 LPD 모듈로 동작하는 USAC 복호화기의 블록도이다. 복호화 과정은 크게 선처리, LPD 복호화, 후처리 과정으로 나눌 수 있다. 선처리 모듈은 비트스트림으로부터 LPD 복호화를 위한 Linear Prediction Coefficient (LPC) 복호화 모듈과, 양자화된 spectral 계수를 수신하기 위한 CAAC 복호화 모듈로 이루어진다. LPD 모듈은 복호화 된 모드 정보(ACELP, wLPT256, wLPT512, wLPT1024)에 따라 ACELP 혹은 wLPT 복호화를 수행한다.

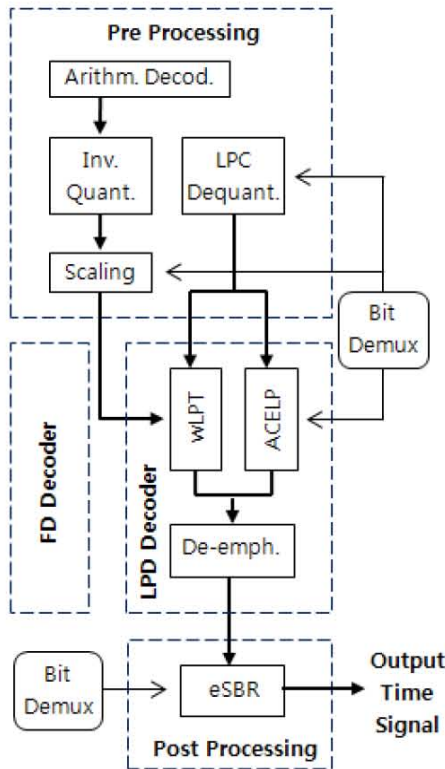


그림 1. LPD 모드를 이용하는 USAC 복호화기 블록도

USAC의 LPD 모듈은 1024샘플의 슈퍼프레임 내에서 256샘플의 서브프레임 단위로 동작하며, LPC 복호화기는 서브프레임의 모드 정보에 따라 16차 LPC 필터 계수를 복원한다. 비트스트림에서 추출되는 양자화된 Line Spectral Frequency (LSF)를 First-stage Approximation과 Algebraic Vector Quantization 복호화기를 거쳐서 역 양자화 하고, interpolation을 통해 각각의 서브프레임에 해당하는 LPC 계수로 변환한다. 복원된 LPC 계수는 ACELP의 합성 필터나 wLPT의 spectrum 이득 계산에 사용된다.

가. LPD Decoder : ACELP

ACELP 모듈은 그림 2와 같이 AMR-WB+와 동일하다. Fixed codebook은 수신된 codebook 정보를 통해 pulse의 이득과 위치를 결정하는 반면, Adaptive codebook은 수신된 pitch 정보로부터 pitch lag를 계산하고 역 Long Term Prediction (LTP) 필터를 거쳐 합성한다. 최종 excitation은 각각의 code vector와 이득으로부터 복원하는데,

복원된 excitation 신호를 LPC 합성필터에 통과시킴으로써 시간영역 합성 신호를 만들어 낼 수 있다.

고정 소수점 구현 과정에서 code vector와 각각의 이득은 고정된 16비트 Q-format을 사용하고, excitation과 시간영역 합성 신호는 동적 범위를 고려하여 16비트 가변 Q-format을 사용한다. 가변 Q-format이란 가상 소수점의 위치를 변수화 하는 방법으로, 추가적인 메모리가 소요되는 단점이 있으나 넓은 동적 범위를 갖는 신호를 효율적으로 제어할 수 있다는 장점이 있다.

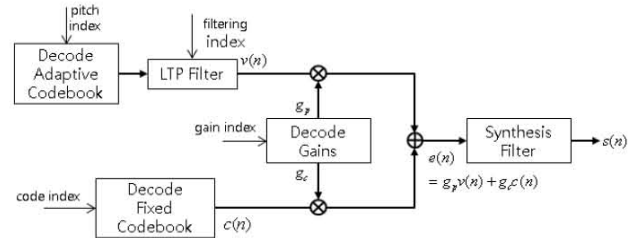


그림 2. ACELP 복호화기의 블록도

나. LPD Decoder : wLPT

USAC의 wLPT 모듈은 AMR-WB+의 Transform Code excitation (TCX)와 차이가 있다. 기존 TCX 모듈의 Discrete Fourier Transform (DCT) 변환을 Modified Discrete Cosine Transform (MDCT)로 대체 하여 주파수 응답을 항상 시키고 프레임간의 부드러운 전이를 가능하게 한다. 또한 비트 효율을 높이기 위하여 기존의 Lattice Vector Quantizer 대신 FD 모드와 동일한 CAAC 모듈을 이용한다. 그림 3은 wLPT 복호화기의 블록도이다. CAAC 복호화기로부터 수신된 spectral 계수에 compont noise를 적용하여 잔여신호의 스펙트럼을 만들고, Adaptive Low Frequency De-emphasis 이득과 LPC 필터 이득을 이용하여 스펙트럼의 레벨을 조절한다. 이와 같이 재합성된 스펙트럼을 Inverse MDCT 변환하면 시간영역 신호를 얻을 수 있다.

변수들의 동적 범위를 고려하여, 재합성된 스펙트럼 계수와 시간영역 신호는 가변 Q-format을 적용한다. Inverse MDCT는 Pre-twiddle과 Post-twiddle을 이용한 Radix-2 IFFT 알고리즘으로 구현하였으며, wLPT 모드의 서브프레임 길이에 따라 256, 512, 1024 샘플 단위로 수행한다 [8].

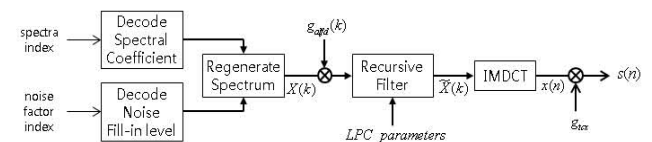


그림 3. wLPT 복호화기의 블록도

4. 성능 평가

본 장에서는 고정소수점 USAC 복호화기의 성능을 분석한다. 테스트 신호는 USAC 평가지침 (Evaluation guideline)에서 12개의 모노 신호를 이용하였다 [9]. 12개의 테스트 신호는 음성 신호 4개, 음악 신호

호 4개, 음악과 음성이 섞인 신호 4개로 구성되어 있다. 각각의 신호는 12kbps, 16kbps, 20kbps, 24kbps의 비트레이트로 부호화, 복호화 되었으며 비트스트림은 USAC의 참조 부호화기를 이용하였다. 또한 비트스트림은 LPD 모듈의 성능을 직접적으로 측정하기 위하여 강제적으로 LPD 모드만 동작하도록 구성하였다. 성능평가는 객관적 평가와 복잡도로 측정한다.

가. Objective Quality : Segmental FFER

고정소수점으로 구현한 USAC LPD 복호화기의 성능을 정량적으로 측정하기 위하여 다음과 같은 FFER (Floating-to-Fixed Error Ratio)를 이용한다.

$$SegFFER = \frac{10}{L} \sum_{l=0}^{L-1} \log_{10} \frac{\sum_{n=0}^{N-1} |s[n,l]|^2}{\sum_{n=0}^{N-1} |d[n,l]|^2} \quad eq.1$$

L은 프레임 개수, N은 프레임 길이, $s[n,l]$ 은 부동소수점 USAC LPD로 복호화 된 신호, $d[n,l]$ 은 고정소수점과 부동소수점 복호화기로부터 얻은 두 신호의 차이를 나타낸다. 표 1은 각각의 비트레이트에서 측정된 12개의 테스트 신호의 FFER이다. 12개의 테스트 신호 모두 30dB 이상의 우수한 FFER이 측정되었다.

표 1. Segmental FFER

SegFFER(dB)	Speech	Music	Mixed
12kbps mono	33.71	35.59	34.03
16kbps mono	33.21	36.29	37.34
20kbps mono	33.93	37.71	38.61
24kbps mono	35.26	34.34	38.41

나. Complexity : WMOPS

ITU-T 고정소수점 표준 라이브러리는 DSP를 구현하기 위한 basic operator를 제공한다 [6]. 각각의 basic operator에는 Weighted Million Operations Per Second (WMOPS)를 측정하기 위한 weight 값이 설정되어 있으며, basic operator를 호출할 때 마다 각각의 weight값을 누적하여 복잡도를 계산한다.

표 2는 각각의 비트레이트에 해당하는 WMOPS의 평균값과 최고치이다. ACELP 모드와 wLPT모드의 WMOPS를 개별적으로 측정하기 위하여 강제적으로 ACELP 모드와 혹은 wLPT 모드만 작동하는 비트스트림을 이용하였다. wLPT는 wLPT256, wLPT512, wLPT1024 중 복잡도가 가장 높은 wLPT256 환경에서 복잡도를 측정하였다.

24kbps에서 ACELP와 wLPT의 최고 복잡도는 각각 10.623와 17.269의 준수한 WMOPS를 갖는다. 모든 비트레이트에서 ACELP에 비해 wLPT가 복잡도가 높게 나오는 경향이 있는데, 이는 wLPT의 inverse MDCT변환과 CAAC 복호화 모듈의 영향으로 볼 수 있다.

표 2. WMOPS

Bitrate	Decoding mode	Average	Worst
12kbps	ACELP	5.382	7.425
	wLPT256	9.620	12.650
16kbps	ACELP	6.094	8.924
	wLPT256	10.809	14.519
20kbps	ACELP	6.793	9.275
	wLPT256	12.098	16.009
24kbps	ACELP	7.397	10.623
	wLPT256	12.845	17.269

5. 결론

본 논문에서는 MPEG-D Unified Speech and Audio Coding의 Linear Prediction Domain 복호화기를 고정소수점 알고리즘으로 구현하고 그 성능을 제시하였다. 제안된 고정소수점 알고리즘은 기존의 부동소수점 USAC 복호화기와 동일한 성능 결과가 나타났으며, 복잡도 측면에서 준수한 WMOPS 값이 측정되었다. 추후 연구에서는 USAC의 Frequency domain 복호화기 모듈과 eSBR 후처리 모듈을 고정소수점 알고리즘을 구현 할 것이며, USAC 서비스화를 위한 DSP 구현도 함께 이루어 질 것이다.

6. 참고 문헌

[1] 3GPP, "Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband(AMR-WB+) codec; Transcoding functions," 3GPP TS 26.290, 2004.

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

[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] ISO/IEC SC29 WG11 N12232, "USAC Verification Test Report," 97th MPEG meeting, Torino, Italy, July 2011.

[5] 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.

[6] International Telecommunication Union, "Software Tools for Speech and Audio Coding Standards," ITU-T, Recommendation G.191, Geneva, Switzerland, 1994.

[7] M. Dietz, L. Liljeryd, K. Kjolring, and O. Kunz, "Spectral band replication, a novel approach in audio coding," in Proc. 112th AES Convention, Munich, Germany, May 2002.

[8] Y. C. Hou and S. D. You, "Implementation of IMDCT for MPEG2/4 AAC on 16 bit Fixed-point Digital Signal Processors," IEEE Asia Pacific Conference on Circuit and Systems, December 2004

[9] ISO/IEC SC29 WG11 N9638, "Evaluation Guidelines for Unified Speech and Audio Proposals," 83rd MPEG meeting, Antalya, Turkey, January 2008.