

복소 공간큐를 활용한 다채널 오디오 코딩 기술

백승권, 임우택, 이태진
한국전자통신연구원

{skbeack, wtlim, tjlee}@etri.re.kr

Complex Spatial Cue based Channel Audio Coding

Beack, Seungkwon Lim, Wootae Lee, Taejin

Electronics and Telecommunications Research Institute (ETRI)

요약

본 논문에서는 복소(complex) 공간큐를 활용한 다채널 오디오 부호화 기술을 제안한다. 복소 공간큐 방식의 다채널 오디오 부호화 기술은 시간영역에서 수행된다. 시간영역의 오디오 채널 신호를 복소 데이터로 변환하여 각 오디오 채널 간의 상관관계를 복소 공간큐로 표현하고, 이를 활용하여 채널 부호화를 수행하기 위한 오디오 채널 신호를 생성한다. 참조 기술로는 최고 성능의 오디오 코덱인 USAC의 예측 부호화 방식의 다채널 오디오 부호화 기술과 비교하여 정보량 감축 측면에 있어서 평균 2.24 dB 이상의 높은 SNR을 나타냄을 관측할 수 있었다.

1. 서론

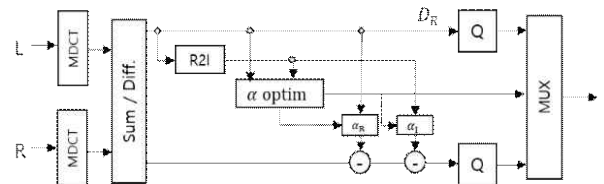
다채널 오디오 부호화 기술은 오디오 콘텐츠가 하나 이상의 채널로 구성되었을 때 오디오 채널 간의 상관관계를 이용하여 부호화 효율을 높이기 위한 기술이다. 따라서 다채널 오디오 부호화 기술은 종전의 오디오 부호화 기술의 전/후처리 모듈로 적용된다 [1]. 다채널 오디오 부호화 기술은 MPEG 오디오 표준 기술의 진화에 따라 발전되어 왔다 [2, 3, 4]. 가장 높은 전송효율을 달성한 PS(Parametric Stereo) 부호화 기술은 스테레오 오디오 신호에 대한 부호화 효율을 최대 50%까지 제공하는 기술로 채널 간의 상관성을 공간 큐(spatial cue)로 표현하여 압축 전송하는 기술이다 [2]. PS 기술의 성능을 개선하고 그 기능을 확장한 기술인 MPS(MPEG Surround) 기술은 현재 최신 오디오 코덱 기술인 USAC(Unified Speech and Audio Coding)과 MPEG-H 3D Audio 기술에도 적용되어 활용되고 있는 기술이다 [2, 3]. MPS 기술의 가장 큰 특징은 PS 기술을 두 개 채널 이상에도 적용 가능하도록 확장하였고 고압축 모드에서 발생하는 음질 열화 문제를 새로운 decorrelator 기술을 적용하여 개선하였다 [3]. PS와 MPS는 주로 고압축 다채널 오디오 부호화 기술로 활용되는 반면, M/S 코딩 방식이나, 예측 기반 다채널 부호화 기술은 고품질 다채널 오디오 부호화 기술로 활용된다 [5, 6, 7].

본 논문에서는 고품질 다채널 오디오 부호화 기술을 위한 복소 공간 큐 기반 다채널 오디오 부호화 기술을 제안한다. 복소 공간큐는 시간영역에서 예측되는 공간큐로, 복소 공간큐를 활용하여 능동적 주요 채널 신호 Mid 신호와 부 채널 신호 Side 신호를 생성한다. 제안 방식은 종전의 주파수 영역의 예측 방식과 달리 시간영역 오디오 채널 신호를 분석하여 추출된다. 종전의 기술 대비 그 효용성 검증은 각 기술 별로 생성하는 부 채널의 정보량이 얼마나 줄었는지를 SNR 수치로 측정하여 판별한다 [6]. SNR이 클수록 부 성분의 정보량이 줄게 되며 이에 할당되는 비

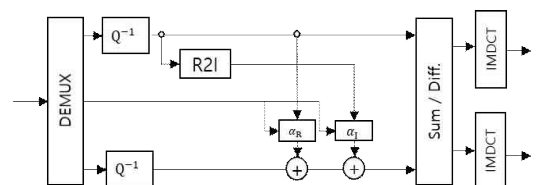
트량을 줄일 수 있어 부호화 효율을 높일 수 있다. 본 논문에서는 스테레오 오디오 신호를 기준으로 제안 기술을 설명하고 검증하였으며 비교 시스템으로는 최고 성능의 USAC의 예측 기반 스테레오 오디오 부호화 기술과 성능을 비교하였다.

2. 예측 기반 USAC 스테레오 부호화 기술

예측 기반 USAC 스테레오 코딩 기술 [6]은 MDCT(Modified Discrete Cosine Transform) 영역에서 수행된다. 그림 1은 이에 대한 인코딩과 디코딩 과정을 그림으로 나타낸 것이다. 인코딩 과정을 살펴보면 스테레오 채널을 MDCT 변환 후, 종전의 M/S 코딩 방식을 수행하기 위하여 'sum/diff' 연산을 수행한다.



(a) 예측 기반 스테레오 인코더



(b) 예측 기반 스테레오 디코더

그림 1. 예측 기반 USAC 스테레오 부호화 기술 블록도

MDCT는 실수부만 존재하므로, 예측 파라미터를 추출하기 위해 허수부를 생성하기 위하여 'R2I (Real to Imaginary)' 블록을 수행한다. 이는 신호의 주파수 영역에서의 선형예측 계수를 실수부 허수부에 대해서 각각 추출함으로써, 부 채널 신호의 정보량을 보다 더 줄이기 위함으로 해석된다. 상대적으로 부 채널의 에너지가 작아진다면, 즉 정보량이 적어지면 양자화 시 주 성분인 D_R 에 보다 효과적으로 비트를 할당하여 양자화를 수행할 수 있다. 이러한 시도는 단순 'sum/diff' 연산을 통해 수행되던 M/S 스테레오 코딩 방식보다 능동적인 예측 계수를 프레임마다 추출하여 능동적 sum/diff 연산을 수행함으로써 보다 적극적으로 부 채널 신호인 side 신호와 주 채널 신호인 mid 신호 간의 정보량 격차를 벌릴 수 있으며, 이러한 차이는 비트 할당 과정에서 부호화 효율을 효과적으로 높일 수 있다.

3. 시간영역 복소 공간큐 스테레오 코딩 기술

상대적으로 비트율이 높은 64kbps 이상의 압축률에서는 예측 기반 스테레오 코딩 기술이 활용되며 가장 최신의 기술은 2장에서 설명한 예측 기반 USAC 스테레오 코딩방식이다. 앞서 살펴본 바와 같이 예측 기반 USAC 스테레오 코딩 기술의 주요 목적은 부 채널 신호인 side 신호의 정보량 및 에너지를 예측 계수를 활용하여 효과적으로 최대한 줄이는 것이다. 따라서 M/S 방식에 기초한 스테레오 부호화 기술은 부 채널 신호의 정보량과 에너지를 낮출 수 있다면 다양한 방법이 시도될 수 있겠다.

본 논문에서 제안하는 방식은 시간 영역에서 복소 공간 큐를 활용하여 부 채널 신호의 파워 및 정보량을 낮추는 방식이다. 주파수 영역이 아닌 시간 축에서 수행이 가능한 이유는 부 채널 신호의 파워는 주파수 성분을 고려하지 않더라도 시간영역에서 충분히 감축될 수 있기 때문이다. 또한 시간 영역의 데이터는 연속적이므로, 임의의 예측 계수를 활용하더라도 보간법 등이 용이하게 적용될 수 있는 장점이 있다. 단점으로는 시간 축에서 스테레오 부호화 기술이 수행될 경우, 블록 단위로 처리시 연속성에 위배되는 왜곡이 발생하지 않도록 후처리 과정이 필요하다는 것이다.

본 논문에서 제안하는 방식은 예측 계수로 오디오 복소 공간 큐를 활용한다. 대표적인 오디오 공간 큐는 CLD(Channel Level Difference), IPD(Inter-channel Phase Difference), ICC(Inter-channel Coherence Correlation) 등이 있다. 제안 방식에서 활용하는 공간 큐는 시간영역에서 IPD와 ICC 파라미터를 새롭게 제안하고 활용한다. 이를 추출하기 위하여 시간영역 신호를 복소 영역으로 변환해야 하는데, 이때 Hilbert transform을 활용하여 신호를 복소수의 분석 신호(analytic signal)로 변환한다 [8].

$$x_a(n) = x(n) + jx_h(n) \quad (1)$$

복소수의 분석 신호로부터 IPD와 ICC 파라미터를 추출할 수 있으며, 종전의 주파수 영역의 파라메트릭 부호화 방식과는 달리 시간영역에서 서브프레임 단위로 추출된다. 시간영역의 서브프레임은 오버랩 구간 없이 IPD의 경우 5 msec 단위로 추출된다. IPD는 4bit로 양자화되며 양자화 인덱스에 대한 엔트로피(entropy) 비트를 측정하면 0.8 kbps의

정보량이 발생한다. 만일 ICC 파라미터까지 전송한다면 1.5 kbps 정도의 부가정보 전송이 필요하다. 이는 USAC의 예측 기반 스테레오 부호화 톨의 부가 정보량보다 2 kbps 정도 낮은 비트율이다 [6].

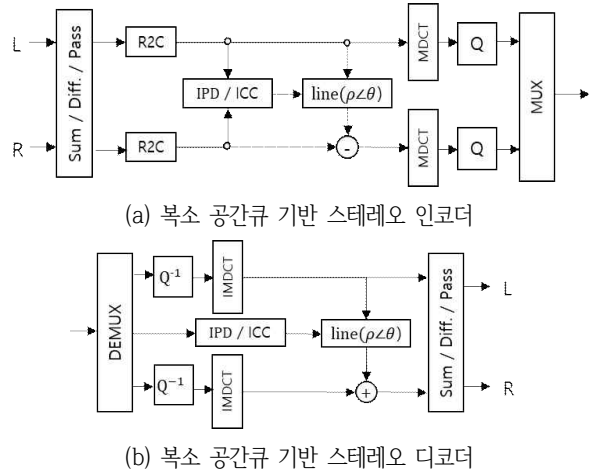


그림 2. 복소 공간큐를 활용한 스테레오 부호화 기술 블록도

그림 2는 제안하는 복소 공간큐를 활용한 다채널 오디오 부호화기의 블록도이다. 인코딩 과정을 살펴보면, 입력 시간영역 오디오 채널 신호는 'sum/diff/pass' 연산을 먼저 수행한다. 여기서 추가적인 'pass' 연산은 L/R이 그대로 활용되는 경우이다. 각각 변환된 시간영역 오디오 채널 신호는 'R2C' 블록을 통해 수식 1과 같이 복소수 오디오 데이터로 변환된다. 각 채널의 복소수 오디오 데이터로부터 복소 공간큐를 추출하고 공간큐로부터 채널 간의 위상 차 θ 값과 채널 상관도에 기인하는 이득 값 ρ 를 계산한다. 연산자 line(*)는 보간법을 수행하는 함수로 이전 프레임의 복소 공간큐 계수와 현재 프레임의 계수 간의 ρ, θ 에 대해서 선형 보간법을 수행한다. 디코딩 과정은 인코딩 과정의 역 과정을 수행한다. 그림 3은 제안 기술을 적용하여 생성한 부 채널 신호의 파형을 나타낸 것이다. 활용된 콘텐츠는 통상적인 스테레오 부호화가 용이하지 않은 critical 테스트 아이템으로, 단순한 M/S 연산으로는 부 채널의 정보량을 줄일 수 없다. 반면에 제안 기술을 활용할 경우, 부 채널의 파워가 효과적으로 감소되고 있음을 확인할 수 있다.

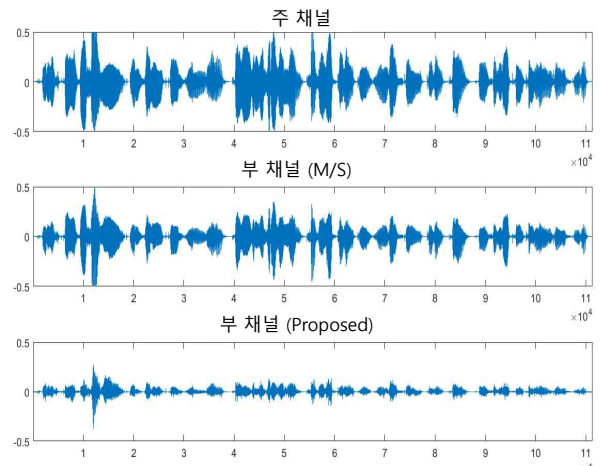


그림 3. 제안 기술을 적용하여 생성한 부 채널 신호의 파형 비교

4. 실험결과

본 논문에서 제안된 방법의 성능을 검증하기 위하여 객관적 평가를 수행하였다. 객관적 평가는 M/S 방식 기준으로 생성되는 부 채널 신호와 제안 방식으로 생성되는 부 채널 신호의 파워 값의 차이를 SNR 값으로 표현하였다. 즉, SNR 값이 클수록 제안 방식의 부 채널 신호의 파워 값이 M/S 방식의 부 채널보다 상대적으로 낮음을 나타내며, 이는 부채널의 정보량을 보다 효과적으로 감축시킨 것으로 M/S 방식보다 부호화 효율을 높일 수 있을 것으로 판단할 수 있다. 종전 USAC의 예측 기반 기술과 성능을 비교하기 위하여 동일한 스테레오 오디오 콘텐츠를 활용하여 실험을 수행하였으며, 모든 콘텐츠는 MPEG Audio 그룹에서 평가를 위해 제공된 오디오 콘텐츠를 활용하였다 [6]. 실험 결과는 그림 4와 같다. 여기서 ‘usac’ 시스템은 USAC의 예측 기반 스테레오 부호화 방식의 성능을 측정하기 위하여 anchor 시스템인 M/S의 부 채널과 예측 기반 스테레오 부호화 방식의 부 채널의 파워를 비교하여 산출한 SNR 수치를 나타내고, ‘t-angle’ 시스템은 제안 기술에서 IPD만을 활용하여 생성한 부채널과 비교한 SNR 결과이며, ‘t-complex’는 IPD와 ICC를 모두 활용한 스테레오 부호화 기술에 대한 부채널의 파워를 M/S의 부 채널과 비교한 SNR 성능을 나타낸다. ‘usac’의 성능은 참조 논문의 데이터를 기반으로 근사화 되어 추출된 데이터로 정확한 수치는 아니나, 논문에서 제시한 성능치를 충분히 반영한 근사 값으로 참조하였다 [6]. ‘usac’의 예측 기반 스테레오 부호화의 경우 3 kbps 이상의 부가정보를 활용하는 반면에, ‘t-angle’은 0.8 kbps, ‘t-complex’는 1.5 kbps의 정보량을 활용한다. 평가 결과에서 관측할 수 있는 것처럼, ‘t-angle’이 ‘usac’보다 다소 성능이 떨어지나 유사한 결과를 나타냄을 관측할 수 있었으며, 본 논문에서 제안한 복소 공간큐를 모두 활용한 ‘t-complex’ 시스템은 종전 ‘usac’ 기술보다 높은 효율의 SNR 성능을 나타냄을 확인할 수 있다.

5. 결론

본 논문에서는 다채널 오디오 부호화 기술에 대해서 소개하고 고품질 다채널 오디오 압축을 위한 스테레오 오디오 부호화 기술을 제안하였다. 제안된 시간영역 복소 공간큐 방식의 오디오 부호화 기술의 성능은 종전의 경쟁 기술 대비 효율적인 부 채널 오디오 신호를 생성할 수 있음을 확인할 수 있었으며, 이는 최종 효율적 오디오 부호화를 수행함으로써 음질의 개선을 달성할 수 있을 것으로 예상된다.

감사의 글

본 연구는 한국전자통신연구원 연구운영비지원사업의 일환으로 수행되었음. [22ZH1200, 초실감 입체공간 미디어·콘텐츠 원천기술 연구]

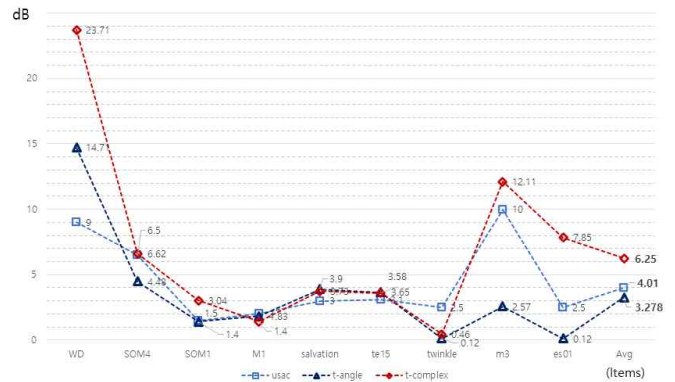


그림 4. 각 스테레오 부호화 방식에 따라 생성된 부 채널과 M/S 방식의 부 채널을 비교한 SNR 성능

참고문헌

- [1] M. Bosi and E. Goldberg, Introduction to Digital Audio Coding and Standards. Norwell, MA: Kluwer, 2002.
- [2] H. Purnhagen: "Low Complexity Parametric Stereo Coding in MPEG-4", 7th International Conference on Audio Effects (DAFX-04), Naples, Italy, October 2004.
- [3] J. Herre et al., "MPEG Surround—The ISO/MPEG Standard for Efficient and Compatible Multichannel Audio Coding," Journal of the Audio Engineering Society, vol. 56, no. 11, pp. 932-955, 2008.
- [4] J. Herre, J. Hilpert, A. Kuntz, and J. Plogsties, "MPEG-H audio—the new standard for universal spatial/3D audio coding," in Proc. 137th AES Conv., Los Angeles, CA, USA, 2014.
- [5] J. Herre et al., "Extending the MPEG-4 AAC codec by perceptual noise substitution." Audio Engineering Society Convention 104. Audio Engineering Society, 1998.
- [6] C. R. Helmrich et al., "Efficient transform coding of two-channel audio signals by means of complex-valued stereo prediction," in Proc. IEEE ICASSP-11, Prague, pp. 497-500, May 2011.
- [7] M. Bosi, et al. "ISO/IEC MPEG-2 advanced audio coding." Journal of the Audio engineering society 45.10 (1997): 789-814.
- [8] Cizek, "Discrete Hilbert transform," IEEE Trans. Audio Electroacoust., vol. 18, no. 4, pp. 340-343, Dec. 1970.