

오디오 채널 신호의 압축을 위한 공간 큐의 효율적 표현 방법

*백승권 김민제 이태진 장대영 강경욱

한국전자통신연구원

*skbeack@etri.re.kr

Efficient Representation method of Spatial cues for audio coding

* Beack, Seungkwon Kim, Minje Lee, Taejin Jang, Daeyoung Kang, Kyeongok

Electronics and Telecommunications Research Institute

요약

본 논문은 공간영역에서의 오디오 채널 신호의 압축 방법이 있어서, 공간 파라미터의 효율적인 표현 방법을 제안하려 한다. 대상이 되는 공간 파라미터는 인간청각의 ILD(Interaural Level Difference) 인지와 관련한 공간 파라미터에 관한 것으로 ICLD(Inter-Channel Level Difference) 파라미터의 표현방법 관한 것이다. 본 논문의 목적은, ICLD의 통계적 특성을 분석하고 이에 충실한 표현방법을 제안함으로써, 양자화 시 기존 표현 방법보다 왜곡을 개선시킴으로써 복원된 오디오 신호의 충실도를 높이는 것을 목적으로 한다. 따라서 본 논문에서는, 새로운 ICLD 표현 방법을 소개하고 이에 대한 이론적 통계적 근거를 제시하며, 실험결과로서 기존 방법과 비교된 왜곡 측정(distortion measure) 결과를 제시하여 제안된 방법의 우수성을 입증한다.

1 서론

오디오 신호를 부호화 하는데 있어서 가장 큰 목적은 신호의 잉여성분(redundancy)을 최소화 하고 남아있는 잔여성분 및 대표 성분으로부터 원 신호를 복원하는 것이다. 잉여성분을 최소화 하는 방법에 있어서 기본적으로 수행할 수 있는 것이 시간적(temporal) 접근방법으로, 시간영역 프레임 단위 신호를 분석하고 상관성에 기반하여 잉여성분을 제거하는 것이다[1][2]. 이때 주로, 시간영역 신호를 주파수영역으로 변환하여 분석을 수행하는데, 여기서 대표적으로 적용될 수 있는 방법이 청각심리음향모델을 적용함으로써, 인간이 인지하지 못하는 신호들을 잉여성분으로 간주하고 제거시킴으로써 원신호를 압축할 수 있는 기술이다[1].

최근 들어, 오디오 신호는 고품질 미디어 서비스를 위하여 스테레오 신호를 넘어 다채널 레이아웃(layout)을 갖는 오디오 콘텐츠에 대한 요구 및 소비가 증가하고 있다. 이는 오디오 신호에 대한 대역폭이 채널 수에 비례해서 증가한다는 것과, 기존 시간적 접근방식에 의거한 압축방법으로는 압축효율을 높이는 데 한계가 있음을 의미한다. 공간큐 기반 다채널 오디오 부호화 기술(SAC: Spatial Audio Coding)은 이러한 다채널 오디오 신호의 압축효율을 보다 극대화 시키기 위해 최근에 새롭게 소개된 기술이다[3][4][5]. 이는 인간 청각의 인지적 특성을 공간적인 특성까지 고려하여 다채널 신호를 파라미터화할 수 있으며 이를 통하여 채널 수를 줄임으로써 원신호를 압축할 수 있는 기술이다.

본 논문에서는 이러한 SAC 기술에 있어서 공간큐(spatial cue)로 활용되는 파라미터의 성능을 개선하고자 한다. 이는 공간큐를 보다 효율적으로 표현할 수 있는 방법을 제안함으로써 전송을 위한 양자화 과정에서 양자화 에러를 개선하고, 결과적으로 보다 강인한 공간 큐 파라미터를 통한 원신호의 복원을 가능하게 하기 위함이다.

2 공간 큐 기반 오디오 코덱의 개요

가. 공간 큐(Spatial cue)

공간적으로 사람이 인지하는 음원의 특성은 사람이 두 귀를 이용한다는 것에 기인하여 결정된다. 즉, 다수의 음원이 존재한다 하더라도, 결국 사람의 두 귀를 통해 인지되므로, 모든 음원으로 부터 발생하는 오디오 신호는 인간에게 전달될 때, 청각의 인지적 특성을 고려하여 두 채널의 binaural 신호로 표현 가능하다[5]. 여기서, 두 귀를 이용하여 공간상의 음원의 위치를 구별할 수 있는데, 그 구별을 위한 특징 인자로서, ILD (Interaural Level Difference)와 ITD (Interaural Time Difference)가 대표적 인자이다. 즉, 두 귀가 인지하는 신호의 레벨 차이, 시간 지연 차이가 그 음원의 위치 및 특징을 인간이 인지하는 범위 내에서 나타나게 할 수 있다. 또 하나의 중요 인자로 공간감을 결정하는 IC(Interaural Correlation) 인자를 들 수 있다. 이는 서로 다른 두 음원이 동일한 레벨 차로 전파될 지라도, 상호간의 신호의 상관성 정도에 따라 인지되는 공간감이 달라지는 현상을 표현할 수 있는 파라미터이다.

이러한 주요 인자를 바탕으로 SAC 코덱에 사용되는 대표적 공간 큐는 다음과 같이 정의된다.

- ICLD (Inter Channel Level Difference): 상호 채널(음원)간의 소리의 레벨차
- ICTD (Inter Channel Time Difference): 상호 채널간의 소리의 지연차
- ICC (Inter Channel Coherence): 채널간의 상호 상관성 정도

이는, 두귀로 인지되는 레벨차, 지연차, 공간감은 결국 상호 음원(채널)간의 레벨차, 지연차, 상관도로 매핑 될 수 있음을 의미한다[5]. 따라서 채널 오디오 신호가 음원이 될 때, 각 채널간의

레벨차, 지연차 및 상관도를 분석함으로써, ICLD 및 ICTD 등을 예측할 수 있게 된다.

나. SAC 기본 구조

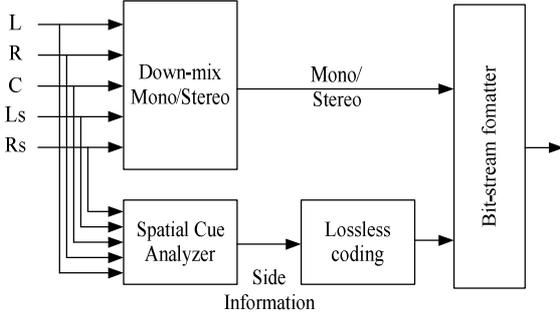


그림 1. SAC 기반 다채널 오디오 부호화기

SAC 기반의 다채널 오디오 코더의 부호화기는 그림 1 과 같이 간략하게 나타낼 수 있다.

다채널 오디오 음원은 다운믹스되며, 이때 각 채널 신호를 분석하여 공간 큐를 추출한다. 다운믹스된 오디오 신호는 모노/스테레오로 기존 오디오 채널을 통해 전송될 수 있도록 하며, 부가 데이터 영역으로 공간 큐들에 대한 정보가 부가정보(side information)으로써 전송된다. 부가정보는 다운믹스 오디오 신호의 부호화된 비트스트림의 잉여영역(ancillary region)에 전송될 수 있으며, 그 데이터량이 채널당 3kbps 정도 소요된다.

결국, SAC 기반 다채널 오디오 복호화기는 이러한 공간 큐를 이용하여 전송된 모노/스테레오 신호로부터 멀티채널 신호를 복원한다.

3. 기존 CLD 표현방법의 문제점

부가정보에서 가장 큰 부분을 차지하는 공간 큐는 ICLD 이다. ICTD 은 헤드폰 재생이 아닐 경우, 다채널 재생 공간상에서 ITD 로 매핑되기가 쉽지 않으며, ICC 의 경우 기본적으로 하나의 대표 값을 주파수 밴드 별로 전송하고, 대역폭에 여유가 있을 때 그 수를 증가시킨다. 따라서 기본적 채널 신호의 복원은 ICLD 로부터 기인한다고 할 수 있으며, 보다 정확한 ICLD 값은 복원되는 오디오 신호의 충실도를 보장할 수 있게 된다.

ICLD 는 주파수영역에서 서브밴드 별로 추출되며 서브밴드의 수는 대개 20~28 개의 ERB(Equivalent Rectangular Bandwidth) 밴드를 활용한다[1]. ICLD 도 전송을 위해서는 양자화 되어야 하며 무손실 부호화 과정을 거치게 된다. 무손실 부호화 방법으로는 일반적으로 허프만(Huffman coding) 방식이 적용된다. 여기서 양자화 방법은 ICLD 의 확률분포를 분석하고 확률분포에 따라 그 양자화기를 설계한다. 일반적으로 ICLD 는 0 dB 를 중심으로 대칭적인 가우시안(Gaussian) 분포를 갖는 것으로 보고 비선형 양자화기를 적용한다[2][3]. 허프만 방식의 무손실 부호화 방법은 양자화기 인덱스의 확률 분포가 라플라시안(Laplacian)과 같이 가파른(sharp)형태 를 갖게 될 경우 그 압축효율이 높다.

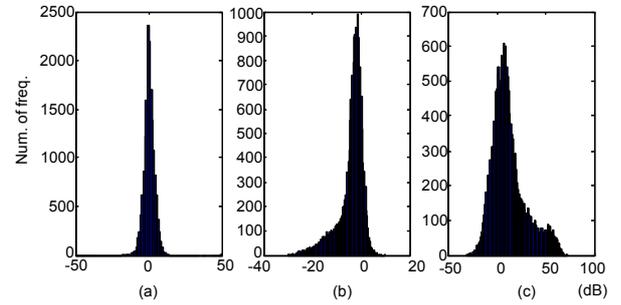


그림 2. 멀티채널 콘텐츠에 대한 기존 ICLD 의 histograms:
(a) applause, (b) classical music, (c) popular music.

그림 2 는 일반 5.1 채널 콘텐츠에서 추출한 ICLD 의 확률분포함수(probability density function)이다. 그림 2 에서 알 수 있듯이, 일부 주변(ambient) 음원에 대해서는 ICLD 가 대칭적 분포를 가지고 있으나, 대부분의 오디오 콘텐츠에서 분석된 ICLD 값은 대칭적이지 않다. 이는 곧 적절한 양자화기를 설계하기가 쉽지 않고, 입력 콘텐츠의 공간 큐의 확률분포함수에 종속적으로 양자화기가 설계 및 적용되어야만 그 양자화 효율을 극대화 시킬 수 있음을 의미한다. 따라서, 고정되었고 한정된 양자화 레벨을 가지고 효율적으로 양자화를 수행하기 위해서는 모든 다채널 콘텐츠에 대한 ICLD 의 확률분포가 일정하여야 한다. 또한 무손실 부호화기의 압축이득을 극대화 하기 위해서는 그 확률분포의 경사도(sharpness)가 높아야 한다.

4. 제안하는 ICLD 표현 방법

양자화 효율을 개선하기 위해서는 우선적으로 ICLD 의 확률분포함수가 가파른 경사도를 가질 수 있도록 설계되어야 한다. 본 논문에서는 H-ICLD (Half region - ICLD)를 소개한다. 이는 ICLD 를 추출할 때, 최대 파워를 갖는 채널 정보를 활용하여 ICLD 를 추출함으로써, 추출된 ICLD 는 항상 음의 값을 가지며, 그 경사도 또한 보장받을 수 있다. 먼저 기준채널에 대한 인덱스 정보 rf_b 는 다음과 같이 얻을 수 있다.

$$rf_b = \arg \max_{1 \leq c \leq C} P_{c,b}, \quad (1)$$

여기서 $P_{c,b}$ 는 각 채널 c 의 b 번째 서브밴드에 대한 파워로 다음과 같이 정의한다.

$$P_{c,b} = \sum_{n=A_b}^{A_{b+1}-1} |S_{c,n}|^2. \quad (2)$$

여기서 A_b 는 서브밴드의 경계정보이다. ICLD 값은 다음과 같이 rf_b 를 활용하여 추출된다.

$$\Delta L_{c,b}^h = 10 \log_{10} \left(\frac{P_{c,b}}{P_{rf_b,b} + \alpha} \right), \quad rf_b > c \quad (3)$$

$$\Delta L_{c-1,b}^h = 10 \log_{10} \left(\frac{P_{c,b}}{P_{rf_b,b} + \alpha} \right), \quad rf_b < c.$$

$\Delta L_{c,b}^h$ 는 새롭게 추출된 ICLD 값이며, 기준채널은 수학적 식 1 에 의해서 결정된다. 이러한 ICLD 값은 양자화 되고, 허프만 부호화 된 후 복호화기에서 다음의 과정을 거쳐 채널 신호의 정보를 복원한다.

먼저 각 채널의 서브밴드의 파워 이득은 수학적 4.5 를 이용하여 추출될 수 있다.

$$G_{ref,b} = \frac{1}{\sqrt{1 + \sum_{i=1}^{C-1} 10^{\Delta L_{i,b}^h/10}}}. \quad (4)$$

$$G_{c,b} = 10^{\Delta L_{c,b}^h/10} G_{rfb,b}, \quad c < rfb \quad (5)$$

$$G_{c,b} = 10^{\Delta L_{c-1,b}^h/10} G_{rfb,b}, \quad c > rfb$$

여기서, $G_{c,b}$ 채널 c 의 서브밴드 b 에서의 파워이득이다. 이러한 파워이득은 서브밴드 별로, 전송된 다운믹스 모노/스테레오 오디오 신호에 적용되며, 이로부터 각 채널의 신호의 파워를 얻는다.

$$U_{c,k} = G_{c,b} S'_k, \quad A_b \leq k \leq A_{b+1} - 1, \quad (6)$$

여기서 $U_{c,k}$ 는 복원된 채널 c 의 주파수 계수 k 성분을 나타내며, S'_k 는 전송되는 다운믹스 신호의 k 번째 주파수 성분을 나타낸다.

이렇게 추출된 H-ICLD 값은 항상 음의 값을 가지므로, 단방향(single-side) 양자화 테이블을 이용할 수 있으며, rfb 를 활용함으로써 0 dB 를 기준으로 높은 경사도의 확률분포 를 얻을 수 있다. 단방향 양자화 테이블을 이용함으로써 비트전송률을 줄일 수 있을 것으로 기대되나, 추가적인 rfb 를 전송해야 하므로 전체 비트전송률은 유사할 것이다. rfb 는 run-length-coding 방식으로 전송된다. 표 1 은 H-ICLD 와 기존의 ICLD 의 양자화 인덱스 표준편차를 나타낸다. 여기서, H-ICLD 가 보다 작은 표준편차를 갖기 때문에 양자화 테이블의 설계 시 보다 작은 dynamic range 안에서 그 값을 표현할 수 있음을 알 수 있다.

표 1. 기존 ICLD vs. H-ICLD 의 표준편차.

Contents Cue(dB)	Ambience sound	Classical music	Popular music
C-ICLD	3.653	5.334	17.148
H-ICLD	2.673	4.806	12.030

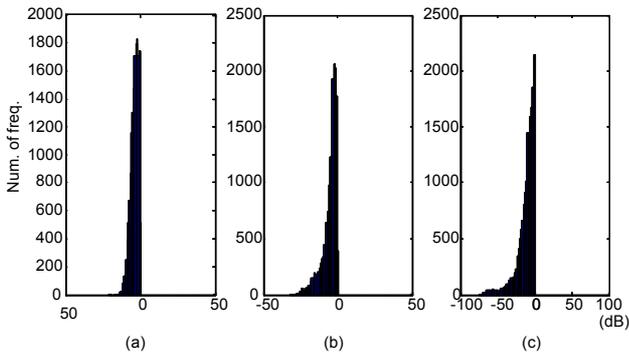


그림 3. 멀티채널 콘텐츠에 대한 H-ICLD 의 histograms:
(a) applause, (b) classical music, (c) popular music.

5. 실험결과

가. 주파수 왜곡을 측정

객관적 실험지표로 주파수 왜곡을 측정하였다. 왜곡을 Kullback-Leibler 거리측정방법을 사용하였다[12].

$$D_{SKL} = \int (P(\omega) - Q(\omega)) \log \frac{P(\omega)}{Q(\omega)} d\omega. \quad (9)$$

여기서 $P(\omega)$ 는 원신호의 주파수 파워를, $Q(\omega)$ 는 복호화된 신호의 주파수 파워를 나타낸다. 왜곡을 프레임별로 측정한 뒤 전체 평균값을 취하였다. 실험콘텐츠는 MPEG 오디오 표준화 그룹에서 사용한 MPEG Surround 테스트 아이템 11 개를 사용하였다. ΔL_{max} 는 양자화 수행시 dynamic range 로, 18, 24, 30dB 로 변화하여 양자화 에러를 측정하였다. 양자화 레벨은 7, 15, 31 로 변화하여 측정하였으며, 기존 방법에 대해서는 대칭적 midread 양자화기를 적용하였고, 제안한 방법에 대해서는 음의 값의 비대칭적 midread 양자화기를 적용하였다. 그림 3 은 실험 결과이다. 실험 결과에서 알 수 있듯이, 모든 양자화 레벨, dynamic range 에 대해서 제안한 H-ICLD 에 의해 합성된 오디오 신호가 원신호 대비 평균 주파수 왜곡이 적음을 알 수 있다.

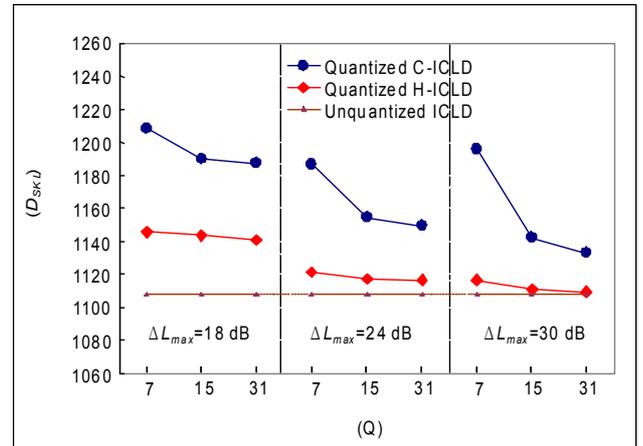


그림 3. 측정된 D_{SKL} 기반의 성능

나. 비트전송률 측정

기존 ICLD 및 제안한 H-ICLD 의 비트전송률을 각각 측정하였다. 기존 ICLD 의 비트전송률은, 11 개의 멀티채널 콘텐츠를 기반으로 설계한 허프만 테이블을 이용하는 무손실 코딩방식을 적용했을 경우의 비트전송률 결과를 조사하였다. 본 허프만 테이블은 기존 ICLD 의 통계적 특성을 기반으로 하여 설계된 것으로, 그 성능이 최적화 되어 있다고 볼 수 있다. 제안한 방법의 비트전송률은, 우선 H-ICLD 의 양자화 인덱스는 기존 ICLD 와 같은 방법으로 양자화 테이블을 이용하였으며, 이전 값과의 차이(difference)를 구하여 그 인덱스에 대한 비트전송률을 구한 것이다. 기준채널 인덱스에 대해서는 run-length coding 을 수행하였다. 수행결과, 제안한 방법의 H-ICLD 의 비트전송률이 기존 ICLD 의 같은 레벨에서 비트전송률보다 증가함을 알 수 있다. 그러나 H-ICLD 는 단면의 대칭적 양자화기를 사용하므로, 실제 그 비트전송률은 한단계 위의 기존 ICLD 의 비트전송률과 비교해야 분해능이 동일하다. 예를 들어, H-ICLD 7Q 는 ICLD 의 15Q 와 같은 양자화 분해능을 갖는다고 볼 수 있으며, 이를 비교하는 것이 적합하다. 따라서, 표 2 에서 알 수 있듯이, 같은 양자화 레벨에서는 제안한 방법의 비트전송률이 크나, 동일한 분해능을 갖는

양자화기 레벨을 고려해 보면, 오히려 3~5kbps 정도의 비트전송률 감소가 있음을 알 수 있다.

[12] E. Klabbers and R. Veldhuis, "Reducing Audible Spectral Discontinuities," *IEEE Trans. On. Speech and Audio Proc.* vol. 9, no. 1, JAN. 2001, pp. 39 – 51.

표 2. 기존ICLD vs. H-ICLD 의 비트레이트

Q(kbps)	7Q	15Q	31Q	64Q
C-ICLD	12.47	15.62	24.86	31.05
H-ICLD	15.77	18.61	27.72	34.05

6. 결론

본 논문에서는 SAC 기반의 다채널 오디오 코딩을 위한 공간 큐의 표현방법에 대해서 연구하였다. 특히 공간 큐로 가장 큰 부분을 차지하고 있는 ICLD의 표현방법 및 양자화 방법에 대해서 언급하였다. 실험결과 기존의 ICLD 표현 방법 및 양자화 방법보다 양자화 왜곡이 적음을 실험결과를 통해 알 수 있었다. 향후 H-ICLD 를 위한 비선형 양자화기를 보다 면밀히 설계하고 이를 기반으로 하는 허프만 테이블을 구축함으로써, 비트전송률에 대한 성능 개선도 추가적으로 기대할 수 있을 것이다. 또한 주관적 음질 평가를 통해 음질 향상 여부도 조사할 예정이다.

7. References

- [1] M. Bosi, K. Brandenburg, S. Quackenbush, L. Fielder, K. Akagiri, H. Fuchs, M. Dietz, J. Herre, G. Davidson, and Y. Oikawa, "ISO/IEC MPEG-2 advanced audio coding," *J. Audio Eng. Soc.*, vol. 45, pp. 789-812, Oct. 1997.
- [2] J. D. Johnston, S. R. Quackenbush, J. Herre and B. Grill, "Review of MPEG-4 general audio coding," *Multimedia, Systems, Standards, and Networks*, pp. 131-155, A. Puri and T. Chen (ed.), Marcel Dekker, Inc. 2000.
- [3] J. Herre, H. Purnhagen, J. Breebaart, C. Faller, S. Disch, K. Kjörling, E. Schuijers, J. Hilpert, and F. Myburg, "The reference model architecture for MPEG Spatial Audio Coding," *AES 118th Convention*, Barcelona, Spain, May 2005.
- [4] J. Breebaar, J. Herre, C. Faller, J. Rödén, F. Myburg, S. Disch, H. Purnhagen, G. Hotho, M. Neusinger, K. Kjörling, W. Oomen, "MPEG spatial audio coding / MPEG Surround: overview and current status," *AES 119th Convention*, New York, NY, Oct. 2005
- [5] C. Faller and F. Baumgarte, "Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression," in Preprint 112th, Conv. Aud. Eng. Soc., May 2002.
- [6] C. Faller and F. Baumgarte, "Binaural Cue Coding Applied to Audio Compression with Flexible Rendering," in Preprint 115th, Conv. Aud. Eng. Soc., Oct. 2002.
- [7] C. Faller and F. Baumgarte, "Binaural Cue Coding-Part II: Schemes and Application," *IEEE Trans. on Speech and Audio Proc.*, vol. 11. no. 6, Nov. 2003.
- [8] C. Faller, "Parametric Coding of Spatial Audio," *Proc. 7th Int. Conf. on Digital Audio Effects*, Naples, Italy, Oct. 2004, pp. 151-156.
- [9] ISO/IEC JTC1/SC29/WG11 (MPEG), "Call for Proposal on Spatial Audio Coding," Document N6455, March 2004.
- [10] ISO/IEC JTC1/SC29/WG11 (MPEG), "Procedures for the Evaluation of Spatial Audio Coding Systems," Document N6691, Redmond, July 2004.
- [11] E. Zwicker and H. Fastl, *Psychoacoustics*, Spreinger, Verlag, Berlin Heidelberg, 1999.