

Binaural Cue Coding 기술을 이용한 오디오 코덱 구현

서정일*, 강경옥*, 이병화**, 한민수**
*한국전자통신연구원, **한국정보통신대학교

The Development of audio codec using binaural cue coding technologies

Jeongil Seo*, Kyeongok Kang*, Byonghwa Lee**, Minsoo Hahn**
*ETRI, **ICU

요약

낮은 대역폭에서 다채널 다객체 오디오 신호를 전송하기 위해 새롭게 제안된 Spatial Audio Coding 기술은 멀티채널 오디오 신호를 다운믹싱하고 나머지 채널은 음향공간상의 위치정보를 나타내는 파라미터들로 압축하여 표현하는 파라메트릭 압축 방식이다. 본 논문에서는 Spatial Audio Coding 기술중의 하나인 BCC 기술을 이용하여 스테레오 오디오 코덱을 구현하고, 주관듣기평가 실험을 통하여 AAC와 비슷한 성능을 나타내면서도 높은 압축율을 얻을 수 있음을 확인하였다.

I. 서론

마스킹 효과와 같은 심리음향모델을 이용한 저각오디오 압축방식은 디지털방송과 DVD 등의 매체에서 멀티채널 오디오 신호를 압축하는데 적용되고 있다. 사용자의 다양한 요구에 부응하기 위하여 5.1채널로 대표되던 멀티채널 재생환경이 6.1채널 7.1채널로 확장되고 있으며, 향후 파면음장합성(Wave Field Synthesis)과 같이 수백개의 채널을 가지는 재생환경으로 발전해가고 있다. 또한 다수의 오디오 객체들로 오디오 장면을 구성하고 오디오 객체들간의 상호연동을 이용하여 대화형의 오디오 장면을 제공하는 규격(MPEG, SMIL 등)이 제공됨에 따라 재생 채널뿐만 아니라 오디오 장면을 구성하는 오디오 객체의 수도 증가하고 있다.

3차원 오디오 신호와 같이 다채널 다객체의 환경에서는 오디오 신호의 채널수나 객체수에 비례하여 오디오

신호를 전송하기 위한 대역폭이 커지게 된다. 이와 같은 오디오 신호를 하나 또는 두개의 채널로 다운믹싱하여 압축하고 객체와 채널에 대한 정보는 별도의 파라미터로 표현하여 전송하면 압축효율을 극대화 할 수 있다. 복호화기에서는 양 귀간의 시간차와 크기차인 ITD와 ILD 등과 같이 사람이 음원의 위치와 거리를 결정하는 파라미터를 Spatial Cue 파라미터라 정의하고, 채널간의 Spatial Cue를 각 채널간의 정보로 추출하고 이를 다운믹싱하여 압축한 비트열에 추가하여 전송하고, 복호화기에서는 다운믹싱된 비트열로부터 다운믹싱된 신호를 복구한 후 Spatial Cue 파라미터를 이용하여 남은 채널 신호들을 복구하게 된다.

Faller와 Baumgarte에 의해 제안된 BCC(Binaural Cue Coding) 기술은 Spatial Cue 파라미터를 Binaural Cue 파라미터로 명하고, ITD와 ILD를 각각 ICTD와 ICLD로 대체하여 Binaural Cue 파라미터로 사용하였다. 본 논문에서는 BCC 코덱의 개요를 먼저 설명하고, ICLD만을 이용하여 구현한 결과를 기술한다.

II. Binaural Cue Coding

2.1 음원의 공간 이미지를 표현하기 위한 파라미터

양 귀사이의 신호 크기차(ILD, Interaural Level Difference)와 시간 지연차(ITD, Interaural Time Difference)는 수평면에서 사운드 소스의 방위각을 결정하며, HRTF(Head Related Transfer Function)는 고막에서 사운드 소스까지의 전달 특성을 특정한 길이의 FIR 필터의 시간응답으로 나타낸 것이다. 이와 같은 파라미

터들은 헤드폰 재생에서 양귀에서의 방향신호를 직접적으로 결정하며, 스피커 재생에서는 간접적으로 영향을 미치게 된다.

2.2 BCC를 이용한 공간이미지의 렌더링

BCC의 개념은 공간 지각정보와 오디오 콘텐츠를 위한 정보를 분리하는 것이다. BCC는 공간 오디오 신호를 모노 오디오 신호와 BCC 파라미터로 나타낸다. BCC 합성기에서 생성되는 모노 오디오 신호는 다채널의 공간 오디오 신호로부터 전달된 모든 신호를 다운믹싱한 신호이며, BCC 파라미터들은 다채널 공간 오디오 신호의 생성에 필요한 공간 지각정보들(예를들면 공간신호를 생성하기 위해 필요한 시간축과 주파수축에서의 특징 파라미터들)이다.

아래의 그림1과 같이 BCC에서의 다채널 신호 추출과정은 시간/주파수축 상에서 소스들이 겹치지 않는 지역에 존재할 때 가능하다. 보통 시간/주파수축 상에서 소스들이 겹쳐지지만 BCC에서는 이런 겹침이 일어나는 것을 무시하고, 시간/주파수 평면상에서 특정 지역에서 가장 지배적인(dominant) 소스 신호에 대해서만 고려하여 처리한다.

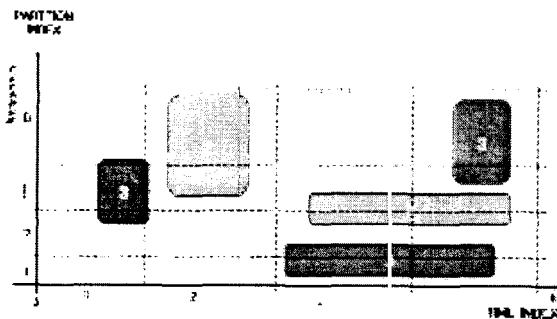


그림 1. 시간/주파수 평면상에서의 동원 소스의 분포

2.3. BCC 기반 오디오 코덱 시스템

BCC를 이용한 오디오 코덱 시스템의 블록도는 아래 그림 2와 같다. 그림 2에서 보이는 것과 같이 BCC 오디오 코덱 시스템은 음향 공간 파라미터들을 추출하고 처리하기 위한 BCC 해석기와 합성기, 모노 오디오 신호를 부호화하는 오디오 부호화기와 복호화기 부분으로 나뉜다.

멀티채널 오디오 신호가 들어오면 BCC 해석기는 음향공간 파라미터들을 뽑아내어, BCC 합성기에 보내고, 합성기는 오디오 복호화기에서 전달받은 모노 오디오

신호와 BCC 음향공간 파라미터들을 조합해 입력 신호와 유사한 오디오 신호를 만들어 낸다. 일반적인 오디오 인코더와 디코더에 대해서는 기존의 방법과 동일하므로, 본 논문에서는 BCC 해석기와 합성기를 중심으로 기술한다.

2.3.1. BCC 해석기

BCC 해석기의 처리과정은 그림 3과 같이 BCC해석기에 입력된 멀티채널 오디오 신호는 시간-주파수 변환기를 거쳐 각각 주파수 형태로 변환되어 주파수 영역의 파티션으로 분할되고, 이 신호는 BCC 해석기의 ICLD(Inter-channel level difference) 추정기와 ICTD(Inter-channel time difference) 추정기를 거쳐 각 파티션에 대한 ICLD와 ICTD를 생성하게 된다. 이렇게 생성된 ICLD와 ICTD는 양자화 및 부호화기를 거쳐 부호화됨으로써 BCC 음향 공간 파라미터를 생성한다.

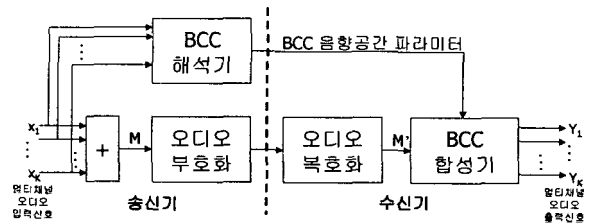


그림 2. BCC 기반 오디오 코덱 시스템

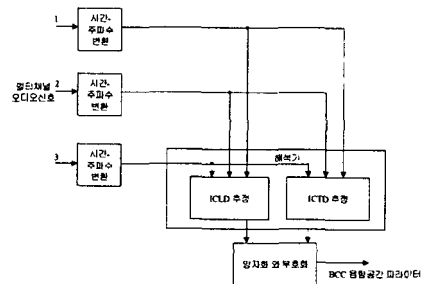


그림 3. BCC 해석기의 블록도

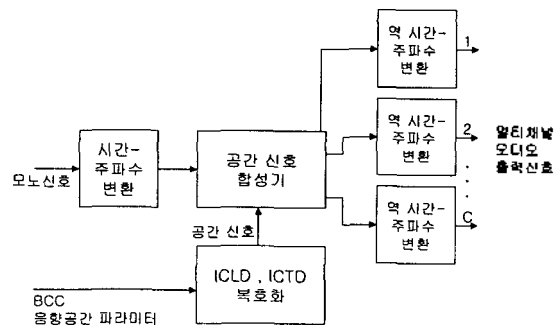


그림 4. BCC 합성기의 블록도

2.3.2. BCC 합성기

BCC 합성기의 처리과정은 그림 4와 같다. 모노 오디오 신호는 시간-주파수 변환기를 거쳐 BCC 해석기에서와 같이 주파수 영역의 파티션으로 분할되며, BCC 파라미터를 입력 받은 ICLD/ICTD 부호화기에서는 양자화된 BCC 파라미터를 복호화 한다. 이렇게 주파수 영역의 파티션으로 변환된 모노 오디오 신호와 ICLD, ICTD는 공간신호 합성기를 거치면서 재조합 된다. 이것을 다시 각각의 역 시간-주파수 변환기에서 시간 영역으로 변환함으로써 최종적인 멀티채널 오디오 신호를 생성한다

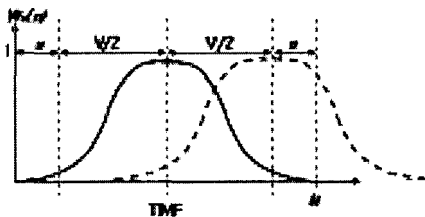
다채널 다객체 환경에서 압축효율을 극대화할 수 있는 방안으로 제시된 BCC 알고리즘은 스테레오 환경에서의 MPEG-4 High Quality 파라메트릭 부호화방식에도 일부 적용되었으며, 멀티채널에 적용하기 위한 방안이 "Spatial Audio Coding"이란 이름으로 MPEG에서 표준화가 진행중이다.

III. BCC를 이용한 오디오 코덱 구현

3.1 시간-주파수 변환

입력을 1024포인트씩 읽어 들어서 프레임단위로 처리를 한다. 읽어 들인 입력에 윈도우 함수를 곱해하는데 그 함수는 아래와 같다.

$$w_v(n) = \begin{cases} 0 & , 0 \leq n < \sigma \text{ or } N - \sigma \leq n < N \\ \sin^2\left(-\frac{(n - \sigma)\pi}{V}\right) & , \sigma \leq n < \sigma + V \end{cases}$$



위와 같은 윈도우를 곱하는데 V/2만큼씩 옮겨가면서 처리를 하게 된다. σ 는 시간 축 상에서의 알리어싱을 방지하기 위해서 0을 채워 넣는 구간이다. 우리의 실험에서는 $V=896$, $\sigma=64$ 로 하였다. 입력을 위의 윈도우 함수로 곱한 후 $N=1024$ 포인트의 퓨리에 변환을 한다. 위와 같은 과정은 수신단에서 채널 합 신호에도 동일하게 적용한다.

3.2 BCC 파라미터

본 논문에서는 스테레오 신호를 대상으로 하였고

BCC 파라미터는 간단한 구현을 위해 ITLD를 제외한 ICLD만을 사용 하였다.

3.2.1 BCC 해석기

BCC 해석기에서 ICLD 파라미터를 구하는데 그 과정은 다음과 같다. 1024포인트로 퓨리에 변환 한 것을 $B_1 \sim B_{19}$ ({3, 6, 9, 13, 18, 24, 32, 42, 54, 69, 88, 111, 140, 176, 220, 275, 343, 427, 513})의 19개 밴드로 나누어서 계산했다. spectral 계수는 대칭적으로 $(N/2)+1$ 개만 고려해주면 된다. $S_{c,n}$ 을 각 채널(c)의 n번째 spectral 계수라 하면 채널(c)의 b번째 밴드의 에너지는 다음과 같이 구해진다.

$$E_{c,b} = \sum_{n=N_b}^{N_{b+1}-1} |S_{c,n}|^2$$

각 채널의 밴드별 에너지를 구한 후 기준 채널에 대한 각 채널의 ICLD를 구하게 된다. 보통 기준 채널은 채널 1로 한다. 본 논문은 스테레오 신호를 대상으로 하기 때문에 기준채널은 left채널로 했다.

$$I_{c-1,n} = 10 \log_{10} \left(\frac{E_{c,n}}{E_{1,n}} \right)$$

이와 같이 구해진 ICLD를 모든 채널을 단순히 더한 모노 신호와 함께 전송한다.

3.2.2 BCC 복호기

수신단에서 받은 모노 신호의 spectral 계수를 s_i 라 하면 각 채널 신호의 spectral 계수 $s_{i,n}$ 은 다음과 같이 구해진다.

$$\tilde{s}_{i,n} = L_{i,n} s_n$$

$L_{i,n}$ 은 ICLD를 반영하는 요소로 기준채널과의 에너지 비를 나타낸다.

$$L_{i,n} = 10^{(I_{i,n} - I_{1,n})/20}$$

$I_{1,n}$ 은 채널 합인 모노 신호에서의 기준 채널의 비를 나타낸다.

$$I_{1,n} = \frac{1}{\sqrt{1 + \sum_{i=2}^N 10^{(I_{i,n} - I_{1,n})/10}}}$$

따라서 복호기에서는 모노신호를 퓨리에 변환 한 것을 각 밴드별로 ICLD를 위와 같이 적용해서 다시 역 퓨리에 변환 하여 각 채널별로 저장하여 재생해 준다.

IV. 실험 및 결과

4.1 실험

실험은 9개의 콘텐츠에 대해 각각 (1)표본화 주파수가 44.1kHz인 소스, (2)MPEG 4 AAC를 이용한 96kbps 스테레오, (3)BCC를 사용하여 채널합이된 모노신호를 MPEG 4 AAC 64kbps로 보내어 BCC로 복원한 파일, (4)2kHz로 bandlimited된 파일에 대해 10명의 MOS 테스트를 실시했다.

테스트는 헤드폰(SENNHEISER HD25-1)을 통해 각 콘텐츠마다 위의 4종류 파일을 랜덤하게 들려주고 음질을 측정하게 하였다. 음질측정은 5등급(1:bad, 2:poor, 3:fair, 4:good, 5:excellent)의 5단계로 측정되었다.

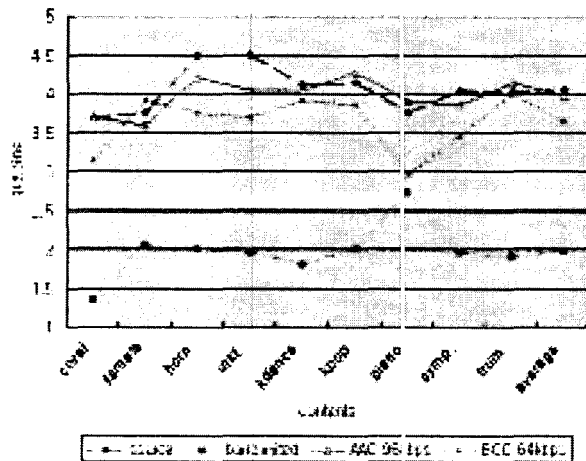


그림 5. MOS 테스트 결과

실험결과는 음질 측정 단계에서 기인차에 따라 값의 변화도가 컸지만 대략적인 그래프의 모습에서 소스파일의 음질이 가장 좋게 나왔다. 그 다음이 MPEG4 AAC 스테레오 96kbps로 된 파일이고 BCC를 사용한 것이 그 다음이었다. 당연한 결과로 2kHz 대역 제한 된 신호가 가장 낮은 음질로 평가되었다. 값의 차이는 약간 있었으나 실제 측정에서 10명의 테스트 참가자들은 2kHz로 대역제한 된 파일을 제외하고 나머지 3가지의 경우에 대해서 음질차이를 크게 인식하지 못하였다.

결론

제한된 대역폭에 다채널/다객체의 오디오 신호를 효과적으로 압축하기 위하여 제안된 기술이 Spatial Audio Coding 의 하나인 BCC 기술을 이용하여 스테레오 오디오 코덱을 구현하고, 주관평가실험을 통하여 우수한 성

능을 보임을 확인하였다. BCC로 대표되는 Spatial Audio Coding 기술은 현재 MPEG에서 멀티채널 오디오 신호에 대한 파라메트릭 압축방식으로 표준화가 진행중이며 5.1채널과 멀티채널 신호를 적은 비트율로 전송하기에 적절한 방식으로 대두되고 있으므로 지속적인 연구가 필요하다.

감사의 글

본 논문은 정보통신부의 연구사업인 "지능형 통합 정보 방송 기술 개발"과제의 일환으로 수행한 결과로서 정보통신부 담당자 및 관련 연구원들의 노력에 감사를 드립니다.

참고문헌

- [1] C. Faller and F. Baumgarte, "Binaural Cue Coding Applied to Stereo and Multi-Channel Audio Compression", in Audio Engineering Society Convention Paper presented at the 112th convention May .10-13. 2002.
- [2] C. Faller and F. Baumgarte, "Efficient representation of spatial audio using perceptual parameterization", in IEEE Workshop Appl. Of Sig. Proc. To Audio and Acoust., Oct . 2001.
- [3] C. Faller and F. Baumgarte, "Binaural Cue Coding Applied to Audio Compression with Flexible Rendering", in Audio Engineering Society Convention Paper 5686 Presented at the 113thConvention. Oct. 5-8. 2002.
- [4] Audio Subgroup, "Draft Call for Proposals on Spatial Audio Coding," N6151, Hawaii, Dec. 2003.
- [5] 남민희, 정은구, 김해광, 서정일, 강경욱, "바이노럴 큐를 이용한 멀티채널 오디오 부호화 방법", 한국음향학회 추계학술발표대회, 2003.