

멀티채널 오디오 재생 시스템에서 가상 음원의 위치 정보를 이용한 압축 재생 방법

은한길, 전상배, 이정석, 성광모, 서정일*
서울대학교, ETRI*

Multi-Channel Audio Coding Method with Virtual Source Location Information

Han-gil Moon

Seoul National Univ. fullmoon@acoustics.snu.ac.kr

요약

본 논문은 방송 및 통신 환경을 이용한 멀티채널 음향 재생 환경에서 다수 객체의 음상 경위를 보다 효과적이고 효율적으로 하기 위한 방법에 관한 것이다. 본 논문에서는 전송되는 정보의 양을 최소화 하면서도 재생되는 음향공간에서는 다수의 음향 객체들이 충실하게 재생되어 자연스러운 음향공간이 개현할 수 있는 방법을 제시하고자 한다. 기존 방법의 경우, 전송 선로를 통해 음원을 압축하여 전송하기 위해서 먼저 멀티채널 신호를 합한 모노신호와 채널 신호사이의 음량차이(ICLD), 시간지연 차이(ICTD), 상관도(ICC)등을 전송하는 양귀단서 신호화 기술(Binaural Cue Coding)을 이용하고 있다. 본 논문에서는 멀티채널 음원을 분석하여, 음원의 가상 위치정보를 벡터적으로 표현하고, 이 위치벡터와 멀티채널 음원을 하나의 모노 음원으로 다운 믹스한 신호를 전송함으로써 전송 효율을 극대화 한 압축 재생 방법을 제시한다.

1. 서론

멀티채널 오디오 신호를 압축하기 위한 연구는 1990년대 중반 MPEG 을 중심으로 이루어졌다. DVD를 중심으로 한 멀티채널 콘텐츠의 폭발적인 증가와 유저들의 멀티채널에 대한 욕구증가 그리고 통신 환경 등을 통한 멀티채널 오디오 서비스의 필요성 증가가 멀티채널 오디오 압축 기술에 대한 연구를 더욱 부추리게 하였다. 이러한 이유로 MPEG에서는 MPEG-2 BC, MPEG-2 AAC, MPEG-4 AAC 등의 멀티채널 오디오 압축 기술이 표준화되어 상업적으로 큰 실효를 거두었으며, Dolby사의 AC-3, DTS 등의 멀티채널 오디오 압축기술이 DVD와 디지털 방송을 위한 오디오

압축기술로 채택되어 사용되고 있다. 최근에서는 BCC (Binaural Cue Coding)로 대표되는 보다 혁신적인 멀티채널 오디오 신호 압축방식에 대한 연구가 활발히 진행되고 있는데(C. Faller, 2002 & 2003; F.Baumgarte, 2001 & 2002). 이를 통해서 보다 적은 정보량을 이용해서 보다 현실감 있는 오디오 데이터의 전달을 꾀하고 있다. BCC는 음향공간에서 사람이 느낄 수 있는 공간감이 주로 양귀효과(Binaural Effect) 때문인 것을 이용하여, 멀티채널 오디오 신호를 압축하는 기술이다. 사람은 특정 음원의 위치를 양귀에 도달하는 소리의 크기차이(Interaural Level Difference)와 시간 지연차이(Interaural Time Difference)를 이용하여 인지한다고 알려져 있으므로, BCC는 멀티채널 오디오 신호를 모노 또는 스테레오 신호로 다운믹스하고, 나머지 채널에 대한 정보는 기준채널과의 신호 크기차이와 시간 지연등의 바이노럴 인지 파라미터(Binaural Cue Parameter)들로 표현한다. 본 논문에서는 멀티채널 음원을 분석하여, 음원의 가상 위치정보를 벡터적으로 표현하고, 이 위치벡터와 멀티채널 음원을 하나의 모노 음원으로 다운 믹스한 신호를 전송함으로써 전송 효율을 극대화 한 압축 재생 방법을 제시한다.

2. BCC (Binaural Cue Coding)

BCC는 멀티채널이 다운믹스(downmix)된 모노 및 스테레오 소스와 각각의 채널의 부가정보를 통하여 본래 채널을 복원하는 다채널 오디오 부호화의 한 방식이다. 여기에서 부가정보란 다채널 신호를 분석하여 추출되는 각각의 채널에 대한 정보로서 후에

복호기에서 다채널 오디오 데이터를 재구성하는데 사용된다.

BCC에서 사용하는 부가정보에는 ICLD(Inter Channel Level Difference), ICTD(Inter Channel Time Difference), ICC(Inter Channel Correlation)가 있다. 이 중에서 ICLD는 각각의 멀티채널 신호와 참조하는 다운믹스 신호 사이의 레벨 차이를 나타내고, ICTD는 특정 음원에 의한 오디오 신호가 시간차를 갖고 채널에 나타나는 경우, 그 시간차에 해당하며, ICC는 멀티채널 신호가 참조하는 다운믹스 신호와 얼마나 상관성이 있는지를 나타낸다.

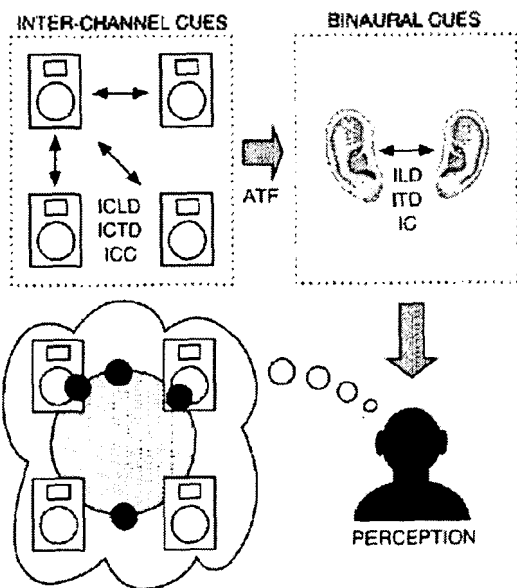


그림 1. BCC 개념과 부가정보

2.1 BCC의 구조

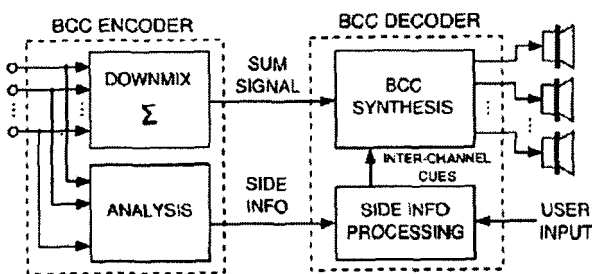


그림 2. BCC의 개념도

그림 2 에서 정의된 것처럼 BCC의 구조는 크게 송신기(Transmitter)와 수신기(Receiver)로 구분할 수 있는데, 송신기에는 부가정보(binaural cue)를 추출해내는 BCC 분석기(Analyzer)가, 수신기에는 다운믹스된 신호와 추출된 부가정보(Binaural cue)를 사용하여 각각의

채널을 다시 재합성하는 BCC 합성기(Synthesizer)블럭이 포함된다.

2.2 부가정보(Binaural Cues)

부가정보는 멀티채널 오디오 신호에 공간감을 부여하는 파라미터들을 말한다. 이 부가정보는 BCC 분석기에서 추출되어 송신되고, BCC 합성기에서 다운믹스된 오디오 신호와 합성된다. 이러한 부가정보에는 ICLD, ICTD, ICC등이 있다. 다음 그림 3.는 2채널(Stereo)의 경우에 각각의 부가정보를 추출하는 방법의 예이다.

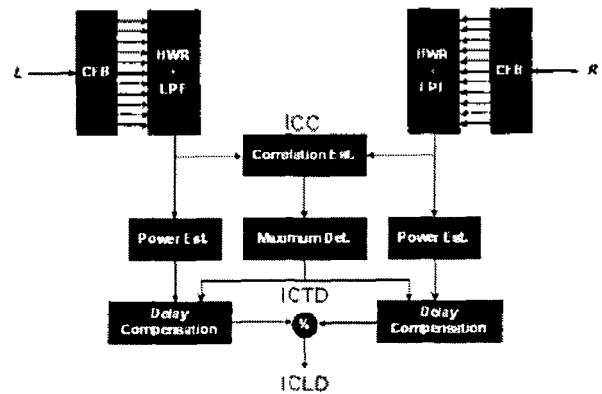


그림 3. 2채널(Stereo) BCC 분석기

3. VSLI(Virtual Source Location Information)

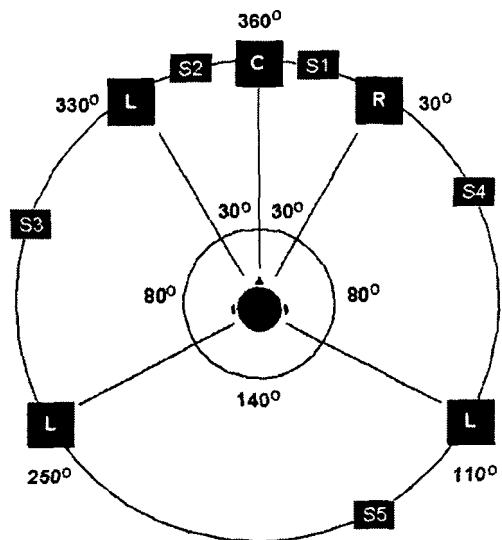


그림 4. BCC의 개념도

가상 음원의 위치정보(Virtual Source Location

Information, VSLI)는 멀티채널 재생환경(ITU-R 권고안)에서 만들어지는 음상의 위치를 벡터적으로 표현한 것이다. 그림 4와 같은 일반적인 5채널 재생 환경에서 생길 수 있는 일반적인 음상의 위치는 S1,S2,S3,S4,S5이다.

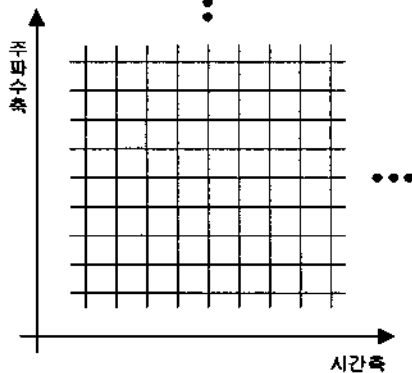


그림 5. ERB 필터뱅크를 이용한 시간/주파수 격자

각각의 음상의 위치는 멀티채널 스피커에서 재생되는 신호를 주파수 분석하여, 인간 청각기관의 주파수 분해능과 같은 ERB(Equivalent Rectangular Band) 필터뱅크(그림 5)로 분해한 후 인접 채널(스피커)들의 벡터합을 계산함으로써 추정할 수 있다. 이렇게 추정된 위치 벡터들은 전송효율과 재생과정시의 편리함, 그리고 전체 채널의 파워를 효과적으로 표현하기 위해 그림 6에서와 같이 Ga(Global angle), LHa(Left Half Plane angle), RHa(Right Half Plane angle), LSa(left subsequent angle), RSa(right subsequent angle)등의 벡터로 변환하여 전송한다.

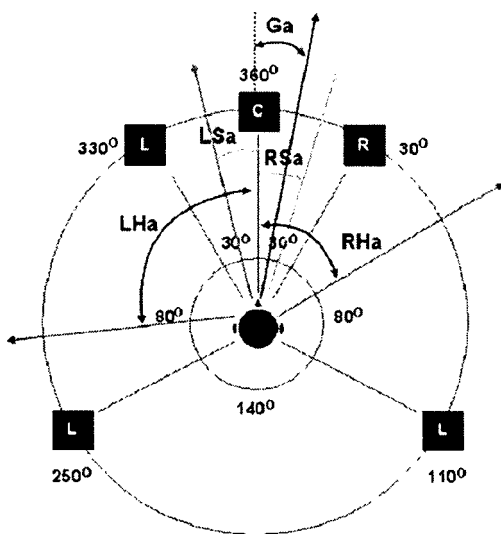


그림 6. VSLI 인자의 벡터 변환

4. VSLI를 이용한 SAC 시스템

3장에서 정의된 VSLI 정보를 이용하여 SAC(Spatial Audio Coding) 시스템을 설계하면, 그림 7-8과 같은 부호화기(Encoder)와 그림 9-10과 같은 복호화기(Decoder)를 만들 수 있다.

부호화기(그림 7)에서는 멀티채널 오디오 데이터를 받아 각 채널의 데이터를 DFT 수행 후 Critical Band를 근사화 한 ERB 필터뱅크로 격자화 한다. 이후 VSLI 분석기(그림 9)에서 격자화된 주파수 정보를 바탕으로 프레임과 밴드별로 에너지 벡터를 추출한다. 이렇게 추출된 프레임/밴드별 에너지 벡터와 CPP 법칙[Pulkki]를 이용하여, 프레임/밴드별 가상음원의 위치를 각도로 환산한 후 양자화 한다. 양자화 된 VSLI는 다운믹스된 오디오 데이터를 AAC로 인코딩한 비트스트림과 함께 전송된다.

복호화기(그림 9)에서는 전송된 VSLI+AAC 비트스트림에서 VSLI를 추출하고, VSLI와 재생단 채널(스피커) 레이아웃 정보를 이용하여, 밴드별 이득값을 계산한다. VSLI 합성기(그림 10)에서는 이 밴드별 이득값과 다운믹스된 오디오 신호의 주파수 데이터를 이용하여, 압축 이전 N 채널의 오디오 스펙트럼을 재구성한다. 이렇게 재구성된 N 채널의 오디오 스펙트럼 자료로부터 최종적인 오디오 데이터가 복원된다.

참고문헌

1. Faller, C. and Baumgarte, F. "Efficient representation of spatial audio using perceptual parametrization," Proc. 2001 IEEE WASPAA, New Paltz, NY, Oct. 2001
2. Baumgarte, F. "A computationally efficient cochlear filter bank for perceptual audio coding," Proc. ICASSP 2001, Salt Lake City, May 2001
3. Baumgarte, F. "A psychoacoustic model for audio coding based on a cochlear filter bank," Proc. 2001 IEEE WASPAA, New Paltz, NY, Oct. 2001
4. Faller, C. and Baumgarte, F. "Binaural cue coding: A novel and efficient representation of spatial audio," Proc. ICASSP 2002, Orlando, Florida, May, 2002

6. Faller, C. and Baumgarte, F. "Estimation of auditory spatial cues for binaural cue coding," Proc. ICASSP 2002, Orlando, Florida, May, 2002
7. V.Pulkki. "Virtual sound source positioning using vector based amplitude panning." Journal of the Audio Engineering Society, 45(6) pp.456-466, June 1997.

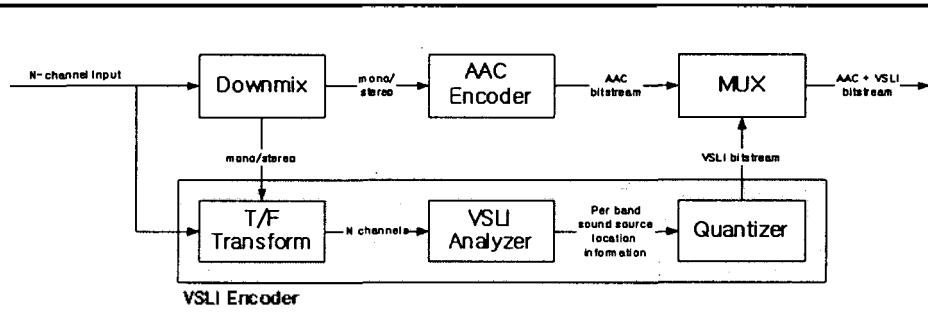


그림 7. VSLI를 이용한 SAC 부호화기

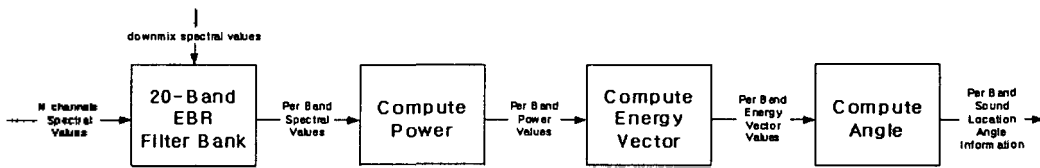


그림 8. VSLI 분석기

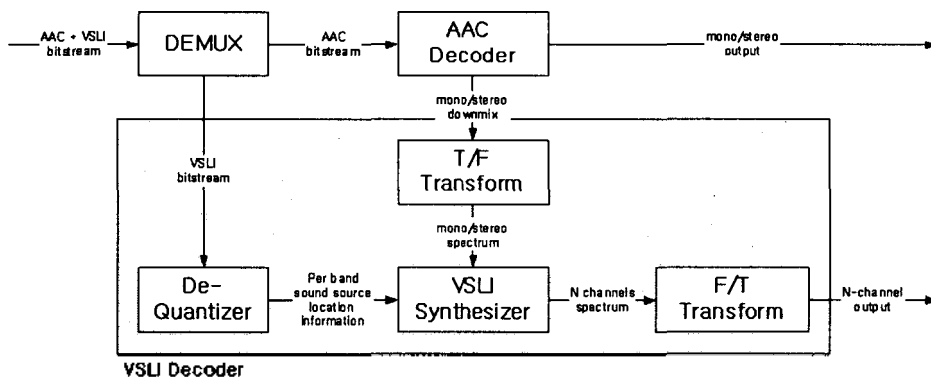


그림 9. VSLI를 이용한 SAC 복호화기

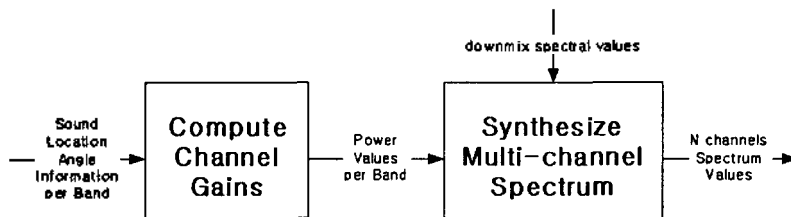


그림 10. VSLI 합성기