

# 채널상관성에 기반한 효율적인 멀티채널 spatial audio coding 방법

이병화\*, 백승권\*, 서정일\*\*, 한민수\*

한국정보통신대학교\*, 한국전자통신연구원\*\* , {bhlee@icu.ac.kr}

## An efficient multichannel spatial audio coding method based on inter channel correlation

Byonghwa Lee\*, Seungkwon Beack\*, Jeongil Seo\*\*, Minsoo Hahn\*

Information and Communications University\*, Electronics and Telecommunications Research Institute\*\*

### 요약

Spatial Audio Coding 방법 중 하나인 Binaural Cue Coding 방법은 다채널 다객체 오디오 신호를 모노나스테레오로 다운 믹스한 신호와 spatial 큐를 전송해 디코더에서 복원하는 기술로 작은 비트 율로 다채널 오디오 신호를 전송 복원해 내는 기술이다.

본 논문은 BCC 코딩 방법에서 채널 상관도를 나타내는 ICC 파라미터에 따라 spatial cue 종류를 달리함으로써 전송되는 부가정보의 비트 율을 줄이는 방법을 제안한다.

### 1. 서론

오디오 신호를 좀 더 작은 비트 율을 가지고 효과적으로 전송하기 위해 Sum/Difference(S/D) 코딩[1], Intensity 코딩[2] 등 많은 연구가 있어왔다. 이러한 방법의 코딩으로 어느 정도 비트 율을 줄일 수는 있지만 멀티 채널 오디오 소스와 같이 채널이 늘어남에 따른 높은 비트 율을 줄이는 방법으로는 부족하다.

최근에 소개된 Spatial Audio Coding 기술 중 하나인 Binaural Cue Coding 방법은 오디오 신호를 모노나스테레오 신호로 다운 믹스한 후 사이드 정보를 이용해 디코더에서 복원해내는 기술로 비트 율을 현저히 줄일 수 있다[3][4][5]. 본 논문에서는 BCC 에서 채널 상관성에 따라 코딩 방법을 달리하는 방법으로 사이드

정보량(side information)을 좀더 줄일 수 있는 방안을 제시하고 있다.

### 2. Binaural Cue Coding (BCC)

BCC의 대략적인 구조는 아래와 같다.

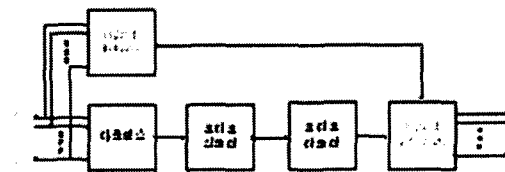


그림 1. BCC의 구조

#### 2.1 Encoding

입력으로 들어온 멀티채널 신호는 그림 1에서 보듯이 모노 신호로 다운믹스 된다[3][6]. BCC 파라미터는 그림 1의 BCC 분석기에서 구해지는데 주파수 영역에서 몇 개의 서브 밴드 (equivalent rectangular bandwidth)로 나누어서 구한다[3][7]. 여기서 구해지는 파라미터는 BCC의 type에 따라 그 종류가 다르다[3][8].

Type I에 사용되는 BCC 파라미터는 소스의 인덱스 정보이다. 한 밴드 내에는 하나의 소스 신호만 있다는 가정으로 입력된 각 소스의 밴드 별 에너지를 비교 후 가장 큰 값을 지니는 소스의 인덱스를 파라미터로 전송한다.

Type II의 경우 사용되는 BCC 파라미터는 ICLD(Inter Channel Level Difference), ICTD(Inter Channel Time Difference)와 ICC(Inter Channel Correlation)이다. ICLD,

ICTD 는 파라미터들은 서브밴드 별로 구하고, ICC 는 가장 큰 에너지를 갖는 두 밴드에 대해서만 ICC 를 구한다[3].

### 2.2 Decoding

Type I 의 경우 BCC 합성부는 매우 간단한 구조를 지닌다. 수신된 모노 신호의 주파수 밴드 신호를 BCC 파라미터인 인덱스 정보에 따라 해당되는 출력 채널의 밴드에 단순히 할당하는 역할을 한다. 이렇게 되면 할당되지 않은 나머지 채널에는 그 밴드의 정보가 비게 되므로 소리의 열화를 초래한다. 이러한 부분은 이후 수신 단에서 사용자가 정한 조건에 맞게 렌더링 과정을 거치면서 해결될 수 있다.

Type II 의 경우는 파라미터로 ICLD, ICTD 를 사용하며, 전송되는 모노신호에 적용하여 다채널의 오디오 원 신호를 만들어 낸다[3].

## 3. Hybrid BCC coding method

본 논문에서 제안하고자 하는 방법은 채널간 상관성에 큐파라미터 추출 전략을 달리함으로써 전송되는 파라미터 정보량을 줄이기 위한 방법이다.

Type II 의 경우 전송되는 ICLD 의 값은 에너지의 비를 나타내는 실수로 할당되는 비트수가 많고 채널이 늘어남에 따라 그에 비례해 전송되는 개수가 증가한다. 이와는 달리 type I 의 경우 사이드 정보가 채널의 인덱스로 정수형의 정보를 가지므로 비트수가 ICLD 의 경우보다 적고 채널이 증가하는 경우에도 가장 큰 에너지 값을 가지는 채널만을 인덱스 정보로 전송하므로 채널 수에 상관없이 항상 하나의 사이드 정보를 가진다. 따라서 채널간 상관도가 높은 경우에는 type II 의 경우를 사용하고 채널간의 상관도가 낮은 경우에는 type I 방법으로 코딩 함으로써 전송되는 사이드 정보량을 줄일 수 있다.

### 3.1 Encoding

전체적인 BCC 의 구조와 파라미터의 종류는 동일하며 채널 상관도에 따라 코딩 방법을 달리한다.

Type II 의 코딩에서 ICC 를 통해 채널간 상관도를 측정하고 측정된 상관도가 낮을 때 좀더 적은 비트

율을 가지는 파라미터인 인덱스 정보를 보냄으로써 사이드 정보의 비트 율을 줄인다. Type I 으로 코딩 타입이 변화 되는 기준은 다음과 같다.

$$ICC_{i_b} < (1 - \text{mean}(ICC_{i_1}^{N_1}, ICC_{i_2}^{N_2} + ICC_{i_3}^{N_3})) \times 0.3 + d(ICC_{i_1}^{N_1}, ICC_{i_2}^{N_2} + ICC_{i_3}^{N_3}) \quad (1)$$

$N_{tot}$  는 전체밴드 수,  $[x]$ 는  $x$  보다 작은 가장 큰 정수, mean, d 는 각각 평균과 분산을 의미한다.

### 3.2 Decoding

ICC 값을 모니터링하여 디코딩 타입을 결정하고 그에 맞는 디코딩 과정을 거친다. 여기서 문제가 되는 것은 type I 으로 디코딩 했을 경우 인덱싱 되지 않은 부분의 밴드에 신호가 할당되지 않아 해당 채널의 빈 스펙트럼으로 인한 음질의 열화이다. 이 문제를 해결하기 위해 짧은 시간 안에서는 신호가 천천히 변화한다는 것과 캐스터네츠와 같은 임펄스성 신호, 잡음과 유사한 신호에 대한 주파수 보상방법을 위해 그림 2 와 같이 시간영역 모델링과 주파수 영역에서의 모델링을 복합적으로 사용하였다.

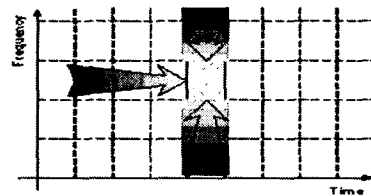


그림 2. 시간과 주파수 복합적으로 보상

#### 3.2.1 시간 축으로의 보상

먼저 시간 축 방향으로 프레임의 비어있는 주파수 영역을 복원한다. 신호는 짧은 시간 내에서 천천히 변한다는 것을 바탕으로 이전 프레임의 다운믹스된 모노 신호와 각 채널의 비가 현재 프레임에서의 에너지 비와 유사할 것이라는 가정으로 시간 축으로의 비어있는 주파수영역을 채우게 된다.

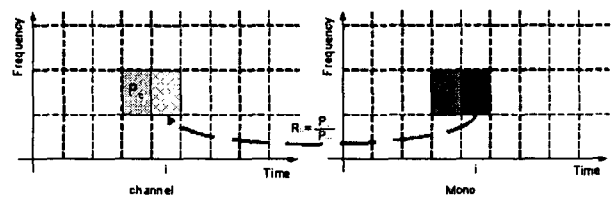


그림 3. 시간 영역에서의 보상

현재 보상하려는 채널에서 이전 프레임의 밴드 파워가  $P_c(i-1,b)$  이고 다운믹스 된 모노 신호의 이전 프레임의 밴드 파워를  $P_m(i-1,b)$  라 하면 시간 축으로의 보상 비율  $R_h(i,b)$  는 아래와 같은 식이 된다. 여기서 소리의 세기가 갈수록 감소하는 것을 나타내기 위해 state 를 두어서  $R_h(i,b)$  에 반영한다. 이전 밴드의 에너지가 정해진 값보다 작으면 state 의 값을 1 씩 증가 시키고 그렇지 않을 경우 0 으로 만들어준다. 수평축으로의 보상 비율은 다음의 식과 같다.

$$R_h(i,b) = (0.99)^{\text{state}} \frac{P_c(i-1,b)}{P_m(i-1,b)} \quad (2)$$

따라서 시간 축으로 보상 되는 주파수 밴드는  $Y_k$  는  $Y_{i,b}^c(k) = U_{i,b}(k) * R_h(i,b)$ ,  $k \in \{A_{b-1}, A_{b-1}+1, \dots, A_b-1\}$  (3) 가 된다. 여기서  $A_b$  는 밴드 내 각 포인트를 나타내고 이 영역은 사람의 인지 특성을 반영한 밴드이다[7].

### 3.2.2 주파수 축으로의 보상

시간 영역으로의 보상 후 주파수 축으로 보상을 한다. 주파수 축에서 보상하고자 하는 영역 앞 뒤 밴드의 다운믹스 된 모노 신호와 채널 간 에너지 비를 계산한다.  $R_v(i,b-1)$  와  $R_v(i,b+1)$  는 각각 이전 밴드의 에너지 비와 다음 밴드의 에너지 비를 나타낸다.

$$R_v(i,b-1) = \frac{P_c(i,b-1)}{P_m(i,b-1)}, \quad R_v(i,b+1) = \frac{P_c(i,b+1)}{P_m(i,b+1)} \quad (4)$$

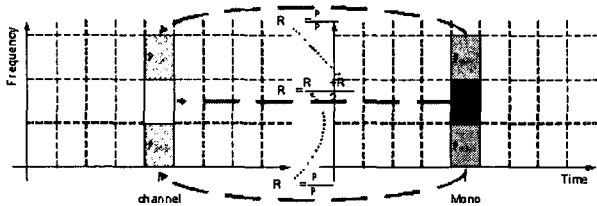


그림 4. 주파수 영역에서의 보상

이 두 밴드의 에너지 비를 평균한 값이 주파수 영역으로의 보상 비( $R_v(i,b)$ )가 된다.

$$R_v(i,b) = \frac{R_v(i,b-1) + R_v(i,b+1)}{2} \quad (5)$$

앞에서 구해진 시간 축으로의 보상비( $R_h(i,b)$ )와 주파수 축으로의 보상비( $R_v(i,b)$ )를 평균한 값( $rP$ )이 최종적으로 다운 믹스된 모노 신호로부터 그 채널의 빈 영역 에너지비가 된다.

$$R_i(i,b) = \frac{R_h(i,b) + R_v(i,b)}{2} \quad (6)$$

여기서 캐스터네츠 소리와 같은 임펄스성 신호에 대해서 위와 같은 보상법을 적용하면 두 가지 보상비율의 평균 과정에서 시간 축으로의 작은 보상 비로 인해 적절한 보상이 이루어 지지 않게 된다. 임펄스성 신호의 스펙트럼 보상법을 위해  $R_v(i,b)$  가  $R_h(i,b)$  에 비해 큰 값을 가질 경우에는 두 값을 평균 하지 않고  $R_v(i,b)$  를 최종 에너지 비로 한다.

$$R_i(i,b) = \begin{cases} R_v(i,b) & , \text{ if } R_v(i,b) > 2R_h(i,b) \\ \frac{R_h(i,b) + R_v(i,b)}{2} & , \text{ if } R_v(i,b) \leq 2R_h(i,b) \end{cases} \quad (7)$$

최종적으로 위에서 구한 스펙트럼의 에너지 비율을 모노 신호에 곱하여 빈 영역의 스펙트럼을 보상해 낸다.

$$Y_{i,b}(k) = U_{i,b}(k) * R_i(i,b), k \in \{A_{b-1}, A_{b-1}+1, \dots, A_b-1\} \quad (8)$$

## 4. 실험 및 결과

실험은 5 가지의 콘텐츠에 대해서 실시하였는데 BCC 와 hybrid 코더 방식에 대한 10 명의 선호도 테스트와 스펙트럼의 Kullback Leibler distance 측정을 하였다.

표 1. 감소된 정보량

	Hybrid rate(%)	Reduced info. no.
Fountain_music	44.7407	52803 (17601)
Glock	5.5354	6141 (2047)
Jackson1	18.4046	17637 (5879)
Poulenc	8.9426	10544 (3518)
Stomp	25.8312	30486 (10162)

표 1 은 hybrid 의 방법을 통해 얻을 수 있는 사이드 정보량을 보여주고 있다. 괄호 안의 수는 hybrid 방식으로 코딩된 밴드 수를 나타내고 이로 인해 줄어드는 사이드 정보의 개수를 보여주고 있다.

표 2 는 원 소스와 각 코딩 방법으로 나온 결과에 대한 채널별 스펙트럼 Kullback Leibler distance 를 측정 한 것이다. Hybrid\_nocom 은 hybrid 방법을 쓸 때 스펙트럼 보상을 하지 않은 값이다. 스펙트럼 보상법을 통해 많은 distance 향상을 있는 것을 알 수 있다.

표 2. Kullback-Leibler distance

	BCC	Hybrid	Hybrid_nocom
Left	73.170005	75.082333	83.026382
Right	69.153935	70.917093	77.859571
Center	63.388749	67.541202	79.42926
Rear left	64.989916	68.998695	76.069841
Rear right	65.547561	68.869933	75.84862
Total	336.250166	351.409256	392.233674

표 3은 테스트들이 더 좋다고 선택한 수를 나타낸다.

표 3. 콘텐츠에 따른 선호를 답한 테스트의 수

	BCC	Hybrid	Same
Fountain_music	2	2	6
Glock	5	1	4
Jackson1	8	1	1
Poulenc	3	3	4
Stomp	8	1	1

Jackson1 과 Stomp 를 제외한 콘텐츠에서는 두 가지의 음질 차를 잘 느끼지 못하는 것으로 나왔다.

그림 5 는 테스트 결과를 평균한 값으로 0 기준으로 양의 값은 BCC 로 코딩된 콘텐츠를 더 선호 하는 것이고 음의 값은 hybrid 를 더 선호한다는 의미이다.

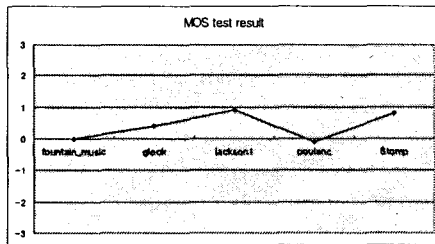


그림 5. 선호도 조사 결과

테스트 참가자들은 전체적으로 BCC 의 음질을 좋게 평가하였으나 음질의 차이를 크게 인식하지는 못하였다. 위와 같은 결과에서 채널 상관성에 따른 코딩 전략을 달리 하는 방법을 통해 큰 음질의 저하 없이 전송되는 사이드 정보를 줄일 수 있음을 알 수 있다.

## 5. 결론

채널의 상관성이 클 경우는 ICLD 를 그냥 사용하고 작을 경우 가장 에너지가 큰 채널의 인덱스만을

보냄으로써 음질의 큰 저하 없이 사이드 정보량을 줄일 수 있다. 인덱스만을 보냈을 때 음질의 열화를 막기 위한 스펙트럼 보상법은 BCC 와의 큰 음질 차이를 느끼지는 못할 만큼의 성능을 보여주었다. 그러나 몇몇 콘텐츠에서는 어느 정도 열화정도가 느껴졌는데 앞으로 이러한 콘텐츠에 대해서도 좋은 결과를 얻을수 있는 적절한 보상법을 찾아 보완해야 할 것이다.

## 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] J. D. Johnston and A. J. Ferreira, "Sum-difference stereo transform coding", in ICASSP-92 Conference Record, 1992, pp. 569-572
- [2] J. Herre, K. Brandenburg, and D. Lederer, "Intensity stereo coding", in Proc. AES 96<sup>th</sup> Convention, Feb. 1994
- [3] C. Faller and F. Baumgarte, "Binaural Cue Coding - Part II: Schemes and Applications", IEEE Trans. Speech and Audio Proc. vol. 11, no. 6, pp. 520-531, Nov.2003
- [4] C. Faller and F. Baumgarte, "Binaural Cue Coding applied to stereo and multi-channel audio compression", in Preprint 112<sup>th</sup> Conv. Aud. Eng. Soc., May 2002
- [5] C. Faller and F. Baumgarte, "Binaural Cue Coding applied to audio compression with flexible rendering." 113<sup>th</sup> AES Conv., Los Angeles, Oct, 2002
- [6] F. Baumgarte and C. Faller, "Audio Coder Enhancement using Scalable Binaural Cue Coding with Equalized Mixing", AES 116<sup>th</sup> convention, May. 2004
- [7] B.R. Glasberg and B. C. J. Moore, "Derivation of auditory filter shapes from notched-noise data", Hear. Res., vol. 47, pp.103-138, 1990
- [8] C. Faller and F. Baumgarte. "Binaural cue coding : A novel and efficient representation of spatial audio",Proc. ICASSP 2002, Orlando, Florida, May 2002