



## Analysis of Spectral Characteristics of Audio Object for Enhancement of Spatial Audio Object Coding

Kwangki Kim\*

*Department of Digital Contents, Korea Nazarene University*

### ABSTRACT

This paper describes a method that can enhance the performance of spatial audio object coding (SAOC) with two-step coding structure to make a multi-object based audio service with users' interaction. In the proposed method, frequency bands and bit-rate of a residual signal can be variably decided by analysing the spectral characteristics of a specific audio object such as vocal object to efficiently handle the residual signal which is used to precisely control the specific audio object. The proposed method divided the frequency band of the specific audio object into five frequency regions including silence according to the spectral characteristics of the specific audio object. From the analysis result, most audio objects including vocal object have about more than 20 percents low frequency region compared with the original fixed frequency band to be 0 to 5.5 kHz. Consequently, the proposed method showed the high bit-rate reduction of about 28.6 percents for test items and about more than 10 percents for audio objects. Especially, since the bit-rate of the vocal object which is mainly used in KARAOKE application scenario can be greatly reduced as about 15 percents, it can be confirmed that the proposed method can enhance the original SAOC with two-step coding structure.

© 2015 KKITS All rights reserved

**KEYWORDS** : Spatial Audio Object Coding, Multi-Objects based Service, Residual Signals, Specific Audio Object, Vocal Objects, KARAOKE

**ARTICLE INFO**: Received 22 December 2014, Revised 13 February 2015, Accepted 13 February 2015.

### 1. 서론

\*Corresponding author is with Department of Digital Contents, Korea Nazarene University, 331-718, Wolbong Ro 48 Seobuk-gu Cheonan-si Chungcheongnamdo KOREA. E-mail address: k2kim@kornu.ac.kr

사용자가 자신만의 오디오 신호를 생성하여 즐

길 수 있는 능동성을 부여하는 다객체 오디오 서비스는 공간 큐 기반의 다객체 오디오 부호화기의 등장과 함께 서비스되는 형태의 다양화가 가능하게 되었다 [1-3]. 이는 기존의 다객체 오디오 서비스가 높은 비트율로 인해 CD와 같은 대용량 저장 매체를 이용해야 된다는 제약이 높은 압축율을 제공하는 공간 큐 기반의 다객체 오디오 부호화기의 도움으로 해결됨으로써 일반적인 유무선 통신이나 네트워크 환경에서도 가능하게 되었기 때문이다.

그러나, 공간 큐 기반의 다객체 오디오 부호화기는 높은 압축율을 갖는 반면에 음질의 열화라는 근본적인 문제점을 지니고 있어 다객체 오디오 서비스에 바로 적용되기에는 어렵다. 이는 공간 큐 기반의 다객체 오디오 부호화기를 이용하여 다객체 오디오 신호를 압축/복원할 때 각각의 오디오 객체 신호가 하나의 다운믹스 신호로 합쳐졌다가 다시 각각의 오디오 객체 신호로 분리되는 과정에서 복원된 오디오 객체 신호에 다른 객체 신호 일부가 포함되어 있기 때문이다. 이러한 문제는 다양한 오디오 객체 신호를 믹싱해서 자신만의 오디오 신호를 생성하는 일반적인 응용 시나리오의 경우에는 크게 문제가 되지 않으나, 특정 오디오 객체 신호를 완벽히 제거해야 되는 가라오케와 같은 응용 시나리오에서는 음질의 열화가 두드러지게 나타나게 된다. 따라서 진정한 다객체 오디오 서비스를 위해 공간 큐 기반의 다객체 오디오 부호화기에서 특정 객체 신호를 완벽히 제어하기 위한 여러 연구가 진행되었으며, 이단 처리 구조를 갖는 공간 큐 기반의 다객체 오디오 부호화기가 제안되었다 [4-6].

이단 처리 구조를 갖는 공간 큐 기반의 다객체 오디오 부호화기는 입력되는 오디오 객체 신호를 일반 객체 신호와 보컬과 같이 완벽한 제어를 필요로 하는 특정 객체 신호로 나뉘서 순차적으로 부호화한다. 일반적인 객체 신호는 원래의 공간 큐

기반의 다객체 오디오 부호화기를 이용하여 하나의 다운믹스 신호와 각각의 객체 정보를 나타내는 공간 큐로서 부호화 되며, 특정 객체 신호는 일반적인 객체 신호의 부호화 과정에서 생성된 다운믹스 신호와 함께 두 신호의 완벽한 분리를 위해 개선된 공간 큐 기반의 다객체 오디오 부호화기를 이용하여 부호화된다. 여기에서 일반적인 객체 신호의 다운믹스 신호와 특정 객체 신호의 부호화 과정을 통해 최종적인 다운믹스 신호가 생성되며, 두 신호의 분리를 위해 추가적인 공간 큐 파라미터가 추출이 된다. 일반적인 다운믹스 신호와 특정 객체 신호의 분리를 위해 사용된 파라미터는 최종적인 다운믹스 신호로부터 복원된 일반적인 다운믹스 신호와 특정 객체 신호가 원래의 일반적인 다운믹스 신호와 특정 객체 신호와의 차이를 나타내는 잔차 신호 (residual signal)를 이용한다 [7].

잔차 신호는 원래의 객체 신호와 복원된 객체 신호의 차로 계산되기 때문에 원래의 객체 신호와 같은 주파수 전 대역에서 분포한다. 따라서 잔차 신호를 처리하기 위해서는 MP3나 AAC (Advanced Audio Coding)과 같은 부호화기를 이용하여 64 kbps의 높은 비트율로 부호화해야 된다 [8]. 기존의 연구에서는 이러한 잔차 신호의 비트율을 줄이기 위해 인간의 청각이 저주파 대역에 민감하다는 특성과 특정 객체 신호를 보컬 신호로 한정하여 잔차 신호의 주파수 대역을 특정 대역 (0-5.5 kHz)으로 고정하여 약 15 kbps로 부호화하였다 [4, 9]. 그러나 기존의 연구는 잔차 신호의 주파수 대역을 0에서 5.5 kHz로 고정하여 부호화함으로써 특정 객체 신호로 사용될 수 있는 다양한 객체 신호 고유의 주파수 특성을 반영하지 못한다. 예를 들어 베이스와 같이 저주파 대역을 갖는 객체 신호는 추출된 잔차 신호가 약 2 kHz 이하의 주파수 대역만을 가지기 때문에 기존의 0에서 5.5 kHz로 고정된 주파수 대역 부호화를 위해 사용되는 비트율을 그

대로 사용하는 경우 비트율의 낭비가 발생하게 된다. 더욱이 다양한 객체 신호의 믹싱을 통해 생성되는 오디오 채널 신호 - 일반적인 객체 신호의 부호화 과정에서 생성되는 다운믹스 신호 - 와 달리 오디오 객체 신호는 상당한 묵음 구간을 지니고 있기 때문에 특정 객체 신호와 다운믹스 신호의 분리를 위한 잔차 신호 자체가 존재하지 않는 경우가 많기 때문에 잔차 신호의 부호화가 필요하지 않은 경우가 발생한다. 따라서 특정 객체 신호의 제어를 위한 잔차 신호의 처리 방법에서의 비트율의 효율적인 소모를 위한 개선 방법에 대한 추가적인 연구가 필요하다. 본 논문에서는 기존의 이단 처리 구조를 갖는 공간 큐 기반의 다객체 오디오 부호화기에서의 잔차 신호 부호화를 효율적으로 수행하기 위한 개선 방안을 제시하기 위하여 다양한 객체 신호의 특성 분석을 통한 잔차 신호 부호화의 가변적 비트율 할당에 대하여 제안하도록 한다.

## 2. 특정 오디오 객체의 제어를 위한 잔차 신호 추출

이단 처리 구조를 갖는 공간 큐 기반의 다객체 오디오 부호화기에서 개선된 다객체 오디오 부호화기는 일반적인 다객체 오디오 부호화기 다음에 동작하며, 특정 객체 신호를 완벽히 제어해야 되는 가라오케와 같은 응용 시나리오를 위해 적용된다. 개선된 다객체 오디오 부호화기는 일반적인 다객체 오디오 부호화기로부터 생성되는 다운믹스 신호와 특정 객체 신호를 입력으로 하여 동작되며, 두 신호의 완벽한 분리를 위하여 잔차 신호를 포함하는 또 다른 공간 큐 파라미터와 최종적인 다운믹스 신호를 추출한다. 개선된 다객체 오디오 부호화기에서 추출하는 공간 큐 파라미터로 식 1을 이용하여 다운믹스 신호와 특정객체 신호간의 파워비인 OPR(object power ratio)을 <표 1>에서 보는

바와 같이 오디오 객체 신호가 갖는 주파수 대역 (0-22,050 Hz)을 28개로 나눈 서브밴드에서 추출한다 [10].

$$OPR(b) = \frac{\min[P_D(b), P_T(b)]}{\max[P_D(b), P_T(b)]}, \text{ for } 1 \leq b \leq 28 \quad (1)$$

표 1. 서브밴드 주파수 대역 (Hz)  
Table 1. Frequency region of sub-band (Hz)

서브밴드1	서브밴드2	서브밴드3	서브밴드4
0-43	44-129	130-215	216-301
서브밴드5	서브밴드6	서브밴드7	서브밴드8
302-387	388-474	475-559	560-645
서브밴드9	서브밴드10	서브밴드11	서브밴드12
646-818	819-990	991-1162	1163-1335
서브밴드13	서브밴드14	서브밴드15	서브밴드16
1336-1680	1681-2024	2025-2368	2369-2713
서브밴드17	서브밴드18	서브밴드19	서브밴드20
2714-3402	3403-4091	4092-4780	4781-5469
서브밴드21	서브밴드22	서브밴드23	서브밴드24
5470-6158	6159-6847	6848-7881	7882-8914
서브밴드25	서브밴드26	서브밴드27	서브밴드28
8915-10293	10294-12015	12016-14083	14084-22050

식 1에서  $b$ 는 서브밴드 인덱스를 나타내며,  $P_D(b)$ 와  $P_T(b)$ 는 서브밴드  $b$ 에서 다운믹스 신호와 특정 객체 신호의 파워를 나타낸다. 이렇게 계산된 OPR 파라미터를 이용하여 <표 2>와 같이 일반 다운믹스 신호와 특정 객체 신호를 복원하며, 복원된 일반 다운믹스 신호와 특정 객체 신호를 식 2에 적용하여 잔차 신호를 추출하게 된다.

$$Res(b) = D_T(b) \cdot \frac{G_T(b)}{G_D(b) + G_T(b)} - D_D(b) \cdot \frac{G_D(b)}{G_D(b) + G_T(b)} \quad (2)$$

여기에서  $D_T(b)$ 는 다객체 오디오의 부호화 과정에서 일반적인 다운믹스 신호와 특정 객체의 합으로

생성한 최종 다운믹스 신호를 나타낸다. 이렇게 추출된 잔차 신호는 오디오 객체 신호와 같은 전 대역 주파수 성분을 포함하고 있기 때문에 그대로 사용하였을 경우 많은 비트율을 소모하게 되어 다객체

오디오 부호화기의 압축 성능이 떨어지게 된다.

표 2. OPR 파라미터를 이용한 다운믹스와 특정 객체신호의 이득

Table 2. Estimated gain of down-mix and target object using OPR parameter

	일반적인 다운믹스 신호의 이득	특정 객체의 이득
특정 객체 의 파워가 클 경우	$G_b(b) = \sqrt{\frac{OPR(b)}{1+OPR(b)}}$	$G_r(b) = \sqrt{\frac{1}{1+OPR(b)}}$
일반 다운믹스의 파워가 클 경우	$G_b(b) = \sqrt{\frac{1}{1+OPR(b)}}$	$G_r(b) = \sqrt{\frac{OPR(b)}{1+OPR(b)}}$

이러한 문제를 해결하기 위하여 앞에서 설명한 바와 같이 특정 객체 신호의 주파수 특성의 변화에 따라 잔차 신호의 유효구간을 정해서 비트율을 효율적으로 소모하도록 해야 된다.

### 3. 객체 신호의 주파수 특성 분석과 잔차 신호의 부호화 비트율 할당

#### 3.1 분석 방법 및 환경

다양한 객체 신호의 주파수 특성은 잔차 신호의 부호화와 연관을 짓기 위하여 별도의 과정을 거치지 않고 이미 추출되어 있는 OPR 파라미터를 활용하도록 한다. 이는 식 2에서 보는 바와 같이 잔차 신호가 OPR 파라미터를 이용하여 추출되기 때문에 부호화기의 연산량을 최소화하기 위함이다. 또한 객체 신호는 일반적으로 가라오케 응용 시나리오를 위해 고려되는 보컬 객체 뿐 만 아니라 기존의 음악에서 자신의 연주로 대체하고자 하는 악기 연주자들을 위해 베이스, 기타, 피아노 등과 같은 다양한 객체 신호도 특정 객체로 사용되는 것을 고

려하여 곡 한곡을 구성하는 모든 객체 신호의 주파수 특성에 대한 분석을 하도록 한다.

OPR 파라미터는 식 1에서 보는 바와 같이 일반적인 객체 신호로 생성한 다운믹스 신호와 특정 객체 신호의 파워비로 계산되며, <표 1>과 같이 전체 주파수 대역을 28개로 나눈 서브밴드에서 추출된다. 따라서 28개의 서브밴드에서 추출된 OPR 파라미터 값을 이용하여 특정 객체 신호의 주파수 특성을 파악할 수 있다. 추출된 OPR 파라미터는 0에서 1사이의 값을 가지고 있으며, 특정 서브밴드에서 추출된 OPR 파라미터 값이 0일 경우 특정 객체 신호가 존재하기 때문에 해당 서브밴드에서 추출되는 잔차 신호 또한 존재하지 않게 된다. 이러한 특성을 이용하여 28개의 서브밴드 중 OPR 파라미터 값을 0으로 갖는 서브밴드를 선별하여 특정 객체 신호의 주파수 특성을 파악할 수 있게 된다.

일반적으로 객체 신호의 주파수 성분은 0 Hz에서 특정 주파수 X Hz 까지 분포하며 특정 주파수 X Hz에서 Y Hz까지 분포하는 경우는 거의 없기 때문에 객체 신호의 특성을 분석할 때 최상위 주파수 대역을 나타내는 28번째 서브밴드에서 아래쪽 서브밴드로 이동하면서 OPR 파라미터의 값이 0이 아닌 첫 번째 서브밴드를 찾는다. 이에 따라 특정 객체 신호가 분포하는 주파수 대역은 0 Hz에서 OPR 파라미터 값이 0이 아닌 첫 번째 서브밴드까지로 정할 수 있으며, 잔차 신호 역시 이와 동일한 주파수 대역을 가지게 된다. 이와 같은 방법으로 객체 신호의 주파수 특성을 분석하면 이론상 객체 신호는 OPR 파라미터를 추출하는 서브밴드 수와 같은 주파수 대역 분포를 가지며, 잔차 신호가 가지는 주파수 대역 역시 OPR 파라미터를 추출하는 서브밴드의 수와 같다. 그러나, 기존의 다객체 오디오 부호화에서 사용되는 잔차 신호의 주파수 대역은 5.5 kHz로 한정되어 있으므로 <표 2>의 20번째 서브밴드까지만 사용하며 잔차 신호의 부호화

의 복잡도를 고려하여 20개의 서브밴드를 4개의 주파수 대역으로 다시 나눠서 분석한다. 또한 앞에서 언급한 바와 같이 오디오 객체 신호는 상당한 묵음 구간을 지니고 있기 때문에 특정 객체 신호와 다운믹스 신호의 분리를 위한 잔차 신호 자체가 존재하지 않는 경우가 많기 때문에 잔차 신호의 부호화가 필요하지 않은 경우를 추가하여 <표 3>과 같이 5개의 영역으로 나눠서 객체 신호의 주파수 특성을 분석하도록 한다. 분석을 위해 <표 4>에서 보는 바와 같은 대중가요 5곡을 사용하였으며, 각 곡은 4개에서 6개의 오디오 객체로 이루어져있다. 또한 각각의 오디오 객체 신호는 16 비트 양자화와 44.1 kHz 샘플링 주파수를 갖는다.

표 3 객체 신호 특성 분석을 위한 주파수 대역 분할  
Table 3. Partitions of frequency bands for analysis of spectral characteristic of audio objects

영역	해당 서브밴드	주파수대역 (Hz)
0	없음	없음
1	1에서 12 번째	0-1,335
2	1에서 16 번째	0-2,713
3	1에서 18 번째	0-4,091
4	1에서 20 번째	0-5,469

표 4 테스트 아이템  
Table 4. Test items

인덱스	아이템	오디오 객체	길이
A	Hajiman	Guitar, bass, keyboard, rhythm, chorus, vocal	240초
B	Braves	Guitar, bass, keyboard, rhythm, chorus, vocal	204초
C	Snow	Guitar, bass, strings, rhythm, chorus, vocal	235초
D	LaLaLa	Strings, bass, drum, vocal	275초
E	Sulpun	Guitar, bass, piano&brass, rhythm, chorus, vocal	214초

### 3.2 객체 신호의 주파수 특성 분석 결과

<표 5>와 <그림 1>은 각각의 테스트 아이템에서

의 <표 3>의 주파수 대역 분포 비율을 보여준다. <표 5>와 <그림 1>에서 보는 바와 같이 모든 테스트 아이템은 평균 19 %의 상당한 비율의 묵음 구간을 지니고 있으며, 기존의 잔차 신호의 주파수 대역 0에서 5.5 kHz인 영역 4는 약 58 %로 실제 잔차 신호의 약 42 %는 그 이하의 주파수 대역을 지니고 있음을 알 수 있다. <표 6>과 <그림 2>는 각각의 오디오 객체별 주파수 대역의 분포 비율을 보여준다. 대부분의 오디오 객체는 약 10 % 이상의 많은 묵음 구간을 지니고 있으며, 각 오디오 객체 별 특성에 따라 다양한 주파수 대역을 지니고 있음을 알 수 있다. 특히, 베이스 객체는 묵음을 포함하여 2 kHz 이하의 저주파 대역의 비율이 약 85 % 정도로 매우 높으며, 코러스 객체의 경우 묵음 구간이 약 60 % 이상을 차지하고 있다. 또한, 다른 객체의 경우에도 기존의 잔차 신호의 주파수 대역인 영역 4보다 낮은 주파수 대역이 평균 20 %이상의 비율을 차지한다. 이와 같은 오디오 객체의 주파수 특성 분석 결과를 통해 기존의 잔차 신호의 주파수 대역을 0에서 5.5 kHz로 고정하고 약 15 kbps의 비트율을 이용하여 부호화 하는 것의 비효율성을 확인할 수 있다. 결론적으로 객체의 주파수 특성에 따라 잔차 신호의 주파수 대역과 그에 따른 잔차 신호의 부호화 비트율을 정할 필요가 있다.

표 5. 테스트 아이템별 주파수 대역 비율 (%)  
Table 5. Portion of spectral bands of test items (%)

아이템	주파수 대역				
	0	1	2	3	4
A	20.41	16.86	5.08	2.61	55.04
B	16.07	16.43	5.87	4.98	56.65
C	22.40	14.26	2.37	3.49	57.48
D	14.61	13.01	7.56	4.74	60.08
E	20.96	11.23	4.58	2.92	60.31
Total average	19.13	14.44	4.96	3.68	57.78

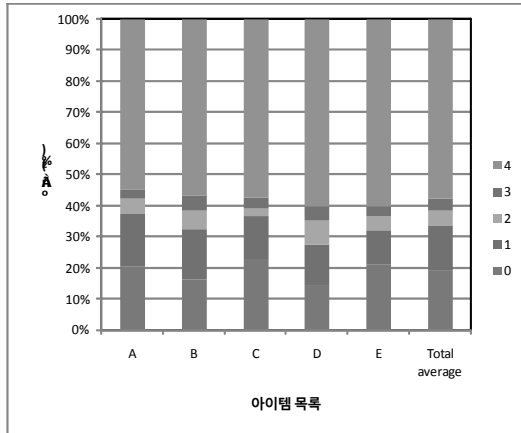


그림 1. 테스트 아이템별 주파수 대역 비율.  
Fig 1. Portion of spectral bands of test items.

표 6. 오디오 객체별 주파수 대역 비율 (%)  
Table 6. Portion of spectral bands of audio objects (%)

객체	주파수대역				
	0	1	2	3	4
Guitar	5.05	5.05	6.66	5.29	77.94
Bass	15.10	56.35	13.54	4.92	10.09
Vocal	13.84	0.98	1.38	4.39	79.41
Rhythm	12.72	6.88	0.85	1.09	78.46
Piano/strings	12.21	7.37	5.31	4.73	70.38
Chorus	62.82	5.76	1.67	1.55	28.21

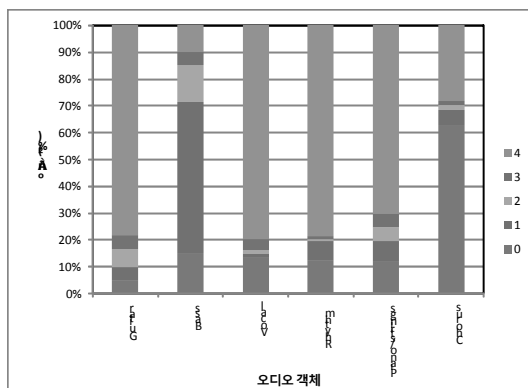


그림 2. 오디오 객체별 주파수 대역 비율.  
Fig 2. Portion of spectral bands of audio objects.

### 3.3 잔차 신호의 부호화 비트율 할당

앞에서 살펴본 오디오 객체의 주파수 특성에 따라 잔차 신호의 주파수 대역을 정하고, 이에 따라 잔차 신호의 부호화 비트율을 가변적으로 할당할 수 있다. 즉, <표 7>에서 보는 바와 같이 특정 오디오 객체의 주파수 대역에 따라 잔차 신호의 주파수 대역과 부호화 비트율이 결정된다. 여기에서 저주파 대역은 인간의 청각 특성 상 가장 민감한 구역이므로 음질의 열화를 피하기 위하여 비트율 할당을 높게 한다.

표 7. 잔차 신호의 주파수 대역 결정과 비트율 할당  
Table 7. Frequency band decision and bit-rate allocation of residual signal

오디오 객체 주파수 영역	잔차 신호 주파수 대역 (Hz)	비트율 (kbps)
0	0	0.0
1	0-1,335	7.5
2	0-2,713	10.0
3	0-4,091	12.5
4	0-5,469	15.0

표 8. 아이템별 잔차 신호의 가변적 비트율과 감소율  
Table 8. Variable bit-rate and reduction rate of residual signal coding of test items.

아이템	고정비트율 (kbps)	가변 비트율 (kbps)	감소율 (%)
A	15	10.36	30.97
B	15	10.94	27.07
C	15	10.36	30.91
D	15	11.34	24.42
E	15	10.71	28.58
Total average	15	10.71	28.62

표 9. 객체별 잔차 신호의 가변적 비트율과 감소율  
Table 9. Variable bit-rate and reduction rate of residual signal coding of audio objects.

	고정 비트율 (kbps)	가변 비트율 (kbps)	감소율 (%)
Guitar	15	13.40	10.68
Bass	15	7.71	48.61
Vocal	15	12.67	15.52
Rhythm	15	12.51	16.63
Piano/strings	15	12.23	18.46
Chorus	15	5.02	66.51

<표 7>과 같이 결정된 잔차 신호의 주파수 대역과 비트율을 이용하여 실제 잔차 신호의 부호화를 수행하면 <표 8>과 <표 9>와 같은 비트율의 감소를 얻을 수 있다. 테스트 아이템 별로는 <표 8>과 같이 잔차 신호의 부호화를 위해 평균 약 10.7 kbps의 비트율이 할당되어 약 28.6 %의 비트율 감소를 얻을 수 있다. 또한 오디오 객체 별로는 가라오케 응용시나리오에서 일반적으로 특정 객체로 사용되는 보컬 객체의 경우 약 15 % 이상의 비트율 감소를 얻을 수 있으며, 특히 저주파 대역으로 구성되어 있는 베이스 객체의 경우 약 48 % 이상의 비트율 감소를 얻을 수 있다. 결론적으로 제안된 객체 신호의 주파수 특성에 따라 잔차 신호의 주파수 대역과 부호화 비트율을 가변적으로 정하는 방법에 의해 높은 비트율 감소를 얻을 수 있음을 확인하였다.

#### 4. 결론

본 논문에서는 사용자에게 능동성을 부여하는 다객체 오디오 서비스가 통신이나 네트워크 환경에서도 가능하도록 하는 이단 처리 구조를 갖는 공간 큐 기반의 다객체 오디오 부호화기의 성능 향상을 위한 방법에 대하여 다루었다. 제안된 방법은 특정 객체 신호를 완벽하게 제어하기 위하여 사용되는 잔차 신호를 효율적으로 다루기 위하여 객체 신호의 주파수 특성을 분석하고 이에 따른 잔차 신호의 주파수 대역과 비트율을 가변적으로 할당하도록 하였다.

제안된 방법에 따라 기존의 방법과 비교하여 아이템별로는 약 28.6 %의 높은 비트율 감소를 얻을 수 있었으며, 객체 별로는 전체적으로 약 10 % 이상의 비트율 감소를 얻을 수 있었다. 특히 가라오케 응용 서비스에서 주로 활용되는 보컬 객체의 경우 15 % 이상의 비트율 감소를 보여 기존의 다

객체 오디오 부호화기의 성능을 크게 향상 시킬 수 있음을 알 수 있다.

제안된 방법은 기존의 방법과 비교하여 음질에는 영향을 미치지 않고 비트율을 크게 감소시키는 것에 초점을 맞추고 있으므로 제안된 방법에 의해 생성된 출력 신호의 음질에 대한 평가를 통한 제안된 방법의 성능 검증이 향후 과제로 남아있다.

#### References

- [1] D.-Y Jang, T.-J Lee, Y.-J Lee, and J.-H Yoo, *A personalized preset-based audio system for interactive service*, in Proc. 121st AES Conv., San Francisco, CA, Preprint 6904, 2006.
- [2] *Consideration of interactive music service*, ISO/IEC JTC1/SC29/WG11 (MPEG), Archamps, Apr. 2008, Document M15390.
- [3] ISO/IEC 2003-2 *Information technology -MPEG audiotecnologies—Part 2: Sptial audio object coding (SAOC)*, 2010.
- [4] K.-K Kim J.-I Seo, S.-K Beack, K.-O Kang, and M.-S Hahn, *Spatial audio object coding with two-step coding structure for interactive audio service*, IEEE Trans. Multimedia, Vol. 13, No. 6, pp. 1208-1216, Dec. 2011.
- [5] K.-K Kim, and B.-O Jang, *Multiple audio object coding for interactive audio service*, Journal of The Korea Knowledge Information Technology Society, Vol. 8, No. 6, pp. 63-70, Dec. 2013.
- [6] J.-H Park, K.-K Kim, and M.-S Hahn, *Vocal removal from multiobject audio using harmonic information for karaoke service*, IEEE Trans. Audio Speech and Language Processing, Vol. 21, Apr. 2013.
- [7] J. Breebaart, S. Disch, C. Faller, J. Herre, J. Hilpert, K. Kjolring, F. Myburg, H.

Purnhagen, and E. Schuijers, *The reference model architecture for MPEG spatial audio coding*, in Proc. 118th AES Conv., Barcelona, Spain, Preprint 6447, 2005.

- [8] ISO/IEC 14496-3:2005, *Information technology - coding of audio-visual objects - part 3: Audio*, 2006.
- [9] Zwicker, Eberhard, and Hugo Fastl. *Psychoacoustics: Facts and models*, Vol. 2. Berlin: Springer, 1999.
- [10] C. Faller, and F. Baumgarte, *Binaural cue coding - part II: Schemes and applications*, IEEE Trans. Speech Audio Processing, Vol. 11, pp. 520-531, Nov. 2003.

## 다객체 오디오 부호화기 성능 향상을 위한 오디오 객체 신호 특성 분석

김광기

나사렛대학교 디지털콘텐츠학과

### 요약

본 논문에서는 사용자에게 능동성을 부여하는 다객체 오디오 서비스가 통신이나 네트워크 환경에서도 가능하도록 하는 이단 처리 구조를 갖는 공간 큐 기반의 다객체 오디오부호화기의 성능 향상을 위한 방법에 대하여 다룬다. 제안된 방법은 특정 객체 신호를 완벽하게 제어하기 위하여 사용되는 잔차 신호를 효율적으로 다루기 위하여 객체 신호의 주파수 특성을 분석하고 이에 따른 잔차 신호의 주파수 대역과 비트율을 가변적으로 할당하도록 하였다. 제안된 방법은 객체 신호의 주파수 대역을 묶음 구간을 포함하여 5개의 영역으로 나뉘어 분석하였으며, 기존의 고정된 고주파 대역과 비교하여 대부분의 오디오 객체에서 약 20% 이상의 저주파 영역을 포함하고 있음을 확인할 수 있다. 따라서 제안된 방법에 따라 기존의 방법과 비교하여 아이템 별로는 약 28.6%의 높은 비트율

감소를 얻을 수 있었으며, 객체 별로는 전체적으로 약 10% 이상의 비트율 감소를 얻을 수 있었다. 특히 가라오케 응용 서비스에서 주로 활용되는 보컬 객체의 경우 15% 이상의 비트율 감소를 보여 기존의 다객체 오디오 부호화기의 성능을 크게 향상시킬 수 있음을 알 수 있다.

### 감사의 글

본 논문은 2014년도 나사렛대학교 교비학술연구조성비 지원에 의해 연구되었음.



**Kwangki Kim** received the B.S. degree in electronic engineering from Korea Aviation University, Koyang, South Korea, in 2002, the M.S. degree and the Ph.D. degree in department of

Information and Communications Engineering at Korea Advanced Institute of Science and Technology, Daejeon, South Korea, in 2004 and 2011, respectively. He worked as a researcher in Samsung DMC R&D center in 2012. Currently, He is a professor in Korea Nazarene University, Chon-an, Korea. His research interests include multi-channel/-object audio coding, spatial audio, 3D sound, and their applications.

*E-mail address:* k2kim@kornu.ac.kr