

지상파 DTV 와 지상파 DMB 방송을 위한 오디오 트랜스코딩 알고리즘

방경호*, 이재성*, 이창준*, 박영철**, 서정일***

*연세대학교 전기전자공학과, **연세대학교 정보기술학부, ***한국전자통신연구원

Audio Transcoding Algorithm for Terrestrial DTV and Terrestrial DMB Systems

Kyoung Ho Bang*, Jae Seong Lee*, Chang Joon Lee*, Young Cheol Park**, Jeong Il Seo***

{euphony, *dream7070, *yesjoon}@mcsp.yonsei.ac.kr, **young00@dragon.yonsei.ac.kr, ***seoji@etri.re.kr

요약

본 논문에서는 지상파 DTV 의 저작물을 지상파 DMB 방송에 활용할 수 있는 오디오 트랜스코딩 기법에 대해 제안한다. 지상파 DTV 에서는 오디오 신호를 AC-3 방식으로 압축하는 반면, 지상파 DMB 에서는 MPEG-4 BSAC 방식을 사용한다. 각 알고리즘이 사용하는 주파수 변환 방식과 심리음향모델에 의한 비트할당 기법이라는 유사성을 이용하면, 두 방식간의 트랜스코딩 효율을 향상시킬 수 있다. 실시간 변환을 요구하는 경우나 휴대기기를 위한 응용분야에서는 지연시간과 전력소모를 줄일 수 있는 잇점을 갖는다.

1. 서론

국내 디지털 TV 방송에 대한 이동형/고정형 전송방식에 대한 논쟁이 고정수신 방식으로 결정이 되면서, 기술적으로 취약했던 DTV 방송의 이동수신 영역을 지상파 DMB 가 담당하게 될 전망이어서 국내 지상파 DMB 의 상용화가 시급한 상황이다. 현재 지상파 DTV 방송의 오디오 기술 규격은 AC-3[1]를 사용하고, 지상파 DMB 방송에서는 MPEG-4 BSAC (Bit Sliced Arithmetic Coding) [2]을 사용하기 때문에,

DTV 방송을 DMB 방송으로 재전송할 경우 서로 다른 오디오 방식으로 인해 직접적인 전송이 불가능하다.

이를 해결하기 위해서는 서로 다른 오디오 방식의 신호를 변환해 주는 기술이 필요하다. 변환 기술은 가져작물의 오디오 데이터를 PCM 신호로 변환한 후 다시 압축하는 탄델 (tandem) 기법과 비트열 혹은 주파수 영역의 오디오 신호에서 포맷을 전환해주는 트랜스코딩 (transcoding) 기법이 있다.

본 논문에서는 AC-3 방식과 MPEG-4 BSAC 방식의 기술 분석을 통해, 트랜스코딩에 사용 가능한 블럭을 정의하고, DTV 방송을 위해 저작된 오디오 데이터를 지상파 DMB 방송에 사용할 수 있도록 변환해 주는 트랜스코딩 기술에 대해 설명하겠다.

2. AC-3 와 MPEG-4 BSAC 기술 분석

AC-3 에서 MPEG-4 BSAC 으로 트랜스코딩을 위해서 각 알고리즘이 사용하는 압축 기법에 대해서 설명한다.

2.1 AC-3

AC-3 오디오 부호화 알고리즘 [3]은 북미 지역의 HDTV 오디오 압축 기술의 표준안 [4]으로 Dolby 연구소가 1991 년에 제안한 방식이다. SMPTE

권고안인 5.1 채널을 수용하고, 32~640kbps 의 비트율을 갖는다. 압축비를 높이기 위해서 채널간 혹은 채널 내의 마스킹 특성을 이용하며, 고주파 대역의 채널 커플링을 이용하여 보다 낮은 비트율을 제공한다. 기본적인 부호화 방법은 시간축 에일리어징 제거 방법에 기초한 MDCT 변환 부호화 방식을 사용하였다. 다음은 AC-3 부호화기 및 복호화기의 블록 다이어그램이다.

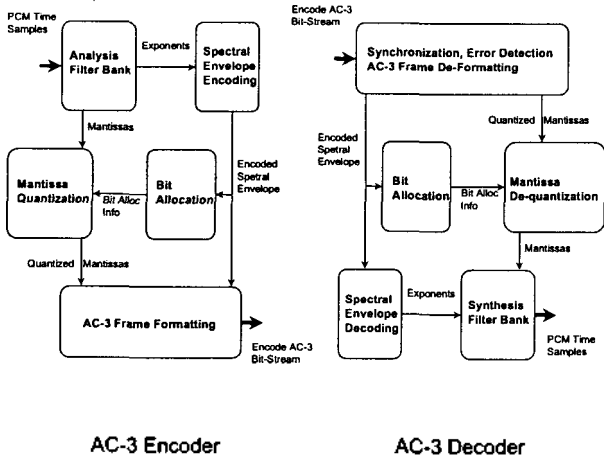


그림 1. AC-3 부호화기 및 복호화기

입력 신호는 윈도우를 씌운 후 MDCT 를 이용하여 주파수 영역으로 변환한다. MDCT 는 시간 영역 에일리어징을 제거하는 특성을 가질 뿐만 아니라 임계 샘플링을 보장하고 계산량을 크게 줄일 수 있는 고속 알고리즘을 적용할 수 있다. 또한 주파수 선택성이 매우 향상된 특성을 갖는다.

MDCT 계수들은 지수부와 가수부로 나뉘어져 부호화 되는데, 이는 간단한 비트 연산을 통하여 추출하므로 실제 구현시 계산량이 많은 로그 계산을 효과적으로 줄일 수 있는 장점이 있다. 또한 지수값을 사용하여 거칠게나마 신호의 스펙트럼을 구할 수 있기 때문에 심리음향 모델을 구현하기 위하여 따로 FFT 와 같은 계산을 필요로 하지 않아 복잡도를 크게 줄일 수 있다. 부호화시 지수부는 고정 비트로 할당하여 전송하되, 전송시 비트의 낭비를 막기 위해서 몇 개의 지수를 묶어 부호화하는 그룹화 방법을 사용한다.

2.2 MPEG-4 BSAC

MPEG-4 BSAC 은 MPEG-4 AAC 의 Huffman Coding 기법에 대한 차선택 중의 하나로써, fine granule scalability 를 제공한다. 채널당 16kbps 에서 64kbps 까지 1kbps 단위로 small step scalability 를 제공할 수 있다.

BSAC 의 기본계층에는 일반적인 프레임 헤더정보와 부가정보가 전송되고, 강화계층에는 각 강화계층별 부가정보와 실제 부호화된 데이터가 포함된다. 강화계층은 1kbps/ch 단위로 계층을 만들 수 있다. BSAC 은 무손실 부호화 부분을 제외하고 나머지 부분은 AAC 와 동일한 툴을 사용한다. 무손실 부호화 부분에서 fine granule scalability 를 가능하게 하는 핵심 기술은 다음과 같다.

- ① 계층 구조를 이용한다. 계층 구조를 정의하는 과정에서 계층별 가용 비트 수, 주파수 대역 등이 결정된다.
- ② Bit-Sliced Arithmetic Coding 을 통하여 양자화 샘플을 비트단위로 부호화한다.
- ③ 양자화 정보와 같은 부가정보를 계층별로 나누어 부호화한다. 계층별로 해당 계층을 복호화 하기 위해 필요한 부가 정보만을 부호화함으로써 낮은 계층에서의 음질이 보장된다.

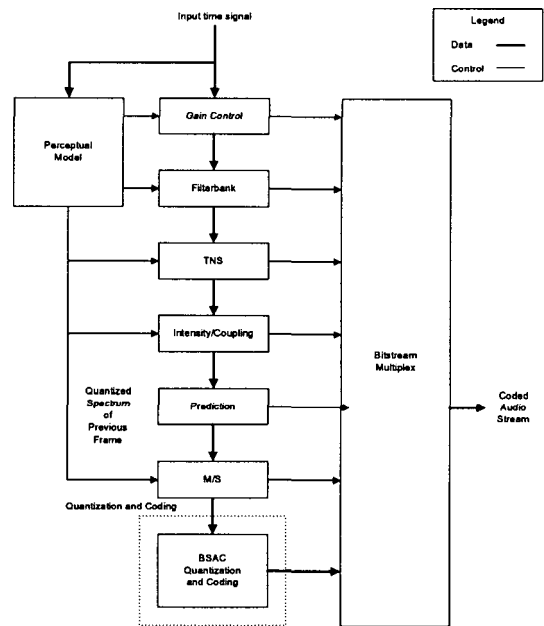


그림 2. MPEG-4 BSAC 부호화기

그림 2 는 MPEG-4 BSAC 부호화기의 블록 다이어그램을 나타내고 있다. 주파수 변환을 위해서는 MDCT 변환을 사용하고, 심리음향모델에서 계산된 마스킹 곡선을 이용하여 양자화를 수행한다.

3. AC-3 에서 MPEG-4 BSAC 으로의 트랜스코딩 기법

AC-3 에서 BSAC 으로의 트랜스코딩을 위하여 크게 두가지 방향으로 접근하였다.

3.1 T/F 변환

AC-3 와 BSAC 은 시간-주파수 변환을 위하여 공통적으로 MDCT 변환을 사용한다는 점에 착안하여, AC-3 복호화 과정의 IMDCT 와 BSAC 부호화 과정의 MDCT 를 제거하고, MDCT 도메인의 신호를 이용하여 트랜스코딩이 가능한지 분석하였다. AC-3 에서는 256 크기의 블록 6 개를 사용하여 하나의 프레임을 구성하고, BSAC 에서는 1024 크기의 블록을 이용하여 하나의 프레임을 구성한다.

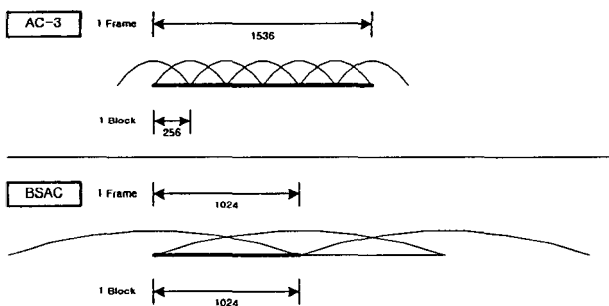


그림 3. AC-3와 BSAC의 T/F 변환 블록

그림 3 은 AC-3 와 BSAC 에서 사용하고 있는 시간-주파수 변환블럭이다. 즉 AC-3 에서는 1536 개의 시간영역의 신호를 512 변환블럭을 50% 겹쳐서 6 개의 블록을 형성하고 있으며, BSAC 에서는 2048 변환블럭을 50% 겹쳐서 하나의 블록을 형성하고 있다. AC-3 에서는 6 개의 변환 블록이 서로 다른 시간으로 간주되어 계산 되고, BSAC 에서는 1024 개의 시간영역 신호가 하나의 시간으로 간주되어 변환된다. 주파수 영역에서 바라보면, AC-3 는 256 의 주파수 해상도를 가지는 반면, BSAC 에서는 1024 의 주파수 해상도를

갖게 된다. 트랜스코딩이 가능하려면, 저해상도의 데이터를 이용하여 보간을 해야하는 문제점이 발생하게 된다. 또한 시간 영역에서의 지연시간 문제도 발생하게 되어 MDCT 영역에서의 트랜스코딩은 불가능한 것으로 분석된다.

3.2 심리음향모델을 이용한 비트할당 과정

AC-3 와 BSAC 은 심리음향모델에 기반한 비트할당 과정을 이용한다는 또하나의 공통점을 지니고 있다. AC-3 의 심리음향 계산 과정은 다음과 같다.

- ① Spectral envelope (=exponent)을 decoding
- ② 252 power spectral density line 생성 (exponent, bandwidth, sample rate 에 따라 line 수 가변)
- ③ 252 psd line 을 이용하여, 64 band 를 형성 (low freq.: band size 1, high freq.: band size 16)
- ④ 각 band 의 psd line 에 대해서, spreading function 을 convolution
- ⑤ Fast decaying masking curve 와 slowly decaying masking curve 을 적용하여 log domain 계산
- ⑥ Band 별 predicted masking 을 구함

Predicted masking 과 Exponent 값을 비교하여 bit allocation 을 위한 SNR 계산.

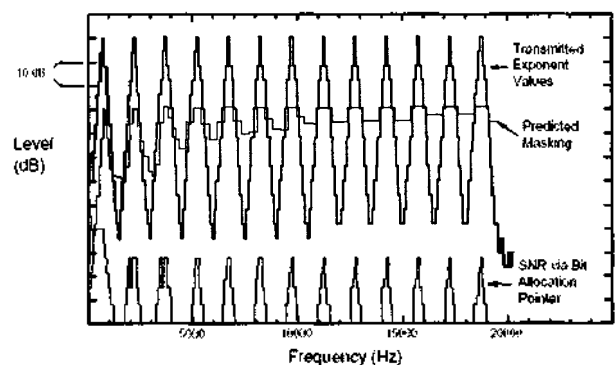


그림 4. AC-3의 비트할당 정보

그림 4 는 AC-3 에서 사용하는 비트할당 정보이다. 이 정보는 6 개의 AC-3 블록에 공통적으로 사용되는 것으로 프레임의 특성을 반영한다. 이것을 이용하여,

BSAC 의 심리음향모델 과정을 대체할 수 있다면, BSAC 부호화 과정의 약 30%를 차지하는 연산량을 감소시킬 수 있다.

4. 실험 및 결과

AC-3 에서 와 BSAC 으로 트랜스코딩을 위하여, T/F 주파수 변환과정의 시간 영역과 주파수 영역에서의 해상도를 비교한 결과가 표 1 과 같다.

표 1 AC-3와 BSAC의 시간-주파수 해상도 비교

	Time	Frequency
AC-3	10.66 msec	93.7 Hz
BSAC	42.66 msec	23.4 Hz

저해상도의 블록을 이용하여 고해상도의 주파수 영역 신호를 재생해야 하는 문제점이 발생하고 시간 영역에서 지연시간 문제로 인하여 MDCT 영역에서의 트랜스코딩은 효율성이 떨어지는 결과를 보이고 있다.

다음으로 AC-3 의 심리음향 분석과정을 통하여 얻어진 비트할당 정보인 SNR 은 BSAC 의 심리음향 계산 과정을 대체 할 수 있는 지 실험하였다.

AC-3 에서는 64 개의 밴드에 대해서 각각 하나씩의 SNR 정보를 계산하게 된다. 이 정보는 그림 4 에서와 같이 전송된 지수값과 예측된 마스킹 값으로부터 구해지게 된다. 한편 BSAC 에서는 49 개의 스케일팩터 밴드를 사용하고, 각각의 밴드에서 마스킹 임계값이 계산되고, 그 값과 양자화잡음과의 비교를 통하여 스케일팩터를 이용한 비트할당을 하게 된다.

이 정보들로부터 다음과 같은 과정을 통하여 트랜스코딩을 진행한다.

- ① Band Mapping (64 band → 49 band)
- ② SNR for bit allocation → initialization of scalefactor-set

SNR 정보로부터 구한 마스킹 임계치를 이용하여, 스케일팩터를 계산하면, BSAC 의 심리음향모델을 이용한 마스킹 임계치와 유사한 결과를 얻을 수 있었다.

MPEG-4 BSAC 부호화 과정에서 심리음향모델 계산과정과 중첩반복루프에 의한 비트할당 과정은 각각 전체 연산량의 30%와 50%를 차지하는 과정으로, 이 두 과정에 필요한 연산량을 제거할 수 있기 때문에, 탠덤 방식에 비해서 MPEG-4 BSAC 부호화 과정의 약 70%의 연산량을 감소시킬 수 있다.

5. 결론

본 논문에서는 지상파 DTV 와 지상파 DMB 를 위해 저작된 오디오 데이터들 서로 변환해주기 위한 트랜스코딩 기법에 대해서 분석하고, 방법론을 제시하였다. 먼저 T/F 변환 관계에서 각각의 오디오 압축 알고리즘이 사용하는 기술의 유사성을 이용하여, MDCT 영역에서의 트랜스코딩이 가능한지 분석하였다. 다음으로 심리음향모델에 기반한 비트할당 과정이라는 유사성을 이용하여, 연산량을 줄일 수 있는지 분석하고, 새로운 변환 과정을 제시하였다. 본 논문에서 제시한 알고리즘은 심도있는 실험적 보안을 통하여 DTV 에서 저작된 오디오 데이터들 지상파 DMB 에서 실시간으로 활용할 수 있을 것이다.

참고문헌

- [1] ISO/IEC JTC1/SC29/WG11 No.972, "Dolby AC-3 Multi-channel Digital Audio Compression System Algorithm Description", Nov., 1993.
- [2] ISO/IEC JTC1/SC29/WG11 No.2203TF, "Information Technology – Coding of AudioVisual Objects Part3: Audio, Subpart4: Time/Frequency Coding, MPEG-4", 1996.
- [3] Craig C. Todd, et al. "AC-3: Flexible Perceptual Coding for Audio Transmission and Storage," Audio Engineering Society, 1993.
- [4] D. Meares "High-Definition Sound of High Definition Television" AES/SMPTE Joint Television Conference, Detroit, Feb 1-2, 1991.