

지상파 및 위성 DMB에서의 오디오 부호화 기법

□ 김인철 / 서울시립대학교 전자전기컴퓨터공학부

1. 서론

오디오[1]는 CD, DAT 등 음악 혹은 음향 정보를 표현할 뿐만 아니라, 비디오 등 다른 정보들과 결합되어 복합적인 멀티미디어 정보를 표현하는데 필수적이다. 오디오는 단독 혹은 다른 미디어와 함께 사용되어 방향성 및 공간감을 높여 상황을 보다 효과적으로 묘사하는데 도움을 준다. 특히, 자연스러운 오디오의 재생은 음악의 시청 뿐만 아니라, 시청각 환경의 생동감을 높여주는 데에도 필수적이다. 그러나 자연스러운 오디오를 표현하기 위해서는 20-20000Hz 범위의 가청주파수 영역을 모두 포함할 뿐만 아니라 2채널 스테레오 이상의 채널 수를 확보하여야 한다. 이 경우 오디오 정보량이 매우 방대해 지므로 오디오 정보를 효과적으로 감축시키는 오디오 부호화 기법에 대한 연구가 활발히 진행되어 왔다[1,2].

오디오 신호는 비정상 신호 (non-stationary signal)로 모델링되기 때문에 통계적인 특성을 이용한 일반적인 파형부호화(waveform coding) 기법으로 그 정보를 압축하는데 한계가 있다. 1975년 Blauert와 Tritthart가 음성 부호화에 도입하고, 1979년에 Krasner[3]가 오디오 부호화에 적용한 이래 사람 귀의 특성을 고려하여 부호화하는 청각 오디오 부호화 (perceptual audio coding) 기법은 일반적인 추세가 되었다. 청각 오디오 부호화의 특성을 가장 잘 설명하는 것이 “13 dB miracle”이다. 이것은 SNR (signal to noise ratio)이 13dB가 되도록 잡음을 섞은 오디오는 심각하게 손상되어 청취되지만, 청각 오디오 부호화를 통해 원음과 거의 구별할 수 없는 정도로 부호화된 오디오 또한 SNR이 13dB로 측정된다는 것이다.

사람 귀의 특성을 고려하는 심리 음향 모델 (psycho-acoustic model)과 함께 오디오 부호화의

성능을 크게 향상시킨 방법은 필터 뱅크를 이용한 주파수 대역신호로의 분석 및 합성 기법이다. 이 방법은 완전 복원이 가능한 MDCT (modified discrete cosine transform)[4]로 발전하였다. 필터 뱅크를 이용한 분석 합성 및 MDCT는 오디오 신호를 주파수 영역에서 세분화된 대역 신호들로 분해하는 시간-주파수 변환을 수행한다. 시간 주파수 변환 기법들은 주파수 선택 필터 (frequency selective filter)로 모델링되는 사람 귀의 특성을 가장 효과적으로 이용하여 오디오 신호를 부호화할 수 있게 한다.

심리음향모델과 시간-주파수 변환을 이용한 최초의 오디오 부호화 표준은 MPEG-1 오디오[5]이다. MPEG-1 오디오는 전송률 및 부호화 복잡도에 따라 layer I, II 및 III 등으로 구분되는데, layer II는 유럽의 DAB (digital audio broadcasting) 표준으로 채택되었고, 통상적으로 MP3로 알려진 layer III는 현재 가장 널리 사용되고 있는 오디오 표현 방식이다. MPEG-1 오디오는 2채널 오디오를 CD 수준의 품질로 부호화하는 방법으로 layer III의 경우 약 13:1의 압축률을 보인다. MPEG-2 오디오[6]에서는 5.1채널 등 다채널 오디오를 수용할 수 있도록 MPEG-1 오디오를 확장하여 방송용으로 사용할 수 있게 하였다. 그렇지만, 미국에서 채택된 HDTV 표준[7,8]에는 MPEG-2 오디오가 아니라 Dolby사의 AC-3가 채택되었다. 이에 자극받아 1997년 MPEG에서는 더욱더 향상된 성능을 보이는 오디오 부호화 방법인 AAC (advanced audio coding)를 MPEG-2 제7부[9]로 제정하였다. MPEG-2 AAC는 긴 길이를 가지는 MDCT를 시간-주파수 변환 기법으로 채택함으로써 변환 효율성을 향상시키고, TNS (temporal noise shaping), 예측 (prediction) 등 새로운 부호화 도구 (coding tool)들을 도입하였

다. 이와 동시에 양자화 및 호프만 부호화 방법을 개선시켜 기존의 오디오 부호화 기법에 비해 성능을 크게 향상시켰다. MPEG-4 오디오[10]에서는 LTP (long-term prediction), PNS(perceptual noise substitution) 등 새로운 부호화 도구 (coding tool)를 도입하는 등 MPEG-2 AAC를 확장하여 일반 오디오 부호화 (generic audio coding) 기법으로 채택하였다. 특히, BSAC (binary arithmetic coding) 등 오류에 대한 강인성을 높이거나, 계층 부호화를 위한 도구들이 도입되었다. 2003년에 MPEG-4 오디오에서는 수정 표준 1 (amendment 1)[11]로서 부호화 효율을 대폭 향상시키는 SBR (spectral band replication) 기법을 새로이 규정하였다.

2005년부터 서비스되고 있는 지상파 및 위성 DMB (digital multimedia broadcasting)[12-14]에서는 한정된 비트량으로 오디오 뿐만 아니라 비디오, 데이터 등 멀티미디어 서비스를 제공함을 목표로 하고 있다. 이 가운데 지상파 및 위성 DMB에서의 오디오 서비스에 대한 요구 조건은 동일하게 정의되어 있는데, 그 요구 조건을 표 1에 요약하였다 [12,13]. 표 1을 참조하면, 오디오 서비스는 128Kbps 이하의 전송율에서 CD 수준의 품질을 유지해야 할 수 있다. 지금까지 소개된 오디오 부호화 표준들은 다양한 응용분야에서 선택적으로 사용되고 있지만, DMB 오디오 요구 조건을 만족시키는 오디오

(표 1) DMB에서의 오디오 서비스 요구사항

파라미터	값
표본화주파수	48KHz 이하
채널 수	mono 혹은 stereo
비트율	128Kbps 이하
오디오만의 품질	CD 수준의 품질
비디오서비스에서 오디오의 품질	아날로그 FM 이상의 품질

부호화 표준은 AAC 밖에 없다. 이에 따라 지상파 DMB[12]에서는 오디오 표준으로 MPEG-4 BSAC 을 채택하였고, 위성 DMB[13]에서는 MPEG-2 AAC에 SBR을 적용하는 기법을 채택하였다.

본 고에서는 지상파 및 위성 DMB를 위한 오디오 부호화 기법에 대하여 살펴보겠다. DMB 오디오 부호화 기법은 MPEG 표준에 근거하므로 MPEG 표준과 DMB 오디오와의 차이점을 중심으로 살펴보겠다. 그리고 압축된 오디오 정보를 실제 전송하기 위해 다른 정보들과 결합되는 multiplexing 방법을 설명하여 압축된 오디오 정보를 실제 전송하기 위해 필요한 사항들을 검토하겠다. 먼저, 지상파 및 위성 DMB에서 공통적으로 사용되는 AAC 기법에 대해 제2절에서 살펴보겠다. 제3절에서는 BSAC 및 지상파 DMB 오디오 규격을 설명하고, 제4절에서는 SBR 및 위성 DMB 오디오 규격을 설명하겠다. 마지막으로 제5절에서는 성능 비교 및 결론을 제시하겠다.

II. Advanced audio coding

본 절에서는 지상파 및 위성 DMB 오디오의 근간이 되는 MPEG-2/4 오디오에 대해 살펴보겠다. 먼저, MPEG-2 AAC와 MPEG-4 AAC에 대해 설명하고, AAC의 기본적인 부호화 도구들에 대해 간략히 설명하겠다.

1. MPEG-2/4 AAC

AAC는 MPEG-1 오디오와의 후방 호환성을 제거하여 보다 효율적으로 오디오 신호를 압축하기 위한 방법으로 MPEG-2의 part 7[9]으로 표준화되

었다. 그 후, MPEG-4 오디오[10]에서 PNS 등 몇 가지 tool들이 추가되어 일반 오디오 부호화 표준으로 정착되었다. MPEG-2 AAC는 채널 당 약 64Kbps의 비트율로 ITU-R의 방송 품질 조건을 만족시킬 수 있는 고품질 오디오 부호화 방식이다. MPEG-2 AAC는 8 ~ 96KHz의 표본화주파수를 가지는 모노, 스테레오 뿐만 아니라 5.1, 7.1 혹은 그 이상의 다채널 오디오 신호를 채널 당 최대 288Kbps (48KHz 표본화주파수 기준)까지의 전송률로 부호화할 수 있다. MPEG-2 AAC에서는 필터뱅크, TNS, 예측 등 12개의 부호화 도구 (coding tool)들이 정의되어 있는데, 이들의 조합에 따라 main, low-complexity (LC), scalable sampling rate (SSR) 등 3개의 프로파일 (profile)을 정의하고 있다.

MPEG-4 오디오[10]에서는 MPEG-2 AAC에서 정의된 것들을 포함하여 약 20여개의 부호화 도구들로 확장하여, 보다 다양한 응용에 적용될 수 있도록 하였다. 즉, PNS (perceptual noise substitution), LTP (long term prediction) 등의 부호화 도구들을 추가하였고, 이들의 조합에 따라 AAC main, AAC-LC, AAC SSR, AAC LTP AAC scalable 등의 부호화 객체 (coding object)들을 구성한다. 또한, 유사한 기능 및 응용 분야를 가지는 객체들로 구성되는 프로파일들을 정의하는데, AAC 관련 객체들은 natural audio profile에 속하게 된다. 각 프로파일은 CPU 및 메모리 사용도 등에 따라 다수의 레벨 (level)로 세분된다. 예를 들어 natural audio profile에는 표본화 주파수 및 CPU 사용도 등에 따라 4개의 서로 다른 레벨들이 정의되어 있다. 즉, MPEG-4 오디오에서는 객체, 프로파일, 레벨 등 복잡도와 사용하는 부호화 도구 등에 따라 오디오 부호화 기법들을 분류해 놓았다. 한편, MPEG-4

오디오 수정규정 [11]에서는 추가적으로 SBR 객체를 도입하고, AAC 프로파일과 HE-AAC 프로파일을 추가적으로 정의하였다. AAC 프로파일은 AAC LC 객체만으로 구성되고, HE-AAC 프로파일은 AAC 프로파일의 모집합 (super set)으로서 AAC-LC와 SBR 객체로 구성된다. 그리고 이 두 프로파일들은 복잡도에 따라 5개의 레벨로 나누어진다.

2. AAC의 주요 부호화 도구들

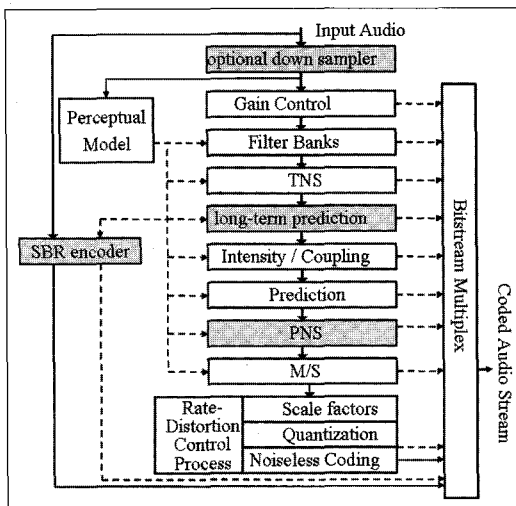
〈그림 1〉에서는 MPEG-4 AAC 부호기의 구조를 제시하였다. 〈그림 1〉에서 색칠된 블록은 MPEG-4 오디오에서 새롭게 도입된 부호화 도구를 나타낸다. 특히, 간축기 (down sampler)와 SBR 부호화기는 SBR 객체를 위한 부분이다. 그리고 이득 제어 (gain control)과 예측 도구는 각각 AAC SSR 및 AAC main에서만 사용된다. 본 고에서는 기본적인 부호화 도구인 심리 음향 모델 (perceptual model), 필터 बैं크, TNS, intensity/MS 스테레오, 그리고

양자화 및 비손실 부호화 부분을 중심으로 설명하겠다.

입력된 오디오 신호에 대해 먼저 심리 음향 모델에서 최소 차폐 문턱 값 (minimum of masking threshold)과 심리음향 엔트로피 (psycho-acoustic entropy)를 측정한다. 심리 음향 모델에서는 큰 주파수 성분이 있을 때 인접한 작은 크기의 주파수 성분은 차폐되어 청취되지 않는다는 사람 귀의 차폐 효과 (masking effect)를 이용한다. 최소 차폐 문턱 값은 사람 귀의 특성을 고려하여, 주파수 영역에서 각 임계 대역 (critical band)에서의 입력 크기에 따른 주변 임계 대역에서의 잡음의 가청 문턱값 (audible threshold)을 계산하고, 이를 임계 대역 별로 합을 구함으로써 얻어진다. 심리 음향 엔트로피는 최소 차폐 문턱값을 고려하였을 때 입력 신호의 엔트로피를 구한 것이다. 최소 차폐 문턱값은 양자화에서 주파수 가중 비트 할당 (frequency weighted bit allocation)에 사용하여 양자화 잡음이 차폐 효과로 인해 사람 귀에 청취되지 않도록 하는 역할을 한다. 그리고 심리 음향 엔트로피는 입력 신호의 non-stationarity와 밀접한 관계가 있어 이 값의 크기에 따라 MDCT의 변환 길이를 결정한다.

필터 बैं크에서는 MDCT를 이용하여 시간 영역 데이터를 주파수 영역 데이터로 변환한다. 먼저, 시간 영역에서 직전 프레임과 1/2 만큼 중복이 되도록 2048 샘플을 취하고, 블록 경계 효과를 감소시키기 위해 윈도우를 취한다. 그런 다음 2048 샘플로부터 1024 주파수 계수를 얻는 MDCT를 수행한다. 이때, 심리 음향 엔트로피의 크기에 따라 길이가 짧은 윈도우 및 변환을 수행할 수도 있다.

다음으로, 변환 계수들에 대해 TNS를 적용한다. AAC와 같이 블록 단위로 부호화하는 방식에서 양자화 잡음은 프레임 전체에 동일한 크기로 분포하



〈그림 1〉 MPEG-4 AAC의 개략적인 구조

게 된다. 만약 프레임 내에서 신호의 크기가 급격하게 변한다면 신호의 크기가 작은 부분에서는 양자화 잡음의 크기가 신호의 크기보다 커질 수 있다. 특히, 작은 크기의 신호 뒤에 큰 신호가 나올 경우, 사람 귀의 1-2msec로 비교적 짧은 사전 차폐 (pre-masking) 범위 밖에서 큰 양자화 잡음이 발생할 수 있다. 이러한 양자화 잡음은 오디오의 품질을 심각하게 저하시키게 된다. TNS는 주파수 영역에서 개루프 (open loop) DPCM을 수행하는 것으로, 이를 통해 시간 영역에서 신호의 포락선에 비례하는 양자화 잡음을 산출시킨다. 즉, 신호가 급격하게 변할 때, 신호의 크기가 큰 부분에서는 비교적 큰 양자화 잡음을 발생시키지만, 신호의 크기가 작은 부분에서는 신호의 크기에 비례하여 작은 크기의 양자화 잡음을 발생시킨다. 이러한 과정을 통해 TNS는 사전 차폐 (pre-masking) 범위 밖에서 큰 양자화 잡음이 발생하는 것을 방지한다.

TNS에서 출력된 데이터는 intensity 및 M/S 부호화 도구에서 스테레오 부호화 기법이 적용된다. 심리 음향 모델에 따르면 오디오 신호의 방향을 감지하는 방법은 주파수에 따라 달라진다. 즉, 약 4000Hz 이상의 주파수 신호에 대해서는 두 채널 사이의 크기 차 (intensity difference)로 방향을 감지하며, 그 이하의 주파수 신호에 대해서는 시간 차 (time difference)로 방향을 감지한다. 이런 사실을 참조하여, 일정 주파수 이상의 변환 계수에 대해서는 왼쪽 채널에만 변환 계수들을 표현하고, 오른쪽 채널은 왼쪽 채널과의 크기 차이만을 표현한다. 이러한 과정은 intensity 부호화 도구에서 수행된다. Intensity 부호화 되지 않은 저주파 성분은 M/S 부호화 도구에서 두 채널의 합으로 표현되는 middle 과 두 채널의 차로 표현되는 side 신호로 변환되어 부호화된다.

앞에서 설명한 과정에 따라 처리된 변환 계수들은 마지막으로 양자화 및 비손실 부호화를 통해 이진 비트열로 표현되고, 다른 부가 정보들과 합해져서 오디오 비트열이 산출된다. 양자화는 심리 음향 모델에서 얻은 최소 차폐 문턱값 이하로 양자화 잡음을 유지시킬 수 있도록 수행된다. 양자화에서는 복잡도를 감소시키기 위해 변환 계수들을 scalefactor와 양자화된 변환 데이터 등 2가지 종류로 나눈다. Scalefactor는 변환 계수의 크기 및 양자화 step size에 관한 정보를 나타내는데, 전체 변환 계수들을 11-51개의 그룹으로 나누고 그룹 단위로 하나의 값을 할당하고, DPCM 후 호프만 부호화한다. 양자화된 변환 데이터는 그 크기 등에 따라 2개 혹은 4개씩 묶은 다음, 11개의 호프만 부호책 (codebook) 중의 하나를 선택하여 호프만 부호화한다.

III. BSAC과 지상파 DMB 오디오

지상파 DMB 오디오에서는 MPEG-4 ER BSAC [10]을 표준으로 사용한다. 지상파 DMB 오디오는 몇 가지 제한 조건을 두었지만, 비교적 충실히 MPEG-4 규격을 준수하고 있어, 국내외 다른 멀티미디어 서비스와의 호환성이 높으며, 부가서비스를 추가적으로 결합시키기 유리하다는 장점이 있다. 본 절에서는 BSAC에 대하여 살펴보고, 지상파 DMB 오디오 규격에 대해 알아보겠다.

1. BSAC

MPEG-4 오디오 표준 ISO/IEC 14496-3:2001[10]의 제4절 일반 오디오 부호화에는 AAC

및 twin-VQ와 함께 BSAC이 포함되어 있다. BSAC은 부호화기의 마지막 부분인 비손실 부호화 부분을 제외하면 모든 부분이 AAC와 동일하다.

BSAC은 AAC에 fine-grain scalability를 제공한다. 이전의 오디오 계층 부호화 (scalable coding)에서는 다른 종류의 부호화기를 종속적으로 연결하거나 동일한 부호화기에서 여러 개의 양자화기를 설치하여, 계층이 추가됨에 따라 향상된 품질을 제공할 수 있도록 하는 방법들이었다. 이러한 방법에서 계층별 산출되는 비트량의 크기 즉 granularity가 큰 large step scalability를 제공한다. 일반적으로 이러한 계층 부호화에서는 각 계층에서 발생하는 비트량은 8Kbps 이상이다. 반면에 BSAC에서는 각 계층에서 발생하는 비트량을 1Kbps 단위로 조절할 수 있는 fine-grain scalability를 제공한다.

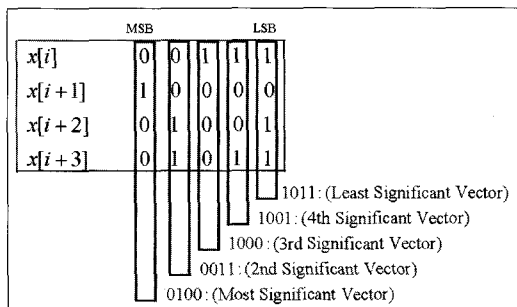
구체적으로 살펴보면, BSAC에서는 fine-grain scalability를 구현하기 위해 양자화된 계수들을 비트 층 (bit plane)으로 분해하고, 각 비트 층들을 산술부호화 (arithmetic coding)한다. <그림 2>에서는 $x[i]$ 부터 $x[i+3]$ 까지 4개의 양자화된 계수를 부호화할 때 BSAC의 기본 개념을 도시하였다. AAC에서는 $x[i]$ 부터 $x[i+3]$ 까지 4개의 양자화된 계수들을 2개 혹은 4개씩 묶어 해당하는 호프만 코드를 할당함

으로써 부호화한다. 이에 반해 BSAC에서는 먼저, 각 양자화된 계수들을 MSB부터 LSB까지 비트 단위로 분해하여 <그림 2>의 아래 부분에 도시한 바와 같이 MS (most significant) 벡터부터 LS (least significant) 벡터까지 bit-sliced 벡터들을 구성한다. 그런 다음, 각 bit-sliced 벡터들을 개별적으로 이진 산술 부호화 (binary arithmetic coding)한다.

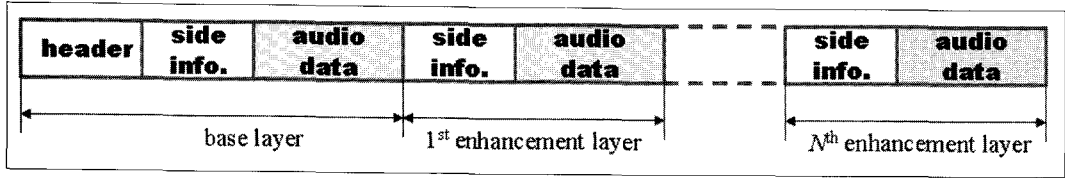
BSAC 압축 비트열은 <그림 3>에 제시한 바와 같이 MSB 층(plane)을 포함하는 기저 계층 (base layer)과 MSB 층을 제외한 다른 비트 층으로 구성되는 향상 계층 (enhancement layer)들로 구성된다. 이때, 향상 계층 데이터는 채널당 1Kbps 단위로 분할하여 표현할 수 있다. 따라서 BSAC을 이용한다면 채널 트래픽 상황에 따라 미세하게 전송 데이터량을 조절할 수 있는 fine-grain scalability를 성취할 수 있다. 또한, 오디오 신호를 계층화된 비트열로 표현하므로, 전송 오류가 발생하였을 때, 전송 오류를 해당 계층에 제한시킬 수 있다. 이러한 점에서 BSAC은 전송 오류가 발생하였을 때에도 오디오 품질을 일정 수준 이상 유지시킬 수 있어, 전송 오류 발생 확률이 높은 환경에 적합한 방법이라고 알려져 있다.

2. 지상파 오디오 규격

초단파 디지털 라디오 방송, 즉 지상파 DMB[12]에서는 MPEG-1 layer II와 함께 앞에서 설명한 BSAC을 이용하여 오디오 정보를 표현한다. 지상파 오디오에서는 ISO/IEC 14496-3:2001[10] 표준 중 객체형 식별자 (object type ID) 22인 ER BSAC에 관한 규정을 준수하되 다음과 같은 몇 가지 제한 조건을 둔다. 첫 번째로 오류 보호 (error protection) 및 오류 강인성 (error resilience) 도구



<그림 2> BSAC의 bit-slicing 개념



〈그림 3〉 BSAC의 계층별 비트열 구성

를 사용하지 않는다. 두 번째로, 프레임의 크기는 1024 샘플로 고정하고, 부호화기 내에서 표본화 주파수는 항상 동일하다고 가정한다. 마지막으로 MPEG-4 오디오 제4절 일반 오디오 부호화에 정의된 부호화 도구 중에서 복잡한 계산량을 요구하는 LTP는 사용하지 않는다. 이때 사용할 수 있는 부호화 도구는 필터뱅크, TNS, 예측, PNS, intensity 및 M/S 스테레오 기법 등이다.

지상파 DMB에서 1024 샘플로 이루어진 1 프레임을 BSAC으로 압축한 데이터인 raw_data_block()은 하나의 SL (synchronization layer) 패킷 [15]으로 패킷화 된다. 그런 다음, MPEG-2 PES (packetized elementary stream) 패킷으로 변환되고 MPEG-2 TS (transport stream) 패킷이라는 전송 양식으로 표현되어 전송된다. 이때, 1 프레임에 대한 압축된 비트열인 raw_data_block()에는 부호화 형식, 표본화주파수, 채널수 등 오디오 비트열 복호화에 필요한 복호기 구성 정보들이 없다. 이러한 복호기 구성 정보는 MPEG-4 시스템[15]의 OD (object descriptor) 및 IOD (initial OD) 내에 정의되는 ES 서술자 (descriptor)의 일부인 DecoderConfiguration Descriptor에 서술된다. 이것은 오디오의 경우 AudioSpecificConfig() 서술자로 표현되며 이 속에 GASpecificConfig() 서술자가 포함된다. AudioSpecificConfig 서술자에는 AAC LC와 BSAC 등 오디오 부호화에 적용된 MPEG-4

오디오 객체형, 표본화주파수, 채널수, 오류 보호 사용여부 등의 정보가 수록된다. GASpecificConfig 서술자에는 프레임 길이와 BSAC 복호화에 필수적인 서브프레임 개수(number of sub-frames) 및 계층의 평균 길이 등의 정보가 수록된다.

따라서 오디오 복호화를 수행하기 위해서는 먼저, TS 비트열에서 PMT의 일부로 포함된 IOD를 검색하여 AudioSpecificConfig() 및 GASpecificConfig() 서술자를 해석한다. 이로부터 복호기 구성 정보를 취득한 다음, 오디오 TS 패킷을 모아 PES 패킷을 구성한다. 그리고 PES 패킷과 SL 패킷 순서로 복호화하여 BSAC raw_data_block()을 재구성한 다음, 이를 복호화하여 복원된 PCM 오디오 데이터를 얻는다.

IV. SBR 기법과 위성 DMB 오디오

위성 DMB 오디오에서는 ISO/IEC 14496-3:2001 Amd. 1[11]에 정의된 HE-AAC (high efficiency AAC) 프로파일에 관한 규정을 따르되, 몇 가지 제한 조건을 두었다. 그리고 오디오 객체형을 알려주는데 필요한 MPEG-4 시스템 규정을 사용하지 않기 때문에 수신된 오디오 비트열은 반드시 HE-AAC 규정에 적합한 것이어야 한다. 또한, 다른 미디어와의 호환성이 없고, 부가 서비스

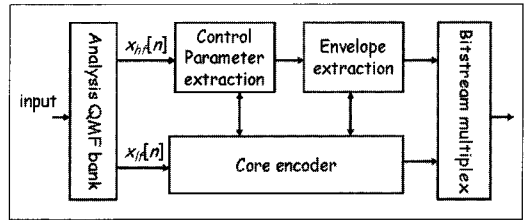
등을 추가하기가 어렵다는 단점이 있다. 본 절에서는 위성 DMB 오디오의 근간이 되는 SBR 기법을 살펴보고, 위성 DMB 오디오 규격을 구체적으로 설명한다.

1. SBR 기법

SBR 알고리즘은 독일의 Coding Technology사에서 개발한 방법으로, 대역폭 확장 (bandwidth extension) 개념을 사용하여 청각 오디오 부호화기의 압축 효율을 큰 폭으로 향상시키는 것으로 알려져 있다.

SBR 기술의 기본원리는 오디오 신호의 고주파 성분은 그에 대응되는 저주파 성분과 매우 높은 상관성을 가진다는 가정에 기반을 둔다. 즉, 사람 목소리와 악기들은 각각 성분(vocal cords) 및 현(string) 등의 진동에 의해 넓은 주파수 범위에서 분포하는 고조파 성분(harmonics)들을 포함한다. 이 고조파 성분들은 준정상 여기 신호(quasi-stationary excitation signal)로 모델링 되며, 사람의 발성 기관 등 공명 장치에 의해 필터링 되어 음색(tone color) 등 고유의 특성을 가지게 된다. 만약, 대역 제한에 의해 높은 주파수 영역의 고조파 성분들이 삭제된다면, 건조하고 풍부하지 못한 음향이 되어버린다. 그러나 사람 목소리와 악기들이 고조파 성분들로 구성된다는 점을 참조한다면, 저주파 성분들을 이용하여 효과적으로 고주파 성분들을 추정할 수 있다.

<그림 4>에서는 SBR을 이용한 부호화기의 개략적인 구조를 나타내었다. <그림 4>에 보인 바와 같이, 먼저 QMF (quadrature mirror filter) 분석을 통해 입력 오디오 신호는 대역폭을 반으로 나누어 저주파 성분 및 고주파 성분으로 분해된다. 여기서,



<그림 4> SBR 기반 부호화기의 블록선도

고주파 성분은 대역폭 확장을 통해 저주파 성분을 재사용하여 표현한다. 즉, 저주파 성분을 주파수 영역에서 고주파 부분으로 이동시킨다. 그런 다음, 몇 가지 부가 정보를 추가함으로써 이동된 신호가 원래의 고주파 성분과 동일한 특성을 갖도록 조정한다. 좀 더 구체적으로 살펴보면, <그림 4>의 제어 요소 추출(control parameter extraction) 블록에서는 저주파 성분을 고주파 성분으로 전치(transposition) 하는데 필요한 제어정보들을 추출한다. 제어정보로는 전치하고자 하는 저주파 및 고주파 성분의 위치들과 저주파 성분으로부터 추출할 수 없는 고주파 성분의 특징 등을 포함한다. 다음으로, 포락선 추출(envelope extraction) 블록에서는 고주파 성분의 포락선을 추출한다. 여기서 추출된 포락선을 이용하여 전치된 신호의 크기를 조정한다.

저주파 성분은 MP3 혹은 AAC 등 중심 부호화기(core encoder)로 부호화한다. 이때, 중심 부호화기에서 처리되는 오디오 신호는 전체 대역 신호가 아니라 저주파 대역만으로 구성되는 간축된(downsampled) 신호이기 때문에 적은 비트량으로 부호화할 수 있게 된다. 또한, 고주파 성분은 SBR과 관련된 제어 정보 및 포락선 정보 등을 포함하는 부가 정보만으로 구성되므로 채널당 약 1-3Kbps라는 아주 적은 비트량으로 표현된다. 따라서 SBR 기반 부호화기는 기존의 MP-3 및 AAC 부호화기

와 함께 사용하여 부호화 효율을 약 2배 정도 향상시키는 것으로 알려져 있다.

〈그림 5〉에서는 중심부호화기로 AAC를 사용하였을 때 SBR 기반 복호화기의 구조를 제시하였다. SBR 기반 복호화기에서 SBR 부분은 중심 복호기에 대한 후처리기와 같이 작동한다. 먼저 부호화된 오디오 비트열이 입력되면 bitstream payload deformatter에서 입력 비트스트림을 SBR 비트열과 AAC 비트열로 분리한다. 그런 다음, AAC 중심 복호화기에서는 AAC 비트열로부터 주파수 영역 오디오 신호를 복원하고, 분석 QMF बैं크에서는 고주파 대역으로의 전치를 위해 저주파 대역 성분을 계산한다. 한편, SBR 비트열은 bitstream parser에 의해 SBR 데이터로 복원된다. SBR 데이터는 호프만 복호화 및 역양자화에서 고주파 대역의 전치 및 포락선 조정에 필요한 정보가 계산되고, 이 정보들은 HF 발생기와 포락선 조정기로 전달된다. HF 발생기에서는 전치를 이용해 고주파 영역 신호를 생성하고, 이는 다시 포락선 조정기에서 포락선이 조정된다. 이렇게 생성된 고주파 영역 신호는 AAC 복호화기에서 얻어진 저주파 신호와 합성 QMF बैं크

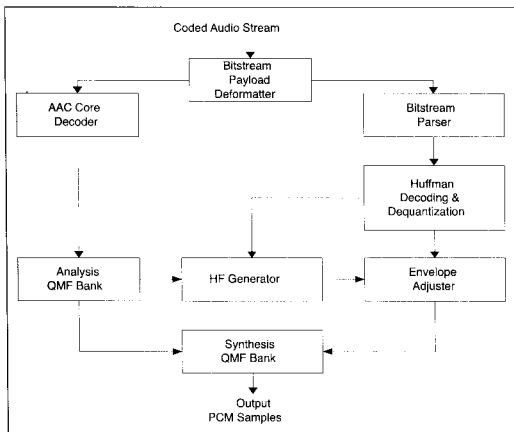
크를 통해 결합되어 시간영역 전대역 신호로 산출된다.

MPEG-4 오디오에서는 2003년 ISO/IEC 14496-3 Amd 1[1]에 HE-AAC 프로파일과 SBR 객체를 정의함으로써 SBR 기법을 일반 오디오 부호화 표준의 하나로 포함시켰다. HE-AAC 프로파일에는 AAC LC 객체와 SBR 객체만으로 구성되는데, 즉, SBR 객체를 사용하여 부호화 할 때 저주파 성분을 부호화하는 객체는 반드시 AAC LC 객체이어야 한다.

2. 위성 DMB 규격

위성 디지털멀티미디어 방송[13], 즉 위성 DMB를 위한 오디오는 ISO/IEC 13818-7[9]에 정의된 AAC LC 프로파일과 ISO/IEC 14496-3: 2001 Amd. 1[1]에 정의된 SBR 객체 규정을 따른다. ISO/IEC 14496-3: 2001 Amd. 1에서는 AAC LC 객체와 SBR 객체로 구성되는 HE-AAC 프로파일을 정의하고 있는데, SBR 객체는 AAC LC 객체와 함께 사용하도록 제한하고 있다. 그런데, MPEG-4 AAC LC 객체는 MPEG-2의 AAC LC 프로파일과 비교하여 PNS가 추가된 것이다. 따라서 MPEG-2 AAC LC 프로파일과 SBR 객체를 사용하도록 정의하고 있는 위성 DMB 오디오 규정은 MPEG-4 HE-AAC 프로파일 규정에 부합된다고 볼 수 있다.

위성 DMB 오디오에서는 중심 부호화기로 MPEG-2 AAC LC 프로파일을 사용하고, 고주파 성분을 효과적으로 부호화하기 위해 선택적으로 SBR 객체를 사용한다. 채널당 1024 샘플로 구성되는 access unit 1개에 해당되는 오디오 데이터들은 raw_data_block() 이라는 순수 오디오 비트열로 표현된다. 이때, AAC LC 프로파일에 의해 2채널 오



〈그림 5〉 SBR 기반 복호화기의 블록선도

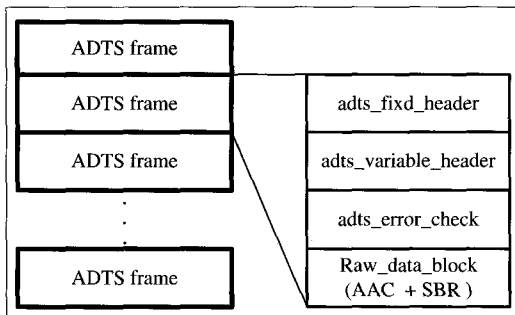
디오는 2개의 single_channel_element() 혹은 1개의 channel_pair_element()로 표현된다. 만약 SBR 객체가 사용된다면 SBR 데이터는 raw_data_block()의 한 요소인 fill_element() 내에 정의되는 extension_payload의 하나로 포함된다. 즉, SBR 데이터는 별개의 정보로 표현되는 것이 아니라 AAC 데이터의 일부로서 표현된다.

그런 다음, 1024 샘플 × 채널수의 오디오 데이터를 압축한 비트열인 1개의 raw_data_block()은 1개의 ADTS (audio data transport stream) 프레임으로 패킷화 된다. ADTS는 ISO/IEC 13818-7[9]에 규정되어 있고, ADTS로 패킷화된 오디오 데이터의 양식은 <그림 6>에 도시하였다. <그림 6>에 도시한 바와 같이 1개의 ADTS 프레임은 ADTS 고정 헤더 (fixed header)와 가변 헤더 (variable header), 오류 확인 부호 (error check code), 그리고 AAC 데이터 및 SBR 데이터를 포함하는 1개의 raw_data_block()으로 구성된다. ADTS 고정 헤더에는 MPEG-2 AAC임을 나타내는 플래그, 표본화 주파수, 채널수 등에 관한 정보들이 있으며, 가변 헤더에는 ADTS 프레임의 바이트 단위 길이와 현재 ADTS 프레임에 포함된 raw_data_block()의 개수 등의 정보들이 포함된다. 그리고 선택적으로 헤더의 오류

유무를 확인하기 위한 CRC_check이 포함된다.

<그림 6>과 같은 형태로 표현된 ADTS 프레임들은 PES 패킷 및 TS 패킷으로 순차적으로 변환되어 전송된다. 오디오 TS 패킷의 식별은 PSI (program specific information)의 하나인 PMT (program map table)에 스트림 형을 MPEG-2 AAC audio with ADTS (0x0F)로 정의한 식별자로 구별하며, PMT에서 오디오 스트림을 정의할 때 표 2와 같은 구조의 AAC_descriptor를 포함시킬 수 있다. AAC_descriptor에는 오디오 스트림 전체의 전송률, AAC의 표본화 주파수, 채널수, 그리고 SBR 사용여부 등에 관한 정보들이 포함된다.

위성 DMB에서는 1채널 혹은 2채널 오디오를 표본화 주파수 24, 32, 혹은 48KHz로 표본화하여 부호화한다. 오디오 비트율은 표 3에 제시한 바와 같이 채널수 및 표본화 주파수에 따라 달라지는데, 128Kbps를 넘지 않는 것으로 규정된다. 부호화는 MPEG-2 AAC LC 혹은 MPEG-2 AAC LC + SBR로 이루어지는데, 사용된 부호화 기법의 식별은 MPEG-4에 규정된 묵시적 신호 방식 (implicit signaling of SBR)에 따른다. 즉, 오디오 비트열에 extension_payload의 일부로 SBR 데이터가 포함되어 있으면, MPEG-2 AAC LC + SBR로 부호화



<그림 6> ADTS 프레임의 구조

<표 2> 위성 DMB에서의 AAC_descriptor.

syntax	no. of bits	value
AAC_descriptor () {		
descriptor_tag	8	201
descriptor_length	8	2
sampling_frequency_index	4	0x3, 5, or 6
channel_number	4	1 or 2
SBR_flag	1	0 or 1
reserved	7	
}		

된 것이고, 그렇지 않으면 MPEG-2 AAC LC 만으로 부호화된 것으로 판별한다. 만약, SBR 데이터가 포함되어 있다면, 최종적인 오디오 표본화 주파수는 ADTS 헤더 및 AAC_descriptor에 명시된 표본화 주파수의 2배가 된다.

V. 부호화기의 성능 및 결론

본 고에서는 지상파 및 위성 DMB를 위한 오디오 부호화 기법에 대하여 살펴보았다. DMB에서 오디오 서비스 요구 사항은 표 1에 제시한 바와 같이 2채널 스테레오 오디오를 128Kbps 이하로 표현할 수 있어야 한다는 것이다. 이와 같은 요구 조건을 만족시키기 위해 지상파 DMB에서는 AAC에서 비손실 부호화 기법을 호프만 부호화에서 산술 부호화로 바꾼 MPEG-4 BSAC을 오디오 부호화 기법으로 규정하였다. 그리고 위성 DMB에서는 MPEG-2 AAC 및 선택적으로 MPEG-4 SBR을 사용하는 방

〈표 3〉 위성 DMB 오디오의 전송률

sampling frequency (Hz)	모노(Kbps)	스테레오(Kbps)
48000	24-96	56-128
32000	24-96	40-128
24000	16-56	24-56

법을 오디오 부호화 기법으로 규정하였다. 그러나 지상파 및 위성 DMB에서 실제 사용되는 부호화 기법은 MPEG-4 오디오 규격과는 약간의 차이를 보인다. 표 4에서는 DMB 오디오 부호화 기법과 관련된 MPEG-4 오디오 부호화 객체들과 지상파 및 위성 DMB 오디오에서 사용 가능한 부호화 도구들을 표기하였다. 표 4를 참조하면 지상파 DMB 오디오는 LTP와 오류 보호 및 오류 강인성 구문(syntax)을 사용하지 않는 것을 제외하면, MPEG-4 BSAC과 동일하다. 위성 DMB 오디오는 MPEG-4 AAC LC 객체 및 SBR을 결합한 것에 비해 PNS를 사용하지 않는 것만 차이난다는 것을 알 수 있다. 그리고 지상파 및 위성 DMB에서 공히 프레임 크기는 1024

〈표 4〉 오디오 부호화 객체에 따른 부호화 도구 사용

tool/modules object type	gain control	block switching	filter bank-standard	filter bank SSR	frame length of 960	TNS	LTP	Intensity stereo	coupling	prediction	PNS	MS stereo	AAC noiseless decoding	BSAC noiseless decoding	SBR	ER payload syntax	EP tool	MPEG-4 object type ID
AAC LC		x	x		x	x		x	x		x	x	x					2
AAC main		x	x		x	x		x	x	x	x	x	x					1
AAC SSR	x	x		x	x	x		x	x		x	x	x					3
BSAC		x	x		x	x	x	x			x	x		x		x	x	22
SBR															x			5
지상파 DMB		x	x			x		x			x	x		x				22
위성DMB		x	x			x		x	x			x	x		x			?

〈표 5〉 스테레오 오디오 신호에 대한 부호화기별 비트율

부호화 기법	MPEG-1 layer II	MPEG-1 layer III	MPEG-2 AAC	MPEG-4 AAC BSAC	AAC + SBR
cd quality		192Kbps	128Kbps		
broadcast quality	192Kbps	128Kbps	96Kbps	64Kbps	48Kbps

샘플로 제한하여 960 샘플 크기 프레임은 사용하지 않는다.

각 부호화 방법의 전송률에 따른 오디오 품질은 표 5에 제시하였다. 여기서, 지상파 DMB는 MPEG-4 BSAC, 그리고 위성 DMB는 AAC + SBR과 동등한 수준의 품질을 보인다. 표 5에 따르면 지상파 및 위성 DMB는 128Kbps 이하에서 CD 수준의 오디오

품질을 유지할 수 있다. 특히, 아날로그 FM 수준의 방송용 품질은 지상파 DMB에서는 약 64Kbps, 그리고 위성 DMB에서는 약 48Kbps에서 성취할 수 있다. 따라서 지상파 및 위성 DMB에서는 제한된 전송률로 오디오 서비스 요구사항을 만족시킬 수 있음을 알 수 있다.

● 참고 문헌 ●

- [1] K.C. Pohlmann, Principles of Digital Audio, 4th Ed., McGraw-Hill, 1999.
- [2] D.T. Yang, C. Kyriakakis, and C.C. Jay Kuo, High-Fidelity Multichannel Audio Coding, Hindawi Pub. Co., 2004.
- [3] M.A. Krasner, "Digital encoding of speech and audio signals based on the perceptual requirements of the auditory system, Technical report 535 of the MIT Lincoln Lab., Lexington, 1979.
- [4] J. Princen and A. Bradley, "Analysis/ synthesis filter bank design based on time domain aliasing cancellation," IEEE Trans. on ASSP, vol. 34, no. 5, pp. 1153-1161, Oct. 1986.
- [5] ISO/IEC 11172-3, Information Technology - Coding of moving pictures and associated audio for digital storage media at up to about 1.5 Mbit/s - Part 3: Audio, 1993.
- [6] ISO/IEC 13818-3, Information Technology - Generic coding of moving pictures and associated audio information - Part 3: Audio, 1994.
- [7] ATSC standard A/53, ATSC Digital Television Standard, 1995.
- [8] 김인철 "DTV 기술 개요" 대한전자공학회지, 제28권, 제11호, pp. 16-22, 2001년11월.
- [9] ISO/IEC 13818-7, Information Technology - Generic coding of moving pictures and associated audio information - Part 7: Advanced Audio Coding, 3rd Ed., 2004.
- [10] ISO/IEC 14496-3, Information technology - Coding of audio-visual objects - Part 3: Audio, 2001.
- [11] ISO/IEC 14496-3, Information technology - Coding of audio-visual objects - Part 3: Audio, Amd. 1: Bandwidth extension, 2003.
- [12] TTAS.KO-07.0026 초단파 디지털라디오방송 비디오송수신 정합표준, 한국정보통신기술협회, 2004년 8월.
- [13] TTAS.KO-07.0027, 위성 디지털멀티미디어방송 송수신 정합표준, 한국정보통신기술협회, 2004년 9월.
- [14] 박진수, 김지희, 이승준, "모바일 방송용 비디오 및 오디오 압축기술," 대한전자공학회지, 제 32권 제5호, pp. 561-569, 2005년 5월.
- [15] ISO/IEC 14496-1, Information technology - Coding of audio-visual objects - Part 1: Systems, 2002.

필자 소개



김인철

- 1985년 : 서울대학교 제어계측공학과 학사 졸업.
- 1987년 : 서울대학교 제어계측공학과 석사 졸업.
- 1992년 : 서울대학교 제어계측공학과 박사 졸업.
- 1992~1994년 : 대우전자(주) 영상연구소 선임연구원
- 1994~1999년 : 한성대학교 정보전산학부 조교수
- 1999~현재 : 서울시립대학교 전자전기컴퓨터공학부 교수
- 주관심분야 : 신호처리, 오디오 신호처리, 오디오/비디오 압축