

디지털 라디오 오디오 코덱 기술의 과거와 미래

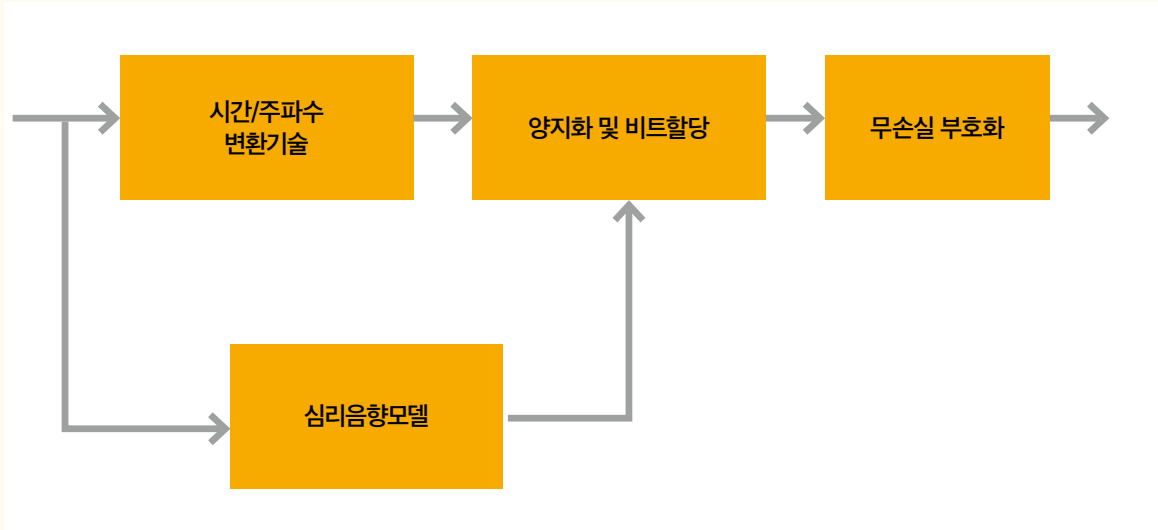
백승권 한국전자통신연구원 오디오연구실

방송의 디지털 전환으로 많은 사람들은 기존의 아날로그 방식과는 차별되는 고품질의 콘텐츠 서비스를 받을 수 있기를 기대한다. 방송사업자의 경우 디지털 전환을 통해 새로운 수익을 창출할 수 있기를 기대할 것이다. 국외에서 먼저 추진된 디지털 라디오 서비스의 도입은 초창기에는 전자보다 후자의 요구사항을 먼저 충족시키고자 하였다. 이는 전통적인 아날로그 방식의 라디오는 FM과 AM을 통해 서로 다른 주파수로 방송되며, 특정 방송을 듣기 위해서 고유의 주파수 값에 맞게 튜닝을 해야 했고, 적은 수의 방송 서비스를 위해서 많은 주파수 대역이 필요하여 사용자들이 들을 수 있는 라디오 방송의 개수는 한정되었다. 그러나 라디오 방송의 디지털 전환은 첫 번째로 이러한 제약을 극복하고 동일한 주파수 대역을 사용하더라도 다수의 라디오 방송 서비스를 제공할 수 있게 되었다. 이는 디지털 전송 방식을 활용하면 다중화(Multiplexing)와 오디오 압축 기술을 통해 하나의 앙상블을 이용하여 다수의 라디오 방송 서비스 전송이 가능해졌기 때문이다. 따라서 디지털 라디오 기술의 핵심은 전송 방식과 오디오 압축 기술이라 할 수 있겠다.

초창기 디지털 라디오 전송 방식은 DAB(Digital Audio Broadcasting) 혹은 유레카(Eureka) 147이라는 명칭으로 1981년에 고안되어 1988년에 독일에서 최초로 전송이 이루어졌다. 이 당시 오디오 코덱 기술은 개념이 정립된 단계로부터 구현이 이루어진 기술로 그 성능이 기대할만한 수준이 아니었다. 또한 적절한 음질을 내기 위해 필요한 연산량을 수행할만한 하드웨어 개발도 더딘 상태였다.

그럼 본격적으로 디지털 라디오에 활용된 오디오 코딩 기술에 대해서 알아보기로 하겠다. 먼저 초창기 적용되었던 오디오 코딩 개념에 대해서 간략하게 살펴보기로 한다.

오디오 코딩 기술은 기본적으로 단방향 오디오 서비스를 고려하여 오디오 신호를 압축/전송하기 위한 기술로 인간이 청각적으로 어떻게 들리느냐에 초점을 맞추고 있다. 이는 음성 코딩 기술이 상호 간에 정보전달을 위해 개발된 것과는 다른 동기를 가진다. [그림 1]은 MPEG에서 개발된 오디오 코딩 기술의 기본 구조를 나타낸 그림이다. 먼저 시간영역의 디지털 오디오 입력 신호는 시간/주파수 변환 기술에 의해 주파수 영역으로 변환된다. 이는 심리음향 모델의 분석 및 적용이 주파수 영역에서 용이하기 때문이다. MPEG에서 주로 사용하는 시간/주파수 변환 기술은 MDCT(Modified Discrete Cosine Transform) 혹은 QMF(Quadrature Mirror Filterbank)이다. MDCT는 DCT 변환기술의 일종으로 DFT(Discrete Fourier Transform)보다 양자화 해야 할 계수의 수를 절반으로 줄일 수 있다. 즉, 단순히 DFT대신 MDCT를 적용한다고 하면, 압축효율이 2배로 증가하는 것이다. 이러한 현상이 발생하는 이유는 오디오 코딩 절차에서는 반드시 수행해야 하는 과정이 이웃한 오디오 신호 프레임 간에 윈도우잉(windowing)과 중첩가중(overlap-add)을 수행해야만 한다. 중첩가중 과정은 프레임 신호의 절반을 중첩시켜야 하는 것으로, 그만큼의 추가적인 데이터 전송을 야기시킨다. 그러나 MDCT는 이러한 중첩과정을 거치더라도 데이터양이 증가하지 않는 특징을 가지고 있다.



[그림 1] MPEG 오디오 부호화기 기본 구조

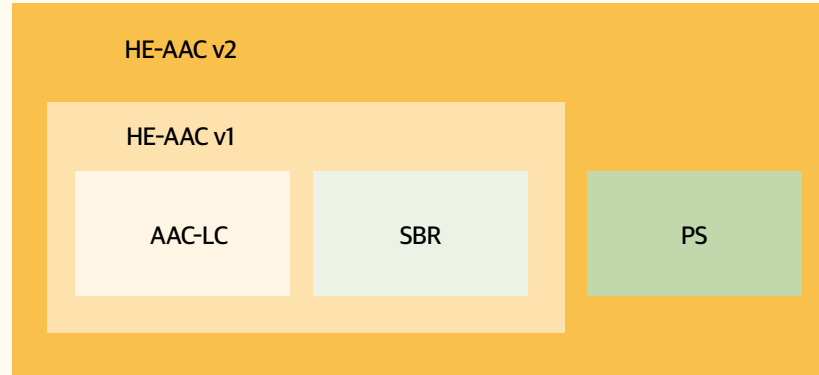
또 다른 변환과정으로 QMF 역시 주파수 변환 기술이지만, 시간 축의 정보도 동시에 지니고 있어서 양자화 시 이를 고려하여 비트 할당을 수행할 수 있다. [그림 1]의 심리음향 모델 역시 입력된 오디오 신호를 주파수 영역으로 변환하여 분석하지만, 분석 이후에 주파수 계수를 활용하지 않기 때문에 분석이 용이한 DFT(Discrete Fourier Transform)를 이용하여 주파수 변환을 수행한다. 인간의 청각은 주파수 영역에서 특정 서브밴드(sub-band) 단위로 인지하며, 그 인지 능력에 맞추어 양자화기를 설계한다. 즉 인간이 인지 못 할 정도의 양자화 잡음이 인위적으로 발생하도록 할 수 있으며, 이를 통해 할당되는 비트량을 줄일 수 있다. 심리음향 모델에서는 양자화를 위한 스케일 팩터(scale factor) 정보를 전달하며 주파수 신호는 스케일 팩터로 나누어진 뒤 주어진 비트 내에서 비트를 할당받게 된다. 예를 들어 스케일 팩터 값이 상대적으로 크다면 양자화 잡음이 많이 발생할 수 있다는 것이며 비트할당을 적게 할 수 있다는 것이고, 심리음향 모델에 의거하여 인간이 인지하지 못하는 수준까지 허용될 수 있다. 마지막으로 양자화된 주파수 계수의 값들은 무손실 부호화를 거쳐 전송된다.

MUSICAM과 같은 초창기 오디오 코딩 기술은 이러한 기본적인 원리를 기반으로 만들어진 코덱 기술이며 압축효율이나 음질이 뛰어나지 못했다. 이러한 이유로, 초창기 DAB 기반의 디지털 라디오 방송 서비스는 디지털 전환에도 불구하고 음질에 대한 문제가 자주 거론되었는데, 일차적으로는 사업자들이 주어진 채널당 다수의 방송 서비스를 위하여 라디오 방송 별로 할당되는 비트율을 낮추었고, 이차적으로는 사용된 오디오 코덱 기술인 MUSICAM이 현재 일반적으로 활용되고 있는 MP3(MPEG Layer III) 코덱보다 그 성능이 매우 떨어지기 때문에 정상적인 음질을 내기 위해서 많은 비트량이 필요하였다.¹⁾

결국 디지털 라디오 방송이 사업자에게는 새로운 수익을 창출하고, 사용자에게는 보다 나은 품질의 오디오 서비스를 제공하기 위해서는 진보된 오디오 코덱 기술이 필요하다.

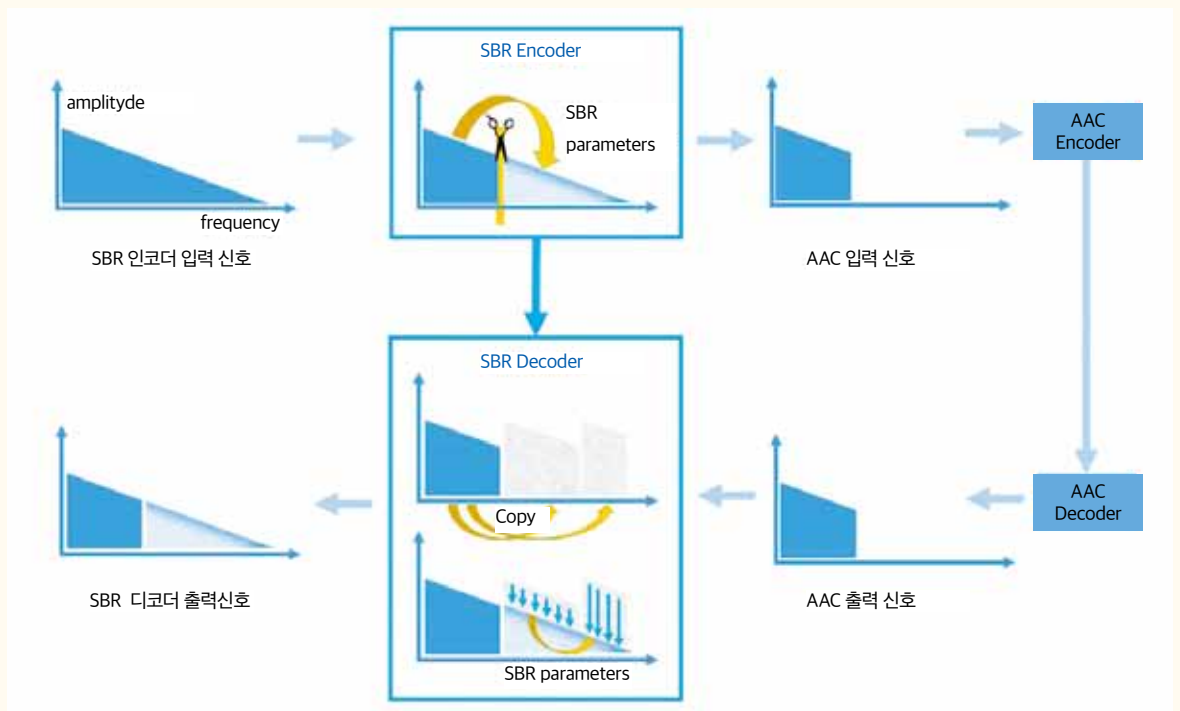
1) MUSICAM이 CD-like한 음질을 내기 위해서는 160kbps 이상의 비트할당이 필요한 것으로 보고되고 있다.

DAB+는 다년간의 이러한 요구사항들을 반영하여 제정된 디지털 오디오 방송 표준의 개정안이며, HE-AAC v2(High Efficiency Advanced Audio Coding version2)라는 새로운 오디오 코덱 기술을 채택하였다. MUSICAM과 HE-AAC v2는 성능 측면에서 매우 큰 차이를 보일 수밖에 없는데, 이는 MUSICAM이 1980년대 활용된 기술인 것에 비해, HE-AAC v2는 2000년대에 표준이 개정된 기술인 것을 감안하여야 할 것이다. 이는 20년간의 오디오 코딩 기술의 진보된 결과가 HE-AAC v2에 반영되어 있다고 보면 되며, MUSICAM과 음질 및 압축 성능에서 큰 차이를 보일 수밖에 없는 것이다.



[그림 2] HE-AAC v2 프로파일

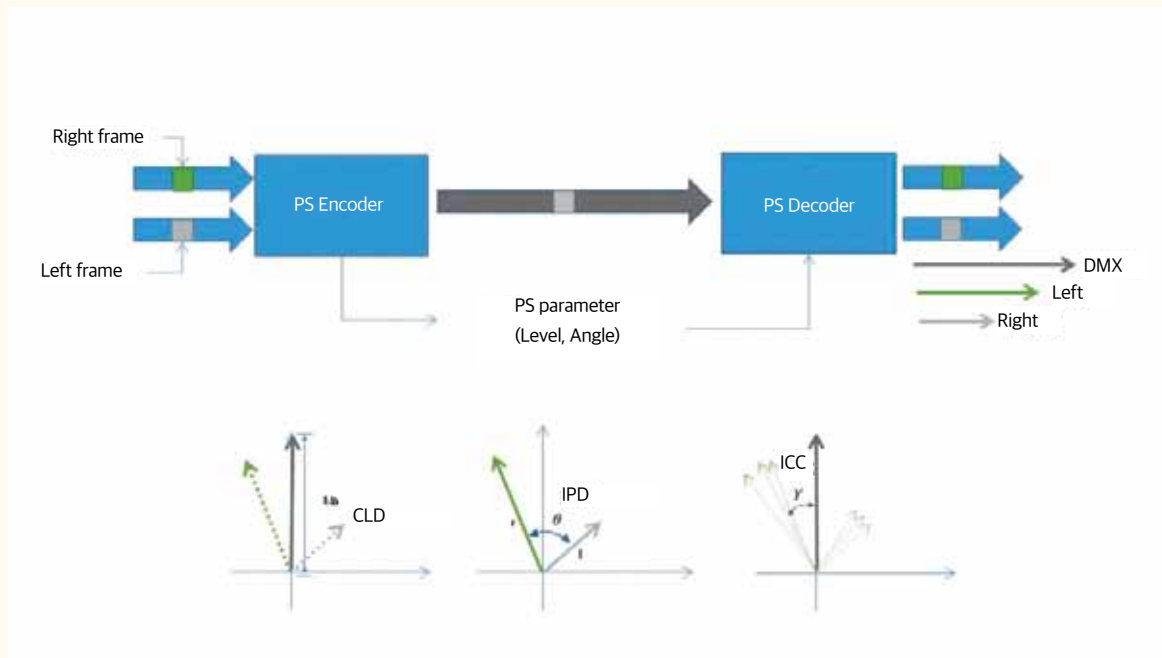
그럼 HE-AAC v2가 어떤 진보된 코딩 기술이 포함되어 있는지 살펴보기로 한다. HE-AAC v2가 version 2인 것을 고려한다면, 이전에 AAC, HE-AAC v1 등이 있었으리라는 것을 예측할 수 있다. 실제로 [그림 2]와 같이 HE-AAC v2 프로파일은 AAC와 HE-AAC v1의 프로파일을 포함하고 있다. AAC는 MP3 이후에 개발된 오디오 코딩 기술로 MP3보다 효율이 높은 오디오 코딩 기술로 이해하면 되겠다. AAC가 MP3 보다 향상된 음질을 보이는 이유는 심리음향 모델 및 양자화기 성능이 개선된 것 뿐만 아니라, 프레임마다 가변적인 비트레이트 할당이 가능해 졌기 때문이다. 이는 오디오 신호에서 비트를 많이 필요로 하는 구간에 적절하게 비트를 가변적으로 할당시킴으로써 적은 비트 할당으로 인하여 발생하는 음질 열화를 개선할 수 있었다. 구조적으로 가장 큰 차이점은 MP3가 MPEG-1 Layer I, II에서 진화한 것으로 호환 구조를 갖기 위해서 QMF와 MDCT를 동시에 쓸 수밖에 없는 하이브리드 형태의 시간/주파수 변환기술을 사용한 것에 반하여, AAC는 MDCT만을 이용하여 양자화 전략을 수행함으로써, 주파수 영역에서 신호의 분해능을 개선하고 신뢰도 높은 양자화를 수행할 수 있었다.



[그림 3] HE-AAC v1 인코딩 디코딩 과정

HE-AAC v1은 AAC에 고주파 대역을 합성할 수 있는 SBR(Spectral Band Replication) 툴을 결합시킴으로써 탄생한 코딩 기술이다. 산술적으로는 AAC보다 2배 압축효율이 높은 코딩 기술이다. SBR 기술은 인간이 덜 민감하게 느끼는 고주파 대역²⁾의 신호는 실제 신호를 양자화 하지 않고, 저대역 신호를 통해 파라미터를 이용하여 합성하는 코딩 기술이다. [그림 3]은 HE-AAC v1의 인코딩 디코딩 과정을 개념적으로 나타낸 것이다. 입력된 신호는 SBR 인코더에서 고주파 신호에 대한 파라미터 정보를 추출하고, SBR 인코더는 저대역 신호만을 출력하여 AAC 인코더의 입력으로 전달한다. SBR 디코딩 과정에서는 AAC 디코더에서 저대역 복원 신호를 입력받아 SBR 파라미터를 통해 저대역 신호를 이용하여 고대역 신호를 만들어 간다. 마치 저대역 신호를 오려 붙이는 것과 같은 역할을 SBR 디코더에서는 수행하며 이는 전송된 SBR 파라미터들이 그 방법을 지시하는 것으로 이해하면 되겠다.

2) SBR에서 다루는 고주파 대역은 입력 신호 샘플링 주파수에 따라 달리 정의되며, 최소 3.5 kHz부터 정의된다.

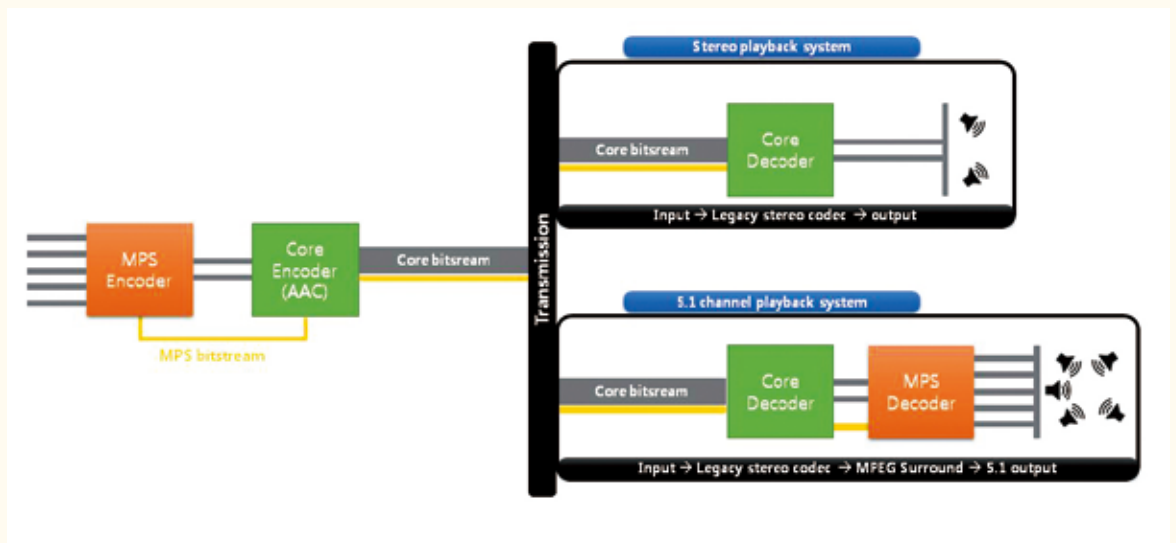


[그림 4] Parametric Stereo 인코딩 디코딩 과정

HE-AAC v2는 스테레오 신호에 대해서 HE-AAC v1보다 2배 압축효율이 높은 오디오 코딩 기술로 산술적으로는 AAC보다 4배 압축이 가능한 오디오 코딩 기술이다. SBR과 유사한 접근 방식으로 스테레오 신호를 압축할 때, 파라미터를 이용하여 채널 신호를 압축하는 기술이다. 이를 PS(Parametric Stereo)라 하며 동작 원리는 간략하게 [그림 4]와 같이 표현하였다. 스테레오 신호를 분석하기 위하여 PS는 입력된 스테레오 신호를 주파수 영역에서 서브밴드 단위로 신호를 분석한다. 각 서브밴드마다 신호 크기 차이, 위상 차이, 상관도 차이를 구한다. 신호 크기 차이는 CLD(Channel Level Difference), 위상차이는 IPD(Interchannel Phase Difference), 상관도 차이는 ICC(InterChannel Coherence)와 같은 파라미터로 정의된다. 이렇게 파라미터를 구하고 나면 스테레오 신호는 모노 신호로 다운믹스된다. 따라서 실제 HE-AAC v1 구조로 부호화되는 신호는 다운믹스된 모노신호만이다. 결국 한 채널 신호를 파라미터로 압축 표현하며 대신하는 것으로, 표현된 파라미터의 정보량도 매우 적기 때문에 HE-AAC v1보다도 2배 높은 압축 효율을 보인다고 할 수 있겠다.

DAB+가 HE-AAC v2를 오디오 코딩 기술로 채택하여 지원할 수 있게 된 것은 사용자 입장에서는 보다 개선된 디지털 라디오 서비스를 즐길 수 있으며, 사업자 입장에서는 고효율 압축 코덱을 사용함으로써, 낮은 비트율에서 사용자의 기대 음질을 만족 시킬 수 있으므로, 하나의 전송 대역폭에 다중화 및 앙상블 기술을 활용하여 여러 개의 라디오 방송을 서비스할 수 있게 되었다. 또 하나의

차별화된 서비스로, DAB+가 채택한 오디오 코덱 기술은 다채널 오디오 서비스를 위한 MPS(MPEG Surround) 기술이다. 이는 일반 라디오 청취자들이 대부분 차량 공간에서 많은 시간을 보내는 것을 고려한다면 이미 다채널 스피커가 내장된 차량 내에서 다채널 오디오 서비스를 제공하는 것은 큰 어려움 없이 차별화된 오디오 서비스를 제공할 수 있을 것이다. 여기서 하나의 제약 사항은 DAB나 DAB+ 전송방식에서 지원하는 오디오 포맷은 모노 혹은 스테레오 신호이므로, 스테레오 채널 이상의 오디오 신호를 전송하는 것은 불가능하다. MPS는 이러한 제약 사항을 준수하면서 모노 혹은 스테레오 신호와 역호환성을 제공할 수 있으며, MPS 디코더가 장착된 단말에서는 다채널 오디오 서비스가 가능하도록 하였다. [그림 5]는 MPS를 이용한 다채널 오디오 서비스를 지원할 때 오디오 코덱 구성에 대한 그림이다. MPS 인코더에서 생성된 부가데이터는 수 kbps로 오디오 비트스트림의 부가데이터로 따로 전송될 수 있으며, MPS 인코더의 오디오 출력은 5.1 채널로부터 만들어진 스테레오 혹은 모노 신호가 된다. 이는 MPS 디코더가 장착되지 않은 일반 라디오 디바이스에서는 종전의 모노/스테레오 서비스가 가능하며, MPS 디코더가 지원된다면 사용자의 선택에 따라 5.1 채널 오디오 신호 재생이 가능하다.



[그림 5] MPEG Surround를 이용한 오디오 서비스

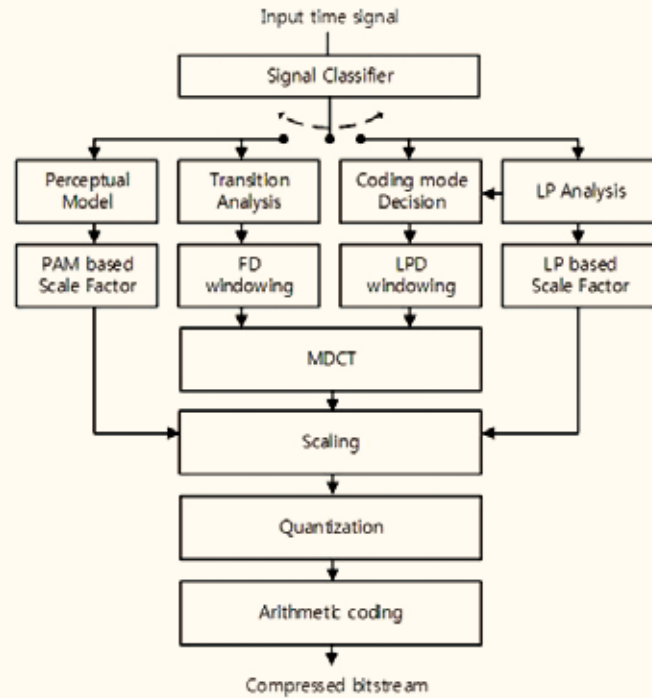
DAB+가 HE-AAC v2와 MPS를 새로운 오디오 코덱 기술로 채택하여 활용함으로써, 디지털 라디오 음질에 대한 불만사항은 많이 감소하였다. 현재 서비스가 활발히 진행되고 있는 유럽 지역에서의 사용자들이 디지털 라디오 방송의 품질이 FM과 동등하거나 더 나은 음질이라고 인식하는 경우가 94%에 이른다고 한다. 그러나 여전히 FM과 비교하여 월등한 음질을 제공한다고는 말하기 힘들며, 지속적인 사업자들의 요구에 의하여 방송의 개수가 늘어날 경우 저 비트레이트에서 음질 보장이 힘들어질 수밖에 없다. 이는 HE-AAC v2가 MUSICAM이나 AAC에 비하여 월등한 압축 및 음질 성능을 보이는 것은 사실이나, 앞서 언급한 바와 같이 오디오 신호에 특화된 코딩 기술이라, 저비트율에서 음성에 대한 품질이 만족스럽지 못할 수 있다. 이러한 오디오 코덱의 단점을 보완하고 음성 및 음악 모두에 만족할 만한 음질을 제공할 수 있다면, 디지털 라디오 서비스도 한층 더 진보된 서비스를 제공할 수 있을 것이다.

이러한 요구사항을 바탕으로 MPEG에서는 2008년부터 새로운 오디오 코딩 기술에 대한 표준화를 시작하여 2012년 2월에 그 표준을 출간하였다. 이 새로운 코덱 기술을 USAC(Unified Speech and Audio Coding)이라 하며, USAC은 음악뿐만 아니라, 음성에 대해서도 높은 음질 성능을 보이며 종전의 어떤 오디오 혹은 음성 코덱 기술보다 뛰어난 압축 효율 및 음질을 제공할 수 있는 기술

이다. 즉, 종전의 최고 기술이었던 HE-AAC v2보다 저 비트율(~64kbps)에서는 콘텐츠에 따라 20~40%의 음질 향상을 달성하였다. USAC이 음성과 음악에 대해서 우수한 성능을 달성할 수 있었던 것은 USAC이 기본적으로 음성과 음악에 대해서 특화된 코딩 모드를 탑재하고 있으며, 입력 신호에 따라 스위칭이 가능한 구조를 가지고 있기 때문이다.

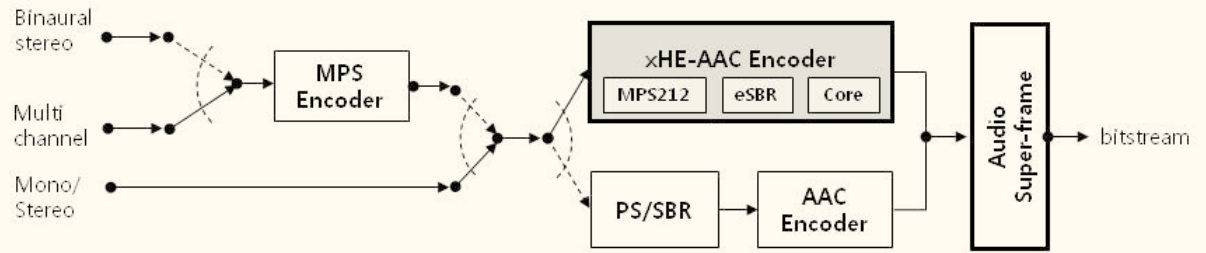
USAC의 구조를 살펴보면 [그림 6]과 같다. USAC도 HE-AAC v2와 마찬가지로 SBR이 포함되어 있으며 그 성능이 개선되었고, PS를 대신하여 MPS212를 활용하였다. 활용된 MPS212 기술은 스테레오 신호만을 생성할 수 있는 MPS 버전으로 PS와 유사한 기능을 수행한다. [그림 6]에서는 보다 분명한 이해를 돕기 위해 SBR과 MPS의 동작과정은 배제하고 스위칭 기반의 코어코딩 과정을 표현하였다. 동작과정을 살펴보면, 입력 신호는 신호 분류기에 의하여 음성인지 음악 신호인지 판별되며, 음악신호로 판별될 경우 AAC와 유사하게 양자화를 수행하여 비트를 할당한다. 이를 USAC에서는 FD(Frequency Domain) 코딩 모드라고 정의하였다. 반대로 입력신호가 음성으로 판별되면 LPD(Linear Prediction Domain) 코딩 모드가 활용되는데 이는 LPC(Linear Predictive Coding) 기반의 잔차신호에 대해서 양자화 하는 코딩 모드로 음성 코딩 분야에서 주로 활용되는 접근 방식이다. 이러한 스위칭 구조의 오디오 코딩 기술을 적용함으로써

음성이든 음악이든 콘텐츠의 특성에 상관없이 좋은 품질의 오디오 신호를 제공할 수 있다. 그러나 음성 및 오디오 판별이 쉽지 않고 빈번한 스위칭이 발생한다면 음질 열화를 초래할 수도 있을 것이다. USAC에 포함된 기술에서 가장 공을 들인 기술도 바로 이러한 스위칭이 빈번히 발생했을 때 음질 열화를 최소화할 수 있는 보정 기술이다. 결국 USAC은 스위칭 시스템이 갖는 우려사항을 추가적인 보정 기술로 이를 해결하였으며, 최종적으로 최고의 품질 및 압축 성능을 제공하는 오디오 코딩 기술을 실현할 수 있게 되었다.

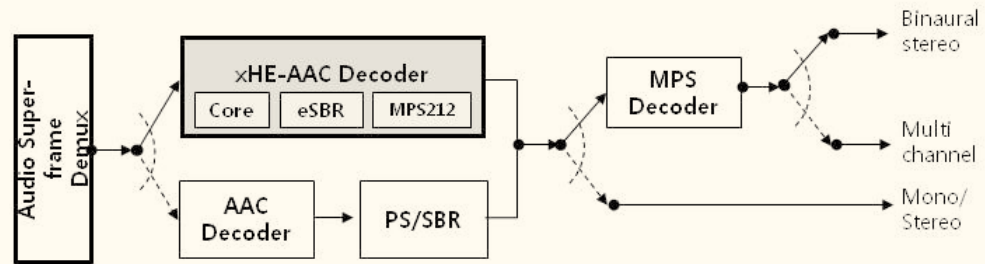


[그림 6] USAC의 코어 코딩 구조

이러한 USAC의 장점은 디지털 라디오 서비스를 한 단계 업그레이드 시킬 수 있을 것이라는 것을 모두가 예상할 수 있을 것이다. 디지털 라디오 전송 방식 중의 하나인 DRM+(Digital Radio Mondiale plus) 표준에서는 USAC을 오디오 코덱으로 활용하는 것을 적극 검토하여 이를 반영하였고, 지난 2014년 1월에 표준 문서를 출간하였다. 그러나 실제로는 USAC이라는 명칭이 HE-AAC 계열의 오디오 코덱을 활용하고 있는 단말기 제조업체나 시장에서 거부감을 줄 수 있기 때문에, USAC 기술을 독립적으로 적용하기 보다는, xHE-AAC(Extended HE-AAC)라는 명칭으로 새로운 프로파일을 정의하여 USAC과 HE-AAC v2를 동시에 지원하는 방법을 모색하였다. xHE-AAC v2는 기본적으로 HE-AAC v2를 지원하는 디코더 형상으로 USAC의 모노와 스테레오 신호에 대해서 추가적으로 지원이 가능한 디코더 구현을 위한 프로파일이다. DRM+에서 xHE-AAC v2를 적용한 구조는 [그림 7]에 나타내었다. DRM+도 DAB+와 마찬가지로 MPS를 이용한 다채널 서비스가 가능하며, xHE-AAC 디코더 형상을 도입하여 USAC 오디오 코딩 기술을 활용할 수 있게 되었다.



(a) DRM + Audio Encoder



(b) DRM + Audio Decoder

[그림 7] DRM+ 방식에서 xHE-AAC가 적용된 인코더 디코더 구조

본 글을 통하여 디지털 라디오 방송을 위해 활용되었던 오디오 코덱 기술들의 기본적인 특징 및 성능에 대해서 살펴보았다. 초창기 MUSICAM과 같은 오디오 코덱은 압축 성능이나 음질이 뛰어나지 못하여 FM보다 뛰어난 음질을 제공하지 못하거나 많은 전송비트를 필요로 하였다. 그러나 하드웨어의 발전과 더불어 오디오 코딩 기술도 많이 진보되어 HE-AAC v2와 같은 고효율 오디오 코덱 기술이 개발되었으며 이를 디지털 라디오 방송에 적용하여 활용함으로써 서비스 개선을 성공적으로 달성하였다.

특히 최근 MPEG에서 표준화가 완료된 USAC 오디오 코덱 기술은 종전의 어떤 오디오 코덱보다도 월등한 성능을 보유하면서 음성과 음악신호의 특성에 상관없이 좋은 성능을 나타내는 코딩 기술로, DRM+에서 최초로 이를 채택하였다. 향후 DAB+에서도 USAC 사용여부가 주목된다. 물론 가장 최신의 오디오 코덱 기술을 활용하는 것이 타당하기는 하나, 추가적으로 고려해야 할 사항은 시스템의 안정성이다. 기존 오디오 코덱과 더불어 모든 전송 메커니즘과 시스템 안정성이 확보된 상황에서 새로운 기술의 도입은 다소 모험이 될 수 있는 요소가 있으며, 새로운 기술 도입에 따른 시스템 안정성 검증 여부가 선행되어야 할 것이다.

한국전자통신연구원에서는 지난 2013년부터 DAB+에 USAC 적용방안에 대한 연구를 진행해 왔으며, 2017년 초까지 USAC을 오디오 코덱으로 DAB+ 기반의 디지털 라디오 방송 서비스에 대한 실험 및 안정성을 검증하고자 한다. 이는 이미 DRM+에는 USAC이 적용되어 다수의 실험방송을 통해 그 안정성이 검증되었지만, DAB+에서도 USAC이라는 새로운 코덱이 성공적으로 적용될 수 있는지에 대한 가능성을 검증하는 절차가 필요하기 때문이다. 

본 연구는 미래창조과학부 및 정보통신기술연구진흥센터의 정보통신·방송 연구개발사업의 일환으로 수행하였음.

[R0166-15-1028, 디지털 라디오 고도화 서비스 표준개발]