

MPEG 오디오 표준 기술 : USAC

□ 박호중, 박영철 / * 경운대학교 전자공학과, **연세대학교 컴퓨터정보통신공학부

1. 서론

최근 방송/통신이 융합되고 개인 무선 통신 및 멀티 미디어 산업이 발달함에 따라 기존 음성 중심의 무선 통신망으로 멀티미디어 정보의 전송이 요구되고 있다. 그에 따라 음성 통신과 오디오 방송으로 구분되던 서비스 구조가 깨지고, 더 이상 음성과 오디오 신호를 별도의 콘텐츠로 분리하는 것이 어렵게 되었다. 이와 같은 서비스의 융합에 따라 하나의 부호화기로 다양한 특성의 음성과 오디오 정보를 동시에 처리하는 통합화가 필요하다. 그러나 기존의 음성 부호화 또는 오디오 부호화 기술은 각각 음성 또는 오디오 신호의 특징을 기반으로 서로 다른 목표에 따라 설계되어 하나의 방법으로 음성과 오디오를 통합적으로 부호화하는데 한계를 가진다. 예로, 음성에 대하여 최고의 성능을 가지는 AMR(adaptive multi-rate) 음성 부호화기로 복잡한 오디오 신호를 부호화 하면 성능이 크게 저하

되고[1, 2], AAC (advanced audio coding) 및 HE-AAC v2(high efficiency AAC version 2) 오디오 부호화기로 음성 신호를 20kbps 이하의 낮은 전송율로 부호화 하면 AMR에 비하여 성능이 저하된다[3, 4].

최초의 의미있는 통합 부호화기로서 AMR-WB+가 개발되었다[5]. 이는 기존의 음성과 오디오 부호화 기술을 혼합한 개념이며, 입력 신호의 특성에 따라 AMR과 변환 부호화(transform coding) 중에서 하나를 선택하는 구조이며, 고효율의 대역폭 확장 기술과 스테레오 처리 기술을 포함한다. 그러나 AMR-WB+는 기존의 첨단 오디오 부호화기에 비하여 오디오 부호화 성능이 저하되어 아직 통합 부호화기로 사용하기에 부족한 부분이 있다.

이와 같은 음성 및 오디오 부호화에서의 새로운 요구에 따라 MPEG에서는 음성과 오디오 신호를 통합적으로 부호화 하는 Unified Speech and Audio Coding(USAC) 기술의 표준화를 2007년에 시작하였

다[6]. USAC이 요구하는 성능 조건은 음성과 오디오에 대하여 각각 최고 성능을 가지는 AMR-WB+와 HE-AAC v2 보다 우수한 품질을 제공하고 12kbps 부터 64kbps까지의 전송률을 지원하는 것이다. 이에 따라 제84차 MPEG 회의에서 전 세계의 7개 업체가 표준화 경쟁에 참여하였고, 2008년 제85차 MPEG 회의에서 성능 평가 결과를 공개하여 최고의 성능을 가지고 Call for Proposal에서 요구하였던 모든 성능 조건을 만족하는 Fraunhofer와 VoiceAge 연합팀의 기술이 USAC Reference Model 0 (RM0)로 결정되었다 [7]. 그에 따라 제86차 MPEG 회의에서 RM0의 Reference Decoder Code가 공개되고 Working Draft(WD)가 완성되었다[8]. 그 이후에, 제87차 회의에서 WD의 무손실 부호화 테이블 변경을 거쳐 현재의 WD2가 완성되었고[9], 현재 MPEG에서는 Core Experiment(CE) 과정을 통하여 계속적으로 성능 향상을 위한 새로운 기술 연구가 진행 중이다.

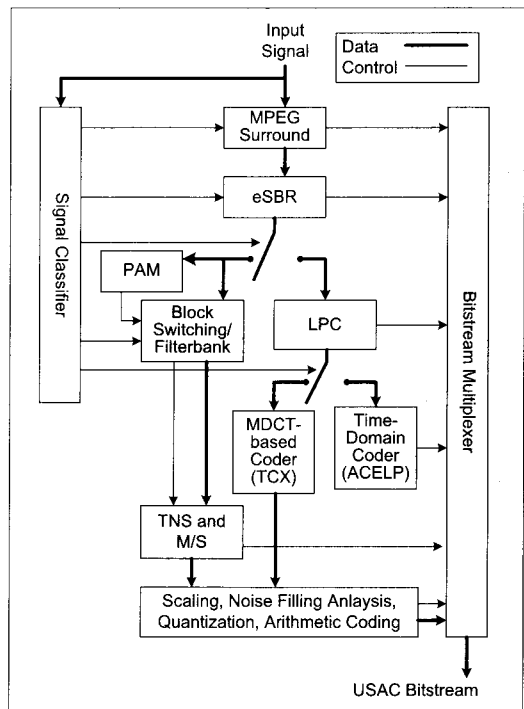
본 기고에서는 USAC WD2를 기준으로 USAC 각 모듈의 동작과 기술을 소개한다. 먼저, USAC의 전체 구조를 정리하여 특징을 분석하고, 각 모듈의 세부 동작을 설명한다. 또한, USAC의 성능을 기존의 표준 부호화기와 비교하여 정리한다.

II. USAC 구조

USAC은 음성과 오디오 신호의 통합 처리를 위하여 완전히 새로운 부호화 방법을 사용하는 것이 아니라, 기존의 음성 및 오디오 부호화 기술을 효과적으로 융합하고 선택적으로 사용하는 방법에 기반을 둔다. 즉, 음성 부호화를 위한 AMR 기술[1, 2], 오디오 부호화를 위한 AAC 기술[3], 저전송율 혼합 신호 부호화에 강점을 가지는 TCX(transform-coded excitation)

기술[5], 고대역 신호를 파라미터로 부호화 하는 SBR(spectral band replication) 기술[10], 스테레오를 위한 MPEG Surround 기술[11], 그리고 서로 다른 동작 사이를 왜곡 없이 연결하기 위한 천이 기술들을 효과적으로 융합하여 전체 USAC을 완성한다.

<그림 1>이 USAC 인코더(encoder)의 전체 동작 구조를 보여준다. MPEG Surround에서 파라미터 스테레오 기법에 따라 스테레오 정보를 추출하고 모노 신호를 출력하고, eSBR에서 고대역 부호화를 위한 파라미터를 추출한다. 이 때 eSBR의 출력 신호는 모노 저대역 신호가 되고, 이 신호를 압축하는 코어 압축(core coding) 동작이 진행된다. 코어 압축 과정을 위하여 AAC, TCX, ACELP 등 3 가지 압축 방법이 제공되고, AAC와 관련된 부분을 회색 박스로 표시하였다. AAC를 FD(frequency domain) 모드,



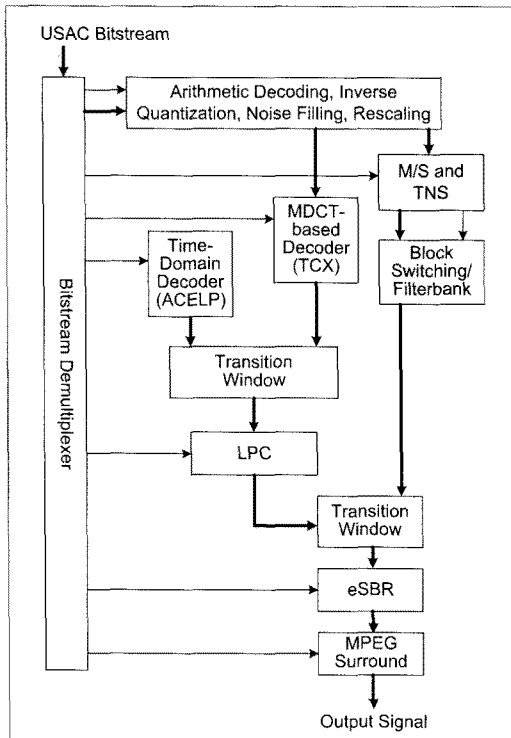
<그림 1> USAC Encoder의 동작 흐름도

<표 1> USAC Core Codec의 3가지 압축 방법

압축기 명칭	특징	주요 대상 신호
AAC	MDCT 기반의 변환 부호화, 기존 AAC-LC와 동일하게 Block Switching을 사용하고 PAM 기반으로 계수 양자화, 새로운 무손실 부호화 방법 사용	고전송률 오디오 신호
ACELP	LPC 기반으로 시간 영역에서 부호화, 기존 AMR-WB의 ACELP 모듈과 동등	음성 신호
TCX	LPC Residual 신호에 대해 MDCT 기반의 변환 부호화, AMR-WB+와 유사한 이득 정규화와 Noise Fill 사용	혼합 신호와 저전송률 오디오 신호

ACELP와 TCX를 LPD(linear prediction domain) 모드라 정의한다. Signal Classifier에서 입력 신호의 특성을 분석하여 최적의 압축이 가능한 압축 방법을 선택한다. 물론, 선택되는 압축 방법을 프레임 단위로 계속 변경할 수 있다. USAC에서 사용하는 3가지의 압축 방법의 특성은 표 1에 정리되어 있다.

<그림 2>는 USAC 디코더(decoder)의 동작 구조



<그림 2> USAC Decoder의 동작 흐름도

를 보여준다. 압축 과정에 대응하여 복원 동작을 수행하고, 3가지 코어 복호기 중에서 인코더에서 선택한 모드에 따라 복원 동작을 수행한다. 디코더에서의 핵심 기술은 프레임 단위로 모드가 변경될 때, 부호화 방식의 차이에 의한 불연속성과 윈도우 연결을 왜곡 없이 진행하는 것이다. 특히 MDCT를 사용하는 모드와 이를 사용하지 않는 모드 사이의 변경에서 TDAC(time domain aliasing cancellation)를 보장하기 위하여 특별히 설계된 윈도우를 사용한다. 또한, LPC(linear predictive coding) 사용 여부가 변경될 때 이로 인한 왜곡이 발생하지 않도록 하는 기술도 사용한다.

III. USAC의 세부 동작

1. AAC

USAC 인코더에서 AAC 동작은 HE-AAC v2의 코어 부호화기와 유사하다 [4]. 다만 TCX 및 ACELP 블록과 연결되어 TDAC를 구현할 수 있는 새로운 형태의 윈도우들이 추가된 것이 가장 큰 차이점이다. 먼저 심리음향 모델(psycho-acoustic model)을 이용하여 신호가 가지는 청각적인 중복성을 제거하기 위한 마스킹 레벨을 결정한다. 기본적인 기능은 기존의 AAC와 동일하다. 블록 스위칭과

필터뱅크 모듈은 입력신호를 주파수 영역으로 변환하기 위한 모듈이다. 블록스위칭은 입력신호에 따라서 변환 윈도우의 크기를 달리함으로써 급격한 변화를 가지는 신호와 상대적으로 변화폭이 적은 신호에 대하여 효율적으로 부호화 할 수가 있다. 윈도우 함수는 필터뱅크 주파수 응답에 중요한 영향을 미치므로 AAC 필터뱅크에서는 입력 신호에 잘 적응하기 위해 윈도우의 모양을 변형시킬 수 있도록 설계되어 있다. 윈도우의 모양은 다양한 입력 신호에 대해 필터뱅크가 입력 스펙트럼 성분을 효율적으로 분리할 수 있도록 인코더와 디코더에서 동시에 변화한다. 2048개의 시간축 샘플을 사용한 변환은 복잡한 스펙트럼을 가진 신호에 존재하는 통계적인 중복성을 제거함으로써 전체적인 부호화 효율을 높여준다.

블록 크기가 큰 장변환(long transform)은 과도기에 있는 신호의 부호화에 대해서는 비효율적이다. 그러므로 천이 구간 신호는 단변환(short transform)으로 부호화해야 한다. 그러나 단변환은 주파수 분해 능력이 작기 때문에 정상 상태 신호에 대해서는 부호화 효율이 좋지 않다.

USAC 필터뱅크는 MDCT(modified discrete cosine transform)를 사용하는데 LPC 기반의 코딩 모듈(LPD)과의 프레임간의 연결을 고려하여 윈도우의 종류가 다양하다. 기존 AAC에서 사용하는 윈도우(LONG_WINDOW, LONG_START_WINDOW, 8 SHORT_WINDOW, LONG_STOP_WINDOW)에 LPD와의 프레임 연결을 위하여 4가지 윈도우

(STOP_START_WINDOW, START_WINDOW_LPD, STOP_WINDOW_1152, STOP_START_WINDOW_1152)가 추가되었다. 이들 윈도우는 기존 start 및 stop 윈도우와 중첩 구간은 같은 모양을 가지나 평탄한 구간(크기가 1인 구간)의 길이가 다르다. 이는 LPD와 MDCT의 Aliasing Cancellation을 만족시키기 위하여 고안된 구조이다. 추가된 3개의 윈도우 중에서 하나(START_WINDOW_LPD)는 단구간 중첩 구간이 64개로 다른 2개의 윈도우(128개)와 다르다. 사용되는 MDCT의 수식은 식 (1)과 같다[4].

변환 결과의 기대칭(odd symmetric) 특성 때문에 0에서 $N/2-1$ 까지의 계수들만으로 의미를 갖게 된다. 한편 디코더에서 수행되는 IMDCT(Inverse MDCT)는 식 (2)와 같다[4].

USAC RM0에서 사용하는 TNS 툴과 M/S 스테레오 툴도 기본적인 기능은 기존의 HE-AAC v2와 같다. TNS 블록은 시간 축에서의 선형 예측 계수를 이용하여 입력 신호의 주파수 스펙트럼의 포락선을 구할 수 있듯이, 주파수 영역에서 선형 예측을 통해 시간축 신호의 포락선을 모델링 할 수 있다는 이중성(duality) 성질을 이용한 것이다. 평탄하지 않은 스펙트럼을 가진 신호는 직접 스펙트럼의 값을 부호화(변환 부호화)하거나 시간축 신호에 예측 부호화 방법을 적용함으로써 효율적으로 부호화 할 수 있다. 반대로, 시간-주파수 쌍대의 원리에 의해 시간축에서 평탄하지 않은 신호(즉, 천이 구간)는 시간축에서 부호화를 하거나 주파수 축 스펙트럼 계수에 대

$$X_{i,k} = 2 \sum_{n=0}^{N-1} z_{i,n} \cos\left(\frac{2\pi}{N}(n+n_0)(k+0.5)\right), \quad 0 \leq k < N/2 \quad (1)$$

$$x_{i,n} = \frac{2}{N} \sum_{k=0}^{N/2-1} X_{i,k} \cos\left(\frac{2\pi}{N}(n+n_0)(k+0.5)\right), \quad 0 \leq n < N \quad (2)$$

해 예측을 통해 부호화 할 때 그 효율이 높게 될 것이다. 다시 말해, 시간축에서의 예측은 부호화기의 스펙트럼 해상도를 증가시키는 반면, 주파수축에서의 예측은 부호화기의 시간 분해능을 향상시킨다.

M/S 스테레오 부호화는 원신호의 스테레오 음상에 대응하도록 부호화 잡음의 상(imaging)을 제어하는 역할을 한다. 이 방법을 사용하면 공간적으로 마스킹되지 않아서 생기는 잡음의 상(imaging) 문제를 해결하는데 도움을 준다. MS 스테레오에서는 좌/우 채널의 MDCT 계수의 합과 차로 Mid와 Side 채널을 정의하여 부호화 한다.

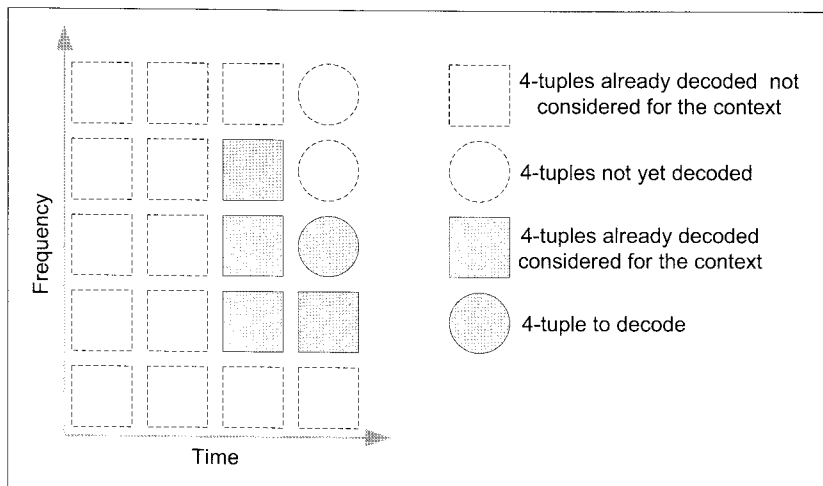
USAC RMO에서는 또한 Time-Warped 필터뱅크를 이용하여 시간 영역에서 서로 다른 해상도를 갖는 주파수 분해 기능도 가지고 있다. 그러나 이 블록의 사용은 매우 제한적이며 사용에 의한 효과도 특별히 관찰되지 않는다.

USAC AAC에서는 무손실 부호화(lossless coding)을 위하여 MPEG-4 AAC에서의 허프만 코딩(huffman coding)을 하지 않고 새로운 Context 기

반의 적응 Arithmetic 코딩을 사용한다. 이를 위해 <그림 3>과 같이 주파수/시간 영역에서 인접한 4개의 스펙트럼 성분으로 누적된 주파수 테이블을 이용하여 현재 State를 결정하여 Arithmetic 코딩을 한다. 이 방법은 이웃한 시간/주파수 Context 정보를 활용하므로 기존의 허프만 코딩에 비하여 많은 비트 감소 효과를 제공한다.

2. ACELP와 TCX

ACELP와 TCX는 모두 LPC 기반으로 신호를 부호화 한다. AMR-WB+와 동일하게 동작의 기본단위는 1024-샘플의 Superframe이고, 세부 단위로 256-샘플의 프레임 사용한다. TCX의 동작 모드는 256, 512, 1024의 3가지 다른 길이를 사용하고, LPC 분석 모듈에서 256-샘플 프레임 단위로 LPC 분석을 실시하여 총 4 LPC Set를 구하고 각각을 LPC1, LPC2, LPC3, LPC4라 정의한다. 실제로 전송할 LPC Set는 선택된 동작 모드에 따라 AMR-



<그림 3> State 연산을 위한 Context

WB+와 동일한 방법으로 결정된다.

USAC는 AMR-WB+와 다른 새로운 LPC 계수 양자화기를 사용한다. 다양한 특성의 계수를 효과적으로 양자화 하기 위하여 4가지의 양자화기 중에서 하나는 선택하는 기법을 사용하고, 표 2에 각 양자화기의 규격이 정리되어 있고 Differential 양자화예의 Reference LPC는 표 3으로 정의된다. 단, LPC4는

〈표 2〉 LPC 양자화기

양자화기	비트수	규격
3-Stage Quantizer	46	1st Stage : 16D VQ, 6 Bits 2nd Stage : Split VQ, 9D+7D Split, 8+8 Bits 3rd Stage : Split VQ, 5D+4D+7D, 8+8+8 Bits
Differential Quantizer #1	0	Bit 전송 없이 Reference LPC를 그대로 사용
Differential Quantizer #2	8	Reference LPC와의 차이를 8 Bit 16D VQ
Differential Quantizer #3	24	Reference LPC와의 차이를 5D+4D+7D Split VQ, 8+8+8 Bits

〈표 3〉 LPC의 Differential 양자화기에서 사용하는 Reference LPC

Reference LPC	
LPC1	LPC2
LPC2	LPC4
LPC3	LPC2와 LPC4의 평균

〈표 4〉 USAC TCX에서 사용하는 윈도우 규격

이전 모드	현재 모드	MDCT 포인트	MDCT 계수수	Zero	천이	Top	천이	Zero
ACELP	TCX256	640	320	160	0	256	128	96
ACELP	TCX512	1152	576	288	0	512	128	224
ACELP	TCX1024	2304	1152	512	128	1024	128	512
TCX	TCX256	512	256	64	128	128	128	64
TCX	TCX512	1024	512	192	128	384	128	192
TCX	TCX1024	2048	1024	448	128	896	128	448

항상 3-Stage 양자화기를 사용하여야 한다. USAC LPC 양자화기는 이웃한 LPC Set 사이에 큰 변화가 없을 경우에 양자화 성능을 향상시키기 위한 기법이 며, 시간에 따라 LPC 변화가 심하지 않는 오디오 신호에서 양자화 성능을 크게 향상시킬 수 있다.

LPC Residual 신호를 ACELP와 TCX 중에서 한 가지 방법으로 처리한다. ACELP 방법은 기존의 AMR-WB와 거의 유사한 방법을 사용하므로 자세한 설명은 생략한다. TCX 방법은 AMR-WB+의 TCX와 유사한 개념이지만 AMR-WB+에서의 DFT 대신에 MDCT를 사용하여 주파수 변환을 하는 근본적 차이점을 가진다. 나머지 프레임 구조, 주파수 계수의 이득 정규화, 이득 양자화, Noise Fill 등은 AMR-WB+와 동일하다.

USAC TCX 동작에서의 핵심 기술은 TCX 모드 변환을 위한 윈도우 스위칭 기술이다. MDCT는 TDAC를 보장하여야 하므로 기존 DFT와 같은 간단한 윈도우를 사용하지 못하므로, 이전과 현재 프레임의 모드 조합에 따라 TDAC를 위하여 새롭게 설계된 윈도우를 사용한다. 표 4에 전체 윈도우의 규격이 정리되어 있고, 이해를 돕기 위하여 <그림 4>에 두 개의 윈도우를 예로 보여준다. 여기서, 천이 구간에서는 실제 Sine 윈도우 모양을 가지는데 그림에서는 간단히 직선으로 나타내었다. 예로, 이전 모드가 TCX256이고 현재 모드가 TCX256일 때, 512-샘

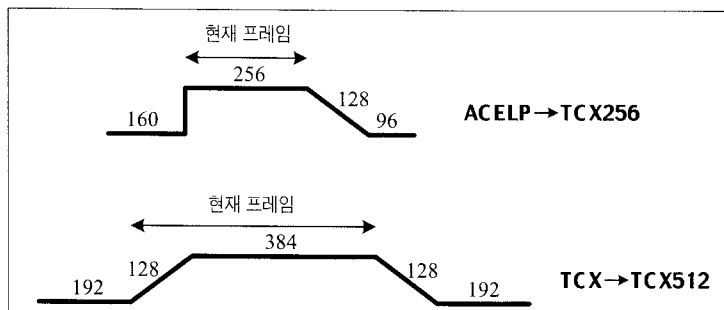
플 신호에 대하여 512-포인트 MDCT를 통하여 256개의 MDCT 계수를 구하여 전달한다. 만일 이전 모드가 ACELP이면, ACELP와의 정상적인 천이 동작을 위하여 윈도우가 640 샘플로 길어지고 640-포인트 MDCT를 실시하고 그에 따라 MDCT 계수의 수가 320개로 증가한다. 결국, ACELP에서 TCX 모드로 천이될 때, 동작의 연결을 위하여 추가 MDCT 계수가 필요하다. 마지막으로 <그림 5>에 동작 모드가 TCX256→ACELP→TCX512→TCX256와 같이 변경될 때, 프레임 구조와 각 모드에서 사용하는 윈도우의 모양, 그리고 윈도우가 겹치는 형태를 보여준다.

이와 같은 윈도우를 사용하여 LPC Residual 신호에 대한 MDCT 계수를 구한 후, Scalar 양자화 하고 AAC 모듈과 동일한 Arithmetic 코딩한다. 이 때, 코

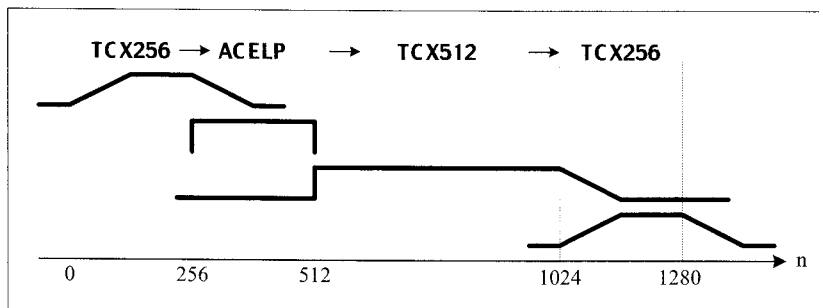
딩에 사용되는 비트 수가 가용 비트 수를 넘지 않기 위하여 양자화 이전에 먼저 MDCT 계수를 Scaling 하는데, 전 대역의 MDCT 계수를 항상 동일한 값으로 Scaling 한다. 마지막으로 MDCT 계수 이득값과 Noise Fill Level을 구하고 양자화 하여 전달한다.

3. SBR

USAC는 고대역 신호를 효율적으로 부호화하기 위하여 저대역 정보와 보정 파라미터를 사용하여 고대역을 부호화 하는 SBR 기술을 사용한다. 기존 HE-AAC v2에서 사용하는 SBR 기술을 많이 차용하고 약간의 규격 수정과 성능 향상을 위한 새로운 선택 사항을 제공하고 이를 eSBR(enhanced SBR)로 정의한다. 기존 SBR은 T/F Grid 단위로 저대역의



<그림 4> TCX에서 사용하는 윈도우의 예



<그림 5> TCX 모드 천이에 따른 윈도우 사용의 예

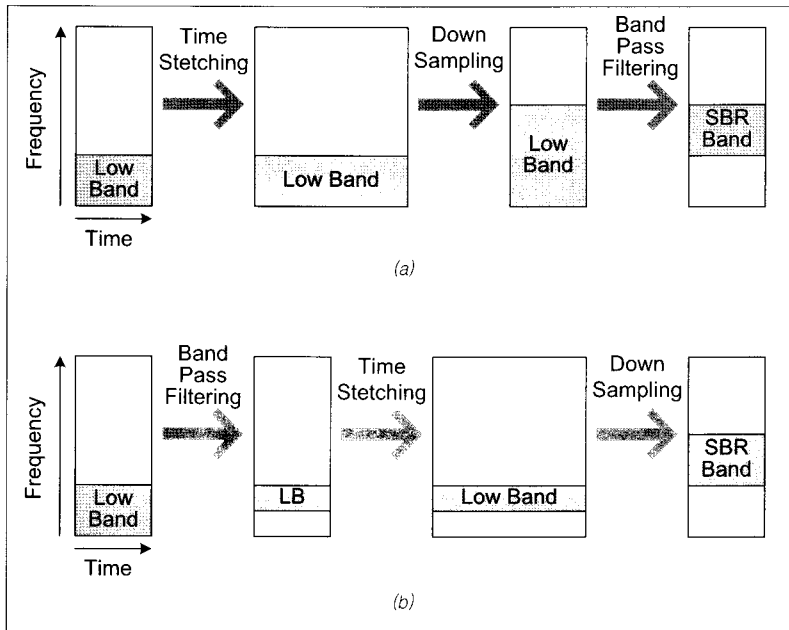
정보를 고대역으로 Patching하여 고대역(또는 SBR 대역) 정보를 생성한다. 반면, eSBR은 SBR 대역 생성을 위한 새로운 기술로 하모닉 SBR 기술을 제공하고, 기존 Patching 방법과 하모닉 SBR 중에서 하나를 선택하여 사용한다.

하모닉 SBR은 SBR 대역 생성을 Phase Vocoder 기법을 통하여 구현하는데, 이 과정을 <그림 6>에 따라 T/F 영역에서 설명하면 다음과 같다. (a)와 같이 저대역 성분만을 가지는 입력 신호를 Phase Vocoder 기술을 활용하여 시간 영역에서 N배 확장된(time-stretched) 신호를 구하고, 그 신호를 N배 Down-Sampling 하여 출력 신호를 구한다. Down-Sampling 과정에 의하여 스펙트럼이 확장되므로 입력 신호에서는 성분이 없었던 고대역에 새로운 성분이 생성된다. 마지막으로 SBR에서 원하는 영역의 신호만을 선택하기 위하여 대역 통과 필터를 적용시

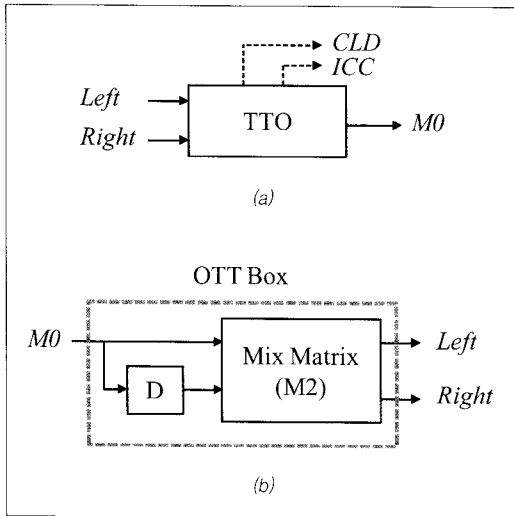
키면 최종 SBR 대역 신호를 생성한다. 또는 동작의 순서를 바꾸어 (b)와 같이 먼저 저대역 성분만을 가지는 입력 신호에 대역 통과 필터를 적용한 후, 시간 확장과 Down-Sampling을 하여 최종 SBR이 요구하는 SBR 대역 신호를 구할 수 있다. 하모닉 SBR 기술을 통하여 SBR 대역을 생성하면 저대역의 하모닉 성분이 SBR 대역으로 그대로 확장되어 하모닉 성분이 강한 신호에 대하여 정확한 SBR 대역의 생성이 가능하다. 마지막으로 기존 SBR에서 사용하는 Envelope, Noise Floor 등의 보정 파라미터를 추출하여 최종 고대역 신호를 부호화 한다.

4. Stereo

USAC는 MPEG Surround 부호화 기술에 기반한 스테레오 톨을 사용하고 있다. MPEG Surround 인코



<그림 6> eSBR의 하모닉 SBR 기술로 SBR 대역을 생성하는 과정



〈그림 7〉 MPEG Surround의 TTO 모듈과 OTT 모듈

당의 기본 구성요소는 TTO(Two-To-One) 블록이다 [11]. 〈그림 7(a)〉에 도시된 TTO 모듈은 두 개의 입력 신호에 대해 하나의 다운믹스 신호와 공간 파라미터를 얻는 과정을 수행한다. 이는 기존 HE-AAC v2의 옵션인 PS(parametric stereo)에 대응된다[12].

TTO 부호화 과정은 주파수 축에서 밴드별로 수행되며 이때 적용되는 주파수 해상도는 인간의 공간 인지에 대한 청각특성을 반영한 Critical 대역폭을 사용한다. PS에서 시간/주파수 매핑을 위해 사용하는 주파수 변환은 Hybrid QMF 분석 필터뱅크로서, 이는 SBR에서 사용하는 64 대역 QMF를 통하여 얻은 저주파 대역에 대해 더 높은 해상도를 얻기 위해 추가적인 필터링 과정이 뒤따르는 것이다.

각각의 파라미터 밴드에 대해 두 가지의 공간 파라미터들이 추출된다. 두 개의 입력 채널에 대한 전력비를 dB 단위로 표현한 CLD(channel level difference) 파라미터와 두 채널 신호 사이의 상관관계를 표현하는 ICC(inter-channel correlation) 파라미터를 추출한다. TTO 모듈은 입력 신호에 대해 레벨 패닝

(panning)된 주요성분과 서로 반대위상을 가지는 잔향성분의 합으로 모델링하고, 이러한 입력 신호에 대해서 TTO 모듈의 입력신호와 출력신호의 에너지가 보존되는 조건에 따라 모노 다운믹스를 수행한다.

디코더에서는 다운믹스된 모노 신호를 먼저 복원한 다음 추출된 CLD와 ICC를 이용하여 스테레오 신호를 복원한다. 〈그림 7(b)〉은 디코더 단에서 사용되는 OTT(One-to-Two) 모듈의 구조이다. 그림에서 복원된 모노 신호와 이를 Decorrelator(그림에서 D로 표시된 블록)를 거쳐 만들어낸 잔향신호를 Mix Matrix를 사용하여 혼합함으로써 스테레오 신호를 만들어 낸다. USAC에서 사용하는 Decorrelator는 MPEG Surround와 마찬가지로 주파수 별로 달라지는 Pre-Delay와 전역통과필터가 연결된 구조로 구성된다. USAC에서는 기존의 MPEG 서라운드에서 지원하는 모드에 스테레오 신호의 다운믹싱-업믹싱을 위한 새로운 2-1-2 모드를 추가하였는데, 2-1-2 구조에서 Decorrelator는 주파수 영역을 4 개로 나누어 서로 다른 차수와 위상 특성을 갖는 전역통과필터를 구성하고 있다. 2-1-2 모드를 이용한 MPEG 서라운드 툴은 주로 낮은 비트율에서 동작하며, 64kbps와 같은 높은 비트율에서 코어 코더는 M/S 스테레오나 독립적인 채널 코딩을 수행하기 때문에 MPEG Surround는 낮은 비트율에서만 동작한다.

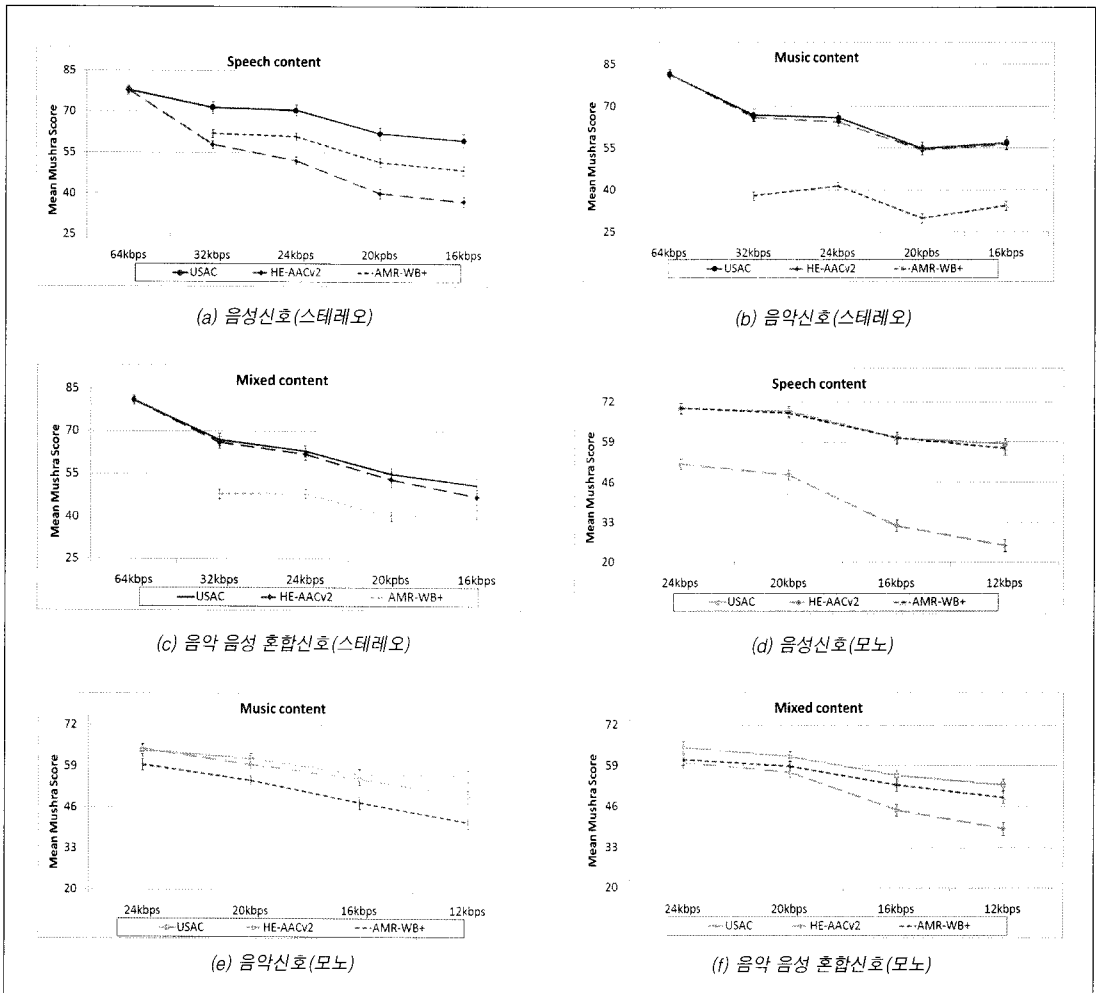
IV. USAC 성능

USAC RM 선정시 청취 평가 성능이 기존 코덱인 HE-AAC v2 [4]와 AMR-WB+ [5] 비해 모든 신호와 모든 전송률에서 적어도 나쁘지 않거나 더 좋은 성능을 보여야만 한다는 조건이 부가되었다[6]. 청취 평가 결과, Fraunhofer와 VoiceAge에서 제출한

코덱만이 이 조건을 만족하는 것으로 판단되었고, 현재의 USAC RMO는 Fraunhofer와 VoiceAge에서 제출한 코덱의 통합된 버전이다. 그러나 USAC 표준화 초기 단계부터 두 회사는 공동으로 코덱을 개발하여 왔기 때문에 두 코덱의 통합은 쉽게 이뤄졌다.

USAC RMO 선정을 위한 성능평가는 다양한 동작 모드에서 MUSHRA 테스트를 이용한 주관적인 청취평가로 수행되었다. 실험에는 음성, 음악, 음성 음

악 혼합 등의 세 가지 범주로 속하는 네 개 씩의 아이টে็ม으로 구성된 총 12개의 아이টে็ม들을 사용하였다. 테스트 아이টে็ม들은 각각 AMR-WB+, HE-AAC v2, USAC RMO로 부호화하여 품질을 비교 평가하였으며, Hidden Reference (HIR), 3.5kHz(LP35), 7.0kHz(LP70)로 저역통과 된 앵커가 포함되었다. 64kbps 모드는 AMR-WB+를 지원하지 않기 때문에 비교 대상에 포함되어 있지 않다. 모노 아이টে็ม에



<그림 8> USAC RMO 청취평가 결과

대해서는 12, 16, 20, 24kbps에서 성능이 평가되었으며, 스테레오 아이টে에 대해서는 16, 20, 24, 32, 64kbps에서 성능 평가가 실시되었다. 실험은 10개 기관에서 다수의 청취자를 대상으로 진행되었다. 여기서 소개하는 실험 결과는 USAC RMO 선정 시 얻어진 결과의 일부로서, 성능 비교의 편의를 위하여 USAC RMO의 성능을 AMR-WB+와 HE-AAC v2와 비교하여 <그림 8>에 정리하였다.

평가 결과는 무엇보다도 AMR-WB+와 HE-AAC v2의 특성을 명확하게 보여준다. 음악 범주에 속하는 아이টে에의 경우, HE-AAC v2가 AMR-WB+에 비해 일관성 있게 우수한 성능을 보여준다. 반면, 음성 아이টে에의 경우에는 AMR-WB+가 HE-AAC v2에 비해 월등한 성능을 보여준다. 이 결과는 ACELP 기반의 음성 부호화기와 MDCT 기반의 오디오 부호화기가 각기 다른 영역에서 최선의 부호화 기임을 보여준다. 따라서 최상의 USAC은 결국 두 가지 이종의 코덱을 얼마나 잘 조화되도록 통합하느냐에 달려있음을 분명하게 알 수 있다. 한편 음악과 음성이 섞여있는 아이টে에의 경우에도 AMR-WB+가 HE-AAC v2보다 우수한 성능을 보여주지만 성능 차이는 음성 신호의 경우보다 더 작아짐을 알 수 있다. 또 한 가지 주목할 사항은 HE-AAC v2가 모노 아이টে에보다는 스테레오 아이টে에에서 AMR-WB+의 성능을 압도하거나(음악, 혼합), 성능 격차가 더 줄어드는 것(음성)을 알 수 있다. 이는 HE-AAC v2에서 사용하고 있는 스테레오 톨이 AMR-WB+보다 더 우수한 성능을 가짐을 보여주는 결과라고 할 수 있다.

이에 비해 USAC RMO는 모든 아이টে에과 모든 전송률에서 평균적인 성능이 AMR-WB+나 HE-AAC v2보다 적어도 떨어지지 않거나 우수함을 알 수 있다. 특히 몇몇 경우에는 두 가지 기준 코덱에 비해 월등한

성능을 보여준다(스테레오 음성신호의 경우와 모노 혼합 신호의 경우). 청취 결과를 종합하면 USAC RMO는 음성에 대해서 우수한 성능을 갖는 AMR-WB+의 장점과 음악에 대해 우수한 HE-AAC v2의 장점을 잘 결합한 부호화기라고 할 수 있다.

V. 결론

개인화된 무선 통신 및 멀티미디어 기기 산업의 발달에 따라 낮은 전송률에서 신호의 특성과 무관하게 일관적인 성능을 보여주는 음성/오디오 통합 부호화기에 대한 필요성은 더욱 커지고 있다. 특히 현존하는 부호화기들이 입력 신호의 분류, 즉, 음성인지 음악인지에 따라 성능이 크게 변화하기 때문에 이 두 가지 범주의 신호에 대해 기존 부호화기가 갖고 있는 최상의 성능을 보여줄 수 있는 USAC에 대한 표준화가 진행되고 있다. 현재 표준화는 RMO가 선정되었으며 RMO의 성능을 개선하기 위한 CE(core experiment) 단계에 있다. 현재 선정된 RMO는 음성 신호 부호화에 효율적인 ACELP 기술과 음악 신호를 효율적으로 부호화하는 MDCT 기반의 AAC 부호화 기술이 결합된 형태로 구성되어 있으며, MPEG 오디오에서 사용하는 SBR, 스테레오 톨들을 채택하고 있다. USAC RMO의 성능은 기존 코더인 AMR-WB+와 HE-AAC v2에 비해 음악 및 음성 범주에 걸쳐 우수하거나 월등한 성능을 보이고 있으며, 추가적인 성능 향상이 계속 진행되고 있다. 따라서 향후 낮은 비트율의 멀티미디어 서비스 산업에 적극적인 활용이 기대된다. 그러나 현재 USAC RMO는 시스템의 복잡도나 부호화를 위한 시간 지연 문제 등 실제 서비스에 적용되기 위해서 해결해야 할 과제들도 안고 있다.

참고 문헌

- [1] 3GPP TS 26.090, "Adaptive Multi-Rate (AMR) Speech Codec : Transcoding Functions," 2002.
- [2] 3GPP TS 26.190, "Adaptive Multi-Rate Wideband (AMR-WB) Speech Codec : Transcoding Functions," 2001.
- [3] ISO/IEC 14496-3, "Coding of Audio-Visual Objects - Part 3 : Audio, Subpart 4 : General Audio Coding - AAC, TwinVQ, BSAC," 2001.
- [4] ISO/IEC 14496-3:2005, "Coding of Audio-Visual Objects, Part 3: Audio," 2005.
- [5] 3GPP TS 26.290, "Extended Adaptive Multi-Rate Wideband (AMR-WB+) codec : Transcoding Functions," 2005.
- [6] ISO/IEC SC29 WG11 N9519, "Call for Proposals on Unified Speech and Audio Coding," 82nd MPEG Meeting, Shenzhen, China, October, 2007.
- [7] ISO/IEC SC29 WG11 M15621 "Technical Description of the Fraunhofer IIS Submission for the CiP on USAC," 85th MPEG Meeting, Hannover, Germany, July, 2008.
- [8] ISO/IEC SC29 WG11 N10215, "WD on Unified Speech and Audio Coding," 86th MPEG Meeting, Busan, Korea, October, 2008.
- [9] ISO/IEC SC29 WG11 N10418, "WD2 of USAC," 87th MPEG Meeting, Lausanne, Swiss, February, 2009.
- [10] ISO/IEC 14496-3, AMD. 1, "Bandwidth Extension," 2003.
- [11] ISO/IEC FCD 23003-1, "MPEG-D (MPEG audio technologies), Part 1: MPEG Surround," 2006.
- [12] ISO/IEC 14496-3, AMD. 2, "Parametric Coding of High Quality Audio," 2004.

필자 소개



박호중

- 1982년 ~ 1986년 : 서울대학교 전자공학과 학사
- 1986년 ~ 1987년 : University of Wisconsin-Madison 전기 및 컴퓨터공학과 석사
- 1989년 ~ 1993년 : University of Wisconsin-Madison 전기 및 컴퓨터공학과 박사
- 1993년 ~ 1997년 : 삼성전자 정보통신본부 선임연구원
- 1997년 ~ 현재 : 광운대학교 전자공학과 교수
- 주관심분야 : 음성 부호화, 오디오 부호화, 멀티미디어 방송 시스템, 음성 품질 향상



박영철

- 1982년 ~ 1986년 : 연세대학교 전자공학과 학사
- 1986년 ~ 1988년 : 연세대학교 전자공학과 석사
- 1988년 ~ 1993년 : 연세대학교 전자공학과 박사
- 1993년 ~ 1995년 : Pennsylvania State University, Graduate Program in Acoustics, PostDoc.
- 1996년 ~ 1997년 : 삼성전자 반도체사업부 선임연구원
- 1997년 ~ 1998년 : 삼성의료원 의공학센터 전문연구원
- 1998년 ~ 2001년 : (주)인타임 시스템 LSI 연구소 연구소장
- 2002년 ~ 현재 : 연세대학교 컴퓨터정보통신공학부 교수
- 주관심분야 : 오디오 부호화, 3D 오디오, 디지털 보청기