

논문 2010-47SP-5-9

통합 음성/오디오 부호화를 위한 새로운 MPEG 참조 모델

(A New MPEG Reference Model for Unified Speech and Audio Coding)

송 정 욱*, 오 현 오**, 강 홍 구*

(Jeongook Song, Hyen-O Oh, and Hong-Goo Kang)

요 약

음성 및 오디오 코덱은 각 신호의 특성 및 응용 분야가 다르기 때문에 오랜 기간 동안 각기 다른 부호화 방법을 기반으로 개발되고 발전되어 왔다. 최근 방송 및 통신 시스템이 융합되는 흐름에 발맞추어 3GPP 및 ISO/IEC MPEG 등의 표준화 기관에서는 두 신호를 하나의 통합 코덱을 이용하여 압축 전송하기 위한 노력을 지속해 왔다. 그 일환으로 MPEG에서는 그 간의 표준화된 기술들을 통합하고, 다양한 주관적 음질 평가 결과를 기반으로 USAC (Unified speech and audio coding)이라고 불리는 코덱의 표준화를 진행 중이다. 그러나 USAC RM (Reference model) 소프트웨어의 구조적인 복잡성, 사용되지 않는 수많은 모듈로 인한 용량의 비대합, 그리고 부호화기의 열악한 성능 등으로 인하여 기존 RM을 개선하고자 하는 필요성이 지속적으로 제기되었다. 본 논문에서는 USAC에 포함된 주요 기술을 설명하고, 이러한 문제를 최소화하기 위해 오픈 소스 기반으로 새롭게 설계된 RM 소프트웨어를 제안한다. 이는 2010년 4월 MPEG 회의에서 발표되었으며, 6월 모든 참여 기관을 위해 소스 코드가 공개되었다.

Abstract

Speech and audio codecs have been developed based on different type of coding technologies since they have different characteristics of signal and applications. In harmony with a convergence between broadcasting and telecommunication system, international organizations for standardization such as 3GPP and ISO/IEC MPEG have tried to compress and transmit multimedia signals using unified codecs. MPEG recently initiated an activity to standardize the USAC (Unified speech and audio coding). However, USAC RM (Reference model) software has been problematic since it has a complex hierarchy, many useless source codes and poor quality of the encoder. To solve these problems, this paper introduces a new RM software designed with an open source paradigm. It was presented at the MPEG meeting in April, 2010 and the source code was released in June.

Keywords : Unified speech and audio coding, USAC Reference Encoder

I. 서 론

음성 코덱은 음성 발생 모델에 기반 하여 소스-필터 (Source-filter) 구조를 바탕으로 발전하여 왔으며, 유선 전화에서부터 무선 이동 통신에 이르기 까지 양방향 음성 통신을 목적으로 폭넓게 사용 되어 왔다. 이와 달리 오디오 코덱은 인체 청각 모델(human hearing system)

에 기반 하여 오디오 신호에 대하여 높은 인지적 음질 (perceptual quality)을 표현할 수 있도록 개발되어 디지털 방송 등과 같이 단 방향 스트리밍의 폭 넓은 서비스 분야에 주로 이용되어 왔다. 이처럼 기존 코덱들은 각 신호의 특성과 응용 분야의 요구에 따라 서로 다른 방식으로 연구 개발 되어 왔다.

그러나 통신과 방송을 융합하기 위한 노력이 진행되고, 다양한 멀티미디어 응용 분야가 생성됨에 따라, 최근에는 음성과 오디오에 대한 구분 없이 단말기의 동일한 입력 음원으로 사용되고 있다. 하지만 오디오 코덱에서는 음성 입력 신호에 대하여, 그리고 음성 코덱에서는 오디오 입력 신호에 대하여 음질이 저하되는

* 정회원, 연세대학교 전기전자공학과
(School of Electrical and Electronic Engineering, Yonsei University)

** 정회원, LG 전자
(LG electronics Inc.)

접수일자: 2010년7월5일, 수정완료일: 2010년8월9일

결과가 나타났으며, 이러한 현상은 저 전송률에서 더욱 심각하였다. 따라서 기존의 음성과 오디오 코덱을 대체할 수 있는 통합 코덱에 대한 요구가 한층 증가하게 되었다.

3GPP는 이러한 요구에 발 맞춰 AMR-WB^[1]라는 통합 코덱을 표준화 하였다. 기존 AMR-WB^[2] 음성 코덱에, 주파수 영역 부호화 기법인 TCX (Transform coded excitation) coding과 대역폭 확장(Bandwidth extension)을 추가하여 만들어진 이 코덱은 오디오 신호를 부호화하는데 있어 상당한 개선이 있었지만, 대표적인 오디오 코덱인 HE-AAC^[3]만큼 음질을 향상시킬 수는 없었다.

HE-AAC를 표준화한 ISO/MPEG 또한 새로운 통합 코덱을 표준화하기 위하여 2007년 10월부터 논의를 시작하여 USAC(Unified speech and audio coding)^[4]이라는 이름의 코덱을 설계하였다. 이는 입력 신호의 특성을 구분하여 음성과 오디오 코덱을 선택적으로 동작하는 방식으로서 기존 기술들에 대한 개선작업을 통하여 표준 모델인 RM (Reference model)으로 발표하였다. 또한 이 RM은 MPEG 회의에서 다루어지는 다양한 기술 기고를 포함하여, 지속적인 성능 향상 과정에 있다^[5-6]. 기본적으로 음성과 오디오의 대표적 코덱인 AMR-WB와 HE-AAC를 결합하여 만든 USAC는 입력 신호의 특성에 따라 선택적으로 동작한다. 신호분류기(Signal classifier)가 전처리단에 위치하여 이와 같은 결정을 하며, 분류기 성능에 비해 최종 음성 품질 역시 많은 차이가 발생한다. 음성 신호의 경우 ACELP (algebraic code excitation linear prediction)^[7] 코딩 방식을 통하여 부호화 되며, 오디오 신호의 경우 심리 음향 모델(psychoacoustic model)을 기반으로 한 AAC (Advanced audio coding)방식을 통하여 부호화 된다.

현재 MPEG에서 제공하는 RM 소프트웨어에는 적절한 신호분류기나 심리 음향 모델 등 부호화기의 성능에 큰 영향을 주는 주요한 모듈들이 포함되어 있지 않다. 이로 인하여 현재의 부호화기는 성능 열화도가 매우 심해졌으며, 이를 해결하기 위하여 기존 RM 부호화기에 대한 성능 개선 프로젝트가 진행 중이다^[8]. 이 프로젝트의 일환으로 지난 92번째 MPEG 미팅에서 오픈 소스 기반의 새로운 RM 소프트웨어인 JAME이 소개 되었다^[9]. 본 논문에서는 USAC에 포함된 주요 기술과 오픈 소스 기반으로 새롭게 설계된 JAME에 대하여 상세히 설명한다.

II. USAC의 전체 구조

그림 1과 2는 USAC의 부호화 및 복호화 과정에 대한 전체 블록도이다^[6]. 시스템은 크게 4개의 파트로 구성되는데, 첫 번째 상위파트에서는 멀티채널 코딩, 대역폭 확장, 신호 분류를 담당한다. 그리고 음악신호를 위한 FD코딩영역, 음성 신호를 위한 TD코더 블록, 양자화 그룹이 추가되어 전체 시스템을 구성하고 있다. 입력신호를 받아 음성인지 오디오 신호인지 판단하는 부분이 신호분류기인데, 그 결과에 따라 TD코더와 FD코더가 선택적으로 동작한다. 전처리 그룹의 MPEG surround는 멀티채널 부호화를 위한 모듈이고, eSBR 과정을 통해 보다 효율적 코딩을 위한 대역폭 확장 기술이 적용된다. FD코더는 대표적인 오디오 코덱인 AAC의 구조와 유사하다. 그 중 Block-switching, AAC MDCT filter bank, 심리 음향 모델(Psychoacoustic

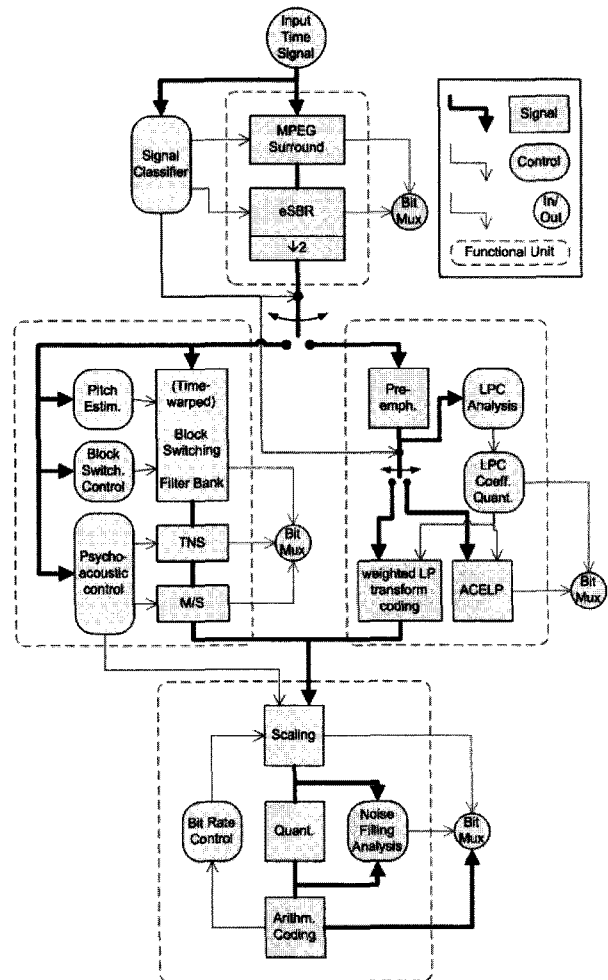


그림 1. USAC 부호화기(Encoder) 블록도
Fig. 1. USAC Encoder.

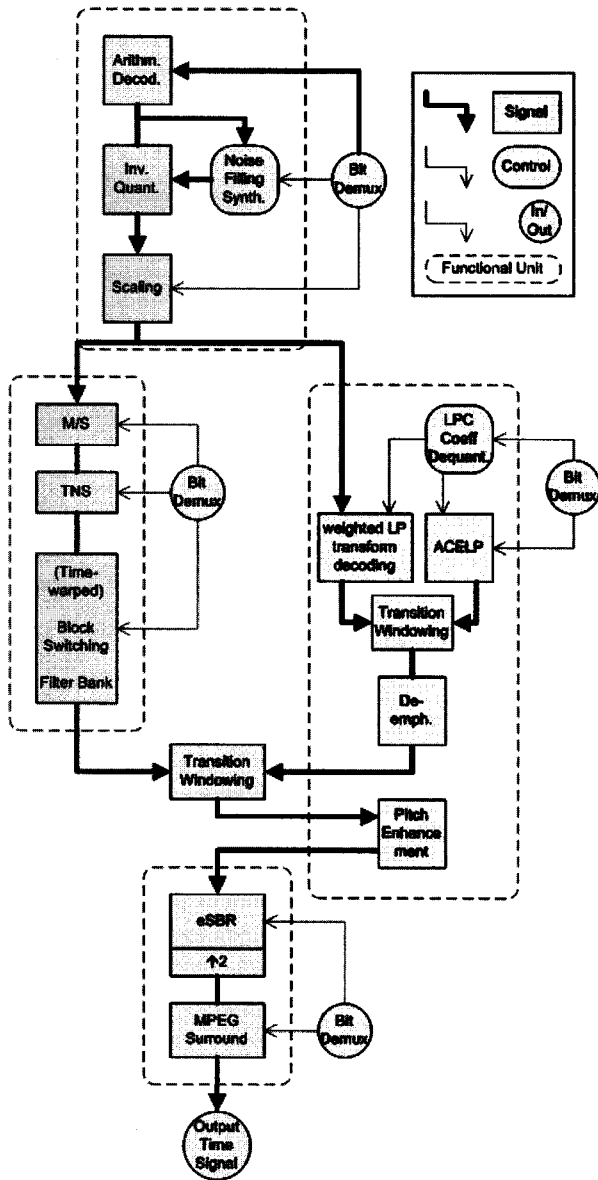


그림 2. USAC 복호화기(Decoder) 블록도^[6]
Fig. 2. USAC Decoder.

model), TNS(Temporal noise shaping), M/S (Middle-side stereo coding) 등은 심리 음향 모델을 이용하여 각 서브 밴드에서 원음 AAC와 동일한 기술을 상속받아 사용하는 것이다. 심리음향 모델에 의해 마스킹(Masking) 되어 들을 수 없는 최대의 잡음 레벨을 결정할 수 있으며, 이를 이용해 각 서브밴드에 최적의 비트를 할당한다. TNS 기법은 필터뱅크 윈도우 내에서 시간영역의 양자화 잡음을 적절하게 배치하여 청각적으로 들리지 않도록 하는 역할을 한다. 멀티 채널 신호에 대해서는 전처리 그룹의 MPEG Surround를 사용할 수도 있으며, FD 내부의 M/S 스테레오 부호화 기법을 사용할 수도 있다.

기존 AAC와는 달리 USAC에서 새로 채택한 기술로는 TW-MDCT(Time-warped modified discrete cosine transform), Noise Filling, CAAC 등이 있다. TW-MDCT는 추정된 피치 정보를 이용하여 신호를 시간영역에서 warping 하는데, 이는 주파수 영역에서의 주요 하모닉 성분들에 대한 코딩 효율을 향상시키기 위함이다^[10]. 또한 Noise filling은 스펙트럼 홀이 발생하는 구간에 잡음 신호를 채워 주는 방식으로 AAC의 PNS(Perceptual noise substitution)의 기능과 동일하다. CAAC는 FD 코더뿐만 아니라 TD 코더에서 추출된 파라미터를 양자화하는 방식인데, 과거 Huffman coding 기법 보다 복잡도는 높으나 압축 효율이 우수하여 채택되었다.

TD 코더는 기존의 AMR-WB+와 거의 유사한 형태를 가진다. LPC 필터링 후 생성된 잔여신호에 대해 음성(speech) 신호의 경우 ACELP(Algebraic CELP)를, 비음성(Non-speech)신호는 wLPT(weighted LP transform coding)방식을 이용하여 부호화한다. AMR-WB+의 ACELP 부분과 USAC의 ACELP 부분은 동일하며, USAC의 wLPT는 AMR-WB+의 TCX가 변형된 형태이다. wLPT는 FD 코더와 같이 MDCT를 사용하며, 잔여신호의 스펙트럼은 CAAC 기법을 사용하여 비트열로 만들어진다.

그림 2의 USAC 복호화기는 부호화기 그룹과 반대 순서로 비트열 복호화 그룹, TD 복호화기, FD 복호화기, 후처리 그룹으로 이루어진다. 비트열 복호화 그룹은 CAAC와 Noise filling으로 생성된 파라미터를 바탕으로 복호화 하며, TD 복호화기나 FD 복호화기를 선택적으로 거쳐, 후처리단을 통과한 후 원 신호를 복원한다.

III. USAC 새로운 참조 모델 : JAME

USAC 표준 코더와 알고리즘에 대한 이해를 쉽게 하여 가독성을 높이고, 새로운 기술을 포함하거나 기존 알고리즘의 수정을 편리하게 하도록 접근성을 높이기 위하여, RM의 체계적 구조 조정 및 RM의 부호화기 향상을 위한 프로젝트가 시작되었다^[8]. 현재, 각 MPEG 회의에서 논의되었던 중요 기술 및 수정 사항들이 RM에 포함되어 MPEG 버전 제어 저장소(MPEG Subversion repository)를 통하여 업데이트 되고 있다^[11]. 이 프로젝트의 일환으로 연세대학교와 LG전자에서는 RM을 분석하여 성능을 향상하고자 두 가지 주요한

연구 방향을 찾아 진행 중이다. 그 첫 번째로 불필요하고 중복된 파일과 모듈을 제거하여 코덱 구조를 간단하고 체계적으로 만드는 것이고 다른 하나는 다양한 모듈을 적절히 연동하여 성능 향상을 하는 것이다. 이러한 과정을 거쳐 설계된 새로운 RM은 오픈 소스 패러다임을 기반하고 있으며, 성능 향상을 위하여 누구라도 참여하여 이 소프트웨어를 활용할 수 있다. 새로운 RM의 이름은 JAME 이라 부르며, Windows 계열의 MSVC 6.0와 Linux 계열의 GCC 4.3.4에서 동작한다. JAME a.b.c 에서 a는 0와 1의 값을 가지며, 0의 경우 RM의 간략화 구조를 가지는 모델이며 1의 경우는 이를 바탕으로 성능 개선 및 최적화를 한 모델이다. b는 RM 또는 WD과 상응하는 버전을 의미하며, c는 내부 하위 버전을 의미한다. 현재 JAME 버전 0.7이 완성되었으며, 이는 MPEG SVN을 통하여 접근가능하다^[12].

1. 구조적 간결화

현재 RM은 USAC와 관련이 없거나 사용되지 않는 폴더와 파일 그리고 함수들을 포함하고 있다. 이는 과거 MPEG 표준 코더인 MPEG-2 AAC 와 MPEG-4 Audio standard 소프트웨어 모듈 위에 불필요한 부분을 제거하지 않고 구현되었기 때문이다. 소프트웨어의 가독성 향상을 위해 JAME은 사용되지 않는 불필요한 모듈을 매우 조심스럽게 제거하였다. 표 1에 정리한 것과 같이 상당히 많은 폴더와 파일들이 제거되어 프로그램 구조가 체계적이고 일관성을 가지게 되었다.

표 1에서 언급한 간략화 작업과 대응하여 전체 소프트웨어 플랫폼을 구조적으로 변경하였다. JAME은 Linux 시스템의 GCC와 Windows 시스템의 MSVC 6.0을 지원하며, 제거된 폴더 및 파일에 대한 프로젝트 관리 파일을 변경하여 일관성 있고 체계적인 구조를 가지게 되었다.

표 1. JAME 0.5와 RM5의 비교

Table 1. comparison of JAME 0.5 vs. RM5.

	RM5	JAME0.5	비율(%)
소스, 라이브러리 크기	20.9MB	14.9MB	71 %
소스 크기	11.7MB	5.8MB	50 %
폴더 수	152	19	13 %
파일 수	1144	385	34 %
C 파일 수	412	120	29 %
함수 수	23xx	8xx	3x %

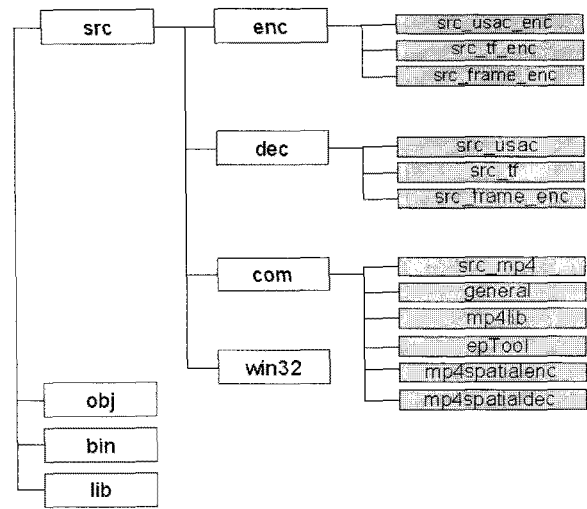


그림 3. JAME 0.7의 디렉터리 구조
Fig. 3. Structure of directory folder in JAME 0.7.

그림 3.은 JAME 0.7에서 제안되고 구현되어진 디렉토리 구조를 나타내고 있다. 소스 폴더(폴더 이름 : src) 외부의 lib, obj, bin, 는 각각 라이브러리, 오브젝트, 바이너리 실행파일 등이 생성되는 폴더이며, 플랫폼이 빌드 되어 질 때, 자동으로 생성된다.

제안된 JAME의 소스 구조는 2개의 계층으로 재개편 되었으며, 기존 RM과의 호환성을 위하여 그림에서 회색으로 칠해진 최하위 계층의 디렉토리 이름은 기존의 RM 이름과 동일하게 두었다. 152개나 되는 복잡한 RM의 디렉토리 구조와는 달리 첫 번째 계층은 부호화기(폴더 이름 : enc), 복호화기(폴더 이름 : dec), 공통 라이브러리(폴더 이름 : com), 그리고 윈도우 계열 지원 폴더(폴더 이름 : win32)로 나누었다. 또한 두 번째 계층에서 부호화기와 복호화기 폴더는 각각 TD 코더를 포함한 USAC 관련 폴더(폴더 이름 : src_usac, src_usac_enc), FD 코더 (폴더 이름 : src_tf, src_tf_enc) 프레임 관련 입출력 폴더(폴더명, src_frame, src_frame_enc)등으로 나누어 전체 소프트웨어 플랫폼을 대칭적이며 일관적으로 만들었다. 간략화 작업을 통하여 생기는 예상치 못한 문제들을 방지하기 위하여, JAME 0.x은 RMx을 통하여 생성된 결과 파일과 정확한 비트열을 생성하는지 확인하고 있다.

2. 성능 향상

JAME은 현재 RM의 플랫폼 구조를 간략화 하는 것 뿐만 아니라, 다양한 입력 신호에 대하여 부호화기의 성능을 높이려는 시도를 지속적으로 하고 있다. 따라서

다양한 모듈이나 툴, 알고리즘, 그리고 소스 등을 포함하고 있으며, 이는 앞으로도 지속적으로 확장될 것이다. 현재 3GPP와 같은 다른 표준화 기관에 의해 배포되는 소프트웨어 모듈 및 USAC의 부호화 관련 기고문들^[8]을 포함하고 있으며, 앞으로 성능 향상에 도움이 되는 다양한 학술 논문과 기술적 보고서, 접근 가능한 오픈 소스들을 받아들여 성능 개선을 할 것이다.

각 알고리즘 통합 후 높은 부호화 품질을 얻는지 확인하기 위하여, 참조 비트열(Reference bitstream)에 의해 얻어진 샘플과의 음질 비교를 통해 수정된 JAME의 성능을 입증할 수 있다. 참조 비트열은 RM 부호화기에서 생성된 것이 아니라, USAC의 RM을 배포하는 단체가 보유하고 있으며 공개하지 않고 있는 부호화기를 통하여 생성된 비트열을 말하며, 이러한 부호화기를 RQE (Reference Quality Encoder)라고 부른다.

가. 심리 음향 모델 적용

JAME 프로젝트의 궁극적인 목적은 실제 알고리즘 수정 및 첨부를 통해 성능을 향상시키는 것이며, 현재까지의 버전인 JAME 1.x에도 이미 다양한 기술들이 포함되어 있다. FD 코더의 성능향상을 위해, 3GPP의 EAAC+(Enhanced aacPlus) 코더의 MDCT(Modified discrete cosine transform)에 기반한 심리 음향 모델과 왜곡 제어 루프(Distortion control loop)가 그 대표적인 예이다. 이러한 모듈은 USAC의 각 블록에 주의 깊게 적절한 위치에 적용되고 최적화되었다. 심리 음향 모델이란 사람이 인지할 수 있는 최고의 음압 레벨과 최저의 음압 레벨사이의 관계를 말하는 것^[13]으로 적절한 문턱 값(Threshold)과 입력 신호의 인지적 엔트로피(Perceptual Entropy)를 구하는데 이용되며, 다음과 같은 수식을 이용하여 얻을 수 있다.

$$PE = \frac{1}{2\pi} \int_0^\pi \max[0, \log_2 \frac{X(e^{j\omega})}{T(e^{j\omega})}] d\omega \text{ (bit/sample)} \quad (1)$$

여기서, $T(e^{j\omega})$ 은 마스킹 문턱 값을 말하며, $X(e^{j\omega})$ 은 입력 신호의 에너지를 말한다. 수식 (1)에서 알 수 있듯이 적절한 심리 음향 모델을 적용하기 위하여 마스킹 문턱 값을 얻어야 하며, 이것은 아래와 같은 절차를 통하여 얻는다.

(1) 스프레딩 (Spreading)

주파수 도메인으로 전이된 신호는 밴드 별로 이전 밴

드와 다음 밴드의 신호 크기와의 비교를 통하여, 아래 두 수식과 같이 문턱 값을 얻는다.

$$thr'_{spr}(n) = \max(\frac{en(n)}{SNR}, S_h(n) \cdot \frac{en(n-1)}{SNR}) \quad (2)$$

$$thr_{spr}(n) = \max(thr'_{spr}(n), S_l(n)thr'_{spr}(n+1)) \quad (3)$$

여기서 $S_h(n), S_l(n)$ 은 주파수 밴드에 의존하는 상수 값이다.

(2) 프리-에코 (pre-echo)제어

프리에코 현상을 피하기 위하여 문턱 값은 아래 수식과 같이 이전 프레임의 문턱 값과의 비교 후에 얻어진다.

$$thr(n) = \max(0.01thr_{spr}(n), \min(thr_{spr}(n), 2thr_{spr,-1}(n))) \quad (4)$$

이와 같이 얻어진 문턱 값은 왜곡 제어 루프를 통하여, 주어진 비트-레이트에 맞게, 인지적 엔트로피 값에 따라 각 밴드별로 비트를 할당한다.

그림 4는 16kbps에서 심리음향 모델을 적용한 JAME 1.x의 성능을 보여준다. MUSHRA (Multiple Stimuli with Hidden Reference and Anchor) 와 유사한 형태의 테스트가 시행되었으며, eSBR과 MPS를 사용하지 않은 FD코더와 TD 코더에 의해 복호화된 신호를 사용하였다. 또한 숨겨진 참조신호(hidden reference)로 FD코더

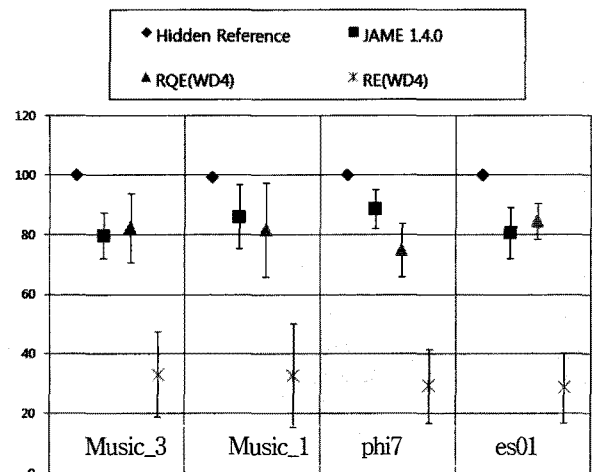


그림 4. JAME 1.x의 성능 비교 (16kbps mono)
Fig. 4. Quality comparison of core codec in the development of JAME 1.x (16kbps mono).

나 TD코더에 의해 복원되는 대역폭까지 필터링 시킨 음원을 사용하였으며, 3.5kHz 저대역 통과 신호인 앵커(anchor)없이 시행되었다. 95%의 신뢰구간을 가지며, RE는 RM 부호화기로 만들어진 음원을, RQE는 Reference bitstream에서 만들어진 음원을 의미한다. 테스트 음원으로는 USAC 테스트 샘플 가운데 3개의 음악 샘플 (Music_3, Music_1, phi7)과 하나의 음성 샘플 (es01)이 사용되었다. 그림에서 볼 수 있듯이 4개의 컨테츠에 대하여 JAME 1.x의 성능이 RQE에 비해 떨어지지 않으며, WD 보다 훨씬 뛰어난 성능을 나타내는 것을 알 수 있다.

나. eSBR 대역폭 확장

RM 부호화기를 거쳐 얻어진 음원의 대역폭은 RQE의 것에 비해 상당히 낮다. FD코더의 대역폭이 비트레이트가 상승함에 따라 커져야함에도 불구하고, 제한되어 있으며, 또한 모노 신호의 경우 eSBR의 시작 주파수(Start frequency)와 정지 주파수(Stop frequency) 모드와 상관없이 고정되어 있다. 따라서 RQE의 대역폭 설정에 맞추어 JAME의 eSBR의 시작 주파수와 정지 주파수 그리고 FD 코더의 대역폭을 표2와 같이 모드별로 변경하였다.

그 결과 그림 5와 같이 9.5kHz에서 대역이 제한된 신호가 그림 6과 같이 13.5kHz에서 대역이 제한된 신호를 얻을 수 있었으며, 확연히 음질 향상 효과를 얻을 수 있었다.

표 2. SBR table
Table 2. SBR table.

mode	FD코더	eSBR
	대역폭(kHz)	대역폭(kHz)
12kbps mono	3.8	12
16kbps mono	3.825	13.5
20kbps mono	4.25	16
24kbps mono	5.3	16
16kbps stereo	3.825	13.5
20kbps stereo	4.25	15.25
24kbps stereo	5.6	16
32kbps stereo	6.5	17.5
64kbps stereo	9.75	22.125

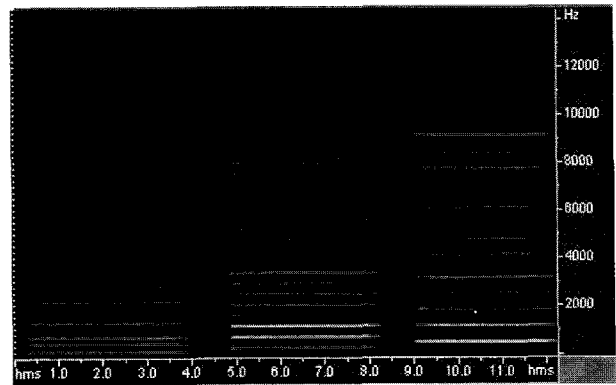


그림 5. phi7의 eSBR 대역폭 수정 전 스펙트럼 (16kbps mono)

Fig. 5. The spectrogram without modified the bandwidth of eSBR.

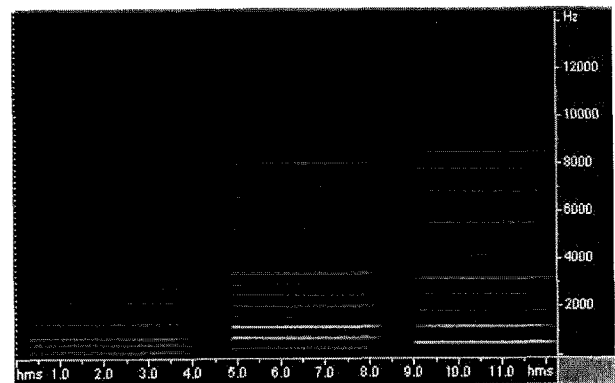


그림 6. phi7의 eSBR 대역폭 수정 후 스펙트럼 (16kbps mono)

Fig. 6. The spectrogram with modified the bandwidth of eSBR.

V. 결 론

본 논문에서는 최근에 제정된 MPEG 표준인 Unified Speech and Audio Coding과 그 오픈 소스 프로그램에 대하여 소개 하였다. USAC은 음성, 음악, 혼합(Mixed signal) 등의 신호에 대하여 우수한 음질을 가지는 새로운 표준 코덱을 완성하기 위해 노력하고 있다. 음성과 오디오 신호를 효과적으로 부호화하기 위하여 최신의 음성과 오디오 코덱을 선택적으로 사용한다. 각각의 기술들은 더 효과적이고 혁신적인 방법으로 보완되어, 다양한 방식으로 결합되고 있다. 또한 USAC RM 소프트웨어의 성능을 향상시키기 위한 활동으로 오픈 소스 기반의 USAC 참조 소프트웨어인 JAME을 소개하였다. 현재 MPEG SVN을 통하여 배포되고 있는 JAME은 간결하고 체계적인 구조를 가지고 있으며, 부호화기의 성능이 향상되고 있어 표준화 과정에서 사용하기 용이하

다. 표준화 완료 이후에는 신뢰할 수 있는 성능의 USAC 개발 소프트웨어로써 제품화 등 시장 활용을 기대할 수 있다.

참고 문헌

- [1] 3GPP, "Adaptive Multi-Rate - Wideband (AMR-WB) speech codec; General description," 2002, 3GPP TS 26.171.
- [2] 3GPP, "Audio codec processing functions; Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions," 2004, 3GPP TS 26.290.
- [3] ISO/IEC 14496-3:2009, "Coding of Audio-Visual Objects, Part 3: Audio," 2009.
- [4] ISO/IEC JTC1/SC29/WG11, "Call for Proposals on Unified Speech and Audio Coding," Shenzhen, China, Oct. 2007, MPEG2007/N9519.
- [5] ISO/IEC JTC1/SC29/WG11, "WD7 of USAC," Max Neuendorf, 2010.
- [6] Max Neuendorf et al., "A Novel Scheme for Low Bitrate Unified Speech and Audio Coding - MPEG RM0," in 126th AES Convention, M'unchen, Germany, May 2009.
- [7] M. Schroeder and B. Atal, "Code-excited linear prediction(celp):High-quality speech at very low bit rates," in Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing, pp. 937~940, 1984.
- [8] ISO/IEC JTC1/SC29/WG11 (MPEG), Document N11042, "Workplan on MPEG USAC Reference Encoder", 90th MPEG Meeting, Xi'an, China, October 2009
- [9] ISO/IEC JTC1/SC29/WG11 (MPEG), Document M17571, "Yonsei-LG Contribution to USAC Reference Software ", 92th MPEG Meeting, Dresden, Germany, April 2010
- [10] Bernd Edler, Sascha Disch, Stefan Bayer, GuillaumeFuchs, and Ralf Geiger, "A Time-Warped MDCT Approach to Speech Transform Coding," in 126th AES Convention, M'unchen, Germany, May 2009.
- [11] USAC SVN repository :
<http://wg11.sc29.org/svn/repos/MPEG-D/trunk/USAC/>
- [12] JAME SVN repository :
<http://wg11.sc29.org/svn/repos/MPEG-D/branches/JAME>
- [13] E.zwicker H.Fastl, "psychoacoustics," Springer, pp. 60, 1990.

저자 소개



송 정 욱(정회원)-교신저자
2004년 연세대학교 전기전자
공학과 학사 졸업.
2008년 연세대학교 전기전자
공학과 석사 졸업.
2008년~현재 연세대학교 전기
전자공학과 박사과정.

<주관심분야 : 디지털 신호처리, 음성 신호처리>



오 현 오(정회원)
1996년 연세대학교 전자공학과
학사 졸업.
1998년 연세대학교 전자공학과
석사 졸업.
2002년 연세대학교 전자공학과
박사 졸업.

2002년~현재 LG전자 Digital TV연구소
책임연구원

<주관심 분야: 오디오 신호처리, 심리음향, 오디오/음성 코덱 표준화>



강 흥 구(정회원)
1989년 연세대학교 전자공학과
학사 졸업.
1991년 연세대학교 전자공학과
석사 졸업.
1995년 연세대학교 전자공학과
박사 졸업.

현재 연세대학교 전기전자공학과 교수

<주관심분야 : 디지털 신호처리, 음성 신호처리, 오디오 신호처리>