

# MPEG 오디오 최신 표준: USAC 기술

이 태 진<sup>a)‡</sup>, 강 경 옥<sup>a)</sup>, 김 환 우<sup>b)</sup>

## MPEG Audio New Standard: USAC Technology

Taejin Lee<sup>a)‡</sup>, Kyeongok Kang<sup>a)</sup>, and Whanwoo Kim<sup>b)</sup>

### 요 약

다양한 기능을 가지는 모바일 기기들이 하나로 융합되어 가는 방향으로 기술이 발전함에 따라, 음성 및 오디오 모두에 대해 우수한 음질을 제공하는 부호화 기술에 대한 요구사항이 증대되고 있다. 이와 같은 새로운 부호화 기술에 대한 요구사항에 따라, MPEG에서는 2007년 10월 82차 회의에서 CFP를 시작으로 USAC 표준화를 시작하였고, 2011년 7월 97차 회의에서 FDIS까지 승인하였다. MPEG-D USAC 기술은 최신 음성 부호화기인 AMR-WB+와 최신 오디오 부호화기인 HE-AAC V2를 융합한 기술로 입력 신호의 특성에 따라 코어 대역 부호화로 AAC, ACELP, TCX 등 다양한 방법 중 하나를 선택하여 부호화를 수행하고, 고대역 부호화 기술로는 SBR, 스테레오 부호화 기술로는 MPEG-Surround를 이용하며, 다양한 코어 코더 사이의 천이를 위한 윈도우 기술을 이용한다. USAC 기술은 음성과 음악 신호 모두에 대해 우수한 음질을 제공할 수 있으며, 모바일 기기로의 멀티미디어 콘텐츠 다운로드, 디지털 라디오, 모바일 TV 및 오디오 북 등에서 응용이 가능하다.

### Abstract

As mobile devices become multi-functional, and converge into a single platform, there is a strong need for a codec that is able to provide consistent quality for speech and music contents. MPEG-D USAC standardization activities started at the 82nd MPEG meeting with a CFP and approved Study on DIS at the 96th MPEG meeting. MPEG-D USAC is converged technology of AMR-WB+ and HE-AAC V2. Specifically, USAC utilizes three core codecs (AAC, ACELP, and TCX) for low frequency regions, SBR for high frequency regions, the MPEG Surround for stereo information, and window transition technology for smoothing transition between various core coder. USAC can provide consistent sound quality for both speech and music contents and can be applied to various applications such as multi-media download to mobile devices, digital radio, mobile TV and audio books.

Keyword : MPEG-D USAC, AAC, AMR-WB+

## 1. 서 론

모바일 기기가 다양한 기능을 가지고, 다양한 기기를 하나의 모바일 기기로서 융합하는 방향으로 기술이 발전하고, 디지털 라디오, 오디오 북 등 음성과 음악신호 모두를 이용하는 응용분야의 시장이 커지면서, 음성과 오디오 신호 모두에 대해 우수한 품질을 제공하는 새로운 부호화 기술에

a) 한국전자통신연구원 방통융합미디어연구부  
ETRI(Electronics and Telecommunications Research Institute)

b) 충남대학교 전자공학과  
Chungnam National University

‡ 교신저자 : 이태진 (tjlee@etri.re.kr)

※ 본 연구는 방송통신위원회의 “무안경 다시점 3D 지원 UHDTV 방송 기술 개발”과제로 수행한 연구로부터 도출된 것입니다.

· 접수일(2011년6월24일), 수정일(2011년8월8일), 게재확정일(2011년8월8일)

대해 시장의 요구가 증대되고 있다<sup>[1]</sup>.

인간의 음성 생성 모델을 기반으로 하고 있는 음성 부호화 기술과, 인간의 청각 모델을 기반으로 하고 있는 오디오 부호화 기술은, 음성 통신과 음악 방송 등 각각의 독립적인 서비스 영역에서 독자적으로 기술 발전을 이루어 왔다. 하지만 최근 방송과 통신이 융합하는 방향으로 기술이 발전하면서, 음성 통신과 음악 방송으로 구분되던 서비스 구조가 깨어지고, 더 이상 음성과 오디오 신호를 별도의 콘텐츠로 분리하는 것이 어렵게 되었다<sup>[2]</sup>.

따라서 하나의 부호화기로 음성과 오디오 신호 모두를 처리할 수 있는 새로운 부호화 기술이 필요하게 되었다. 하지만 기존의 부호화 기술은 음성이나 오디오 신호의 특성을 기반으로 서로 다른 목표에 따라 개발되었기 때문에, 음성과 오디오 신호 모두에 대해 우수한 음질을 제공할 수 없다. 이에 따라 새로운 부호화 기술에 대한 표준이 요구되었고, MPEG에서는 음성과 오디오 신호를 통합적으로 부호화 하는 MPEG-D USAC (Unified Speech and Audio Coding) 기술의 표준화를 2007년 10월 Cfp (Call for Proposals)를 시작으로 진행하고 있으며, 2011년 7월 97차 MPEG 미팅에서 FDIS(Final Draft for International Standard)를 승인하였다<sup>[2][3][4]</sup>.

본 논문에서는 현재까지의 MPEG-D USAC 표준화 진행 상황과 RM (Reference Model) 기술에 대해 설명한다. 본 논문의 구성은 다음과 같다. 먼저 II장에서는 현재까지의 USAC 표준화 현황에 대해 기술하고, III장에서는 USAC의 특징 및 구조에 대해 기술한다. IV장에서는 USAC RM의 각 모듈별 특징에 대해 기술하고, V장에서는 주관적 청취평가 결과를 기반으로 USAC의 성능을 분석한다. 마지막으로 VI장에서 결론을 맺는다.

## II. MPEG-D USAC 표준화 현황

### 1. MPEG-D USAC Cfp 및 Evaluation Guideline

MPEG 오디오 서브그룹에서는 음성과 오디오 신호 모두에 대해 우수한 음질을 제공하는 새로운 부호화 기술의 개

발을 위해, 최신 음성 부호화 기술인 AMR-WB+<sup>[5]</sup>와 최신 오디오 부호화 기술인 HE-AAC V2<sup>[6]</sup>를 기본 부호화기 (Reference Codec)로 선정하였다. 청취평가에 사용할 각 아이템 별로 두 최신 부호화기 중 우수한 음질을 제공하는 부호화기를 VC (Virtual Codec)로 선정하여, 새로운 기술(NT: New Technology)은 동일한 비트율에서 VC보다 성능이 나쁘지 않아야 한다는 요구사항을 기본으로 표준화를 진행하고 있다<sup>[3][7]</sup>. 주관적 음질 청취평가 결과 AMR-WB+는 음성 신호에 대해서는 우수한 음질을 제공하지만, 오디오 신호에 대해서는 우수한 음질을 제공하지 못하고 있다. 또한 HE-AAC V2는 오디오 신호에 대해서는 우수한 음질을 제공하지만, 음성 신호에 대해서는 우수한 음질을 제공하지 못한다. USAC 기술개발의 목표는 그림 1과 같이 음성, 오디오 신호 모두에 대해 최신 음성과 오디오 부호화기만큼의 성능을 제공하는 것으로 시작하였고, 2011년 7월 현재, 산술 부호화(Arithmetic Coding) 튜, FAC(Forward Aliasing Cancellation) 튜, SBR(Spectral Bandwidth Replication) 튜 등 다양한 튜들의 성능을 개선하여 음성 신호에 대해 AMR-WB+ 보다 우수한 음질을 제공하고, 음악 신호에 대해 HE-AAC V2보다 우수한 음질을 제공한다.

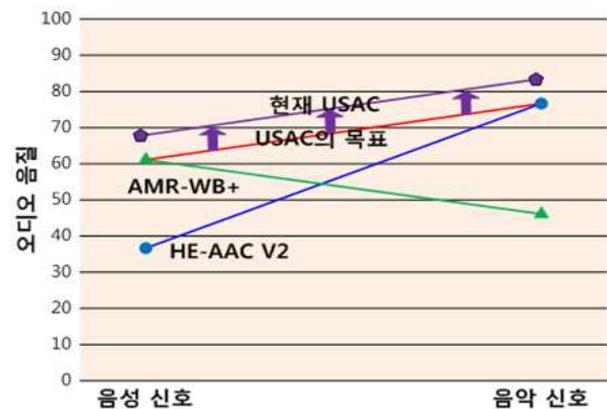


그림 1. MPEG-D USAC 표준화의 목표

Fig. 1. The purpose of USAC standardization

MPEG 오디오 서브그룹에서는 2007년 10월 MPEG-D USAC Cfp를 시작으로 본격적으로 새로운 부호화 기술에

대한 표준화를 진행하였고, 2008년 7월 RM0 선정을 거쳐서 2011년 3월 회의에서 Study on DIS를 승인하였고, 2011년 7월 97차 회의에서 FDIS를 승인하였으며, 2011년 11월에 IS (International Standard) 승인을 예정하고 있다. 표 1은 현재까지 진행된 각각의 MPEG 회의 및 관련 결과물을 요약한 것이다.

표 1. MPEG-D USAC 표준화 진행 상황  
Table 1. The progress of MPEG-D USAC Standardization

Meeting	Action
82차 MPEG 2007년 10월, 중국	MPEG-D USAC CfP issue Evaluation Guideline draft
83차 MPEG 2008년 1월 터키	Evaluation Guideline issue Workplan for Evaluation draft
84차 MPEG 2008년 4월 프랑스	Workplan for evaluation issue Exploration item selection
85차 MPEG 2008년 7월 독일	MPEG-D USAC CfP response Reference Model 0 selection
86 ~ 92차 MPEG 2008년 10월 한국 ~ 2010년 4월 독일	MPEG-D USAC WD1 ~ WD7 MPEG-D USAC Reference S/W - Arithmetic Coder Update - Unified Stereo Coding - Unification of window transition
93차 MPEG 2010년 7월 스위스	MPEG-D USAC CD - Improved Noiseless Coding
94차 MPEG 2010년 10월 중국	MPEG-D USAC Study on CD - Improved SBR/Harmonic Transposer
95차 MPEG 2011년 1월 한국	MPEG-D USAC DIS - Improved SBR structure
96차 MPEG 2011년 3월 스위스	MPEG-D USAC Study on DIS
97차 MPEG 2011년 7월 이탈리아	MPEG-D USAC FDIS

MPEG-D USAC Evaluation guideline<sup>[7]</sup>은 CfP에 참가하는 각 기관의 기술을 평가하는 방법을 기술한 문서로 청취평가 결과가 각각의 동작 모드(12kbps 모노 ~ 64kbps 스테레오) 및 카테고리(음성신호, 음악신호, 혼합신호)에서 VC

보다 음질이 나쁘지 않아야 한다는 기본 요구사항을 만족하는 기술에 대해 전체 청취평가 평균 점수, 기술의 복잡도 (Complexity), 지연시간(Delay) 등을 이용하여 RM0를 선정하는 방법을 기술하고 있다.

MPEG-D USAC Workplan for Subjective Testing of USAC Proposals<sup>[8]</sup>는 기술 기고 및 청취평가에 참여할 기관에 대한 정보와 표준화 일정에 대해 기술하고 있는 문서이다. USAC 표준화에는 국내에서는 ETRI, 삼성, LG가 참여하였고, 국외에서는 HE-AAC V2 기술을 가지고 있는 FhG, Dolby, Phillips와 AMR-WB+ 기술을 가지고 있는 VoiceAge, Thomson 등이 참여하였다.

## 2. MPEG-D USAC RM0 선정 및 WD 승인

2008년 7월 85차 MPEG 회의에서 각 기관의 기술기고 내용과 청취평가 결과를 분석하여 RM0를 선정하였다. 표 2는 기술 기고에 참여한 각 기관별 기술의 특징을 요약한 것이다<sup>[3]</sup>. 대부분의 기관이 음성 신호와 음악 신호에 대해 별도의 부호화 방식을 사용하였고, 대역확장을 위한 SBR<sup>[9]</sup>, 스테레오 부호화를 위한 PS (Parametric Stereo)<sup>[10]</sup> 기술을 적용하였다.

표 2. MPEG-D USAC 기술 기고 기관별 특징  
Table 2. The characteristics of USAC proposals for each proponents

	ETRI	Thomson	Samsung	LG	Dolby/Phillips	FhG/VoiceAge
Coding Method	AAC AMR-WB+	AAC Residual	CELP Frequency domain Coding	AAC AMR-WB+ Residual	AAC LPC/LTP	AAC AMR-WB+
Bandwidth Extension	SBR	SBR	SBR	SBR	SBR	SBR
Stereo Processing	Parametric Stereo	Parametric Stereo	Parametric Stereo	Parametric Stereo	Parametric Stereo	MPEG Surround
Speech Coding	ACELP	N/A	CELP based	ACELP	LPC, LTP in MDCT domain	ACELP
Lossless	AAC based	AAC based	Arithmetic coding	AAC based	AAC based	Arithmetic Coding

USAC RM0 선정을 위한 주관적 음질 평가에는 총 11개 기관에서 참여하였다. 청취평가 결과, 총 평균에서 FhG가 1위를 하였고, VoiceAge가 2위를 하였다(그림 2 참조). FhG와 VoiceAge의 기술 모두 VC보다 나쁘지 않아야 한다

는 요구사항을 만족하였으나, 두 기관은 공동 작업을 통해 동일한 복호화기를 이용하였고 따라서 Voice-Age 기술이 FhG 기술의 서브셋으로 볼 수 있어서, MPEG 오디오 그룹은 FhG/VoiceAge 기술을 RM0로 선정하였다<sup>[11]</sup>.

FhG/VoiceAge는 제 86차 MPEG 회의에서 USAC에 사용한 각 모듈별 기능 및 비트스트림을 기술한 WD와 RM0의 참조 부호화기/복호화기(Reference Encoder/Decoder)를 공개하였다<sup>[12]</sup>. 그 후 제 87차에서 92차 MPEG 회의까지 WD2에서 WD7까지를 승인하였는데, 산술 부호화 틀과 스테레오 부호화 틀, 통합 윈도우 천이 틀 등의 추가 및 성능 개선이 이루어 졌다. 이후 93차에서 CD (Committee Draft), 95차에서 DIS, 97차 MPEG 회의에서 FDIS를 승인하였고, 2011년 11월에 최종 IS를 승인할 예정이다.

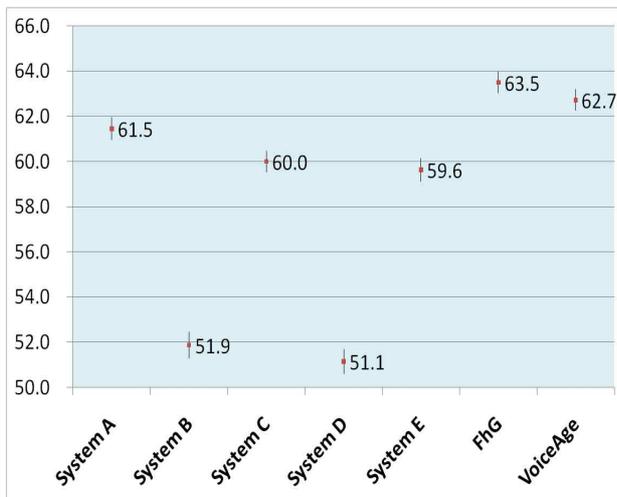


그림 2. MPEG-D USAC Evaluation 청취평가 결과  
Fig. 2. The listening test results for USAC Evaluation

### III. MPEG-D USAC 기술

#### 1. MPEG-D USAC 기술 특징

USAC 부호화기는 그림 3과 같이, HE-AAC V2와 AMR-WB+의 장점을 적절하게 융합한 코덱으로서, 입력신호의 특성에 따라 적절한 코어 코덱을 선택하여 동작한다.

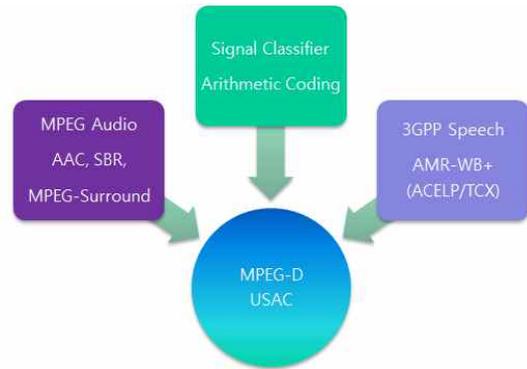


그림 3. MPEG-D USAC 특징  
Fig. 3. The characteristics of MPEG-D USAC

USAC은 MPEG Audio 부호화 기술에서 음악 특성 신호를 부호화하기 위한 AAC(Advanced Audio Coding), 고주파 대역을 부호화하기 위한 SBR, 스테레오 신호를 부호화하기 위한 MPEG-Surround 틀을 선택하였다. 다음 3GPP Speech 부호화 기술인 AMR-WB+의 ACELP(Algebraic Code-Excited Linear Predictor)와 TCX(Transform Codex eXcitation)를 이용하여 음성 특성 신호를 부호화 한다. 그리고 음성 특성 신호와 음악 특성 신호를 분류하기 위한 신호 분류 틀(Signal Classifier), 다양한 특성의 부호화 틀 사이의 부드러운 전이를 위한 모드 변환 틀 등을 이용하여 USAC은 음성과 음악 신호 모두에 대해 우수한 음질을 제공할 수 있다. 특히 USAC에서는 기존 AAC에 있는 허프만(Huffman) 방식의 무손실 부호화 틀을 산술 부호화 방식으로 변경하여 성능을 향상시켰으며, AMR-WB+의 TCX를 MDCT 기반으로 변경하고, AAC와 동일한 산술 부호화 방식으로 무손실 부호화를 수행한다.

#### 2. MPEG-D USAC 구조

USAC 부호화기는 그림 4와 같은 구조를 가지고 있다. 먼저 입력 신호는 신호 분류 틀을 이용하여 음성 특성 신호와 음악 특성 신호로 분류하여 적절한 부호화 방법을 선택한다. 음성 특성 신호인 경우 AMR-WB+와 같이 프레임별로 Segmental SNR을 계산하여 ACELP나 TCX를 선택하여 부호화 하는 LPD 모듈을 이용하여 부호화 하고, 음악

특성 신호인 경우 AAC를 개선한 FD 모듈을 이용하여 부호화를 수행한다.

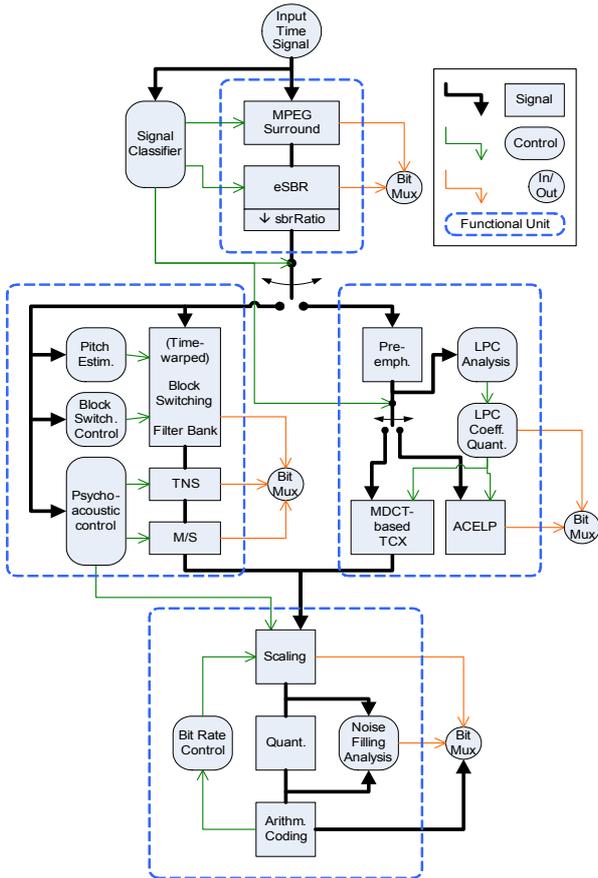


그림 4. MPEG-D USAC 구조  
Fig. 4. The structure of MPEG-D USAC

#### IV. MPEG-D USAC 툴 별 특징

##### 1. MPEG-Surround 툴

USAC의 스테레오 신호 부호화 툴은 MPEG-Surround 기술을 개선하여 이용한다. USAC은 기본적으로 매우 낮은 비트율에서 스테레오 부호화 방법을 정의하므로, MPEG-Surround 툴을 파라메트릭 스테레오 툴로써 사용하는데, 이는 MPEG Surround에 원래 정의되지 않은 2-1-2 모드를 기존 OTT box를 이용하여 구현하여 사용하고 있다.

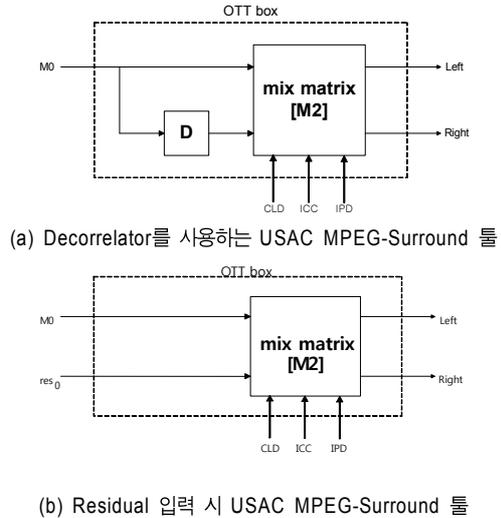


그림 5. USAC MPEG-Surround 툴  
Fig. 5. The MPEG-Surround tool for USAC

그림 5는 OTT box를 2-1-2 모드에 맞추어 설명한 그림이다. 기존 MPEG-Surround와 달리 M1 matrix가 필요 없으며, decorrelator는 1개만 있으면 된다. 또한 그림 5.(b)와 같이 residual 신호도 입력 받을 수 있다. USAC의 MPEG-Surround 2-1-2에서 새롭게 추가된 툴로 Unified Stereo Coding 툴이 있는데, 이는 residual 신호가 사용되는 stereo coding 모드에서 기존의 sum 신호(모노 신호)를 통해 부호화 할 수 있는 만큼 최대한 부호화 한 후, 남은 신호를 residual 신호로 표현하여 효과적으로 전송하여 음질을 향상하기 위한 툴이다.

USAC에서는 Unified Stereo Coding의 개념을 MPEG-Surround 툴이 사용되지 않는 고 비트율 모드에서도 적용할 수 있도록 확장한 툴을 도입하였다. 그림 6은 MPEG-Surround를 사용하지 않는 고 비트율 모드(64kbps 이상)에서 기존의 AAC 복호화 과정에서 사용하는 M/S stereo 대

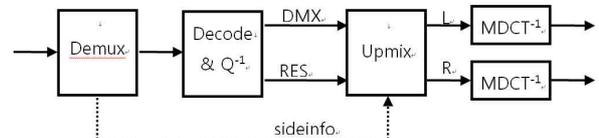
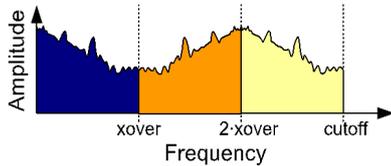


그림 6. 고 비트율에서 스테레오 모듈 성능 향상  
Fig. 6. The improved stereo decoding process in high bitrate

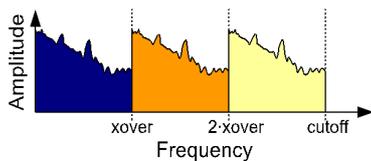
신에 DMX와 RES를 생성하듯 부호화하고, 이를 MPEG-Surround 파라미터를 이용하여 upmix 하듯이 L, R 신호를 생성함으로써 L/R 혹은 M/S의 양자택일 방식의 부호화 방법보다 개선된 부호화 효과를 기대할 수 있게 되었다.

### 2. eSBR 틀

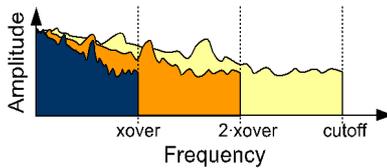
USAC에서는 고주파 대역 부호화 기술로 기존 SBR 기술을 개선한 eSBR(enhanced SBR) 틀을 이용한다. eSBR 기술은 MPEG-4 SBR을 확장한 것으로 특히 harmonic transposer를 추가적으로 사용할 수 있도록 한 것이 주요 특징이다. SBR에서는 고주파 대역을 저주파 대역의 정보를 활용하여 생성하는데 저주파 대역을 transposition 과정을 통해 고주파 신호로 복제한다. Transposition 방법으로는 그림 7과 같이 upsampling을 통한 mirroring 효과에 의해 생성하는 방법, 저주파수를 복제하거나 modulation해서 다시 붙이는 patch 방법, 그리고 phase vocoder를 이용한



(a) Upsampling (mirroring).



(b) Copying or modulation.



(c) Stretching (phase vocoder).

그림 7. SBR transposition 방법  
Fig. 7. The transposition method of SBR too

stretching과 decimation으로 설명할 수 있는 harmonic transposer 등의 여러 가지 방법으로 가능하다<sup>[13]</sup>. USAC eSBR은 그림 7(c)와 같은 phase vocoder 기반의 transition 방법을 추가하여 저 비트율에서 기존 SBR 보다 우수한 압축효과를 제공한다.

Phase vocoder를 이용한 SBR은 그림 8과 같이 저대역 신호를 Phase Vocoder를 이용하여 시간영역에서 N배 확장한 신호(time stretched)를 구하고, 이를 N배 다운 샘플링하여 출력신호를 구한다. 다운 샘플링 과정에서 스펙트럼이 확장되어 고대역에 새로운 성분을 생성하고, 이를 대역 통과 필터를 이용하여 SBR 대역 신호를 생성한다<sup>[3]</sup>.

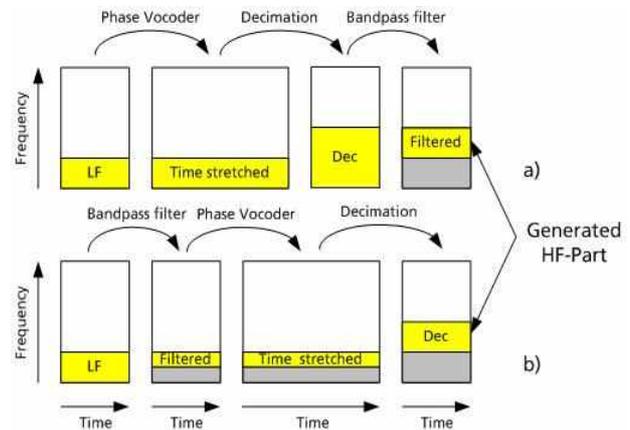


그림 8. SBR의 harmonic transposition  
Fig. 8. The harmonic transposition of SBR tool

### 3. FD 모드 틀

USAC의 FD 모드 틀은 MPEG-4 AAC<sup>[6]</sup>를 개선한 것으로, 무손실 부호화 방법으로 허프만 부호화 틀이 아닌 Context 기반의 산술부호화 틀을 이용하는 것과 LPD 모드와의 연결을 위한 새로운 윈도우 및 관련 틀을 추가한 것, Time-Warping 필터 बैं크를 이용하여 시간영역에서 서로 다른 해상도를 갖는 주파수 분해 기능을 제공하는 것 등이 특징이다. Time-Warping 필터 बैं크는 LPD 모드를 사용하지 않는 고 비트율에서 음성 신호가 입력될 경우 음성의 피치에 비례하여 코딩블록이 가변적으로 적용할 수 있도록 만든 틀로, 피치의 퍼짐에 인한 왜곡을 줄일 수 있다.

Time-Warping 필터 बैं크 변환 과정은 그림 9와 같다.

피치 윤곽을 추정하는 모듈로부터 변하는 피치값을 입력 받으면 이를 상수화시킬 수 있는 샘플링율을 규정하고 이에 따라 resample할 position을 계산한다. 또한 MDCT의 TDAC 구조를 보장하기 위한 윈도우 형태도 보정할 필요가 있다. 이와 같은 정보를 바탕으로 입력 신호를 resampling하고 변경된 윈도우를 처리한다. 이때 변환 블록 크기는 종래의 일반적인 MDCT와 동일한 2048 샘플이다<sup>[14]</sup>.

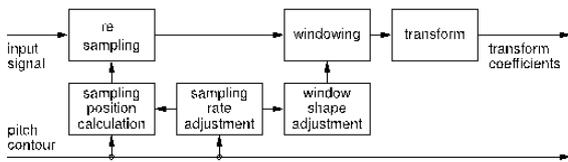


그림 9. Time-warping filterbank  
Fig. 9. Time-warping filterbank

USAC에서는 그림 10과 같이 기존 AAC에서 사용하던 윈도우(LONG\_WINDOW, LONG\_START\_WINDOW, 8\*SHORT\_WINDOW, LONG\_STOP\_WINDOW) 외에 LPD 프레임과의 연결을 위해 윈도우가 추가적으로 필요하며, Short 및 Long 프레임의 유연한 스위칭을 위해 STOP\_START\_WINDOW도 추가로 디자인하였다.

Value	Window	#coeffs	Window Shape
0	ONLY_LONG_SEQUENCE = LONG_WINDOW	1024/960	
1	LONG_START_SEQUENCE = LONG_START_WINDOW	1024/960	
2	EIGHT_SHORT_SEQUENCE = 8 * SHORT_WINDOW	8*(128/120)	
3	LONG_STOP_SEQUENCE = LONG_STOP_WINDOW	1024/960	
1	STOP_START_SEQUENCE = STOP_START_WINDOW	1024/960	
Note	Dashed lines indicate window shape when the adjacent window sequence is LPD_SEQUENCE		

그림 10. USAC에서 사용하는 윈도우  
Fig. 10. The window sequence of USAC

USAC에서는 noise filling이라는 이름으로 PNS를 대체한 틀을 FD 모드에서 사용한다. 기존 PNS 틀은 PNS가 on-

set되면 특정 밴드 이상의 모든 MDCT 계수는 0으로 세팅되고 해당 밴드들이 PNS로 대체되는데 반해, noise filling에서는 어떤 밴드 이상을 noise filling 틀로 대체하되, 해당 주파수 구간 안에 정상적으로 부호화된 스펙트럼 값이 중첩될 수 있는 것이 다르다.

USAC에서 사용하는 심리음향 모델(Psycho-acoustic model), TNS(Temporal Noise Shaping) 틀, M/S 스테레오 틀 등의 동작은 기존 AAC와 동일하다.

#### 4. LPD 모드 틀

USAC LPD 모드는 음성 특성 신호를 부호화하기 위한 것으로 LPC를 기반으로 부호화를 수행한다. LPD 모드는 Closed-loop Segmental SNR을 이용하여 ACELP나 TCX 틀을 이용하여 부호화를 수행하는데, AMR-WB+와 같이 1024 샘플의 슈퍼 프레임(super frame)을 입력 받아서, 256 샘플의 프레임 단위로 동작한다. LPD 모드에서는 입력 슈퍼 프레임을 256 샘플 프레임 단위로 LPC 분석을 실시하여 LPC1, LPC2, LPC3, LPC4의 4개의 LPC set을 계산하고, ACELP나 TCX 등 슈퍼 프레임의 모드에 따라 전송할 LPC set을 결정한다. USAC에서는 AMR-WB+와는 다른 새로운 방식으로 LPC를 양자화하는데, 이는 LPC set 사이에 변화가 적을 때 양자화 성능을 향상시키기 위한 것이다.

LPC set을 계산한 후, AMR-WB+처럼, 각 256 프레임 단위로 ACELP나, TCX256, TCX512 및 TCX1024 분석을 통해 슈퍼 프레임의 모드를 결정한다. USAC의 ACELP는 AMR-WB+와 유사하지만, TCX는 DFT 대신 MDCT를 사용하여 주파수 변환을 수행하고, AAC와 동일한 산술 부호화 방식으로 양자화를 수행한다.

MDCT 기반의 TCX 틀은 프레임의 TCX 모드에 따라 3가지 형태의 윈도우를 가질 수 있는데, 윈도우는 그림 11과



그림 11. USAC에서 사용하는 MDCT 기반의 TCX 윈도우  
Fig. 11. The MDCT based TCX window of USAC

같이 좌측 오버랩 L 샘플, 중간 M 샘플, 우측 오버랩 R 샘플과 좌/우측 zero 영역 ZL/ZR로 구성한다.

TCX 윈도우의 기본 오버랩 크기는 표 3에서 알 수 있듯이 256 샘플이지만, SHORT\_WINDOW와의 천이인 경우 L, R 오버랩 구간의 크기는 128이 된다.

표 3. 프레임 모드에 따른 TCX 윈도우

Table 3. The TCX window sequence based on frame mode

value of mod[x]	spectral coefficients	ZL	L	M	R	ZR
1	256	0	256	0	256	0
2	512	128	256	256	256	128
3	1024	384	256	768	256	384

### 5. 무손실 부호화 틀

USAC에서 무손실 부호화 방식으로 사용하는 산술 부호화 틀은 양자화된 스펙트럼에서 추가적으로 redundancy를 제거하기 위한 것으로, dynamically adapted context를 이용한다. 이를 위해 그림 12와 같이 시간과 주파수를 모두 고려한, 인접한 2개의 스펙트럼 성분으로 context 기반의 누적된 주파수 테이블을 이용하여 현재 상태를 결정하여 산술 부호화를 수행한다. USAC에 새로 추가된 산술 부호화 틀은 위와 같은 특징에 의해 기존의 허프만 부호화 방식에 비해 우수한 비트 감소 효과를 제공한다. FD 모드와 LPD 모드의 TCX가 사용되는 경우에 무손실 부호화 틀을 이용한다.

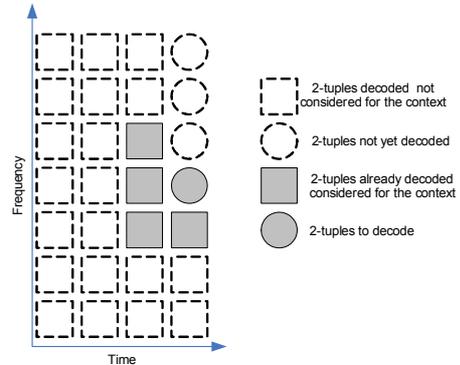


그림 12. State 계산을 위한 Context  
Fig. 12. The Context for the state calculation

### 6. USAC 윈도우 천이 틀

USAC은 FD와 LPD의 2개의 코어 모듈을 이용하여 부호화를 수행하며, FD 모듈은 Long 윈도우와 Short 윈도우가 있고, LPD 모드는 ACELP와 TCX 틀이 있어서 4가지의 다른 코딩 방식에 따른 총 16가지의 윈도우 천이가 발생한다. WD5 까지는 이러한 16개의 윈도우 천이를 처리하기 위해 특이한 형태의 윈도우를 정의하여 처리하였으나 TCX/ACELP에서 Short 윈도우 FD 모드로의 천이가 불가능하였고, TCX와 ACELP 사이의 천이 시 변환한 샘플을 버리는 등의 문제가 있었다. 이를 해결하기 위해 FAC 틀을 새로 추가하였는데, FAC 틀은 MDCT 기반의 코어 코더(FD

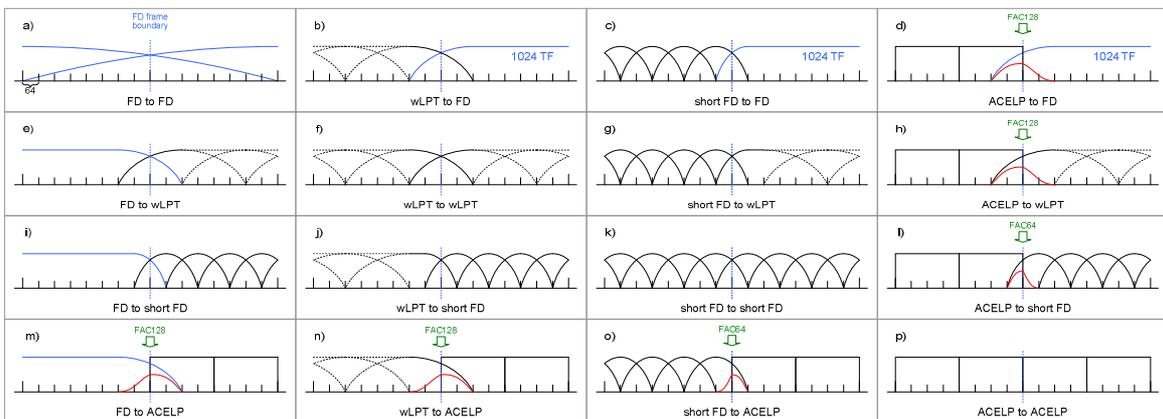


그림 13. USAC의 윈도우 천이  
Fig. 13. The window transition for USAC

모드, TCX)와 ACELP 사이의 천이 시, FAC 윈도우를 이용하여 TDAC를 제공하기 위한 것이다. FAC 톨을 이용하여 불가능하던 Short FD로의 천이가 가능해 졌고(그림 13 (l)), TCX와 ACELP 사이의 천이 시 샘플을 버리지 않고 TDAC 적용이 가능해 졌다(그림 13 (h), (n)). 또한 기존 FD와 TCX 모두 MDCT 기반의 변환 부호화 톨이지만, TCX 톨의 경우 LPC 필터 다음에 MDCT를 수행하였기 때문에 TDAC를 만족하지 못하던 문제점을 해결하기 위해 FDNS(Frequency Domain Noise Shaping)를 이용하여 TCX도 원 신호에 MDCT를 적용하여 FD와 TCX 변환 사이에 TDAC를 만족하게 되었다(그림 13 (b),(e),(g),(j)).

### V. MPEG-D USAC 성능

MPEG 오디오 서브그룹에서는 USAC 기술에 대한 verification test를 진행 하였다. USAC의 verification test는 청취평가 시간을 단축하기 위해 표 4와 같이 모노 저 비트율(8, 12, 16, 24kbps), 스테레오 저 비트율(16, 20, 24kbps), 스테레오 고 비트율(32, 48, 64, 96kbps)로 나누어서 청취평가를 진행하였고, RM0 선택 때와는 달리 각각의 카테고리에 8개의 청취평가 아이템을 이용하여, 총 24개의 아이템을 이용하여 청취평가를 수행하였다<sup>[15]</sup>.

그림 14/15/16은 각각 Mono low bitrate/Stereo low

표 4. USAC Verification test  
Table 4. USAC Verification

Mono at Low rates		Stereo at Low rates		Stereo at High rates	
Condition	Label	Condition	Label	Condition	Label
Hidden reference	HR	Hidden reference	HR	Hidden reference	HR
LP anchor at 3.5 kHz	LP3500	LP anchor at 3.5 kHz	LP3500	LP anchor at 3.5 kHz	LP3500
LP anchor at 7 kHz	LP7000	LP anchor at 7 kHz	LP7000	LP anchor at 7 kHz	LP7000
AMR-WB+ at 8 kbps	AMR-8	AMR-WB+ at 16 kbps	AMR-16	AMR-WB+ at 32 kbps	AMR-32
AMR-WB+ at 12 kbps	AMR-12	AMR-WB+ at 24 kbps	AMR-24	HE-AAC v2 at 32 kbps	HE-AAC-32
AMR-WB+ at 24 kbps	AMR-24	HE-AAC v2 at 16 kbps	HE-AAC-16	HE-AAC v2 at 64 kbps	HE-AAC-64
HE-AAC v2 at 12 kbps	HE-AAC-12	HE-AAC v2 at 24 kbps	HE-AAC-24	HE-AAC v2 at 96 kbps	HE-AAC-96
HE-AAC v2 at 24 kbps	HE-AAC-24	USAC at 16 kbps	USAC-16	USAC at 32 kbps	USAC-32
USAC at 8 kbps	USAC-8	USAC at 20 kbps	USAC-20	USAC at 48 kbps	USAC-48
USAC at 12 kbps	USAC-12	USAC at 24 kbps	USAC-24	USAC at 64 kbps	USAC-64
USAC at 16 kbps	USAC-16			USAC at 96 kbps	USAC-96
USAC at 24 kbps	USAC-24				

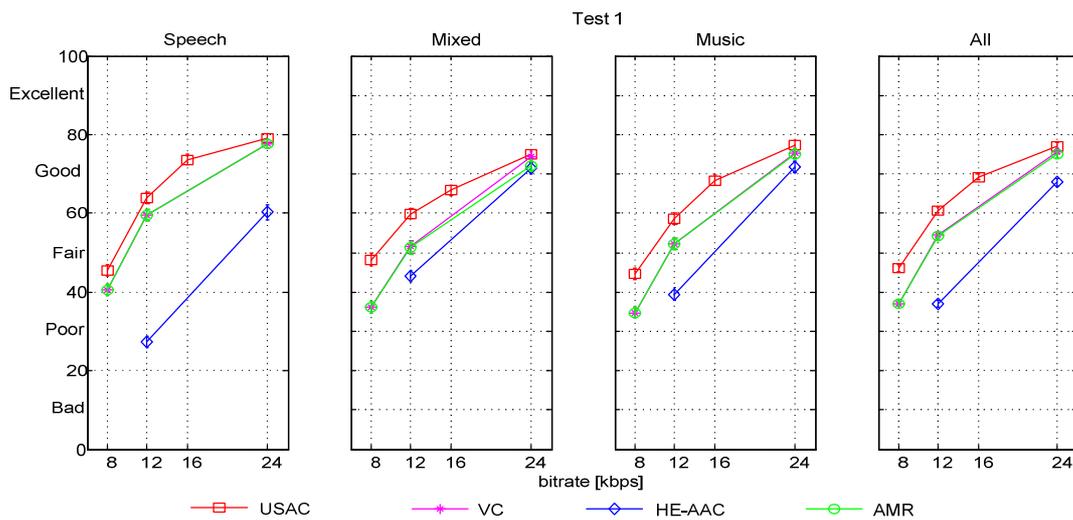


그림 14. USAC Verification Test 결과: Mono low bitrate  
Fig. 14. USAC Verification Test results: Mono low bitrate

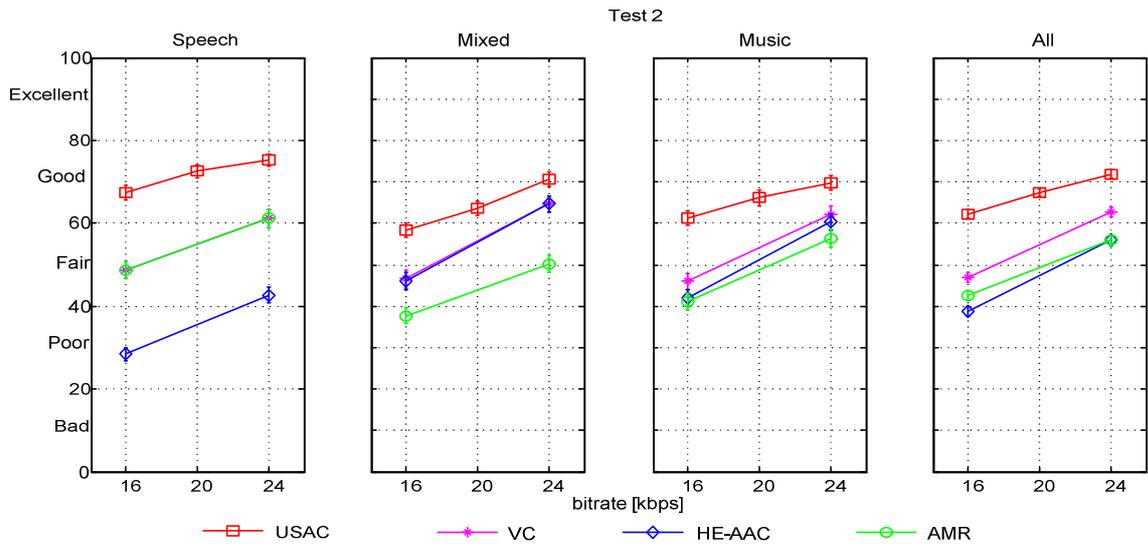


그림 15 USAC Verification Test 결과: Stereo low bitrate  
 Fig. 15. USAC Verification Test results: Stereo low bitrate

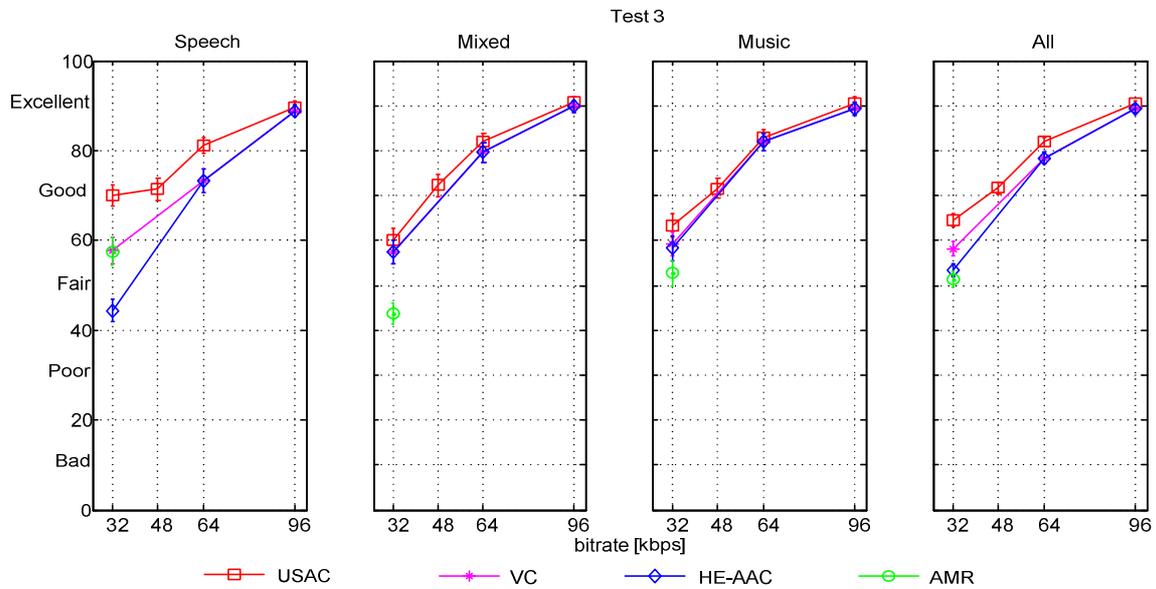


그림 16 USAC Verification Test 결과: Stereo high bitrate  
 Fig. 16. USAC Verification Test results: Stereo high bitrate

bitrate/Stereo High bitrate에 대한 USAC Verification Test 결과를 보여준다. 그림에서 알 수 있듯이 USAC 기술은 음성/혼합/음악 신호 모두에 대해 HE-AAC V2와 AMR-WB+보다 나쁘지 않은 음질을 보여주고 있으므로, USAC의 기본 요구사항을 만족하고 있다. 특히 USAC 기술은 기존의

최신 오디오 부호화기나 최신 음성 부호화기가 입력 신호의 특성에 따른 성능의 변화 폭이 큰 것에 비해 입력 신호의 특성에 관계없이 일정한 음질을 보여주고 있으며, 저 비트율을 뿐 아니라 고 비트율에서도 기존 부호화기만큼의 우수한 음질을 보여주고 있다.

## VI. 결 론

본 논문에서는 MPEG 오디오 서브그룹에서 표준화가 진행 중인 MPEG-D USAC 기술관련 표준화 현황 및 RM0에 대해 기술하였다. 현재 USAC 표준화는 2011년 7월에 FDIS까지 승인되었다.

USAC은 음성과 음악 신호 모두에 대해 우수한 음질을 제공하는 기술로, 최신 음성 부호화기인 AMR-WB+와 최신 오디오 부호화기인 HE-AAC V2 기술을 기반으로 개발되어, 저대역 부호화 방식으로 음악 특성 신호를 위한 FD 모드, 음성 특성 신호를 위한 LPD 모드를 이용하고, 고대역 부호화 방식으로 SBR, 스테레오 부호화 방식으로 MPEG-Surround를 이용한다. USAC 기술은 표준화 초기에 최신 오디오 부호화기와 최신 음성 부호화기보다 나쁘지 않은 음질을 보여야한다는 기본 요구사항을 바탕으로 표준화가 진행되어왔고, 현재 음성과 음악 신호 모두에 대해 최신 부호화 기술보다 우수한 음질을 제공하고 있다.

음성과 음악 신호 각각에 특화되어 개발된 기존의 부호화 기술과 달리 USAC 기술은 음성과 음악신호 모두에 대해 우수한 음질을 제공하기 때문에, 디지털 라디오, 오디오 북, 모바일 방송 등 음성과 음악 신호가 혼합되어 서비스가 필요한 다양한 분야에서 응용이 가능하다.

## 참 고 문 헌

- [1] ISO/IEC SC29 WG11 N9519, Call for Proposals on Unified Speech and Audio Coding, 82nd MPEG Meeting, October, 2007.
- [2] 박호종, 박영철, "MPEG 오디오 표준기술: USAC," 방송공학회지, 14권, 2호, 63~74쪽, 2009.
- [3] 이태진, 강경욱, 김환우, "MPEG-D USAC: 통합 음성 오디오 부호화 기술", 한국음향학회 논문지, 제 28권, 제 7호, pp. 589-598, 2009.
- [4] ISO/IEC SC29 WG11 N12013, Study on ISO/IEC 23003-3:201x/DIS of Unified Speech and Audio Coding, 96th MPEG Meeting, March, 2011.
- [5] 3GPP TS 26.290 V6.3.0, Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec, 2007.
- [6] ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio, 2005
- [7] ISO/IEC SC29 WG11 N9638, Evaluation Guidelines for Unified Speech and Audio Proposals, 83rd MPEG Meeting, Antalya, Turkey, January, 2008.
- [8] ISO/IEC SC29 WG11 N9927, Workplan for Subjective Testing of Unified Speech and Audio Coding Proposals, 84th MPEG Meeting, Archamp, France, April, 2008.
- [9] ISO/IEC 14496-3, AMD. 1, Bandwidth Extension, 2003.
- [10] ISO/IEC 14496-3, AMD. 2, Parametric Coding of High Quality Audio, 2004.
- [11] ISO/IEC SC29 WG11 N10047, Report on Unified Speech and Audio Coding Call for Proposals Submissions, 85th MPEG Meeting, July, 2008.
- [12] ISO/IEC SC29 WG11 N10215, WD on Unified Speech and Audio Coding, 86th MPEG Meeting, October, 2008.
- [13] Frederik Nagel, Sascha Disch, and Nikolaus Rettelbach, "A PHASE VOCODER DRIVEN BANDWIDTH EXTENSION METHOD WITH NOVEL TRANSIENT HANDLING FOR AUDIO CODECS," in 127thAESConvention, New York, USA, Oct. 2009.
- [14] 이태진, 강경욱, 오현오, MPEG-D USAC 표준화 동향 및 기술 추적 보고서, MPEG 포럼, 2010
- [15] ISO/IEC SC29 WG11 N12027, Workplan for Verification Testing of USAC, 964th MPEG Meeting, Geneva, Switzerland, March, 2011.

## 저 자 소 개



이 태 진

- 1996년 2월 : 전북대학교 전자공학과 (공학사)
- 1998년 2월 : 전북대학교 대학원 전자공학과 (공학석사)
- 1998년 2월 ~ 2000년 5월 : Mobens, Co., Ltd. 영상음성기술연구팀 연구원
- 2002년 10월 ~ 2003년 9월 : 일본 Tokyo Denki University, 방문연구원
- 2000년 5월 ~ 현재 : ETRI 실감음향연구팀 선임연구원
- 2009년 3월 ~ 현재 : 충남대학교 전자전파정보통신공학과 박사과정
- 주관심분야 : 오디오 신호처리 및 부호화, 3D 오디오

---

저 자 소 개

---



강 경 옥

- 1985년 2월 : 부산대학교 물리학과 (이학사)
- 1988년 2월 : 부산대학교 대학원 물리학과 (이학석사)
- 2004년 2월 : 한국항공대학교 대학원 항공전자공학과 (공학박사)
- 2006년 4월 ~ 12월 : 영국 Southampton University, ISVR 방문연구원
- 1991년 2월 ~ 현재 : ETRI 실감음향연구팀장, 책임연구원
- 주관심분야 : 오디오 신호처리 및 부호화, 3D 오디오



김 환 우

- 1977년 : 서울대학교 전자공학과 (공학사)
- 1979년 : 한국과학기술원 전기 및 전자공학과 (공학석사)
- 1988년 : University of Utah (공학박사)
- 1980년 ~ 현재 : 충남대학교 전자공학과 교수
- 주관심분야 : 신호처리, 디지털통신, 유무선 통신