

# MPEG-D USAC: 통합 음성 오디오 부호화 기술

## MPEG-D USAC: Unified Speech and Audio Coding Technology

이 태 진\*, 강 경 옥\*, 김 환 우\*\*  
(Taejin Lee\*, Kyeongok Kang\*, Whan-Woo Kim\*\*)

\*한국전자통신연구원 방통융합미디어연구부, \*\*충남대학교 전자전파정보통신공학과  
(접수일자: 2009년 8월 28일; 채택일자: 2009년 9월 30일)

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, 2009년 4월 88차 회의에서 WD3까지 완성되었다. 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 이용한다. USAC 기술은 음성과 음악 신호 모두에 대해 모두 우수한 음질을 제공할 수 있으며, 모바일 기기로의 멀티미디어 콘텐츠 다운로드, 디지털 라디오, 모바일 TV 및 오디오 북등에서 응용이 가능하다.

**핵심용어:** MPEG-D USAC, HE-AAC V2, AMR-WB+

**투고분야:** 뉴미디어 분야 (13)

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 content. MPEG-D USAC standardization activities started at the 82nd MPEG meeting with a CFP and approved WD3 at the 88th 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, and the MPEG Surround tool for stereo information. USAC can provide consistent sound quality for both speech and music content and can be applied to various applications such as multi-media download to mobile device, Digital radio, Mobile TV and audio books.

**Keywords:** MPEG-D USAC, HE-AAC V2, AMR-WB+

**ASK subject classification:** New Media (13)

### I. 서론

모바일 기기가 다양한 기능을 가지고, 다양한 기기를 하나의 모바일 기기로 융합하는 방향으로 기술이 발전하면서, 음성 및 오디오 신호 모두에 대해 우수한 품질을 제공하는 새로운 부호화 기술에 대해 시장의 요구가 증대되고 있다 [1].

각각 인간의 음성 생성 모델과 청각 모델을 기반으로 사용하는 음성 부호화 기술과 오디오 부호화 기술은, 음성 통신, 음악 방송 등 각각의 서비스 영역에서 독자적으로 기술 및 서비스의 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 [2].

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 있다.

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 USAC RM의 각 모듈별 특징에 대해 기술한다. IV장에서는 주관적 청취평가 결과를 기반으로 USAC의 성능을 분석하고, 마지막으로 V장에서 결론을 맺는다.

## II. MPEG-D USAC 표준화 현황

### 2.1. MPEG-D USAC CFP

MPEG 오디오 그룹에서는 음성과 오디오 신호 모두에 대해 우수한 음질을 제공하는 새로운 부호화 기술의 개발을 위해, 최신 음성 부호화 기술인 AMR-WB+ [3]와 최신 오디오 부호화 기술인 HE-AAC V2 [4]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 [1]. 두 최신 부호화 기술에 대한 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 것이다.

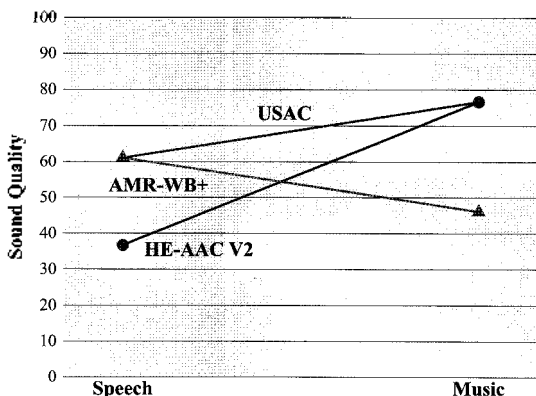


그림 1. MPEG-D USAC 표준화의 목표  
Fig. 1. The purpose of USAC standardization.

MPEG 오디오 그룹은 USAC의 동작모드를 모노 입력 신호에 대해, 12 kbps, 16 kbps, 20 kbps, 24 kbps, 스테레오 입력 신호에 대해 16 kbps, 20 kbps, 24 kbps, 32 kbps, 64 kbps 등 총 9 가지로 구성하였고, 청취평가에 사용할 아이템은 음성 신호 4개, 오디오 신호 4개, 음성과 오디오가 혼합된 신호 4개를 선정하였다.

MPEG 오디오 그룹에서는 2007년 10월 MPEG-D USAC CFP를 시작으로 본격적으로 새로운 부호화 기술에 대한 표준화를 진행하였고, 2008년 7월 RMO 선정을 거쳐서 2009년 7월 회의에서 WD3 (Working Draft 3)를 승인하였다. 표 1은 현재까지 진행된 각각의 MPEG 회의 및 관련 결과물을 요약한 것이다.

MPEG-D USAC Evaluation guideline [5]은 CFP에 참가하는 각 기관의 기술을 평가하는 방법을 기술한 문서로 청취평가 결과가 각각의 동작 모드 (12 kbps 모노 ~ 64 kbps 스테레오) 및 카테고리 (음성신호, 음악신호, 혼합신호)에서 VC보다 음질이 나쁘지 않아야 한다는 기본 요구사항을 만족하는 기술에 대해 전체 청취평가 평균 점수, 기술의 복잡도 (Complexity), 지연시간 (Delay) 등을 이용하여 RMO를 선정하는 방법을 기술하고 있다.

MPEG-D USAC Workplan for Subjective Testing of USAC Proposals [6]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.

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
86차 MPEG 2008년 10월 한국	MPEG-D USAC WD1 MPEG-D USAC Reference S/W
87차 MPEG 2009년 1월 스위스	MPEG-D USAC WD2 Workplan for USAC CEs
88차 MPEG 2009년 4월 미국	MPEG-D USAC WD3 Workplan for USAC CEs
89차 MPEG 2009년 6월 영국	Workplan for USAC CEs

## 2.2. MPEG-D USAC RMO 선정 및 WD 승인

USAC RMO 선정을 위한 주관적 음질 평가에는 기술을 기고한 기관과 France Telecom, Huawei, Ericsson 등 총 11개 기관에서 참여하였다. 2008년 7월 85차 MPEG 회의에서 각 기관의 기술기고 내용과 청취평가 결과를 분석하여 RMO를 선정하였다. 표 2는 기술 기고에 참여한 각 기관별 기술의 특징을 요약한 것이다. 대부분의 기관이 음성 신호와 오디오 신호에 대해 별도의 부호화 방식을 사용하였고, 대역확장을 위한 SBR (Spectral Bandwidth Extension) [7], 스테레오 부호화를 위한 PS (Parametric Stereo) [8] 기술을 적용하였다.

9가지의 USAC 동작모드에 대한 청취평가 결과, 총 평균에서 FhG가 1위를 하였고, VoiceAge가 2위를 하였다 (그림 2 참조). FhG와 VoiceAge의 기술 모두 VC보다 나쁘지 않아야 한다는 요구사항을 만족하였으나, 두 기관은 공동 작업을 통해 동일한 복호화기를 이용하였고 따라서 VoiceAge 기술이 FhG 기술의 서브셋으로 볼 수 있어서, MPEG 오디오 그룹은 FhG 기술을 RMO로 선정하였다 [9].

RMO로 선정된 FhG는 제 86차 MPEG 회의에서 USAC

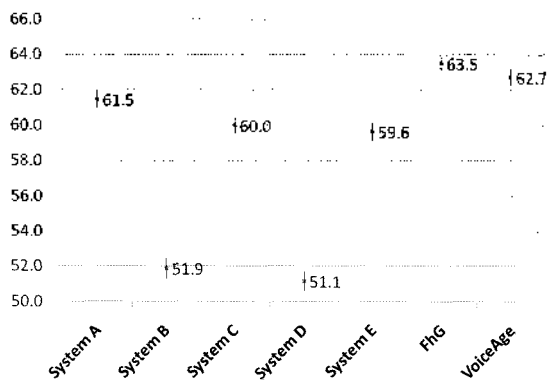


그림 2. MPEG-D USAC Evaluation 청취평가 결과  
Fig. 2. The listening test results for USAC Evaluation.

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

에 사용한 각 모듈별 기능 및 비트스트림을 기술한 WD와 RMO의 Reference Encoder/Decoder를 공개하였다 [10]. 그 이후에 제 87차 MPEG 회의에서 RMO의 산술 부호화 (Arithmetic coding) 부분과 변환 윈도우 (Transform window)를 개선한 WD2 [11]가 완성되었고, 제 88차 회의에서 위상 정보 기반 스테레오 부호화 (Phase Coding), AVQ (Algebraic Vector Quantization) 기반의 LPC 양자화기 등을 추가한 WD3 [12]가 완성되었다. 현재 MPEG에서는 TCX window transition, eSBR의 harmonic transition과 Enhanced Temporal Envelope Shaping, ACELP의 새로운 Pulse indexing scheme 등의 CE (Core Experiment)가 진행 중이며 [13], 2010년 1월 91차 회의에서 CD (Committee Draft)가 완성될 예정이다.

## III. MPEG-D USAC RM 기술

### 3.1. MPEG-D USAC RM 구조

USAC 부호화기는 그림 3과 같이 스테레오 부호화를 위한 MPEG-Surround 틀, 고대역 부호화를 위한 enhanced SBR 틀, 입력신호의 특성을 분석하여 저대역 부호화 방법을 선택하기 위한 신호 분류 틀 (Signal Classifier), 입력 신호의 특성에 따라 저대역 신호를 부호화하기 위한 ACELP, TCX (Transform Coded eXcitation) 및 AAC 틀, 서로 다른 저대역 부호화 방식 사이를 왜곡 없이 연결하기 위한 천이기술 등으로 이루어진다.

입력신호가 스테레오인 경우 MPEG-Surround 틀에 의해 스테레오 관련 파라미터를 추출한 다음, eSBR 틀에서 고대역 부호화를 위한 파라미터를 추출한다. USAC에서는 저대역 부호화 방식으로 AAC, ACELP, TCX의 3가지 방법이 있는데, 입력 신호 분류 틀에 의해 먼저, 음악

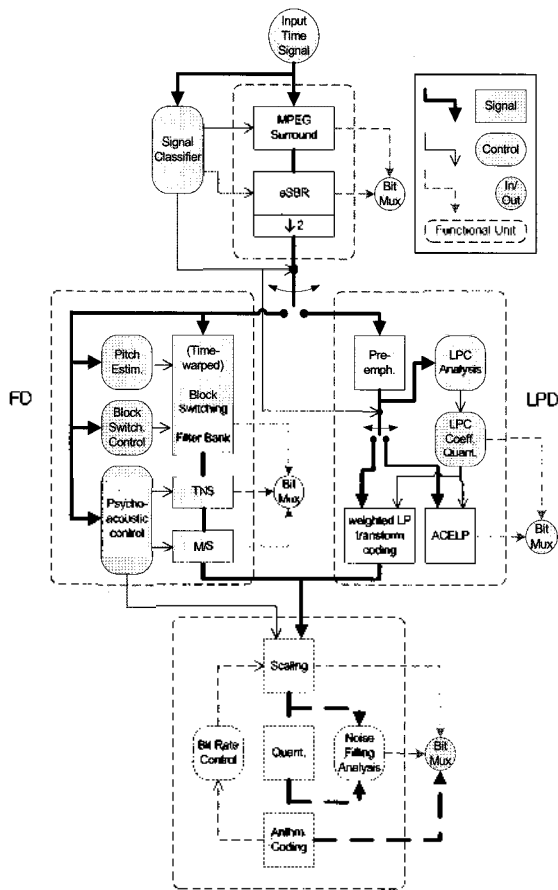


그림 3. MPEG-D USAC RMO 부호화기 구조  
Fig. 3. The encoder structure of USAC RMO.

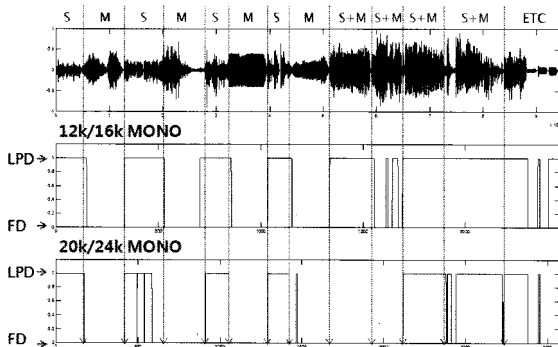


그림 4. MPEG-D USAC RMO: Signal Classifier 결과  
Fig. 4. The results of USAC Signal Classifier tool.

특성이 강한 경우 FD (Frequency Domain) 모드, 음성 특성이 강한 경우 LPD (Linear Prediction Domain) 모드로 분류된다. FD 모드는 AAC 기반의 저대역 부호화 방식이고, LPD 모드는 ACELP나 TCX 기반의 저대역 부호화 방식이다. LPD 모드 내에서 ACELP나 TCX의 선택은 lose-loop mode decision 방식에 따라 (서브 프레임의 SNR (Signal to Noise Ratio)을 이LP다. 그림 4 DoUSAC의 전체 청취평가 아이템 별로 신호 분류 툴의 출력을 보여준다. 그림에서 S는 음성 아이템, M은 음악 아이템,

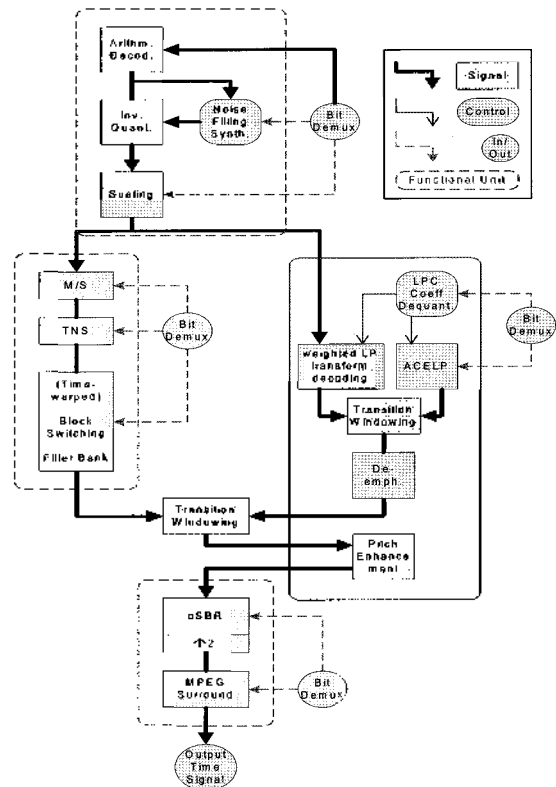


그림 5. MPEG-D USAC RMO 디코더 구조  
Fig. 5. The decoder structure of USAC RMO.

S+M은 음성과 음악 혼합 아이템을 의미한다. 그림에서 음성 신호와 혼합신호는 대부분 LPD 모드로 동작하고, 음악 신호는 FD 모드로 동작함을 알 수 있다.

그림 5는 USAC 복호화기의 구조이다. USAC 복호화기에서는 부호화기의 역순으로 복호화 과정을 통해 복원 동작을 수행한다. 입력 비트스트림은 FD나 LPD 등 부호화기의 동작 모드에 따라 복호화 툴을 동작시켜서 출력신호를 생성하고, 동작모드 사이의 천이를 왜곡 없이 연결하기 위해 특별히 설계한 윈도우를 이용한다.

### 3.2. MPEG-D USAC RMO 툴 별 특징

#### 3.2.1. AAC 툴

USAC의 AAC 툴은 HE-AAC V2의 AAC를 개선한 것으로, 무손실 부호화 (Lossless Coding) 방식으로 허프만 부호화가 아닌 새로운 Context 기반의 산술 부호화 방식을 이용하는 것과, LPD 모드와의 연결을 위한 새로운 윈도우를 추가한 것, Time-Warping 필터 뱅크를 이용하여 시간 영역에서 서로 다른 해상도를 갖는 주파수 분해 기능을 제공하는 것 등이 특징이라고 할 수 있다.

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

Value	Window	ecoeffs	Followed by an FD frame	Followed by an LPD frame
0	ONLY_LONG_SEQUENCE = LONG_WINDOW	1024/960		
1	LONG_START_SEQUENCE = LONG_START_WINDOW	1024/960		
2	EIGHT_SHORT_SEQUENCE = 8*SHORT_WINDOW	6*(128/120)		
3	LONG_STOP_SEQUENCE = LONG_STOP_WINDOW	1024/960		
1	STOP_START_SEQUENCE = STOP_START_WINDOW	1024/960		
3	STOP_1152_SEQUENCE = STOP_WINDOW_1152	1152/1080		
1	STOP_START_1152_SEQUENCE = STOP_START_WINDOW_1152	1152/1080		

그림 6. MPEG-D USAC AAC에서 사용하는 윈도우  
Fig. 6. The window shapes for USAC AAC tool.

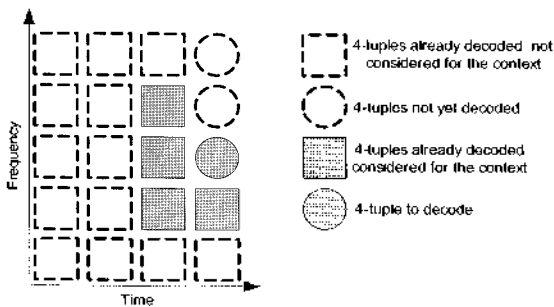


그림 7. 산술 부호화 틀의 상태 연산을 위한 Context  
Fig. 7. The Context for state calculation of USAC Arithmetic Coding tool.

USAC에서는 기존 AAC에서 사용하던 윈도우 (LONG\_WINDOW, LONG\_START\_WINDOW, 8\*SHORT\_WINDOW, LONG\_STOP\_WINDOW) 외에 LPD 프레임과의 연결을 위해 그림 6과 같이, 우측 중첩구간의 크기를 64로 변경한 새로운 윈도우를 이용한다.

USAC AAC에서 부손실 부호화 방식으로 사용하는 산술 부호화 틀은 양자화된 스펙트럼에서 추가적으로 redundancy를 제거하기 위한 것으로, dynamically adapted context를 이용한다. 이를 위해 그림 7과 같이 시간과 주파수를 모두 고려한, 인접한 4개의 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 저비트 부호화 방식에 비해 우수한 비트 감소 효과를 제공한다.

### 3.2.2. ACELP와 TCX 틀

USAC에서 사용하는 ACELP와 TCX는 LPC를 기반으로 신호를 부호화 하는 틀로써, AMR-WB+와 같이 1024 샘플의 슈퍼 프레임 (super frame)을 입력받아서, 256 샘플의 프레임 단위로 동작한다. 먼저 입력 슈퍼 프레임을 256 샘플 프레임 단위로 LPC 분석을 실시하여 LPC1,

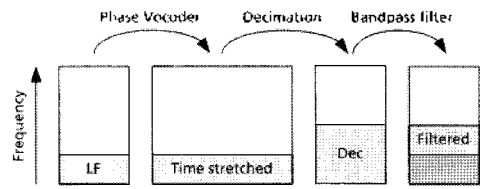


그림 8. MPEG-D USAC RM0: 하모닉 SBR  
Fig. 8. Harmonic SBR for USAC RM0.

LPC2, LPC3, LPC4의 4개의 LPC set을 계산하고, ACELP 나 TCX 등 슈퍼 프레임의 모드에 따라 전송 할 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 계산 한 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.

USAC의 TCX는 MDCT 기반이기 때문에, TDAC (Time Domain Aliasing Cancellation)를 위해 각 TCX 모드에 따른 다양한 형태의 천이 윈도우를 사용한다. TCX 윈도우에 대한 자세한 내용은 5절에서 자세하게 기술한다.

### 3.2.3. eSBR 틀

USAC에서 사용하는 고대역 부호화 기술은 HE-AAC V2의 SBR과 유사하며, 저 비트율에서 하모닉을 이용하는 기술을 추가하여 eSBR이라 한다.

하모닉 SBR은 그림 8과 같이 Phase Vocoder를 이용하여 고대역을 생성하는 방법이다. 저대역 신호를 Phase Vocoder를 이용하여 시간영역에서 N배 확장된 신호 (time stretched)를 구하고, 이를 N배 다운 샘플링 하여 출력신호를 구한다. 다운 샘플링 과정에서 스펙트럼이 확장되어 고대역에 새로운 성분을 생성하고, 이를 대역통과 필터를 이용하여 SBR 대역 신호를 생성한다 [10].

그림 9는 하모닉 SBR의 효과를 보여 주는 것으로, 일반적으로 SBR에서 사용하는 Patch 방법을 이용하는 경우, 저대역에 있는 하모닉 성분을 효과적으로 표현할 수 없지만, 하모닉 SBR에서는 효과적으로 하모닉 성분을 표현할 수 있음을 보여준다. 하모닉 SBR은 USAC의 9개 동작 모드 중에서 12 kbps, 16 kbps 모노와 16 kbps 스테레오 등 저 비트율에서만 동작하며, 기존 SBR대비 50% 정도의 비트율 감소가 가능하므로, 저 비트율에서 저대역에 더 많은 비트를 할당할 수 있다.

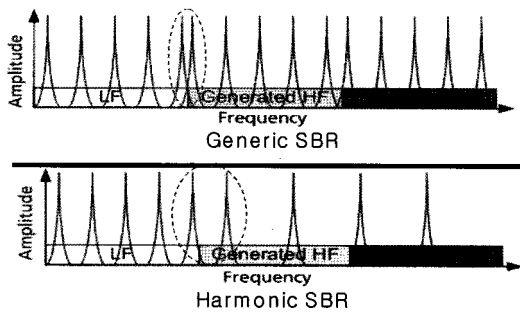


그림 9. MPEG-D USAC RM0: 하모닉 SBR 효과  
Fig. 9. The effect of USAC Harmonic SBR tool.

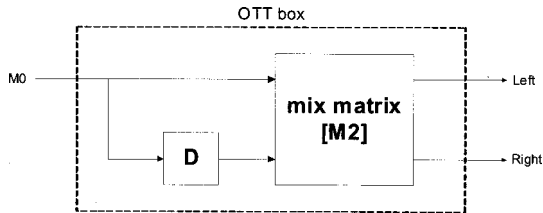


그림 10. MPEG-D USAC RM0: MPEG-Surround 툴  
Fig. 10. MPEG-Surround tool for USAC RM0.

3.2.4. MPEG-Surround 툴

USAC에서 스테레오 신호에 대한 부호화는 MPEG-Surround 기술을 변경하여 이용한다. 이는 MPEG-Surround의 5X5, 7X7 동작모드에 그림 10과 같이 단일 OTT (One-To-Two) 블록을 이용하여, 2-1-2 모드를 추가한 것이다. USAC 부호화기에서는 TTO (Two-To-One) 블록을 이용하여, 입력 스테레오 신호에서 각각의 파라미터 밴드에서 두 입력채널에 대한 전력비를 표현한 CLD (Channel Level Difference)와 상관관계를 표현한 ICC (Inter-Channel Correlation) 파라미터를 추출한다. USAC에서 MPEG-Surround 툴은 64 kbps 스테레오를 제외한 저 비트율에서 동작하고, 64 kbps에서는 M/S 스테레오나 독립적인 채널 부호화를 수행한다 [14] [15].

3.2.5. USAC의 윈도우 천이

USAC에서는 코어 내역인 저대역 부호화를 위해 AAC, ACELP, TCX 3개의 부호화 방식을 이용하기 때문에, 이들 사이를 천이하는 경우가 9가지이다. 따라서 이들 각각의 모드 사이의 천이 시 왜곡을 제거하는 것이 USAC의 핵심 기술 중 하나이다.

그림 11은 AAC, ACELP, TCX 모드에서 AAC 모드로의 천이를 보여주는 것으로, AAC에서 AAC로의 천이는 그림 11의 (a)와 같이, 기존의 AAC 천이와 같이 TDAC folding point에서 윈도우를 겹침으로써, TDAC를 수행한다. ACELP 나 TCX 등 LPD 모드에서 AAC로의 천이 시에는 그림 11의 (b) (c)와 같이 가상의 TDAC를 위해 복호화기에서 AAC

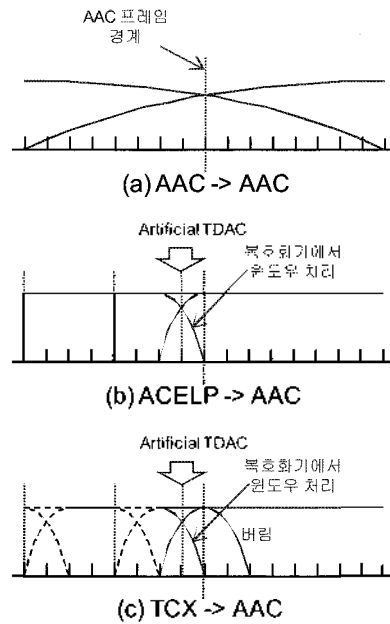


그림 11. MPEG-D USAC RM0: AAC 모드로의 천이  
Fig. 11. The transition to AAC mode for USAC RM0.

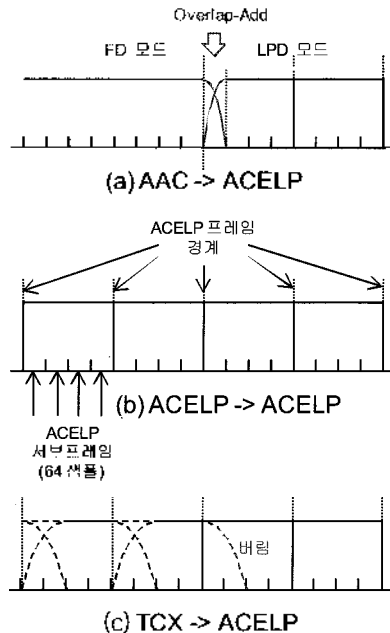


그림 12. MPEG-D USAC RM0: ACELP 모드로의 천이  
Fig. 12. The transition to ACELP mode for USAC RM0.

프레임에 대해 특별한 윈도우를 이용한다. 이때 TCX에서 AAC로의 천이 시에는 마지막 TCX의 우측 부분을 버리고, 가상의 TDAC를 수행한다. 이는 TCX와 AAC 모두 변환방법으로 MDCT를 사용하지만, TCX는 LPC 처리한 잔차 신호에 대해 MDCT를 적용하기 때문에, AAC 사이의 연결과 같은 방법을 사용하지 못하기 때문이다.

그림 12는 세 종류의 저대역 부호화 방식에서 ACELP 모드로의 천이 윈도우이다. 먼저 AAC에서 ACELP로의

천이 시는, 그림 12의 (a)와 같이, 64 샘플 중첩 가산 (Overlap-Add) 형태로 AAC와 ACELP 프레임을 연결한다. ACELP와 ACELP의 연결은 그림 12의 (b)와 같이, ACELP의 직각 윈도우 (Rectangular window)를 그대로 사용한다. TCX와 ACELP 모드의 연결 시에는 그림 12의 (c)와 같이, TDA가 발생하는 TCX의 우측 128 샘플 천이 구간을 버리고, 직각 형태의 ACELP 윈도우와 연결한다.

AAC, ACELP, TCX 모드와 TCX 모드 사이의 연결에는 그림 13과 같은 형태의 천이 윈도우를 사용한다. 먼저 AAC 모드에서 TCX 모드로의 천이 시에는, 그림 13의 (a)와 같이, 64 샘플 중첩 가산 방법을 이용한다. 실제 TCX 윈도우의 형태는 TCX256, TCX512는 좌측 천이 구간이 직각의 형태이기 때문에, 복호화기에서 임의로 겹치는 구간을 만들어서 AAC 프레임과 연결한다. TCX1024의 경우 좌측 천이 구간이 128 샘플이기 때문에, 이 구간을 버

리고, 겹치는 구간을 생성하여 연결한다.

ACELP 모드와 TCX 모드의 연결은 그림 13의 (b)와 같이 TCX 모드에 따라 다양한 방법을 이용한다. ACELP와 TCX256, TCX512를 연결할 때는 좌측 천이 구간이 직각인 특변한 형태의 윈도우를 이용하고, 이때 직각 형태의 윈도우에 의한 왜곡을 감소시키기 위해 ZIR (Zero Input Response)을 제거해 준다. ACELP 모드와 TCX1024 모드의 연결 시에는, TCX1024의 256 샘플 좌측 천이 구간을 버리고, ZIR을 제거하지 않고, 직접 ACELP의 직각 형태의 윈도우와 연결한다. TCX와 TCX 모드의 연결 시에는 모든 TCX의 좌/우측 천이구간의 크기가 128 샘플이기 때문에, 그림 13의 (c)와 같이, 128 샘플의 중첩 구간을 이용하여 TDA를 수행하는 형태로 연결한다.

지금까지 기술한 바와 같이 USAC에서는 AAC, ACELP, TCX 등 다양한 저대역 부호화 방식 사이의 효과적인 천이를 위해 다양한 형태의 윈도우를 사용하고 있다. 하지만, TCX와의 연결 시, 직각 형태의 윈도우를 이용하여 연결하거나, 복호화기에서 버려지는 부분이 있는 등 비효율적인 면이 있기 때문에, 현재 ETR1, VoiceAge 및 LG 등에서 좀 더 효과적인 천이 윈도우 관련 CB를 수행하고 있다.

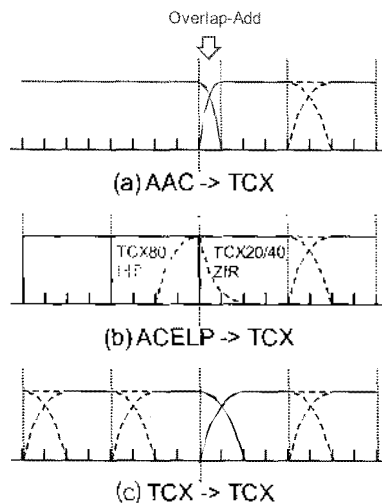


그림 13. MPEG-D USAC RMO: TCX 모드로의 천이.  
Fig. 13. The transition to TCX mode for USAC RMO.

## IV. MPEG-D USAC의 성능

### 4.1. MPEG-D USAC 청취평가

USAC 표준화 과정에서, 각 기관의 기술을 평가하기 위해 MUSHRA (Multiple Stimulus Hidden Reference and Anchor) 테스트를 이용한 주관적인 음질평가를 수행하였다. USAC의 청취평가에는 표 3에 기술한 12개의 아이템을 이용하였다 [5]. 12개의 아이템은 음성, 음악, 음

표 3. MPEG-D USAC 청취평가 아이템  
Table 3. The listening test items for USAC.

Item	Filename	Category	Duration	Order
1	es01	S	10.121	1
2	louis_raquin_15	S	15.000	3
3	Wedding_speech	S	8.640	5
4	te1_mg54_speech	S	7.965	7
5	twinkle_ff51	S+M	16.294	9
6	speechOverMusic_1	S+M	12.363	10
7	speechOverMusic_4	S+M	15.984	11
8	Harry Potter	S+M	15.033	12
9	salvation	M	16.400	2
10	te15_s	M	16.542	4
11	Music_1_s	M	15.649	6
12	Music_3	M	16.250	8

성 및 음악 혼합신호 각각 4개씩으로 이루어져 있다. 각각의 아이템은 표 3의 순서 (Order)에 따라 하나의 파일 형태로 구성되어 부호화 및 복호화를 수행한 후, 다시 각각의 아이템 별로 분리하여, 청취평가를 수행하였다.

**4.2. MPEG-D USAC 청취평가 결과 분석**

그림 14는 9개의 동작모드 별 USAC RM 청취평가 결과이다 [9]. USAC 기술은 그림에서 알 수 있듯이 저 비트율인 12 kbps에서는 55점대의 평가 점수를 얻었고, 고 비트율인 64 kbps에서는 80점 정도의 평가 점수를 얻었다.

II장에서 기술한 바와 같이, USAC 기술의 기본 요구사항은, 최신 음성 부호화기와 오디오 부호화기 중 음질이 우수한 것으로 선택되는 VC와 비교하여 음질이 나쁘지 않아야 된다는 것이다. 이러한 요구사항의 검토를 위해 음성, 음악, 음성 및 음악 혼합 신호별로 VC와의 차이를 비교해 보았다.

먼저 표 4는 9개의 동작모드와 12개의 아이템 별로 선

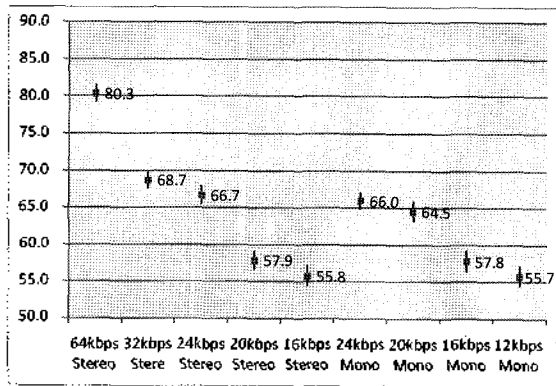


그림 14. MPEG-D USAC RMO 청취평가 결과  
Fig. 14. The listening test results for USAC RMO.

택된 VC를 보여준다 [9]. 표에서 알 수 있듯이 음악 신호에 대해서는 대부분 HE-AAC V2가 VC로 선택되었고, 음성 신호에 대해서는 AMR-WB+가 VC로 선택되었다. 음성과 음악 혼합 신호에 대해서는 저 비트율에서는 AMR-WB+, 고 비트율에서는 HE-AAC V2가 VC로 많이 선택되었다. VC 선택 결과를 보면 두 부호화기의 특성을 분명하게 보여준다. 특히 스테레오 신호인 경우, AMR-WB+에 비해 HE-AAC V2가 압도적으로 많이 VC로 선택되었는데, 이는 AMR-WB+보다 HE-AAC V2가 더 효과적으로 스테레오 신호를 부호화 할 수 있다는 것을 보여준다.

그림 15는 음성 신호에 대해 9개의 동작모드 별로 USAC과 VC와의 차이에 대한 그래프이다 [9]. 그림에서 알 수 있듯이 음성 신호에 대해 모든 동작모드에서 USAC 기술은 VC와 비슷하거나 우수한 음질을 제공한다. 다른 카테고리에 비해, 음성 신호에 대해 VC보다 많은 성능이 향상되었는데, USAC이 HE-AAC V2와 AMR-WB+에서 우수한 톨을 조합하였기 때문이다. 즉 음성 신호에 대해 저대역 부호화 방식으로 AMR-WB+의 ACELP나 TCX를 이용하고, HE-AAC V2의 SBR, MPEG-Surround를 이용한 스테레오 처리 등 우수한 톨을 조합하였기 때문에,

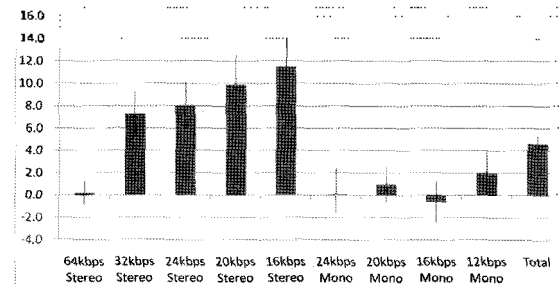


그림 15. USAC RMO 청취평가 결과 - 음성신호  
Fig. 15. The listening test results for USAC - Speech.

표 4. 동작모드 및 아이টে에 따라 선택된 VC  
Table 4. The selected VC for each operation modes.

	Test 1	Test 2	Test 3	Test 4	Test 5	Test 6	Test 7	Test 8	Test 9
음성 1	AAC	AMR	AMR	AMR	AAC	AAC	AMR	AMR	AMR
혼합 1	AAC	AMR	AMR	AMR	AMR	AMR	AMR	AMR	AMR
음성 2	AAC	AMR	AMR	AMR	AMR	AMR	AMR	AMR	AMR
음악 1	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AAC
음악 2	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AAC
음악 3	AAC	AAC	AAC	AAC	AAC	AMR	AMR	AMR	AAC
혼합 2	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AMR	AMR
혼합 3	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AMR	AMR
음성 3	AAC	AAC	AMR	AMR	AMR	AMR	AMR	AMR	AMR
음악 4	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AMR	AMR
혼합 4	AAC	AAC	AAC	AAC	AAC	AAC	AAC	AMR	AMR
음성 4	AAC	AAC	AAC	AAC	AAC	AMR	AMR	AMR	AMR



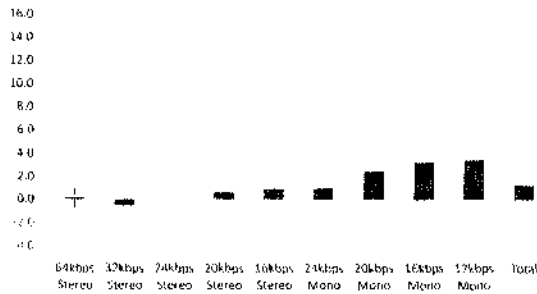


그림 16. USAC RMO 청취평가 결과 - 혼합신호  
 Fig. 16. The listening test results for USAC - Mixed.

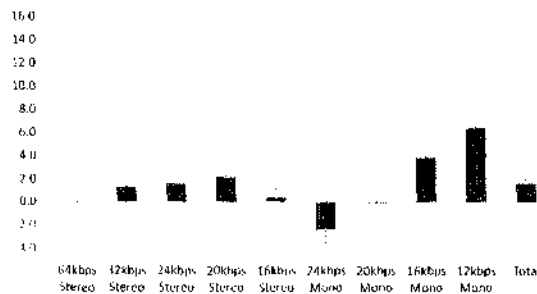


그림 17. USAC RMO 청취평가 결과 - 음악신호  
 Fig. 17. The listening test results for USAC - Music.

VC에 비해 성능이 많이 향상되었다.

그림 16은 음성과 음악 혼합 신호에 대해 9개의 동작모드 별로 USAC과 VC와의 차이에 대한 그래프이다 [9]. 음성과 음악 혼합 신호에 대해 모든 동작모드에서 USAC 기술은 VC와 비슷하거나 우수한 음질을 제공한다.

그림 17은 음악 신호에 대해 9개의 동작모드 별로 USAC과 VC와의 차이에 대한 그래프이다 [9]. 음악 신호에 대해 모든 동작모드에서 USAC 기술은 VC와 비슷하거나 우수한 음질을 제공할 수 있다. 그림에서 24 kbps 모노 신호에서는 USAC 기술이 VC보다 아래에 위치하지만, 신뢰구간이 겹치기 때문에, 나쁘지 않아야 한다는 USAC의 기본 요구사항은 만족한다.

VC와 USAC 기술의 카테고리별 청취평가 결과를 분석해 보면, 특히 저 비트율에서 (12 kbps, 16 kbps) 기존 기술에 비해 성능이 뚜렷하게 향상되었음을 알 수 있다. 이는 SBR이나 산술 부호화기 등의 기술을 이용하여 비트를 감소시켜서, 낮은 비트율 코디대역인 저대역 부호화 방식에 이용할 수 있기 때문이다.

청취평가 결과를 종합해 보면, USAC 기술은 음성과 음악 신호 모두에 대해 VC와 비슷하거나 더 나은 성능을 제공하기 때문에, 음성 특성 신호에 대해 강점이 있는 AMR-WB+와 음악 특성 신호에 대해 강점이 있는 HE-AAC V2의 장점을 효과적으로 결합한 부호화기라고 할 수 있다.

## V. 결론

본 논문에서는 MPEG 오디오 그룹에서 표준화가 진행 중인 MPEG-D USAC 기술관련 표준화 현황 및 LRM에 대해 기술하였다. 현재 USAC 표준화는 RMO가 선정되었고, WD3까지 승인되었으며, 2010년 1월에 CD가 승인될 예정이다. 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 AAC, ACELP, TCX를 이용하고, 고대역 부호화 방식으로 SBR, 스테레오 부호화 방식으로 MPXG-Surround를 이용한다. USAC 기술은 음성과 음악신호 모두에 대해 우수한 음질을 제공하기 때문에, 디지털 라디오, 오디오 북, 모바일 방송 등 다양한 형태의 응용이 가능하다.

현재까지의 MPEG-D USAC 기술은 음질을 위주로 개발해 왔기 때문에 복잡도나 지연 시간에 대해서는 많은 고려를 하고 있지 않다. 특히 저대역 부호화 방식으로 AAC, ACELP, TCX 등 3가지의 독립적인 방식을 이용하기 때문에, 기존 기술에 비해 복잡도가 많이 증가되었다. 또한 통신환경을 고려하고 있지 않기 때문에 지연시간 역시 큰 편이다. 따라서 향후 표준화가 진행됨에 따라 지연 시간 등의 개선이 필요하다.

## 감사의 글

본 연구는 방송통신위원회, 지식경제부 및 한국산업기술평가관리원의 IT 원천기술개발사업의 일환으로 수행한 연구로부터 도출된 것입니다. [과제관리번호: 2008-F-011, 과제명: 차세대 DTV 핵심기술 개발]

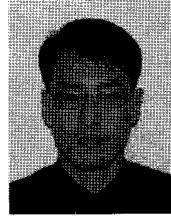
## 참고 문헌

1. ISO/IEC SC29 WG11 N9519, *Call for Proposals on Unified Speech and Audio Coding*, 82nd MPEG Meeting, October, 2007.
2. 박호중, 박영철, "MPEG 오디오 표준기술: USAC," *방송공학회지*, 14권, 2호, 63-74쪽, 2009.
3. 3GPP TS 26.290 V6.3.0, *Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec; Transcoding functions*, 2007.
4. ISO/IEC 14496-3:2005, *Information technology - Coding of audio-visual objects Part 3: Audio*, 2005
5. ISO/IEC SC29 WG11 N9638, *Evaluation Guidelines for Unified Speech and Audio Proposals*, 83rd MPEG Meeting, Antalya,

- Turkey, January, 2008.
6. ISO/IEC SC29 WG11 N9927, *Workplan for Subjective Testing of Unified Speech and Audio Coding Proposals*, 84th MPEG Meeting, April, 2008.
  7. ISO/IEC 14496-3, AMD, 1, *Bandwidth Extension*, 2003.
  8. ISO/IEC 14496-3, AMD, 2, *Parametric Coding of High Quality Audio*, 2004.
  9. ISO/IEC SC29 WG11 N10047, *Report on Unified Speech and Audio Coding Call for Proposals Submissions*, 85th MPEG Meeting, July, 2008.
  10. ISO/IEC SC29 WG11 N10215, *WD on Unified Speech and Audio Coding*, 86th MPEG Meeting, October, 2008.
  11. ISO/IEC SC29 WG11 N10418, *WD2 on Unified Speech and Audio Coding*, 87th MPEG Meeting, January, 2009.
  12. ISO/IEC SC29 WG11 N10661, *WD3 on Unified Speech and Audio Coding*, 88th MPEG Meeting, April, 2009.
  13. ISO/IEC SC29 WG11 N10848, *Workplan for USAC CEs*, 88th MPEG Meeting, July, 2009.
  14. Max Neuendorf et al., "Unified Speech and Audio Coding Scheme for High Quality at Low bitrates," ICASSP 2009.
  15. ISO/IEC 23003-1:2007, *Information technology - MPEG audio technologies - Part 1: MPEG Surround*, 2007.

## 저자 약력

### •이 태 진 (Taejin Lee)



1996.2: 전북대학교 전자공학과 (공학사)  
 1998.2: 전북대학교 대학원 전자공학과 (공학석사)  
 1998.2~2000.5: Mobens, Co., Ltd. 영상음성기술 연구팀 연구원  
 2002.10~2003.9: 일본 Tokyo Denki University, 방문연구원  
 2000.5~현재: ETRI 미디어응용연구팀 선임연구원  
 2009.3~현재: 충남대학교 전자전자정보통신공학과 박사과정  
 ※관심분야: 오디오 신호처리 및 부호화, 3D 오디오

### •강 경 옥 (Kyeongok Kang)



1985. 2: 부산대학교 물리학과 (이학사)  
 1988. 2: 부산대학교 대학원 물리학과 (이학석사)  
 2004. 2: 한국항공대학교 대학원 항공전자공학과 (공학박사)  
 2006.4~12: 영국 Southampton University, ISVR 방문연구원  
 1991. 2~현재: ETRI 미디어응용연구팀장, 책임연구원  
 ※관심분야: 오디오 신호처리 및 부호화, 3D 오디오

### •김 환 우 (Whan-Woo Kim)



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