

## 통합 음성 오디오 부호화 기술

이태진, 백승권, 강경옥, \*김환우  
한국전자통신연구원, \*충남대학교

[tjlee@etri.re.kr](mailto:tjlee@etri.re.kr), [skbeack@etri.re.kr](mailto:skbeack@etri.re.kr), [kokang@etri.re.kr](mailto:kokang@etri.re.kr), [\\*wwkim@cnu.ac.kr](mailto:*wwkim@cnu.ac.kr)

### Unified Speech and Audio Coding Technology

Taejin Lee, Seungkwon Beack, Kyeongok Kang, \*Whan-Woo Kim  
Electronics and Telecommunications Research Institute \*Chungnam National University

#### 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. MPEG 에서는 2008 년 10 월부터 MPEG-D USAC 기술에 대해 CfP 를 시작으로 본격적으로 표준화를 진행하고 있으며, 2011 년 3 월 96 차 미팅에서 Study on DIS 까지 승인하였다. 본 논문에서는 LPD 모드의 TCX 윈도우의 변경을 통한 USAC 성능향상 방법은 제안한다. TCX 프레임의 연결에 고정된 크기의 중첩만을 이용하는 현재의 방식과는 달리, 이전 TCX 모드와 다음 TCX 모드, transient 의 존재 유무에 따라 적절하게 TCX 윈도우 중첩 크기를 조절하여 음악 특성 신호에 대해 LPD 모드의 음질을 개선할 수 있다.

#### 1. 서론

모바일 기기가 다양한 기능을 가지고, 다양한 기기를 하나의 모바일 기기로 융합하는 방향으로 기술이 발전하고, 디지털 라디오, 오디오 북 등 음성과 음악신호 모두를 이용하는 다양한 응용분야의 시장이 커지면서, 음성과 오디오 신호 모두에 대해 우수한 품질을 제공하는 새로운 부호화 기술에 대해 시장의 요구가 증대되고 있다[1].

이에 따라 새로운 부호화 기술에 대한 표준이 요구되었고, MPEG 에서는 음성과 오디오 신호를 통합적으로 부호화 하는 MPEG-D USAC (Unified Speech and Audio Coding) 기술의 표준화를 2007 년 10 월 CfP (Call for Proposals)를 시작으로 진행하고 있고, 2011 년 3 월 96 차 MPEG 미팅에서 Study on DIS(Draft for International Standard)를 승인하였다[2] [3].

본 논문에서는 MPEG-D USAC 을 구성하고 있는 FD(Frequency Domain) 모드와 LPD(Linear Prediction Domain) 중 LPD 모드의 음질을 개선하기 위한 TCX 윈도우 기술을 제안하고, 청취평가 결과를 기술한다.

#### 2. MPEG-D USAC

USAC 부호화기는 그림 1 과 같이, 최신 오디오 부호화기인 HE-AAC V2[4]와 최신 음성 부호화기인 AMR-WB+[5] 의 장점을 적절하게 융합한 코덱으로서, 입력신호의 특성에 따라 적절한 코어 코덱을 선택하여 동작한다.

USAC 에서는 MPEG Audio 부호화 기술에서 음악 특성 신호를 부호화 하기 위한 AAC(Advanced Audio Coding), 고주파 대역을 부호화 하기 위한 SBR, 스테레오 신호를 부호화 하기 위한 MPEG-Surround 툴을 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 툴, 다양한 특성의 부호화 툴 사이의 부드러운 전이를 위한 모드 변환 툴 등을 이용하여 USAC 은 음성 과 음악 신호 모두에 대해 우수한 음질을 제공할 수 있다.

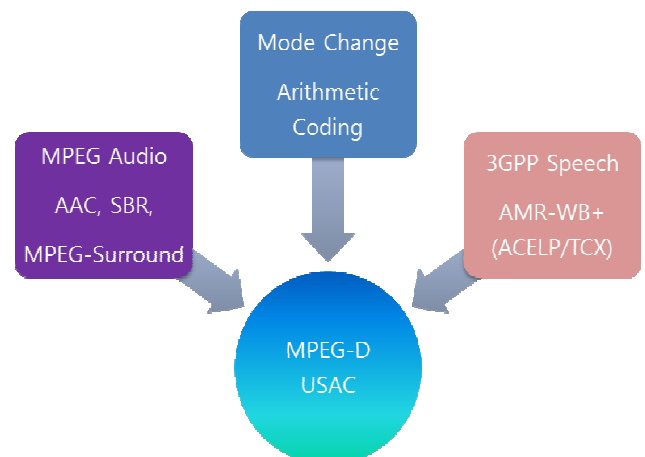


그림 1. MPEG-D USAC 기술 특성

USAC 부호화기는 그림 2 와 같은 형태를 가진다. 입력신호의 특성을 분석하여 음악 특성 신호인 경우 저대역 신호를 FD 모드로 부호화 하고, 음성 특성 신호인 경우 LPD 모드로 부호화를 수행한다. FD 모드는 기존 AAC 에서 사용한 툴을 개선한 것이고, LPD 모드는 AMR-WB+의 툴을 개선한 것이다. LPD 모드는 입력 신호를 ACELP 나 TCX 로 분석한 후 Closed-loop 방식으로 SNR 을 비교하여 부호화 모드를 선택한다. USAC 은 AAC 의 Huffman 방식의 무손실 부호화 툴을 Arithmetic 방식으로 개선하여 무손실 부호화 툴의 성능을 향상시켰으며, TCX 툴에서는 DFT 를 이용한 방식에서 MDCT 를 이용한 방식으로 변경하고, TCX 와 FD 모드 모두 동일한 Arithmetic 방식의 무손실 부호화 툴을 사용하고 있다[3].

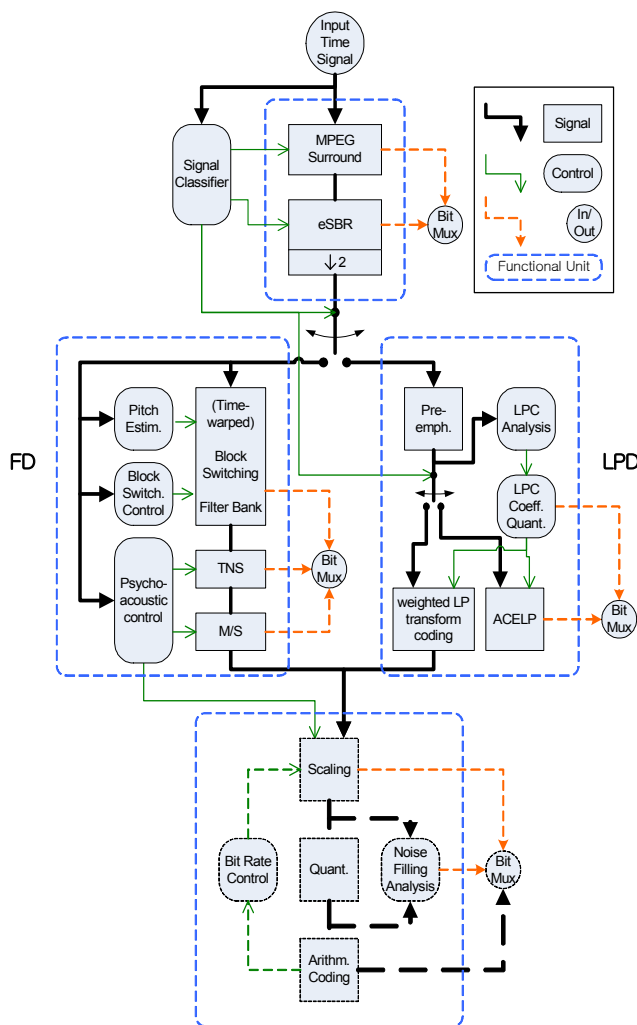


그림 2. MPEG-D USAC 부호화기 구조[3]

USAC 에서 사용하는 TCX 기술은 LPC 를 기반으로 신호를 부호화 하는 툴로써, AMR-WB+와 같이 1024 샘플의 슈퍼 프레임(super frame)을 입력 받아서, 256 샘플의 프레임 단위로 동작한다. 먼저 입력 슈퍼 프레임을 256 샘플 프레임 단위로 LPC 분석을 실시하여 LPC1, LPC2, LPC3, LPC4 의 4 개 LPC set 을 계산하고, ACELP 나 TCX 등 슈퍼 프레임의 모드에 따라 전송 할 LPC set 을 결정한다. USAC 에서는

AMR-WB+와는 다른 새로운 방식으로 LPC 를 양자화하는데, 이는 LPC set 사이에 변화가 적을 때 양자화 성능을 향상시키기 위한 것이다[6].

LPC set 을 계산 한 후, AMR-WB+처럼, 각 256 프레임 단위로 ACELP 나, TCX256(TCX20), TCX512(TCX40) 및 TCX1024(TCX80) 분석을 통해 슈퍼 프레임의 모드를 결정한다. USAC 의 TCX 는 DFT 대신 MDCT 를 사용하여 주파수 변환을 수행하고, AAC 와 동일한 산술 부호화(Arithmetic coding) 방식으로 양자화를 수행하는 점이 AMR-WB+와 차이가 있다.

USAC 의 TCX 는 MDCT 기반으로 TCX 프레임과의 연결을 위해 그림 3 과 같이 TCX20, TCX40, TCX80 모든 프레임에 대해 256 샘플의 고정된 크기의 중첩 윈도우를 이용한다. 현재 USAC 에서 이용하고 있는 TCX 윈도우는 50% 중첩 윈도우 특성이 가능한 MDCT 변환의 특성을 효율적으로 이용하지 못하고 있으며, 프레임 사이에서 TCX 모드의 변화나 transient 가 존재하는 경우에 대한 고려가 없는 구조이다.

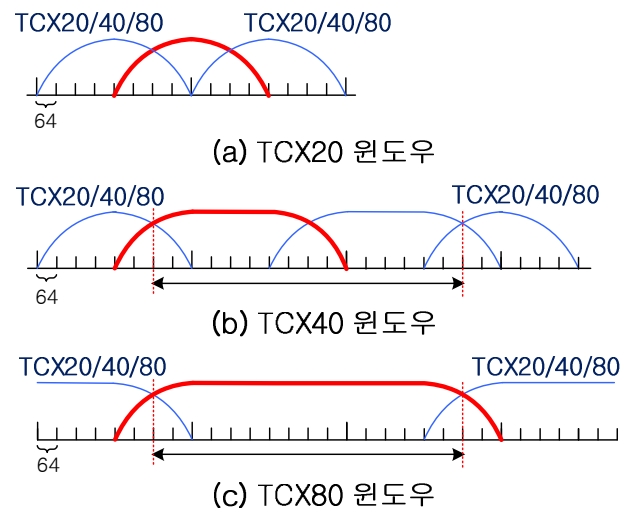


그림 3. MPEG-D USAC TCX 윈도우

### 3. MPEG-D USAC LPD 모드 성능향상 방안

본 논문에서는 USAC 의 LPD 모드의 성능을 향상시키기 위한 방법으로 TCX 프레임 연결 시, 이전 프레임과 현 프레임의 특성에 따라 중첩 윈도우 크기를 변경하는 방법을 제안한다.

먼저 TCX20 윈도우의 경우, 그림 3.(a)와 같이 현재 USAC 윈도우를 그대로 이용한다. TCX40 의 경우 그림 4 와 같이 이전 프레임과 현 프레임, 다음 프레임의 모드에 따라 다양한 형태로 변경한다. 이전 프레임이나 다음 프레임이 TCX20 인 경우 그림 4.(a)와 같이 현 USAC TCX 와 동일한 형태의 윈도우를 이용한다. 하지만 이전 프레임이나 다음 프레임이 TCX40/80 인 경우 그림 4.(b)와 같이 512 중첩 윈도우를 이용한다. 마지막으로 프레임 사이에 에너지의 변화가 급격하게 일어나는 transient 가 존재하는 경우, 그림 4.(c)와 같이 128 중첩 윈도우를 이용하여 TCX 프레임을 연결한다.

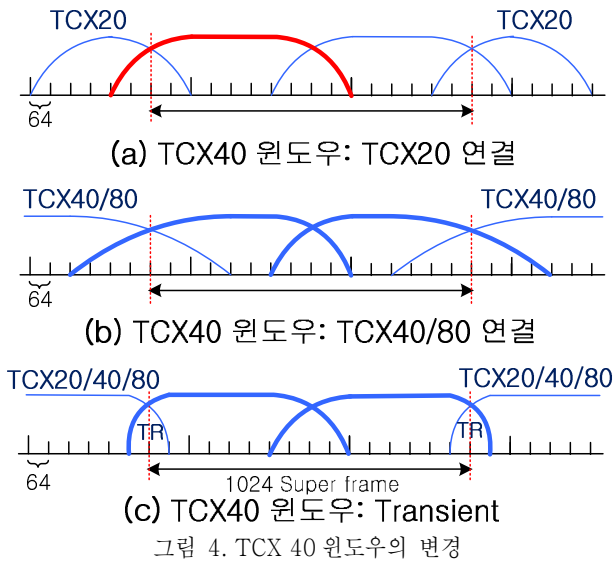


그림 4. TCX 40 윈도우의 변경

그림 5는 본 논문에서 제안하는 TCX80 윈도우의 다양한 형태를 보여준다. 먼저 이전 프레임이나 다음 프레임이 TCX20 인 경우 그림 5.(a)와 같이 현재 USAC TCX80 윈도우와 동일한 형태의 윈도우를 이용한다. 하지만 TCX40 프레임과 연결을 위해서는 그림 5.(b)와 같은 512 중첩 TCX 윈도우를 이용하고, TCX80 과 연결을 위해서는 그림 5.(c)와 같은 1024 중첩 TCX 윈도우를 이용한다. 마지막으로 프레임 사이에 transient 가 존재하는 경우, 그림 5.(d)와 같이 128 중첩 윈도우를 이용한다.

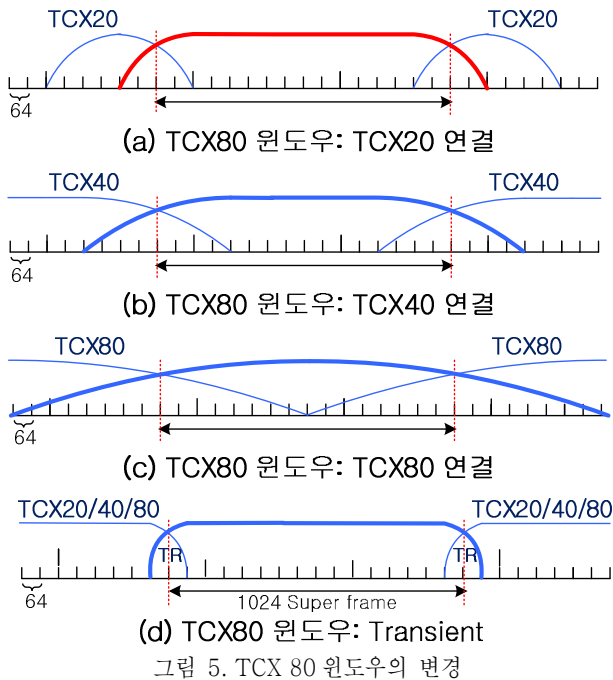
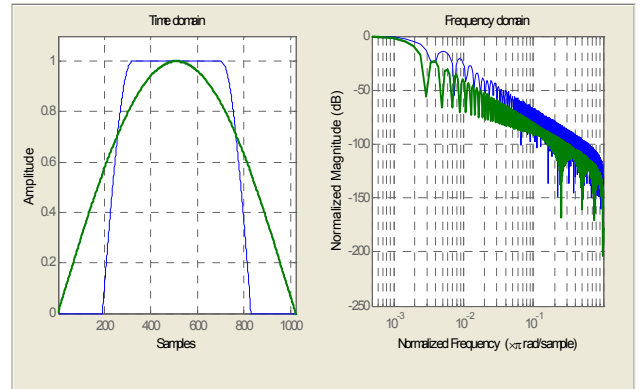


그림 5. TCX 80 윈도우의 변경

본 논문에서 제안한 TCX 윈도우의 특성을 객관적으로 검증하기 위해 윈도우의 특성을 비교하였다. 일반적으로 분석 윈도우는 임펄스 형태를 가져야 하며, 이는 main lobe 가 좁고, side lobe 의 감쇠가 빨라야 함을 의미한다. 이러한 형태의

윈도우가 분석윈도우의 leakage 를 최소화 할 수 있기 때문에 더 효과적인 분석이 가능하다. 본 논문에서 제안하는 다양한 형태의 TCX 윈도우는 현 USAC TCX 윈도우와 윈도우 특성을 객관적으로 비교하였을 때, 더 우수한 주파수 특성을 보인다. 그림 6은 현 USAC 과 제안한 방식의 TCX80 윈도우에 대해 비교한 것으로, 그림에서 주파수 특성을 보면, 제안한 TCX80 윈도우가 더 좁은 main lobe 특성과 더 빠른 side lobe 감쇠 특성을 보임을 알 수 있다.



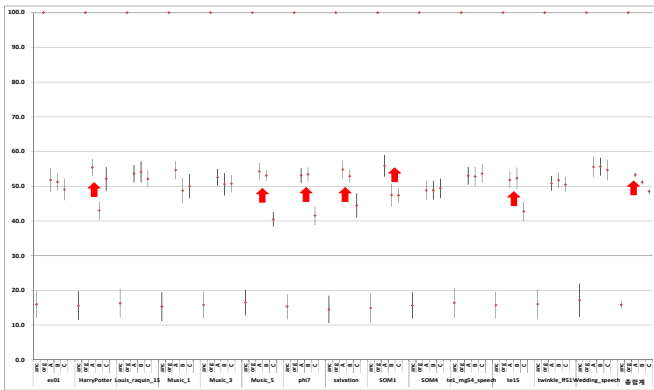
|          | Leakage Factor | Side lobe attenuation | Main lobe width |
|----------|----------------|-----------------------|-----------------|
| Proposed | 0.48%          | -23.1 dB              | 0.00219         |
| USAC     | 4.7%           | -14.4 dB              | 0.00317         |

그림 6. TCX 80 윈도우의 주파수 특성

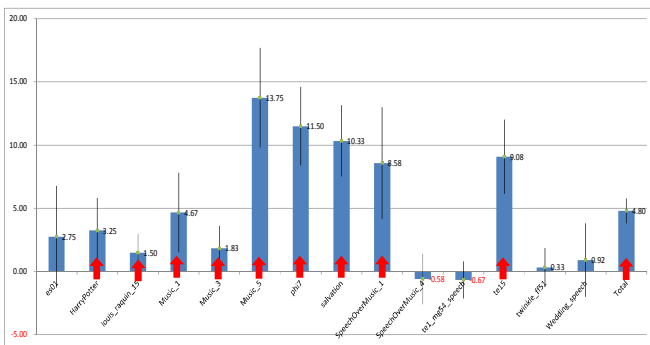
#### 4. LPD 단일모드 MPEG-D USAC 청취평가

본 논문에서 제안한 TCX 윈도우 알고리즘의 성능을 평가하기 위해 12 명이 MUSHRA 청취평가를 수행하였다. 청취평가에는 USAC 표준화에 사용하고 있는 음성 신호 4 개, 음악 신호 6 개, 음성과 음악 혼합 신호 4 개 등 총 14 개의 테스트 아이템[7]을 이용하였고, 비트율은 12kbps 모노, 16kbps 모노를 선택하였다.

그림 7 과 8 은 각각 12kbps 모노, 16kbps 모노 신호에 대해 absolute score 와 제안한 시스템과 현 USAC 의 difference score 를 보여주는 것으로 화살표로 표시한 아이템은 제안한 시스템이 더 우수한 성능을 보이는 것을 의미한다. 12kbps 모노 신호에 대한 청취평가 결과, 제안한 시스템은 음악 신호 4 개, 혼합 신호 2 개에 대해 absolute score 에서 더 우수한 결과를 보였고, difference score 에서는 음성 신호 1 개, 음악 신호 6 개, 혼합 신호 2 개에 대해 우수한 청취평가 결과를 보였다. 16kbps 모노 신호에 대한 청취평가 결과, 제안한 시스템은 음악 신호 4 개에 대해 absolute score 에서 더 우수한 결과를 보였고, difference score 에서는 음악 신호 4 개에 대해 우수한 청취평가 결과를 보였다. 따라서, 본 논문에서 제안한 TCX 윈도우 기술을 이용하면 USAC 의 음질을 효과적으로 개선시킬 수 있음을 알 수 있다.

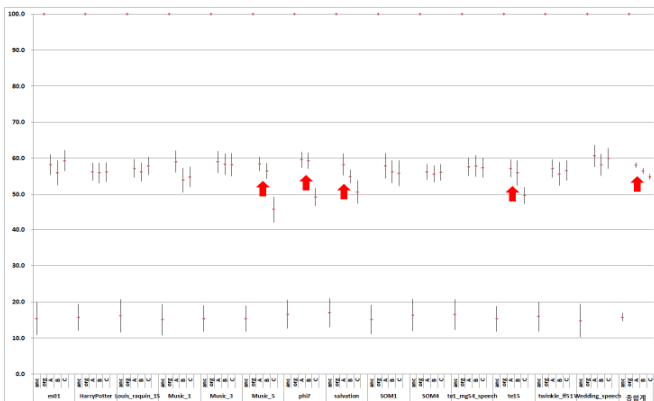


(a) Absolute Score

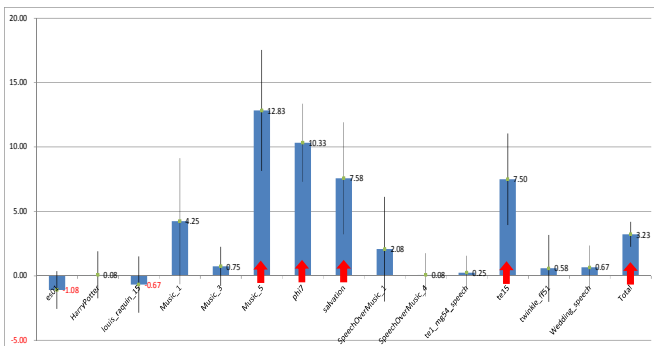


(b) Difference Score

그림 7. 청취평가 결과: 12kbps 모노



(a) Absolute Score



(b) Difference Score

그림 8. 청취평가 결과: 16kbps 모노

## 5. 결론

본 논문에서는 MPEG-D USAC 의 LPD 모드 성능을 개선하기 위한 TCX 윈도우 기술을 제안하였다. 본 기술은 USAC 의 TCX<->TCX 사이의 천이를 효과적으로 하기 위한 윈도우로서, TCX 모드에서 허용 가능한 최대한의 중첩을 이용하여 MDCT 의 특성을 최대한 이용하는 것이 핵심이다. 또한 프레임과 프레임 사이에 transient 신호가 있는 경우, 중첩 윈도우 크기를 줄여 음악 특성 신호에 대해 효과적인 부호화가 가능하다.

제안한 알고리즘의 성능을 평가하기 위해 USAC 표준화에서 사용하는 청취평가 아이템을 이용하여 MUSHRA 방식의 주관적 음질평가를 수행하였다. 청취평가 결과 본 논문에서 제안한 TCX 윈도우 기술이 USAC 의 음질을 효과적으로 향상 시킬 수 있음을 알 수 있었으며, 특히 음악 특성 신호에 대해 효과적으로 음질 향상이 가능하다.

## 감사의 글

본 연구는 방송통신위원회, 한국산업기술평가관리원의 “무안경 다시점 3D 지원 UHD TV 방송 기술 개발” 과제로 수행한 연구로부터 도출된 것입니다.

본 연구수행에 참여하여 도움을 주신 동료에게 감사의 마음을 전합니다.

## 참고문헌

- [1] ISO/IEC SC29 WG11 N9519, Call for Proposals on Unified Speech and Audio Coding, 82<sup>nd</sup> MPEG Meeting, October, 2007.
- [2] 이태진, 강경옥, 김환우, “MPEG-D USAC: 통합 음성 오디오 부호화 기술”, 한국음향학회 논문지, 제 28 권, 제 7 호, pp. 589-598, 2009.
- [3] ISO/IEC SC29 WG11 N12013, Study on ISO/IEC 23003-3:201x/DIS of Unified Speech and Audio Coding, 96<sup>th</sup> MPEG Meeting, March, 2011.
- [4] ISO/IEC 14496-3:2005, Information technology - Coding of audio-visual objects - Part 3: Audio, 2005
- [5] 3GPP TS 26.290 V6.3.0, Extended Adaptive Multi-Rate - Wideband (AMR-WB+) codec, 2007.
- [6] 이태진, 백승권, 강경옥, 이재성, 김환우, “MPEG-D USAC 성능향상을 위한 적응 MDCT TCX 윈도우 기술”, 제 22 회 영상처리 및 이해에 관한 워크샵, 2010.
- [7] ISO/IEC SC29 WG11 N9638, Evaluation Guidelines for Unified Speech and Audio Proposals, 83<sup>rd</sup> MPEG Meeting, Antalya, Turkey, January, 2008.