

논문 2011-48SP-5-17

# MDCT 도메인에서 오디오 신호 특징을 고려한 초광대역 코덱 개선

( Enhancement of Super-wideband Coder by Considering Audio Feature  
in MDCT Domain )

홍 기 봉\*, 정 규 혁\*\*, 이 인 성\*\*\*

( Ki-Bong Hong, Gyu-Hyeok Jeong, and In-Sung Lee )

## 요 약

본 논문에서는 입력신호에 따라 효율적인 부호화를 위해 다중모드를 가지는 초광대역 부호화기를 제안하였다. 최근에 개발된 G.718기반의 초광대역 확장 코덱은 초광대역에 존재하는 오디오 신호를 부호화하기 위하여 Generic모드와 Sinusoidal모드로 신호를 분류하여 부호화 하지만, 현악기, 관현악기와 같은 오디오 신호에 존재하는 멀티플 피치 및 하모닉 성분과 타악기와 같은 오디오 신호에 존재하는 Individual-Line 성분을 효율적으로 부호화하지 못하였다. 제안하는 방법은 오디오 신호에 존재하는 하모닉과 Individual-Line성분의 특징을 고려하여 모드를 세분화하였다. 성능평가를 위하여 객관적인 평가로 MDCT 도메인에서 SNR을 이용하였으며 MUSHRA 테스트를 통해 주관적인 평가를 하였다. 실험 결과 제안한 방법은 기존의 G.718 초광대역 부호화기보다 Individual-Line신호 경우 약 3dB, 하모닉 신호의 경우 약 0.8dB 높은 SNR을 보였으며 MUSHRA 테스트 역시 평균 5점 음질 향상을 확인하였다.

## Abstract

This paper presents the coding method that have multi-mode and efficiency of audio codecs using the feature of audio signal. Recently, the developed extension super-wideband codec based on G.718 wideband divides two mode between Generic and Sinusoidal. So codec efficiently encode audio signal exist in super-wideband. But the codec is not as efficient coding for harmonic component of wind instrument and string instrument and individual-Line component of percussion instrument. The proposed method are modeling and encoding multiple pitch and individual-line feature using multi mode coding. For the performance evaluation, we used SNR in MDCT domain for objective test and MUSHRA test for subjective test. As a result, the performance of SNR and MUSHRA test of the proposed method have better performance than the G.718 super-wideband codec.

**Keywords :** Super-Wideband codec, MDCT domain, Multi-mode coder, Individual-Line, multi pitch

## I. 서 론

최근 음성통신의 전송대역폭은 LTE나 BcN과 같은 차세대 통신망에서 소비자 요구를 충족시키기 위해 오디오 신호를 포함하는 20-14000Hz의 초광대역으로의 확장이 이루어지고 있는 추세이다<sup>[1]</sup>. 현재 광대역 음성 코덱을 기반으로한 대역확장 코덱은 G.711과 G.722 기

\* 학생회원, \*\*\* 정회원-교신저자, 충북대학교 전자통신공학과

(Department of Radio Engineering, ChungBuk University)

\*\* 학생회원, LG전자 (LG Electronics)

※ 이 논문은 엘지전자 연구과제지원으로 수행되었습니다.

접수일자: 2011년3월23일, 수정완료일: 2011년7월12일

반의 융합코덱<sup>[2~3]</sup>과 G.729.1과 G.718기반의 융합코덱<sup>[4]</sup>이 있으며, 이러한 융합 코덱을 통해 유무선 통합(FMC) 환경, 디바이스 통합 환경, 이중 통신망간의 서비스를 준비하고 있다<sup>[5]</sup>. 하지만 현재 음성코덱을 기반으로 한 대역확장 코덱은 대역폭 관점에서만 품질의 개선이 이루어져 왔다. 따라서 부호화 하는 대역폭이 증가하면 서비스의 품질이 높아지는 대역폭 확장 관점에서의 품질 개선이 이루어졌기 때문에, 오디오 신호의 특징을 효율적으로 부호화 하지 못하였다<sup>[6]</sup>. 최근에서야 오디오 신호가 존재하는 7000-14000Hz의 초광대역을 부호화하기 위하여 대역폭 분할 및 확장 계층기반과 오디오 코덱에 많이 사용되는 알고리즘인 MDCT기반의 대역확장을 하였다<sup>[7]</sup>. 최근에 개발된 G.718기반의 초광대역 확장 코덱<sup>[8]</sup>은 초광대역에 존재하는 오디오 신호를 효율적으로 부호화하기 위하여 Generic모드와 Sinusoidal모드로 신호를 분류하여 부호화하지만, 일반적으로 현악기, 관현악기에 존재하는 하모닉성분과 타악기에 존재하는 Individual-Line성분을 효율적으로 부호화하지 못하였다<sup>[9]</sup>.

따라서 본 논문에서는 최근에 발표된 G.718기반의 초광대역 확장 코덱의 MDCT 도메인에서의 오디오 신호 특징<sup>[10]</sup>을 고려한 다중모드 임베디드 초광대역 부호화기를 설계함으로써 오디오 신호의 품질을 개선한다. 본 논문에서는 II장에서 G.718 초광대역 코덱의 방법 및 문제점을 간략하게 설명하고, III장에서 G.718 기반의 초광대역 확장 코덱의 부호화를 개선하기 위한 과정을 설명한다. IV장에서는 제안한 방법의 음질평가 결과를 보며 V장에서 본 논문에 대한 결론을 맺는다.

## II. G.718 초광대역 코덱의 부호화 방법

G.718 초광대역 확장 코덱은 기존 G.718 광대역 코덱<sup>[11]</sup>과의 호환성을 제공하기 위하여 임베디드 가변 비트율 코덱인 G.718 광대역 코덱을 기반으로 기존의 광대역을 제외한 7000-14000Hz의 초광대역 대역을 추가하는 임베디드 구조로 대역을 확장하였다. 기본적인 대역폭 확장 알고리즘은 초광대역 신호를 부호화하기 위하여 2가지 코딩모드를 사용하여 MDCT 도메인에서 대역폭을 확장한다. 대역 확장의 첫 번째 계층은 Generic 모드와 Sinusoidal 모드로 분할되어 부호화된다. 두 모드는 입력되는 초광대역 신호의 Tonality 측정을 통하여 결정된다.

### 1. Generic 모드와 Sinusoidal 모드 부호화

Generic 모드는 현재 프레임이 Non-tonal로 판단되었을 경우 부호화 하는 모드이며 G.718 광대역 코덱의 부호화된 MDCT 신호를 이용하여 부호화한다. 초광대역 신호는 밴드분할되며, 각 부밴드는 부호화된 광대역 신호와의 유사도를 측정하게 된다. 가장 유사하다고 판단되는 광대역 신호의 구간에 선형 도메인과 로그 도메인에서 각 2개의 Scale값을 연산함으로써 초광대역 신호를 부호화 한다.

Sinusoidal 모드는 현재 프레임이 Tonal로 판단되었을 경우 부호화 하는 모드이다. 초광대역의 신호를 부호화하기 위하여 MDCT 계수 중 10개의 펄스를 추출하는 방법으로 부호화한다. 최초의 4개의 펄스는 7000-8600Hz에 해당하는 대역에서 다음 4개의 펄스는 8600-10200Hz, 2개는 각 1개씩 10200-11800Hz와 11800-12600Hz에서 추출한다.

### 2. G.718 초광대역 부호화기의 문제점

Individual-Line 신호는 타악기 및 합주음악에서 많이 볼 수 있는 특성으로써 일반적인 Generic 모드로 부호화시 그림 1과 같이 원신호와 비교시 Individual-Line 이외에 주변 잡음 함께 부호화 된다. Generic 모드는 광대역신호를 참조하여 스케일 값을 곱하여 결국은 에너지가 집중된 계수들 주변에 잡음으로 인식되는 계수들도 함께 참조하기 때문이다.

Sinusoidal 모드 역시 밴드분할하여 초광대역 신호를 미리 결정된 개수의 신호로 표현하기 때문에 하모닉 성분을 효율적으로 부호화 하지 못하는 단점을 가지고 있다. 즉, Sinusoidal 모드는 한정된 개수의 계수만을 추출함으로써 하모닉 구조를 가지는 신호에 효율적이지 못하다.

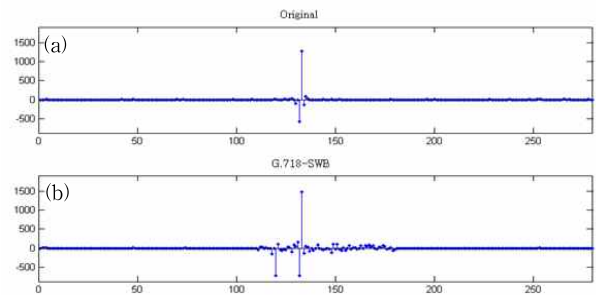


그림 1. Generic 모드의 문제점 : (a)원음, (b)Generic 모드로 부호화된 신호  
 Fig. 1. Problem of the Generic mode : (a) original signal, (b)encoded signal in Generic mode.

### III. 신호 특징에 따른 다중모드 부호화기 설계

본 연구에서 제안하는 방법은 초광대역의 신호의 특성에 따라 적합한 모델을 사용하는 방법을 전제로 한다. 제안하는 방법은 Nokia사의 G.718 초광대역 후보 코덱을 세분화하여 효율적인 부호화에 초점을 맞춘다. 기존 Generic과 Sinusoidal 모드 이외 오디오 신호의 특징을 고려한 Individual-Line 모드와 하모닉 모드를 추가한다. 세분화된 모드를 포함한 초광대역 부호화기는 그림 2와 같다. 편의상 추가된 모드는 점선으로 표현하였다.

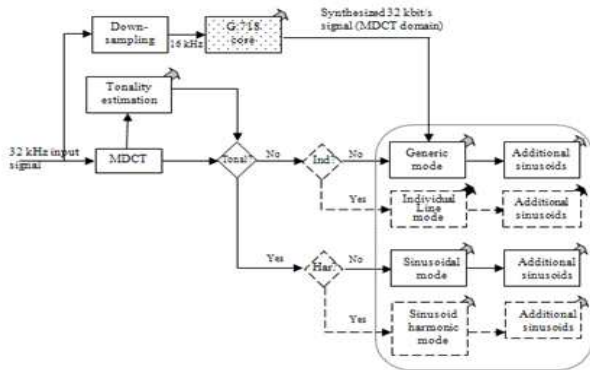


그림 2. 제안하는 초광대역 부호화기의 블록도  
Fig. 2. Structure of proposed super-wideband codec.

#### 1. Individual-Line 모드

제안하는 Individual-Line모드에서는 피치정보보다는 에너지가 큰 부밴드에 따른 반복적 검색을 통해 피크값을 부호화하고 잔여신호에 대해서는 추출한 피크값 이외 신호들을 부호화된 광대역 신호를 참조하여 부호화된다. Individual-Line모드는 다음과 같은 과정으로 진행된다.

##### 가. Generic과 Individual-Line 모드 선택부

제안하는 Individual-Line모드는 tonality에 따라 Generic모드로 구별된 MDCT신호 일부 계수들의 에너지비율에 따라 나뉘어 진다. 다음의 식을 통해 구해진 원신호  $M_{32}(k)$ 에서 에너지가 가진 큰 펄스들의 에너지 합과 전체 에너지와의 비율을 통해 모드를 구별한다.

$$R_{ps} = \frac{\sum_{k=0}^7 \{S_{ps}(k)^2\}}{\sum_{k=0}^{279} \{M_{32}(k)^2\}} \quad (1)$$

이때  $S_{ps}(k)$ 는 현재 프레임의 최대 에너지를 가지는 8개 pulse들이며  $R_{ps}$ 는 원신호의 에너지 및 추출된 peak의 에너지 비율이다.  $R_{ps}$ 이 0.6이상일 경우 Individual-Line 모드로 판단하게 되며 0.6은 실험적으로 구해진 상수이다. 이와 같은 판단기준은 MDCT 도메인상에 Individual-Line신호의 특성이 일부 계수에 대부분의 에너지가 집중되는 현상을 이용한 것이다.

##### 나. MDCT 도메인에서의 부밴드별 최대값 추출 알고리즘

Individual-Line모드로 결정된 프레임의 일부 계수를 제외한 신호의 에너지는 추출하는 계수의 에너지와 비교시 상당히 작은 값을 가진다. 따라서 최대값을 가지는 일부 계수에 대한 부호화가 잘 이루어진다면 Individual-Line 모드의 부호화는 기존의 Generic 모드보다 뛰어난 성능을 보일 수 있다. 따라서 Individual Line 모드는 우선적으로 에너지가 집중된 일부 계수 즉, 최대값을 반복적으로 추출하는 방법을 중심으로 이루어진다. 다음의 식과 같이 초광대역(7000-13400Hz) MDCT신호와 추출한 pulse 크기의 차이값을 정의 한 후 4개의 부밴드  $D_{i,j}(k)$   $j=0, \dots, 3$ 로 분할한다. 이때 최초 정의되는 차이값  $D_0(k)$ 은 추출한 펄스가 존재하지 않으므로 원신호  $M_{32}(k)$ 가 된다.

$$\begin{aligned} D_{i,0}(k) &= D_i(k), & k &= 0, \dots, 63 \\ D_{i,1}(k) &= D_i(k+64), & k &= 0, \dots, 63 \\ D_{i,2}(k) &= D_i(k+128), & k &= 0, \dots, 63 \\ D_{i,3}(k) &= D_i(k+192), & k &= 0, \dots, 63 \end{aligned} \quad (2)$$

$i$ 는 반복횟수를 나타내며 총 5회 반복되며 부밴드의 길이는 각 64개의 MDCT 계수를 가진다. 부밴드별 에너지를 연산 후 에너지가 가장 큰 부밴드  $B^i$ 를 선택하게 된다.

$$B^i = \underset{0 \leq j \leq 3}{\operatorname{argmax}} \left( \sum_{k=0}^{63} \{D_{i,j}(k)^2\} \right) \quad (3)$$

선택된  $B^i$ 에 존재하는 MDCT 계수 중 에너지가 가장 큰 값을 부호화 target신호로 정의 한 후 선택된 최대값 이외에 좌, 우에 위치한 계수의 에너지를 비교하여 펄스를 추가적으로 추출한다. 이는 Individual-Line신호의 특성상 최초 추출된 최대값 주변에 추출된 펄스와 유사한 크기를 가지는 계수가 존재하므로 위치정보의 bit-allocation을 효율적으로 하기위한 방식으로 다음 식과 같이 추출한다.

$$P_{amp}^i(n_1) = \max(|P_{amp}^i(n_0 - 1)|, |P_{amp}^i(n_0 + 1)|) \quad (4)$$

표 1. Individual-Line 모드의 비트할당

Table 1. Bit-allocation for Individual-Line mode.

Parameter	Number of bits
SWB / stereo	1
Generic / Sinusoidal mode	1
Selected band	10
Pulse position	30
Pulse amplitude	16
Pulse sign	8
Wideband position	9
Noise average energy	4

$n_0$ 는 최초에 선택된 펄스의 위치정보이며  $P_{amp}^i(n)$ 는 부밴드  $B^i$ 의  $n$ 번째 MDCT 계수의 크기이다.

위의 과정 후 추출된 2개의 최대값을 제외한 4개의 부밴드별 에너지를 다시 연산하게 된다. 다음의 식과 같이 추출된 최대값들을 제외 후 target신호를 재정의 함으로써 잔여신호 중 에너지가 큰 값을 다시 추출할 수 있게 된다.

$$D_{i+1}(k) = \begin{cases} |D_i(k) - P_{amp}^i(64 \times B^i + n_0) - P_{amp}^i(64 \times B^i + n_1)| & 0 \leq i < 3 \\ |D_i(k) - P_{amp}^i(64 \times B^i + n_0)| & 3 \leq i < 5 \end{cases} \quad (5)$$

총 3회 실시하여 6개의 펄스를 추출 한 후 추출되는 최대값 주변에 존재하지 않지만 큰 에너지를 가지는 단일 펄스를 추출하기 위해 추가적인 최대값 추출과정을 2회 실시한다.

위의 과정을 거친 후 추출되는 펄스는 크기( $P_{amp}^i(n)$ ), 위치( $n_0^i, n_1^i$ ), 부호( $P_{sign}^i(n)$ ) 정보로 양자화되며 펄스 정보 및 선택된 밴드( $B^i$ )대한 비트할당은 표 1과 같다.

#### 다. 잔여신호의 노이즈화

펄스 추출 후 초광대역의 잔여신호는 노이즈신호와 유사한 성질을 가진다. 잔여신호를 가우시안 노이즈와 같이 무작위로 부호화가 가능하나 계층구조인 G.718 초광대역 코덱의 특성상 부호화단과 복호화단에 참조되는 신호의 정보는 공유되어야 한다.

따라서 잔여신호는 Generic 모드와 동일하게 4개의 부밴드로 분할하고 광대역과 유사한 구간을 검색하여 유사성이 가장 큰 위치를 부호화한다. 펄스를 제외한 잔여신호를 부호화하기 위해 유사성이 가장 높은 광대역의 위치정보 이외에 초광대역 신호 중 펄스를 제외한 잔여신호의 에너지를 부호화해야 한다.

잔여신호의 에너지 분포는 부호화된 펄스들의 에너지  $\hat{P}_e$ 와 잔여신호의 에너지  $N_e$ 의 전체적인 비율  $\hat{R}$ 을

이용하여 부호화된다. 다음과 같이 펄스들의 에너지와 잔여신호의 에너지를 구하고 그 비율을 부호화한다.

$$N_e = \sum_{k=0}^{279} \{M_{32}(k)^2\} - \sum_{n=0}^8 \{\hat{P}_{amp}(n)^2\} \quad (6)$$

$$\hat{R} = \frac{\hat{P}_e}{\hat{P}_e + N_e} \times 100 \quad (7)$$

## 2. 하모닉 모드

제안하는 하모닉 모드는 오디오 신호에 자주 발생하는 하모닉 신호를 위한 모델링과 양자화 기법을 이용하여 오디오 신호를 효율적으로 부호화하며, 하모닉 성분은 피치성분을 이용하여 MDCT 도메인 상의 펄스의 위치를 표시하여 양자화에 많은 비트를 할당하여 부호화의 성능을 높이게 된다.

하모닉 모드로 구분된 MDCT신호  $M_{32}(k)$ 는 그림 3과 같이 일정한 기본 주파수의 배수가 되는 지점에 펄스가 위치하는 하모닉 성분을 포함하는 현상을 볼 수 있다. 이러한 많은 하모닉 성분은 Sinusoidal 모드로 표현하기 위해서는 펄스의 위치를 표현하기 위하여 많은 비트를 할당하여야 하며, 상대적으로 추출된 펄스를 양자화 하는데 많은 비트를 할당하지 못하게 된다. 따라서 하모닉 신호의 기본주파수를 이용하여 하모닉의 위치를 표현하게 된다면 상대적으로 펄스의 위치를 표현하기 위하여 적은 비트를 할당하고, 추출된 펄스를 양자화 하는데 많은 비트를 할당 할 수 있게 된다.

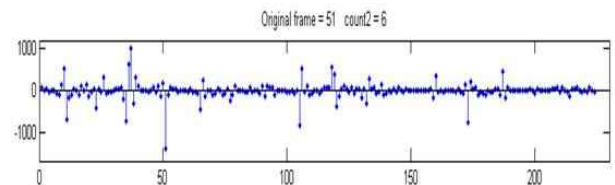


그림 3. MDCT 도메인에서의 harmonic 신호  
Fig. 3. Harmonic signal in MDCT domain.

#### 가. 하모닉 트랙 추출 알고리즘

하모닉 모드의 부호화는 한 프레임에 존재하는 멀티 피치 성분을 효율적으로 모델링하기 위하여 초광대역신호에서 총 4개의 하모닉 트랙을 추출하는 방법으로 부호화한다. 4개의 하모닉 트랙은 7000-12600Hz에서 추출된다. 4개의 하모닉 트랙은 최대 상이한 2개의 피치 성분을 포함할 수 있으며, 하나의 피치 성분에 2개의 하모닉 트랙을 추출하여 MDCT 도메인의 특성을 고려한

모델링 기법을 사용한다. 이를 통해 프레임에 존재하는 여러 악기에 의해서 발생하는 하모닉 성분의 신호를 표현하게 된다.

원신호  $M_{32}(k)$ 는 자기상관 함수를 이용하여 가장 많은 에너지를 포함하는 하모닉 트랙  $D_j$ 의 피치  $P_i$ 를 찾는다. 피치의 범위를 MDCT 계수 20-27개로 제한하여 추출되는 하모닉의 개수를 제한한다.

$$P_i(m) = \sum_{k=0}^{279} (|M_{32}(k)| \times |M_{32}(k+m)|), \quad m = 20, \dots, 27, \quad i = 1, 2 \quad (8)$$

개별 피치  $P_i$ 마다 가장 많은 에너지를 포함하는 총 2개의 하모닉 트랙  $D_j$ 의 시작위치  $PS_i$ 를 계산하여 하모닉 트랙  $D_j$ 을 추출한다. 하모닉 트랙  $D_j$ 의 시작위치  $PS_i$ 의 범위는 추출되는 하모닉의 개수를 포함하여 계산되어지며, MDCT 도메인 신호의 특성을 반영하여 같은 피치  $P_i$ 에 의해 2개의 시작위치  $PS_i$ 에 의하여 총 2개의 하모닉 트랙  $D_j$ 이 추출된다.

$$PS_i(2m-1) = \sum_{n=1}^{\lfloor 280/P_i \rfloor} |M_{32}[(2m-1) + P_i \times n]|, \quad m = 1, \dots, 16$$

$$PS_i(2m) = \sum_{n=1}^{\lfloor 280/P_i \rfloor} |M_{32}[(2m) + P_i \times n]|, \quad m = 1, \dots, 16 \quad (9)$$

추출된 4개의 하모닉 트랙 Track  $D_j$ 의 피치  $P_i$ 와 시작위치  $PS_i$ 의 범위와 개수는 표 2와 같다.

표 2. 하모닉 트랙별 정보  
Table 2. Harmonic track structure.

Pitch	Track	Start position	step size	Length
20-27	1	280	2	16
	2	281	2	16
20-27	3	280	2	16
	4	281	2	16

나. 주요 pulse 및 잔여 pulse 양자화

위의 과정에서 추출되는 4개의 하모닉 트랙은 최대 44개의 펄스를 포함한다. 44개의 펄스의 크기값과 부호정보를 모두 양자화 하기 위해서는 많은 비트를 필요로 한다. 따라서 각각의 하모닉 트랙의 펄스 중 에너지를 많이 포함하는 펄스 2개씩 추출하여 크기 값과 부호정보를 양자화한다. 추출된 8개 펄스를 제외한 하모닉 트랙은 하나의 트랙으로 결합되어 DCT를 이용하여 크기값과 부호정보

표 3. 하모닉 모드의 비트할당

Table 3. Bit-allocation for Harmonic mode.

Parameter	Number of bits
SWB / stereo	1
Generic / Sinusoidal mode	1
Sinusoidal / Harmonic mode	1
Extracted pitch	6
Extracted Harmonic	16
Harmonic position	8
pulse amplitude	24
DCT coefficient	23

를 동시에 양자화 한다. 하모닉 모드의 부호화 정보 및 비트할당은 표 3과 같다.

다. sinusoidal과 하모닉 모드 선택부

하모닉 모드는 Tonal정도에 따라 Sinusoidal 모드로 구별된 MDCT신호의 일부 계수들의 추출된 에너지에 따라 나뉘어진다. 4개의 하모닉 트랙  $D_j$  에너지의 합이 Sinusoidal 모드의 10개의 펄스들의 전체 에너지보다 많은 에너지를 포함하면 현재의 프레임이 하모닉 성분을 포함하는 프레임으로 판단하고, 하모닉 모드로 부호화 한다.

IV. 실험

1. 음질평가 실험 환경

본 장에서는 제안된 알고리즘의 성능평가는 G.718 초광대역 코덱과의 36kbps의 동일한 조건하에서 주관적인 평가 방법으로 오디오 신호의 성능평가에 활용되는 MUSHRA 테스트를 통해 성능을 평가하고, 객관적인 평가 방법으로 MDCT 도메인의 SNR을 비교하였다.

MUSHRA<sup>[12]</sup>은 표 4와 같이 은닉된 원음과 원음을 가공(low-pass-filter)한 앵커음, 비교대상, 그리고 평가대상을 상호비교 후 배점을 주는 형식이다. MUSHRA 테스트 중에는 은닉된 원음을 찾아낸다는 가정하에 테스트 신호 중 1개는 무조건 100점으로 평가하여만 한다. 이와 같은 방법으로 실험에 참여한 청취자 중 원음을 구별하지 못하는 청취자의 결과는 post-Screening을 통해 최종 결과에서 제외한다. 실험에 참여한 청취자는 훈련되지 않은 일반인 12명으로 구성되며 동일한 환경에서 평가가 이루어지기 위해 동일한 헤드폰과 동일한 PC에서 실험을 실시한다. 실험에 사용되는 음원들은 실생활에 발생할 수 있는 신호들로 이루어져있다. 초광대

표 4. MUSHRA 테스트 비교 대상  
Table 4. Compare sound for MUSHRA test.

Test NO.	명 칭	설 명
1	Hidden Reference	원음
2	LPF70	LPF(fc=7kHz) 통과한 원음
3	LPF100	LPF(fc=10kHz) 통과한 원음
4	G.718_36	G.718-SWB layer 6
5	G.718_40	G.718-SWB layer 7
6	G.718_48	G.718-SWB layer 8
7	TSP_36	제안한 알고리즘

역에는 음성보다 악기/음악의 신호가 많이 분포되어 있으므로 실험에 사용되는 신호는 대부분 mixed content / Music신호로 이루어져 있다. 사용되는 음원들은 Generic / Sinusoidal / Individual-Line / Harmonic 성분이 포함된 음원 각 4개씩 16개를 사용하였다.

SNR 계산은 다음의 수식과 같이 원음의 초광대역(7000-14000Hz) MDCT 계수  $M_{32}(k)$ 와 부호화된 신호의 초광대역(7000-14000Hz) MDCT 계수  $\widehat{M}_{32}(k)$ 의 사용하였다.

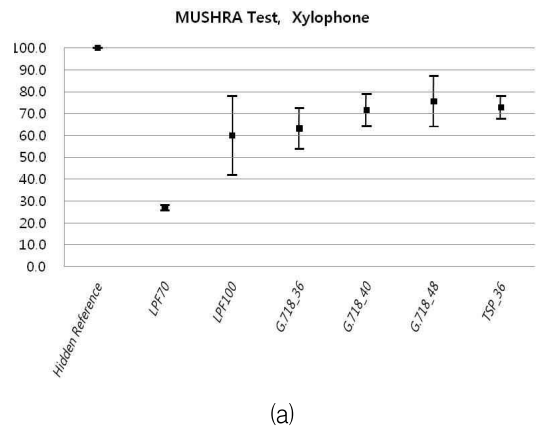
$$SNR(dB) = 10 \log \left[ \frac{\sum_{k=0}^{279} \{M_{32}(k)\}^2}{\sum_{k=0}^{279} \{M_{32}(k) - \widehat{M}_{32}(k)\}^2} \right] \quad (10)$$

2. 음질평가 결과

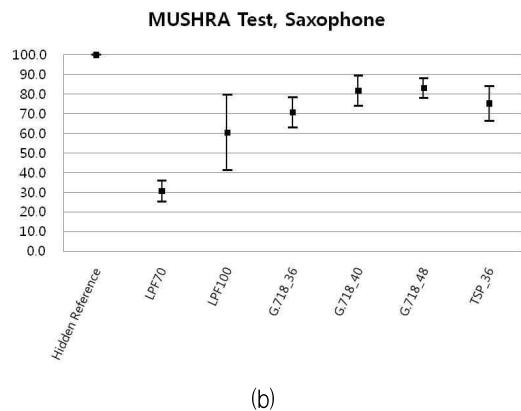
제안한 방법은 Generic 모드와 Sinusoidal 모드와 공존한다. 따라서 Generic 성분과 Sinusoidal 성분이 존재하는 신호는 기존의 모드로 부호화되기 때문에 제안한 방법의 MUSHRA 테스트 결과값은 기존의 G.718 초광대역 코덱과 유사하다.

그러나 신호가 Individual-Line 성분과 하모닉 성분을 포함하고 있을 경우 평균 6점정도 높은 결과를 보였으며 Individual-line 성분을 가지는 신호를 부호화시 같은 전송률을 가지는 G.718 광대역 코덱 6계층의 Generic 모드만 존재하는 경우보다 평균 8점 높은 점수를 얻었다. 또한 그림 4의 (a)와 같이 명확한 Individual-Line성분이 존재하는 실로폰 신호의 경우 Generic 모드로 부호화한 경우보다 월등히 높으며 4kbps가 더 추가된 7계층과 대등한 평가를 얻었다.

하모닉 성분이 강한 신호 역시 같은 전송률을 가지는 6계층의 Sinusoidal 모드보다 평균 5정도 높은 점수를



(a)



(b)

그림 4. MUSHRA test 결과 : (a)실로폰, (b)색소폰  
Fig. 4. Result of MUSHRA test : (a)xylophone, (b)Saxophone

얻었으며 그림 4의 (b)와 같이 하모닉 성분을 많이 포함하는 색소폰 신호의 경우 같은 전송률인 6계층보다 월등히 높은 평가를 얻었다.

이는 그림 5과 그림 6과 같이 스펙트럼으로도 확인 가능하다. 그림 5과 같이 Individual-line모드 경우 Generic 모드로 부호화시 발생하는 주변 잡음이 발생되지 않으며 그림 6과 같이 하모닉 모드 경우 sinusoidal 모드보다 많은 펄스를 추출함으로써 음질을 향상시켰다.

Individual-Line모드의 신호 대 잡음비는 표 5와 같이 기존 Generic 모드보다 높은 값을 가지는 것을 볼 수 있다. 일부 Individual-Line성분이 강한 실험 샘플에 대해서는 약 3.5dB 높은 값을 가진다. 이는 Generic 모드에 비해 에너지가 집중된 계수들을 충실하게 부호화한 결과라 볼 수 있다. 또한 기존의 Generic 모드의 경우 모든 신호를 광대역신호를 참조하는 형식으로 되어 있다. 스케일 값을 곱하지만 결국은 에너지가 집중된 계수들 주변에 잡음으로 인식되는 계수들도 함께 참조된



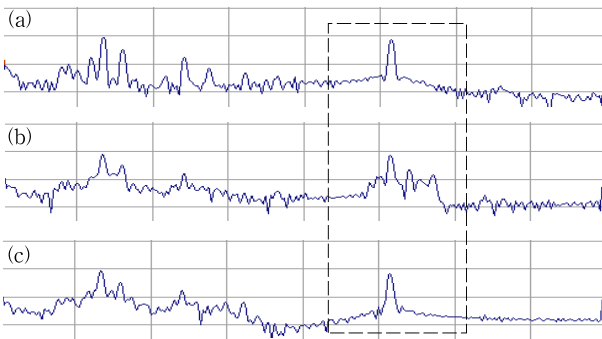


그림 5. Individual-Line신호의 스펙트럼 비교 : (a)원음, (b)Generic 모드로 부호화된 신호, (c)Individual-Line 모드로 부호화된 신호

Fig. 5. Compare spectrum in Individual-Line signal : (a)original signal, (b)encoded signal in generic mode, (c)encoded signal in Individual-Line mode.

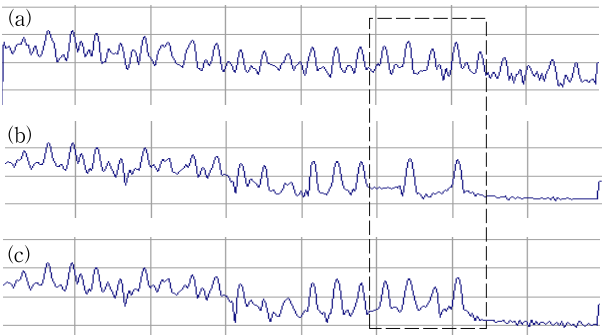


그림 6. Harmonic 신호의 스펙트럼 비교 : (a)원음, (b)Sinusoidal 모드로 부호화된 신호, (c)하모닉 모드로 부호화된 신호

Fig. 6. Compare spectrum in Harmonic signal : (a)original signal, (b)encoded signal in sinusoidal mode, (c)encoded signal in harmonic mode.

표 5. Individual-Line 모드의 SNR 결과  
Table 5. Result of SNR in Individual-Line mode.

Test No.	Generic mode	Individual-Line mode	차이
1	2.8260	6.3182	+3.492
2	2.2387	6.0047	+3.766
3	1.8279	4.6495	+2.822

표 6. 하모닉 모드의 SNR 결과  
Table 6. Result of SNR in Harmonic mode.

Test No.	Sinusoidal mode	Harmonic mode	차이
1	2.4100	3.4100	+1.000
2	4.9238	5.7413	+0.818
3	3.6632	4.4137	+0.751

다. 그러나 Generic 모드와 비교시 제안하는 방법은 에너지가 집중된 계수들만을 별도의 부호화가 이루어지며 잔여신호의 에너지 분포를 원신호의 잔여신호와 유사하게 복원함으로써 원신호와 유사하게 부호화한다.

하모닉 모드 역시 표 6과 같이 기존의 Sinusoidal 모드보다 높은 SNR을 가진다. 이와 같이 SNR이 높은 구간은 기존의 Sinusoidal 모드에 비해 하모닉 성분이 많은 MDCT 계수들을 양자화에 많은 비트를 할당하여 충실하게 부호화한 결과라 볼 수 있다. 오디오 신호에 존재하는 하모닉 성분을 부호화함에 있어서, 기본 주파수를 이용하여 MDCT 도메인상에서 펄스의 위치를 표현하여 양자화에 많은 비트를 할당할 수 있었다.

### V. 결 론

본 연구에서는 오디오 신호의 특징을 고려하여 다중 모드 초광대역 코덱 설계를 제안하였다. 최근에 개발된 G.718기반의 초광대역 확장 코덱은 초광대역에 존재하는 오디오 신호를 부호화하기 위하여 Generic 모드와 Sinusoidal 모드로 신호를 분류하여 부호화하지만, 일반적으로 현악기, 관현악기에 존재하는 하모닉 성분과 타악기에 존재하는 Individual-Line 성분을 효율적으로 부호화하지 못하였다. 따라서 본 연구에서는 광대역 코덱인 G.718기반의 초광대역 확장 코덱의 음질을 개선하고자 오디오신호의 특징을 이용하여 효율적인 부호화기를 제안하였다. Individual-Line 모드의 경우 중요 일부 펄스를 반복적으로 추출하는 구조와 하모닉 모드의 경우 피치를 이용하여 한정된 전송률에서 기존 sinusoidal 모드보다 많은 신호를 부호화하며 오디오 신호에서 많이 발생하는 멀티피치에 효율적인 부호화를 한다.

음질 테스트를 통해 주관적인 평가인 경우 기존의 G.718 초광대역 코덱보다 Individual-Line 신호의 경우 평균 8점 높은 점수를 획득하였으며 하모닉 신호 역시 평균 5점 정도의 높은 점수를 얻었다. 또한 객관적인 평가인 SNR 경우 Individual-Line 모드 경우 기존의 Generic 모드보다 3dB, 하모닉 모드의 경우 기존의 Sinusoidal 모드보다 약 0.8dB 높아짐을 볼 수 있었다. 따라서 차후 LTE와 BcN와 같은 차세대 네트워크에서 고품질의 음성 오디오 서비스를 제공할 수 있을 것으로 기대된다.

## 참고 문헌

- [1] P. Noll, "Wideband speech and audio coding," *IEEE Comm. Maga.*, vol. 26, pp. 34-44, Nov. 1993.
- [2] ETRI, "High-level Description of ETRI Candidate for G.722/G.711 SWB Extension", ITU-T WP3/SG16 AC-0907-Q10-08, Geneva, July. 2009.
- [3] 성종모, 김현우, 김도영, 이병선, 고윤호, "ITU-T G.711.1 및 G.722 슈퍼와이드밴드 확장 후보 코덱 알고리즘," *전자공학논문지*, 제47권 SP편 제5호, 63-73쪽, 2010.
- [4] 김현우 외, "G.729.1 광대역 멀티코덱 표준 기술동향," *전자통신동향분석*, 제21권 제6호, 77-85쪽, 2006.
- [5] ETRI, "Trends of Speech-Based Audio Convergence Codec Technology", *전자통신동향분석*, 제24권, 제5호, 10-19쪽, 2009
- [6] M. Dietz, L. Liljeryd, K. Kjolring and O. Kunz, "Spectral Band Replication, a novel approach in audio coding," *Proc. 112nd AES convention*, May 2002.
- [7] S. A. Ramprashad, "Embedded coding using a mixed speech and audio coding paradigm," *International Jour. of speech tech.*, vol. 2, pp. 359-372, 1999.
- [8] Bernd Geiser et al., "Candidate Proposal for ITU-T Super-wideband Speech and Audio coding", *ICASSP 2009*, pp.4121-4124, Apr. 2009.
- [9] Purnhagen, H.(2000) "HILN - the MPEG-4 parametric audio coding tools", *IEEE international Symposium*, vol. 3, pp. 201-204.
- [10] Duadet, L., M. Sandler,(2004), "MDCT Analysis of sinusoids : exact results and applications to coding artifacts reduction", *IEEE Trans*, vol. 12, pp. 302-312.
- [11] ITU-T G.718, "Frame Error Robust Narrowband and Wideband Embedded Variable Bitrate Coding of Speech and Audio from 8-32kbit/s," 2008. 6.
- [12] Recommendation BS.1534-1: Method for the subjective assessment of intermediate quality levels of coding systems, ITU (International Telecommunication Union), 2003.

## — 저 자 소 개 —



홍 기 봉(학생회원)  
2010년 충북대학교 정보통신  
공학과 학사 졸업  
2010년~현재 충북대학교  
전파통신공학과 석사과정

<주관심분야 : 음성/음악 신호처리, 영상 신호처리>



정 규 혁(학생회원)  
2004년 충북대학교 전기전자  
공학 학사 졸업  
2006년 충북대학교 전파통신  
공학과 석사 졸업  
2011년 충북대학교 전파통신  
공학과 박사 졸업

2010년~현재 LG 전자기술원 미래IT융합연구소  
선임연구원

<주관심분야 : 음성/오디오 신호처리>



이 인 성(정회원)-교신저자  
1983년 연세대학교 전자공학  
학사 졸업  
1985년 연세대학교 전자공학  
석사 졸업  
1992년 Texas A&M University  
전기공학과 박사 졸업

1993년~1995년 한국전자 통신연구원 이동통신  
기술연구단 선임연구원

1995년~현재 충북대학교 전기전자공학부 정교수  
<주관심분야 : 음성/오디오 신호처리, 이동통신,  
적응필터>