

논문 2010-47SP-5-8

ITU-T G.711.1 및 G.722 슈퍼와이드밴드 확장 후보 코덱 알고리즘

(A Candidate Codec Algorithm on Superwideband Extension to ITU-T
G.711.1 and G.722)

성종모*, 김현우*, 김도영*, 이병선*, 고윤호**

(Jongmo Sung, Hyun Woo Kim, Do Young Kim, Byung Sun Lee, and Yun-Ho Ko)

요약

본 논문은 ITU-T SG16 Q.10에서 표준화 진행 중인 G.711.1 및 G.722 슈퍼와이드밴드 확장 코덱의 후보 코덱 알고리즘에 대한 것으로, 제안된 후보 코덱은 기존 ITU-T 광대역 코덱 G.711.1 및 G.722와 비트스트림 호환성을 지원함과 동시에 슈퍼와이드밴드 확장 계층을 통해 50-14,000 Hz에 해당하는 슈퍼와이드밴드 신호를 부호화한다. 본 후보 코덱 알고리즘은 기존 광대역 코덱과 비트스트림 호환성을 위한 핵심 계층 부호화 알고리즘과 선형 예측 기반 정현파 코딩을 이용한 슈퍼와이드밴드 확장 알고리즘으로 이루어져 있다. 제안된 슈퍼와이드밴드 확장 코덱은 5 ms의 프레임에서 동작하며 핵심 코덱에 따라 64, 80, 96 및 112 kbit/s로 이루어진 네 개의 슈퍼와이드밴드 비트율을 제공한다. 각 비트율에 해당하는 비트스트림들은 내장형 구조를 가지고 있어 별도의 상호부호화 과정 없이 간단한 비트스트림 절단을 통해 핵심 코덱 비트스트림으로 변환할 수 있다. 제안된 코덱 알고리즘은 짧은 알고리즘 지연과 낮은 복잡도를 가지며, ITU-T에서 실시된 G.711.1/G.722 슈퍼와이드밴드 확장 코덱 자격 시험을 통과하였다.

Abstract

In this paper we proposed a candidate algorithm on G.711.1 and G.722 superwideband extension codec which is under standardization by ITU-T. The proposed codec not only provides an interoperable bitstream with ITU-T G.711.1 and G.722, but also encodes a superwideband signal with a bandwidth of 50-14,000 Hz using superwideband extension layer. The candidate codec consists of a core layer to provide an interoperability with conventional wideband codecs and superwideband extension layer using linear prediction-based sinusoidal coding. The proposed extension codec operates on 5 ms frame and provides four superwideband bitrates of 64, 80, 96, and 112 kbit/s depending on the core codec. Since the resulting bitstream has an embedded structure, it can be converted into core bitstream by simple truncation without transcoding. The proposed codec has a short algorithmic delay and low complexity and passed the qualification test of G.711.1 and G.722 superwideband extension codec performed by ITU-T.

Keywords: G.711.1, G.722, 슈퍼와이드밴드 확장, 선형 예측 기반 정현파 코딩

I. 서론

과거 음성 통신은 PSTN과 같은 전화망을 기반으로 발전되어 왔으며, 소위 협대역(NB: Narrowband)이라고 불리는 300-3,400 Hz의 제한된 신호 대역폭 내에서 최고의 품질을 제공하는데 집중되었다. 그러나 이러한 대역폭 제한으로 인해 사용자에게 제공할 수 있는 서비스

* 정회원, 한국전자통신연구원
(ETRI)

** 정회원, 충남대학교 메카트로닉스공학과
(Chungnam National University)

※ 본 연구는 지식경제부 및 한국산업기술평가관리원의 IT성장동력기술개발사업의 일환으로 수행하였음. [KI001918, FMC 어커스틱 융합코덱 및 제어 기술 연구]

접수일자: 2010년7월6일, 수정완료일: 2010년8월10일

품질 역시 제약되었다. 근래에 들어 통신망의 기술 발전으로 인해 전송 대역폭은 유선망뿐만 아니라 무선망에서도 급속도로 확대되고 있는 추세이다.

이와 같은 통신망 환경 변화에 따라 ITU-T (International Telecommunication Union - Telecommunication Standardization Sector)에서 1988년에 G.722를 표준화한 이래로, 2000년을 기점으로 최근까지 ITU-T를 비롯한 3GPP (3rd Generation Partnership Project), 3GPP2 등의 국제 표준화 단체는 50-7,000 Hz에 해당하는 광대역 신호(WB: Wideband)를 처리할 수 있는 G.722.2 (AMR-WB), G.729.1, G.718, G.711.1 및 VMR-WB 등의 광대역 코덱들을 표준화하였다. 이 광대역 코덱들은 종래의 협대역 코덱에 비해 보다 우수한 품질의 양방향 통신 서비스를 가능하게 한다. 최근에는 이와 같은 추세에서 더 나아가 50-14,000 Hz에 해당하는 슈퍼와이드밴드 신호(SWB: Super-wideband)를 효과적으로 부호화함으로써 더욱 개선된 현장감과 충실도를 제공할 수 있는 슈퍼와이드밴드 코덱 개발이 활발히 진행 중에 있으며, ITU-T G.722.1 Annex C, G.729.1 Annex E, G.718 Annex B와 Skype사의 SILK 등이 대표적이다.

이러한 음성 코덱의 진화하는 추이를 살펴볼 때, 최근에 개발된 코덱들은 완전히 새로운 코덱을 개발하기 보다는 기존 협대역 혹은 광대역 코덱과의 호환성을 제공함과 동시에 신호 대역폭을 확장하는 경향이 우세하다. 이는 음성 통신 서비스의 변화가 일시에 일어나기 보다는 점진적으로 진화하기 때문이며, 새로운 코덱 설계에 있어 중요한 요소로 작용하게 된다. 즉, 기존 코덱과의 호환성 제공은 별도의 상호부호화 과정 없이 기존 통신망과 연동이 가능한 장점을 갖는다.

대표적인 슈퍼와이드밴드 코덱 개발 활동 중 하나로, ITU-T SG16 (Study Group 16) Q.10(Question 10)은 2007년 10월에 PCM (Pulse Code Modulation) 기반 광대역 코덱인 G.711.1 및 G.722와 호환 가능한 슈퍼와이드밴드 확장 코덱 표준화를 착수하였다. G.711.1은 기존 전화망뿐만 아니라 대부분의 VoIP (Voice over Internet Protocol) 서비스에서 기본 코덱으로 사용 중인 G.711 코덱과 비트스트림 호환성을 제공하며, 총 3개의 계층으로 이루어진 내장형 비트스트림의 조합으로 64, 80 (협대역), 80 (광대역), 96 kbit/s의 비트율로 부호화한다^[1]. 이 G.711.1을 구성하는 알고리즘은 잡음 피드백을 포함하는 저대역 로그 PCM, 저대역에서 핵

심 계층의 품질 향상을 위한 적응형 비트 할당 방법을 가진 내장형 PCM 확장 및 MDCT(Modified Discrete Cosine Transform) 기반의 고대역 가중 벡터 양자화로 이루어져 있다. 한편, G.722는 SB-ADPCM (SubBand-Adaptive Differential PCM) 방식을 이용하여 입력 신호를 48, 56, 64 kbit/s로 부호화한다^[2]. 이 G.722 코덱은 최근 많은 광대역 VoIP 서비스에서 채용되고 있으며, 특히 유럽 차세대 무선 전화 표준인 NG-DECT (Next Generation - Digital Enhanced Cordless Telecommunications)에서 반드시 지원해야 하는 코덱으로 선정되었다^[3]. 이와 같은 G.711.1과 G.722 코덱의 보편성은 두 코덱이 제공하는 광대역 품질뿐만 아니라 다른 광대역 코덱들에 비해 상당히 낮은 복잡도와 짧은 알고리즘 지연에 기인한다. 따라서, 본 G.711.1/G.722 슈퍼와이드밴드 확장 코덱의 요구사항은 이러한 두 코덱의 특성을 훼손하지 않는 범주 내에서 설계되

표 1. ITU-T G.711.1/G.722 SWB 확장 코덱 모노 부 요구사항 개요

Table 1. Summary on monaural part requirements of ITU-T G.711.1/G.722 SWB extension codec.

파라미터	핵심 코덱	요구사항
핵심 코덱	G.722	G.722와 비트스트림 호환 핵심 비트율: 56, 64 kbit/s
	G.711.1	G.711.1과 비트스트림 호환 핵심 비트율: 80, 96 kbit/s
신호 대역폭	공통	광대역: [50, 7000] Hz 슈퍼와이드밴드: [50, 14000] Hz
프레임 길이	공통	5 ms
비트스트림 구조	G.722/ G.711.1	내장형 비트스트림 (그림 1 참조)
음성 품질 (-26/36/16 dBov)	G.722	R1sm: NWT* G.722.1C@24 kbit/s R2sm: NWT G.722.1C@32 kbit/s R3sm: NWT G.722.1C@48 kbit/s
	G.711.1	R3sm: NWT G.722.1C@32 kbit/s R4sm: NWT G.722.1C@48 kbit/s
음악 품질 (-26dBov)	G.722/ G.711.1	음성 품질 조건과 동일
오류 조건에서 음성 품질 (-26 dBov)	G.722/ G.711.1	3 % random 프레임 오류 음성 품질 조건과 동일
배경 잡음 하에서 음성 품질	G.722/ G.711.1	- 25 dB SNR music - 20 dB SNR office noise - 30 dB SNR babble noise - 15 dB SNR interfering talker 음성 품질 조건과 동일
알고리즘 지연	G.722	≤ 15 ms
	G.711.1	≤ 20 ms
계산량	G.722	≤ 30 WMOPS**
	G.711.1	≤ 30 WMOPS

었다.

표 1은 ITU-T G.711.1/G.722 슈퍼와이드밴드 확장 코덱 모노 부의 주요 요구사항을 나타낸 것이다^[4]. 여기서, G.711.1/G.722 슈퍼와이드밴드 확장 코덱의 각 비트율을 표시하는 규칙은 다음과 같다^[4].

Rxyz	
x=비트율	인덱스: '0'(56 kbit/s), '1'(64 kbit/s), '2'(80 kbit/s), '3'(96 kbit/s), '4'(112 kbit/s)
y='n'	(협대역), 'w'(광대역), 's'(슈퍼와이드밴드)
z='m'	(모노), 's'(스테레오)

예를 들어, G.722 R3sm은 96 kbit/s 슈퍼와이드밴드 모노 비트율로, 전체 비트스트림은 G.722 64 kbit/s와 2개의 16 kbit/s 슈퍼와이드밴드 향상 계층으로 이루어진다.

표 1에서 음성 품질 비교를 위한 기준 코덱으로 사용된 G.722.1C^[11]는 ITU-T에서 표준화된 최초의 슈퍼와이드밴드 코덱으로 MLT(Modulated Lapped Transform) 기반의 변환 부호화 방식을 이용하여 슈퍼와이드밴드 대역 신호를 24, 32, 48 kbit/s의 비트율로 부호화한다. 이 G.722.1C 코덱은 매 20 ms 프레임마다 동작하며, 40 ms의 알고리즘 지연과 약 11 WMOPS의 연산량을 가진다.

본 논문에서는 표 1의 요구사항을 기반으로 제안된 G.711.1과 G.722의 슈퍼와이드밴드 확장 코덱의 후보 알고리즘에 대해서 설명한다. 본 논문의 구성은 다음과 같다. II 장과 III 장에서는 G.711.1/G.722 슈퍼와이드밴드 확장 코덱을 위해 제안된 부호화기와 복호화기 알고리즘에 대해서 각각 설명하고, IV 장에서 본 슈퍼와이드밴드 확장 코덱의 알고리즘 지연과 복잡도 및 주관적 코덱 품질 등의 코덱 성능 평가 결과를 보여준다. 끝으로 V 장에서 결론을 맺는다.

II. 부호화기

본 G.711.1/G.722 슈퍼와이드밴드 확장 코덱은 32 kHz로 샘플링된 16 비트 PCM(Pulse Code Modulation) 신호를 부호화기 입력으로 하며, 5 ms의 입력 프레임 상에서 동작한다.

* NWT: 통계 분석에서 "not worse than" 범주를 의미한다.
 ** WMOPS: "Weighted Million Operations Per Second"의 약어로 초당 수행되는 오퍼레이션의 수를 가리킨다.

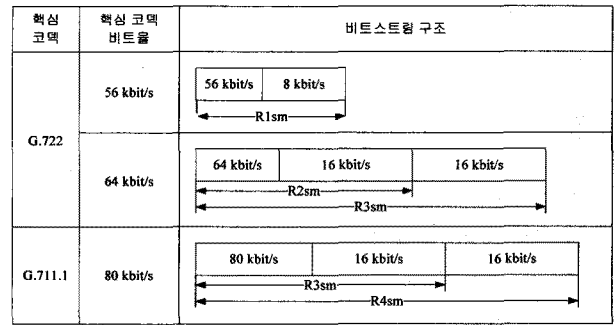


그림 1. G.711.1/G.722 SWB 확장 코덱 비트스트림 구조
 Fig. 1. Bitstream structures of G.711.1/G.722 SWB extension codec.

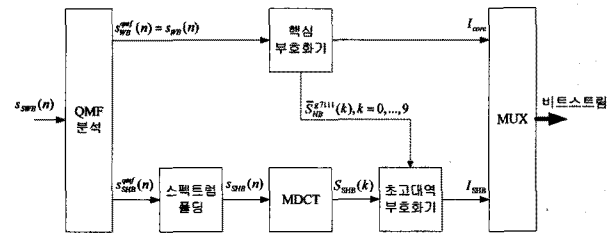


그림 2. G.711.1/G.722 SWB 확장 부호화기
 Fig. 2. G.711.1/G.722 SWB extension encoder.

그림 2는 제안된 G.711.1/G.722 슈퍼와이드밴드 확장 부호화기의 고수준 블록도를 나타낸 것이다. 슈퍼와이드밴드 입력 신호, $s_{SWB}(n)$ 는 64-차 QMF 분석을 통해 16 kHz로 샘플링된 광대역 신호(0-8 kHz)와 초고대역 신호(8-16 kHz)로 분리된다.

핵심 부호화기는 핵심 부호화 모드에 따라 G.711.1 혹은 G.722 부호화기를 이용하여 광대역 신호, $s_{WB}(n)$ 를 부호화하여 각각 80 kbit/s, 56 kbit/s 혹은 64 kbit/s의 핵심 비트스트림, I_{core} 를 출력한다. 초고대역 신호, $s_{SHB}(n)$ 는 QMF 분석의 출력 신호, $s_{SHB}^{gmf}(n)$ 를 스펙트럼 폴딩(spectral folding)함으로써 얻어진다.

$$s_{SHB}(n) = (-1)^n s_{SHB}^{gmf}(n) \quad (1)$$

이 초고대역 신호는 MDCT를 통해 주파수 영역으로 변환되고, 초고대역 부호화기는 초고대역 MDCT 계수를 부호화하여 초고대역 비트스트림, I_{SHB} 을 출력한다. 그리고 핵심 부호화 모드가 G.711.1 80 kbit/s인 경우, 입력 신호의 7-8 kHz 대역에 해당하는 G.711.1 고대역 MDCT 계수, $\bar{S}_{HB}^{g711}(k)$ 는 초고대역 부호화기에서 양자화된다. 핵심 부호화기와 초고대역 부호화기에 의해서 부호화된 각 비트스트림, I_{core} 와 I_{SHB} 는 MUX에 의해서 다중화된다.

1. 핵심 계층 부호화기

본 G.711.1/G.722 슈퍼와이드밴드 확장 코덱의 핵심 부호화기는 입력 광대역 신호를 부호화하여 ITU-T G.722 56 kbit/s, 64 kbit/s 및 G.711.1 80 kbit/s와 호환 가능한 비트스트림을 출력한다.

G.711.1 핵심 부호화기는 기존 G.711.1 부호화기와 동일하다. 그러나 G.711.1의 실제 부호화 대역이 50-7000 Hz이므로, G.711.1 부호화기의 고대역 MDCT 계수 중 7-8 kHz에 해당하는 10 개의 MDCT 계수를 추출하여 G.711.1 고대역 MDCT 이득으로 정규화하여 초고대역 부호화기로 전달한다.

$$\bar{S}_{HB}^{g7111}(k) = \frac{S_{HB}^{g7111}(k)}{g_{HB}^{g7111} + \epsilon_{HB,g}^{g7111}}, k = 0, \dots, 9 \quad (2)$$

여기서, $S_{HB}^{g7111}(k)$ 는 G.711.1 고대역 MDCT 계수, g_{HB}^{g7111} 는 고대역 MDCT 이득이고, $\epsilon_{HB,g}^{g7111}$ 는 0으로 나누어 지는 것을 방지하기 위해 더해지는 상수이다.

기존 ITU-T G.722 부호화기는 입력 광대역 신호를 8 kHz의 저대역과 고대역 신호로 나눈 다음, 각 대역 신호를 ADPCM으로 부호화한다. 이 ADPCM 부호화 방식에서 제한된 샘플 당 부호화 비트 수로 인해 발생하는 양자화 잡음은 사람의 청각기관에 의해 쉽게 인지 되므로 전체적인 음질 저하를 초래한다. 제안된 G.722 핵심 부호화기는 이러한 음질 저하를 개선하기 위해 잡음 피드백을 이용한 잡음 형상화(noise shaping)를 저대역 및 고대역 신호에 각각 적용하였다. 이 G.722 핵심 부호화기의 잡음 형상화는 ITU-T G.711.1 코덱의 저대역 잡음 형상화^[1, 5]와 동일한 방법을 사용하였다. 그림 3은 잡음 형상화를 이용한 G.722 핵심 부호화기의 부대역 부호화부를 나타낸 것이다. 잡음 형상화 필터, $F(z)$ 는 다음과 같이 주어진다.

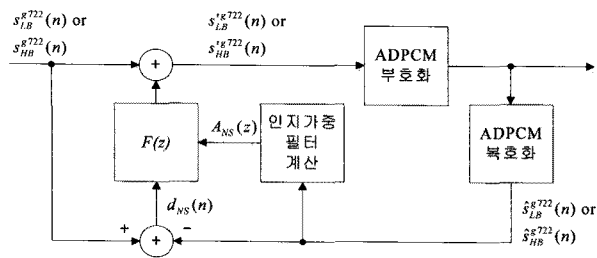


그림 3. G.722 핵심 부호화기의 잡음 형상화를 이용한 부대역 부호화부

Fig. 3. Subband encoding part with noise shaping in G.722 core encoder.

$$F(z) = A_{NS}(z/\gamma_{NS}) - 1 \quad (3)$$

여기서, $A_{NS}(z)$ 는 복호화된 각 G.722 부대역 신호로부터 선형 예측 분석을 통해 계산된 필터이고, γ_{NS} 는 인지 가중치이다. 선형 예측 필터는 과거 2 프레임의 복호화된 부대역 신호로부터 4 차의 선형 예측 분석을 통해 계산된다. 이 잡음 형상화 필터를 이용함으로써 ADPCM 양자화 잡음의 스펙트럼은 $1/A_{NS}(z/\gamma_{NS})$ 로 형상화된다. 이러한 잡음 형상화를 통해 각 부대역의 인지는 품질을 향상시킬 수 있다.

2. 초고대역 부호화기

G.711.1 및 G.722 핵심 부호화 모드에 대해서 입력 신호의 8-14 kHz 대역에 해당하는 초고대역 신호는 선형 예측-기반 정현파 코딩 (sinusoidal coding) 방식을 이용하여 부호화된다.

이를 위해, 먼저 16 kHz로 샘플링된 초고대역 신호는 MDCT 계수로 변환된다. 이 MDCT 변환은 5 ms의 프레임 길이와 5 ms의 lookahead로 이루어진 10 ms의 신호에 대해서 수행되며, 변환 식은 다음과 같다.

$$S_{SHB}(k) = \sum_{n=0}^{159} w_{TF}(n) \cos\left(\frac{\pi}{80}(n+40.5)(k+0.5)\right) s_{SHB}(n), k = 0, \dots, 79 \quad (4)$$

여기서, $w_{TF}(n)$ 은 MDCT 분석 윈도우이다.

$$w_{TF}(n) = \sin\left(\frac{\pi}{160}(n+0.5)\right), n = 0, \dots, 159 \quad (5)$$

그림 4는 G.722/G.711.1 슈퍼와이드밴드 확장 코덱의 초고대역 부호화기를 나타낸 것이다.

초고대역 MDCT 계수는 전체 80 개의 MDCT 계수 중 8 kHz에서 14 kHz 대역에 해당하는 60 개를 각 12 개의 계수로 이루어진 5 개의 부대역으로 나눈 다음, 고주파 영역을 강조하기 위해 주파수 영역에서 다음과

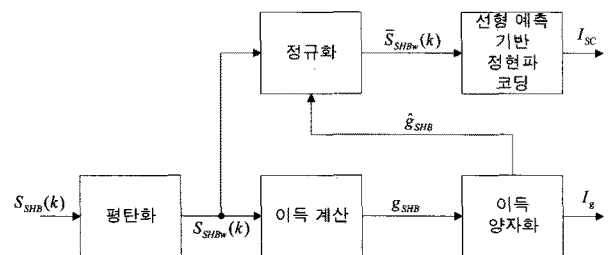


그림 4. 초고대역 부호화기

Fig. 4. Super higher-band encoder.

같이 계수 평탄화 과정을 거친다.

$$S_{SHBw}(k) = \begin{cases} S_{SHB}(k), & 0 \leq k < 12 \\ S_{SHB}(k), & 12 \leq k < 24 \\ (4/3)S_{SHB}(k), & 24 \leq k < 36 \\ (5/3)S_{SHB}(k), & 36 \leq k < 48 \\ (2)S_{SHB}(k), & 48 \leq k < 60 \end{cases} \quad (6)$$

평탄화된 초고대역 MDCT 계수의 전체 이득을 계산하기 위해 RMS(Root Mean Square)를 구한 다음, 로그 영역으로 변환하여 각 핵심 계층 모드에 따라 정해진 비트 수로 스칼라 양자화한다.

$$g_{SHB} = \sqrt{\frac{1}{60} \sum_{k=0}^{59} (S_{SHBw}(k))^2} \quad (7)$$

$$g_{SHB}^{LOG} = \log_{10}(g_{SHB} + \epsilon_{SHB,g})$$

여기서, $\epsilon_{SHB,g}$ 는 로그 항이 0이 되는 것을 방지하기 위해 더해지는 상수이다. 이득 양자화를 위해 G.722 56 kbit/s 모드인 경우에 5 비트를 할당하고, G.722 64 kbit/s와 G.711.1 80 kbit/s는 7 비트를 할당하였다.

이득 양자화를 통해 얻어진 양자화된 이득을 이용하여 초고대역 MDCT 계수를 정규화한다.

$$\bar{S}_{SHBw}(k) = \frac{S_{SHBw}(k)}{\hat{g}_{SHB} + \epsilon_{SHB,g}}, k = 0, \dots, 59 \quad (8)$$

여기서, \hat{g}_{SHB} 는 양자화된 초고대역 이득이다.

정규화된 초고대역 MDCT 계수는 선형 예측 기반 정현파 코딩을 통해 양자화된다.

그림 5는 선형 예측 기반 정현파 코딩을 이용한 MDCT 계수 부호화기의 블록도를 나타낸 것이다.

먼저, 정규화된 MDCT 계수를 기반으로 주파수 영역에서 자기 상관 방법을 이용한 선형 예측 분석을 통해 선형 예측 계수를 구한다. 이 선형 예측 합성 필터의 전달 함수는 다음과 같다.

$$\frac{1}{A_{TF}(z)} = \frac{1}{1 + \sum_{i=1}^{p_{TF}} a_i z^{-i}} \quad (9)$$

여기서, p_{TF} 는 선형 예측 차수이다. 이 선형 예측 계수는 LSF(Line Spectral Frequency)로 변환되어 벡터 양자화된다. 표 2는 각 핵심 코덱 별 선형 예측 차수와 양자화 비트 수를 나타낸 것이다.

정규화된 MDCT 계수를 부호화하기 위한 선형 예측 분석 기반의 정현파 부호화 과정은 3 단계로 수행된다.

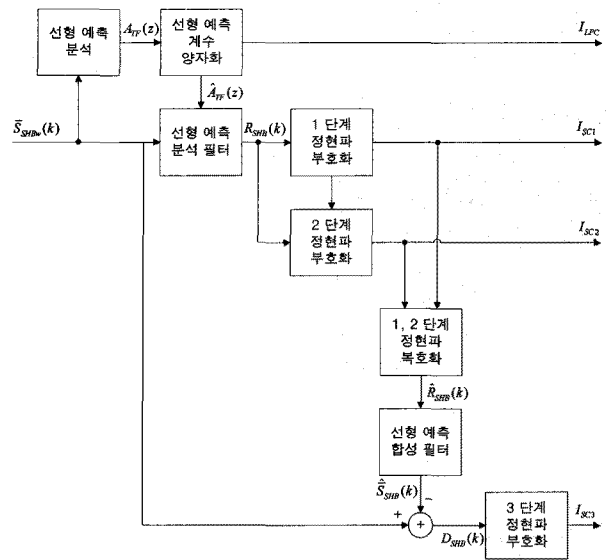


그림 5. 선형 예측 기반 정현파 부호화기
Fig. 5. Linear prediction-based sinusoidal coder.

표 2. 핵심 코덱 별 선형 예측 차수 및 양자화 비트 수

Table 2. Linear prediction order and number of quantization bits for core codecs.

핵심 코덱	선형 예측 차수 (p_{TF})	양자화 비트 수
G.722 56 kbit/s	6	5
G.722 64 kbit/s	8	7
G.711.1 80 kbit/s	8	7

제 1 단계와 2 단계 정현파 부호화는 선형 예측 잔차 MDCT 계수에 대해 적용되고, 3 단계 정현파 부호화는 정규화된 MDCT 계수와 1 단계와 2 단계에서 복원된 계수의 차이 MDCT 계수에 대해서 적용된다.

제 1 단계와 2 단계 정현파 부호화를 위해서 양자화된 선형 예측 계수를 이용하여 선형 예측 분석을 통해 잔차 MDCT 계수, $R_{SHB}(k)$ 를 구한다.

$$R_{SHB}(k) = \bar{S}_{SHBw}(k) + \sum_{i=1}^{p_{TF}} \hat{a}_i \bar{S}_{SHBw}(k-i), \quad k = 0, \dots, 59 \quad (10)$$

여기서, \hat{a}_i 는 양자화된 선형 예측 계수이다.

제 1 단계 정현파 부호화에서 선형 예측 잔차 MDCT 계수는 핵심 계층 모드와 정해진 부대역에 따라 표 3에 정의된 트랙 구조로 교차 삽입된다. 표 3은 해당 부대역의 주파수 대역, 트랙 인덱스, 트랙당 펄스 개수 및 MDCT 계수 인덱스로 표현되는 각 트랙의 위치 정보를 포함하고 있다.

제 2 단계 부호화 과정은 원 잔차 MDCT 계수에서 1 단계 부호화에서 얻어진 양자화된 잔차 계수의 차이,

표 3. 제 1 단계 정현파 부호화를 위한 트랙 구조
Table 3. Track structure of the first stage sinusoidal coding.

(가) G.722 56 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
[8-12.8]	0	i_0	0, 6, 12, 18, 24, 30, 36, 42
	1	i_1	1, 7, 13, 19, 25, 31, 37, 43
	2	i_2	2, 8, 14, 20, 26, 32, 38, 44
	3	i_3	3, 9, 15, 21, 27, 33, 39, 45
	4	i_4	4, 10, 16, 22, 28, 34, 40, 46
	5	i_5	5, 11, 17, 23, 29, 35, 41, 47

(나) G.722 64 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
8-12.8	0	i_0, i_1	0, 6, 12, 18, 24, 30, 36, 42
	1	i_2, i_3	1, 7, 13, 19, 25, 31, 37, 43
	2	i_4, i_5	2, 8, 14, 20, 26, 32, 38, 44
	3	i_6, i_7	3, 9, 15, 21, 27, 33, 39, 45
	4	i_8, i_9	4, 10, 16, 22, 28, 34, 40, 46
	5	i_{10}, i_{11}	5, 11, 17, 23, 29, 35, 41, 47
12.8-13.2	0	i_{12}	48, 49, 50, 51

(다) G.711.1 80 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
8-12.8	0	i_0, i_1	0, 6, 12, 18, 24, 30, 36, 42
	1	i_2, i_3	1, 7, 13, 19, 25, 31, 37, 43
	2	i_4, i_5	2, 8, 14, 20, 26, 32, 38, 44
	3	i_6, i_7	3, 9, 15, 21, 27, 33, 39, 45
	4	i_8, i_9	4, 10, 16, 22, 28, 34, 40, 46
	5	i_{10}, i_{11}	5, 11, 17, 23, 29, 35, 41, 47
12.8-14	0	i_{12}	48, 51, 54, 57
	1	i_{13}	49, 52, 55, 58
	2	i_{14}	50, 53, 56, 59

$D_{LP}(k)$ 를 계산한 다음, 구해진 잔차 계수의 차이에 대해서 표 4의 트랙 구조에 따라 정현파 부호화를 수행한다.

$$D_{LP}(k) = R_{SHB}(k) - \hat{R}_{SC1}(k), k=0, \dots, 59 \quad (11)$$

여기서, $\hat{R}_{SC1}(k)$ 는 1 단계 정현파 부호화에서 복원된 잔차 MDCT 계수이다.

표 4-(가)에서 첫 번째 트랙은 8-12.8 kHz 대역을 8개의 계수를 갖는 6 개 부대역으로 나누었을 때 1 단계 정현파 부호화 과정에서 가장 적은 펄스를 갖는 부대역에 해당하며, 그 트랙의 위치 정보는 선택된 부대역의 첫 번째 계수 인덱스에 대한 상대적인 위치이다.

마지막으로, 제 3 단계 부호화를 위해 1 단계와 2 단계에서 얻어진 양자화된 잔차 계수를 선형 예측 합성

표 4. 제 2 단계 정현파 부호화를 위한 트랙 구조
Table 4. Track structure of the second stage sinusoidal coding.

(가) G.722 64 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
b-(b+0.8)	0	i_{13}	0, 1, 2, 3, 4, 5, 6, 7
8-12.8	0	i_{14}	0, 6, 12, 18, 24, 30, 36, 42
	1	i_{15}	1, 7, 13, 19, 25, 31, 37, 43
	2	i_{16}	2, 8, 14, 20, 26, 32, 38, 44
	3	i_{17}	3, 9, 15, 21, 27, 33, 39, 45
	4	i_{18}	4, 10, 16, 22, 28, 34, 40, 46
	5	i_{19}	5, 11, 17, 23, 29, 35, 41, 47
12.8-14	0	i_{20}, i_{21}	48, 51, 54, 57
	1	i_{22}, i_{23}	49, 52, 55, 58
	2	i_{24}, i_{25}	50, 53, 56, 59

(나) G.711.1 80 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
8-12.8	0	i_{15}	0, 6, 12, 18, 24, 30, 36, 42
	1	i_{16}	1, 7, 13, 19, 25, 31, 37, 43
	2	i_{17}	2, 8, 14, 20, 26, 32, 38, 44
	3	i_{18}	3, 9, 15, 21, 27, 33, 39, 45
	4	i_{19}	4, 10, 16, 22, 28, 34, 40, 46
	5	i_{20}	5, 11, 17, 23, 29, 35, 41, 47

필터를 통해 양자화된 MDCT 계수를 복원한다.

$$\begin{aligned} \hat{R}_{SHB}(k) &= \hat{R}_{SC1}(k) + \hat{R}_{SC2}(k), k=0, \dots, 59 \\ \hat{S}_{SHB}(k) &= \hat{R}_{SHB}(k) - \sum_{i=1}^{PTF} \hat{S}_{SHB}(k-i), \\ k &= 0, \dots, 59 \end{aligned} \quad (12)$$

여기서, $\hat{R}_{SC2}(k)$ 는 2 단계 정현파 부호화에서 복원된 잔차 MDCT 계수이다. 원 정규화된 MDCT 계수와 1 단계와 2 단계 정현파 부호화 과정에서 복원된 MDCT 계수의 차이, $D_{SHB}(k)$ 를 계산한다.

$$D_{SHB}(k) = \bar{S}_{SHBw}(k) - \hat{S}_{SHB}(k), k=0, \dots, 59 \quad (13)$$

제 3 단계 정현파 부호화는 $D_{SHB}(k)$ 에 대해서 표 5에 나타낸 트랙 구조에 따라 부호화 과정을 수행한다.

앞에서 설명한 각 모드 별 트랙 구조를 이용하여 각 트랙에 대한 펄스 검색 과정에서 해당 트랙의 MDCT 계수들을 절대값이 큰 순서로 정렬한 다음 각 트랙 별로 정해진 개수의 가장 큰 펄스들을 선택한다. 선택된 각 펄스들은 펄스의 위치, 부호 및 크기로 이루어진 정현파 파라미터들로 나누어 각각 부호화된다. 펄스의 위치는 수식 (14)와 같이 \log_2 (트랙 당 가능한 위치) 비트

표 5. 제 3 단계 정현파 부호화를 위한 트랙 구조
Table 5. Track structure of the third stage sinusoidal coding.

(가) G.722 64 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
8.6-11	0	i_{26}, i_{27}	6, 9, 12, 15, 18, 21, 24, 27
	1	i_{28}, i_{29}	7, 10, 13, 16, 19, 22, 25, 28
	2	i_{30}, i_{31}	8, 11, 14, 17, 20, 23, 26, 29

(나) G.711.1 80 kbit/s 핵심 계층 모드

부대역 [kHz]	트랙	펄스	위치
8.6-11	0	i_{21}, i_{22}	6, 9, 12, 15, 18, 21, 24, 27
	1	i_{23}, i_{24}	7, 10, 13, 16, 19, 22, 25, 28
	2	i_{25}, i_{26}	8, 11, 14, 17, 20, 23, 26, 29

로 부호화된다.

$$I_{pos}(t, l) = \frac{pos(t, l) - t}{N_{track}}, \quad (14)$$

$$l = 0, \dots, (N_{t,p} - 1), t = 0, \dots, (N_{track} - 1)$$

여기서, $pos(t, l)$ 는 특정 부대역에서 t -번째 트랙의 l -번째 펄스의 위치 정보, $N_{t,p}$ 는 t -번째 트랙의 펄스 개수이고 N_{track} 는 부대역의 트랙 개수이다.

펄스의 부호는 다음과 같이 각 1 비트로 부호화된다.

$$I_{sign}(t, l) = \begin{cases} 1, & T_{SC}(pos(t, l)) \geq 0 \\ 0, & T_{SC}(pos(t, l)) < 0 \end{cases} \quad (15)$$

여기서, $T_{SC}(pos(t, l))$ 는 $pos(t, l)$ 에서 각 정현파 부호화 단계의 목표 신호로 $R_{SHB}(k)$, $D_{LP}(k)$ 및 $D_{SHB}(k)$ 를 가리킨다. 트랙 당 펄스 개수가 2 개인 경우($N_{t,p}=2$), 적절한 펄스 인덱싱을 통해서 1 비트의 부호 비트를 절약할 수 있다^[6]. 즉, 동일 트랙의 두 펄스의 부호 중 첫 번째 펄스의 부호만 전송하고, 나머지 펄스의 부호는 두 펄스의 상대적인 위치로 부호화한다. 복호화기에서 두 번째 펄스의 부호는 두 펄스의 위치 정보와 첫 번째 펄스의 부호로부터 추정할 수 있다.

펄스의 크기는 로그 영역으로 변환된 다음, 각 부대역에 따라 동일 트랙의 펄스 쌍 혹은 트랙 쌍에 해당하는 2-차의 크기 벡터를 2 비트로 양자화한다.

$$amp(t, l) = \log_{10}(|T_{SC}(pos(t, l))| + \epsilon_{SHB,amp}) \quad (16)$$

여기서, $\epsilon_{SHB,amp}$ 는 로그 항이 0이 되는 것을 방지하기 위해 더해진 상수이다.

한편, II-1 절에서 설명한 바와 같이 핵심 부호화 모드

가 G.711.1 80 kbit/s인 경우, G.711.1 핵심 부호화기로부터 전달된 입력 신호의 7-8 kHz 대역에 해당하는 G.711.1 고대역 MDCT 계수는 초고대역 복호화기에서 (3, 3, 4)-차의 부-벡터로 나뉘어 각각 6, 6, 8 비트로 벡터 양자화된다.

III. 복호화 알고리즘

그림 6은 G.711.1/G.722 슈퍼와이드밴드 확장 부호화기의 블록도를 나타낸 것이다.

수신된 G.711.1/G.722 SWB 확장 비트스트림은 역다중화기를 통해 핵심 비트스트림, I_{core} 와 초고대역 비트스트림, I_{SHB} 으로 역다중화된다. 핵심 복호화기는 핵심 비트스트림을 복호화하여 복원된 광대역 신호, $\hat{s}_{WB}(n)$ 를 출력한다. 초고대역 비트스트림은 초고대역 복호화기를 통해 MDCT 계수, $\hat{S}_{SHB}(k)$ 로 복원된다. IMDCT(Inverse MDCT)는 복원된 MDCT 계수를 역변환하여 초고대역 신호, $\hat{s}_{SHB}(n)$ 를 출력하고, 초고대역 신호는 다시 스펙트럼 폴딩을 거쳐 QMF 합성 필터의 입력 신호, $\hat{s}_{SHB}(n)$ 로 변환된다. QMF 합성 필터는 복원된 광대역 신호와 초고대역 신호로부터 슈퍼와이드밴드 신호, $\hat{s}_{SWB}(n)$ 를 합성한다. 만약 프레임이 손실되었다면, 광대역 신호와 초고대역 신호는 각 대역의 오류 은닉 과정을 통해 복구된다.

그리고 핵심 부호화 모드가 G.711.1 80 kbit/s인 경우, 초고대역 복호화기에서 신호의 7-8 kHz 대역에 해당하는 G.711.1 고대역 MDCT 계수, $\hat{S}_{HB}^{711.1}(k)$ 를 복원하여 G.711.1 핵심 부호화기로 전달한다.

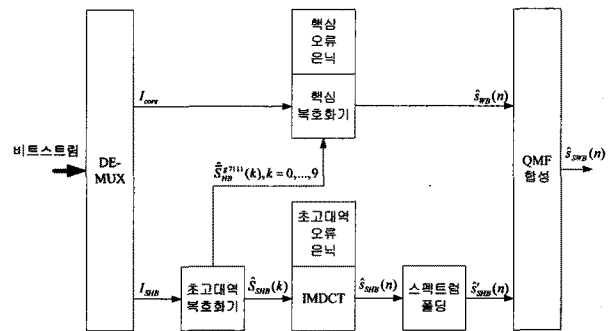


그림 6. G.722/G.711.1 SWB 확장 복호화기
Fig. 6. G.722/G.711.1 SWB extension decoder.

1. 핵심 복호화기

핵심 복호화기는 G.711.1 및 G.722 핵심 부호화 모드 모두 기존 ITU-T G.711.1 및 G.722와 동일한 과정을 거쳐 광대역 신호를 복원한다. 다만, G.711.1 핵심 부호화 모드의 경우에 초고대역 복호화기로부터 전달된 신호의 7-8 kHz 대역 MDCT 계수를 포함하여 G.711.1 고대역 IMDCT 과정을 수행한다.

$$\hat{S}_{HBm}^{g7111}(k) = \hat{g}_{HB}^{g7111} \hat{S}_{HB}^{g7111}(k), k = 0, \dots, 9 \quad (17)$$

여기서, \hat{g}_{HB}^{g7111} 는 양자화된 G.711.1 고대역 MDCT 이득이고, $\hat{S}_{HBm}^{g7111}(k)$ 는 역-가중된 고대역 MDCT 계수이다.

2. 초고대역 복호화기

그림 7은 초고대역 복호화기의 블록도를 나타낸 것이다. 초고대역 복호화기는 초고대역 MDCT 계수의 이득, 선형 예측 계수, 1 단계 정현파 파라미터, 2 단계 정현파 파라미터 및 3 단계 정현파 파라미터를 각각 복호화한다. 복호화된 제 1, 2, 3 단계 정현파 파라미터로부터 표 3, 4, 5의 트랙 구조를 이용하여 각 부대역에 해당하는 MDCT 계수를 복원한다.

$$\hat{T}_{SC}(k) = \begin{cases} \text{sign}(t,l)10^{amp_q(t,l)}, & k = \text{pos}(t,l) \\ 0, & k \neq \text{pos}(t,l) \end{cases} \quad (18)$$

여기서, $amp_q(t,l)$ 은 크기 코드북 인덱스에 의해서 참조되는 양자화된 t -번째 트랙의 l -번째 펄스 크기이다.

양자화된 잔차 MDCT 계수는 수식 (12)와 같이 1, 2 단계 정현파 복호화에서 복원된 두 MDCT 계수의 합으로 계산되며, 이 계수는 다시 선형 예측 합성을 통해

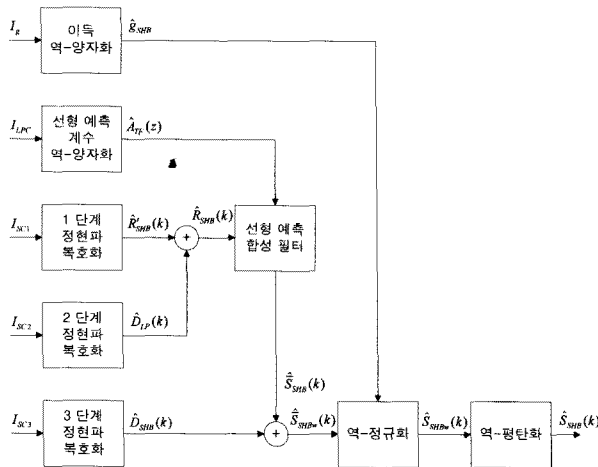


그림 7. 초고대역 복호화기
Fig. 7. Super higher-band decoder.

양자화된 MDCT 계수, $\hat{S}_{SHB}(k)$ 를 복원한다. 그리고, 정규화된 초고대역 MDCT 계수는 $\hat{S}_{SHB}(k)$ 와 3 단계 정현파 복호화를 통해 얻어지는 오류 MDCT 계수, $\hat{D}_{SHB}(k)$ 의 합으로 계산된다.

$$\hat{S}_{SHBw}(k) = \hat{S}_{SHB}(k) + \hat{D}_{SHB}(k), k = 0, \dots, 59 \quad (19)$$

정규화된 MDCT 계수, $\hat{S}_{SHBw}(k)$ 는 이득 역-정규화 및 수식 (6)의 역-과정으로 역-평탄화 과정을 거친 후, IMDCT를 통해 양자화된 MDCT 계수, $\hat{S}_{SHB}(k)$ 를 시간 영역의 초고대역 신호, $\hat{s}_{SHB}(n)$ 로 변환된다.

$$y(n) = \frac{1}{80} w_{TF}(n) \sum_{k=0}^{79} \left(\frac{\pi}{80} (k+0.5)(n+40.5) \right) \hat{S}_{SHB}(k),$$

$$n = 0, \dots, 159$$

$$\hat{s}_{SHB}(n) = y'(n+80) + y(n), n = 0, \dots, 79 \quad (20)$$

여기서, $y(n)$ 와 $y'(n)$ 는 각각 현재 프레임과 이전 프레임의 역-변환된 시간 영역 중간 신호이며, $y'(n)$ 의 후반부와 $y(n)$ 의 전반부를 OLA(Overlap-And-Add)하여 초고대역 신호를 복원한다.

3. 오류 은닉

프레임이 손실되었을 때 각 대역에 대한 오류 은닉 과정이 실행되며, 광대역 신호는 핵심 오류 은닉 과정을 통해 복구되고 초고대역 신호는 초고대역 오류 은닉 과정을 통해 복구된다.

핵심 오류 은닉 과정은 G.711.1의 경우 기존의 G.711.1 오류 은닉^[1]과 동일한 방식을 사용하고, G.722는 G.722 Appendix IV^[7]을 G.711.1/G.722 SWB 확장 코덱의 프레임 길이인 5 ms에 맞도록 수정하여 적용하였다. 기존 G.722 Appendix IV는 10 ms 혹은 20 ms 프레임에 대해서 동작한다.

초고대역 오류 은닉 과정은 감쇄된 지난 정상 프레임의 IMDCT 시간 영역 중간 신호를 손실된 프레임에 대한 시간 영역 중간 신호로 복사한다.

$$y(n) = \alpha_{SHB,FEC} y'(n), n = 0, \dots, 159 \quad (21)$$

여기서, $\alpha_{SHB,FEC}$ 는 감쇄 인자로 0.875이다. 이와 같이 복사된 중간 신호를 이용하여 수식 (20)의 OLA 과정을 수행하여 초고대역 신호를 복원한다.

IV. 성능 평가

1. 알고리즘 지연 및 복잡도

본 논문에서 제안한 G.711.1/G.722 슈퍼와이드밴드 확장 코덱의 알고리즘 지연에 대한 항목별 기여분은 표 6과 같으며, 전체 알고리즘 지연은 핵심 코덱에 따라 각각 11.969 ms와 13.844 ms이다.

표 7과 8은 제안된 G.711.1/G.722 SWB 확장 코덱의 계산량과 메모리 요구량을 각각 나타낸 것이다.

표 6. G.711.1/G.722 SWB 확장 코덱의 항목별 지연
Table 6. Delay contributions of G.711.1/G.722 SWB extension codec.

핵심 코덱	항목	지연
G.722	입력 프레임 길이	5 ms
	MDCT lookahead	5 ms
	QMF 분석/합성 필터	1.969 ms
G.711.1	G.711.1 핵심 코덱	11.875 ms
	QMF 분석/합성 필터	1.969 ms

표 7. G.711.1/G.722 SWB 확장 코덱의 계산량 (단위: WMOPS)

Table 7. Computational complexity of G.711.1/G.722 SWB extension codec (in WMOPS).

핵심 코덱	부호화기	복호화기	합계
G.722	14.96	10.19	25.15
G.711.1	15.19	7.96	23.15

표 8. G.711.1/G.722 SWB 확장 코덱의 메모리 요구량 (단위: 16-비트 워드)

Table 8. Memory requirements of G.711.1/G.722 SWB extension codec (in 16-bit words).

핵심 코덱	메모리 종류	합계
G.722	RAM (static+scratch)	4,694
	Program ROM	4,731
	Data ROM	4,447
G.711.1	RAM (static+scratch)	5,266
	Program ROM	4,869
	Data ROM	5,736

2. 주관적인 품질 평가

본 논문에서 제안된 G.711.1/G.722 슈퍼와이드밴드 확장 코덱의 품질 평가를 위해 주관적인 청취 시험을 실시하였다. 시험 계획서는 ITU-T SG12 Q.7 전문가들에 의해서 준비되었으며^[8], 본 청취 시험의 개요를 표 9에 나타내었다.

실험 1a는 깨끗한 음성에 대한 주관적인 품질을 측정하고, 실험 2는 배경 잡음을 갖는 음성에 대한 품질, 그리고 실험 3a는 음악 및 혼합 콘텐츠에 대한 코덱 품질

표 9. 주관적인 청취 시험 개요

Table 9. Overview of the subjective listening test.

실험	1a	2a/b/c/d	3a
조건	깨끗한 음성	배경 잡음 음성	음악 및 혼합 콘텐츠
배경 잡음	없음	2a:Music@25dB SNR 2b:Office@20dB SNR 2c:Babble@30dB 2d:IT@15dB	없음
기준 코덱	ITU-T G.722.1C 24 kbit/s, 32 kbit/s, 48 kbit/s		
시험 방법	ITU-R BS.1285 triple stimulus/hidden ref./double blind test		
피험자 수	24 명	24 명	24 명
화자 수	4 명	4 명	음악: 2 종 혼합 콘텐츠: 2 종
샘플 수	20	20	20
평가 기관	ETRI	2a/b/c: Dynastat Inc. 2d:ETRI	Dynastat Inc.
언어	한국어	2a/b/c: 북미 영어 2d: 한국어	혼합 콘텐츠: 북미 영어

을 평가한다. 실험 2는 배경 잡음의 종류에 따라 4 개의 부-실험들로 나누어지는데, 각 부-실험은 배경 음악, 사무실 잡음, 사람의 재잘거림 및 간섭 화자가 존재하는 환경에서 음성 품질을 측정한다. 이 주관적인 품질 시험을 위해 ITU-R BS.1285의 “triple stimulus/hidden reference/double blind test” 방법을 사용하였다^[9]. 이 방법은 ‘Ref-A-B’ 방법이라고도 하는데, 시험에 참여하는 피험자는 기준 샘플 “Ref”, 샘플 “A” 및 샘플 “B”로 이루어진 3 개의 오디오 샘플을 각 시행마다 2 회 반복하여 듣게 된다. 이 때, 샘플 “A”와 샘플 “B” 중 하나는 기준 샘플과 동일한 샘플이고 나머지는 코덱에 의해서 처리된 샘플이다. “A”와 “B” 중 기준 샘플은 무작위로 할당되며, 피험자는 “A”와 “B” 중에 기준 샘플을 먼저 식별한 다음, 나머지 샘플에 대해서는 기준 샘플 대비 품질 저하 정도를 1.0에서 4.9의 점수로 평가한다.

표 10은 각 실험 별 주관적인 청취 시험 결과를 나타낸 것으로, 각 시험 조건과 기준 조건에 대한 평균 점수, 표준 편차, 95% 신뢰구간(CI95) 및 품질 요구사항 시험 결과 등을 포함하고 있다. 여기서, 품질 요구사항 시험은 95 %의 신뢰구간에서 Student t-검증을 이용하였다. 표 10에서 본 논문에서 제안된 슈퍼와이드밴드 확장 코덱은 실험 3a의 CuT/G.722 R1sm 조건을 제외한 모든 품질 요구사항을 만족하는 것을 알 수 있다. 실험 3a의 CuT/G.722 R1sm에서 품질 요구사항을 만족하지 못한 것은 8 kbit/s의 제한된 슈퍼와이드밴드 확장 비트율로 인해 선형 예측 기반 정현파 코딩 방식이 음악 및 혼합 콘텐츠에 포함된 복잡한 하모닉 성분을 충분히 표현하지 못한 것으로 판단된다.

표 10. 주관적인 청취 시험 결과

Table 10. Subjective listening test results.

실험	시험 코덱					기준 코덱					Pass/Fail
	시험 조건	FER*	평균	표준편차	CI95	기준 조건	FER	평균	표준 편차	CI95	
1a	CuT**/G.722 R1sm	0 %	4.439	0.486	0.082	G.722.1C@24kbit/s	0 %	3.465	0.954	0.162	Pass
	CuT/G.722 R2sm	0 %	4.738	0.362	0.061	G.722.1C@32kbit/s	0 %	3.933	0.842	0.143	Pass
	CuT/G.722 R3sm	0 %	4.761	0.354	0.060	G.722.1C@48kbit/s	0 %	4.428	0.549	0.093	Pass
	CuT/G.711.1 R3sm	0 %	4.703	0.346	0.059	G.722.1C@32kbit/s	0 %	3.933	0.842	0.143	Pass
	CuT/G.711.1 R4sm	0 %	4.738	0.339	0.058	G.722.1C@48kbit/s	0 %	4.428	0.549	0.093	Pass
	CuT/G.722 R1sm	3 %	3.285	1.047	0.177	G.722.1C@24kbit/s	3 %	2.519	0.983	0.167	Pass
	CuT/G.722 R2sm	3 %	3.465	1.147	0.194	G.722.1C@32kbit/s	3 %	2.768	1.065	0.181	Pass
	CuT/G.722 R3sm	3 %	3.700	1.072	0.182	G.722.1C@48kbit/s	3 %	2.879	1.078	0.183	Pass
	CuT/G.711.1 R3sm	3 %	3.773	1.013	0.172	G.722.1C@32kbit/s	3 %	2.768	1.065	0.181	Pass
2a	CuT/G.722 R1sm	0 %	3.771	0.819	0.164	G.722.1C@24kbit/s	0 %	2.706	1.012	0.203	Pass
	CuT/G.722 R2sm	0 %	4.551	0.461	0.092	G.722.1C@32kbit/s	0 %	3.195	0.903	0.181	Pass
	CuT/G.722 R3sm	0 %	4.579	0.521	0.104	G.722.1C@48kbit/s	0 %	4.180	0.580	0.116	Pass
	CuT/G.711.1 R3sm	0 %	4.392	0.580	0.116	G.722.1C@32kbit/s	0 %	3.195	0.903	0.181	Pass
	CuT/G.711.1 R4sm	0 %	4.789	0.277	0.056	G.722.1C@48kbit/s	0 %	4.180	0.580	0.116	Pass
2b	CuT/G.722 R1sm	0 %	3.774	0.754	0.151	G.722.1C@24kbit/s	0 %	3.083	0.901	0.181	Pass
	CuT/G.722 R2sm	0 %	4.743	0.276	0.055	G.722.1C@32kbit/s	0 %	3.690	0.794	0.159	Pass
	CuT/G.722 R3sm	0 %	4.821	0.210	0.042	G.722.1C@48kbit/s	0 %	4.356	0.528	0.106	Pass
	CuT/G.711.1 R3sm	0 %	4.541	0.479	0.095	G.722.1C@32kbit/s	0 %	3.690	0.794	0.159	Pass
	CuT/G.711.1 R4sm	0 %	4.881	0.164	0.032	G.722.1C@48kbit/s	0 %	4.356	0.528	0.106	Pass
2c	CuT/G.722 R1sm	0 %	3.938	0.681	0.137	G.722.1C@24kbit/s	0 %	2.789	0.902	0.181	Pass
	CuT/G.722 R2sm	0 %	4.501	0.526	0.105	G.722.1C@32kbit/s	0 %	3.414	0.894	0.179	Pass
	CuT/G.722 R3sm	0 %	4.559	0.473	0.095	G.722.1C@48kbit/s	0 %	4.279	0.565	0.113	Pass
	CuT/G.711.1 R3sm	0 %	4.368	0.542	0.109	G.722.1C@32kbit/s	0 %	3.414	0.894	0.179	Pass
	CuT/G.711.1 R4sm	0 %	4.827	0.198	0.039	G.722.1C@48kbit/s	0 %	4.279	0.565	0.113	Pass
2d	CuT/G.722 R1sm	0 %	4.314	0.727	0.123	G.722.1C@24kbit/s	0 %	3.924	0.784	0.133	Pass
	CuT/G.722 R2sm	0 %	4.657	0.463	0.079	G.722.1C@32kbit/s	0 %	4.396	0.630	0.107	Pass
	CuT/G.722 R3sm	0 %	4.722	0.492	0.083	G.722.1C@48kbit/s	0 %	4.649	0.382	0.065	Pass
	CuT/G.711.1 R3sm	0 %	4.679	0.447	0.076	G.722.1C@32kbit/s	0 %	4.396	0.630	0.107	Pass
	CuT/G.711.1 R4sm	0 %	4.763	0.427	0.072	G.722.1C@48kbit/s	0 %	4.649	0.382	0.065	Pass
3a	CuT/G.722 R1sm	0 %	3.831	0.876	0.175	G.722.1C@24kbit/s	0 %	4.065	0.749	0.150	Fail
	CuT/G.722 R2sm	0 %	4.663	0.373	0.075	G.722.1C@32kbit/s	0 %	4.510	0.439	0.087	Pass
	CuT/G.722 R3sm	0 %	4.745	0.354	0.071	G.722.1C@48kbit/s	0 %	4.742	0.296	0.059	Pass
	CuT/G.711.1 R3sm	0 %	4.588	0.388	0.078	G.722.1C@32kbit/s	0 %	4.510	0.439	0.087	Pass
	CuT/G.711.1 R4sm	0 %	4.735	0.312	0.063	G.722.1C@48kbit/s	0 %	4.742	0.296	0.059	Pass

V. 결론

본 논문에서는 ITU-T SG16 Q.10에서 표준화 추진 중인 G.711.1/G.722 슈퍼와이드밴드 확장 코덱을 위한 후보 알고리즘을 제안하였다. 제안된 슈퍼와이드밴드 확장 코덱은 기존 코덱과의 호환성을 위한 G.711.1 및 G.722 핵심 코덱 알고리즘뿐만 아니라, G.722 핵심 코덱의 성능 개선을 위한 잡음 형상화 및 선형 예측 기반 정현파 코딩을 이용한 슈퍼와이드밴드 확장 방법을 채용하였다.

본 확장 코덱은 5 ms의 프레임에서 동작하며, 56 kbit/s, 64 kbit/s 및 80 kbit/s의 기존 ITU-T G.711.1

및 G.722 광대역 코덱과의 비트스트림 호환성을 지원하고 각 핵심 코덱에 따라 R1sm(64 kbit/s), R2sm(80 kbit/s), R3sm(96 kbit/s), R4sm(112 kbit/s)의 슈퍼와이드밴드 비트스트림을 제공한다. 또한, 제안된 코덱의 알고리즘 지연 및 복잡도는 ToR(Terms of Reference)의 해당 요구사항^[4]을 만족하였으며, 주관적인 청취 시험을 통해 ITU-T G.711.1/G.722 슈퍼와이드밴드 확장 코덱 품질 요구사항에 명시된 총 35 개의 시험 항목 중 34 개를 통과하였다. 이러한 결과를 기반으로 본 코덱은 2009년 7월에 개최된 ITU-T SG16 WP3 (Working Party 3) 회의에서 G.711.1/G.722 슈퍼와이드밴드 확장 코덱 표준의 자격 시험(qualification test)을 통과하였으며^[10], 현재 이를 기반으로 최적화/특성화 단계를 진행 중이다.

* FER: "Frame Error Rate"의 약어로 프레임 오류 발생율을 의미한다.

** CuT: "Codec under Test"의 약어로 시험 대상 코덱을 가리킨다.

참고 문헌

- [1] ITU-T Rec. G.711.1, "Wideband embedded extension for G.711 pulse code modulation," March 2008.
- [2] ITU-T Rec. G.722, "7 kHz audio coding within 64 kbit/s," November 1988.
- [3] ETSI TR 102 570, "Digital Enhanced Cordless Telecommunications (DECT); New Generation DECT; Overview and Requirements," March 2007.
- [4] ITU-T TD 109 (WP3/16) Annex Q10.D, "Report of Question 10/16 speech and audio coding and related software tools," November 2009.
- [5] J. Lapiere, R. Lefebvre, B. Bessette, V. Malenovsky, and R. Salami, "Noise shaping in an ITU-T G.711-interoperable embedded codec," in Proc. of EUSIPCO2008, Lausanne, Switzerland, August 2008.
- [6] ITU-T Rec. G.722.2, "Wideband coding of speech at around 16 kbit/s using Adaptive Multi-Rate Wideband (AMR-WB)," July 2003.
- [7] ITU-T Rec. G.722 Appendix IV, "A low complexity packet loss concealment algorithm with G.722," November 2006.
- [8] ITU-T TD 15 (WP1/12), "Superwideband extension to G.711.1 and G.722 qualification quality assessment testplan," March 2009.
- [9] ITU-R Rec. BS.1285, "A pre-selection methods for the subjective assessment of small impairments in audio systems," 1997.
- [10] ITU-T TD 48 (WP3/16), "A Report of Q10/16 Rapporteur Group Meeting (Geneva, 6-9 July, 2009)," 2009.
- [11] ITU-T Rec. G.722.1, "Low-complexity coding at 24 and 32 kbit/s for hands-free operation in systems with low frame loss," May 2005.

저 자 소 개



성 종 모(정회원)
 1995년 부산대학교 전자공학과
 학사 졸업.
 1997년 부산대학교 전자공학과
 석사 졸업.
 1997년~1999년 LG전자 연구원.
 1999년~현재 한국전자통신
 연구원 HD-VoIP연구팀
 선임연구원.

<주관심분야 : 신호처리, 음성/오디오 부호화>



김 현 우(정회원)
 2001년 서울대학교 전기공학부
 학사 졸업.
 2003년 서울대학교 전기컴퓨터
 공학부 석사 졸업.
 2003년~현재 한국전자통신
 연구원 HD-VoIP연구팀
 선임연구원.

<주관심분야 : 신호처리, 음성/오디오 부호화>



김 도 영(정회원)
 1985년 성균관대학교 전자공학과
 학사 졸업.
 1987년 성균관대학교 대학원 전자
 공학과 석사 졸업.
 2007년 충남대학교 대학원 컴퓨터
 공학과 박사 졸업.

1987년~현재 한국전자통신연구원 HD-VoIP연구
 팀장(책임연구원).

<주관심분야 : HD 인터넷전화(VoIP) 코덱 및 제
 어, 실시간 미디어 신호처리, 실감미디어 부호화>



이 병 선(정회원)
 1980년 성균관대학교 수학과 학사
 졸업.
 1982년 동국대학교 전산학과 석사
 졸업.
 2003년 한국과학기술원 전산학과
 박사 졸업.

1982년~2010년 한국전자통신연구원,
 現 서비스플랫폼연구부장.

<주관심분야 : 인터넷 서비스, 서비스 공학>



고 윤 호(정회원)-교신저자
 1995년 부산대학교 전자공학과
 학사 졸업.
 1997년 한국과학기술원 전기및
 전자공학과 졸업.
 2002년 한국과학기술원 전기및전
 자공학과 박사 졸업.

2002년~2004년 하이닉스반도체 선임연구원.

2004년~현재 충남대학교 메카트로닉스공학과
 조교수.

<주관심분야 : 신호처리, 컴퓨터비전, 영상인식,
 영상압축>