

# 심층 신뢰 신경망을 이용한 오푸스 코덱 기반 인공 음성 대역 확장 기술

## Artificial speech bandwidth extension technique based on opus codec using deep belief network

최윤상<sup>1</sup>, 이아성<sup>1</sup>, 강상원<sup>†</sup>

(Yoonsang Choi,<sup>1</sup> Yaxing Li,<sup>1</sup> and Sangwon Kang<sup>1†</sup>)

<sup>1</sup>한양대학교 전자통신공학과

(Received August 30, 2016; revised September 22, 2016; accepted January 25, 2017)

**초 록:** 대역폭 확장 기술은 300 ~ 3,400 Hz 대역의 협대역 음성 신호를 50 ~ 7,000 Hz 대역의 광대역 음성신호로 확장하여 음질, 명료도, 그리고 자연성을 높이는 기술이다. 본 논문에서는 협대역 음성 정보를 이용하여 광대역 음성신호를 추정하는 인공 대역폭 확장 기술을 설계하여, 오푸스(Opus) 오디오 복호화기에 내장시킴으로써, 대역폭 확장 모듈에서의 LPC(Linear Prediction Coding) 분석 및 LSF(Line Spectral Frequencies) 해석과 관련된 계산량을 감소시켰고 알고리즘 지연도 줄였다. 이를 위해 현재 다양한 분야에 적용되고 있는 딥 러닝 기술 중 하나인 심층 신뢰 신경망(Deep Belief Network, DBN) 방식을 스펙트럼 포락선 확장에 도입하여 전통적인 코드북 매핑법보다 더 좋은 품질의 스펙트럼을 만들 수 있었다.

**핵심용어:** 대역폭 확장, 음성 코덱, 음질 개선, 오푸스 코덱, 심층 신뢰 신경망

**ABSTRACT:** Bandwidth extension is a technique to improve speech quality, intelligibility and naturalness, extending from the 300 ~ 3,400 Hz narrowband speech to the 50 ~ 7,000 Hz wideband speech. In this paper, an Artificial Bandwidth Extension (ABE) module embedded in the Opus audio decoder is designed using the information of narrowband speech to reduce the computational complexity of LPC (Linear Prediction Coding) and LSF (Line Spectral Frequencies) analysis and the algorithm delay of the ABE module. We proposed a spectral envelope extension method using DBN (Deep Belief Network), one of deep learning techniques, and the proposed scheme produces better extended spectrum than the traditional codebook mapping method.

**Keywords:** Bandwidth extension, Speech codec, Speech enhancement, Opus codec, Deep belief network

**PACS numbers:** 43.72.Gy, 43.72.Ar

## 1. 서 론

협대역 음성 통화 서비스는 주파수 대역이 300 ~ 3.4 kHz로 제한된다.<sup>[1]</sup> 그럼에도 불구하고 협대역 음성 부호화기는 음성 통화자의 신원확인, 기본적인 청취 이해력 등을 제공할 수 있어 지난 수십 년간 주요 유무선 통신시스템에서 사용하여 왔다. 그러나

보다 발전된 유무선 통신 시스템이 개발되고 다양한 멀티미디어 통신 서비스가 제공되고 있는 요즘, 통신 사용자들은 더욱 높은 수준의 광대역 음성 통화 서비스를 요구하고 있다. 이러한 서비스는 기존의 협대역 음성 통화 서비스와 비교하여 높은 명료도, 자연스러운 음질 등을 제공한다.<sup>[2]</sup>

대역 확장은 스펙트럼 포락선 확장과 여기 신호 확장으로 나눌 수 있고,<sup>[3]</sup> 스펙트럼 포락선 확장은 다시 코드북 매핑 방식,<sup>[3]</sup> 은닉 마르코프 모델(Hidden Markov Model, HMM),<sup>[4]</sup> 가우시안 혼합 모델(Gaussian

<sup>†</sup>Corresponding author: Sangwon Kang (swkang@hanyang.ac.kr)  
Department of Electronics and Communication Engineering, Hanyang University, 55 hanyangdaehak-ro, Sangnok-gu, Ansan-si, Gyeonggi-do 15588, Republic of Korea  
(Tel: 82-31-400-4035, Fax: 82-31-418-8476)

Mixture Model, GMM) 방식<sup>[5]</sup> 등이 많이 연구되어 왔다. 본 연구에서는 최근 음성인식 등 많은 분야에 이용되고 있는 딥 러닝 기술 중 하나의 기법인 심층 신뢰 신경망(Deep Belief Network, DBN)<sup>[6]</sup>을 이용하여 기존 방법에 비하여 대역 확장 성능을 향상시켰고, 잡음이 포함된 훈련 데이터를 사용함으로써 다양한 잡음 환경에서도 좋은 품질의 음성을 출력할 수 있도록 하였다.

본 연구에서 개발한 대역 확장 시스템은 협대역 음성 복호화기 내부에 위치 시켜 협대역 음성 복호화기에서 분석된 계수를 이용함으로써 알고리즘 지연을 줄이고 계산량을 감소시키는 장점이 있다. 기존의 음성 대역 확장 방식인 DTE(Decode Then Extend)<sup>[7]</sup> 방식은 오디오 코덱의 복호화기에서 합성된 음성을 다시 분석하고, 분석된 파라미터를 이용하여 스펙트럼을 확장한다. 그러나 본 연구에서는 오픈스 복호화기에 대역 확장 시스템을 내장함으로써 기존의 DTE 방식에 비해 알고리즘 지연과 계산량을 줄였다.

본 논문은 II장에서 오픈스 코덱의 개요 및 코덱 내의 대역폭 조절 알고리즘에 대하여 설명한다. III장에서는 오픈스 복호화기에 내장된 인공 대역 확장 시스템의 전체 구조에 대하여 살펴보고, 스펙트럼 포락선확장에 사용된 DBN 방식과 네트워크 상태 변화에 따른 음질 열화를 줄이기 위해 개발된 적응형 대역폭 조절 알고리즘을 소개한다. IV장에서는 본 연구에서 개발된 대역 확장 시스템에 대한 실험 및 성능 분석 결과를 보이고, 마지막으로 V장에서 결론을 맺는다.

## II. 오픈스 코덱

### 2.1 오픈스 코덱 개요

오픈스 오디오 코덱은 국제 인터넷 표준화 기구에서 개발한 손실 오디오 압축 코덱으로서 2012년 7월에 표준화되었다.<sup>[8]</sup> 오픈스 코덱은 현재 로열티 없이 많은 어플리케이션에 사용되고 있으며, 특히 인터넷 상의 인터랙티브 실시간 응용 프로그램에 적합하게 만들어졌기 때문에 패킷망을 이용한 음성 서비스에 많이 사용되고 있다. 오픈스 코덱의 구조는 음성과

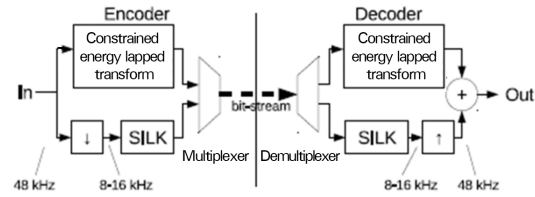


Fig. 1. The overall structure of Opus codec.

Table 1. Recommended bitrate ranges of Opus codec.

Input type	Recommended bitrate range (kbps)	
	Mono	Stereo
Fullband (FB)	28-40	48-72
Super-wideband (SWB)	20-28	36-48
Wideband (WB)	16-20	28-36
Mediumband (MB)	12-16	20-28
Narrowband (NB)	8-12	14-20

오디오 신호의 효율적인 압축을 위하여 Fig. 1과 같이 SILK 계층과 CELT(Constrained Energy Lapped Transform) 계층으로 각각 구성되어 있는데, SILK 계층은 주로 음성 신호를 부호화하기 위하여 LP(Linear Prediction)을 기반으로 하고, CELT 계층은 오디오 신호를 부호화 하기 위하여 MDCT(Modified Discrete Cosine Transform)를 기반으로 한다.<sup>[3]</sup>

오픈스는 6 kbps에서 510 kbps의 광범위한 비트 전송율을 지원하며,<sup>[8]</sup> 입력 오디오 신호의 대역폭에 따라 권장되는 비트 전송율은 Table 1과 같다.<sup>[9]</sup> 오픈스 코덱에서는 비트 전송율과 상관없이 별도의 옵션으로 입력 오디오 신호를 부호화하기 위한 코덱 내부 샘플링 주파수 또는 최대 대역폭을 설정할 수 있는데, 예를 들어 8 kbps 전송율에서도 설정에 따라 광대역 신호로 부호화 및 복호화할 수 있다. 그러나 비트 전송율이 12 kbps 미만인 경우 오픈스 코덱의 광대역 출력은 음질 열화가 많이 발생하므로, 본 연구에서는 이를 해소하기 위해 12 kbps 미만의 비트 전송율에서는 협대역 코덱으로 부호화하고 결과적인 협대역 출력을 인공 대역 확장해서 광대역 출력을 생성하는 알고리즘을 도입하였다.

### 2.2 오픈스 코덱 대역폭 조절 알고리즘

인터넷을 이용한 음성통화 시, 네트워크 상태에 따라 할당되는 비트 전송율을 달리 함으로써 출력되

는 음성 품질을 조절할 수 있다. 별도의 옵션을 선택하지 않는 경우, 오픈스 코덱의 내부 샘플링 주파수와 출력의 최대 대역폭은 비트 전송율에 따라 결정되며 비트 전송율이 높아질수록 협대역에서 전대역까지의 신호를 출력한다.

오픈스 코덱은 네트워크 상태가 변화할 가능성에 대비하여, 비트 전송율이 급변하는 경우 출력 음성의 최대 대역폭이 매끄럽게 변화하도록 하는 대역폭 조절 알고리즘을 제공한다. Fig. 2는 네트워크 상태의 변화를 고려하여 여러 비트 전송율로 음성신호를 부호화 하였을 때의 출력신호를 나타내었다. 이 때 비트 전송율은 실험적으로 8, 12, 16, 24 kbps를 사용하였으며, 별도의 옵션이 없는 경우 각 비트 전송율에 따른 오픈스 코덱 내부 샘플링 주파수와 출력신호의 최대 대역폭을 Table 2에 나타내었다.

Fig. 2(a)와 (b)는 네트워크 상태에 따른 오픈스 코덱의 출력 스펙트로그램을 나타낸 것이다. 먼저 Fig.

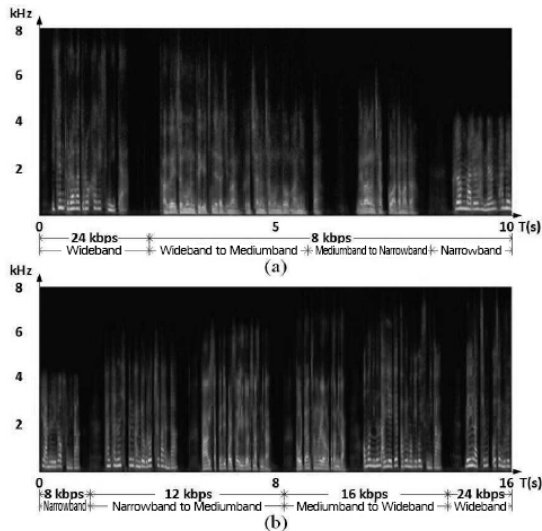


Fig. 2. The output spectrogram of Opus codec for network conditions (a) in network congestion (b) in recovery from congestion.

Table 2. Internal sampling rate and maximum bandwidth of output signal of Opus codec for bitrate.

Bitrate	Internal sampling rate	Maximum bandwidth
8 kbps	8 kHz	4 kHz (NB)
12 kbps	12 kHz	6 kHz (MB)
16 kbps	16 kHz	8 kHz (WB)
24 kbps	16 kHz	8 kHz (WB)

2(a)는 네트워크 상태가 급격히 악화될 경우를 가정하여 비트 전송율을 24 kbps에서 8 kbps로 낮추었을 때의 출력 음성신호이다. 전송율이 줄어든 시점을 시작으로 오픈스 코덱의 내부 샘플링 주파수는 16 kHz로부터 12 kHz, 8 kHz로 단계적으로 감소하며 출력 음성의 대역폭이 8 kHz에서 4 kHz로 서서히 줄어드는 천이구간이 존재함을 알 수 있다.

Fig. 2(b)는 네트워크 상태가 서서히 회복될 경우를 가정하여 비트 전송율을 8 kbps에서 12, 16, 그리고 24 kbps까지 서서히 올렸을 때의 출력 음성을 나타낸 것이다. 이 경우에는 Fig. 2(a)와는 반대로 전송율이 증가한 시점을 기점으로 오픈스 코덱의 내부 샘플링 주파수는 8 kHz로부터 12 kHz, 16 kHz로 단계적으로 증가하며 출력 음성의 대역폭이 4 kHz에서 8 kHz로 서서히 증가함을 알 수 있다.

### III. 임베디드 대역 확장 시스템

#### 3.1 대역 확장 시스템 구조

오픈스 복호화기에 내장된 대역 확장 시스템 구조는 Fig. 3과 같다. 대역확장 시스템은 크게 스펙트럼 포락선 확장과 여기 신호 확장으로 나누어진다. 스펙트럼 포락선 확장에서는, 오픈스 복호화기에서 10차 LSF(Line Spectral Frequencies) 계수  $^{nb}f(n)$ 를 복호화한 후 DBN 방식을 이용하여 최적의 16차 광대역 LSF 계수  $^{wb}f(n)$ 를 구하고, 이를 16차 LPC 계수  $^{wb}a$ 로 변환한다.

여기 신호 확장의 수행과정은 다음과 같다. 오픈스 복호화기에서 복호화된 협대역 여기 신호,  $^{nb}e(n)$ 에 Eq. (1)과 같이 6차 LPC 분석을 수행함으로써 여기 신호의 백색화 과정을 수행한다. 이러한 과정을 통해 평평한 스펙트럼을 갖는 백색화된 협대역 여기 신호,  $^{nb}e'(n)$ 가 생성된다.

$$^{nb}e'(n) = ^{nb}e(n) - \sum_{j=1}^6 a'_j ^{nb}e(n-j). \quad (1)$$

두 번의 LPC 분석을 통해 얻어진 평평한 협대역 여기 스펙트럼을 고 대역으로 확장함으로써 정확하고 대역 여기 스펙트럼을 얻을 수 있다. 즉 한 번의 LPC

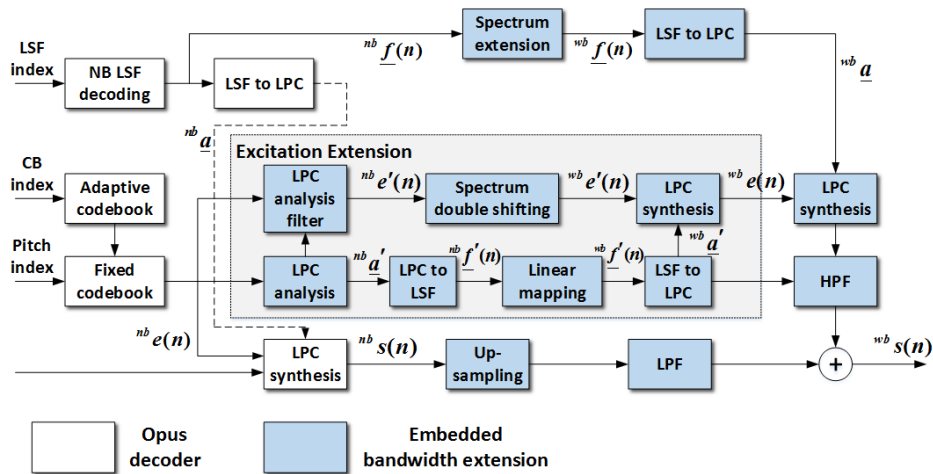


Fig. 3. The block diagram of proposed embedded bandwidth extension scheme.

분석 필터를 통과한 협대역 여기 스펙트럼은 완전히 평평하지 않고 주파수가 높아짐에 따라 아주 완만한 내리막 경사를 갖기 때문에 이를 고 대역으로 확장하면 실제 고 대역 신호보다 큰 값을 갖게 된다.

위에서 구한 백색화된 협대역 여기 신호를 피치적응 스펙트럼 더블 시프팅 방식을 이용하여 고 대역 여기 신호를 생성하고 앞서 추출된 두 번째 LPC 계수  $^{nb}a'$ 를 선형 매핑을 통해 광대역 여기 신호 LPC 계수  $^{wb}a'$ 로 변환한 뒤 합성필터를 통해 광대역 여기 신호를 생성한다.

이렇게 생성된 광대역 여기 신호를 DBN 방식을 통해 얻은 광대역 LPC 합성 필터에 입력하여 광대역 신호를 합성하고, 합성된 광대역 신호를 4.0 kHz의 차단 주파수를 갖는 고 대역 통과 필터에 통과시키고 고 대역의 대역확장 성분(4.0 ~ 8.0 kHz)을 추출한 후 오픈스 복호화기에서 합성된 협대역 출력신호에 더하면 최종 광대역 신호  $^{wb}s(n)$ 가 생성된다. 즉 4.0 kHz 이하의 저 대역 성분은 원래의 협대역 음성출력을 사용하고, 4.0 kHz 이상의 고 대역 성분은 확장 알고리즘을 통하여 추정된 신호를 사용한다. 다음 절에서는 스펙트럼 포락선 확장을 위하여 본 연구에서 사용한 DBN 방식에 대하여 소개한다.

### 3.2 심층 신뢰 신경망을 이용한 스펙트럼 포락선 확장

DBN은 2006년 Geoffrey Hinton 교수에 의해 소개된

알고리즘으로 크게 선행학습 과정과 미세튜닝 과정으로 이루어져 있다. 먼저 제한된 볼츠만 머신 (Restricted Boltzmann Machine, RBM)을 통해 선행학습이 이루어지는데, 미세튜닝을 위한 초기 가중치 값을 설정해 줌으로써 전체적인 학습이 잘 이루어지게 한다. RBM은 기존의 볼츠만 머신(BM)에서 같은 계층 내 노드들 간의 연결강도를 0으로 제한하여 각 노드들 간 관계를 독립시킨 것이다.<sup>[10]</sup> DBN은 Fig. 4와 같이 RBM이 차례대로 연결된 것으로, 아래 단의 은닉 계층이 위단의 가시 계층 역할을 한다.<sup>[11]</sup> Fig. 4를 통해 DBN 학습 방식을 살펴보면, 먼저 가시 계층과 은닉 계층 1을 하나의 RBM으로 보고 학습시킨다. 이 학습이 끝나면 은닉 계층 1의 값을 입력으로 하여 은닉 계층 1과 2로 이루어진 RBM을 학습한다. 이처럼 단계적으로 올라가면서 마지막 층까지 학습을 한다. 본 연구에서는 2개의 은닉 계층을 가지는 DBN 구조를 사용하였다.

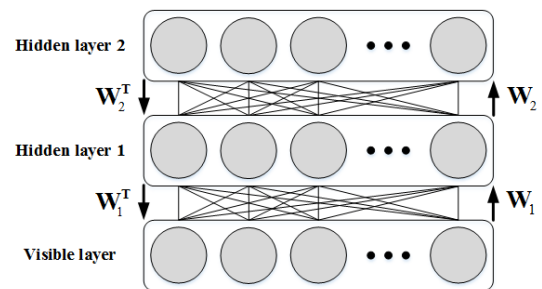


Fig. 4. Deep belief network structure.

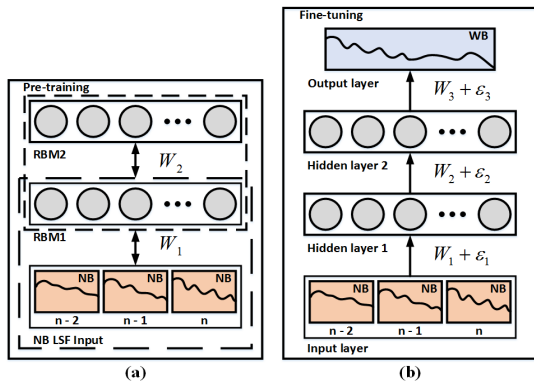


Fig. 5. Deep belief network structure and training (a) pre-training (b) fine-tuning.

본 연구에서는 스펙트럼 포락선 확장에 DBN 방식을 응용하였는데 이를 구체적으로 살펴보면, 먼저 선행학습을 위하여 Fig. 5(a)와 같이 10차 협대역 LSF 벡터를 입력으로 하여 두 계층으로 이루어진 RBM을 학습한다. 이 때 협대역 LSF 벡터는 현재 프레임과 두 개의 이전 프레임을 사용하였는데, 이와 같이 과거 정보를 함께 이용함으로써 현재 프레임만 사용하는 것보다 향상된 학습결과를 얻을 수 있다. 선행학습이 끝나면 미세튜닝 과정을 거치게 되는데, LSF 확장을 위한 미세튜닝 과정은 Fig. 5(b)와 같다. 미세튜닝 과정에서는 예측된 16차 광대역 LSF 계수와 원음성의 광대역 LSF 계수 사이의 유클리디안(Euclidean) 거리를 최소화 하는 최적의 광대역 LSF 계수를 찾기 위하여 MMSE(Minimum Mean Square Error) 기법을 사용하였다.

### 3.3 대역 확장을 고려한 오푸스 코덱 대역폭 조절 알고리즘

II장에서 오푸스 코덱에서 제공하는 대역폭 조절 알고리즘에 대하여 설명하였다. 본 연구에서 개발한 대역 확장 시스템은 내부 샘플링 주파수가 8 kHz인 협대역 출력 신호를 대상으로 하였기 때문에, 이를 오푸스 코덱에 그대로 적용하는 경우 Fig. 6에 보여진 바와 같이 문제가 발생한다. 즉 네트워크 상태가 변화할 때 출력신호는 대역이 바뀌는 천이구간 또는 중간대역(MB)에서 대역 확장이 적용되지 않기 때문에, Fig. 6에서 대역이 확장되는 구간(3)을 전후로 대역 간극이 발생하는 것을 확인할 수 있다. 사람의 귀

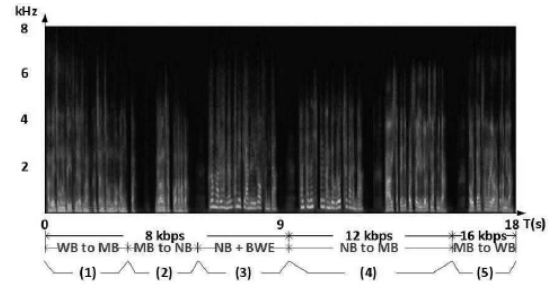


Fig. 6. The output spectrogram of proposed bandwidth extension system for Opus codec with network condition.

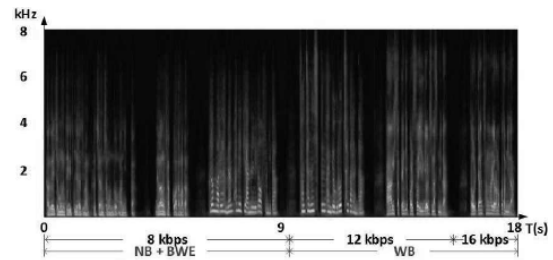


Fig. 7. The output spectrogram of proposed bandwidth extension system for Opus codec with network condition after modifying bandwidth control algorithm.

는 이러한 대역간극을 인지할 수 있으므로 음성 품질의 저하를 초래하게 된다. 따라서 본 연구에서는 대역간극이 발생되지 않도록 오푸스 코덱의 대역폭 조절 알고리즘을 다음과 같이 수정하였다.

중간대역(MB) 출력 신호의 경우는 대역 확장을 하지 않음으로서 협대역 신호에 대역 확장 시스템을 적용하였을 경우 중간대역과 광대역 간의 대역 간극이 초래된다. 비트 전송율이 12 kbps 이상일 경우 오푸스 코덱의 광대역 출력의 품질은 우수하기 때문에 대역확장을 위한 기준 전송율을 12 kbps로 하였다. 그래서 12 kbps 미만의 비트 전송율을 사용하는 경우에는 협대역 신호를 출력해서 대역확장 시스템을 적용하였고, 12 kbps 이상일 경우에는 광대역 신호를 출력하도록 함으로서 대역 간극이 없도록 하였다.

Fig. 7은 네트워크 상태가 변화하는 경우 대역폭 조절 알고리즘이 수정된 오푸스 코덱에 대역 확장 시스템을 적용한 결과를 나타내고 있다. Fig. 7에서 확인할 수 있듯이 전송율 8 kbps 구간에서는 오푸스 협대역 출력을 이용하여 대역을 확장하고, 12 kbps 이상 구간에서는 오푸스 코덱 내에서 광대역 신호를 출력함으로써, 전 구간에서 광대역 신호가 출력

된다. 따라서 Fig. 6과 달리 대역 간극이 없음을 알 수 있다.

### IV. 실험 및 결과

제안된 대역 확장 시스템의 성능 평가를 위하여 16 kHz로 샘플링 된 TIMIT 데이터베이스를 이용하였다. 전체 4,200 문장 중 약 3시간 20분에 해당하는 3,800 문장은 훈련에 이용되었고 나머지 약 20분에 해당하는 400 문장은 실험에 이용되었다. 또한 잡음 환경에서의 성능 실험을 위하여 핑크 잡음과 다중주 화잡음(babble noise)을 이용하였다.

Table 3은 스펙트럼 확장으로 코드북 매핑 방식을 이용하는 대역확장 시스템과 본 논문에서 제안하는 DBN을 이용한 대역 확장 시스템의 성능 차이를 비교한 것이다. 성능 비교에는 LSD(Log Spectral Distance) 방식을 사용하였고 Eq. (2)와 같이 N-점 퓨리에 변환을 이용하여 계산하였다. 본 연구에서는 N = 256을 사용하였다.

$$LSD = \sqrt{\frac{1}{N/2} \sum_{k=0}^{N/2-1} [10\log_{10}|A'(k)|^2 - 10\log_{10}|A(k)|^2]^2}, \quad (2)$$

$$S(w) = 1/|A(w)|^2, S'(w) = 1/|A'(w)|^2, \quad (3)$$

여기서  $S(w)$  및  $S'(w)$ 는 원래의 LSF 및 양자화된 LSF로부터 유도된 음성신호의 주파수 응답이다.

Table 3. Comparison of log spectral distance performance between codebook mapping and proposed deep belief network technique for spectrum extension.

	Codebook mapping		DBN
	L = 256	L = 512	
Full-band LSD (dB)	4.901	4.706	4.102
High-band LSD (dB)	4.981	4.915	4.621

Table 4. Comparison of complexity between codebook mapping and proposed deep belief network technique for spectrum extension.

	Codebook mapping		DBN
	L = 256	L = 512	
WMOPS	0.438	0.874	0.729

전체대역의 경우  $\Delta F$ 는 0 ~ 8.0 kHz에 해당하며, 고대역의 경우  $\Delta F$ 는 4.0 ~ 8.0 kHz에 해당한다. Table 3에서 확인할 수 있듯이, 스펙트럼 확장 시 코드북 길이(L)가 256 및 512인 코드북 매핑 방식에 비하여 DBN을 이용한 대역확장 방식을 이용하였을 때 대역 확장 시스템의 성능이 전체대역 LSD 기준으로 0.799 및 0.604 dB 만큼 향상되었고, 고 대역 LSD 기준으로 0.360 및 0.294 dB 향상되었다.

Table 4는 스펙트럼 확장으로 코드북 매핑 방식을 이용하는 대역확장 시스템과 본 논문에서 제안하는 DBN을 이용한 대역 확장 시스템의 계산량을 비교하였다. 계산량 비교에는 오디오 코덱 연산량 측정에 일반적으로 사용되는 WMOPS(Weighed Million Operations Per Second) 방식을 이용하였으며, 이 방식은 ITU-T G.191에 정의되어있다.<sup>[12]</sup> 코드북 길이가 512인 코드북 매핑 방식과 비교 시 DBN 방식의 WMOPS가 0.145(16.6%) 만큼 감소하였다.

본 연구에서는 기존 오픈스 코덱과 제안된 대역확장 시스템이 내장된 오픈스 코덱에 대한 주관적인 음질 성능을 비교하는 CCR(Comparison Category Rating) 테스트를 수행하였다. CCR 테스트는 A와 B, 두 개의 음성을 무작위로 번갈아 듣고 선호하는 음성을 선택하는 방법인데, 두 개 음성에 대해 매우 정밀한 음성 성능 측정이 가능하다. CCR 측정 기준을 Table 5에 나타내었다.

본 연구에서는 두 가지 CCR 테스트를 수행하였으며 그 결과를 Fig. 8(a)과 (b)에 각각 나타내었다. Fig. 8(a)은 Opus@8 kbps 협대역 출력과 이 협대역 출력을 이용하여 확장된 광대역 출력의 음질을 비교한 결과이고, Fig. 8(b)은 네트워크 상태 변화를 가정하여 코덱 내의 비트 전송율을 변화시키면서 기존 오픈스

Table 5. Comparison category rating scale.

CCR score	Performance
3	A is much better than B
2	A is better than B
1	A is slightly better than B
0	A is same as B
-1	A is slightly worse than B
-2	A is worse than B
-3	A is much worse than B

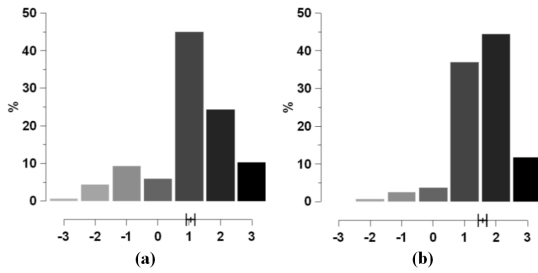


Fig. 8. Comparison category rating test results for proposed bandwidth extension system (a) Opus@8 kbps narrow band vs. Opus@8 kbps narrow band + proposed artificial bandwidth extension (b) output of Opus codec vs. Opus codec + proposed ABE for network condition.

코덱의 성능과 제안된 대역확장 시스템이 내장된 오픈스 코덱의 성능을 비교한 것이다. 이 실험에는 9명의 한국 남성과 여성의 음성 샘플들이 각각 사용되었으며, 청취 실험에는 10명의 한국 학생이 참여하였다.

Fig. 8(a)에서 알 수 있듯이, 대역 확장 음성을 선호하는 비율이 약 80%, 협대역 음성을 선호하는 비율이 약 14%, 두 가지 음성의 차이를 느끼지 못한 비율이 약 6%였으며, CCR 테스트 평균 점수는 1.05로 측정되었다. Fig. 8(b)에서는, 제안된 대역확장 시스템이 내장된 오픈스 코덱을 선호하는 비율이 약 93%, 기존 오픈스 코덱을 선호하는 비율이 약 3%, 두 가지 코덱의 차이를 느끼지 못한 비율이 약 4%였으며, CCR 테스트 평균 점수는 1.57로 측정되었다. 두 가지 CCR 테스트에서 선호도와 CCR 테스트 평균 점수를 근거로 분석해보면 앞서 객관적 평가 결과와 마찬가지로 제안된 대역 확장 시스템이 내장된 오픈스 코덱이 기존 오픈스 코덱에 비하여 우수한 성능을 나타냄을 알 수 있다.

## V. 결론

본 논문에서는 DBN 방식으로 스펙트럼을 확장하는 인공 대역확장 시스템을 설계하고, 이를 국제 인터넷 표준화 기구에서 개발된 오픈스 오디오 코덱의 디코더 내부에 장착한 “오픈스 코덱 기반 인공 대역 확장 시스템”을 제안 하였다. 제안된 시스템은 코드북(L = 512) 매핑 방식을 이용하는 기존 시스템 보다

고 대역 LSD 기준 0.294 dB 만큼 성능을 향상시켰고, 계산량은 16.6% 정도 감소시켰다.

CCR 테스트에서는 8 kbps 오픈스 코덱의 협대역 출력과 제안된 대역확장 시스템이 내장된 8 kbps 오픈스 코덱의 확장된 광대역 출력의 성능 비교 시, 대역 확장 음성을 선호하는 비율이 약 80%, 협대역 음성을 선호하는 비율이 약 14%, 두 가지 음성의 차이를 느끼지 못한 비율이 약 6%이며, CCR 테스트 평균 점수는 1.05로 측정되었다. 네트워크 상태를 고려한 시스템 성능 비교에서는, 제안된 대역확장 시스템이 내장된 오픈스 코덱을 선호하는 비율이 약 93%, 기존 오픈스 코덱을 선호하는 비율이 약 3%, 두 가지 시스템의 차이를 느끼지 못한 비율이 약 4%이며, CCR 테스트 평균 점수는 1.57로 측정되었다.

## 감사의 글

이 연구는 방위사업청 및 국방과학연구소의 재원에 의해 설립된 신호정보 특화연구센터 사업의 지원을 받아 수행되었음.

## References

1. ITU-T Recommendation, G.712, *Performance characteristics of PCM channels between 4-wire interfaces at voice frequencies*, 1988.
2. P. Noll, “Wideband speech and audio coding,” *IEEE Communications Magazine*, **31**, 34-44 (1993).
3. U. Kornagel, “Techniques for artificial bandwidth extension of telephone speech,” *Signal Processing*, **86**, 1276-1306 (2006).
4. P. Jax and P. Vary, “On artificial bandwidth extension of telephone speech,” *Signal Processing*, **83**, 1707-1719 (2003).
5. K. Park and H. Kim, “Narrowband to wideband conversion of speech using GMM based transformation,” in *Proc. ICASSP*, 1843-1846 (2000).
6. G. Hinton, S. Osindero, and Y. Teh, “A fast learning algorithm for deep belief nets,” *Neural Computation*, **19**, 1527-1554 (2006).
7. Jax, Peter and Peter Vary, “Bandwidth extension of speech signals: A catalyst for the introduction of wideband speech coding?” *IEEE Communications Magazine*, **44**, 106-111 (2006).
8. *Definition of the Opus Audio Codec*, RFC 6716,

<http://www.ietf.org/rfc/rfc6716.txt>, 2012.

9. K. Vos, K. Sorensen, S. Jensen, and J.-M. Valin "Voice coding with opus," in proc. the AES 135<sup>th</sup> Convention (2013).
10. G. Hinton, "A practical guide to training restricted boltzmann machines," Toronto Univ., Tech. Rep. 6(1), 2010.
11. G. Hinton and R. Salakhutdinov, "Reducing the dimensionality of data with neural networks," *Science*, **313**, 504-507 (2006).
12. ITU-T Recommendation, G.191, *Software tools for speech and audio coding standardization*, 2010.
13. K. Li and C.-H. Lee, "A deep neural network approach to speech bandwidth expansion," in Proc. ICASSP, 4395-4399 (2015).

### 저자 약력

▶ 최 윤 상 (Yoonsang Choi)



2006년 ~ 2013년: 한양대학교 전자통신공학과(공학사)  
 2014년 ~ 2016년: 한양대학교 전자통신공학과(공학석사)

▶ 이 아 성 (Yaxing Li)



2007년 ~ 2011년: 하얼빈 공업대학 전자공학과(공학사)  
 2011년 ~ 현재: 한양대학교 전자통신공학과 석박사통합과정

▶ 강 상 원 (Sangwon Kang)



1976년 ~ 1980년: 한양대학교 전자공학과(공학사)  
 1980년 ~ 1982년: 서울대학교 전자공학과(공학석사)  
 1985년 ~ 1990년: Texas A&M 대학교 전기공학과(공학박사)  
 1982년 ~ 1994년: 한국전자통신연구원 신호처리 연구실  
 1994년 ~ 현재: 한양대학교 전자공학부 교수