

논문 2013-50-12-23

하모닉 구조 확장과 NMF 기반의 인공 대역 확장 기술

(Artificial Bandwidth Extension Based on
Harmonic Structure Extension and NMF)

김 기준*, 박 호 중**

(Kijun Kim and Hochong Park[©])

요 약

본 논문에서는 주파수 영역에서 협대역 신호를 광대역으로 확장하는 새로운 인공 대역 확장 기술을 제안한다. 제안한 기술은 협대역 신호를 여기 신호와 스펙트럼 포락선 성분으로 분리하고, 주파수 영역에서 각각 독립적인 방법으로 확장한다. 여기 신호는 저대역의 하모닉 구조가 고대역에서 유지되도록 확장하고, 스펙트럼 포락선은 부대역별 에너지를 기반으로 NMF 방법으로 확장한다. 마지막으로 시간 축에서 프레임 사이의 상관관계를 기반으로 스펙트럼 위상을 결정하여 최종 광대역 신호를 생성한다. 주관적 청취 평가를 통하여 제안한 방법으로 대역 확장된 신호가 원 협대역 신호보다 음질이 향상된 것을 확인하였다.

Abstract

In this paper, we propose a new method for artificial bandwidth extension of narrow-band signal in frequency domain. In the proposed method, a narrow-band signal is decomposed into excitation signal and spectral envelope, which are extended independently in frequency domain. The excitation signal is extended such that low-band harmonic structure is maintained in high band, and the spectral envelope is extended based on sub-band energy using NMF. Finally, the spectral phase is determined based on signal correlation between frames in time domain, resulting in the final wide-band signal. The subjective evaluation verified that the wide-band signal generated by the proposed method has a higher quality than the original narrow-band signal.

Keywords: 인공 대역 확장(artificial bandwidth extension), NMF(non-negative matrix factorization), 하모닉 구조(harmonic structure), 스펙트럼 포락선(spectral envelope), 위상(phase)

I. 서 론

기존의 협대역 (0.3~3.4kHz) 음성 통신 서비스는 고대역 정보의 손실로 인하여 고품질의 서비스를 제공하

는데 한계를 가진다. 이를 해결하기 위하여 최근 광대역 (0.05~7kHz) 음성 통신 서비스가 시도되고 이를 위한 광대역 음성 부호화기가 개발되었으며, 초광대역 (0.05~14kHz) 통신을 위한 부호화기도 개발되었다^[1]. 광대역 및 초광대역 음성 서비스는 협대역 서비스 보다 음성의 명료성과 자연성 등에서 뛰어난 품질을 제공한다. 그러나 새로운 서비스를 제공하기 위해 통신 네트워크 변경이 필요하고 전송 데이터가 증가하므로, 기존 통신 서비스를 일시에 광대역으로 변환하는 것은 한계가 있다. 따라서 동일한 데이터 전송량을 사용하고 네트워크 변경 없이 광대역 음성 서비스를 제공하기 위한

* 학생회원, ** 정회원, 광운대학교 전자공학과
(Dept. of Electronics Engineering, Kwangwoon University)

© Corresponding Author(E-mail: hcpark@kw.ac.kr)

※ 본 연구는 미래창조과학부 및 정보통신산업진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음(NIPA-2013-H0301-13-4007).

접수일자: 2013년10월16일, 수정완료일: 2013년12월2일

방법으로 인공 대역 확장 (artificial bandwidth extension, ABE)을 사용할 수 있다^[2].

인공 대역 확장은 추가적인 정보의 전송 없이 주어진 협대역 신호로부터 고대역 정보를 추정하여 광대역 신호를 합성하는 기술이다. 인공 대역 확장 기술은 소스 필터 모델 (source-filter model)을 사용하여 구현할 수 있다^[2]. 이 방법에서 협대역 음성 신호는 선형 예측 분석에 의하여 여기 신호 (excitation signal)와 스펙트럼 포락선 (envelope)으로 모델링 되며, 각각의 고대역 성분을 추정하여 대역을 확장하고, 확장된 여기 신호와 스펙트럼 포락선을 이용하여 최종 광대역 음성을 합성한다.

여기 신호의 대역을 확장하는 방법으로 주파수 영역에서의 스펙트럼 폴딩 (spectral folding) 기법이 제안되었다^[3]. 그러나 이 방법은 고대역의 하모닉 구조를 원본과 유사하게 생성하지 못하므로 복원된 신호의 음성 품질을 저하시키는 문제를 가진다.

스펙트럼 포락선의 대역 확장 방법으로 HMM (hidden Markov model) 또는 GMM (Gaussian mixture model) 등과 같은 확률 모델 기반 방법이 개발되었다^[4, 5]. 이 방법들은 신호의 여러 특성을 모델링하기 위하여 다수 모델을 사용하고, 각 프레임 신호의 특성에 해당하는 모델을 선택한다. 그러나 음성 신호의 특성이 모호하거나 여러 특성이 혼합된 경우가 자주 발생하므로, 만일 특성 분류가 잘못되어 잘못된 모델을 선택하면 심각한 성능 저하가 발생할 수 있다. 또한, 최종 광대역 합성 과정에서 스펙트럼 위상에 대한 조정 과정이 없으므로 잘못된 위상에 의한 신호 왜곡이 발생하여 성능이 저하되는 요인으로 작용한다.

본 논문에서는 이와 같은 기존 인공 대역 확장 기술의 문제점을 해결하고 우수한 품질의 고대역 신호 합성을 위하여 하모닉 구조를 유지하는 여기 신호 확장 기술, 부대역 (sub-band)별 에너지를 기반으로 NMF (non-negative matrix factorization) 방법을 사용하는 스펙트럼 포락선 확장 기술, 그리고 시간 영역에서 프레임 사이의 상관관계를 기반으로 위상을 결정하는 기술을 포함하는 새로운 인공 대역 확장 기술을 제안한다. 제안하는 여기 신호 확장 기술은 저대역 하모닉 구조가 고대역으로 확장되도록 하여 대역 확장 성능을 향상시키며, NMF 방법으로 고대역 에너지를 추정하여 특성 분류 없이 하나의 기저 행렬로 포락선을 확장한

다. 위상은 중첩된 프레임 영역에서의 상관관계를 높이도록 결정하여 합성된 신호의 왜곡을 감소시킨다. 주관적 청취 평가를 통하여 제안한 방법으로 확장된 신호의 품질이 협대역 신호보다 우수한 것을 확인하였다.

II. 제안한 인공 대역 확장 기술

1. 개요

그림 1은 본 논문에서 제안하는 인공 대역 확장 기술의 전체 구조를 나타낸다. 입력되는 협대역 신호는 샘플링 주파수가 8kHz이고 AMR-NB (adaptive multi-rate narrow-band) 음성 부호화기로 부호화 및 복호화 된 신호이다. 대역 확장 기술은 통신 단말기의 최종 출력단에 적용되고, 현재 가장 대표적인 협대역 통신 부호화기가 AMR-NB이므로, AMR-NB로 복원된 신호를 입력으로 사용한다. 전체 대역 확장 동작은 50% 중첩 (overlap) 되는 프레임 단위로 수행되며, 프레임 길이 20ms와 프레임율 (frame rate) 10ms를 사용한다.

협대역 입력 신호는 선형 예측기를 사용하여 여기 신호와 16차 LPC 계수로 분리하고, 각각에 256-point DFT (discrete Fourier transform)를 적용하여 각각의 스펙트럼을 구하고, 각 스펙트럼을 주파수 영역에서 독립적인 방법으로 고대역으로 확장한다. 여기서, LPC 계

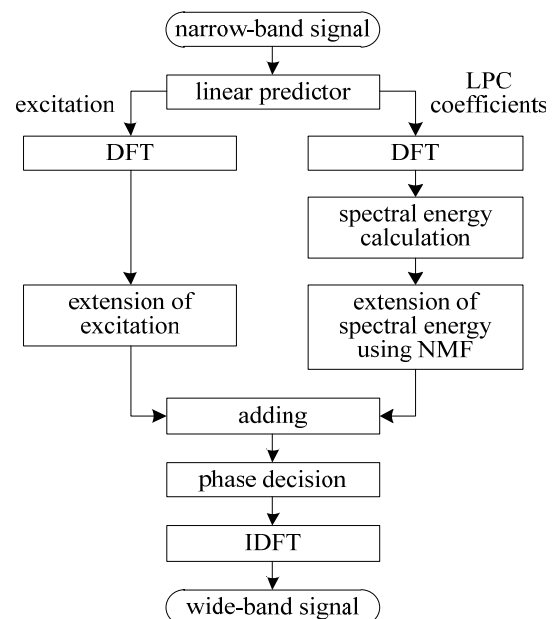


그림 1. 제안한 인공 대역 확장 기술의 전체 구조
Fig. 1. Overall structure of proposed artificial bandwidth extension method.

수의 스펙트럼이 입력 신호의 스펙트럼 포락선에 해당한다. 고대역의 여기 신호와 스펙트럼 포락선이 결정되면 최종적으로 고대역 위상을 결정하여 광대역 신호를 생성한다.

본 논문에서 제안하는 인공 대역 확장 기술의 여기 신호와 스펙트럼 포락선의 고대역 확장 방법, 그리고 고대역 위상 결정 방법은 각각 2절, 3절, 4절에서 자세히 설명한다.

2. 여기 신호의 고대역 확장

하모닉 구조는 음질에 큰 영향을 미치므로 여기 신호의 고대역 확장은 저대역의 하모닉 구조가 고대역으로 불연속 없이 확장되는 것을 목표로 한다^[6]. 하모닉 구조는 주파수 영역에서 주기적인 스펙트럼 피크 형태로 나타난다. 따라서 고대역에 동일한 주기로 유사한 모양의 스펙트럼 피크를 생성하면 되고, 저대역과 고대역 경계에 불연속이 없어야 한다. 이와 같은 배경에 따라 본 논문에서는 상관관계 기반으로 스펙트럼 중첩-가산을 사용하는 방법을 제안하며, 그림 2를 기반으로 상세한 동작을 설명하면 다음과 같다.

그림 2의 첫 그림에서, 저대역 여기 신호 스펙트럼의 마지막 구간에 $N/2$ 의 길이를 가지는 구간 A 를 설정한다. 본 논문에서 N 은 1kHz로 설정하였으며 이 값은 실험에 의한 값이므로 변경할 수 있다. 다음, 나머지 스펙트럼 영역에서 A 와 크기 상관도 (magnitude correlation)가 가장 높은 구간을 검색하고, 검색된 구간의 시작 지점에서 N 의 길이를 가지는 구간 B 를 설정한다. 그리고 그림과 같이 B 의 첫 $0 \sim N/2$ 구간과 A 에 윈도우를 적용하고 중첩-가산을 수행하여 스펙트럼 길이를 $N/2$ 만큼 증가시킨다(빗금 친 영역). A 와 B 가 높은 상관도를 가지므로 B 는 A 와 유사한 스펙트럼 모양을 가지고, 따라서 빗금 친 영역은 A 와 유사한 하모닉 구조를 가지며 불연속 없이 연결된다.

위에서 확장된 $N/2$ 영역을 다시 새로운 A 로 설정하고, A 와 가장 높은 상관도를 가지는 B 를 찾고, 위와 동일한 방법으로 중첩-가산을 수행하여 추가로 $N/2$ 영역 만큼 길이를 확장한다. 이렇게 확장된 영역은 역시 이전 스펙트럼과 불연속 없이 유사한 하모닉 구조를 가지게 된다. 이상의 과정을 확장된 구간이 7kHz가 될 때까지 계속 반복하여 확장된 여기 신호의 스펙트럼을 구한다. 이 방법을 사용하면 피치 주파수를 검색하지 않고

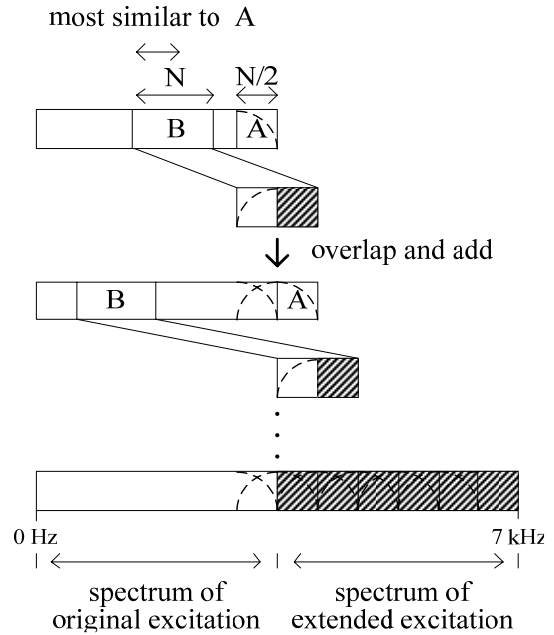


그림 2. 스펙트럼 중첩-가산에 의한 여기 신호의 고대역 확장 방법

Fig. 2. High-band extension of excitation signal by spectrum overlap-and-add.

자동으로 해당 피치 주파수를 가지는 고대역 스펙트럼을 생성할 수 있다.

하모닉 구조가 없는 무성음 (unvoiced) 신호일 경우, 스펙트럼은 주기성이 없는 불규칙 모양을 가진다. 이 경우, 위 방법에 따라 높은 상관도를 가지는 B 영역을 사용하여 대역을 확장하면 불규칙 모양이 자동적으로 고대역으로 확장되므로 고대역에서도 무성음 스펙트럼이 얻어진다. 따라서 제안하는 방법을 사용하면 유성음/무성음 구분이 필요 없고, 모든 신호에 동일한 방법을 적용하여 원하는 성질의 스펙트럼을 고대역으로 확장할 수 있다.

3. 스펙트럼 포락선의 고대역 확장

본 논문에서 제안하는 스펙트럼 포락선의 고대역 확장은 NMF 방법을 사용한다^[7, 8]. 기존의 NMF 방법은 고대역의 포락선을 DFT 주파수 단위로 확장하지만 본 논문에서는 부대역 단위로 NMF를 수행하여 고대역을 확장하는 방법을 제안한다. 제안한 방법을 사용하면 기존의 방법에 비하여 매우 많은 계산량을 감소시킬 수 있다.

NMF는 (1)과 같이 프레임 t 에서 0 이상의 원소값을 가지는 벡터 X_t 를 기저 (basis) 벡터 b_j 와 이득 (gain)

$g_{j,t}$ 의 조합으로 가장 근사적으로 표현한다.

$$X_t \approx \sum_{j=1}^J g_{j,t} b_j \quad (1)$$

여기서 J 는 기저 벡터의 개수를 나타내며, $g_{j,t}$ 는 프레임 t 에서 j 번째 기저 벡터의 이득을 나타내며, g 와 b 의 모든 원소값은 0 이상이다.

모든 프레임의 X_t 을 동일한 기저 벡터로 모델링하기 위하여 (1)의 모든 t ($1 \leq t \leq K$)를 결합하여 (2)와 같은 행렬 표현식을 얻을 수 있다.

$$X \approx BG \quad (2)$$

여기서 $X = [X_1 \cdots X_K]$, $B = [b_1 \cdots b_j]$, $[G]_{j,t} = g_{j,t}$ 를 나타낸다. X 가 주어질 때, X 를 (2)와 같이 기저 행렬과 이득의 조합으로 가장 근사적으로 표현하기 위하여 (3)과 (4)로 표시되는 갱신 방법을 사용할 수 있다^[7].

$$B \leftarrow B \times \frac{X}{BG} G^T \quad (3)$$

$$G \leftarrow G \times \frac{B^T X}{B^T 1} \quad (4)$$

여기서 1은 모든 원소값이 1이고 행과 열의 길이가 X 와 같은 행렬을 나타낸다.

NMF 기반으로 포락선을 확장하기 위해 전체 광대역 훈련 데이터의 프레임별 스펙트럼 포락선 X_t 로 구성된 X 에 대하여 B 와 G 를 구하고, 각 프레임에서 포락선 확장할 때 B 를 활용해야 한다. 그러나 NMF 행렬의 크기가 증가할수록 계산량이 급격하게 증가한다. 따라서 본 논문에서는 DFT 주파수 단위의 스펙트럼 포락선으로 X 를 표현하지 않고, 일정한 부대역별로 포락선 에너지를 구하여 X 로 사용한다. 부대역별 에너지로 주파수 포락선을 조정하는 방법은 SBR (spectral bandwidth replication) 기술에서 사용되어 우수한 성능을 나타내고 있으므로 제안한 방법도 이와 유사한 방식을 사용한다^[9].

그림 3은 제안한 스펙트럼 포락선의 고대역 확장 방법을 나타낸다. Step 1은 훈련 과정이고, 훈련에 사용되는 광대역 신호는 샘플링 주파수가 16kHz이고 최대 주파수가 7kHz인 신호이고, 총 K 개의 프레임을 가진다.

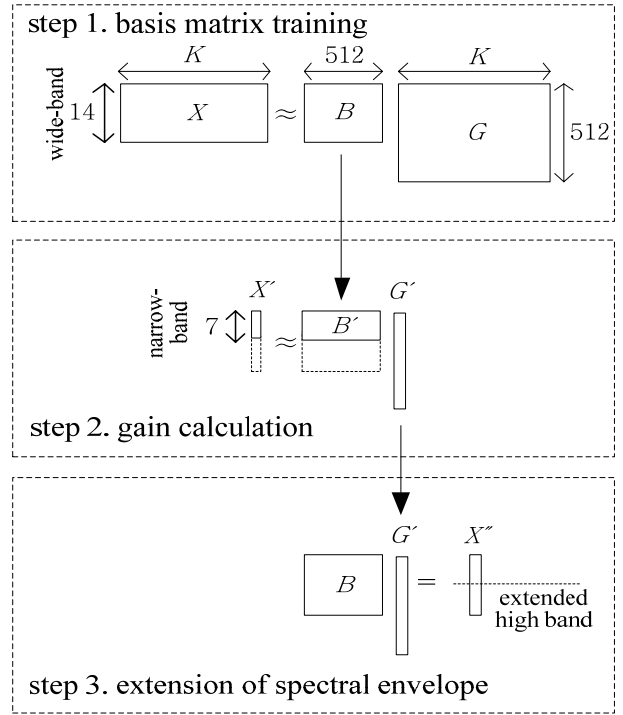


그림 3. NMF 기반의 스펙트럼 포락선의 고대역 확장 방법
Fig. 3. The method of high-band extension of spectral envelope based on NMF.

먼저, 각 프레임 신호의 16차 LPC 계수를 구한 후 LPC 계수에 대한 512-포인트 DFT를 수행하여 광대역 스펙트럼 포락선을 얻는다. 스펙트럼 포락선은 500Hz 대역폭을 가지는 총 14개의 부대역으로 구분하여 각 부대역의 에너지를 계산한다. 이와 같은 방법으로 $14 \times K$ 크기의 X 가 정의되고, (3)과 (4)에 따라 B 와 G 를 구한다. 본 논문에서는 총 512개의 기저 벡터를 사용하여 B 의 크기가 14×512 가 되고, (5)와 같이 표현된다. 이렇게 구해진 B 는 스펙트럼 포락선을 고대역으로 확장할 때 모델로 사용된다.

$$\begin{bmatrix} x_{1,1} & \cdots & x_{1,K} \\ \vdots & \ddots & \vdots \\ x_{14,1} & \cdots & x_{14,K} \end{bmatrix} \approx \begin{bmatrix} b_{1,1} & \cdots & b_{1,512} \\ \vdots & \ddots & \vdots \\ b_{14,1} & \cdots & b_{14,512} \end{bmatrix} \times \begin{bmatrix} g_{1,1} & \cdots & g_{1,K} \\ \vdots & \ddots & \vdots \\ g_{512,1} & \cdots & g_{512,K} \end{bmatrix} \quad (5)$$

input data basis gain

NMF 기반의 스펙트럼 포락선 확장은 그림 3의 step 2와 3 단계로 수행한다. 현재 프레임의 협대역 스펙트럼 포락선이 주어지면 500Hz 단위의 부대역별 에너지를 구하여 X' 를 정의하며, 협대역은 총 7개 부대역을 가지므로 X' 는 7×1 크기의 벡터이다. 다음, step 2와 같이 B 의 첫 7개의 열에 해당하는 7×512 크기의 B'

를 정의하고, (4)를 사용하여 $X' \approx B'G'$ 를 만족하는 512×1 크기 G' 를 구하고, 이 관계는 (6)으로 표현된다. G' 는 협대역 기저 벡터를 어떻게 결합하면 주어진 협대역 포락선 에너지가 되는지를 정의하며, 결국 협대역 스펙트럼 포락선 에너지의 특성을 통합적으로 보여준다.

$$\begin{bmatrix} x'_1 \\ \vdots \\ x'_7 \end{bmatrix} \approx \begin{bmatrix} b_{1,1} & \cdots & b_{1,512} \\ \vdots & \ddots & \vdots \\ b_{7,1} & \cdots & b_{7,512} \end{bmatrix} \times \begin{bmatrix} g'_1 \\ \vdots \\ g'_{512} \end{bmatrix} \quad (6)$$

마지막으로, Step 3과 같이 $B'G'$ 를 구하여 14×1 크기의 X'' 를 구하면 (7)이 된다. 즉, 고대역에 해당하는 기저 벡터 성분을 포함하여 전체 기저 벡터를 G' 에 따라 결합한 결과가 X'' 이며, 이는 협대역의 기저 벡터 결합 구조를 광대역에 동일하게 적용한 결과이므로, X'' 의 고대역 영역 원소값이 원하는 고대역 포락선의 부대역별 에너지가 된다.

$$\begin{bmatrix} b_{1,1} & \cdots & b_{1,512} \\ \vdots & \ddots & \vdots \\ b_{14,1} & \cdots & b_{14,512} \end{bmatrix} \times \begin{bmatrix} g'_1 \\ \vdots \\ g'_{512} \end{bmatrix} = \begin{bmatrix} x''_1 \\ \vdots \\ x''_{14} \end{bmatrix} \quad (7)$$

4. 고대역 위상 결정 방법

고대역 여기 신호의 스펙트럼과 고대역 스펙트럼 포락선 에너지를 결합하면 최종 고대역 스펙트럼이 결정되고, IDFT (inverse discrete Fourier transform)를 수행하기 위하여 최종적으로 고대역 스펙트럼의 위상을 결정해야 한다.

저대역과 고대역 위상 사이에는 의미 있는 상관관계가 존재하지 않으므로 저대역으로부터 고대역 위상을 추정하는 기법은 사용하지 못한다. 기존의 위상 결정 방법에는 하모닉 부호화에서 사용하는 프레임 위상 연속성을 보장하는 방법과, SBR에서 사용하는 저대역 위상 복사 방법 등이 있다^[9]. 본 논문에서는 프레임 연속성을 높여 왜곡을 감소시키는 것을 목표로 위상을 결정한다.

제안하는 인공 대역 확장 방법은 프레임 단위로 구현되고, 인접한 프레임 간의 신호 연속성을 높이기 위해 50% 중첩 가산을 사용한다. 그러나 중첩 영역에서 두 신호의 상관도가 낮으면 중첩 영역에서 복원된 신호의 왜곡이 발생한다. 또한, 위상은 주파수별 성분이 결합될 때 시간 축에서 각 성분의 상대적 위치를 결정하고, 그

결과 최종 시간 축 포락선 (temporal envelope)을 결정한다. 따라서 제안하는 방법은 대역 확장에서 프레임 사이의 신호 연속성을 최대 하도록 위상을 정한다.

IDFT를 수행하기 직전 단계의 최종 광대역 스펙트럼을 500Hz 부대역으로 구분하며, 14개의 부대역 중 저대역 7개의 부대역은 입력 협대역 신호이고, 고대역 7개의 부대역은 확장된 신호를 나타낸다. 저대역의 위상을 구하고 500Hz 부대역 단위로 벡터로 표현하여 코드 벡터를 정의하고, 그에 따라 7개의 코드 벡터로 구성된 위상 코드북 (codebook)을 생성한다. 다음, 아래 설명과 같이 고대역의 각 부대역 단위로 코드북에서 최적의 위상을 선택한다.

고대역 위상은 각 부대역 단위로 독립적으로 결정한다. 이전 프레임에서 합성한 고대역의 각 부대역 단위별 시간 축 신호 $p(n)$ 를 구한다. 여기서, n 은 시간 축 샘플 인덱스이다. 또한, 현재 프레임에서 고대역의 각 부대역 단위별로 k 번째 위상 코드 벡터를 사용할 경우의 시간 축 신호 $q_k(n)$ 를 구한다. 다음, 인접한 두 프레임의 중첩 영역에서 각 부대역 단위별로 $p(n)$ 와 $q_k(n)$ 의 상관관계를 구하고, 최대값을 가지는 k 를 검색하여 최종 부대역 위상을 결정한다. 모든 고대역 위상을 결정한 후에 최종적으로 IDFT를 수행하고 중첩-가산 동작을 수행하여 대역이 확장된 신호를 출력한다.

위상을 주파수 샘플 단위가 아니라 부대역 단위로 결정하는 이유는, 시간 축 파형의 포락선이 인접한 주파수 위상의 상호 관계에 의해 결정되기 때문에 주파수 샘플 단위로 위상을 결정할 경우 시간 축 포락선을 조정할 수 없기 때문이다. 또한, 부대역 단위로 위상을 결정하면 원본 신호가 가지는 인접한 주파수 위상의 상호

표 1. 프레임 중첩 영역에서 부대역 신호의 상관도 비교

Table 1. Comparison of sub-band signal correlation value in the overlapped region.

frequency (Hz)	phase copying in SBR	proposed method
3500 ~ 4000	0.01	0.33
4000 ~ 4500	-0.01	0.34
4500 ~ 5000	-0.06	0.35
5000 ~ 5500	-0.04	0.35
5500 ~ 6000	0.00	0.36
6000 ~ 6500	0.02	0.38
6500 ~ 7000	0.03	0.37

관계를 유지할 수 있는 장점을 가진다. 저대역 위상으로부터 위상 코드북을 정의할 때, 코드 벡터 사이의 중첩을 허용하여 코드북의 크기를 증가시키면 위상 선택 성능을 향상시킬 수 있으나, 선택을 위한 계산량이 증가하므로 본 논문에서는 7개의 코드 벡터만 사용한다.

표 1은 SBR의 저대역 위상 복사 방법과 제안한 위상 결정 방법으로 복원된 신호에 대하여, 인접한 두 프레임의 중첩 영역에서 부대역 단위로 신호의 평균 상관도를 보여준다. 제안한 방법을 사용한 경우에 상관도가 더 높아지는 것을 확인할 수 있다.

III. 성능 평가

주관적 청취 평가를 통하여 제안한 방법의 성능을 평가하였다. 성능 평가는 제안한 방법으로 대역 확장된 광대역 신호, AMR-NB로 복원된 협대역 신호, AMR-WB로 복원된 광대역 신호를 비교하여 진행한다. 성능 평가에는 NTT speech DB의 Korean 음성 신호를 사용하였고, 각 신호는 8~9초의 길이를 가진다. 총 96개의 신호 중 72개의 신호는 NMF 기저 행렬을 구하는 훈련 단계에 사용하였고, 나머지 24개 신호에 대한 대

역 확장 성능을 평가하였다. 평가에 참여한 평가자는 총 5명이고, 조용한 사무실 환경에서 헤드폰을 통하여 청취하였다.

그림 4는 원본 광대역 신호와 제안한 방법으로 고대역이 확장된 신호의 스펙트로그램을 보여준다. 그림 4 (b)에서 보듯이 대역 확장된 신호는 (a)의 하모닉 구조를 매우 유사하게 표현하고 있고, 저대역과 고대역 사이의 하모닉 구조 연속성을 유지하는 것도 확인할 수 있다.

청취 평가는 CCR (comparison category rating) 방식을 사용하여 수행한다^[10]. CCR 평가 방식은 청취자에게 두 개의 음성 샘플을 들려주고 첫 번째 샘플에 대한 두 번째 샘플의 상대적 성능을 평가한다. 성능은 총 7개의 단계로 평가하며, 3은 'Much Better', 2는 'Better',

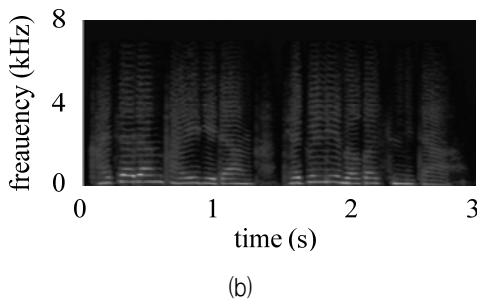
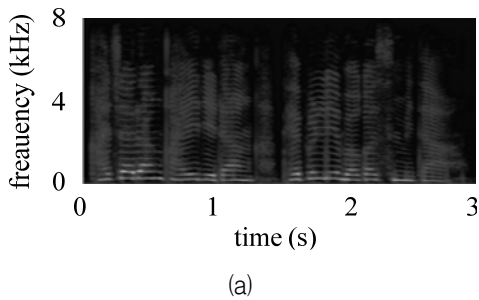


그림 4. 원본 광대역 신호 (a)와 제안한 방법으로 고대역이 확장된 신호 (b)의 스펙트로그램.

Fig. 4. Spectrogram of original wide-band signal (a) and high-band-extended signal by the proposed method (b).

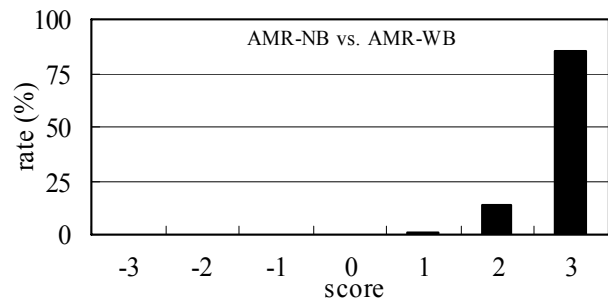
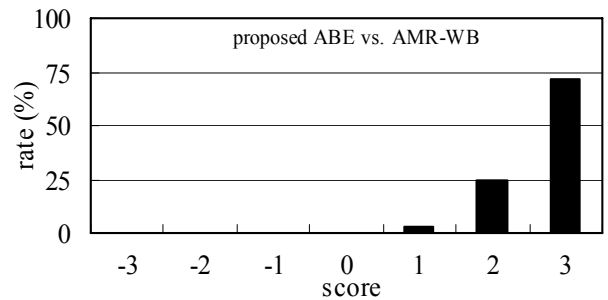
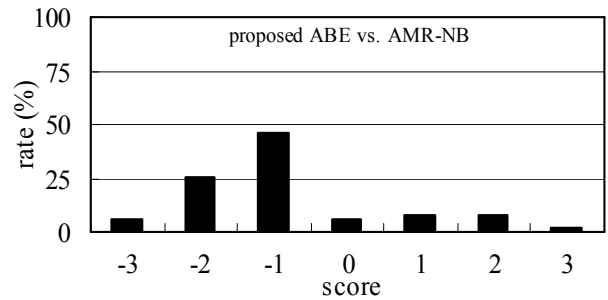


그림 5. 제안한 기술에 대한 주관적 청취 성능 평가 결과

Fig. 5. Result of subjective performance evaluation of the proposed method.

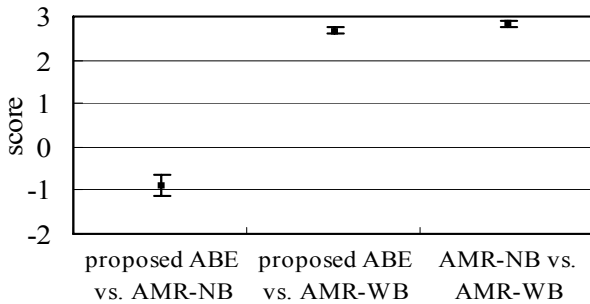


그림 6. 평가 항목별 주관적 청취 성능 평가 결과
Fig. 6. Result of subjective performance evaluation for each category.

1은 ‘Slightly Better’, 0은 ‘About the Same’, -1은 ‘Slightly Worse’, -2는 ‘Worse’, -3은 ‘Much Worse’를 의미한다.

그림 5는 제안한 기술을 이용하여 광대역으로 확장된 신호의 품질을 CCR 평가 방식을 통하여 측정된 결과이며, y 축의 rate는 전체 평가 신호와 모든 평가자에 대하여 각 점수가 선택된 개수의 백분율을 나타낸다. 각 그래프의 A vs. B에서 양수 값은 B가 A보다 음질이 좋은 것을 의미한다.

그림 6은 각 평가 항목별로 전체 신호에 대한 평균을 95% 신뢰구간과 함께 보여준다. “proposed ABE vs. AMR-NB” 항목에서 신뢰구간이 0과 겹치지 않고 -1과 겹치므로, 제안한 방법의 음질이 AMR-NB보다 청각적으로 뚜렷하게 우수한 것을 보여준다. 또한, “proposed ABE vs. AMR-WB” 점수가 “AMR-NB vs. AMR-WB” 점수보다 신뢰구간의 겹침 없이 0에 더 근접하므로, AMR-NB 보다 제안한 방법의 음질이 AMR-WB에 청각적으로 더 근접함을 알 수 있다.

이상의 주관적 평가로부터 제안한 방법으로 대역 확장된 신호의 음질이 원 협대역보다 우수한 것을 확인하였다. 또한, AMR-WB, 제안한 방법, AMR-NB 순으로 성능이 우수한 것을 확인할 수 있다. 물론, 제안한 방법의 성능이 AMR-WB보다 많이 낮은 것을 알 수 있으며, 이는 저대역 정보만으로 고대역 정보를 추정하여 신호를 합성하므로 성능 향상에 근본적인 한계가 있기 때문이다.

IV. 결 론

본 논문에서는 주파수 영역에서 하모닉 구조를 유지

하는 NMF 기반의 인공 대역 확장 기술을 제안하였다. 제안한 기술은 입력된 협대역 신호를 여기 신호와 스펙트럼 포락선으로 분리하고, 각각 독립적인 방법을 사용하여 대역을 확장한다. 여기 신호는 하모닉 구조가 유지되도록 확장하며, 스펙트럼 포락선은 NMF 기반으로 확장하고, 위상은 시간 축 신호의 프레임 상관관계를 기반으로 결정한다. NMF 기반의 고대역 포락선 확장에서 부대역 단위의 에너지 값을 사용하여 성능 저하 없이 NMF 계산의 복잡도를 감소시켰다.

제안한 방법의 성능을 주관적 청취 평가를 통하여 측정하였고, 제안한 방법으로 고대역 확장된 신호는 원 협대역 신호보다 청각적으로 우수한 음질을 가지는 것을 확인하였다.

REFERENCES

- [1] J. Sung, H. W. Kim, D. Y. Kim, B. S. Lee and Y. H. Ko, “A candidate codec algorithm on superwideband extension to ITU-T G.711.1 and G.722,” *J. Institute of Electronics Engineers of Korea*, vol. SP-47, no. 5, pp. 62-73, 2010. 9.
- [2] P. Jax and P. Vary, “On artificial bandwidth extension of telephone speech,” *Signal Processing*, vol. 83, no. 8, pp. 1707-1719, August 2003.
- [3] S. Chennoukh, A. Gerrits, G. Miet and R. Sluijter, “Speech enhancement via frequency bandwidth extension using line spectral frequencies,” in *Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing*, pp. 665-668, Salt Lake City, Utah, USA, May 2001.
- [4] P. Jax and P. Vary, “Artificial bandwidth extension of speech signals using MMSE estimation based on a hidden Markov model,” in *Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing*, pp. 680-683, Hong Kong, China, April 2003.
- [5] K. Y. Park and H. S. Kim, “Narrowband to wideband conversion of speech using GMM based transformation,” in *Proc. IEEE Conf. on Acoustics, Speech, and Signal Processing*, pp. 1843-1846, Istanbul, Turkey, June 2000.
- [6] K. B. Hong, G. H. Jeong and I. S. Lee, “Enhancement of super-wideband coder by considering audio feature in MDCT domain,” *J. Institute of Electronics Engineers of Korea*, vol. SP-48, no. 5, pp. 129-136, 2011.9.

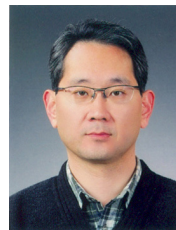
- [7] D. D. Lee and H. S. Seung. "Learning the parts of objects by non-negative matrix factorization," *Nature*, vol. 401, pp. 788-791, August 1999.
- [8] D. Bansal, B. Raj and P. Smaragdis, "Bandwidth expansion of narrowband speech using non-negative matrix factorization," in *Proc. Interspeech*, pp. 1505-1508, Lisbon, Portugal, September 2005.
- [9] M. Dietz, L. Liljeryd, K. Kjørling and O. Kunz, "Spectral band replication, a novel approach in audio coding," in *Proc. 112th AES Convention*, pp. 10-13, Munich, Germany, May 2002.
- [10] ITU-T Rec. P.800, "Methods for subjective determination of transmission quality," August 1996.

 저 자 소 개



김 기 준(학생회원)
 2011년 2월 광운대학교
 전자공학과 학사
 2011년 3월~현재 광운대학교
 전자공학과
 석박사통합과정

<주관심분야 : 음성/오디오 신호처리, 멀티미디어
 신호처리>



박 호 중(정회원)-교신저자
 1986년 서울대학교 전자공학과
 학사
 1987년 Univ. of Wisconsin-
 Madison 전기 및 컴퓨터
 공학과 석사

1993년 Univ. of Wisconsin- Madison 전기 및
 컴퓨터공학과 박사
 1993년~1997년 삼성전자 선임연구원
 1997년~현재 광운대학교 전자공학과 교수
 <주관심분야 : 음성/오디오 신호처리, 멀티미디어
 신호처리>