

2차원 이진 마스크를 이용한 적응형 음성향상 잡음 제거기

이기현^{*}, 이정현^{**}, 조진호^{***}, 김명남^{****}

Adaptive Noise Canceller for Speech Enhancement Using 2-D Binary Mask

Gihyoun Lee[†], Jyung Hyun Lee^{**}, Jin-Ho Cho^{***}, Myoung Nam Kim^{****}

ABSTRACT

Speech enhancement algorithm plays an important role in numerous speech signal processing applications. Over the last few decades, many algorithms have been studied for speech enhancement. The algorithms are based on spectral subtraction, Wiener filter, and subspace method etc. They have good performance of speech enhancement, but the performance can be deteriorated in specific noises or low SNR environment. In this paper, a new speech enhancement algorithms are proposed based on adaptive noise canceller. And the proposed algorithm improved performance of adaptive noise cancelling using 2-D binary mask. From objective experimental index, it is confirmed that the proposed algorithm is useful and has better performance than recently proposed speech enhancement algorithms.

Key words: Speech Enhancement, Adaptive Noise Canceller, 2-D Binary Mask

1. 서 론

최근 멀티미디어 기기의 발달과 함께 음성 신호처리에 관한 많은 연구가 이루어지고 있다. 특히 음성 신호처리 시스템에서 환경 잡음으로 인한 시스템 성능의 저하 현상은 해결되어야 할 중요한 문제로 인식되고 있다. 따라서 잡음이 시스템에 미치는 영향을

줄이기 위해 다양한 잡음 감쇄 기법과 음성향상 기법이 연구되어 왔으며 다양한 음성 신호처리 분야에 사용되고 있다[1].

음성향상은 음성신호가 주변 잡음에 의해 오염되어 입력되었을 때 음성 신호에서 잡음을 제거하고 음성을 강화하여 음성 신호를 향상 시키는 기법으로 극한의 작업환경이나 군사 작전 중에 사용되는 음성

※ Corresponding Author : Myoung Nam Kim, Address: (41566) Gukchaebosang-ro 680, Jung-gu, Daegu, Korea, TEL : +82-53-200-5266, FAX : +82-53-200-5264, E-mail : kimmn@knu.ac.kr

Receipt date : May 24, 2016, Approval date : June 7, 2016

[†] Dept. of Medical & Biological Eng., Graduate School, Kyungpook National University
(E-mail : gihyounlee@gmail.com)

^{**} Dept. of Biomedical Eng., School of Medicine, Kyungpook National University
(E-mail : pung@knu.ac.kr)

^{***} School of Electronics Eng., College of IT Engineering, Kyungpook National University
(E-mail : jhcho@ee.knu.ac.kr)

^{****} Dept. of Biomedical Eng., School of Medicine, Kyungpook National University

※ This research was financially supported by the Ministry of Trade, Industry and Energy (MOTIE) and Korea Institute for Advancement of Technology (KIAT), DaeGu Institute for Regional Program Evaluation (IRPE) through the Research and Development for Regional Industry (R0004188) and this work was supported by a grant from the National Research Foundation of Korea (NRF), which was funded by the Korean government (MSIP) (No. 2015R1A2A2A03006113).

통신 기기의 통신 품질을 향상 시키거나 여러 가지 스마트 장비나 의료기기에서 인간-기기 상호작용 시 음성 인식이나 화자 인식 성능을 높일 수 있다. 또한 헤드셋과 디지털 보청기와 같은 음향기기에 사용하여 배경 잡음을 억제하고 음질을 향상시키기 위해 사용될 수 있다[1,2].

기존의 고전적인 음성향상 알고리즘은 대부분 주파수 영역에서의 잡음 제거 방법, 통계적 모델 (statistic model)에 기반한 필터, 부분공간 (subspace)을 이용한 방법 등을 사용하였다[1-3]. 주파수 영역에서의 잡음 제거 방법으로는 주파수 차감법 (spectral subtraction)[4]이 있고 통계적 모델에 기반한 방법으로는 Wiener 필터[5]가 있으며, 부분공간을 이용한 방법으로는 마스크 필터링 (mask filtering) 방법이 있다. 주파수 차감법은 푸리에 변환을 이용해 변환된 주파수 영역에서 잡음의 스펙트럼을 추정하여 제거하는 방법으로 우수한 잡음 제거 능력이 있지만 음성과 비슷한 주파수 특성을 가진 잡음에 대해서는 좋은 성능을 보여주지 못하고 위상 검출 부분에서 어려움을 보이는 단점이 있다[6]. Wiener 필터는 잡음 참고신호 (reference signal)와 오차의 통계적 모델에 기반하여 원하는 신호를 추정하고 음성 신호를 향상하는 방법으로 잡음 환경에 맞춰 적용하여 잡음을 제거하는 적응형 잡음 제거기에 널리 사용되고 있다. 적응형 잡음 제거기는 음성향상 및 잡음제거 분야에 널리 사용되고 있으며 잡음제거와 음성향상에 좋은 성능을 보이지만 제거하기 위한 잡음 특성이 반영된 잡음 참고신호를 획득하기 위한 별도의 신호 입력단이 필요하며 낮은 신호 대 잡음비 (signal-to-noise ratio, SNR) 환경에서 성능 저하가 일어나는 문제점이 있다. 그리고 부분공간을 이용한 방법은 높은 잡음 제거 성능을 가졌지만 음성과 비슷한 특징을 가지거나 시간에 따라 통계적 특성이 변하는 불안정한 (unstable) 잡음에 대해서는 성능 저하가 일어나는 문제점을 가지고 있다[7]. 또한 최근 웨이블릿 변환에 기반한 웨이블릿 수축법 (wavelet shrinkage)을 이용한 잡음제거 방법에 관한 연구가 수행되고 있다. 하지만 웨이블릿 문턱치 구간에서 신호의 불연속이 생기는 문제를 가지고 있으며, 음성구간 내의 잡음을 제거하지 못하는 단점이 있다[8].

본 논문에서는 기존의 음성향상 알고리즘들이 가지고 있는 문제점을 해결하기 위해 잡음환경 적응형

잡음 제거기에 기반한 새로운 음성향상 알고리즘을 제안한다. 적응형 잡음 제거기에서 요구되는 참고신호를 제안한 2차원 이진 마스크를 이용하여 추출한다. 제안한 2차원 이진 마스크는 웨이블릿 패킷 분해를 통해 분해된 음성신호를 시간-주파수 2차원 영역에서 잡음영역을 분리하여 잡음을 추출한다. 이 잡음을 적응형 잡음 제거기의 참고신호에 적용하여 음성신호의 손실을 최소화 하며 음성을 향상 시킨다. 제안한 알고리즘의 음성향상 결과를 객관적으로 비교 평가하기 위해 객관적인 음질 평가 지표인 PESQ (perceptual evaluation of speech quality)[9]와 향상된 음성신호의 정확도를 객관적 수치로 나타낸 SNR_{seg} (segmental SNR)[1]를 이용하였다. 기존의 Wiener 필터와 마스크 필터링 기법을 개선하여 최근 제안된 Wiener WT (wavelet thresholding)[10], IBM (ideal binary mask)[11] 음성향상 알고리즘과 지표로서 비교하였다. 그 결과, 모든 지표에서 제안한 음성향상 방법이 기존의 음성향상 알고리즘들 보다 개선된 성능을 보였으며, 제안한 음성향상 방법이 대부분의 잡음 환경에서 높은 음질과 정확도를 보이는 것을 확인하였다.

2. 제안하는 음성향상 알고리즘

2.1 제안한 알고리즘의 개요

본 논문에서는 잡음이 섞인 음성신호에서 잡음을 제거하여 음성향상하기 위한 새로운 적응형 잡음 제거기를 제안한다. 제안된 적응형 잡음 제거기는 잡음 추정과정에서 한계를 보이는 기존의 적응형 잡음 제거기의 단점을 보완하기 위해 인지적 웨이블릿 패킷 분해 (perceptual wavelet packet decomposition, PWPD)를 이용해 분해된 음성신호에서 2차원 이진 마스크 필터링을 통해 추출된 잡음 밴드를 참고신호로 활용한다[12]. Fig. 1은 제안하는 음성향상 알고리즘의 전체적인 흐름도이다.

먼저, 잡음이 섞인 음성신호가 입력되면 PWPD를 이용해 17개의 웨이블릿 서브밴드로 분해된다. 분해된 신호는 17개의 주파수 정보를 가진 시간영역 신호의 형태를 가지고 있으며, 시간과 주파수의 정보를 모두 나타내어 2차원 행렬로 나타낼 수 있다. 2차원 행렬 신호를 제안하는 2차원 이진 마스크를 이용한 필터링과정을 거치게 되어 음성 밴드와 잡음 밴드를

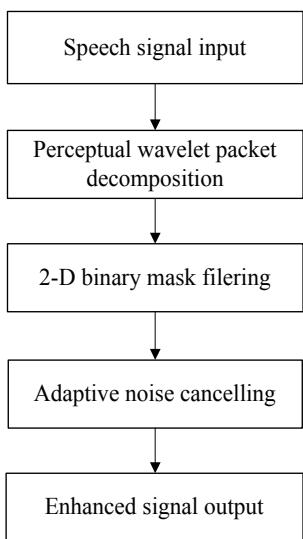


Fig. 1. Block diagram of the proposed speech enhancement algorithm.

분리하게 된다. 분리된 잡음 밴드를 이용하여 적응형 잡음 제거기에 쓰이는 잡음 참고신호를 추출하게 된다. 추출된 참고신호로 잡음을 추정하는 적응형 잡음 제거 과정을 거친 후, 최종적으로 향상된 음성 신호를 얻을 수 있다.

2.2 인지적 웨이블릿 패킷 분해 (PWPD)

신호처리 분야에서 주파수 영역의 정보를 얻기 위해 널리 사용되고 있는 웨이블릿 변환 (wavelet

transform)은 다중 해상도를 이용하여 신호의 영역 변환과정에서 발생하는 해상도의 문제를 해결하여 푸리에 변환의 단점을 보완 하였다. 그리고 웨이블릿 패킷 분해 (wavelet packet decomposition, WPD)는 웨이블릿 변환을 기반으로 웨이블릿 필터 뱅크 (wavelet filter bank)를 변형한 형태를 가지고 있다 [11]. WPD의 대표적인 분해 방법 중 하나인 PWPD는 기본적인 웨이블릿 패킷 분해 방법인 기존의 WPD 방법과 같이 산술적인 밴드 별 에너지를 기반으로 하지 않고 인간의 음향 청각 모델을 기반으로 인간의 청신경에 자극 되는 에너지의 크기에 맞추어 Fig. 2와 같이 17개의 서브밴드를 가진 웨이블릿 필터뱅크를 구성하였다.

입력된 음성 신호는 PWPD를 통해 Fig. 2의 구조로 17개의 서브밴드를 가지는 웨이블릿 계수 ($w_{j,m}(k)$)로 분해된다. 웨이블릿 계수 ($w_{j,m}(i)$)는 PWPD에서 j 번째 레벨, m 번째 서브밴드의 i 번째 웨이블릿 계수를 나타내며, $j=3, 4, 5, m=1, \dots, 17$ 이다. 본 논문에서는 웨이블릿 계수 ($w_{j,m}(i)$)를 시간과 주파수 영역의 정보를 동시에 처리하기 위해 식 (1)과 같이 2차원 행렬로 나타낸다.

$$\Psi_m(t) = \begin{pmatrix} \psi_1(1) & \psi_1(2) & \dots & \psi_1(t) \\ \psi_2(1) & \psi_2(2) & \dots & \psi_2(t) \\ \vdots & \vdots & \ddots & \vdots \\ \psi_m(1) & \psi_m(1) & \dots & \psi_m(t) \end{pmatrix} \quad (1)$$

여기서 $\Psi_m(t)$ 는 특정시간 t 에서의 m 번째 서브밴드의 웨이블릿 계수를 나타낸다.

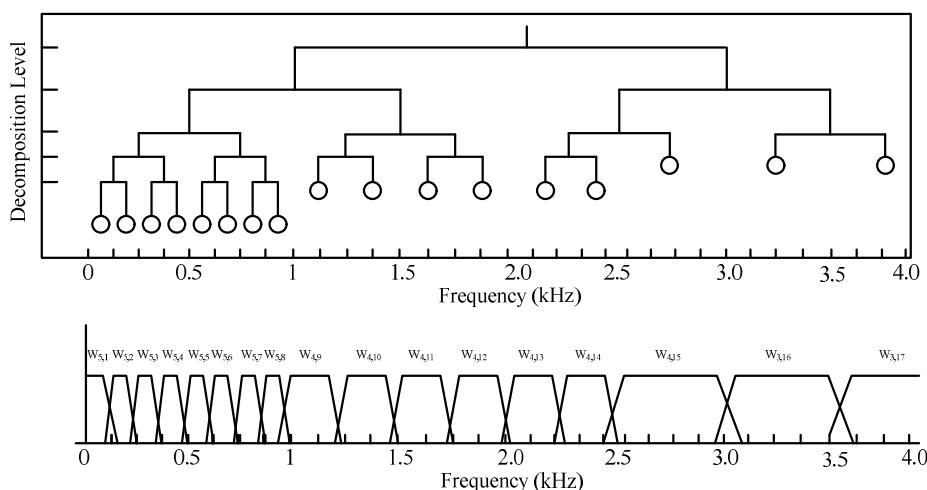


Fig. 2. A Structure of PWPD [11].

2.3 제안한 2차원 이진 마스크

기존의 적응형 잡음 제거기는 다채널의 입력신호들로부터 획득한 참고신호나 선형적 신호 대 잡음비, 웨이블릿 문턱치 등을 이용해 추정된 잡음을 사용하였다. 따라서 기존의 방법은 여러 개의 채널을 요구하거나 높은 연산량을 가지는 단점이 있었다. 이러한 단점을 보완한 제안하는 적응형 음성향상 잡음 제거기는 단일 채널 입력 신호를 이용하여 잡음을 제거하고 음성을 향상 시킨다. 적응형 잡음 제거기에 사용될 잡음 참고신호를 추정하기 위해 2차원 이진 마스크를 제안한다. 제안한 2차원 이진 마스크는 전 단계의 PWPD를 이용해 분해된 음성신호를 음성밴드와 잡음밴드로 분리하는 필터 역할을 한다. 2차원 이진 마스크는 분해된 첫 잡음 프레임의 밴드별 웨이블릿 계수의 편차에 기반한 마스크 특징 벡터를 통해 추출된다. 마스크 특징 벡터는 다음의 식 (2)와 (3)의 과정을 통해 계산된다.

$$\nu_m = \frac{1}{N} \sqrt{\frac{1}{N} \sum_{i=1}^N \Psi_m(i)^2 - \left[\frac{1}{N} \sum_{i=1}^N \Psi_m(i) \right]^2} \quad (2)$$

$$\lambda_m = B \nu_m \sqrt{\log(N) \cdot N} \quad (3)$$

여기서 ν_m 은 m 번째 웨이블릿 밴드의 잡음 편차이며, N 은 한 프레임의 샘플 개수이다. 그리고 B 는 웨이블릿 밴드의 개수이며 본 논문에서는 PWPD를 사용하였기 때문에 B 는 17의 값을 가진다. 또한 λ_m 은 마스크 특징 벡터로 2차원 이진 마스크 추출을 위해 m 번째 웨이블릿 밴드의 잡음 특징(feature)을 나타낸다. 다음의 식 (4)에서 각각의 웨이블릿 밴드의 잡음 특징을 이용하여 시간 영역과 주파수 영역의 2차원에 대한 이진 마스크를 추출한다.

$$M_m(k) = \begin{cases} 1, & \bar{\Psi}_m(k) > \lambda_m \\ 0, & otherwise \end{cases} \quad (4)$$

여기서 $M_m(k)$ 는 m 번째 웨이블릿 밴드의 k 번째 이진 마스크를 나타내며, $\bar{\Psi}_m(k)$ 는 m 번째 웨이블릿 밴드의 k 번째 프레임의 웨이블릿 계수 평균값이다. 2차원 이진 마스크 $M_m(k)$ 는 0과 1의 값을 가지며, 분해된 웨이블릿 계수들을 프레임별로 잡음밴드와 음성밴드로 분리하여 필터링하는 역할을 한다. 앞선 과정을 통해 추출된 2차원 이진 마스크의 결과를 Fig. 3에 나타내었다.

Fig. 3 (a)는 백색잡음이 SNR 5dB의 크기로 섞인 음성신호이며, (b)는 (a)의 음성신호를 전단계의 PWPD를 이용해 분해한 결과이고, (c)는 추출된 2차원 이진 마스크이다. (b), (c) 모두 x 축은 웨이블릿 밴드를 나타내고, y 축은 시간을 나타낸다. 그리고 (b)에서 z 축은 웨이블릿 계수를 나타내며, (c)에서 z 축은 이진 마스크값을 나타낸다. (b)에서 17개의 웨이블릿 서브밴드로 분해된 시간영역과 주파수영역의 음성신호 정보를 볼 수 있으며, 큰 웨이블릿 계수를 가지는 음성영역밴드의 구간을 확인 할 수 있다. (c)의 마스크는 0과 1의 값을 가지며 0은 잡음영역밴드를 1은 음성영역밴드를 나타낸다. (b)에서 확인 할 수 있는 음성영역밴드의 구간과 (c)의 음성영역밴드의 구간이 거의 일치하는 것을 볼 수 있으며, 제안한 2차원 이진 마스크로서 음성영역밴드와 잡음영역밴드를 분리할 수 있다. 분리된 잡음영역 밴드의 웨이블릿 계수를 이용하여 다음의 식 (5)와 같이 잡음 참고신호를 추정한다.

$$\mathfrak{J}_m(t) = \begin{cases} \bar{\Psi}_m(t), & M_m(t) = 0 \\ 0, & otherwise \end{cases} \quad (5)$$

여기서 $\mathfrak{J}_m(t)$ 는 시간에 따른 추정된 웨이블릿 잡음 밴드 계수이며, 모든 밴드의 웨이블릿 계수의 합 연산을 통해 잡음 참고신호를 추정한다. 추정된 잡음 참고신호를 Fig. 4에 나타내었다.

Fig. 4 (a)는 백색잡음이 SNR 5dB의 크기로 섞인 음성신호이며, (b)는 제안하는 2차원 이진 마스크를 이용해 추정된 잡음 참고신호이다. (a)에서 전 영역에서 백색잡음이 섞인 음성신호를 볼 수 있으며, 백색잡음은 음성이 없는 구간에는 잡음만 존재하지만 음성이 있는 구간에서는 음성 속에 잡음이 혼재되어 있으므로 잡음을 분리하기가 힘들다. (b)에서 추정된 잡음 참고신호를 볼 때, 잡음만 존재하는 구간뿐만 아니라 음성 구간 내의 잡음까지 추정된 결과를 볼 수 있다. 추정된 잡음 참고신호는 음성 구간과 잡음 구간 모두의 통계적 특징을 보존하며, 음성 신호와의 독립성을 가지므로 제안하는 적응형 잡음 제거기에 사용될 수 있다.

2.4 제안하는 2차원 이진 마스크를 이용한 적응형 잡음 제거기

기존의 적응형 잡음 제거기는 다 채널 입력신호를 요구하거나 잡음 참고신호를 추정하기 위해 웨이블

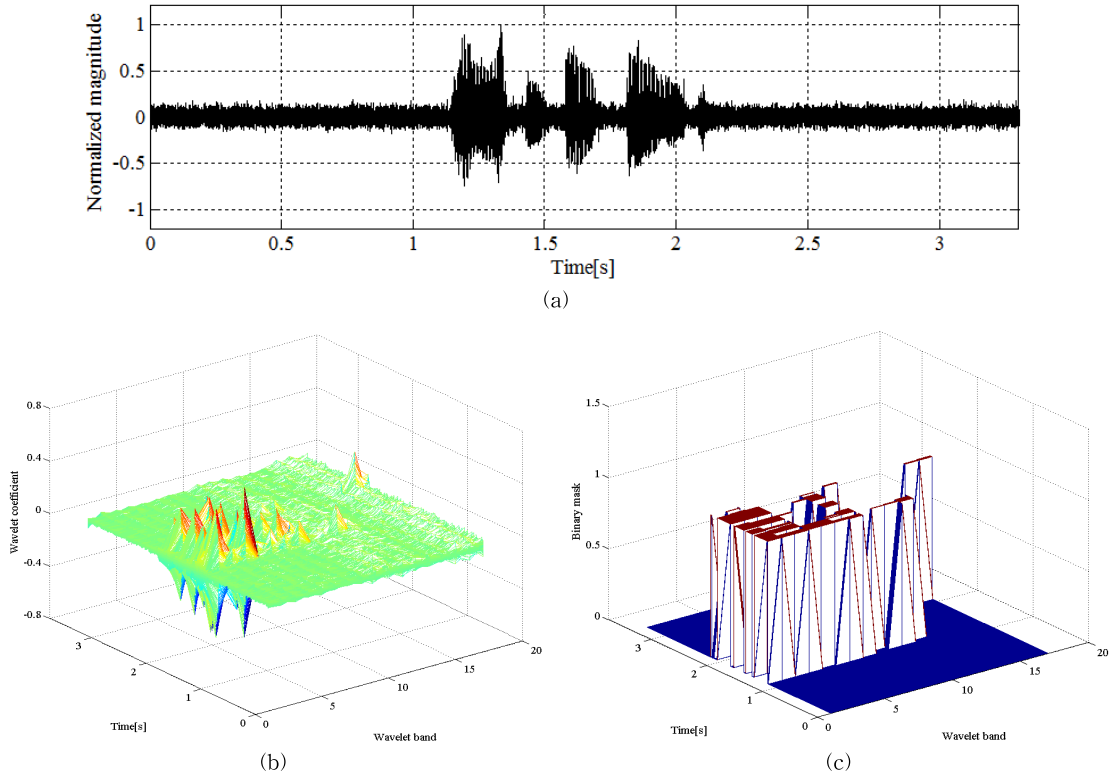


Fig. 3. 2-D binary mask, (a) a speech signal mixed with white noise (b) decomposed speech signal using PWD, and (c) the extracted 2-D binary mask.

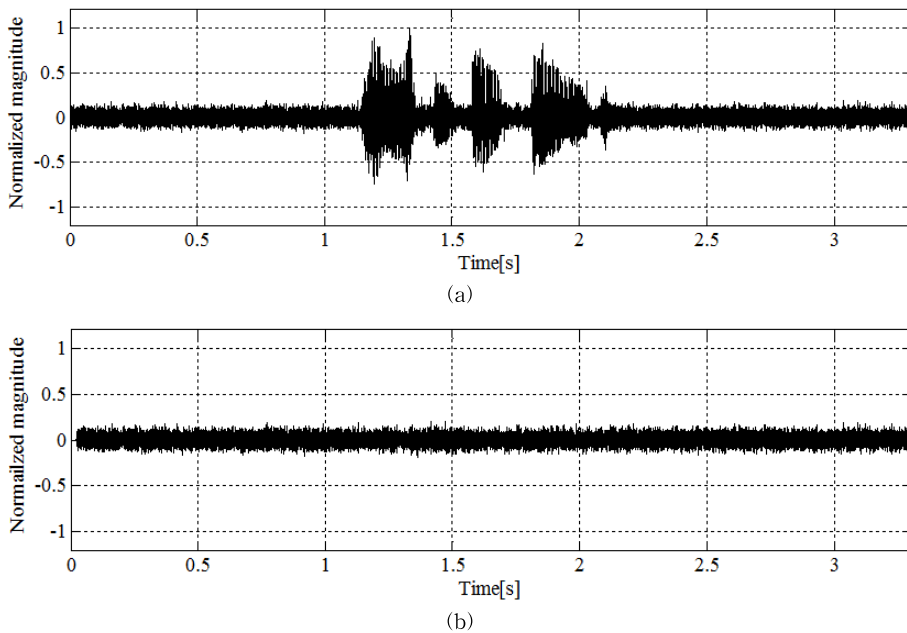


Fig. 4. Estimation of noise reference signal, (a) a speech signal mixed with white noise, and (b) the estimated noise reference signal.

릿 문턱치, 선형적 신호 대 잡음비 등의 방법을 이용하기 때문에 단일채널 입력신호를 가진 음향 기기에서는 사용할 수 없거나 특정 잡음환경에서 성능이 떨어지는 단점이 있었다. 제안하는 적응형 잡음 제거기는 Fig. 5와 같은 구조를 가지고 있으며, 전 단계에서 추정된 잡음 참고신호를 이용하여 단일채널 입력 신호만으로 잡음을 제거하여 음성을 향상 시키며, 모든 잡음 환경에 적용할 수 있도록 설계하였다.

Fig. 5에서 볼 수 있듯이, 입력된 음성신호는 PWPDP를 이용해 웨이블릿 패킷 분해된다. 그리고 2차원 이진 마스크로 필터링 되어 잡음밴드와 음성밴드로 분리되며 잡음밴드의 웨이블릿 신호로 참고신호를 추출한다. 입력된 음성신호는 연산을 위한 시간과 한 프레임의 데이터 축적 시간만큼 지연된 시간으로 입력되며 추출된 참고신호와 함께 시스템이 최소의 오차를 가지도록 적응 필터의 값을 갱신하여 효율적으로 잡음을 제거하는 과정을 거쳐 입력된 음성신호의 잡음을 제거하고 음성을 향상시킨다.

3. 실험 결과 및 고찰

제안한 방법의 유효성을 검증하기 위하여 공인된 데이터베이스에서 임의 추출한 신호 샘플을 사용하였고, 다양한 잡음환경을 만들어 실험하였다. 음성신호의 샘플은 TIMIT[13] 데이터베이스에서 추출하였으며, 잡음신호의 샘플은 NOISEX-92[14] 데이터

베이스에서 추출하였다. 데이터 샘플은 16bit의 비트 심도, 16kHz의 샘플링레이트, 그리고 256kbps의 비트율을 가진다. 또한 음성신호는 다양한 사람들이 발음한 120개의 음성 신호 샘플을 임의 추출하였으며, 다양한 잡음환경에서 실험을 수행하기 위해, 백색 잡음 (white noise), 자동차 잡음 (car noise), 웅성거림 잡음 (babble noise), 공장 잡음 (factory noise), 그리고 탱크엔진 잡음 (Leopard noise)을 다양한 SNR (0dB, 5dB, 10dB, 15dB)로 음성 신호와 섞어 실험 환경을 구축하였다. 제안한 알고리즘의 성능을 비교 평가하기 위해 기존의 고전적인 적응형 잡음제거 알고리즘인 Wiener 필터에 웨이블릿 문턱치를 적용하여 참고신호를 추정하는 Wiener WT 알고리즘과 Wiener 필터와 이진 마스크를 이용한 IBM 알고리즘과 비교하였다. SNR 5dB의 백색 잡음이 섞인 환경에서의 제안한 알고리즘과 비교 알고리즘들을 음성향상 결과를 Fig. 6에 그래프로 나타내었다.

Fig. 6 (a)는 잡음이 섞이지 않은 깨끗한 음성신호이며, (b)는 백색잡음이 SNR 5dB의 크기로 섞인 음성신호이다. (c)는 기존의 음성향상 알고리즘인 Wiener WT의 음성향상 결과로 주변부의 잡음을 많이 감쇄시켜 음성을 향상 시킨 결과를 보였다. 하지만 잡음을 확실하게 제거하지 못하여 음성구간 주변부에 잡음이 많이 남아 있는 결과를 보인다. (d)는 기존의 음성향상 알고리즘인 IBM을 이용한 결과로 높은 잡음 제거 성능을 보여준다. 하지만 너무 많은

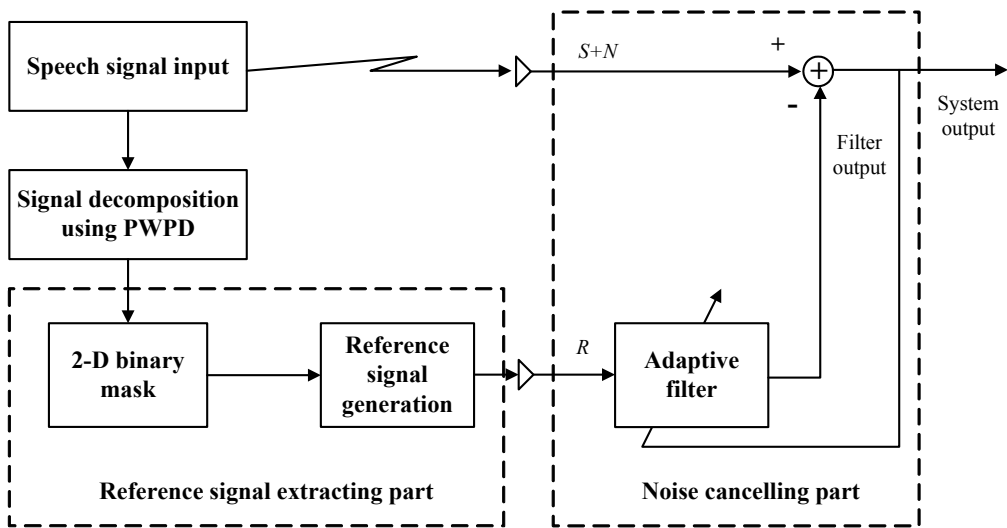


Fig. 5. The structure of proposed adaptive noise canceller.

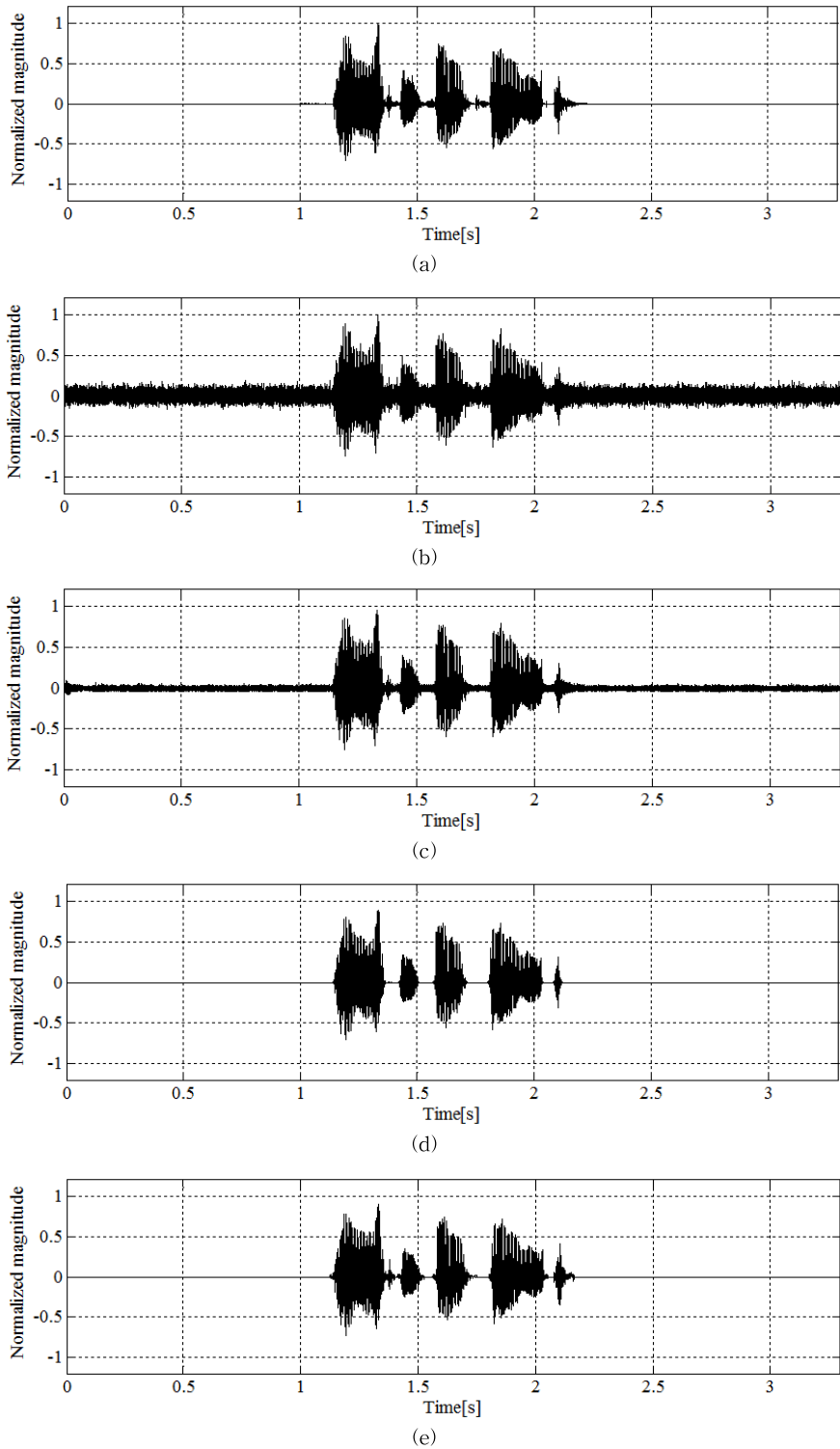


Fig. 6. The speech enhancement results (a) clean signal, (b) contaminated input signal, (c) Wiener WT, (d) IBM, and (e) proposed algorithm.

잡음을 제거하여 음성신호의 손실이 있는 것을 볼 수 있다. (e)는 제안한 적응형 잡음 제거기를 이용하여 음성을 향상 시킨 결과로 주변부의 잡음이 깨끗하게 제거되고 음성신호도 유지된 것을 확인 할 수 있다.

음성향상 성능을 객관적으로 평가하기 위해 ITU-T recommendation P.862에 채택된 PESQ를 평가 지표로 사용하였다.[9] P.862는 전화 통신 및 음성 코덱의 객관적 평가 지표로 제안된 표준이며, PESQ는 음성의 크기, 활성화도, 지연, 에코, 그리고 패턴 등을 모두 감안하여 모든 언어에 활용 가능하도록 디자인 되어 현재 가장 널리 사용되고 있는 객관적 음질 평가 지표이다. 또한 향상된 음성신호의 각 프레임별 평균 SNR을 계산하여 신호 전체에 대해 기하평균으로 음성신호의 정확도를 평가하는 SNR_{seg} 을 이용해 정확도를 평가하였다. 제안한 알고리즘을 앞서 설명한 5가지 잡음 전체의 평균값을 SNR 크기에 따라 객관적 음성향상 성능 평가 지표인 PESQ와 SNR_{seg} 의 결과를 Table 1에 나타내었다.

Table 1에서 기존의 음성향상 알고리즘들과 제안하는 알고리즘의 객관적 음성향상 성능을 비교하여 볼 수 있다. SNR 0dB의 잡음 환경은 음성신호의 에너지와 잡음신호의 에너지가 같은 환경으로 매우 잡음이 심한 환경이라고 볼 수 있다. 이러한 0dB의 환경에서는 기존의 알고리즘은 좋은 성능을 보여주지 못한다. 특히 음질의 지표인 PESQ의 결과에서 1이하의 매우 낮은 음질을 결과를 보였다. 그리고 SNR이 증가하여 잡음이 줄어든 환경에서도 잡음 제거와 함께 음성신호의 손실을 가져와 15dB 잡음 환경에서 SNR_{seg} 의 결과가 15dB보다 작은 결과를 보였다. 반면, 제안한 알고리즘은 0dB 환경에서 7dB 이상의 향상을 보였고, 다른 알고리즘은 15dB 환경에서도 3.7dB의 향상된 SNR_{seg} 의 결과를 보였다. 그리고 높은 잡음 제거 성능을 보였던 IBM은 너무 많은 신호

를 제거하여 향상된 음성신호의 정확성의 지표인 SNR_{seg} 의 결과는 Wiener WT와 비슷한 결과를 보이지만 잡음 제거과정에서의 음성신호 손실로 인해 음질의 지표인 PESQ 결과에서 가장 낮은 음질을 가지는 것을 확인 할 수 있다. 반면, 제안한 알고리즘은 0dB의 환경에서도 2 이상의 우수한 PESQ 결과를 보여 음질의 지표에서도 좋은 결과를 보였다. 이상과 같이, 음질과 정확성 측면에서 제안한 알고리즘과 기존 알고리즘의 음성향상 성능을 비교한 결과, 제안한 알고리즘이 모든 잡음 환경에서 기존의 알고리즘보다 개선된 음성향상 성능을 보였으며, 특히 잡음이 강하게 나타나는 낮은 SNR 환경에서 더 좋은 결과를 보였다.

4. 결론

본 논문에서는 음성 통신, 인간-기기 상호 작용, 디지털 보청기 등 다양한 음성 신호 처리 분야에서 중요하게 사용되는 새로운 음성향상 방법을 제안하였다. 제안한 음성향상 방법은 기존의 주파수 영역 변환 방법인 푸리에 변환에서 나타나는 영역 변환 과정에서의 해상도 문제를 개선한 웨이블릿 변환을 이용하여 음성 신호의 시간 영역의 정보와 주파수 영역의 정보를 모두 활용하였으며, 기존의 전처리 과정에서 많이 사용되던 웨이블릿 변환을 기반으로 변형된 형태를 가진 PWPD를 사용하여 음성신호를 분해하였다. 또한 기존의 이진 마스크를 이용한 기법에서 발생하는 음성신호의 손실 문제를 개선한 새로운 2차원 이진 마스크를 이용하여 분리된 잡음 밴드로써 잡음 구간의 잡음뿐만 아니라 음성 구간 내의 잡음을 추정하여 이어지는 적응형 잡음 제거기에 잡음 참고신호로 사용하여 잡음을 제거하였다.

제안한 음성향상 알고리즘의 효용성을 검증하기

Table 1. The results of speech enhancement

	Wiener WT		IBM		Proposed	
	SNR_{seg}	PESQ	SNR_{seg}	PESQ	SNR_{seg}	PESQ
0dB	5.72dB	0.97	4.83dB	0.86	7.62dB	2.02
5dB	8.54dB	1.08	7.65dB	0.87	11.12dB	2.48
10dB	11.43dB	1.33	10.97dB	0.95	14.62dB	2.89
15dB	14.05dB	1.75	14.74dB	1.16	18.70dB	3.26
Average	9.93dB	1.28	9.55dB	0.96	13.02dB	2.66

위해 기존의 음성향상 알고리즘을 개선하여 최근 제안된 음성향상 알고리즘들과 다양한 음성 샘플과 잡음 환경에서 비교 실험을 하였다. 그리고 음성향상 결과를 객관적으로 비교평가하기 위해 전화통신 및 음성 코덱의 객관적인 음질 평가 지표로 널리 사용되는 PESQ를 사용하여 음질을 객관적인 지표로 평가하였으며, SNRseg를 이용하여 향상된 음성 신호와 깨끗한 원 신호의 정확도를 객관적으로 평가 하였다. 그 결과, 모든 지표에서 제안한 음성향상 방법이 비교 음성향상 알고리즘 중에 가장 좋은 성능을 보였으며, 다양한 잡음 종류와 SNR 변화 환경에서 높은 음질과 정확도를 보이는 것을 확인하였다.

제안한 음성향상 알고리즘은 통신을 위한 음성향상 분야뿐만 아니라 스마트기기, 디지털 보청기 등 거의 모든 음성 신호처리 분야에 유용하게 적용될 수 있을 것으로 예상된다.

REFERENCE

- [1] P.C. Loizou, *Speech Enhancement: Theory and Practice*, 2nd ed., CRC Press, Boca Raton, Florida, 2013.
- [2] M. Grimm and K. Kroschel, *Robust Speech Recognition and Understanding*, I-Tech Education and Publishing, Vienna Austria, 2007.
- [3] J. Proakis and D. Manolakis, *Digital Signal Processing*, 3rd ed., Prentice Hall, Upper Saddle Rive, NJ, 1996.
- [4] S.F. Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," *IEEE Transaction on Acoustics Speech Signal Processing*, Vol. 27, No. 2, pp. 113-120, 1979.
- [5] N. Wiener, *Extrapolation, Interpolation, and Smoothing of Stationary Time Series Vol. 2*, MIT Press, Cambridge, 1949.
- [6] J. Beh and H. Ko. "A Novel Spectral Subtraction Scheme for Robust Speech Recognition: Spectral Subtraction Using Spectral Harmonics of Speech," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1. pp. I-64, 2003.
- [7] Y. Hu and P. Loizou, "A Generalized Subspace Approach for Enhancing Speech Corrupted by Colored Noise," *IEEE Transaction on Speech and Audio Processing*, Vol. 11, No. 4, pp. 334-341, 2003.
- [8] J.F. Zhu and Y.D. Huang, "Improved Threshold Function of Wavelet Domain Signal Denoising," *Proceeding of International Conference on Wavelet Analysis and Pattern Recognition*, pp. 14-17, 2013.
- [9] ITU, *Perceptual Evaluation of Speech Quality (PESQ), and Objective Method for End-to-End Speech Quality Assessment of Narrow-band Telephone Networks and Speech Codecs*, ITU-T Recommendation P.862, 2000.
- [10] Y. Hu and P. Loizou, "Speech Enhancement Based on Wavelet Thresholding the Multitaper Spectrum," *IEEE Transaction on Speech and Audio Processing*, Vol. 12, No. 1, pp. 59-67, 2004.
- [11] Y. Li and D. Wang, "On the Optimality of Ideal Binary Time-Frequency Masks," *Speech Communication*, Vol. 51, No. 3, pp. 230-239, 2009.
- [12] G.H. Lee, Y.J. Lee, J.H. Cho, M.N. Kim, "Voice Activity Detection Algorithm Using Fuzzy Membership Shifted C-means Clustering in Low SNR Environment," *Journal of the Korea Multimedia Society*, Vol. 17, No. 3, pp. 312-323, 2014.
- [13] J.J. Godfrey, C.E. Holliman, and J. McDaniel, "SWITCHBOARD: Telephone Speech Corpus for Research and Development," *Proceedings of International Conference on Acoustics, Speech, and Signal Processing*, Vol. 1, pp. 517-520, 1992.
- [14] A. Varga and J.M.S. Herman, "Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems," *Speech Communication*, Vol. 12, No. 3, pp. 247-251, 1993.



이 기 현

2009년 8월 경북대학교 천문대기
과학과(이학사)
2012년 2월 경북대학교 대학원
의용생체공학과(공학석사)
2012년 3월~현재 경북대학교 대
학원 의용생체공학과 박
사과정

관심분야: 생체신호처리, 의용전자기기



조 진 호

1977년 2월 경북대학교 공업교육
학과 졸업(공학사)
1979년 2월 경북대학교 대학원 전
자공학과 졸업(공학석사)
1988년 2월 경북대학교 대학원 전
자공학과 졸업(공학박사)

1984년~현재 경북대학교 IT대학 전자공학부 교수
관심분야: 의용전자, 생체신호처리, 이식형 보청기



이 정 현

2002년 2월 경북대학교 전자전기
공학부(공학사)
2004년 2월 경북대학교 대학원
전자공학과(공학석사)
2008년 2월 경북대학교 대학원
전자공학과(공학박사)

2008년~현재 칠곡 경북대학교병원 의공학과 조교수
관심분야: 생체 텔레메트리 시스템, 양이 보청기



김 명 남

1988년 2월 경북대학교 전자공학
과(공학사)
1990년 2월 경북대학교 대학원
전자공학과(공학석사)
1995년 2월 경북대학교 대학원
전자공학과(공학박사)

1996년~현재 경북대학교 의학전문대학원 의공학교실
주임교수
관심분야: 생체신호처리시스템, 의학영상처리