

# Multi-band Power Subtraction과 Wavelet Packets Decomposition을 이용한 개선된 음성 향상 방법

정회원 이윤창\*, 준회원곽정훈\*\*, 종신회원 안상식\*\*\*

## Improved Speech Enhancement Algorithm employing Multi-band Power Subtraction and Wavelet Packets Decomposition

Yoon-Chang Lee *Regular Member*, Jeong-Hoon Kwak *Associate Member*,  
Sang-Sik Ahn *Lifelong Member*

### 요약

잡음은 음성과 관련된 시스템의 성능을 제한하는 주된 원인이기 때문에 음성향상과 관련된 연구는 꾸준히 계속 되어왔다. 전통적인 음성향상 방법은 무성음과 잡음을 구분하지 않기 때문에 잡음제거 과정에서 무성음이 함께 제거되는 단점이 있으며, 웨이블릿 기반의 전통적인 잡음제거 방법은 각 대역마다 동일한 문턱값을 사용하기 때문에 시변 환경에서 성능이 떨어지는 단점이 있다. 이 단점들을 개선하기 위해 다중대역 파워 차감법과 Perceptual 웨이블릿 패킷 분해를 이용한 웨이블릿 기반의 개선된 음성향상 방법을 제안한다. 전처리 과정으로 다중대역 파워 차감법을 사용하여 광대역 잡음을 제거하고 뮤지컬 잡음의 발생을 줄이며, psycho-acoustic 모델 기반 Perceptual 웨이블릿 패킷으로 신호를 분해한 후 각 웨이블릿 노드의 엔트로피 비율과 음성검출을 이용하여 무성음/유성음/잡음을 구분한다. 구분된 신호에 따라 각 웨이블릿 노드마다의 문턱값을 기준으로 웨이블릿 Shrinkage를 적용하여 잡음을 제거하고 무성음이나 파워가 작은 유성음이 제거되는 오류를 최소화 한다. 또한 잡음 파워 추정 과정에 적응적으로 망각 계수를 선택하여 잡음 파워 추정 오류를 최소화한다.

**Key Words** : speech enhancement; multiband processing; wavelet packets; voice activity detection; wavelet shrinkage

### I. 서론

인간이 언어를 만든 이후 음성은 다른 사람에게 정보를 전달하는 직접적인 방법이 되었으며, 통신 네트워크에서 음성신호는 전화, 라디오, TV, 그리고 인터넷과 같은 분야로 서비스가 확장되고 있다. 그 결과 음성신호를 처리하기 위한 많은 연구가 이루어졌으며, 이런 알고리즘을 이용하여 음성 부호화기나 음성인식 시스템의 성능을 향상시키는 것이 음성 향상을 하는 중요한 목적이다.

전통적인 음성 향상 알고리즘인 파워 차감법은 DFT (Discrete Fourier Transform)를 이용하여 주파수 영역에서 적용되었다<sup>[1]</sup>. 그러나 이러한 전통적

인 음성 향상 방법은 잡음과 음성신호가 비정적 (non-stationary)일 경우에는 적용하기에 적합하지 않다. 이 문제점은 음성이나 잡음이 거의 정적이라고 가정할 수 있을 정도의 짧은 시간 구간으로 나누어서 처리하는 STFT (Short-Time Fourier Transform)을 사용해서 해결할 수 있었다<sup>[2]</sup>. 그렇지만 STFT은 시간-주파수 영역에서 각 해상도에 서로 상충 관계 (trade-off)가 존재하는 문제점이 있다. 즉, 나누는 시간 구간이 짧아질수록 시간 해상도는 높아지지만 주파수 해상도는 낮아지게 되어서 신호를 분해하는데 적당하지 않게 되며, 시간 구간이 길어질수록 주파수 해상도는 높아지게 되어서 신호를 보다 정확하게 분해할 수 있지만 신호가 비정적이 될 수 있

\* 고려대학교 전자및정보공학부 신호처리연구실 (ychlee@korea.ac.kr), \*\* LG전자 DMP연구소 Video 개발그룹 (ainayui@korea.ac.kr)

\*\*\* 고려대학교 전자및정보공학부 (sahn@korea.ac.kr)

논문번호: KICS2006-03-126, 접수일자: 2006년 3월 15일, 최종논문접수일자: 2006년 6월 16일

다. 또한 STFT 기반의 파워 차감법은 낮은 시간 해상도 때문에 발생하는 부정확한 잡음 추정으로 인해 원하지 않는 스펙트럼 편차가 존재하여 뮤지컬 잡음(musical noise)을 발생시키는 문제점이 있다<sup>3)</sup>. 이런 문제점들을 해결하기 위해 다중 대역 파워 차감법<sup>4)</sup>과 웨이블릿(Wavelet) 공간에서의 신호 처리 방법들이 사용되고 있다<sup>5)</sup>.

먼저 다중대역 파워 차감법은 잡음이 음성신호 스펙트럼의 각 대역마다 서로 다른 영향을 주기 때문에 각 대역에서 영향 받은 만큼의 잡음 스펙트럼을 차감해주는 방법이며, 이를 위해 각 대역마다 적절한 파라미터를 추정한다. 또한 스펙트럼 차감과정에서 잡음 추정 오류로 인해 음의 값을 갖는 스펙트럼 대역에는 0 또는 매우 작은 스펙트럼 바닥 인자(spectral floor factor)를 할당하여 뮤지컬 잡음의 발생을 최소화 하는 방법이다. 그리고 웨이블릿 공간에서 신호 처리하는 방법인 웨이블릿 분해는 비정적인 신호를 정적인 신호로 만들기 위한 창함수의 길이가 고정되어 있는 STFT과 달리 창함수 길이 선택에 많은 유연성을 갖고 있어서 음성신호와 같은 비정적인 신호를 처리하는데 있어 보다 더 효과적인 신호처리 방법이다.

웨이블릿의 종류에는 Biorthogonal, Daubecheis, Symlets 그리고 Coiflets과 같은 다양한 종류가 있으며<sup>6)</sup>, 웨이블릿 변환 후 얻어진 계수들 중 특정 값 이하의 작은 계수들을 제거하더라도 원 신호를 왜곡 없이 복원할 수 있는 기본적인 잡음 제거 특징을 내재하고 있으며, 이 특징을 이용한 여러 가지 웨이블릿 Shrinkage 방법들을 이용하여 음성 향상 및 잡음 제거의 효과를 얻을 수 있기 때문에 음성신호처리에 많이 사용되고 있다<sup>7,8)</sup>.

웨이블릿 Shrinkage는 웨이블릿 변환 후 음성신호의 특징이 낮은 주파수 대역에 많이 포함되어 있는 특징을 이용하여 높은 주파수 대역 내의 일정 크기 이하의 작은 값을 가지는 계수들을 제거하여 잡음을 제거하는 방법이다. 그렇지만 일반적인 웨이블릿 변환은 낮은 주파수 대역의 신호를 주된 관점으로 하고 있어서 높은 주파수 특성을 갖는 무성음에 대한 음성 향상에는 한계가 있다. 이러한 문제점을 해결하기 위해 WPD (Wavelet Packets Decomposition) 이 제안되었고<sup>9)</sup>, 이는 인간의 소리를 받아들이는 구조와 비슷한 Psycho-acoustic 모델을 기반으로 하는 PWPD (Perceptual Wavelet Packets Decomposition) 으로 발전되었다<sup>9)</sup>.

전처리 과정으로 다중 대역 파워 차감법을 사용

하여 SNR을 높이고 파워 차감법의 뮤지컬 잡음이 발생하는 문제점을 개선한 후, PWPD으로 분해한 각 밴드마다 무성음, 유성음 그리고 잡음의 형태를 구분하여 각 형태에 적합한 웨이블릿 Shrinkage 를 적용하여 무성음이 음성 향상 과정에서 제거되는 문제점을 개선한다.

음성과 잡음을 구분하기 위해서 에너지 레벨을 기준으로 하는 음성검출(Voice Activity Detection)의 도움을 받으며, PWPD 후 각 웨이블릿 노드마다의 엔트로피 비율을 기준으로 무성음과 유성음/잡음을 구분한다.

본 논문은 다음과 같이 구성되었다. 먼저 2장에서는 전통적인 음성 향상 방법인 파워 차감법의 문제점에 대해서 알아보고 3장에서는 웨이블릿 분해 방법과 웨이블릿 Shrinkage에 대해서 알아본다. 4장에서 기존의 음성 향상 방법들이 가지고 있던 문제점들을 해결한 개선된 음성 향상 알고리즘을 제안한다. 5장에서는 기존의 알고리즘과 제안한 알고리즘을 컴퓨터 시뮬레이션을 통하여 비교하며, 결론에서 그 결과를 분석하고 제안하는 알고리즘의 우수성을 검증한다.

## II. 전통적인 잡음 제거 방법

잡음 환경에서 음성인식을 목적으로 잡음을 제거하고 음성을 향상시키기 위한 방법이 지속적으로 연구되어 왔으며, 음성 향상은 음성인식은 물론 음성압축 분야에도 중요한 역할을 해왔다. 전통적인 음성 향상 방법에는 Wiener filtering, Magnitude subtraction, Spectral subtraction 등이 있으며, 이들 알고리즘은 STFT와 같은 선형적인 신호처리 과정을 이용한다<sup>10)</sup>. 이들 알고리즘 중 가장 간단하고 성능이 좋은 파워 차감법은 잡음이 섞인 음성신호의 스펙트럼에서 잡음의 스펙트럼을 제거함으로써 깨끗한 음성신호의 스펙트럼을 얻고자 하는 것이다. 깨끗한 음성신호  $s_n$ 에 잡음 신호  $v_n$ 가 섞인 신호를  $x_n$ 라 하면  $x_n$ 에 대해 STFT 계수는 다음과 같이 표현된다.

$$x_n = s_n + v_n \tag{1}$$

$$X_{n,k} = S_{n,k} + V_{n,k} \tag{2}$$

여기서  $S_{n,k}$ ,  $V_{n,k}$  그리고  $X_{n,k}$ 는 각각 음성신호와 잡음 신호 그리고 잡음이 섞인 음성신호의  $n$  번째 프레임의  $k$  번째 STFT 계수를 나타낸다. 이들 계수로부터 향상된 음성신호의 스펙트럼을 찾기 위

한 파워 차감법은 다음과 같다<sup>[1]</sup>.

$$|\hat{S}_{n,k}|^2 = |X_{n,k}|^2 - |\hat{V}_{n,k}|^2 \quad (3)$$

여기서 잡음의 파워  $|\hat{V}_{n,k}|^2$  는 직접적으로 얻어질 수 없으므로 음성신호가 존재하지 않는 구간에서 다음과 같은 수식으로 추정된다.

$$|\hat{V}_{n+1,k}|^2 = \mu |\hat{V}_{n,k}|^2 + (1 - \mu) |X_{n,k}|^2 \quad (4)$$

망각 계수  $\mu$ 는 일반적으로 0.90 ~ 0.95 값을 가지며, 비정적이고 시변 특성을 갖는 잡음 환경에서는 작은 값을 가질수록 좋은 잡음 추정 특성을 보인다. 잡음의 특성이 음성신호보다 느리게 변하는 특징이 있기 때문에 위와 같은 잡음 갱신 방법이 주로 사용되며, 추정된 음성신호의 파워에 신호의 위상정보를 주어 향상된 음성신호를 복원하면 다음과 같다.

$$\begin{aligned} \hat{S}_{n,k} &= |\hat{S}_{n,k}| \angle X_{n,k} \\ &= \sqrt{||X_{n,k}|^2 - |\hat{V}_{n,k}|^2|} \frac{X_{n,k}}{|X_{n,k}|} \end{aligned} \quad (5)$$

파워 차감법은 잡음 추정의 정확도에 많은 영향을 받게 되므로 시변 환경에서는 성능이 떨어지는 단점이 있으며, 스펙트럼 차감 과정에서 발생하는 매우 큰 값 (spectral peak)이나 작은 값 (spectral valley)의 영향으로 뮤지컬 잡음이 발생한다. 또한 큰 파워를 갖는 유성음의 경우 잘못 추정된 잡음에 의한 영향이 적지만 무성음이나 파워가 작은 유성음의 경우에는 적은 잡음 추정 오류에도 광대역 잡음에 묻혀서 쉽게 제거되는 문제점도 있다. 이 문제점들을 해결하기 위해 다중 대역 파워 차감법이 이용된다.

### III. 웨이블릿 변환

#### 3.1 웨이블릿 분해

Fourier 변환은 사용되는 기저 함수인 삼각함수의 특성상 국부적인 변화를 추출하기보다는 전반적인 주기성을 분석하기에 유용하기 때문에 시변 신호를 분석하는데 적당하지 않다. 웨이블릿 해석도 신호를 기저 함수들의 선형적인 결합으로 나타내는 분석 방법 [5]이지만, Fourier 해석과는 달리 다양한 형태의 기저함수를 사용할 수 있고 유동적인 창함수(window function)를 사용할 수 있기 때문에 신

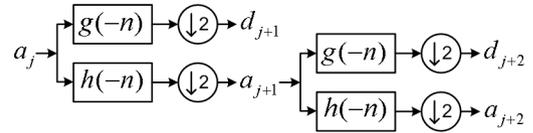


그림 1. 필터 बैं크 구조의 웨이블릿 분해 블록선도

호가 가지고 있는 비정적 성질들을 표현하는데 웨이블릿 해석이 유용한 특징을 가지고 있다.

STFT는 신호를 시간-주파수의 2차원 함수에 대응시킬 때 창함수의 길이에 많이 의존하며, 한번 그 길이가 결정되면 시간-주파수 해상도는 전체 시간-주파수에 대해 제한되고, 이러한 제한이 STFT를 이용하여 신호를 해석하는데 단점이 된다. 그래서 이 한계점을 극복하기 위해서 창함수의 길이를 다양하게 변화시킬 수 있는 유용성이 필요하게 되었으며, 웨이블릿은 높은 주파수에서는 짧은 길이의 창함수를 사용하고, 낮은 주파수에서는 긴 길이의 창함수를 사용하는 다중 해상도 특징을 가진 효과적인 신호 해석 방법이다. 그렇기 때문에 STFT으로 표현할 수 없었던 신호의 단절이나 불연속성을 표현할 수 있는 장점을 가지게 되었다.

웨이블릿 분해는 그림 1과 같이 저역통과 필터  $h(n)$ 과 고역통과 필터  $g(n)$ 으로 신호를 필터링하는 연속적인 2-채널 필터뱅크로 구현하며, 저역통과 필터와 고역통과 필터를 거친 후 down-sampling이 이루어진다 [5]. 이러한 필터뱅크 구조는 완벽한 재생 (perfect reconstruction)을 구현하는 대역분할 방법의 기본적인 구조이다.

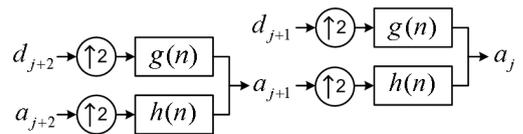


그림 2. 필터 बैं크 구조의 웨이블릿 합성 블록선도

여기서  $a_j$ 와  $d_j$ 는 각각  $j$ 번째 웨이블릿 분해 레벨의 계수들을 나타낸다. Down-sampling은 필터링을 거친 신호의 짝수 번째 성분만을 취하고 나머지는 버리는 과정을 수행하며, 데이터량이 증가되는 것을 막기 위해 사용된다. 또한 웨이블릿 합성은 그림 2와 같이 웨이블릿 분해의 역 과정을 수행한다.

#### 3.2 웨이블릿 Shrinkage

대부분의 음성 향상 알고리즘들은 신호의 분석을 위해 다중 해상도 분석 방법을 이용하며, 최근에는

비정적인 특성을 갖는 음성신호에 가장 적합한 특징이 있는 웨이블릿 분해 방법이 폭넓게 사용되고 있다.

웨이블릿을 이용한 가장 간단하고 효과적인 잡음 제거 방법이 웨이블릿 **Shrinkage** 이며, 웨이블릿 분해 후 생성된 웨이블릿 계수들 중 특정 값보다 작은 계수들을 잡음의 특성을 나타내는 계수라고 판단하여 0으로 만들어주는 기법을 말한다. 그리고 이 값의 기준을 정하는 방법에 따라서 **Hard**, **Soft**, **Semisoft thresholding** 과 같은 전통적인 방법들이 있으며 아래와 같은 변환 함수로 표현 된다<sup>[7]</sup>.

**Hard Thresholding:**

$$T\{\omega, \lambda_1\} = \begin{cases} \omega - \lambda_1 \text{sgn}(\omega), & |\omega| \geq \lambda_1 \\ 0, & \text{otherwise} \end{cases} \quad (6)$$

**Soft Thresholding:**

$$T\{\omega, \lambda_1\} = \begin{cases} (|\omega| - \lambda_1) \text{sgn}(\omega), & |\omega| \geq \lambda_1 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

**Semisoft Thresholding:**

$$T\{\omega, \lambda_1, \lambda_2\} = \begin{cases} 0, & \lambda_1 \geq |\omega| \\ \lambda_2 \frac{|\omega| - \lambda_1}{\lambda_2 - \lambda_1} \text{sgn}(\omega), & \lambda_1 < |\omega| < \lambda_2 \\ \omega, & \lambda_2 \leq |\omega| \end{cases} \quad (8)$$

$\omega$ 와  $T\{\}$ 는 각각 웨이블릿 계수와 웨이블릿 **Shrinkage** 변환 함수를 나타내며,  $\text{sgn}()$ 은 부호 함수를 나타낸다. 그리고 일반적으로 웨이블릿 계수들을 **thresholding** 하는 기준이 되는 문턱값들은 잡음의 표준편차  $\sigma$ 를 이용하여 각각  $\lambda_1 = \sigma\sqrt{2\log_{10}(N\log_2 N)}$ ,  $\lambda_2 = \lambda_1\sqrt{2}$  와 같이 계산된다. 여기서  $N$ 은 프레임에 포함된 샘플의 수를 나타낸다.

**Hard thresholding**은 문턱값이 클수록 계수간의 불연속성이 발생하여 신호의 분산이 커지고 원신호에 많은 왜곡을 가져오는 단점이 있으며, 이 문제점을 개선한 방법이 **Soft thresholding** 이지만 여전히 원신호의 왜곡이 존재한다. 이를 개선하기 위해 두 방법을 절충한 방법이 두 개의 문턱값을 이용하는 **Semisoft thresholding** 방법이며, 두 문턱값 사이에서 부드러운 변화를 가져서 **Hard thresholding** 방법의 불연속성을 피할 수 있으며 복원 후 원신호의 왜곡을 줄일 수 있는 특징이 있다.

웨이블릿 **Shrinkage** 방법을 이용한 음성 향상 방법은 간단하지만 개선해야 할 중요한 문제점이 있

다. 음성신호는 시간에 따라 크게 변화하는 특징이 있기 때문에 동일한 문턱값을 기준으로 **thresholding** 을 적용해서는 음성신호의 왜곡을 피할 수 없다. 즉 모든 주파수 대역에서 동일한 문턱값을 사용할 경우 잡음을 제거할 뿐만 아니라 잡음과 비슷한 특성을 갖는 무성음과 파위가 작은 유성음도 함께 제거되기 때문에 실제적인 잡음 환경에 적용하기에는 한계가 있다. 이러한 문제점을 개선하기 위해서 각 주파수 대역마다 독립적인 문턱값을 가지도록 하는 방법을 사용한다.

**3.3 Perceptual 웨이블릿 패킷 분해(PWPD)**

일반적인 웨이블릿 분해/합성 방법은 그림 1과 그림 2처럼 2-채널 필터뱅크 구조로 구현되며 각 레벨마다 저역통과 분기(branch)에 대해서만 연속적으로 필터 뱅크가 적용된다. 그러나 이런 웨이블릿 분해 방법은 로그 급수 해상도를 갖기 때문에 고주파 영역의 해상도가 낮아서 음성신호에 적용하기에 적당하지 않다. 이러한 문제점을 해결하기 위해서 그림 3과 그림 4에서와 같이 저역통과 분기뿐만 아니라 고역통과 분기에도 2-채널 필터 뱅크 구조를 반복 적용한다. 이러한 웨이블릿 분해 방법이 웨이블릿 패킷 분해이다 [5] [6]. 웨이블릿 패킷 분해는 주파수 대역을 균등하게 분해하여 낮은 주파수 대역은 물론 높은 주파수 대역까지 높은 분해능을 갖게 하여 신호를 더 효과적으로 분해할 수 있다.

위 그림에서  $j$ 와  $l$ 은 각각 웨이블릿 분해 레벨

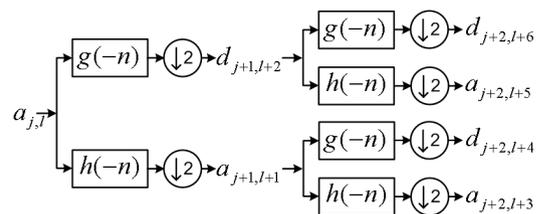


그림 3. 2-레벨 웨이블릿 패킷 분해 블록선도

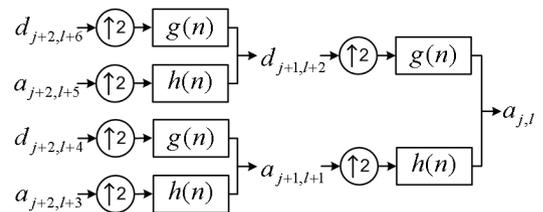


그림 4. 2-레벨 웨이블릿 패킷 합성 블록선도

### IV. 개선된 음성 향상 방법

#### 4.1 다중 대역 파워 차감법

전통적인 음성 향상 방법인 파워 차감법은 음성 향상 처리 후 뮤지컬 잡음이 발생하게 된다. 또한 시변 잡음 환경에서 성능이 떨어지고 향상된 음성이 많이 왜곡되는 문제점이 있다. 이러한 문제점들을 개선하기 위해 다중 대역 파워 차감법을 이용한다. 실제 환경에서의 대부분의 잡음은 주파수 공간에서 모든 대역에 동일하게 영향을 주는 형태가 아닌 특정 주파수 대역에 영향을 주는 형태 (colored noise)이다. 이와 같은 특징 때문에 전통적인 파워 차감법의 경우 모든 주파수 대역에 동일한 가중치를 가지고 추정된 잡음을 차감하게 되어서 음성신호에 왜곡이 생기고 뮤지컬 잡음이 발생한다. 이러한 문제점은 각 주파수 대역별로 잡음의 특성에 따른 가중치를 기준으로 추정된 잡음을 차감함으로써 해결할 수 있다. 본 논문에서는, 유선전화 채널에서와 같이 음성신호가 주로 300~3400 Hz 대역에서 존재하고 3~4개의 대역에서 파워 집중대 (formant)가 형성되는 점을 반영하여 PWPD 대역을 기준으로 표 1과 같이 전체 주파수 대역을 다섯 개의 다중 대역으로 나눈다<sup>[13]</sup>.

각 대역마다 서로 다른 가중치를 사용하는 다중 대역 파워 차감법은 다음과 같다.

$$|\hat{S}_{n,k}^i|^2 = |X_{n,k}^i|^2 - \alpha^i |\hat{V}_{n,k}^i|^2 \tag{9}$$

여기서  $i$ 번째 다중 주파수 대역의 과차감 인자 (over-subtraction factor)  $\alpha^i$ 는 해당 주파수 대역마다의 SNR을 기준으로 다음 식과 같이 계산된다.

$$\alpha^i = \begin{cases} 4.75, & -5 \geq \zeta^i \\ A(\zeta^i - 15)^2 + 1, & -5 < \zeta^i \leq 15 \\ 1.00, & 15 < \zeta^i \end{cases} \tag{10}$$

표 1. 각 다중 대역 밴드의 주파수 범위

다중 밴드 인덱스	주파수 대역 [Hz]
Band #1	0 ~ 250
Band #2	250 ~ 750
Band #3	750 ~ 2000
Band #4	2000 ~ 3000
Band #5	3000 ~ 4000

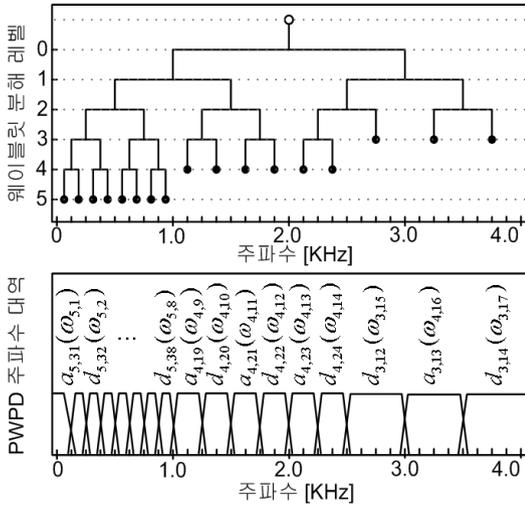


그림 5. PWPD 의 구조와 주파수 대역

과 웨이블릿 노드 번호를 나타낸다. 웨이블릿 패킷 분해 방법은 주파수 공간을 동일한 대역으로 나누고 있고 서로 직교하는 기저함수를 사용한다는 특징 때문에 STFT을 근사함은 물론 동일한 계산량을 갖는 특징이 있어서 많은 신호처리 알고리즘에서 STFT을 대신해서 사용되기도 한다. 그러나 웨이블릿 패킷 분해 방법도 음성신호를 효과적으로 표현하기에는 부족한 면이 있고 이 부족한 면을 채워주는 방법이 PWPD 이다. PWPD 은 2-채널 필터뱅크 구조를 인간이 소리를 받아들이는 구조와 비슷한 Psycho-acoustic 모델에 적용한 웨이블릿 분해 방법이다. 표본화 주파수가 8 KHz 일 때 PWPD 으로 분해한 17 개의 주파수 대역은 그림 5와 같다<sup>[11]</sup>. 이러한 주파수 대역들은 신호의 주된 주파수 성분들이 집중되어 있는 대역 (critical band)을 기준으로 나누어지며 가청주파수 대역에서 Bark scale을 이용하여 계산한다<sup>[12]</sup>.

그림 5와 같이 신호는 PWPD 으로 17 개의 노드로 분해되어지며 각 노드는 웨이블릿 계수  $w_{j,l}(k)$ 로 이루어져있다. 여기서  $w_{j,l}(k)$ 는  $j$ 번째 웨이블릿 분해 레벨의  $l$ 번째 웨이블릿 노드의  $k$ 번째 계수를 나타낸다. PWPD에서의 주파수 대역을 보면 음성 신호가 많이 포함되어 있는 노드에 높은 해상도를 부여하여 WPD 보다 분해하는 가지가 적어서 계산량을 줄일 수 있으면서도 음성 신호에 대한 분석을 효과적으로 할 수 있는 것이 특징이다.

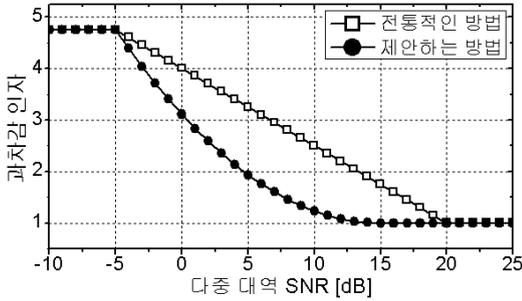


그림 6. 과차감 인자 선택 곡선

여기서  $\zeta^i$ 는  $i$ 번째 주파수 대역의 SNR을 나타내며  $A = 0.009375$ 이다.

전통적인 다중 대역 파워 차감법에서 사용하는 선형적인 과차감 인자 선택 방법은 15dB 이상의 높은 SNR 환경에서도 추정된 잡음을 과도하게 차감하기 때문에 특히 무성음과 같은 잡음과 유사한 음성 신호 성분에 왜곡을 가져오는 문제점이 있었다<sup>[14]</sup>. 이 문제점을 개선하기 위해 그림 6과 같은 비선형적인 과차감 인자 선택 방법을 제안한다.

높은 SNR 환경에서는 잡음 추정이 용이하여 전통적인 파워 차감법에 의한 음성 향상 성능이 좋기 때문에 다중 대역 SNR이 상대적으로 높은 대역에서는 전통적인 파워 차감법과 동일한 방법을 사용하고, 다중 대역 SNR이 낮은 대역에서는 추정된 잡음 보다 더 많은 양을 빼주도록 하여 음성신호가 주된 주파수 대역에서 음성신호의 왜곡을 최소화하고 잡음을 많이 제거하는 효과를 얻을 수 있다. 다중 대역 파워 차감법을 이용해서 향상된 음성신호는 다음과 같다.

$$\hat{S}_{n,k}^i = |\hat{S}_{n,k}^i| \angle X_{n,k}^i \quad (11)$$

$$= \frac{\sqrt{|X_{n,k}^i|^2 - \alpha^i |\hat{V}_{n,k}^i|^2}}{|X_{n,k}^i|} X_{n,k}^i$$

여기서 제공근 내부의  $|X_{n,k}^i|^2 - \alpha^i |\hat{V}_{n,k}^i|^2$  가 음의 값을 갖는 경우 뮤지컬 잡음이 발생되는 원인이 되므로 다음과 같이 수정한다.

$$|\hat{S}_{n,k}^i|^2 = |X_{n,k}^i|^2 - \alpha^i |\hat{V}_{n,k}^i|^2$$

$$= \begin{cases} |X_{n,k}^i|^2 - \alpha^i |\hat{V}_{n,k}^i|^2, & |X_{n,k}^i|^2 - \alpha^i |\hat{V}_{n,k}^i|^2 > 0 \\ \beta |X_{n,k}^i|^2, & \text{otherwise} \end{cases} \quad (12)$$

여기서  $\beta$ 는 스펙트럼 바닥 인자를 나타낸다.  $\alpha$ 와  $\beta$ 를 선택함에 있어서 음성신호의 왜곡과 뮤지컬 잡음의 제거 사이에는 상충관계가 존재한다. 즉,  $\alpha$ 가 커질수록 뮤지컬 잡음 및 광대역 (broad band) 잡음 제거의 효과가 큰 반면 유성음 성분이 제거되는 양이 많아져서 음성신호의 왜곡을 가져오게 된다. 또한  $\beta$ 의 값이 커질수록 뮤지컬 잡음이 제거되는 효과는 커지지만 광대역 잡음 제거 효과가 줄어드는 특징이 있다. 따라서 음성신호의 왜곡 정도와 뮤지컬 잡음의 제거 성능을 고려하여 실험적으로 선택한 스펙트럼 바닥 인자  $\beta$ 는 0.0024이다.

#### 4.2 무성음/유성음/잡음 구분 방법

기존의 음성 향상 알고리즘에서는 음성 향상 후 잡음과 함께 무성음도 함께 제거되는 문제점이 있었다. 이는 잡음과 무성음의 특징이 비슷하기 때문이며, 이 문제점을 개선하기 위해 본 논문에서는 무성음과 유성음 그리고 잡음을 구분하여 각 특성에 맞는 음성 향상 처리를 한다.

먼저 무성음과 유성음/잡음을 구분하기 위해서 각 웨이블릿 노드마다의 엔트로피  $E_{j,l}$ 를 측정하여 아래의 수식과 같이 무성음 여부를 판단 한다<sup>[15]</sup>.

$$E_{j,l} = - \sum_k \omega_{j,l}^2(k) \log_{10}(\omega_{j,l}^2(k)) \quad (13)$$

$$\frac{E_{j,l}}{\max(E_j)} = \begin{cases} 0.1 \sim 0.9, & \text{무성음} \\ \text{otherwise}, & \text{유성음/잡음} \end{cases} \quad (14)$$

여기서  $\max(\ )$ 는 최대값 함수를 나타내며,  $E_{j,l}$ 은  $j$ 번째 웨이블릿 레벨의  $l$ 번째 노드의 엔트로피를 의미한다. 이와 같은 방법으로 구분된 무성음 그리고 유성음/잡음에서 다시 유성음과 잡음을 구분하기 위해 에너지 레벨 기반 음성검출을 이용한다. 이를 위해 각 프레임마다의 에너지  $A_n$ 을 자기 상관 함수 값을 이용하여 다음 식과 같이 계산한다.

$$A_n = R(0) + \sum_{i=1}^p R(i)^2 \quad (15)$$

여기서  $p$ 는 LPC(linear prediction coding) 필터의 차수를 나타내며,  $R(i)$ 는  $i$ 번째 자기 상관 값을 나타낸다. 자기상관의 첫 번째 값인  $R(0)$ 는 신호의 에너지를 가리키며, 보다 정확한 유성음/잡음

결정을 내리기 위해서 나머지 자기 상관 값의 제곱의 합을 이용하여 해당 프레임의 에너지를 표현하였다. 음성신호는 상대적으로 상관관계가 크기 때문에  $\Lambda_n$ 은 잡음만 존재하는 구간보다 음성신호가 존재하는 구간에서 더 큰 값을 갖게 되어 유성음과 잡음을 구분하는 기준이 될 수 있다. 유성음/잡음 구분을 위한 기준 문턱값을 찾기 위해 약 1 초 정도의 초기화 구간동안에는 잡음만 존재한다고 가정하고, 이 구간동안의 에너지의 평균  $\bar{\Lambda}$ 과 표준편차  $\sigma_{\Lambda}$ 를 이용하여 문턱값  $\eta = \bar{\Lambda} + \gamma\sigma_{\Lambda}$ 를 얻는다. 이를 기준으로 다음과 같이 유성음과 잡음을 구분한다.

$$\Lambda_n = R(0) + \sum_{i=1}^p R(i)^2 \begin{matrix} > \eta & \text{유성음} \\ < \eta & \text{잡음} \end{matrix} \quad (16)$$

여기서 문턱값 가중인자  $\gamma$ 는 시변 잡음 환경에서는 큰 값을 가지며, 시불변 잡음 환경에서는 작은 값을 가져서 유성음과 잡음을 구분하는 성능을 결정짓는 변수로 사용한다.

음성 검출의 성능에 따라서 잡음 추정 성능이 결정되고 이는 음성향상 결과에 직접적인 영향을 미치게 된다. 따라서 낮은 SNR 환경에서도 음성 검출의 성능을 향상시키기 위한 연구 [16]가 이루어지고 있지만, 본 논문에서 제안하는 음성향상 방법은 5 dB 이상의 높은 SNR 환경을 목표로 하고 있기 때문에 식 (16)의 에너지 레벨에 근거한 음성 검출 방법으로도 충분한 성능을 얻을 수 있다.

### 4.3 적응 망각 계수를 이용한 잡음 추정 방법

잡음의 파워를 추정하기 위해 잡음이 음성보다 변화율이 적은 특징을 이용해서 음성신호가 존재하지 않는 구간에서 식 (4)와 같이 잡음의 파워를 갱신한다. 그러나 이 잡음 추정 방법은 음성 검출의 성능에 큰 영향을 받기 때문에 음성 구간을 잡음 구간이라고 잘못 결정내린 구간에서 잡음 갱신이 일어나게 되면 잡음 추정 값에는 큰 오류가 발생하게 된다. 이와 같은 문제점을 개선하기 위해서 잡음을 갱신하는 망각 계수  $\mu$ 를 주어진 잡음 환경에 맞게 그림 7과 같이 적응적으로 선택하여 사용한다. 이를 통해 SNR이 높은 주파수 대역의 경우 음성 성분이 많이 포함되어 있기 때문에 큰 망각 계수 값을 가지도록 하여 음성 성분이 잡음 추정 과정에

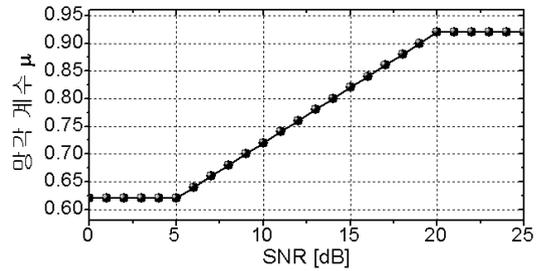


그림 7. 적응 망각 계수 선택 곡선

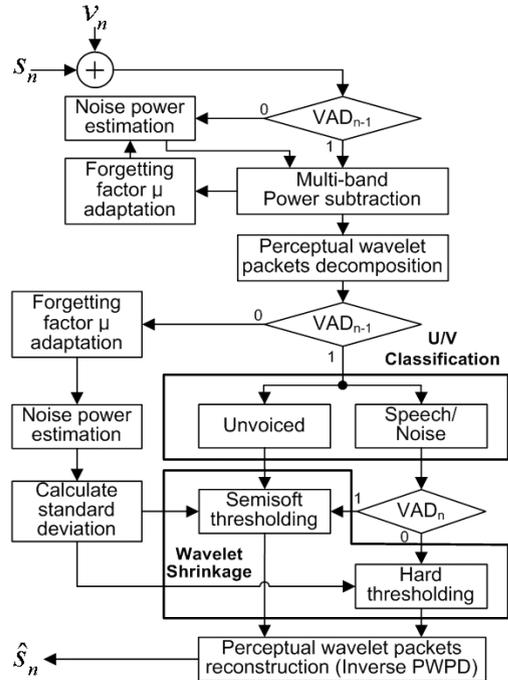


그림 8. 제안하는 개선된 음성 향상 방법

포함되는 양을 최소화 할 수 있으며, 반대로 SNR이 낮은 주파수 대역의 경우 잡음 신호가 우세하기 때문에 작은 망각 계수 값을 가지도록 하여 잡음 추정과정에 현재의 잡음 성분을 많이 포함하도록 하는 것이다. 이러한 적응 망각 계수 선택 방법을 이용한 잡음 추정 방법을 이용하여 제한적인 음성 검출 성능으로 인한 음성 향상 알고리즘의 성능 저하를 방지할 수 있다.

### 4.4 개선된 음성 향상 방법

기존의 음성 향상 방법들은 잡음 제거 시 무성음도 함께 제거되는 문제점이 존재한다. 이는 복원된 음성신호의 명료도가 낮아져서 음성 인식과 같은 음성신호처리 시스템에서 음성 인식 실패율을 높게

만들어서 시스템의 성능이 떨어지는 원인이 된다. 이같은 문제점을 해결하기 위해 본 논문에서는 그림 8과 같은 개선된 음성 향상 알고리즘을 제안한다. 전처리 과정에서 Fourier 주파수 공간의 신호를 다섯 개의 다중 대역으로 나눈 후, 각 대역마다의 SNR 을 기준으로 망각 계수를 선택하고 잡음의 파워를 갱신한다. 그리고 이 잡음 파워를 다중 대역 파워 차감법에 적용하여 각 대역마다의 과차감 인자를 선택하여 다중 대역 파워 차감법을 수행한다. 이를 통해 신호의 SNR을 증가시키고 광대역 잡음은 제거하여 뮤지컬 잡음의 발생을 최소화 한다. 전처리 과정 후 PWPD를 이용하여 신호를 17 개의 웨이블릿 노드로 분해하고 웨이블릿 thresholding 적용을 위한 문턱값  $\lambda_{j,l}$  을 다음과 같이 계산한다.

$$\lambda_{j,l} = \sigma_{j,l} \sqrt{2 \log_{10}(N \log_2 N)} \quad (17)$$

여기서  $\sigma_{j,l}$ 은  $j$ 번째 웨이블릿 분해 레벨의  $l$ 번째 노드에서 추정된 잡음의 표준편차를 나타내며,  $N$ 은 프레임에 포함되어 있는 신호의 샘플 개수를 나타낸다. 각 웨이블릿 노드마다의 엔트로피 비율을 기준으로 무성음과 유성음/잡음을 구분하고 다시 음성 검출의 도움을 받아 유성음과 잡음을 구분한다. 그리고 위에서 계산된 문턱값  $\lambda_{j,l}$ 을 기준으로 유성음과 무성음에는 Semisoft thresholding을, 잡음에는 Hard thresholding을 적용하여 전처리 과정 이후의 잔여 잡음을 제거한다. 이 과정에서 무성음이 잡음과 함께 제거되는 오류를 최소화할 수 있다.

### V. 컴퓨터 시뮬레이션

제안한 음성 향상 알고리즘의 성능을 분석하기 위해서 시뮬레이션에 사용한 음성신호는 SiPro Lab [17] 데이터베이스의 미국인 남성의 데이터이며, 잡음 신호는 NOISEX-92 DB [18]의 White, Babble, Vehicular (Volvo car) 그리고 Factory 데이터이다. 이 신호들은 모두 8 KHz 표본화 주파수로 16 비트 해상도를 가지도록 양자화 시켜서 사용하였다. 또한 제안하는 음성 향상 방법은 음성 인식이나 음성 압축과 같은 높은 SNR을 요구하는 응용분야에 적용하기 위한 알고리즘이므로 적어도 5 dB 이상의 SNR 환경이라고 가정하였다. 신호를 프레임으로 나누기 위해 240 길이를 갖는 Hamming 창함수를 사

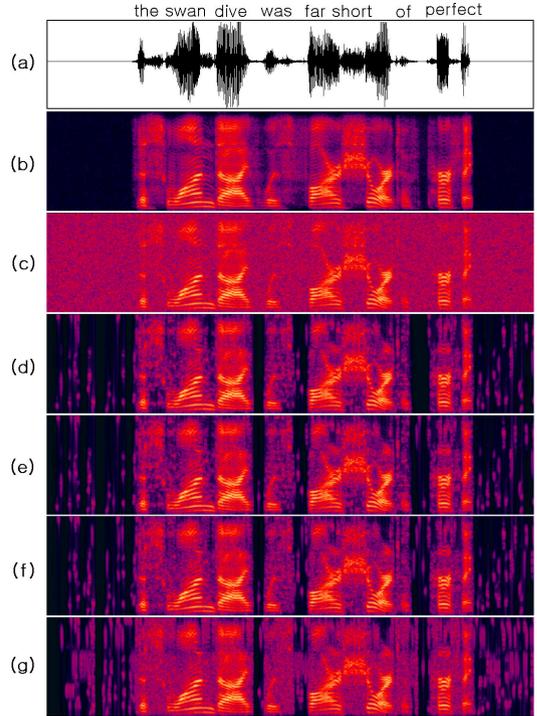


그림 9. 스펙트럼 바닥 인자  $\beta$ 에 따른 다중 대역 파워 차감 후 스펙트럼 (a) 깨끗한 음성신호의 파형 (b) 깨끗한 음성신호의 스펙트럼 (c) 잡음 섞인 음성신호의 스펙트럼 (d)  $\beta=0.0$ , (e)  $\beta=0.0024$ , (f)  $\beta=0.05$ , (g)  $\beta=0.50$

하였으며, 음성 검출에 사용되는 문턱값을 계산하기 위해서 128 프레임 길이의 초기화 구간을 적용하였다.

#### 5.1 스펙트럼 바닥 인자 선택

그림 9는 5 dB White 잡음 환경에서 여러 스펙트럼 바닥 인자  $\beta$ 를 이용하여 다중 대역 파워 차감법을 수행한 후 향상된 신호의 스펙트럼을 보여주고 있다.

작은  $\beta$ 를 가질수록 광대역 잡음은 줄어들지만 좁은 뮤지컬 잡음의 원인이 되는 시간-주파수 영역을 갖는 스펙트럼 성분들이 남아 있으며 음성신호의 왜곡도 심해지는 것을 확인할 수 있다. 또한  $\beta$ 가 커질수록 광대역 잡음이 제거되는 성능이 떨어지는 것을 확인할 수 있다.  $\beta$ 를 0으로 사용하는 기존의 파워 차감법에 비해 0.0024를 사용하는 제안하는 방법이 음성신호의 왜곡이 적은 상태로 광대역 잡음을 줄이게 되어서 음성 향상에 도움이 되는 것을 기대할 수 있다.

최적의 스펙트럼 바닥 인자  $\beta$ 를 찾기 위해 8

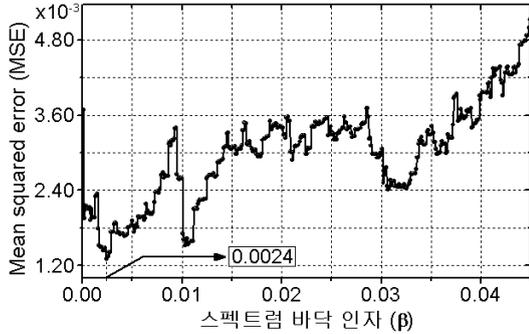


그림 10. 스펙트럼 바닥 인자  $\beta$ 에 따른 다중 대역 파워 차감 후 MSE 값

dB White 잡음 환경에서 다양한  $\beta$  값을 적용하여 다중 대역 파워 차감법을 시행한 후 MSE (mean squared error)를 계산한 결과는 그림 10과 같다.

그림 10에서 스펙트럼 바닥 인자  $\beta$ 가 작아질수록 MSE가 작아져서 원신호의 왜곡이 줄어드는 것을 확인할 수 있지만  $\beta$ 가 매우 작을 경우 오히려 MSE가 증가하는 것을 확인할 수 있다. 따라서 본 논문에서는  $0 < \beta \leq 1$  인 구간에서 0.0024 값을 선택하였다.

### 5.2 웨이블릿 기저 함수 선택

웨이블릿 기반 신호처리를 위한 웨이블릿 기저 함수(웨이블릿 필터)의 선택은 필터뱅크의 시간-주파수 해상도를 결정짓는 중요한 역할을 하며, 필터의 길이와 계산량은 직접적인 연관이 있어서 시간-주파수 해상도와 계산량 사이에는 서로 상충관계가 존재한다. 즉, 긴 길이의 필터를 사용함으로써 좋은 분해 해상도를 얻을 수 있지만 계산량이 증가되는 문제점이 존재한다. 제안한 음성 향상 알고리즘에 적용하기 위한 웨이블릿 기저 함수를 선택하기 위해 8 dB SNR을 가지는 여러 잡음 환경에서 다양한 길이의 웨이블릿 필터를 사용하여 프로세싱 시간과 MSE를 비교하고 이 결과를 표 2에 나타냈다. 여러 웨이블릿 기저 함수들 중 음성신호에 가장 적합한 특성을 갖는 Daubechies 기저 함수가 시뮬레이션에 사용되었다.

표 2를 보면 10차 Daubechies 기저 함수가 8차인 경우와 비교해서 평균 프로세싱 시간과 MSE에는 큰 차이가 없지만 12차인 경우의 평균 프로세싱 시간보다 훨씬 작은 것을 확인할 수 있다. 따라서 위 결과를 근거로 본 논문에서는 웨이블릿 필터로 10차 Daubechies 기저 함수를 선택하였다.

표 2. 웨이블릿 기저 함수 선택을 위한 실험 결과

Noise	Daubechies 8		Daubechies 10		Daubechies 12	
	프로세싱 시간(s)	MSE	프로세싱 시간(s)	MSE	프로세싱 시간(s)	MSE
White	65.656	0.00210	68.297	0.00197	193.56	0.00799
Babble	54.375	0.00162	56.219	0.00196	180.47	0.00165
Volvo	54.453	0.01691	56.391	0.01641	181.95	0.01650
Factory	54.250	0.00072	56.375	0.00076	181.56	0.00086
평균	56.289	0.00366	58.451	0.00362	183.76	0.00359

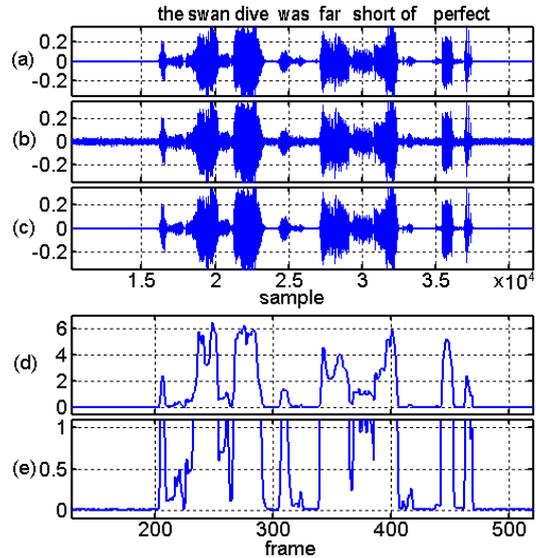


그림 11. 엔트로피 값을 이용한 무성음 검출 과정 (a) 깨끗한 음성신호 (b) 잡음섞인 음성신호 (c) 전처리과정 후 신호 (d) 전처리과정 후 신호의 엔트로피 비율 (e) 엔트로피 비율의 0.0 ~ 1.0 부분 확대

### 5.3 무성음과 유성음/잡음의 구분

그림 11은 10 dB White 잡음 환경에서 웨이블릿 계수의 엔트로피 비율을 이용하여 무성음과 유성음/잡음을 구분하는 방법을 시뮬레이션 한 결과이다. 이 그림을 보면 엔트로피 비율이 0.1 이하인 경우에 대부분의 잡음이 포함되어 있으며, 1.0 이상의 큰 엔트로피 비율을 갖는 부분에 유성음이 포함되어 있는 것을 확인할 수 있다. 따라서 0.1 ~ 0.9 사이의 엔트로피 비율을 가질 경우 무성음이나 파워가 작은 유성음일 가능성이 크므로 이 값을 기준으로 식 (14)와 같이 무성음과 유성음/잡음을 구분할 수 있다.

### 5.4 적응 망각 계수 선택

그림 12 와 13은 8 dB Babble 잡음 환경에서 고정된 망각 계수를 이용한 경우와 본 논문에서 제

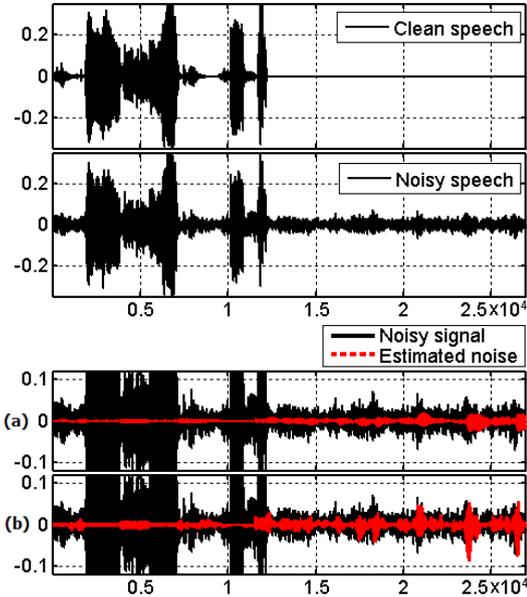


그림 12. 잡음이 섞인 신호와 추정된 잡음의 비교  
(a) 고정된 망각 계수를 이용하여 추정된 잡음 (b) 적응 망각 계수를 이용하여 추정된 잡음

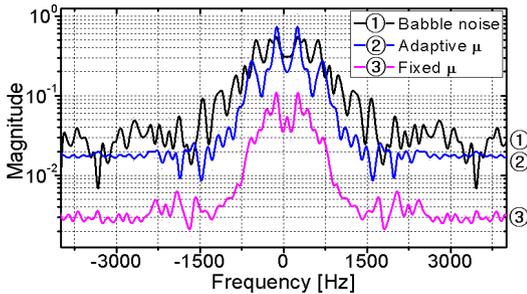


그림 13. 추정된 잡음의 FFT 공간 스펙트럼  
(a) 잡음 신호 (b) 적응 망각 계수를 이용하여 추정된 잡음  
(c) 고정된 망각 계수를 이용하여 추정된 잡음

안하고 있는 적응 망각 계수를 이용한 경우의 잡음 파워 추정 결과를 비교하고 있다.

그림 12의 (a)와 (b)를 비교해보면 적응 망각 계수를 적용하여 잡음 추정을 했을 때 그렇지 않았을 때보다 Babble 잡음과 같은 비정적인 잡음을 잘 추적하고 있음을 확인할 수 있으며, 그림 13으로 부터 적응 망각 계수를 이용하면 잡음의 스펙트럼을 보다 정확하게 추정할 수 있음을 확인할 수 있다. 한편, 그림 14에서 음성 향상 후의 MSE 결과를 다양한 SNR 환경에서 비교하였으며, 고정된 망각 계수를 이용했을 경우보다 SNR을 기준으로 적응적으로 선택된 망각 계수를 이용했을 경우 음성 향상 후 MSE가 더 작은 것을 확인할 수 있다.

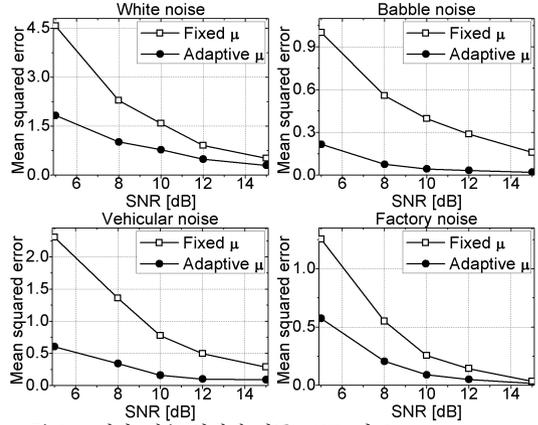


그림 14. 망각 계수 선택에 따른 MSE 비교

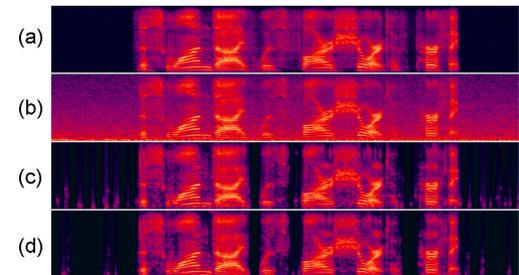
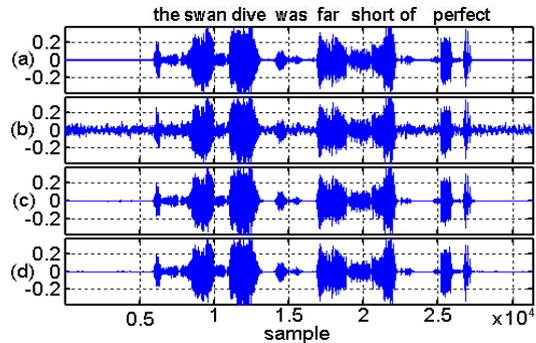


그림 15. 망각 계수의 종류에 따른 음성 향상된 신호의 파형과 스펙트럼  
(a) 깨끗한 음성신호 (b) 잡음이 섞인 신호 (c) 고정된 망각 계수를 이용한 경우 (d) 제안한 망각 계수를 이용한 경우

그림 15는 8 dB Factory 잡음 환경에서 고정된 망각 계수( $\mu = 0.89$ )와 적응 망각 계수를 사용했을 경우의 향상된 음성신호의 파형과 스펙트럼을 보여 주고 있다.

위 그림에서 SNR을 기준으로 적응 망각 계수를 선택하여 잡음 추정과정에 사용하는 제안한 방법이 전처리 과정 이후의 향상된 음성신호에 광대역 잡음이 더 많이 줄어들었을 뿐만 아니라 파워가 작은 유성음 성분도 덜 제거되고, 뮤지컬 잡음도 적게 발생하는 것을 확인할 수 있다.

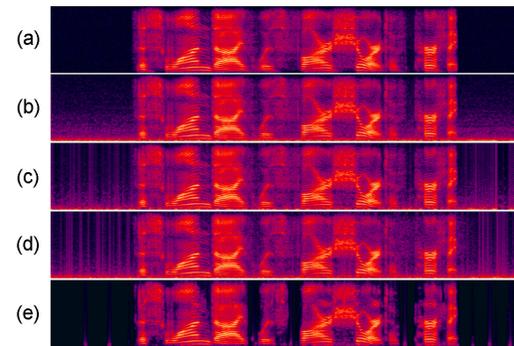
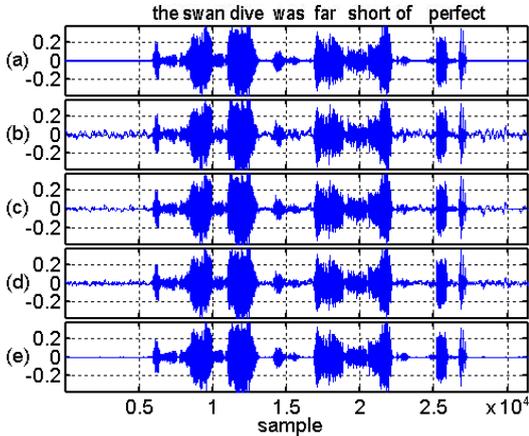


그림 16. 음성 향상 알고리즘들의 성능 비교  
(a) 깨끗한 음성신호 (b) 잡음섞인 음성신호 (c) 파워 차감법 적용 (d) 웨이블릿 기반 스펙트럼 차감법 적용 (e) 제안한 음성 향상 방법 적용

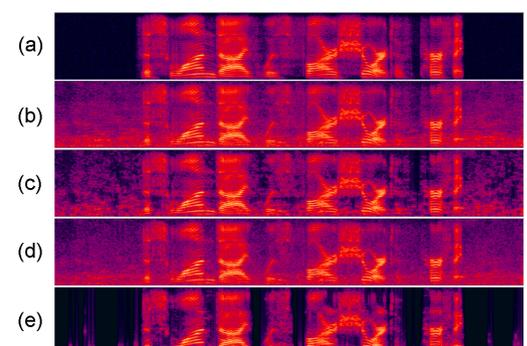
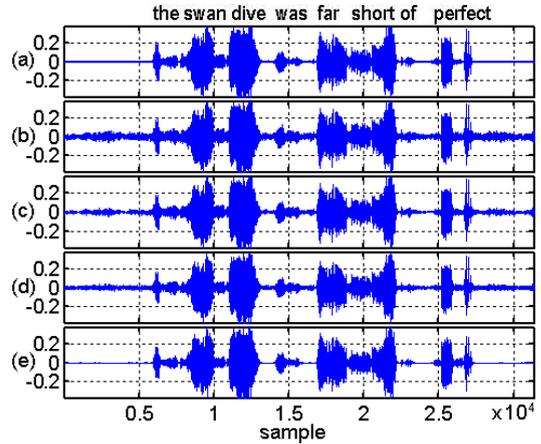


그림 17. 음성 향상 알고리즘들의 성능 비교  
(a) 깨끗한 음성신호 (b) 잡음섞인 음성신호 (c) 파워 차감법 사용 결과 (d) 웨이블릿 기반 스펙트럼 차감법 사용 결과 (e) 제안한 음성 향상 방법 사용 결과

### 5.5 개선된 음성 향상 방법

그림 16은 제안한 음성 향상 알고리즘의 성능을 평가하기 위해 기존의 여러 알고리즘들과 음성 향상 처리 후 파형을 비교한 결과이다. 대표적인 시불변 잡음이며서 낮은 주파수 대역에 에너지의 대부분을 포함하고 있는 Vehicular 잡음이 사용되었으며 8 dB SNR 환경이다. 비교에는 STFT을 이용하는 전통적인 파워 차감법 [10]과 웨이블릿 기반 스펙트럼 차감법 [19]이 사용되었다.

그림 16의 (c)에서 전통적인 파워 차감법을 적용한 경우 잡음 구간에서 뮤지컬 잡음이 존재하며, 웨이블릿 기반 스펙트럼 차감법을 적용한 (d)에서도 잡음 구간에 광대역 잡음이 많이 남아있고 뮤지컬 잡음이 존재하는 것을 확인할 수 있다. 그렇지만 제안한 음성 향상 알고리즘을 적용한 (e)의 경우 광대역 잡음이 대부분 제거되었고 뮤지컬 잡음도 발생되지 않았음을 확인할 수 있다.

그림 17은 8 dB Babble 잡음 환경에서의 여러 음성 향상 알고리즘과 제안한 알고리즘의 성능을

비교한 결과이다. Babble 잡음은 대표적인 시변 잡음이며, 주 음성신호에 많은 주변의 음성신호들이 더해진 형태이기 때문에 원 음성과 잡음을 구분하기 어려워 음성신호처리 분야에서 까다로운 잡음 중 하나이다.

그림 17의 (c)에서 파워 차감법은 잡음 구간에서 광대역 잡음을 제거하지 못할 뿐만 아니라 많은 뮤지컬 잡음이 생성되는 것을 볼 수 있으며, (d)에서 웨이블릿 기반 스펙트럼 차감법도 역시 파워 차감법의 경우보다는 적지만 잡음 구간에서 광대역 잡음과 뮤지컬 잡음이 남아 있는 것을 확인할 수 있다. 그러나 (e)에서 볼 수 있듯이 기존의 음성 향상 알고리즘보다 제안한 알고리즘이 잡음 구간에서 광대역 잡음 및 뮤지컬 잡음을 제거하는 성능이 더 뛰어나며 무성음은 물론 파워가 작은 유성음 부분이 왜곡되거나 제거되지 않고 여전히 남아있는 것을 확인할 수 있다. 이것은 제안한 다중 대역 파워 차감법에서 주변 환경에 맞는 적절한 과차감 인자

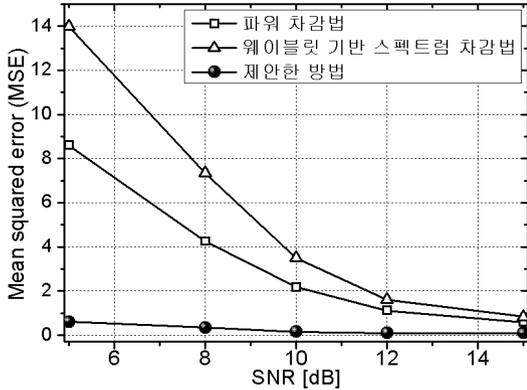


그림 18. Vehicular 잡음 환경에서의 MSE 성능 비교

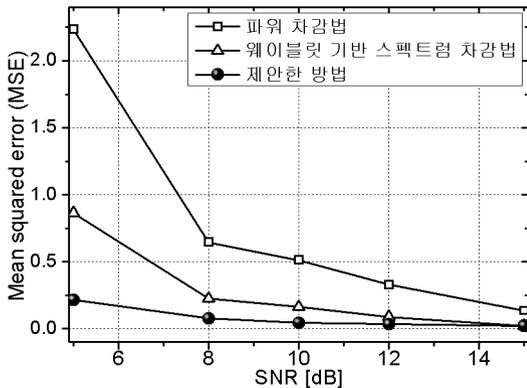


그림 19. Babble 잡음 환경에서의 MSE 성능 비교

와 스펙트럼 바다 인자를 선택했기 때문이다.

그림 18과 그림 19는 각각 Vehicular 잡음과 Babble 잡음 환경에서 제안한 음성 향상 알고리즘을 Fourier 공간에서 STFT을 사용하는 파워 차감법 그리고 웨이블릿 공간에서 웨이블릿 패킷 분해를 사용하는 웨이블릿 기반 스펙트럼 차감법과 성능을 비교하기 위해서 알고리즘 적용 후 MSE를 비교한 결과이다.

위 그림에서 제안한 음성향상 방법이 기존의 방법보다 MSE가 크게 줄어들어서 성능이 개선되었음을 확인할 수 있다. 또한 낮은 주파수 대역에 많은 에너지를 가지고 있는 Vehicular 잡음의 경우 다중 대역 파워 차감법에서 Band #1 대역을 DC 신호라고 판단하여 제거했기 때문에 기존의 방법들보다 더욱더 많은 성능개선을 볼 수 있었으며, Babble 잡음의 경우 음성신호와 비슷한 특징 때문에 음성 검출의 성능도 떨어지고 이로 인해 잡음 추정 오류가 많아서 음성향상에 많은 어려움이 있었지만, 잡음 추정 과정에서 적응 망각 계수를 사용하여 비정

적이거나 시변 잡음 환경에서 성능 제한이 있는 음성 검출의 단점을 보완해주었기 때문이다.

## VI. 결론

전통적인 음성 향상 방법인 파워 차감법과 웨이블릿 Shrinkage의 문제점을 개선하기 위해서 다중 대역 파워 차감법과 Perceptual 웨이블릿 패킷 분해 방법을 적용한 개선된 음성 향상 방법을 제안하였다. 먼저 전처리 과정으로 Fourier 주파수 공간에서 다중 대역 파워 차감법을 사용하여 신호의 SNR을 증가시켰으며, SNR에 대해서 선형적인 과차감 인자를 사용하는 방법에서 나타나는 무성음이나 파워가 작은 신호들이 제거되는 문제점을 해결하기 위하여 비선형적인 과차감 인자 선택 방법을 제안하였다.

또한 스펙트럼 바다 인자를 적절히 조절함으로써 뮤지컬 잡음이 발생하는 것을 최소화 하였다. 이후 성능을 더욱 개선하기 위해서 PVPD를 이용하여 신호를 17 개의 웨이블릿 주파수 대역으로 분해하고 각 웨이블릿 노드마다의 엔트로피 값을 기준으로 신호를 유성음/무성음으로 구분하여 각각의 특성에 맞는 웨이블릿 Shrinkage를 적용하여 전처리 과정 이후의 잔여 잡음을 제거하여 뮤지컬 잡음의 발생을 더욱 더 줄일 수 있었다.

한편, 그림 12에서 볼 수 있듯이 잡음 추정 시 망각 계수를 각 주파수 대역의 SNR에 비례하게 적응적으로 선택하여, 음성 검출의 성능 제한으로 인해 음성 구간을 잡음 구간으로 잘못 판단하는 경우에도 잡음의 파워 추정 과정에 음성신호 성분이 미치는 영향을 최소화 하였다. 즉 음성 구간이 잡음 구간으로 잘못 결정 내려지더라도 SNR 이 높기 때문에 잡음 추정 및 갱신 과정에서 현재 프레임에 대한 정보가 극히 조금만 적용되어서 음성 신호 성분이 파워 차감 과정에서 제거되는 영향이 적다.

끝으로, 여러 가지 환경에서의 시뮬레이션을 통해서 제안한 음성 향상 알고리즘이 비정적이거나 시변 특징을 갖는 실제적인 잡음 환경에서도 높은 성능을 보임을 검증하였다. 이를 통해 음성 인식이나 음성 압축과 같은 높은 명료성을 요구하는 음성 신호처리 분야에 응용될 수 있을 것이라고 판단된다.

앞으로의 연구 분야는 Fourier 공간과 웨이블릿 공간의 선형적인 결합으로 인한 계산량의 증가를 해결하는 것과 제안한 알고리즘에 사용된 여러 가지 인자들을 주변 환경의 변화에 적응하도록 하는 것이다.

참 고 문 헌

- [1] S. Boll, "Suppression of acoustic noise in speech using spectral subtraction," *IEEE Trans. Acoust, Signal Processing*, Vol.27, pp.113-120, 1979.
- [2] L.R. Rabiner and R.W. Schafer, *Digital Processing of Speech Signal*. Prentice-Hall, 1978.
- [3] Z. Goh, K.C. Tan, T.G. Tan, "Postprocessing method for suppressing musical noise generated by spectral subtraction," *IEEE Transactions on Speech and Audio Processing*, Vol.6, Issue 3, pp.287-292, May 1998.
- [4] A. Juneja, O. Deshmukh, C. Espy-Wilson, "A multi-band spectral subtraction method for enhancing speech corrupted by colored noise," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Vol.4, pp.IV-4164, May 2002.
- [5] C.Sidney and Burrus, *Introduction to Wavelets and Wavelet Transforms A Primer*. Prentice-Hall, 1998.
- [6] Stephane Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, pp.321-327, 1998.
- [7] D.L. Donoho, and I.M. Jonstone, "Ideal Spatial Adaptation by Wavelet Shrinkage," *Biometrika*, Vol.81, pp.425-455. 1994
- [8] H.Y. Gao and A.G. Bruce, "WaveShrinkage with Semisoft Shrinkage," *StatSci Division of MathSoft, Inc*. 1995.
- [9] M. Bahoura and J. Rouat, "Wavelet Speech Enhancement based on The Teager Energy Operator," *IEEE Signal Processing Letter*, Vol.8, pp.10-12, 2001.
- [10] J. Yang, "Frequency Domain Noise Suppression Approaches in Mobile Telephone System," *Acoustics, Speech, and Signal Processing, IEEE International Conference on*, Vol.2, pp.363-366, April, 1993.
- [11] S.H. Chen et al., "Robust Voice Activity Detection Algorithm Based on the Perceptual Wavelet Packet Transform," *ISPACS 2005. Proceedings of International Symposium on*, pp.45-48, December, 2005.
- [12] E.Zwicker and E.Terhardt, "Analytical Expressions for Critical Band Rate and Critical Bandwidth as a Function of Frequency," *JASA*, Vol.68, pp.1523-1525, 1980.
- [13] A.M.Kondoz, *Digital Speech coding for low bit rate communiucations systems*, John Wiley & Sons, pp.142-160, 1995.
- [14] M. Berouti, R. Schwartz, J. Makhoul, "Enhancement of speech corrupted by acoustic noise," *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '79*. Vol.4, pp.208-211, April, 1979.
- [15] I.J. Kim, S.I. Yang, Y. Kwon, "Speech enhancement using adaptive wavelet shrinkage," *Proceedings of IEEE International Symposium on Industrial Electronics*, Vol.1, pp.501-504, 2001.
- [16] 이윤창, 안상식, "음성 향상 전처리와 문턱값 갱신을 적용한 향상된 음성검출 방법," *한국통신학회논문지*, Vol.28, No.11C, pp.1161-1168, 2003.
- [17] <http://www.sipro.com>. Sipro Lab. Telecom Inc
- [18] <http://spib.rice.edu>, Rice University. DSP group
- [19] Ryouichi Nishimura, Futoshi Asano, and Yoiti Susuki, "Speech Enhancement Using Spectral Subtraction with Wavelet Transform," *Electronics and communications in Japan, Part3*, Vol.81, No.1, 1998.

이 윤 창 (Yoon-Chang Lee)

정회원



1998년 2월 고려대학교 응용 전  
자공학과 졸업

2000년 2월 고려대학교 전자 정  
보공학과 석사

2000년 3월~현재 고려대학교 전  
자정보공학과 박사과정

<관심분야> 음성 신호처리, 이동

통신 알고리즘

곽 정 훈 (Jeong-Hoon Kwak)

준회원



2004년 2월 배재대학교 정보 통신공학부 졸업  
2006년 2월 고려대학교 전자 정보공학과의 석사  
2006년 6월~현재 LG전자 DMP 연구소 Video 개발 그룹 연구원

<관심분야> 음성 신호처리, 이동 통신 알고리즘

안 상 식 (Sang-Sik Ahn)

종신회원



1983년 2월 고려대학교 전자 공학과 졸업  
1985년 2월 고려대학교 전기 공학과 석사  
1994년 1월 Polytechnic Univ Electrical Engineering, Ph.D.  
1994년 2월~1995년 2월 LG 중

앙 연구소 책임 연구원

1995년 3월~현재 고려대학교 전자 및 정보공학부 부교수

<관심분야> 음성 신호처리, 통신 시스템, DSP 알고리즘의 FPGA 하드웨어 구현