# 적응 웨이블릿 패킷 기반 스펙트럼 차감법을 이용한 음성신호 개선에 관한 연구

김진호, 박정재, 장성욱, 권영현, 양성일
한양대학교 전자전기제어계측공학부, 한양대학교 물리학과

# A study on Speech Enhancement Using Adaptive Wavelet Packet Based Spectral Subtraction

Jinho Kim, Jeong-Jae Park, Sungwook Chang, Y. Kwon, Sung-il Yang
kimjinho@ihanyang.ac.kr, schang@ihanyang.ac.kr, yyhkwon@hanyang.ac.kr, syang@hanyang.ac.kr

## 요약

본 논문에서는 최근에 제안된 음성신호 개선을 위한 적응 웨이블릿 패킷 기반의 스펙트럼 차감법을 이용하여 다양한 측면에서의 성능평가 결과를 제시한다. 사용된 음성신호 개선 방식은 적응 웨이블릿 패킷의 불균등 주파수 해상도와 높은 에너지 집중도로 인해 발생하는 극대, 극소값의 영향을 피하기 위해 기하평균을 이용하는 스펙트럼 추정법을 사용하였다. 다양한 측면의 성능평가를 위해 주관적 평가 척도인 MOS 와 높은 상관도를 갖는 것으로 알려진 log likelihood ratio, log area ratio, segmental SNR, weighted spectral slope 등을 평가 척도로 사용하였다. Fourier 기저를 사용한 방식과의 비교에서 적응 웨이블릿 패킷 방식은 segSNR 과 음성의 명료도를 비교적 잘 반영하는 weighted spectral slope 측면에서 우수한 성능을 보였다.

## 1. 서론

음성 신호처리 응용 분야에서 배경잡음을 감소시키는 것은 매우 중요한 일이므로, 이것을 위해 다양한 방법들이 제안되었다. 종래의 음성신호 개선 방식들은 배경잡음을 백색 가우시안 잡음으로 가정한 것들이다. 백색 가우시안 잡음이 수학적으로는 유용한 개념이지만 실생활에서 발생하는 잡음은 대부분 유색 잡음일 뿐 아니라 시간에 따라 변하는 non-stationary 특성도 함께 지니고 있다. 이와 같은 배경잡음에 대해서 최근 제안된 웨이블릿 패킷 기반의 스펙트럼 차감법을 이용하여 잡음제거 실험을 하고 그 결과를 다양한 방법으로 비교 분석 하고자 한다.

## 2. 스펙트럼 차감법

### 2.1 잡음 추정

배경잡음이 non-stationary 하거나 SNR 이 낮은 경우 잡음에 손상된 음성 프레임의 이전 프레임에서

잡음을 추정하는 방식들이 Hirsh 에 의해 제안되었으나, 제안된 적응웨이블릿 패킷 기반 스펙트럼 차감법에서는 다음 식과 같이 적응 웨이블릿 패킷 계수에 적용한다.

$$\hat{N}_k = \begin{cases} \alpha\hat{N}_{k-1} + (1-\alpha)X_k(n), & X_k(n) \le \beta N_{k-1}(n) \\ \hat{N}_{k-1}(n) & , \quad \text{otherwise} \end{cases} \tag{1}$$

여기서 $X_k(n)$ 은 $k$ 번째 프레임 적응 웨이블릿 필터 뱅크 $n$ 번째 노드의 magnitude 스펙트럼이고, $\hat{N}_k(n)$ 는 추정잡음이다.

## 2.2 적응 스펙트럼 가중치

제안된 방식은 적응 웨이블릿 패킷의 높은 에너지 집중도로 인해 magnitude 스펙트럼에 대해서 극대, 극소값들의 영향이 커지게 된다. 따라서 일반적인 스펙트럼 차감법에서 사용하는 샘플 평균을 이용하게 되면 극한값의 영향으로 잡음 추정과 가중치 추정에 방해 요소로 작용한다. 극한값의 영향을 감소시키기 위해 기하 평균을 이용하고, non-stationary 잡음의 특성을 살릴 수 있도록 적응 스펙트럼 차감법 가중치를 구한다. 각 프레임에 대해 잡음에 손상된 신호의 로그 스케일 기하평균 NSGM 과 추정된 잡음의 로그 스케일 기하평균 NGM 을 다음과 같이 정의한다.

$$NSGM(k) = \log\left(\prod_{n=0}^{N_{node}} |X_k(n)|\right)^{\frac{1}{N_{node}}} \tag{2}$$

$$NGM(k) = \log\left(\prod_{n=0}^{N_{node}} |\hat{N}_k(n)|\right)^{\frac{1}{N_{node}}} \tag{3}$$

여기서 $N_{node}$ 는 노드의 수이다. 이 NSGM 과 NGM 을 이용하여 GNNR(geometric noisy signal to noise ratio)를 정의하고 입력음성에 대한 global 가중치로 사용한다.

$$GNNR = \frac{\sum_{k=0}^{N_{frame}-1} NSGM(k)}{\sum_{k=0}^{N_{frame}-1} NGM(k)} \tag{4}$$

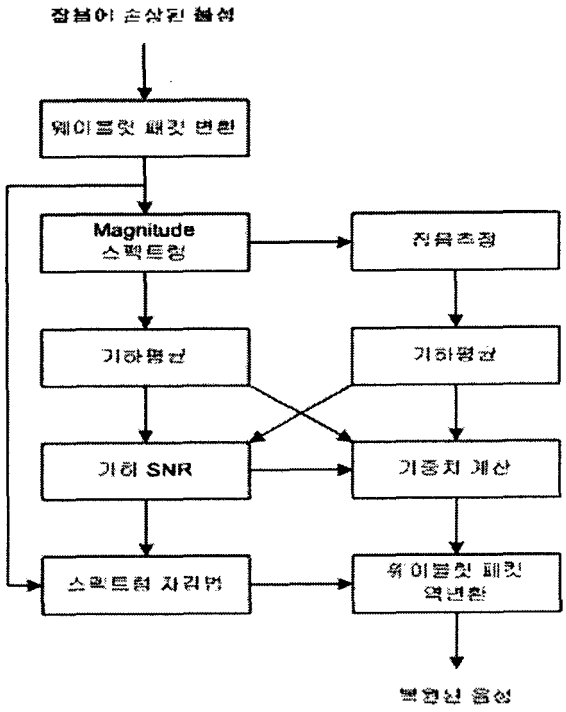


그림 1. 제안된 스펙트럼 차감법의 개요도

여기서 $N_{frame}$ 는 프레임의 수이다. 이 global 가중치와 기하평균을 이용하여 다음 식과 같이 각 프레임 에서의 local 가중치를 구한다.

$$\xi(k) = GNNR\left(\frac{1}{\rho}\ \frac{NSGM_{max} - NSGM(k)}{NSGM_{max} - NSGM_{min}}\right) \tag{5}$$

여기서 $NSGM_{max}$ , $NSGM_{min}$ 은 모든 프레임 중 NSGM 의 최대값과 최소값이다. $\rho$ 는 $2.0 \le \rho \le 3.0$ 사이의 값이며 적응 웨이블릿 패킷의 높은 집중도로 인해 local 가중치가 과다하게 설정되는 것을 막아준다. 따라서, 사용된 적응 웨이블릿 패킷은 SNR 이 낮아질수록 음성신호 보다는 잡음의 특성에 적응하게 되어 잡음에 대한 에너지 집중도가 높아지게 되므로

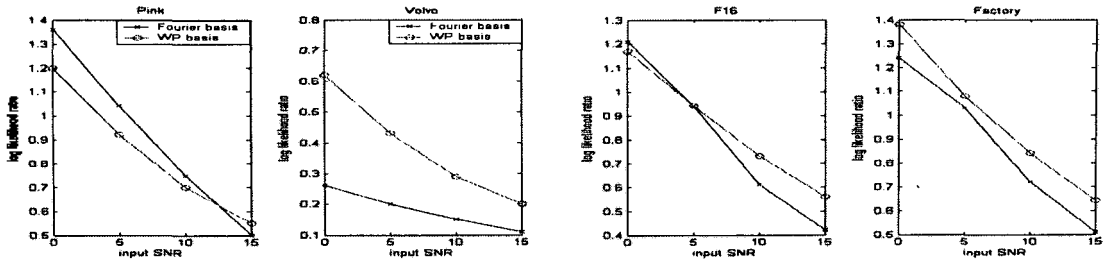


그림 2. 20개의 음성에 대한 log likelihood ratio

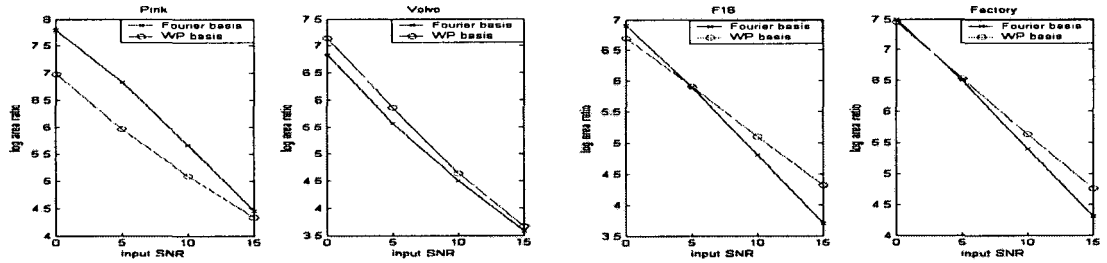


그림 3. 20개의 음성에 대한 log area ratio

음성신호의 포락선에 반비례하는 $\xi(k)$ 의 크기를 이용하면 잡음추정의 정확도가 증가하게 되어 잡음 성분에 대한 에너지 집중도에 적응하는 특성을 가지게 된다.

### 2.3 스펙트럼 차감법

적응 스펙트럼 차감법 가중치를 이용하여 k 번째 프레임 n 번째 노드의 gain $G_k(n)$을 구한다.

$$G_k(n)=\begin{cases}\sqrt{1-\dfrac{(1+\xi(k))\,|\hat{N}_k(n)|}{|X_k(n)|}}, & |X_k(n)|\geq(1+\xi(k))\,|\hat{N}_k(n)| \\ \eta\sqrt{\dfrac{\hat{N}_k(n)}{X_k(n)}} & \text{, otherwise} \quad (0\leq\eta\leq 1)\end{cases} \tag{6}$$

구해진 gain 을 이용하여 웨이블릿 패킷 계수들의 크기를 줄임으로써 음성의 개선이 이루어진다.

$$\hat{x}_{k,n}(i)=x_{k,n}(i)G_k(i) \tag{7}$$

$N_s$ 를 적응 웨이블릿패킷 트리의 노드 사이즈라 했을 때, $0\leq k\leq N_{frame}$, $0\leq n\leq N_{node}$, $0\leq i\leq N_s$ 범위에 있다.

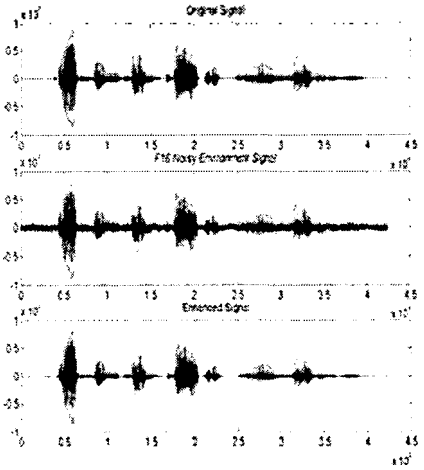


그림 4. 잡음이 제거된 신호와 원신호

## 3. 실험 및 결과

본 실험에서는 음성 데이터로는 TIMIT 데이터 베이스에서 추출한 각각 10명의 남녀 화자의 음성을 사용했으며, 웨이블릿 기저로는 Daubechies 20차 기저를 사용하였고, Noisex-92 데이터 베이스로부터 추출된 F16, Factory, Pink, Volvo, White 잡음이 사용되었다. 또한 다양한 측면에서 성능평가를 위해 log likelihood ratio, log area ratio, segmental SNR, weighted spectral slope 등을 평가 지표로 사용하였다.

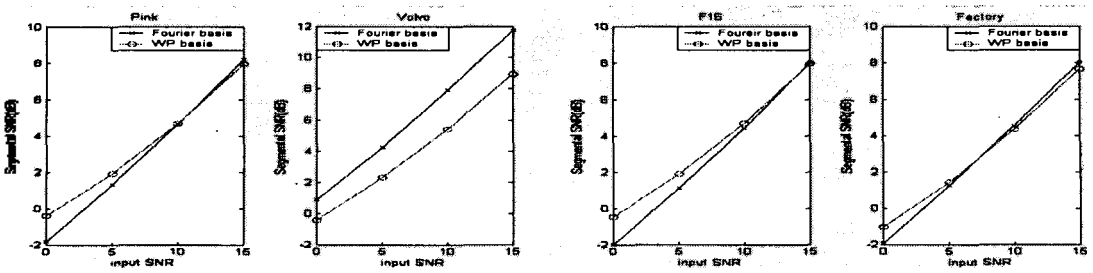


그림 5. 20개의 음성에 대한 segmental SNR

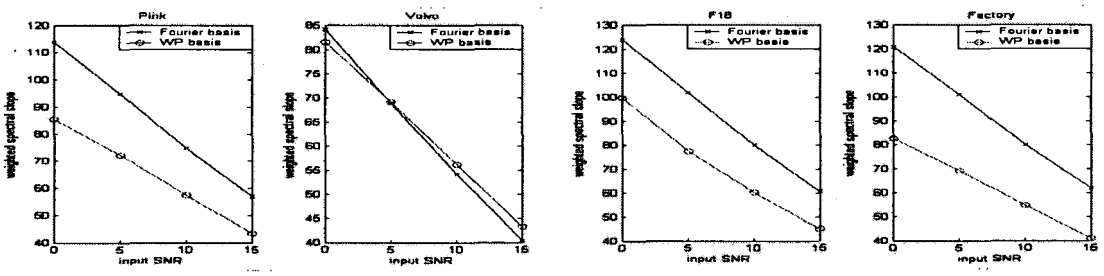


그림 6. 20개의 음성에 대한 weighted spectral slope

log likelihood ratio 의 경우 pink 와 white 배경잡음을 제외한 나머지에서 웨이블릿 패킷 기저를 사용한 경우가 Fourier 기저를 사용할 때 보다 높은 상관도를 보였다. log area ratio 의 경우 white 배경잡음을 제외하고, Fourier 기저와 웨이블릿 패킷 기저를 이용한 결과가 거의 차이가 없음을 볼 수 있었다. segSNR 의 경우 Volvo 배경잡음을 제외하고 웨이블릿 패킷 기저가 우수한 성능을 보였으며, 음성의 명료도를 비교적 잘 반영하는 weighted spectral slope 측면에서 웨이블릿 패킷 기저가 배경잡음의 종류에 상관없이 우수한 성능을 보였다.

## 참고문헌


1. Sungwook Chang, Sung-il Jung, Younghun Kwon, Sung-il Yang, "Adaptive Wavelet Based Speech Enhancement with Robust VAD in Non-stationary Noise Environment". THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA, Vol.22, No.4E, 2003.12.
2. Sungwook Chang, Sung-il Jung, Younghun Kwon, Sung-il Yang, "Speech Enhancement Using Level Adapted Wavelet Packet with Adaptive Noise Estimation", THE JOURNAL OF THE ACOUSTICAL SOCIETY OF KOREA, Vol.22, No.2E, 2003, 6.
3. Sungwook Chang, Younghun Kwon, Sung-il Yang, "Speech Enhancement for Non-Stationary Noise Environment by Adaptive Wavelet Packet", Proc.IEEE Int. Conf. Acoustics, Speech and Signal Processing
4. H. G. Hirsch and C.Ehricher, "Noise estimation techniques for robust speech recognition", Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing, 1995