웨이브렛 변환을 이용한 음성신호의 잡음 제거

황 향 자, 김 종 교
전북대학교 전자정보공학부

Noise Cancellation of Speech Signal Using Wavelet Transform

Hyang-Ja Hwang, Chong-Kyo Kim
Division of Electronics & Information Engineering, Chonbuk National University
E-mail: hhjbluesky@hanmail.net

Abstract

본 논문은 잡음 환경에서의 음성인식을 위하여 음성에 부가된 잡음을 제거하는 방법으로 프레임 단위로 웨이브렛 변환 영역의 표준편차를 활용하여 시간 적응적 임계값을 사용하는 새로운 방법을 제안한다. 웨이브렛 변환영역의 cD1과 cA3의 표준편차 값을 이용하여 임계값을 설정함으로써 음성의 변화에 적응할 수 있도록 하였다. 또한 묵음구간의 잔여 잡음을 제거하기 위한 방법을 제안하였다. 실험을 통하여 제안한 방법이 기존의 웨이브렛 변환과 웨이브렛 패킷 변환을 이용한 잡음제거 방법보다 SNR(Signal to Noise Ratio)과 MSE(Mean Squared Error) 측면에서 향상됨을 확인할 수 있었다.

I. 머리말

음성인식 시스템의 실용화가 늘어남에 따라 최근에는 주변 잡음에 대한 인식 시스템의 성능저하가 문제시되고 있다. 이러한 이유는 잡음이 없거나 비교적 조용한 실험실에서는 우수한 성능을 나타내는 음성인식 시스템의 성능이 잡음환경의 영향을 고려하지 않은 음성인식 시스템의 실제 환경에서는 성능이 급격하게 감소하게 되므로 잡음에 의해 오염된 음성신호에 잡음을 제거하는 기술인 음성 개선은 음성신호처리 분야에서 매우 중요하다고 볼 수 있다.

최근 들어 활발히 연구 되고 있는 웨이브렛 변환(wavelet transform)은 해석하고자 하는 주파수 성분에 따라 가변 할 수 있는 창함수의 크기를 제공한다. 웨이브렛 변환에서 창함수는 기저함수(basis function)라고 불리어진다. 즉 가변하는 기저함수에 의한 웨이브렛 변환의 다중해상도 특성과 시간-주파수 국부성은 통계적 특성을 모르거나 시간적으로 예측하기 힘든 비정상(non-stationary)적인 신호해석에 매우 유용한 것으로 밝혀졌다[1][2].

본 논문에서는 웨이브렛 변환을 이용한 잡음제거 방법에서 가장 중요한 부분인 웨이브렛 계수들을 두 그룹, 즉, 잡음 성분의 영향을 많이 받는 계수들과 그러지 않은 계수들로 나누는 기준인 임계값을 정하는 새로운 방법을 제안한다. 또한 제안한 방법으로 잡음을 제거하고 난 후에도 묵음구간에 잔여 잡음이 존재하게 되는데 묵음구간을 검출하기 위한 방법을 제안하며 이를 이용하여 묵음구간에 존재하는 잔여 잡음을 제거한다. 성능 평가를 위해서 SNR(Signal to Noise Ratio)과 MSE(Mean Squared Error)을 계산하여 기존의 웨이브렛 변환과 웨이브렛 패킷 변환과 비교하였다.

II. 웨이브렛을 이용한 잡음제거

웨이브렛을 이용한 thresholding 방법의 기본 원리는 백색 가우시안 잡음(white gaussian noise)에 의해 오염된 신호를 웨이브렛 변환 했을 때 각 스케일에 포함된 잡음 성분은 신호 성분의 크기보다 상대적으로 작은 값을 가지므로 적절한 문턱값 이하의 값을 제거한 후 다시 합성함으로써 효과적으로 잡음을 제거할 수 있다는 것이다. 이 과정을 요약하면 다음과 같다[3][4].

1) Decompose : 기저 웨이브렛을 선택한 후 잡음에 손상된 음성으로부터 웨이브렛 변환을 취하여 계수를 구한다.

2) Thresholding : 각각의 레벨별로 임계값을 정한 후 soft thresholding을 하여 웨이브렛 계수의 임계값 이

하의 성분을 제거한다.

3) Reconstruction : 추정된 웨이브렛 계수로부터 역 웨이브렛 변환을 취함으로써 잡음이 제거된 음성을 구한다.

잡음이 없는 원 신호 s에 평균값이 '0'이고 표준편차가 σ인 백색 가우시안 잡음 n을 첨가한 잡음신호 x는 식(1)로 모델링 된다.

$$x = s + \sigma n \tag{1}$$

식(1)을 웨이브렛 변환하면 식(2)과 같이 나타내며 여기서 W를 직교 웨이브렛 변환이라 가정한다면 $X = Wx$를 의미한다.

$$X = S + N \tag{2}$$

원음성의 추정값 $\hat{s}$를 얻기 위한 웨이브렛 변환식으로서 $\hat{S}$와 M은 각각 원음성의 웨이브렛 변환 계수 S의 추정값과 W의 역변환 행렬을 의미하며 $\hat{S}$는 잡음음성의 웨이브렛 계수인 X를 shrinking 또는 killing함으로써 얻을 수 있다.

$$\hat{s} = M\hat{S} \tag{3}$$

본 논문에서는 식(4)와 같이 soft threshold 방법으로 웨이브렛 계수들을 제거하는 방법을 사용하며 임계값 λ는 웨이브렛 변환일 경우 식(5)을 사용하며, 웨이브렛 패킷 변환일 경우 식(6)을 사용한다.

$$T_{soft}(X) = \begin{cases} sgn(X)(|X| - \lambda) & ,|X| \geq \lambda \\ 0 & ,|X| < \lambda \end{cases} \tag{4}$$

$$\lambda = \sigma\sqrt{2\log(N)} \tag{5}$$

$$\lambda = \sigma\sqrt{2\log(N\log_2 N)} \tag{6}$$

N은 신호의 샘플수이고, σ는 선택되어진 웨이브렛 계수의 표준편차이다. 표준편차는 잡음이 포함된 웨이브렛 계수의 중간값을 이용하여 식(7)과 같이 계산한다.

$$\sigma = median/0.6745 \tag{7}$$

Ⅲ. 시간 적응적 임계값

웨이브렛 변환으로 얻어지는 웨이브렛 계수가 잡음신호 구간과 잡음이 섞인 음성신호 구간에서 다른 특성을 가진다는 것을 이용하였다. 웨이브렛 계수의 표준편차는 잡음이 섞인 음성신호 구간에서 작은 값을 가지며 음성이 포함된 구간에서는 비교적 큰 값을 가진다. 웨이브렛 변환에서 스케일이 작을수록 빠르게 변화하는 부분에서 정확하게 검출할 수 있어서 고주파 구간에서 웨이브렛 계수의 표준편차는 작아지게 된다. 잡음이 섞인 음성신호의 초기부분과 끝부분은 잡음 신호만이 존재한다고 가정한다[5].

본 논문은 프레임 단위로 이산 웨이브렛 변환을 통해 얻어진 첫 번째 detailed 스케일 계수 cD1과 세 번째 approximation 스케일 계수 cA3의 표준편차를 이용한다. cD1은 고주파 부분을, cA3은 저주파 부분을 잘 반영하는 성질을 이용하여 적응적 임계값을 결정하고 묵음 구간을 검출하는데 사용하였다.

3.1 시간 적응적 임계값을 이용한 잡음 제거

본 논문에서는 주어진 음성 신호를 프레임 단위로 나누어 각각의 프레임에 대하여 웨이브렛 변환한 후 cD1과 cA3를 이용하면 잡음 환경 하에서도 음성 구간을 검출할 수 있음을 이용하여 적응적 임계값을 생성한다. 이는 음성의 시작이나 끝부분에 존재하는 파열음이나 마찰음의 경우 신호의 에너지는 유성음 구간에 비해 상대적으로 작지만 주파수 영역에서 고주파 부분에 많은 에너지를 가지게 되며, 유성음 구간의 경우 저주파 부분에 많은 에너지를 가지게 됨을 이용한 것이다.

이러한 성질을 이용하여 웨이브렛 영역에서 음성검출을 위한 파라미터를 식(8)과 같이 정의하여 사용할 수 있다[5].

$$T^k = S^k_{A3} + \alpha S^k_{D1} \qquad k = 1, \cdots, N. \tag{8}$$

S^k_{A3}은 세 번째 approximation 스케일의 표준편차이며, α은 weighting factor이고, S^k_{D1}은 첫 번째 detailed 스케일의 표준편차이다. k는 프레임의 수이며, 가중치 α값은 실험적으로 적당한 값을 선정하기 위해 실험을 통하여 [5]의 경우와 같이 '6'의 값이 최적의 값임을 알 수 있었다.

이 파라미터 값의 최대값이 '1'이 되도록 정규화 과정(최대값 조정 과정)을 거친다.

$$T^{k'} = T^k \times \frac{1}{\max(T^k)} \tag{9}$$

이렇게 구한 파라미터 $T^{k'}$를 이용하여 시간 적응적 임계값을 다음과 같이 정의한다[3].

$$\lambda^{k'} = \beta\lambda(1 - T^{k'}) \tag{10}$$

여기에서 β는 조정 파라미터이다. 마지막으로 재 샘플링을 통하여 시간 적응적 임계값을 구한다. $\lambda^{k'}$는 원 신호의 샘플수가 아닌 프레임의 수이므로 원 신호의 샘플수로 재 샘플링을 통하여 시간 적응적 임계값을 결정할 수 있다.

잡음 신호와 시간 적응적 임계값의 비교과정을 거치면서 soft thresholding을 이용하여 잡음을 제거한다. 그러나 이러한 방법으로 잡음을 제거하여도 묵음구간에서의 잔여잡음이 존재하게 된다.

3.2 묵음구간 잔여잡음 제거

잡음이 섞인 음성신호의 초기부분과 끝부분은 잡음 신호만이 존재한다고 가정하였으므로 원 신호의 앞부

분과 뒷부분의 일부를 이용하여 묵음구간을 검출하기 위한 파라미터로 사용한다. 본 연구에서는 실험 과정을 거쳐 프레임의 크기를 256으로 정하였으며, 이때 앞부분과 뒷부분 각각 5개의 프레임을 이용하도록 하였다. 만약 프레임 크기가 바뀔 경우 이용되는 프레임 수도 바뀌게 된다.

잡음이 섞인 음성 신호의 음성구간과 묵음 구간의 특성을 살펴보면 음성구간의 경우 해당 프레임의 파워와 표준편차는 크며, 묵음 구간의 경우 해당 프레임의 파워와 표준편차는 작은 특성을 가지고 있다. 이를 이용하여 각 프레임에 해당하는 샘플 값들을 제곱한 후 합산해서 앞에서 구한 T^k 값으로 곱한 값을 SF (silence frame)으로 정하며 식(11)과 같다. 묵음 구간을 검출하기 위한 임계값은 λ_{SF_k} 으로 정하며 식(12)과 같이 설정하여 사용한다.

$$SF_k = T^k \times \sum_{i=1}^{256} (F_i^k)^2 \qquad (11)$$

$$\lambda_{SF_k} = \left(\sum_{k=1}^{5} SF_k + \sum_{k=N-4}^{N} SF_k \right) / 10 \qquad (12)$$

여기에서 F_i^k 는 k번째 프레임의 각각의 샘플 값들을 의미한다.

SF_k 의 처음 5개의 값과 마지막 5개의 값의 평균을 묵음 구간을 검출하기 위한 임계값으로 사용한다. 만약 SF_k 의 값이 λ_{SF_k} 보다 작으면 묵음 구간으로 간주하고 잡음이 제거된 신호에서 이 프레임에 해당하는 샘플 값들은 '0'으로 만들어 묵음 구간의 잔여 잡음을 제거하며, SF_k 의 값이 λ_{SF_k} 보다 크면 음성구간으로 간주하며 원래의 샘플 값을 변경시키지 않는다. 즉, 식(11)과 식(12)와 같은 간단한 식을 이용하여 음성의 시작점과 끝점을 검출하는 끝점 검출을 수행할 수 있다. 끝점 검출의 정확도는 프레임의 크기를 작게 함으로써 높일 수가 있다.

Ⅳ. 실험 및 고찰

4.1 음성 및 잡음의 구성

실험에 사용된 음성은 16kHz 샘플링, 16bits로 양자화된 "청와대"를 발음한 음성신호를 사용하였다. 잡음 음성을 위해서는 원음성에 white gaussian noise를 첨가하여 0dB에서 20dB 까지의 잡음신호를 만들어 실험을 수행하였다. 분석 프레임은 256샘플이고 128샘플단위로 overlap을 수행하였다.

웨이브렛 함수는 다우베치(Daubechies)함수를 사용하였다. 다우베치 웨이브렛 함수를 사용한 이유는 함수 모양이 음성 파형과 유사하고 직교성을 가지기 때문이다. 직교성을 가지고 있는 함수만이 원 신호를 완전히 복원할 수 있다[1].

다우베치 웨이브렛 함수에 의한 이산 웨이브렛 변환 계수의 cD1과 cA3의 표준편차 및 이를 이용한 시간 적응적 임계값을 그림 1에 나타내었다.

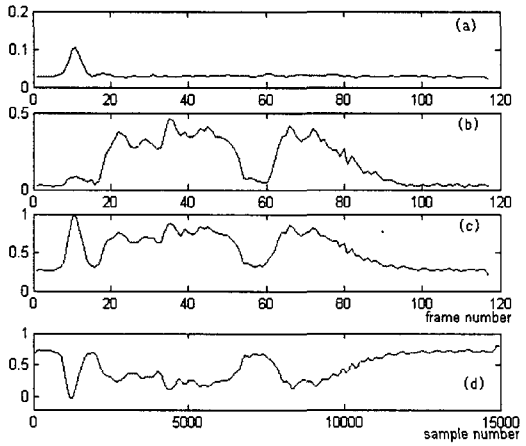

그림 1. (a) S_{D1}^k (b) S_{A3}^k (c) $T^{k'}$ (d) $1-T^{k'}$

그림 2와 그림 3은 각각 10dB와 15dB의 white Gaussian noise를 첨가한 음성신호의 잡음 제거 결과를 기존의 웨이브렛 변환과 웨이브렛 패킷 변환 그리고 제안한 시간 적응적 임계값 방법을 비교하여 보여주고 있다.

그림 2와 3을 통하여 제안한 시간 적응적 임계값 방법과 묵음구간 잔여 잡음 제거를 통하여 효과적으로 원 신호와 동일하게 복원되었음을 알 수 있다. 특히 웨이브렛 변환이나 웨이브렛 패킷 변환의 경우 1100 ~1800 샘플 구간의 파열음을 많이 제거하는 반면 제안한 알고리즘은 파열음 부분이 원 신호와 거의 동일하게 복원함을 볼 수 있다.

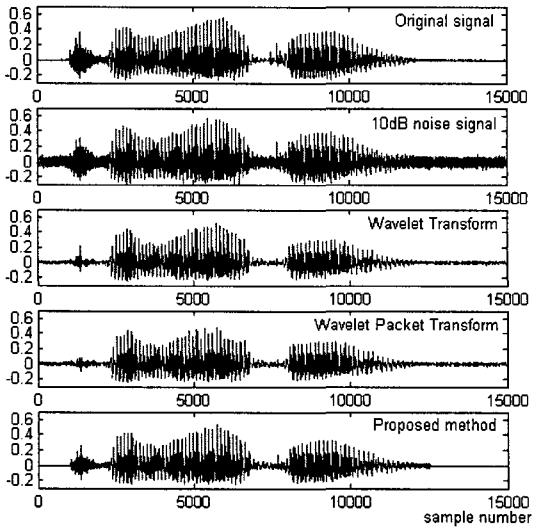

그림 2. 10dB의 white gaussian noise를 첨가한 음성신호의 잡음 제거 비교

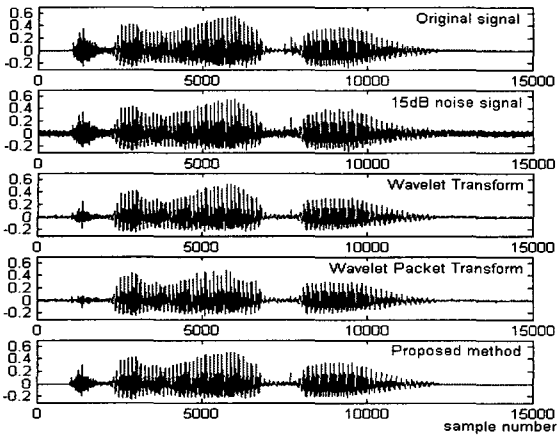

그림 3. 15dB의 white gaussian noise를 첨가한 음성신호의 잡음 제거 비교

4.2 성능 평가

본 논문에서는 잡음 제거의 객관적인 성능 평가를 위하여 SNR(Signal to Noise Ratio)과 MSE(Mean Squared Error)를 계산 한다[6]. 수식으로 나타내면 다음과 같다.

$$SNR(\delta) = 10 \log \frac{\sum_i \hat{s}_i^2}{\sum_i n_i^2} \tag{13}$$

$$MSE(\delta) = \frac{\sum_{i=0}^{N} (s_i - \hat{s}_i)^2}{N} \tag{14}$$

여기에서 s, n, $\hat{s}$ 는 각각 원 신호, 잡음, 원 신호의 추정값을 나타내며, N은 샘플수이다.

SNR 값이 크다는 것은 음성 신호에 개선되는 정도가 크다는 것을 나타내며, MSE 값이 작을수록 원 신호와 거의 동일하게 복원되었음을 나타낸다.

표1과 표2는 SNR과 MSE를 통해 웨이브렛 변환과 웨이브렛 패킷 변환 그리고 제안한 시간 적응적 임계값 방법의 성능을 비교한다.

표 1. SNR 비교

	WT	WPT	proposed
0dB	6.1706	6.1220	6.3984
5dB	8.4284	8.0363	10.2654
10dB	10.1966	8.8437	13.2173
15dB	11.8938	9.1112	15.6220
20dB	13.3708	9.2142	17.5736

표1과 표2를 통해서 제안한 방법이 20dB의 잡음을 첨가한 경우 SNR측면에서는 웨이브렛 변환보다는 4dB, 웨이브렛 패킷 변환보다는 8.3dB정도로 향상되었으며, MSE측면에서도 각각 0.23, 0.81정도의 우수한 성능을 보였다.

표 2. MSE 비교 (단위 : 1.0e-003)

	WT	WPT	proposed
0dB	1.9130	1.9340	1.8150
5dB	1.1380	1.2460	0.7450
10dB	0.7570	1.0340	0.3780
15dB	0.5123	0.9720	0.2172
20dB	0.3645	0.9491	0.1385

0dB의 잡음을 첨가한 경우에 대해서도 기존의 웨이브렛 변환이나 웨이브렛 패킷 변환보다 SNR측면에서는 각각 0.22dB, 0.37dB, MSE측면에서도 각각 0.10, 0.12정도의 개선된 성능을 보였다.

V. 맺음말

본 논문에서는 음성이 white gaussian noise에 의해 손상되었다는 가정에서 이를 제거하기 위하여 웨이브렛 변환 영역에서 시간 적응적 임계값을 설정하는 방법을 제안하였다. 그러나 잡음을 제거하고 난 후에도 묵음구간에 잔여 잡음이 존재하게 되므로 묵음구간을 검출하여 묵음구간의 잔여 잡음 또한 제거하였다. 제안한 방법이 효과적으로 잡음을 제거 할 수 있음을 기존의 웨이브렛 변환이나 웨이브렛 패킷 변환을 이용한 방법과의 비교에서 SNR의 향상과 MSE의 감소로써 증명하였다. 그리고 이러한 방법을 실제 환경에서 적용하기 위해서는 좀더 다양한 잡음에 대한 연구와 적용이 필요할 것으로 본다.

참고문헌

[1] Daubechies, I., "The Wavelet transform, time-frequency localization and signal. analysis", *IEEE Trans. on information theory*, vol. 36, no. 5, pp. 961-1005, 1990.

[2] Michel Misiti, Yves Misiti, Georges Oppenheim, Jean-Michel Poggi, *Wavelet Toolbox for Use with MATLAB®*, The Math Work Inc., 2001.

[3] Bahoura, M., Fouat, J., "Wavelet speech enhancement based on the teager energy operator", *IEEE Signal Process. Lett.* 8(1), pp. 10-12, 2001.

[4] S. W.. Chang, Y. Kwon, S. I. Yang, I. J. Kim, "Speech Enhancement for Non-stationary Noise Environment by adaptive Wavelet Packet" *IEEE International Conference on Acoustics, Speech, and Signal Processing,* v1, pp. I-561-I-564, 2002.

[5] 석종원, "Wavelet Transform-Based Speech Signal Processing: Speech Enhancement and Endpoint Detection", 박사학위논문, 경북대학교, 1999.

[6] 김현기, 이상운, 홍재근, "이산 웨이브렛 변환영역에서의 스펙트럼 차감법을 이용한 잡음제거", 멀티미디어학회 논문지, 제4권, 제4호 , pp. 306-315, 2001.