

웨이브렛 변환을 이용한 음성신호의 잡음제거

Denoising of Speech Signal Using Wavelet Transform

한 미 경*, 배 건 성**

(Mi Kyung Han*, Keun Sung Bac**)

요 약

본 논문에서는 웨이브렛 변환을 이용하여 음성신호의 잡음을 제거하는 방법에 대해 연구하였다. Cycle-Spinning 기법과 비간축 웨이브렛 변환을 음성신호의 잡음제거에 적용하고, 그 결과를 일반적인 웨이브렛 변환을 이용한 방법과 비교·분석하였다. 잡음제거를 위해서 Symlets 8-tap 필터와 피라미드 알고리즘으로 웨이브렛 변환을 수행하고 soft-thresholding 기법을 이용하였다. 성능 평가를 위해서 SNR, 첵스트럼 거리, 주관적인 청취실험을 수행하였고 실험결과 주관적인 청취실험뿐만 아니라 객관적인 성능평가에서도 Cycle-Spinning 기법과 비간축 웨이브렛 변환을 이용한 방식이 일반적인 웨이브렛 변환을 이용한 방식에 비해 우수한 성능을 보였다. 합성음의 청취실험 결과 일반적인 웨이브렛 변환을 이용한 잡음제거 신호의 불연속점에서 나타나는 "clicks" 성분이 많이 감소하여 음질이 향상됨을 확인하였다.

핵심용어: 음성개선, 잡음제거, 웨이브렛 변환, 비간축 웨이브렛 변환, Cycle-Spinning 기법

투고분야: 음성처리 분야(2.3)

ABSTRACT

This paper deals with speech enhancement methods using the wavelet transform. A cycle-spinning scheme and undecimated wavelet transform are used for denoising of speech signals, and then their results are compared with that of the conventional wavelet transform. We apply soft-thresholding technique for removing additive background noise from noisy speech. The symlets 8-tap wavelet and pyramid algorithm are used for the wavelet transform. Performance assessments based on average SNR, cepstral distance and informal subjective listening test are carried out. Experimental results demonstrate that both cycle-spinning denoising(CSD) method and undecimated wavelet denoising(UWD) method outperform conventional wavelet denoising(CWD) method in objective performance measure as well as subjective listening test. The two methods also show less "clicks" that usually appears in the neighborhood of signal discontinuities.

Key words: Speech enhancement, Noise reduction, Wavelet transform, Undecimated wavelet transform, Cycle-Spinning

1. 서 론

잡음환경의 영향을 고려하지 않은 음성신호처리 시스템은 실제 환경에서 성능이 급격하게 감소하게 되므로 잡음에 의해 오염된 음성신호에서 잡음을 제거하는 기술인 음성개선은 음성신호처리 분야에서 매우 중요하다고 볼 수 있다. 이러한 음성개선 기술은 음성인식 시스템의 전처리 단계에서 배경잡음의 제거 및 억제에 사용되고 있으며, 오래된 레코드의 음질 향상과 보청기 등 여러 가지 시스템에 응용되고 있어서 음성개선에 관한 연구는 꾸준히 이루어지고 있다.

일반적으로 시간에 따라 변하는 신호의 주파수 해석을 위하여 식 (1)에 주어진 단구간 푸리에 변환(STFT: short-time Fourier transform)이 이용되고 있다.

$$STFT(\tau, w) = \int f(t)g^*(t-\tau)e^{-j\omega t} dt \quad (1)$$

그러나, 창(window)함수 $g(t)$ 가 고정된 길이를 가지기 때문에 시간-주파수 해상도 역시 고정되므로 음성신호와 같이 비정재성(nonstationary)을 갖는 신호의 특성을 표현하는데 단구간 푸리에 변환이 적합하지 않다고 볼 수 있다[1]. 이러한 한계를 극복하고자 최근에 널리 연구되고 있는 웨이브렛 변환은 다중해상도(multiresolution)를 갖는 신호해석 방법으로서 시간 및 주파수의 국부성을

* 삼성전자

**경북대학교 전자·전기공학부

접수일자: 2000년 5월 9일

가지므로 통계적 특성을 모르거나 시간적으로 예측하기 힘든 신호해석에 유용하다. 그리고 웨이브렛 필터의 성질을 이용하여 신호의 압축 및 잡음제거 등에도 널리 이용되고 있다[6-9].

본 논문에서는 thresholding 기법을 기반으로 하는 음성 개선 알고리즘을 Cycle-Spinning 기법[9]을 이용한 웨이브렛 변환과 비간축 웨이브렛 변환[4,5] 영역에서 적용하여 효과적으로 잡음을 제거할 수 있음을 보였다. 특히, 잡음 제거 과정에서 무성음의 손실을 줄이기 위해서 먼저 무성음 구간을 검출한 다음 유성음과 무성음에 대해 서로 다른 스케일에서 잡음 레벨을 추정하고 다른 방법으로 잡음을 제거하였다. 그리고 Cycle-Spinning 기법을 이용한 방식(CSD: Cycle-Spinning Denoising)과 비간축 웨이브렛 변환을 이용한 방식(UWD: Undecimated Wavelet Denoising)의 성능을 일반적인 웨이브렛 변환을 이용한 방식(CWD: Conventional Wavelet Denoising)과 비교·분석하였다.

본 논문의 구성은 다음과 같다. II장에서는 웨이브렛 이론 및 일반적인 이산 웨이브렛 변환, 비간축 웨이브렛 변환, Cycle-Spinning 기법에 대해서 설명을 하고, III장에서는 음성신호의 특성을 고려하여 제안한 음성개선 알고리즘에 대하여 기술하였다. 그리고 실험 결과 및 고찰에 대해 IV장에서 설명을 하였고, 마지막으로 V장에서 결론을 기술하였다.

II. 웨이브렛 변환(Wavelet Transform)

2.1. 웨이브렛 이론

웨이브렛 이론은 응용수학에서 처음 소개된 후 컴퓨터 비전 분야 등에서 연구되어 온 다중해상도 표현과 연관성이 있음이 밝혀졌으며, 푸리에 변환과 같이 기저함수들의 집합을 이용하여 신호를 분해하는 일종의 신호해석 방법이다[1-3]. 연속 웨이브렛 변환(CWT: Continuous Wavelet Transform)은 식 (2)와 (3)으로 정의된다. 식 (2)에서 $\varphi(t)$ 는 원형(prototype) 웨이브렛이며, $\varphi_{a,r}$ 는 $\varphi(t)$ 를 이동과 확장하여 구해진다. 즉, 고정된 크기의 창함수를 사용하는 푸리에 변환과는 달리 짧은 지속시간을 갖는 고주파신호에 대해서는 짧은 창함수를 사용하고 긴 지속시간을 갖는 저주파신호에 대해서는 긴 창함수를 이용하는 결과가 되어 주파수 영역에 따른 다중해상도를 갖게 된다.

$$CWT(\tau, a) = \int f(t) \varphi_{a,r}^*(t - \tau) dt \tag{2}$$

$$\varphi_{a,r} = \frac{1}{\sqrt{a}} \varphi\left(\frac{t - \tau}{a}\right) \tag{3}$$

연속 웨이브렛 변환은 시간과 스케일 인자가 연속적인 값을 갖기 때문에 신호 분석시에 많은 중복성을 지니게 되므로 실제 응용에서는 a 와 τ 를 적절히 샘플링한 이산적인 형태인 dyadic 웨이브렛 변환(DyWT: Dyadic

Wavelet Transform)이 널리 이용된다. DyWT은 식 (4)와 (5)로 표현된다. 여기서 j 와 k 는 각각 이산 변환에서의 스케일과 시간영역에서의 이동을 가리킨다[1].

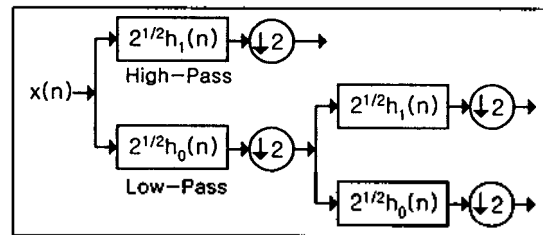
$$d_{j,k} = \int f(t) \varphi_{j,k}^*(t) dt \tag{4}$$

$$\varphi_{j,k}(t) = 2^{-\frac{j}{2}} \varphi(2^{-j}t - kT) \tag{5}$$

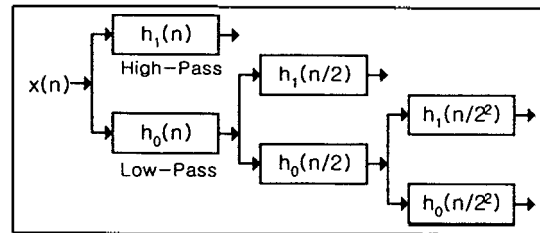
2.2. 일반적인 이산 웨이브렛 변환과 비간축 웨이브렛 변환

웨이브렛 변환은 신호처리 관점에서 대역통과 필터뱅크의 출력으로 볼 수 있으며, 신호를 분할하기 위해서 그림 1과 같이 트리 형태의 필터뱅크를 구성한다. 그림 1에서 (a)는 일반적인 웨이브렛 분해 필터뱅크 구조이다. 입력신호가 저역통과 필터(h_0)와 고역통과 필터(h_1)를 거치고 2배의 간축(decimation)과정을 거치게 되면 한 번의 웨이브렛 변환이 수행되며, 이러한 과정을 원하는 스케일까지 반복적으로 수행하여 웨이브렛 변환된 신호를 얻을 수 있다. 그리고 일반적인 웨이브렛 변환에서 각 스케일은 상위 스케일에서 2배로 간축하여 구해지므로 각 스케일의 샘플수가 상위 스케일의 반이 된다.

그림 1의 (b)는 비간축 웨이브렛 변환을 위한 필터뱅크 구조이다. 비간축 웨이브렛 변환은 필터링된 신호의 길이를 입력신호의 길이와 동일하게 유지하면서 다음 스케일의 신호를 얻기 위해 필터계수들 사이에 0을 삽입하여 신호와 컨볼루션(convolution) 함으로써 간축과정을 수행하지 않는다. 그리고 컨볼루션 수행시 신호의 가장자리는 주기적인 확장(periodic expansion) 방법을 이용하며, 신호에서 급격히 변하는 부분의 위치를 각 스케일에서 잘 보존하는 특징이 있다[4,10].



(a) General discrete wavelet transform



(b) Undecimated wavelet transform

그림 1. 이산 웨이브렛 변환을 위한 dyadic 필터뱅크 구조
Fig. 1. Structure of dyadic filterbank for discrete wavelet transform

2.3. Cycle-Spinning 기법

Coifman[3]에 의해 제안된 Cycle-Spinning 기법은 분석 과정을 수행하기 전에 입력신호를 강제로 왼쪽 또는 오른쪽으로 순환이동을 시키고, 합성과정을 거친 다음 처음과 반대 방향으로 순환이동을 시킴으로써 원래의 신호를 복원할 수 있다는 이론이다. 신호 $x_t (0 \leq t \leq n)$ 가 h 만큼 순환이동된 경우 식 (6)으로 표현할 수 있으며, S_h 는 신호를 h 만큼 왼쪽으로 순환이동을 한다는 의미이다.

$$(S_h x)_t = x_{(t+h) \bmod n} \tag{6}$$

웨이브렛 변환을 기반으로 하는 잡음제거 방식의 경우 주로 불연속점(discontinuity) 주변과 신호가 급격하게 변하는 부분에서 원신호의 레벨을 기준으로 undershoot과 overshoot을 반복하는 pseudo-Gibbs 현상이 나타난다. 특히, 잡음을 제거해야 하는 신호가 청각적인 정보를 가지는 신호인 경우 이러한 현상은 “clicks(또는 blips)” 성분으로서 귀에 거슬리는 소리를 발생시킨다. 웨이브렛 변환 과정에서 서로 컨볼루션 되는 신호와 기저함수간의 mis-alignment가 “clicks” 성분이 발생하게 되는 가장 큰 원인이며, 입력신호의 모양에 따라 출력신호에서 나타나는 잔여잡음의 영향도 달라진다[9]. 따라서 Cycle-Spinning 기법으로 입력신호를 적절히 순환이동 시킨다면 신호와 웨이브렛 기저함수간의 mis-alignment를 해결함으로써 잔여잡음을 효과적으로 감소시킬 수 있다. 그러나 Cycle-Spinning 기법을 적용할 때 입력신호에 여러 개의 불연속점이 있는 경우 특정 불연속점에 대해서는 최적의 이동이 될 수 있지만 다른 불연속점에 대해서는 최악의 이동이 될 수도 있기 때문에 최적의 순환이동 값을 찾는 것은 매우 어렵다고 볼 수 있다. 따라서 본 연구에서는 이동값을 변화시키면서 실험을 하고 그 결과를 비교하였다.

III. 웨이브렛 변환을 이용한 음성개선

웨이브렛 변환을 이용한 thresholding 기법의 기본 원리는 백색 가우시안 잡음에 의해 오염된 신호를 웨이브렛 변환했을 때 각 스케일에 포함된 잡음 성분은 산호 성분의 크기보다 상대적으로 작은 값을 가지므로 적절한 문턱값 이하의 값을 제거한 후 다시 합성함으로써 효과적으로 잡음을 제거할 수 있다는 것이다[6].

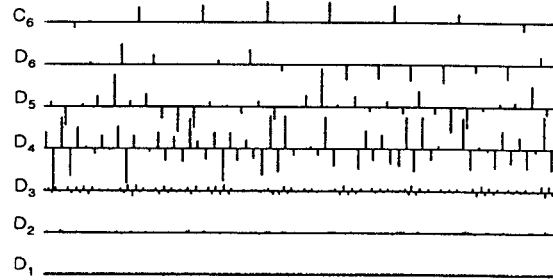
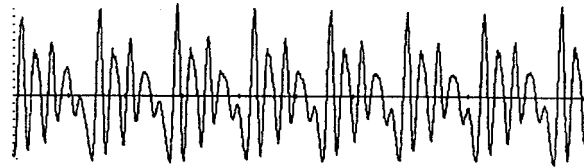
잡음에 의해 오염된 음성신호 x 는 식 (7)과 같이 원신호 s 와 평균이 0이고 표준편차가 σ 인 잡음 n 의 합으로 나타낼 수 있으며, 웨이브렛 변환하면 식 (8)과 같다. 즉, W 를 직교 웨이브렛 변환이라 가정하면, $X = Wx$ 를 의미한다.

$$x = s + \sigma n \tag{7}$$

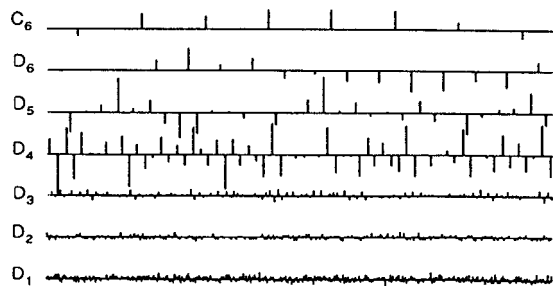
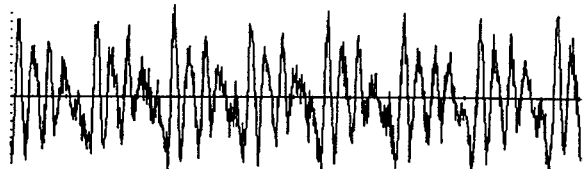
$$X = S + N \tag{8}$$

식 (9)는 원음성의 추정값 \hat{s} 을 얻기 위한 역웨이브렛 변환식으로서 \hat{S} 과 M 은 각각 원음성의 웨이브렛 변환 계수 S 의 추정값과 W 의 역변환 행렬을 의미한다. \hat{S} 은 잡음음성의 웨이브렛 계수인 X 를 shrinking 또는 killing함으로서 얻는다.

$$\hat{s} = M\hat{S} \tag{9}$$



(a) Original speech signal



(b) Noisy speech signal(SNR: 10dB)

그림 2. 음성신호의 각 스케일별 웨이브렛 계수의 예
Fig. 2. An example of wavelet coefficients for each scale.

그림 2는 일반적인 웨이브렛 변환을 실제 음성신호에 적용한 예로서 원음성과 SNR 10dB 잡음음성의 유성음 구간을 스케일 6까지 분해한 경우이다. D_1 에서 D_6 은 스케일 1에서 6까지의 상세(detail)신호를 의미하고, C_6 은 스케일 6의 개략(coarse)신호를 의미한다. 유성음의 경우

신호 성분은 저주파 영역에 해당되는 높은 스케일의 계수에 집중되어 나타나고, 잡음 성분은 모든 스케일에서 고르게 분포하므로 잡음의 영향은 고주파 영역에 해당되는 낮은 스케일의 웨이브렛 계수에서 더 우세하게 나타난다고 볼 수 있다. 그러나 상대적으로 무성음의 경우는 웨이브렛 변환 영역에서 잡음과 유사한 특성을 가지기 때문에 잡음을 제거하고자 할 때 무성음에 대한 특별한 고려가 필요하다[7].

그림 3은 본 연구에서 적용한 음성개선 알고리즘에 대한 블록도이다. 일반적인 신호의 thresholding 기법과의 차이점은 잡음제거 과정에서 발생하는 음성신호의 손실을 줄이기 위해서 음성분류 단계를 포함한다는 것이다.

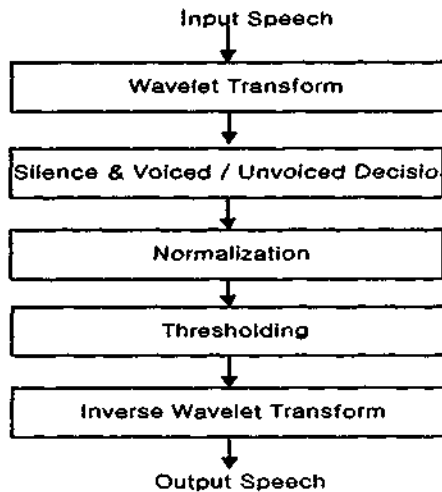


그림 3. 웨이브렛 변환을 이용한 음성개선 블록도
Fig. 3. Block diagram of the speech enhancement algorithm.

잡음제거 과정에서 무성음 성분의 손실을 줄이기 위해서 웨이브렛 계수 스케일 4의 MLR(Maximum Likelihood Ratio)[10]인 MLR_4 와 $D_1 Sum D_2$ 를 사용하여 먼저 무성음을 검출한 다음 유성음이나 묵음과는 다른 잡음처리 과정을 수행하였다. MLR_k 와 $D_1 Sum D_2$ 는 식 (10)과 (11)로 정의하였는데, 여기서 σ_k^2 와 σ_{noise}^2 는 각각 해당 프레임에 대하여 스케일 k 에서 웨이브렛 계수의 분산과 배경잡음에 해당하는 묵음구간의 분산을 가리킨다.

$$MLR_k = \frac{1}{2} \left| \ln \left(\frac{\sigma_k^2}{\sigma_{noise}^2} \right) - \frac{\sigma_k^2}{\sigma_{noise}^2} \right| \quad (10)$$

$$D_1 Sum D_2 = \log(\sigma_1^2 + \sigma_2^2) + MLR_1 \quad (11)$$

그림 4는 무성음 검출 알고리즘을 나타낸다. 그림에서 T_M 과 T_D 는 각각 묵음구간으로 간주한 첫 번째 다섯 프레임의 MLR_4 와 $D_1 Sum D_2$ 의 평균값을 나타내는데,

MLR_4 가 T_M 보다 큰 경우 유성음으로 판정하고, 유성음으로 판정되지 않은 프레임 중에서 $D_1 Sum D_2$ 가 T_D 보다 큰 경우 무성음으로 판정한다.

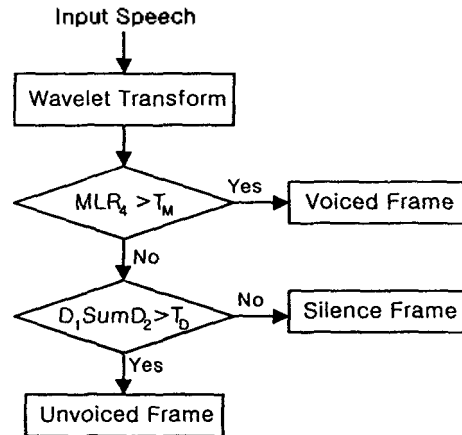


그림 4. 무성음 검출 알고리즘
Fig. 4. Algorithm of the unvoiced region detection.

Thresholding 기법에서 잡음 레벨을 추정하여 모든 스케일의 계수를 추정된 값으로 나누는 과정이 정규화 단계이다. 본 연구에서는 평균이 0인 백색 가우시안 잡음의 경우에 잘 맞는 추정방법으로 알려진 MAD(Median Absolute Deviation)[6,8]를 적용하였다. 저주파 성분을 많이 가지는 유성음은 첫 번째 스케일의 상세신호(D_1)에서, 상대적으로 고주파 성분을 많이 가지는 무성음은 첫 번째 스케일의 개략신호(C_1)에서 음성 성분에 비해 잡음이 더 많이 존재하기 때문에 음성의 특징을 고려하여 무성음의 경우는 C_1 에서 잡음 레벨을 추정하고, 그 외의 프레임은 가장 고주파 성분을 가지는 계수인 D_1 에서 추정하였다.

Thresholding 함수에는 hard, soft, semisoft 함수[6-8]가 있으며, 본 연구에서는 문턱값 λ 이하의 값은 0으로 killing하고, 문턱값 이상의 값은 shrinking하는 기법인 soft-thresholding 함수를 사용하였다. 식 (12)는 soft-thresholding 함수를 표현한 것이며, 문턱값 λ 는 식 (13)을 이용하였다[6].

$$T_{soft}(X) = \begin{cases} \text{sgn}(X)(|X| - \lambda), & |X| \geq \lambda \\ 0, & |X| < \lambda \end{cases} \quad (12)$$

$$\lambda = \hat{\sigma} \sqrt{2 \log(L)/L} \quad (13)$$

여기서, $\hat{\sigma}$ 과 L 은 각각 MAD/0.6745와 신호의 길이를 의미하며, 정규화 과정을 거친 계수에서 MAD를 얻는다. 그리고 무성음에서 D_1 의 경우는 문턱값으로 $\frac{\lambda}{2}$ 를 적용하여 고주파 성분의 손실을 줄이고자 하였고, 그 외

의 경우는 모든 스케일 신호에 대해 동일한 문턱값으로 soft-thresholding을 하였다.

그림 3의 음성개선 알고리즘을 CWD, CSD과 UWD의 각 방식에 적용하는 과정은 다음과 같다.

- ① 분석 프레임 단위로 웨이브렛 변환을 수행한다. CWD과 UWD 방식은 각각 그림 1의 필터뱅크를 구성하여 웨이브렛 변환을 수행하고, CSD 방식은 먼저 식 (6)과 같이 입력신호를 순환이동 시킨 다음 CWD 방식과 같이 간축과정을 포함하는 일반적인 웨이브렛 변환을 수행한다.
- ② 그림 4와 같이 무성음을 검출한다.
- ③ ②과정에서 분류된 프레임의 특성별로 각각 D_1 과 C_1 에서 잡음 레벨을 추정하여 모든 스케일의 계수를 정규화한다.
- ④ Soft-thresholding을 수행한다.
- ⑤ 역웨이브렛 변환을 수행한다. 특히, CSD 방식은 역변환을 한 다음 ①의 반대 방향으로 순환이동을 시켜야한다. 그리고 앞에서도 언급하였듯이 최적의 순환이동 값을 찾는 것이 어렵기 때문에 본 연구에서는 순환이동 범위 H 만큼의 합성음을 평균하는 방법을 적용하였다. 식 (14)는 CSD 방식에서 합성음을 얻는 과정을 표현한 것이다. S_{-k} 는 처음과 반대 방향으로 순환이동을 한다는 의미이며, thresholding 기법을 적용한 잡음제거 과정을 THR로 표현하였다.

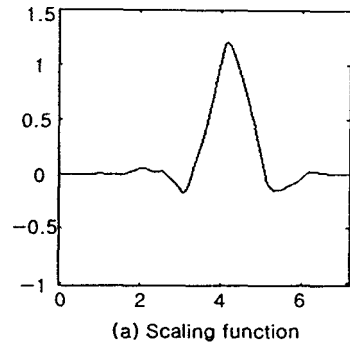
$$\hat{s} = Average_{k \in H} [S_{-k}(THR(S_k(x)))] \quad (14)$$

IV. 실험 과정 및 검토

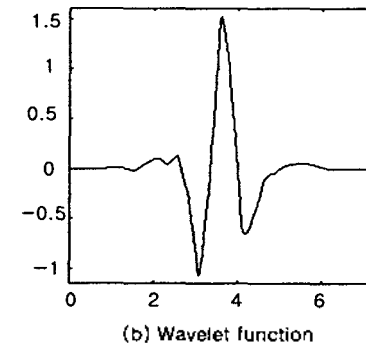
4.1. 실험 데이터 및 조건

본 연구에서는 10kHz로 샘플링 되고 16bits로 양자화된 음성신호에 백색 가우시안 잡음을 첨가하여 SNR이 0에서 15dB인 잡음음성을 대상으로 실험을 하였다. 분석 프레임은 256샘플이고 합성시에는 128 샘플 단위로 overlap-and-add를 수행하였다. Symlets 8-tap 필터[3]를 이용하여 스케일 4까지 웨이브렛 변환을 수행한 다음 각각의 잡음제거 알고리즘을 적용하였다. 그리고 CSD 방식에서 순환이동 범위 H 는 각각 8, 32, 128, 256으로 설정하였다.

그림 5는 실험에서 사용한 Symlets 8-tap 웨이브렛을 나타낸다. Symlets 웨이브렛은 Daubechies[3]가 제안한 웨이브렛으로 모양이 거의 좌우 대칭을 이루며 Daubechies 웨이브렛과 유사한 특성을 가진다. 웨이브렛 변환을 위해서 Symlets 필터 이외에 Daubechies, Coiflets[3] 등의 다른 필터를 사용하여 보았으나 성능에 큰 차이를 보이지는 않았고 필터링시에 계산량을 고려하여 8-tap 필터를 사용하였다. 그리고 각 방식의 성능을 비교하기 위한 척도로 SNR과 LPC 캡스트럼 거리를 이용하였다.



(a) Scaling function



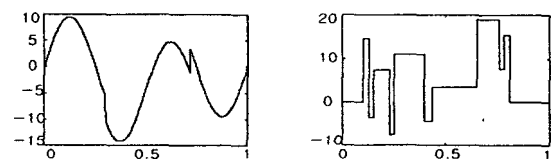
(b) Wavelet function

그림 5. Symlets 8-tap 웨이브렛
Fig. 5. Symlets 8-tap wavelet.

4.2. 실험 결과 및 검토

실제 음성신호에 적용하기 전에 테스트 신호로 널리 사용되고 있는 HeavySine과 Blocks 신호를 이용하여 CWD, CSD, UWD의 세가지 방식에 대해 성능을 비교하여 보았다. 신호의 길이는 2048 샘플이고 첨가된 잡음은 평균이 0이고 분산이 1인 백색 가우시안 잡음이다. 웨이브렛 변환을 수행한 후 개략신호를 제외한 모든 상세신호를 대상으로 soft-thresholding 기법을 적용하였다.

그림 6은 각 방식에 따라 잡음이 제거된 신호의 파형을 비교한 것이다. 신호가 급격하게 변하는 부분에서 원신호와 오차가 크게 나타나고 CSD과 UWD 방식이 불연속점 주변에서 주로 발생하는 "clicks" 성분이 일반적인 웨이브렛 변환을 이용한 경우에 비해 감소함을 볼 수 있다. 특히 (e)의 UWD 방식은 다른 방식에 비해 신호가 스무딩(smoothing)되는 결과를 보였다.



(a) Original signal(left: HeavySine, right: Blocks)

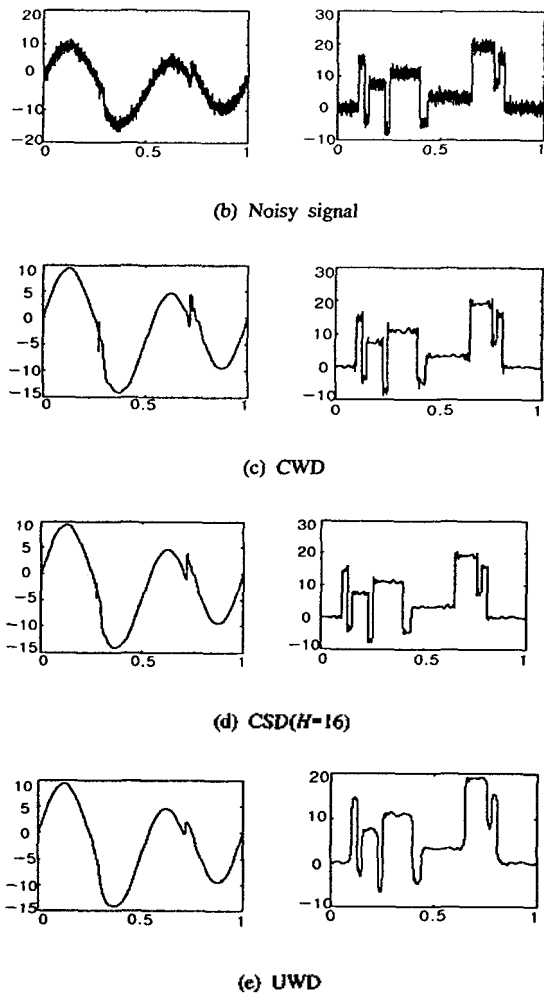


그림 6. 텍스트 신호에 대한 각 방식별 파형 비교
Fig. 6. Waveform comparison of denoised test signal depending on each method.

그림 7은 무성음을 검출하기 위한 파라미터인 MLR_4 와 D_1SumD_2 를 SNR 10dB인 잡음음성 /Should we chase those cowboys?라는 문장에 적용한 결과이다. 그림 (b)의 MLR_4 는 유성음 구간을 잘 검출하고 있으며, 그림 (c)에서 점선으로 표시한 부분이 음성신호에서 무성음에 해당하는 구간이다. 즉, MLR_4 에 의해 유성음으로 판정되지 않은 프레임 중에서 D_1SumD_2 가 묵음구간의 평균 값보다 큰 구간을 무성음으로 판정하게 된다. 실제로 잡음 환경에서는 이러한 유성음과 무성음의 분류가 정확하게 이루어지는 것은 어렵다고 알려져 있다. 실험 결과 유성음의 검출은 대체로 좋은 성능을 보였으나 잡음 레벨이 높아질수록 묵음과 무성음을 정확하게 구분하는 것이 어려웠고, 특히 SNR이 낮은 신호에서 이러한 점이 음질 저하의 큰 원인이 될 수 있다고 생각된다.

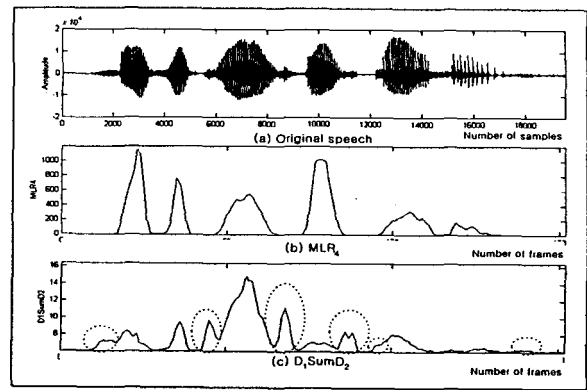
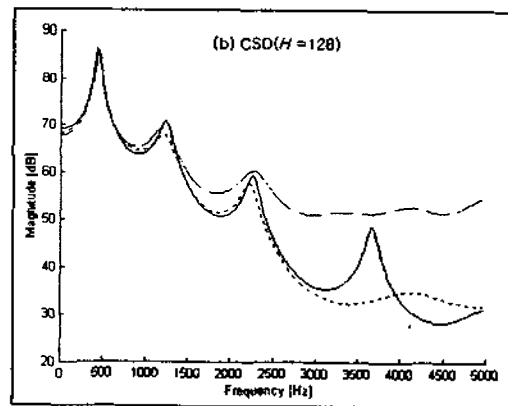
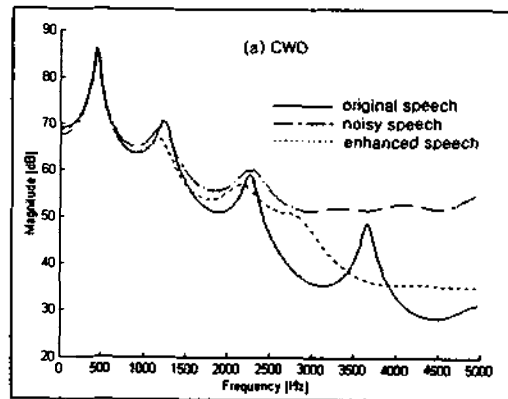


그림 7. 무성음 검출 파라미터
Fig. 7. Contours of parameter for unvoiced sound detection.

그림 8과 9는 각 방식의 합성음에서 유성음 및 무성음 구간에 대한 LPC 스펙트럼을 보여주고 있다. 전체적으로 thresholding 과정에서 고주파 영역에서 다소의 손실이 나타나는데, 이는 thresholding 과정에서 전체 스펙트럼에 미치는 영향이 저주파수 영역보다는 고주파수 영역에서 상대적으로 더 큰 영향을 미쳐서 고주파수 부분이 손실을 입은 것으로 생각이 된다. 그리고 CSD와 UWD 방식이 CWD 방식에 비해 원음성의 스펙트럼을 잘 따라가고 있음을 볼 수 있다.



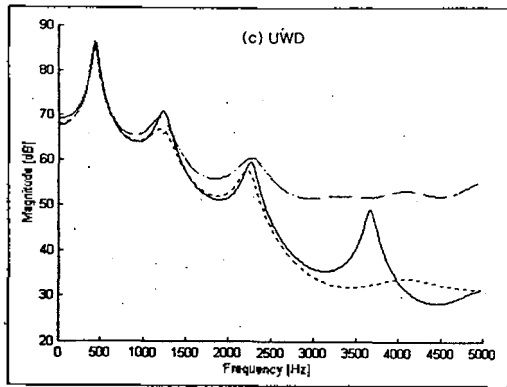


그림 8. 각 방식에 따른 합성음의 LPC 스펙트럼
Fig. 8. LPC spectra of synthesized speech(SNR: 10dB).

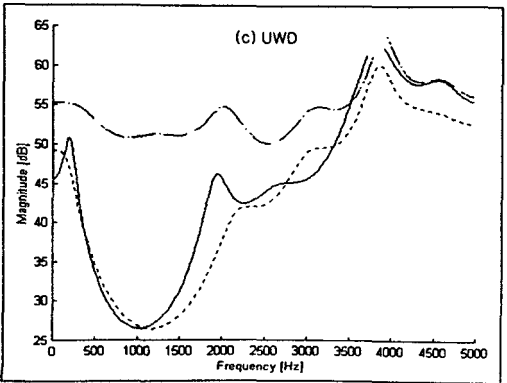
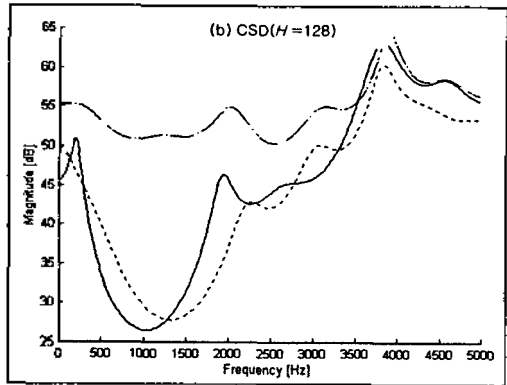
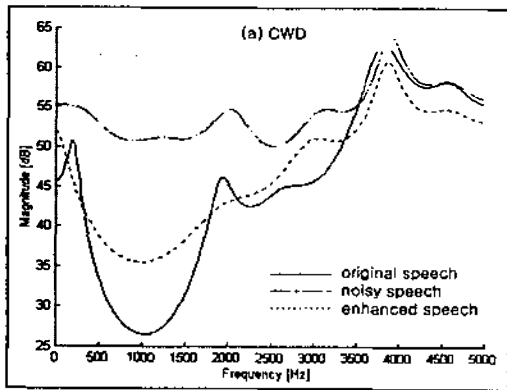


그림 9. 각 방식에 따른 합성음의 LPC 스펙트럼(계속)
Fig. 9. LPC spectra of synthesized speech(continued).

그림 7의 음성신호에 대한 각 방식에 따른 SNR과 켈스트럼 거리에 대한 결과는 표 1과 2에 나타나 있다. 그림 10은 표 1과 2의 결과를 그래프로 표현한 것이다. 표 1에서 알 수 있듯이 Cycle-Spinning 기법을 이용한 CSD 방식과 비간축 웨이브렛 변환을 이용한 UWD 방식이 일반적인 웨이브렛 변환을 이용한 CWD 방식에 비해 평균 1dB이상의 성능 향상을 보였다. SNR에 비해 주관적인 성능평가와 더 높은 상관관계를 갖는 켈스트럼 거리의 경우에도 두 방식이 더 나은 성능을 보였다. 특히, CSD 방식의 경우 순환이동 범위 H 의 증가에 따른 큰 성능 차이는 보이지 않았고 UWD 방식과 거의 비슷한 성능을 보였다.

표 1. 각 방식에 따른 합성음의 SNR 비교

Table 1. Comparison of SNR of synthesized speech depending on each method.

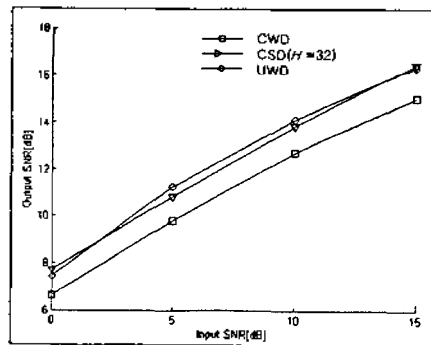
Input SNR	Output SNR[dB]				
	A	B	C	D	E
0dB	6.6304	7.4561	7.7263	7.7135	7.7422
5dB	9.8062	11.2475	10.8563	10.8123	10.8794
10dB	12.6833	14.0779	13.8242	13.8074	13.8310
15dB	15.0107	16.3438	16.4556	16.4388	16.4498

A: CWD B: UWD C: CSD($H=8$)
D: CSD($H=32$) E: CSD($H=256$)

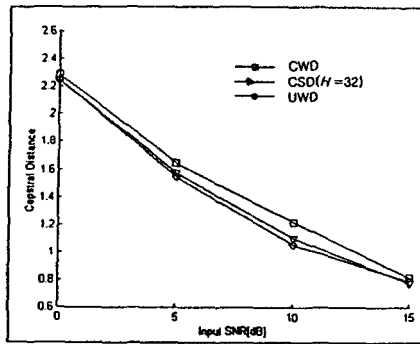
표 2. 각 방식에 따른 합성음의 LPC 켈스트럼 거리 비교

Table 2. Comparison of LPC cepstrum of synthesized speech depending on each method.

Input SNR	Cepstral Distance					
	잡음음성	A	B	C	D	E
0dB	2.7029	2.2900	2.2464	2.2673	2.2500	2.2597
5dB	2.2867	1.6477	1.5522	1.5831	1.5786	1.5763
10dB	1.8357	1.2239	1.0580	1.0960	1.1043	1.0874
15dB	1.3882	0.8147	0.7892	0.7765	0.7735	0.7651



(a) SNR



(b) Cepstral distance

그림 10. SNR과 켈스트럼 거리 비교
Fig. 10. Comparison of SNR and cepstral distance.

V. 결 론

본 연구에서는 Cycle-Spinning 기법을 이용한 웨이브렛 변환과 비간축 웨이브렛 변환을 이용하여 잡음에 의해 오염된 음성신호에서 효과적으로 잡음을 제거할 수 있음을 보였으며, 일반적인 웨이브렛 변환을 이용한 결과와 성능을 비교·분석하였다. 특히, 잡음제거 과정에서 잡음과 유사한 특징을 갖는 무성음의 손실을 줄이기 위하여 먼저 MLR_1 와 $D_1 SumD_2$ 파라미터를 이용하여 무성음을 검출하였다. 그리고 음성신호의 특성을 고려하여 저주파 성분을 많이 가지는 유성음의 경우는 첫 번째 스케일의 상채신호에서, 고주파 성분을 많이 가지는 무성음의 경우는 첫 번째 스케일의 개략신호에서 잡음 레벨을 추정하였으며, soft-thresholding 함수를 이용하여 서로 다른 방법으로 잡음을 제거하였다.

Cycle-Spinning 기법을 이용한 CSD 방식과 비간축 웨이브렛 변환을 이용한 UWD 방식은 일반적인 웨이브렛 변환을 이용한 CWD 방식에 비해 계산량이 많은 단점을 가지고 있으나, 실험결과 SNR과 켈스트럼 거리에서 우수한 성능을 보였다. 그리고 CSD 방식의 경우 순환이동 범위의 증가에 따른 큰 성능 차이는 보이지 않았다. 또한 CSD와 UWD 방식은 청취실험 결과 기존의 방식에 비해 "clicks" 성분이 다소 감소하는 것으로 나타나서 음질이 향상됨을 확인할 수 있었다. 그러나 thresholding 기법을 적용하여 잡음을 처리하는 과정에서 저주파 영역에 비해 고주파 성분의 손실이 나타나는데 이러한 손실을 줄이기 위한 추가적인 연구가 필요하며, 특히 SNR이 낮은 신호에서는 음성분류 과정에서의 에러가 음질 저하의 큰 원인이 되므로 잡음에 강인한 무성음과 묵음 분류 파라미터가 요구된다.

감사의 글

본 연구는 한국과학재단의 특정기초연구과제(과제번호: 1999-2-303-001-3) 연구비 지원으로 수행되었음.

참 고 문 헌

- O. Rioul and M. Vetterli, "Wavelet and Signal Processing," *IEEE Signal Processing Magazine*, pp.14-38, October, 1991.
- I. Daubechies, "The Wavelet Transform Time-Frequency Localization and Signal Analysis," *IEEE Trans. on Information Theory*, Vol.36, No.5, pp.961-1005, 1990.
- I. Daubechies, "Ten Lectures on Wavelets," SIAM, 1992.
- S. G. Mallat and W. Hwang, "Singularity detection and processing with wavelets," *IEEE Trans. on Information Theory*, Vol.38, pp.617-643, 1992.
- Mark J. Shensa, "The Discrete Wavelet Transform: Wedding the A Trouns and Mallat Algorithms," *IEEE Trans. on Signal Processing*, Vol.40, No.10, October, 1992.
- D. L. Donoho, "De-Noising by Soft-Thresholding", *IEEE Trans. on Information Theory*, pp.961-1005, 1995.
- Jong Won Seok and Keun Sung Bae, "Speech Enhancement with Reduction of Noise Components in the Wavelet Domain," *ICASSP'97*, pp.1323-1326, 1997.
- D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation via wavelet shrinkage," *Biometrika*, Vol.81, pp.425-455, 1994.
- R. R. Coifman and D. L. Donoho, "Translation-Invariant De-Noising," Technical Report, Department of Statistics, Stanford University, 1995
- 손영호, 배건성, "웨이브렛 변환을 이용한 유성음/무성음/묵음 분류," *음성통신 및 신호처리 워크샵 논문집*, pp.449-453, 1998.

▲ 한 미 경(Mi Kyung Han)



1998년 2월: 경북대학교 전자공학과 (공학사)
2000년 2월: 경북대학교 전자공학과 (공학석사)
2000년 3월~현재: 삼성전자 정보통신부
* 주관심분야: 디지털 신호처리, 음성신호처리, 웨이브렛 이론

▲ 배 건 성(Keun Sung Bae)



1977년 2월: 서울대학교 전자공학과 (공학사)
1979년 2월: 한국과학기술원 전기 및 전자공학과 (공학석사)
1989년 5월: University of Florida (공학박사)
1979년 3월~현재: 경북대학교 전자공학과 교수

* 주관심분야: 음성분석 및 인식, 디지털 신호처리, 디지털 통신, 음성 부호화, 웨이브렛 이론 등.