

# 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기 설계

## Design of the Noise Suppressor Using the Perceptual Model and Wavelet Packet Transform

김 미 선\*, 박 서 영\*, 김 영 주\*, 이 인 성\*

(Mi Seon Kim\*, Seo Young Park\*, Young Ju Kim\*, In Sung Lee\*)

\*충북대학교 전파공학과

(접수일자: 2006년 9월 19일; 채택일자: 2006년 10월 17일)

본 논문은 인지 모델과 웨이블릿 패킷 변환을 이용하여 단일 채널에서 유색잡음 또는 비정지적 성격의 잡음을 제거하는 데 목적을 두고 있다. 이러한 잡음은 부대역을 나누어 접근해야하며, 잔여잡음과 음성의 왜곡으로 인한 문제를 해결하기 위해 웨이블릿 패킷 변환 후 웨이블릿 계수 문턱값을 적절히 개선해야 한다. 본 논문에서 부대역은 웨이블릿 패킷 변환 후에 스케일과 임계대역을 매칭하여 설계하였으며, 웨이블릿 계수 문턱값은 세그멘탈 신호대잡음비 (seg\_SNR)와 노이즈마스킹 임계값 (Noise Masking Threshold : NMT)을 이용하여 적응적으로 계산했다. 결과적으로 TTA 표준인 EVRC 잡음 제거기와 유사한 성능을 가졌으며, 웨이블릿 변환 후 웨이블릿 계수에 Universal 문턱값을 적용하는 것보다 PESQ-MOS 값이 0.29 높았다. 인코딩과 디코딩 후 PESQ-MOS 값은 EVRC 잡음 제거기보다 0.23 정도 우수한 성능을 가졌다.

**핵심용어:** 음질 개선, 잡음 제거, 웨이블릿 패킷 변환, 인지 모델

**투고분야:** 음성 처리 분야 (2,3)

In this paper, we propose the noise suppressor with the perceptual model and wavelet packet transform. The objective is to enhance speech corrupted colored or non-stationary noise. If corrupted noise is colored, subband approach would be more efficient than whole band one. To avoid serious residual noise and speech distortion, we must adjust the Wavelet Coefficient Threshold (WCT). In this paper, the subband is designed matching with the critical band and WCT is adapted noise masking threshold (NMT) and segmental signal to noise ratio (seg\_SNR). Consequently, it has similar performance with EVRC in PESQ-MOS. But it's better than wavelet packet transform using universal threshold about 0.289 in PESQ-MOS. The important thing is that it's more useful than EVRC in coded speech. In coded speech, PESQ-MOS is higher than EVRC about 0.23.

**Key words:** Speech enhancement, Noise suppression, Wavelet packet transform, Perceptual model

**ASK subject classification:** Speech signal processing (2.3)

### I. 서론

선형예측 (LP) 모델 또는 하모닉 (Harmonic) 모델을 사용하는 저속 보코더는 인간의 성대를 모델링한 구조를 가지고 있다. 따라서 음성과 특성이 다른 주변잡음을 포함하는 음성에 대하여 음질저하가 매우 심하기 때문에

보코더에서의 잡음 제거 과정은 필수적이라고 할 수 있다. 그 중에서도 운송수단 (비행기, 자동차, 전투기) 내부에서 통신 할 때 발생하는 잡음제거에 관한 연구가 많이 진행되고 있으며, 그 외에 의료 생체 신호처리와 같이 정확성이 요구되는 분야에서도 잡음 제거 연구는 필수적이라 할 수 있다.

본 논문에서는 시간에 따른 대역별 주파수 특성변이가 다른 잡음의 제거 성능을 향상시키고자 웨이블릿 패킷 변환 (WPT : Wavelet Packet Transform)을 이용하였

다. 웨이블릿 변환을 이용한 분석 방법은 신호처리를 위해 꾸준히 연구되어 왔으며, 이러한 분석 방법은 신호를 다해상도 필터뱅크 함으로써 쉽게 구현될 수 있다 [1]. 웨이블릿 변환을 이용한 분석법은 영상신호처리나 음성 신호 처리에도 꾸준히 성공적으로 적용되어왔다 [2]-[7].

Donoho와 Johnston은 백색잡음을 제거하기 위해 웨이블릿 계수에 Universal 문턱값을 적용하는 방안을 제안했다 [8][9]. 이 방법은 백색잡음에 적절한 방안이었으나, 유색잡음에는 잘 적용되지 않았다. 유색잡음은 각 부대역 별로 웨이블릿 계수에 다른 문턱값을 적용함으로써 성능향상을 꾀할 수 있었다. Sheikhzadeh Abutaleb는 저대역의 문턱값은 낮추고 고대역의 문턱값은 높여 연성 문턱값을 적용하는 방안을 고안했다 [10]. 하지만 이와 같은 웨이블릿 계수 문턱값 (WCT : Wavelet Coefficient Threshold) 적용 기법은 음성인지 잡음인지를 잘 구분하여 적용해야 한다. Bahoura와 Rouat은 레벨 의존형 문턱값을 제안했으며, 음성구간인지 잡음구간인지를 알기 위해 Teager energy operator를 사용하였다 [11]. 이와 비슷하게 음성과 잡음을 구별하는 연구로 NMI (Noise Magnitude Intensity)와 SDI (Speech Dominant Indicator)를 사용하는 기법이 있다 [12]. 또한 빠른 잡음 변화에 대해 효과적일 수 있도록 묵음구간으로부터 잡음을 추정하지 않고 음성/잡음 우세 결정에 의해 잡음처리를 하는 연구도 이루어졌다 [13].

본 논문에서는 유색잡음이나 비정지적 잡음을 제거하기 위해 5단계의 웨이블릿 패킷 분해를 통해 얻은 웨이블릿 계수들과 임계대역을 매칭하여 부대역을 구조화 하였으며, 잡음 마스킹에 사용되는 문턱값은 세그멘탈 신호대잡음비와 노이즈마스킹 임계값을 이용하여 적용되도록 설계하였다.

이는 사람의 청각적 인지 특성을 이용하여 잡음을 제거함으로써 주관적인 성능의 향상을 기대할 수 있기 때문이다. 또한 음성과 유사한지 잡음과 유사한지를 계산하여 반영하는 특성에 따라 계수의 전체적인 모양을 달리 하는 구조를 가지므로 별도로 유무성음 구별 과정 없이 웨이블릿 계수 문턱값을 자동 조절하도록 설계하였다.

본 논문은 2장에서 제안된 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기의 구조에 대해 설명하고, 3장에서 임계대역을 반영한 웨이블릿 패킷변환과 노이즈마스킹 임계값과 세그멘탈 신호대잡음비를 웨이블릿 계수 문턱값에 적용하는 방법에 대해 설명한다. 4장에는 실험 결과를 설명하며 5장에서 결론을 맺는다.

## II. 제안하는 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기

### 2.1. 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기 구조

입력된 신호  $x[n]$ 을 윈도링 한 후 128 포인트 FFT 취하여 대역 별로 에너지의 편차를 구한다. 이것을 이용하여 잡음인지 음성인지를 판단하며, 잡음일 경우 잡음을 추정하는 알고리즘을 수행한다 [14].

윈도링된 입력된 신호  $x_m[n]$ 과 추정된 잡음을 이용하여 세그멘탈 신호대잡음비를 구하게 되며,  $x_m[n]$ 을 5단계의 웨이블릿 패킷변환 후 추정된 노이즈 마스킹 임계값 (NMT : Noise Masking Threshold)과 함께 최종적인 문턱값 갱신에 사용된다.

$x_m[n]$ 에 개선된 문턱값을 연성 문턱값 형태로 적용해 다시 역변환을 거친 후 이전 프레임과의 연계성을 형성 해주면 음질 개선 과정이 끝나게 된다.

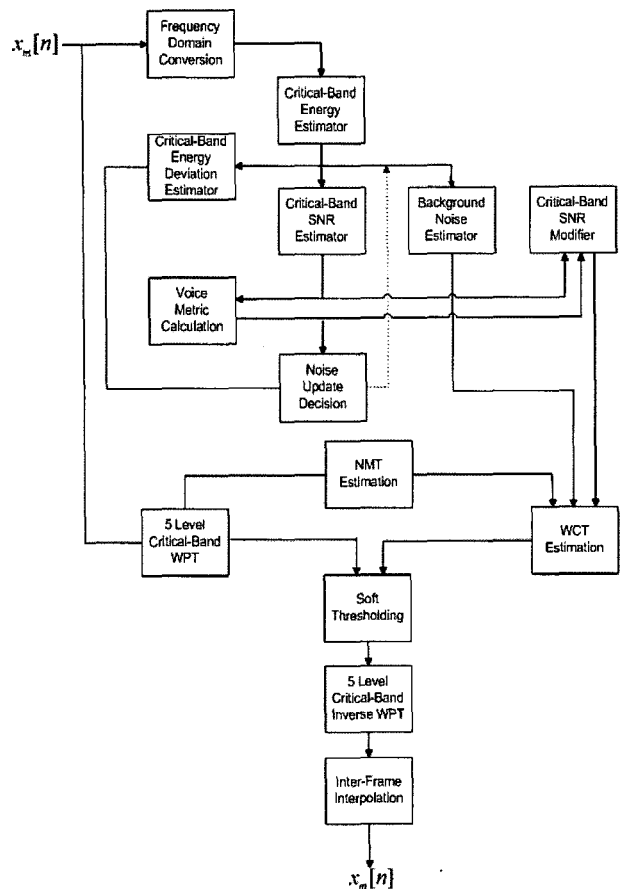


그림 1. 제안하는 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기 구조  
Fig. 1. Structure of proposed noise suppressor using perceptual mode and wavelet packet transform.

### 2.2. 웨이블릿을 이용한 음성신호의 표현

잡음 섞인 음성신호는 식 (1)과 같이 음성신호와 잡음 신호의 합으로 표현할 수 있다.

$$x_m[n] = s_m[n] + w_m[n] \tag{1}$$

여기에서  $n = 0, 1, 2, \dots, N-1$  이며,  $N$ 은 한 프레임의 샘플 수,  $x_m[n], s_m[n], w_m[n]$ 은 각각  $m$  번째 프레임의 잡음 섞인 음성신호, 음성신호, 잡음신호를 나타낸다.

잡음 섞인 음성신호  $x_m[n]$ 의 프레임의 길이를  $[0, 1]$ 로 정규화 하여 웨이블릿 함수와 스케일링 함수로 표현하면 식 (2)와 같다 [1].

$$x[n] = \sum_{j=L+1}^J \sum_{i=1}^{2^{j-1}} \sum_{k=0}^{2^j-1} X_{j,k}^i \cdot \psi_{j,k}^i[n] + \sum_{k=0}^{2^J-1} X_{J,k}^i \cdot \phi_{J,k}[n] \tag{2}$$

여기서  $L = -\log_2 N$ ,  $J$ 는 가장 낮은 레벨의 인덱스,  $i, j$ 는  $j$ 번째 레벨의 부대역의 수,  $i$ 는 부대역의 인덱스,  $k$ 는 레벨을 각각 의미하며  $\psi_{j,k}^i[n]$ 는 웨이블릿 모함수를  $\phi_{j,k}[n]$ 는 스케일링 함수를 의미한다. 또한  $X_{j,k}^i$ 는 웨이블릿 계수,  $X_{J,k}^i$ 는 스케일링 계수를 각각 의미한다. 이들은 식 (3)과 같이 신호와의 내적을 통하여 구하여 진다.

$$X_{j,k}^i = \langle x[n], \psi_{j,k}^i[n] \rangle \tag{3}$$

$$X_{J,k}^i = \langle x[n], \phi_{J,k}[n] \rangle$$

### 2.3. 인지 모델을 이용한 웨이블릿 패킷 변환

청각 마스크는 인접한 신호 요소에 관한 청각적 인지 현상이다. 두 신호 성분이 동시에 발생했을 경우 상대적으로 큰 신호가 작은 신호를 가리게 된다. 이 때 작은 신호는 존재하지만 들리지 않게 된다. 이 때의 마스크 임계값은 바크 스케일 (Bark scale) 별로 찾게 되는데, 그 이유는 바크 스케일 별로 마스크 임계값이 비교적 일정하게 유지되기 때문이다. 이러한 특성을 웨이블릿 계수 임계값으로 이용할 경우 유색잡음의 개선을 기대할 수 있다. 일단, 신호를 임계대역을 반영한 웨이블릿 패킷 변환을 이용하여 웨이블릿 계수 (WC : Wavelet Coefficient)로 변환한다. 각각의 임계대역마다 계산된 노이즈 마스크 임계값을 이용하여 웨이블릿 계수 임계값을 재조정한다. 개선된 웨이블릿 계수를 역으로 합성하면 개선된 음성을 얻게 된다.

이 때의 웨이블릿 계수는 신호를 웨이블릿 직교 기저들로 분해함으로써 얻어지게 된다. 이는 신호를 저대역

통과 필터 (LPF : Low Pass Filter)와 고대역 통과 필터 (HPF : High Pass Filter)한 후 2배로 다운 샘플링 함으로써 얻게 된다. 4 kHz 에 해당하는 입력 신호를 입력으로 하였다면, 0-2kHz와 2-4kHz의 두 대역 신호를 얻게 된다. 이와 같은 과정을 반복함으로써 필터뱅크 (Filter Bank)을 얻을 수 있게 되고, 이 필터뱅크와 임계대역을 매칭하면 인지적인 부대역 설계가 가능해진다.

## III. 웨이블릿 패킷 변환과 인지 모델을 이용한 부대역 설계 및 문턱값 개선

### 3.1. 인지 모델을 반영한 부대역 설계

마스킹 임계값은 주파수 요소의 임계대역과 밀접하게 연관되어 있다. 이 중심주파수는 임계대역을 (Critical Band Rate)  $\xi$ 라고 알고 있는 부대역에 위치하고 식 (4)와 같이 표현할 수 있다.

$$\xi = 13 \arctan(7.6 \times 10^{-4} f) + 3.5 \arctan(1.33 \times 10^{-4} f)^2 [k] \tag{4}$$

여기에서  $f$ 는 주파수 (Hz)이다. 임계대역을  $\xi$ 는 그림 3에 나타나 있으며, 임계대역폭 (CBW : Critical Band Width)는 약 100 Hz에서 500 Hz 사이의 폭을 갖는 필터들이다. 임계대역폭은 식 (5)와 같이 나타낼 수 있다.

$$CBW(f) = 25 + 75(1 + 1.4 \times 10^{-6} f^2)^{0.69} [Hz] \tag{5}$$

여기에서  $f$ 는 중심주파수를 의미하며, 그림 4에 도시하였다.

한 임계대역의 중심 값에서 바로 이웃한 임계대역의 중심 값까지 거리를 1 Bark 라고 한다. 사람의 귀는 20-20000 Hz 정도의 소리를 들을 수 있는데 이는 대략 25 Bark 대역에 해당한다. 이를 웨이블릿 계수와 매칭하여 표 1과 같이 부대역을 설계하였다 [4].

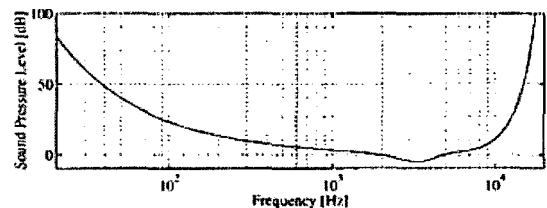


그림 2. 음압 레벨  
Fig. 2. Sound Pressure Level.

표 1. 웨이블릿 계수와 임계대역 매장을 통한 부대역 설계  
Table 1. Subband design with wavelet coefficient and critical band.

Subband $\xi$	$k(\xi)$	Coefficient $i_a - i_b$	Stage $j$
0	1	0-0	5
1	1	1-1	5
2	1	2-2	5
3	1	3-3	5
4	1	4-4	5
5	1	5-5	5
6	1	6-6	5
7	1	7-7	5
8	2	8-9	4
9	2	10-11	4
10	2	12-13	4
11	2	14-15	4
12	2	16-17	4
13	2	18-19	4
14	4	20-23	3
15	4	24-27	3
16	4	28-31	3
17	8	32-39	2
18	8	40-47	2
19	8	48-55	2
20	8	56-63	2

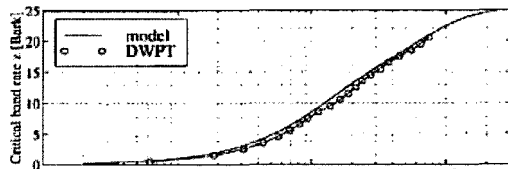


그림 3. 임계대역률  
Fig. 3. Critical Band Rate.

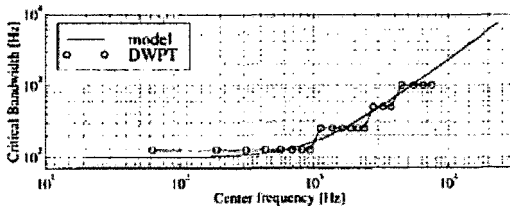


그림 4. 임계대역의 중심 주파수와 웨이블릿 계수와의 관계  
Fig. 4. The relation between critical band center frequency and wavelet transform coefficient.

### 3.2. Universal 문턱값을 이용한 연성 문턱값 적용

Donoho와 Johnstone가 백색잡음을 제거하기 위한 보편적인 웨이블릿 계수의 문턱값을 제안하였다. 이 문턱값은 부분적으로 잡음 섞인 선형신호의 연성 문턱값으로 잘 적용된다. 연성 문턱값은 식 (6)과 같이 정의되며,  $T$ 는 Universal 문턱값을 의미한다 [9].

$$\rho_T(X) = \begin{cases} \text{sign}(X)(|X| - T), & \text{if } |X| \geq T \\ 0, & \text{if } |X| < T \end{cases} \quad (6)$$

$$T = \rho \sqrt{2 \log(N)}$$

$$\rho = \rho_{MAD} / 0.6745$$

Universal 문턱값을 이용해 연성 문턱값을 적용할 경우 백색잡음에 대한 잡음제거 성능은 좋으나 유색잡음이

나 비정지적 잡음제거에 적용하였을 경우 잡음신호의 예측치가 커지게 되어 음성의 왜곡문제가 발생하게 되거나 잡음이 거의 제거되지 않는 문제를 가지게 된다. 이를 개선하기 위해 노이즈 마스킹 임계값과 세그멘탈 신호대 잡음비를 이용하여 문턱값을 개선하는 방안을 제안하고자 한다.

### 3.3. 문턱값 개선

음성 프레임에 대하여 크게 유성음과 무성음으로 나눌 수 있는데, 이는 유성음은 낮은 주파수의 부대역에 에너지 집중도가 높기 때문에 낮은 주파수 대역을 잘 표현할 필요가 있다. 반면에 웨이블릿 변환 계수는 무성음일 경우에 하위 대역의 계수가 큰 값을 가짐을 알 수 있다. 그러므로 웨이블릿 계수 문턱값은 식 (7)과 같이 주어져야 한다[15].

$$T_j^i(m) = K_{SNR}(m) \cdot K_{Perj}^i(m) \cdot \rho \sqrt{2 \log_e N} \quad (7)$$

식 (7)에서  $\rho \sqrt{2 \log_e N}$ 은 Universal 문턱값을 의미하며,  $K_{SNR}$ 은 세그멘탈 신호대잡음비를 반영하여, 과도하게 예측된 문턱값을 줄여주는  $c$ 와 트랜지션 구간의 기울기를 결정하는  $\alpha$ , 트랜지션의 중심으로부터 오프셋을 결정하는  $\gamma$ 로 식 (8)과 같이 표현할 수 있다.

$$K_{SNR}(m) = c [1 - 1 / (1 + e^{-\alpha \cdot \widehat{SNR}(m) + \gamma})] \quad (8)$$

여기서,  $\widehat{SNR}(m)$ 은 식 (9)와 같다.

$$\widehat{SNR}(m) = 10 \cdot \log_{10} \left( \frac{\sum_i |X_m^i|^2}{|X_m^-|^2} \right) \quad (9)$$

본 논문에서 제안하는 노이즈 마스킹 임계값과 세그멘탈 신호대잡음비는 근본적으로 Universal 문턱값이 잡음의 특성에 대한 적응력이 떨어진다는 문제점을 해결하기 위해서이다.

$$K_{Perj}^i = 1 + K_{em}(m) \cdot [U_j^i(m) - \bar{U}(m)] \quad (10)$$

여기서,

$$K_{em}(m) = 1 / (1 + e^{-\alpha \cdot \widehat{SNR}(m) + \gamma}) \quad (11)$$

$K_{Perj}^i(m)$ 은  $K_{em}(m)$ 와  $U_j^i(m)$ ,  $\bar{U}(m)$ 에 관한 식으로 구성된다.  $K_{em}(m)$ 은 음성의 변화율을 따라가기 위한 함수이며, 노이즈 마스킹 임계값이 세그멘탈 신호대잡음비

값이 클 경우에 정확해 지므로  $K_{em}(m)$  값은 세그멘탈 신호대잡음비를 반영하고 있다. 또한 세그멘탈 신호대잡음비가 높고 낮음에 따른 값의 부드러운 연결을 위해 곡선 형태를 가지도록 한다.

$$U_j^i(m) = 1 - \frac{Th_m(\xi)}{\max_{\xi}(Th_m(\xi))} \quad (12)$$

사람 귀의 인지 모델을 반영한 함수인  $U_j^i(m)$ 은 가중치 함수로서 노이즈 마스킹 임계값을 반영하며,  $\bar{U}(m)$ 은 가중치의 평균값을 의미한다.

$$\bar{U}(m) = \frac{1}{M} \sum_{j=L+1}^J \sum_{i=1}^I U_j^i(m), \quad (13)$$

$$M = \sum_{j=L+1}^J \sum_{i=1}^I 1$$

이와 같이 계산된  $T_j^i(m)$ 을 이용하여 연성 문턱값으로 적용한다.

임계대역에 대한 노이즈 마스킹 임계값을 계산하기 위해서 잡음 신호를 웨이블릿 변환 후에 얻어진 웨이블릿 계수를 통해 부대역 에너지  $E_m(\xi)$ 를 계산한다 [16].

$$E_m(\xi) = \sum_{k=l_k}^{h_k} |X_{j,k}^i(m)|^2 \quad (14)$$

여기에서  $m$ 은 프레임 번호를,  $l_k$ ,  $h_k$ 는 각각  $\xi$  부대역의 하위 상위 경계를 의미한다.

$\xi$ 에 대한 상위 하위 경계는 표 1에 잘 나타나 있다.

여기패턴 (Excitation Pattern)  $B_m(\xi)$ 은 내이의 달팽이관에서의 에너지 분포로 생각할 수 있으며, 부대역 에너지  $E_m(\xi)$ 와 산포함수  $F_m(\xi)$ 와의 컨볼루션으로 구할 수 있다.  $F_m(\xi)$ 은 Schroeder에 의해서 1979년에 제안되었다.

$$B_m(\xi) = F_m(\xi) * E_m(\xi) \quad (15)$$

산포함수는 분석에 용이한 형태로 식 (16)와 같이 표현하였다.

$$10\log_{10} F_m(\xi) = 15.81 + 7.53(\xi + 0.474) - 17.53(1 + (\xi + 0.474)^2)^{0.5} \text{ [dB]} \quad (16)$$

$O(\xi)$ 는 신호가 얼마나 음성에 가까운지 혹은 잡음에 가까운지를 알려주는 오프셋이다.

$O(\xi)$ 는 부대역 로그 에너지를 재조정하며 각각의 임계대역에서 항상 음수값을 가진다. 노이즈 마스킹 임계값

$\tilde{B}_m(\xi)$ 는 식 (17)와 같다.

$$\tilde{B}_m(\xi) = 10\log_{10} B_m(\xi) + O(\xi) \quad (17)$$

여기서,

$$O(\xi) = -(14.5 + i) \text{ [dB]} \quad (18)$$

여기서,  $i$ 는 Critical Band Index를 의미한다.

부대역의 에너지와 분포함수를 컨볼루션 했기 때문에 각각의 웨이블릿 부대역 에너지가 증가하게 되고 이를 재정규화하기 위해서 에너지 이득의 역수를 곱하여 최종적인 노이즈 마스킹 임계값을 구하게 된다.

$$Th_m(\xi) = \tilde{B}_m(\xi) - G_m(\xi) \quad (19)$$

$G_m(\xi)$ 은 산포 에너지  $B_m(\xi)$ 와 부대역 에너지  $E_m(\xi)$  간의 이득을 의미한다.

$$G_m(\xi) = 10\log_{10} \left( \frac{B_m(\xi)}{E_m(\xi)} \right) \quad (20)$$

### IV. 실험 및 결과

설계된 잡음 제거기를 평가하기 위하여 웨이블릿 모함수는 길이 4인 Daubechies 웨이블릿을 선택하였으며, 필터계수는 표 2와 같다.  $h$ 와  $g$ 는 각각 분해시의 필터뱅크 저대역 통과필터와 고대역 통과필터를 의미하며,  $\tilde{h}$ 와  $\tilde{g}$ 는 합성시의 필터뱅크의 저대역 통과필터와 고대역 통과필터를 각각 의미한다.

잡음은 각각 버블잡음, 바람소리, 천둥소리, 빗소리를 선택하였으며, 음성과 0 dB, 5 dB, 10 dB로 합성하여 TTA 표준인 EVRC의 잡음제거기와 Universal 문턱값을 이용한 연성 문턱값 적용방법, 제안하는 알고리즘을 서로 비교하여 도시화 하였다.

표 2. Daubiches 웨이블릿 필터계수  
Table 2. Daubiches wavelet filter coefficient.

Filter	Coefficient Order			
	1	2	3	4
$h$	-0.1294	0.2241	0.8365	0.4830
$g$	-0.4830	0.8365	-0.2241	-0.1294
$\tilde{h}$	0.4830	0.8365	0.2241	-0.1294
$\tilde{g}$	-0.1294	-0.2241	0.8365	-0.4830

그림 5와 그림 6에서 볼 수 있듯이 EVRC는 음성과 유사한 버블잡음을 음성으로 판단하여 거의 개선하지 못하고 있음을 알 수 있다. 하지만 제안하는 잡음 제거기는 EVRC 잡음제거기보다 잡음을 감소하고 있음을 알 수 있다. 전체적인 음성의 시간 영역의 모양에서도 크게 차이가 없음을 볼 수 있다. 실제적으로 웨이블릿을 이용해

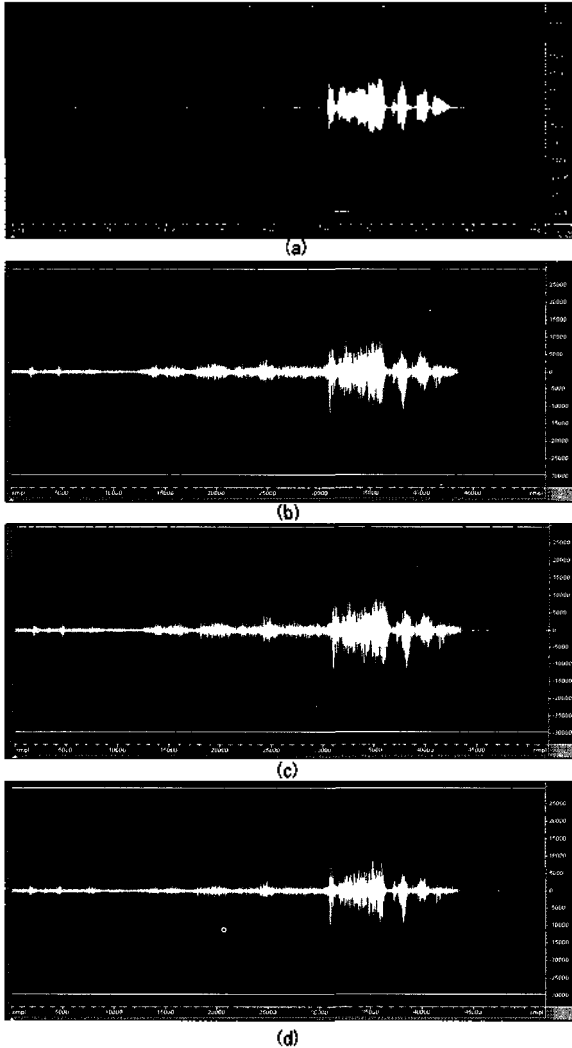


그림 5. 시간영역에서의 웨이블릿 패킷 변환을 이용한 잡음 제거기의 성능 :  
 (a) 바람 잡음과 음성 0dB로 섞은 신호,  
 (b) EVRC (Embedded Variable Rate Coder) 잡음 제거기,  
 (c) 웨이블릿 계수 문턱값 (Universal Threshold)을 이용한 잡음 제거기,  
 (d) 제안하는 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기.  
 Fig. 5. Performance evaluation result of noise suppressor using perceptual model and wavelet packet transform in time (Noisy speech corrupted by additive babble noise (seg\_SNR = 0dB)):  
 (a) Clean speech,  
 (b) EVRC(Embedded Variable Rate Coder) noise suppressor,  
 (c) Noise suppressor using Wavelet Coefficient Threshold (Universal Threshold),  
 (d) Proposed noise suppressor using perceptual model and wavelet packet transform.

좀 더 많은 잡음을 제거할 수 있지만, 이는 음성 왜곡이라는 문제점을 유발하게 된다. 웨이블릿을 이용한 잡음 제거기의 핵심은 음성 왜곡과 잡음개선의 적절한 합의점을 찾는 것이라고 할 수 있다.

결과적으로 TTA 표준인 EVRC 잡음 제거기와 유사한 값이 나왔으며, Universal 문턱값보다 0.29 정도 높았

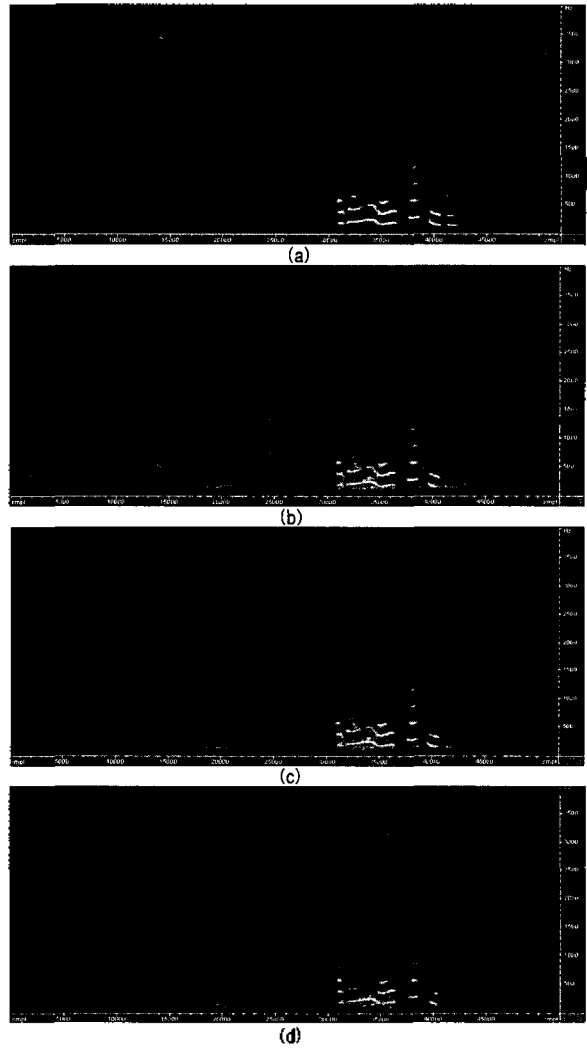


그림 6. 주파수영역에서의 웨이블릿 패킷 변환을 이용한 잡음 제거기의 성능 :  
 (a) 바람 잡음과 음성 0dB로 섞은 신호,  
 (b) EVRC (Embedded Variable Rate Coder) 잡음 제거기,  
 (c) 웨이블릿 계수 문턱값 (Universal Threshold)을 이용한 잡음 제거기,  
 (d) 제안하는 인지 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기.  
 Fig. 6. Performance evaluation result of noise suppressor using perceptual model and wavelet packet transform in time in frequency (Noisy speech corrupted by additive babble noise (seg\_SNR = 0dB)):  
 (a) Clean speech,  
 (b) EVRC(Embedded Variable Rate Coder) noise suppressor,  
 (c) Noise suppressor using Wavelet Coefficient Threshold (Universal Threshold),  
 (d) Proposed noise suppressor using perceptual model and wavelet packet transform.

표 3. 웨이블릿 변환을 이용한 잡음제거기로 전처리된 후 EHSX 코더로 인코딩된 디코딩된 성능평가(TU-T 표준안 p.862 PESQ MOS)  
 Table 3. Performance evaluation result of preprocessing with the noise suppressor using perceptual model and wavelet packet transform and encoding/decoding with EHSX code.

	EVRC 잡음제거기	제안하는 잡음제거기
버블 0 dB	1.712333	1.939771
버블 5 dB	2.090448	2.354719
버블 10 dB	2.367708	2.585542
바람 0 dB	2.157479	2.184688
바람 5 dB	2.421854	2.539729
바람 10 dB	2.569448	2.788646
천둥 0 dB	1.8325	1.991938
천둥 5 dB	2.174385	2.345552
천둥 10 dB	2.416313	2.591604
비 0 dB	1.424958	1.773135
비 5 dB	1.706906	2.208156
비 10 dB	2.082625	2.471583

다. MELP 코더로 인코딩한 후 디코딩 과정을 거쳐 측정된 PESQ-MOS 값은 EVRC 잡음 제거기보다 성능이 우수한 것으로 나타났다.

### V. 결론

제안된 잡음 제거기는 세그멘탈 신호대잡음비와 노이즈 마스킹 임계값을 반영하여 문턱값을 적응적으로 계산하여 연성 문턱값을 적용하는 방법으로 동작하도록 설계하였다. 실험의 결과에서도 볼 수 있듯이 Universal 문턱값을 이용하여 연성 문턱값 적용을 한 방법보다 잔여 잡음이 많이 제거되었으며 음성의 왜곡 면에서도 개선된 음질을 확인할 수 있었다. EVRC의 잡음 제거 알고리즘과의 성능 비교에서는 제안한 방법과 EVRC 잡음 제거 알고리즘이 유사한 성능을 가짐을 알 수 있었다.

실제적으로 음성의 왜곡과 잡음제거를 동시에 할 수는 없으며, 적당한 함의점을 찾는 것이 중요함을 위에서도 언급했다. 그러기 위해서 웨이블릿을 이용한 적절한 부대역의 설계가 이루어져야하고 적절한 문턱값을 가져야 하는 것이다. 우리가 알고 있는 Universal Threshold는 유색잡음에서 좋은 성능을 가지지 못하기 때문이다.

LPC나 하모닉 구조를 가진 보코더에서는 제안된 잡음 제거 방법이 MOS 0.23 정도 높은 값을 가짐을 확인할 수 있었으며, 이는 웨이블릿이 시간해상도를 고려하여 음성의 시간 변이를 잘 묘사하는 데서 나타난 결과로 해

석할 수 있다. 이로써 저속 코더에 기존의 방식보다 더 적합한 잡음제거기를 설계하였다고 결론지을 수 있겠다.

### 감사의 글

"이 논문은 2006년도 교육인적자원부 지방연구중심 대학 육성사업의 지원에 의하여 연구되었음."

### 참고 문헌

1. S. Mallat, *A Wavelet Tour of Signal Processing*(A Harcourt Science and Technology, Academic-Press, SanDiego, 1999)
2. N. R. Chong, I. S. Burnett, J. F. Chicharo, "A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition", *IEEE Trans. speech Audio Process.* 8 345-348, 2000.
3. D. Sinha, A. H. Tewfik, "Low bit rate transparent audio compression using adapted wavelets.", *IEEE Trans. Signal Process.* 41 3463-3479, 1993.
4. J. D. Johnston, "Transform coding of audio signal using perceptual noise criteria.", *IEEE J. Select. Area Commun.* 6, 314-323, 1988.
5. B. Carnero, A. Drygajlo, "Perceptual speech coding and enhancement using frame-synchronized fast wavelet packet transform algorithm.", *IEEE Trans. Signal Process.* 47 1622-1635, 1999.
6. J. Lukasiak, I. S. Burnett, "Exploiting simultaneously masked linear prediction in a WI speech coder.", In: *Int. Conf. on Acoustics, Speech and Signal Processing(ICASSP)*, 11-13, Istanbul, June, 2000.
7. P. Srinivasan, L. Jamieson, "High-quality audio compression using an adaptive wavelet packet decomposition and psychoacoustic modeling.", *IEEE Trans. Signal Process.* 46 1085-1093, 1998.
8. D. L. Donoho, I. M. Johnston, "Ideal spatial adaptation by wavelet shrinkage.", *Biometrika*, 81 425-455, 1994.
9. D. L. Donoho, "De-noising by soft- thresholding", *IEEE Trans. Inform. Theory* 41 613-627, 1995.
10. H. Sheikhzadeh, H. R. Abutalebi, "An improved waveletbased speech enhancement system.", In: *Proc. European Conf. on Speech Communication and Technology*, 1855-1858, Aalborg, September, 2001.
11. M. Bahoura, J. Rouat, "Wavelet speech enhancement based on the teager energy operator.", *IEEE Signal Process. Lett.* 8 10-12, 2001.
12. S. W. Chang, S. I. Jung, Y. H. Kwon, "Adaptive Wavelet Based Speech Enhancement with Robust VAD in Non-stationary Noise Environment", *The Journal of the Acoustical Society of Korea* 22 161-166, 2003.
13. 윤석현, 유창동, "시간-주파수 영역에서 음성/잡음 우세 결정에 의한 새로운 잡음처리", *한국음향학회지* 20 48-55, 2001.
14. S. Yoon, C. D. Yoo, "Speech/Noise- dominant detection for speech enhancement.", In: *Proc. European Conf. on Speech*

Communication and Technology, 1941-1944, 2001.

15. C. T. Lu, H. C. Wang, "Enhancement of single channel speech based on masking property and wavelet transform", Speech Commun. Mag. 41 409-427, 2003.

16. N. Virag, "Single channel speech enhancement based on masking properties of the human auditory system", IEEE Trans. Signal Process. Mag. 7 126-137, 1999.

**저자 약력**

• 김 미 선 (Mi Seon Kim)



2003년 8월: 충북대학교 전기전자공학부 (공학사)  
 2003년 8월~현재: 충북대학교 전파공학과 (석사과정)  
 ※주관심분야: 음성/오디오 부호화, 디지털 신호처리

• 박 서 영 (Seo Young Park)



2005년 2월: 충북대학교 전기전자공학부 (공학사)  
 2005년 3월~현재: 충북대학교 전파공학과 (석사과정)  
 ※주관심분야: 음성/오디오 부호화, 채널 코딩, 디지털 신호처리, VoIP

• 김 영 주 (Young Ju Kim)



1989년 2월: 고려대학교 전자전공학과 (공학사)  
 1996년 2월: 한국과학기술원 정보통신공학과 (공학석사)  
 2001년 2월: 한국과학기술원 전기및전자공학과 졸업 (공학박사)  
 1987.11~1993.12: (주)금성사 평택연구소  
 1996.2~1997.2: 동경공업대학교 연구원  
 2000.4~2000.12: ETRI 위축연구원  
 2001.1~2003.8: (주)엘지전자 UMTS 시스템 연구소  
 2003.9~현재: 충북대학교 정보통신공학과 교수

※주관심분야: 이동통신 시스템, MIMO, 유비쿼터스 네트워크

• 이 인 성 (In Sung Lee)



1983년 2월: 연세대학교 전자공학과 (공학사)  
 1985년 2월: 연세대학교 전자공학과 (공학석사)  
 1992년 2월: Texas A&M University 전기공학과 (공학박사)  
 1993년 2월~1995년 9월: 한국전자 통신연구원 이동통신 기술연구단 선임연구원  
 1995년 10월~현재: 충북대학교 전기전자공학부 정교수  
 ※주관심분야: 음성/영상 신호 압축, 이동통신, 적용필터