

심리음향 모델과 웨이블릿 패킷 변환을 이용한 잡음제거기

김미선, 김영주, 이인성
 충북대학교 전자공학과

miseon80@nate.com, yjkim@cbnu.ac.kr, inslee@cbucc.chungbuk.ac.kr

Noise suppressor Using Psychoacoustic Model and Wavelet Packet Transform

Mi-seon Kim, Young-ju Kim, In-sung Lee

Dept, Information communication engineering, Chungbuk National University

Abstract

In this paper, we propose the noise suppressor with the psychoacoustic model and wavelet packet transform. The objective of the scheme is to enhance speech corrupted by 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.

In this paper, the subband is designed matching with the critical band. And WCT is adapted by noise masking threshold(NMT) and segmental signal to noise ratio(seg_SNR). Consequently this work improve the PESQ-MOS about 0.23 in the case of coded speech.

I. 서론

본 논문에서는 시간에 따른 대역별 주파수 특성변이가 다른 잡음의 제거 성능을 향상시키고자 웨이블릿 패킷 변환(Wavelet Packet Transform)을 이용하고 있다. 웨이블릿은 신호처리를 위해 꾸준히 연구되어 왔으며, 이러한 분석 방법은 신호를 다해상도 필터뱅크 함으로써 쉽게 이루어질 수 있다[1]. 웨이블릿 변환을 이용한 분석법은 영상신호처리나 음성신호 처리에도 성공적으로 적용되어왔다[2]-[4].

본 논문에서는 5단계의 웨이블릿 패킷 분해를 통해 얻은 웨이블릿 계수들과 임계대역을 매칭하여 부대역을 구조화 하였으며, 잡음 마스킹에 사용되는 문턱값은 세그멘탈 신호대 잡음비와 노이즈마스킹 임계값을 이용하여 적응되도록 설계하였다. 음성과 유사한지 잡음과 유사한지를 계산하여 반영하는 특성에 따라 계수의 전체적인 모양을 달리 하는 구조를 가지므로 별도로 유무성음 구별 과정 없이 웨이블릿 계수 문턱값을 자동 조절한다.

II. 본론

2.1 제안하는 심리음향 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.

입력된 신호 $x(n)$ 과 추정된 잡음을 이용하여 신호 대잡음비를 구하게 되며, $x(n)$ 을 5단계의 웨이블릿 패킷변환 후 추정된 노이즈 마스킹 임계값(Noise Masking Threshold)과 함께 최종적인 문턱값 갱신에 사용된다.

$x(n)$ 에 개선된 문턱값을 연성 문턱값 형태로 적용해 다시 역변환을 거친 후 이전 프레임과의 연계성을 형

성해주면 음질 개선 과정이 끝나게 된다.

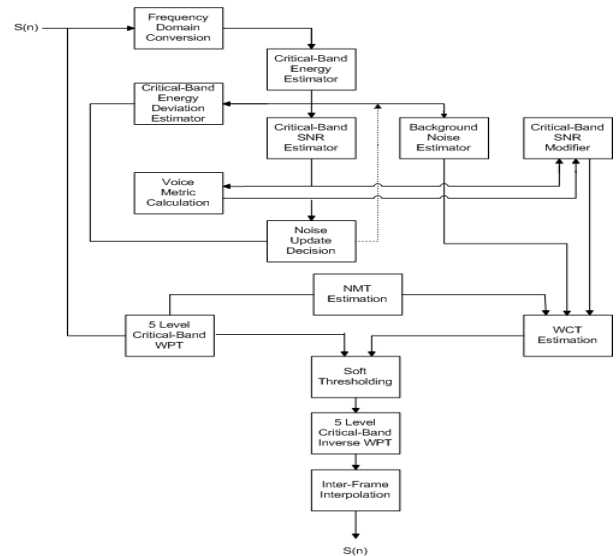


그림 1. 심리음향 모델과 웨이블릿 패킷 변환을 이용한 잡음 제거기 구조

2.2 심리음향 모델을 반영한 부대역 설계

소리의 주파수 요소는 임계대역과 밀접하게 연관되어 있다. 이 중심주파수는 임계대역률(Critical Band Rate) ξ 라고 알고 있는 부대역에 위치하고 식 (1)과 같이 표현할 수 있으며 단위는 [Bark]이다.

$$\xi = 13\arctan(7.6 \times 10^{-4}f) + 3.5\arctan(1.33 \times 10^{-4}f)^2 \quad (1)$$

이를 웨이블릿 계수와 매칭하여 표 1과 같이 부대역을 설계하였다[4].

2.3 문턱값 개선

음성 프레임은 유성음과 무성음으로 나눌 수 있는데, 유성음은 낮은 주파수에 에너지 집중도가 높기 때문에 저주파수 대역을 잘 표현할 필요가 있다. 이와 반대로 무성음의 경우 고주파 대역의 에너지를 잘 표현해야한다. 그러므로 웨이블릿 계수 문턱값은 식 (2)과 같이 주어져야 한다[5].

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

식 (2)에서 $\rho \sqrt{2\log_e N}$ 은 Universal 문턱값을 의미하며, K_{SNR} 은 세그멘탈 신호대 잡음비를 반영하여, 과도하게 예측된 문턱값을 줄여주는 c 와 트랜지션 구간의

기울기를 결정하는 α , 트랜지션의 중심으로부터 오프셋을 결정하는 γ 로 식 (3)과 같이 표현할 수 있다.

표 1. 웨이블릿 계수와 임계대역 매칭해 부대역 설계

Subband ξ	$l(\xi)$	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

$$K_{SVR}^i(m) = c[1 - 1/(1 + e^{-\alpha \cdot \widehat{SNR}(m) + \gamma})] \quad (3)$$

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

$$K_{Prj}^i = 1 + K_{em}(m) \cdot [U_j^i(m) - \bar{U}(m)] \quad (4)$$

여기서,

$$K_{em}(m) = 1/(1 + e^{-\alpha \cdot \widehat{SNR}(m) + \gamma}) \quad (5)$$

$K_{Prj}^i(m)$ 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 신호대 잡음비 값이 클 경우에 정확해 지므로 $K_{em}(m)$ 값은 세그멘탈 신호대 잡음비를 반영하고 있다. 또한 세그멘탈 신호대 잡음비가 높고 낮음에 따른 값의 부드러운 연결을 위해 곡선 형태를 가지도록 한다.

$$U_j^i(m) = 1 - \frac{Th_m(\xi)}{\max_{\xi}(Th_m(\xi))} \quad (6)$$

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

$$\bar{U}(m) = \frac{1}{M} \sum_{j=L+1}^J \sum_{i=1}^{I_j} U_j^i(m), \quad (7)$$

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

이와 같이 계산된 $T_j^i(m)$ 을 연성 문턱값으로 적용한다.

III. 실험결과

본 알고리즘을 평가하기 위하여 웨이블릿 모함수는 길이 4인 Daubechies 웨이블릿을 선택하였다. 잡음은 각각 버블잡음, 바람소리, 천둥소리, 빗소리를 선택하였으며, 음성과 0 dB, 5 dB, 10 dB로 합성하여 TTA 표준인 EVRC의 잡음제거기와 제안하는 알고리즘을 서

로 비교하여 도시화 하였다.

표 2. 웨이블릿 변환을 이용한 잡음제거기로 전처리를 한 후 MELP 코더로 인코딩 디코딩한 성능평가 (ITU-T 표준안 p.862 PESQ MOS)

	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

결과적으로 TTA 표준인 EVRC 잡음 제거기와 유사한 값이 나왔으며, Universal 문턱값보다 0.289 정도 높았다. MELP 코더로 인코딩한 후 디코딩 과정을 거쳐 측정된 PESQ-MOS 값은 EVRC 잡음 제거기보다 성능이 월등히 우수한 것으로 나타났다.

IV. 결론

제안한 논문은 세그멘탈 신호대 잡음비와 노이즈 마스킹 임계값을 반영하여 문턱값을 적응적으로 계산하여, 연성 문턱값을 적용하는 방법으로 동작하는 잡음 제거기이다. 실험의 결과 EVRC의 잡음 제거 알고리즘과의 성능 비교에서는 제안한 방법과 EVRC 잡음 제거 알고리즘이 유사한 성능을 가짐을 알 수 있었다.

LPC나 하모닉 구조를 가진 코더로 인코딩과 디코딩 작업을 거친 음성에서는 제안하는 방법이 MOS 0.23 정도 높은 값을 가짐을 확인할 수 있었으며, 이는 웨이블릿이 시간해상도를 고려하여 음성의 시간 변이를 잘 묘사하는 데서 나타난 결과로 해석할 수 있다. 이로써 저속 코더에 기존의 방식보다 더 적합한 잡음제거기를 설계하였다고 결론지을 수 있겠다.

이 논문은 2006년도 교육인적자원부 지방연구중심대학 육성사업의 지원에 의하여 연구되었음.

참고문헌

- [1] Mallat, S., "A Wavelet Tour of Signal Processing", A Harcourt Science and Technology. Academic-Press, San Diego., 1999.
- [2] Chong, N.R., Burnett, I.S., Chicharo, J.F., "A new waveform interpolation coding scheme based on pitch synchronous wavelet transform decomposition", IEEE Trans. speech Audio Process., Vol. 8, No. 3, pp.345-348
- [3] Sinha, D., Tewfik, A.H., "Low bit rate transparent audio compression using adapted wavelets.", IEEE Trans. Signal Process., Vol. 41, pp.3463 - 3479, 1993.
- [4] Johnston, J.D., "Transform coding of audio signal using perceptual noise criteria.", IEEE J. Select. Area Commun., Vol. 6, pp.314 - 323, 1988.
- [5] Lu, C. T., and Wang, H. C., "Enhancement of single channel speech based on masking property and wavelet transform", Speech Commun. Mag., Vol. 41, No. 2-3, pp.409-427, 2003.