1999년도 한국음향학회 학술발표대회 논문집 제18권 제2(s)호

웨이블릿 패킷을 이용한 잡음에 손상된 음성신호 인식에 관한 연구

°고광현*, 장성욱*, 양성일*, 권영현**
*한양대학교 제어계측공학과, **한양대학교 물리학과

Recognition of Corrupted Speech by Noise using Wavelet Packets

°Kwang-hyun Koh*, Sungwook Chang*,
Sung-il Yang*, Y. Kwon**
*Dept. of Control & Inst. Eng., Hanyang Univ.,
**Dept. of Physics, Hanyang Univ.
E-mail : khkoh@hanmail.net

요약

인식기 훈련과정에서 발생하지 않았던 잡음이 인식과정에서 신호를 손상할 경우 인식률의 저하가 발생한다. 본 논문에서는 음성의 질을 떨어뜨리는 이러한 잡음을 Wavelet Packets을 이용하여 전처리함으로서 인식률을 향상시키는 방법을 제안한다.

인식기로는 Hidden Markov Model을 사용하였고, 시스템에 사용된 특징 파라미터로는 15차 Cepstrum을 사용하였다. 11 kHz로 샘플링된 숫자음에 Additive White Gaussian Noise를 첨가한 손상된 음성신호를 인식실험에 사용하였다. 화자독립으로 진행된 실험에서 잡음에 의해 손상된 SNR 20dB의 음성신호에 대하여 Wavelet Packets로 잡음을 제거한 후 복원된 음성신호의 인식률은 약 10% 향상됨을 확인하였다.

1. 서론

음성인식에 있어 잡음은 정확한 실음성 구간 검출을 어렵게 할 뿐만 아니라 음성을 변화시켜 인식률을 떨어뜨리는 요인이 된다. 잡음 환경에서의 인식률 향상을 위해 사용되는 여러 가지 방법들이 있는데 본 논문에서는 웨이블릿 패킷을 이용한 잡음제거 방법을 전처리 과정에 사용해 음질을 개선하였다. 웨이블릿은 원시자료의 분석과 처리에 있어 효율적인 도구이며, 근래에 신호처리분야에 있어 다양하게 적용되고 있는 변환 방법이다. 엔트로피를 이용하여 최적의 basis에 적응되는 웨이블릿 패킷은 신호의 압축과 잡음제거 부문에 적용되고 있다. 잡음이 없는 환경에서의 음성으로 훈련을 한 음성인식기는 실제의 환경에서 사용될 때 예상치 못했던 잡음으로 인해 인식률의 저하가 일어난다. 이러한 잡음의 영향을 효과적으로 줄여줄 수 있는 음성인식 시스템이 필요한데, 본 논문에서는 전처리기로 웨이블릿 패킷을 사용하는 인식시스템을 제안하였다.

본 논문의 구성은 다음과 같다. 2절에서는 웨이블릿 패킷과 웨이블릿 패킷을 이용한 잡음제거 방법에 대해, 3절에서는 인식시스템에의 잡음제거 적용방법 및 실험과 결과에 대해 서술하였고, 4절에서 결론을 맺었다.

2. 웨이블릿 패킷을 이용한 잡음제거

2.1 웨이블릿 패킷 변환

Ronald Coifman에 의해 제안된 웨이블릿 패킷 시스템은 고주파 대역에서 주파수를 더 자세히 그리고 조정 가능하게 볼 수 있게 한다. 이는 특정 신호에 대해 적응할 수 있는 풍부한 구조를 제공한다. 웨이블릿 패킷 변환은 <그림 1>에서 보는것과 같이 Mallat 알고리즘 트리를 이용하여 신호를 웨이블릿 계수 h_1와 scaling 계수 h_0를 사용하여 각각 고주파통과의 웨이블릿 가지와 저주파통과의 scaling 함수 가지로 나뉘어지는 것을 반복함으로써 이루어 질 수 있다[1]. h_0와 h_1의 필터를 거친 신호를 decimation함으로써 트리구조의 아래로 내려갈수록 패킷의 크기는 1/2로 준다.

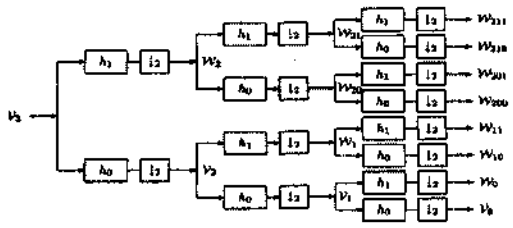

그림 1. 웨이블릿 패킷의 이진 트리 구조

2.2 Best Basis의 결정[2]

<그림 1>과 같은 웨이블릿 패킷 트리로부터 분해한 벡터들의 Shannon-Weaver(SW) 엔트로피를 최소화하는 best othonormal basis를 찾을 수 있다. 엔트로피는 잡음의 성질인 무질서를 표현하기 때문에 이를 최소화하면 분해된 신호의 SNR을 향상시킬 수 있다.

신호 x={x_i}의 SW 엔트로피는

$$H(x) = -\sum p_i \log p_i \qquad (1)$$

단, $p_i = (|x_i|^2)/(||x||^2)$

만약 $p=0$ 이면 $p\log p = 0$

으로 나타낼 수 있다. 식 (1)로부터 additive information cost 함수를 새로 정의할 수 있다.

$$H(x) = ||x||^2 \lambda(x) + \log ||x||^2 \qquad (2)$$

$$\lambda(x) = -\sum_i |x_i|^2 \log |x_i|^2 \qquad (3)$$

식 (2)에서 $\lambda(x)$를 최소로 하는 것은 H(x)도 최소가 된다. 식 (3)을 사용해서 <그림 2>와 같이 아래의 레벨부터 분해된 두 개의 패킷의 λ를 더한 것을 위 레벨의 λ와 비교하여 값이 더 작은 것을 선택하며, 이 과정을 가장 위의 레벨까지 반복하여 best basis를 선택한다.

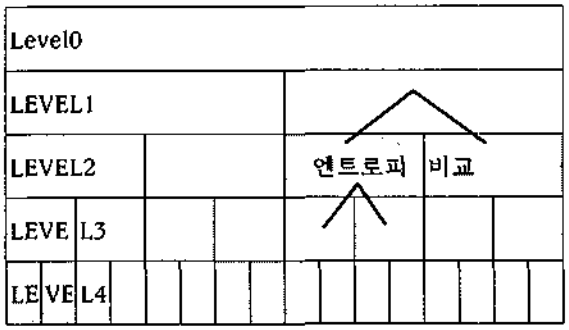

그림 2. 웨이블릿 패킷에서 Best Basis

2.3 임계치의 결정[3]

신호 s_i가 σ의 표준편차를 가지는 i.i.d, zero mean, White Gaussian noise n_i에 의해 손상되었을 때 이를 x_i라 하자[4]. 즉,

$$x_i = s_i + \sigma n_i, \qquad i=1,2,3,\cdots,N \qquad (4)$$

여기서 x_i로부터 s_i를 복구하는 것이 잡음제거의 목적이다. 만약 W와 W^{-1}을 각각 N×N의 웨이블릿 변환과 그 역변환 행렬이라고 하면, x, s, n은 각각

$$X=Wx,\ S=Ws,\ N=Wn \qquad (5)$$

(x, s, n은 x_i, s_i, n_i의 열벡터들)

이 되며, 식 (4)는 아래와 같이 된다.

$$X = S + N \qquad (6)$$

W에 의해 옮겨진 N은 여전히 평균이 0이고 분산이 σ^2 인 White Gaussian의 형태를 가지며, 반면에 S는 단지 몇 안되는 큰값의 웨이블릿 계수들로 나타난다. 그러므로 신호 s를 추정하기 위해서 X에 적당한 임계치를 주어 그 값보다 작은 계수들을 없애면 된다. 대각 필터링 연산, Δ은 이러한 thresholding을 표현할 수 있다.

$$\Delta = \mathrm{diag}[l_1, l_2, \cdots l_N] \qquad (7)$$

이 연산을 사용하여 원래신호의 추정치 $\hat{s}$를 표현하면,

$$\hat{s} = W^{-1} \Delta W s \qquad (8)$$

가 된다.

웨이블릿 threshold 필터는 두가지의 기본적인 형태로 나눌 수 있다. Hard threshold 필터 l_h는 잡음의 분산 σ에 의해 정해진 threshold τ의 이상이 되는 α_i (웨이블릿 변환된 계수)값들 만을 보존한다.

$$i_h(i) = \begin{cases} 1 & \text{if} \quad |\alpha_i| > \tau \\ 0 & otherwise \end{cases} \qquad (9)$$

Hard thresholding은 존재하지 않는 진동을 만들 수 있지만, l_2 error를 줄이는데 있어서는 Soft의 경우보다 더 나은 결과를 가져온다[1]. Soft threshold 필터 l_s는 임계치보다 작은 값들을 0으로 대체하면서, 웨이블릿 계수값들을 아래 식 (10)에 의해 줄인다.

$$l_h(i)=\begin{cases}\dfrac{sgn[\alpha_i][|\alpha_i|-\tau]}{\alpha_i} & \text{if} \quad |\alpha_i|>\tau \\ 0 & otherwise\end{cases} \qquad (10)$$

Soft thresholding은 잡음을 제거한 신호의 Smoothness를 원래의 신호의 Smoothness와 같게 한다.

임계치 ρ는 각 신호에 대해서 4가지 임계치 추정 방법이 있는데, 이 중 Stein's unbiased risk estimate (SURE)에서의 임계치(universal threshold)은 아래의 방법으로 결정된다[5].

$$\rho=\sqrt{(2\log_e(N\times\log_2(N)))} \qquad (11)$$

N : 신호벡터안에서의 샘플의 개수

식(10)에서 결정된 ρ을 입력신호의 잡음 성분을 고려해 실제 사용하는 임계치를 결정한다[6].

$$\tau=\rho\times median(|d^1|)/0.6745 \qquad (12)$$

d^1 : 입력벡터의 웨이블릿 패킷 첫 번째 레벨에서의 details값

3. 실험 및 결과

3.1 SNR에 따른 에 따른 잡음 제거 방식의 결정방법

잡음에 의해 손상받기 전의 음성의 인식률	94.6 %

원음성과 잡음과의 SNR	잡음 제거를 하지 않은 경우의 인식률	임계치 τ로 잡음제거를 한 경우의 인식률	임계치 ρ로 잡음제거를 한 경우의 인식률
40 dB	94.4 %	92.6 %	83.8 %
30 dB	92.8 %	85.4 %	84.8 %
25 dB	84.0 %	86.8 %	85.4 %
20 dB	73.4 %	77.4 %	83.8 %
15 dB	45.2 %	60.4 %	63.2 %
10 dB	9.0 %	45.8 %	12.2 %
5 dB	8.4 %	37.0 %	11.0 %

표 1. SNR별 잡음 제거 전과 제거 후의 인식률 비교

앞 절에서 제시된 잡음 제거 방식은 입력 신호에서 잡음을 제거하는 데에서 적절한 효과를 나타내지만, 실제 음성인식에 있어서는 동일한 결과를 보장하지 않는다. 잡음에 의해 손상되기 전의 신호와 잡음과의 신호 대 잡음비(SNR)가 30dB 이상이 되는 입력 음성에서는 잡음이 제거 되지 않은 경우는 잡음이 없는 음성과 인식결과가 거의 차이가 나지 않았지만, 잡음을 제거한 경우는 오히려 인식률이 떨어지는 결과가 나타났다. SNR이 5dB 이상부터 30dB 미만의 입력 음성에 대해서는 잡음을 제거한 경우가 제거하지 않은 경우보다 더 나은 결과를 보였다. 그러나 신호에 더해진 잡음에 맞게 설정된 임계치 τ로 잡음을 제거하였을 경우 15 dB부터 25 dB 미만까지는 universal threshold ρ로 잡음을 제거한 경우 보다 더 낮은 인식률을 보였다. 이는 잡음제거에 의해 원래의 음성과 가장 오차가 작은 음성을 만들어 내는 것이 가장 좋은 인식률을 보장하지는 않음을 보여주고 있다. 다시 말해 더 많은 잡음을 제거할수록 더 많은 원래음성의 성분이 같이 제거가 되므로 오히려 잡음을 적게 제거하고 원래음성의 성분을 조금 더 보호하는 것이 더 인식률을 높일 수도 있다는 것을 의미한다.

위의 결과로부터 입력음성의 신호 대 잡음비를 추정하여 각 경우에 대해 다른 잡음제거 방식을 취하는 전처리기를 구성하였다.

입력음성의 신호대 잡음비	처리 방법
30 dB 이상	잡음 제거를 하지 않음
15 dB ~ 30dB	임계치 ρ를 사용하여 잡음제거
15dB 미만	임계치 τ를 사용하여 잡음제거

표 2. 전처리기의 SNR별 잡음제거 방법

3.2 실험 데이터 구성

실험에 사용된 음성데이터는 실험실 환경에서 녹음된 남성화자 18명이 10회씩 발음한 숫자음을 11 kHz, 8 bits/sample로 샘플링한 것이다. 이중 8명의 데이터가 학습에 사용되었고, 나머지 10명의 음성이 인식실험에 사용되었다.

부가잡음으로는 위의 음성데이타에 대해 5, 10, 15, 20, 25, 30, 40 dB의 Additive White Gaussian Noise(AWDN)을 생성시켰으며, 이를 음성데이타에 더해 손상된 실험음성데이타를 만들었다.

3.3 인식시스템의 구성

잡음 제거에 사용한 웨이블릿 basis는 Daubechies 12차를 사용했으며, Hard Thresholding을 사용하였고, 가장 작은 웨이블릿 패킷의 크기는 32 샘플로 하였다.

음성의 특징 추출에 사용된 특징 벡터로는 15차의 Cepstrum이 256 레벨로 VQ가 되어 사용되었으며, 인식에 left-to-right HMM이 사용되었다. 잡음에 손상된 음성의 인식실험에 잡음에 손상되지 않은 음성으로 훈련된 모델 특징벡터들이 사용되었다.

3.4 결과

입력신호의 신호 대 잡음비에 따른 다른 잡음제거 방식을 취한 음성 인식 시스템의 인식결과는 아래와 같다.

원음성과 잡음과의 SNR	잡음 제거를 하지 않은 경우의 인식률	SNR에 따른 잡음제거를 한 경우의 인식률
40 dB	94.4 %	94.4 %
30 dB	92.8 %	92.8 %
25 dB	84.0 %	85.4 %
20 dB	73.4 %	83.8 %
15 dB	45.2 %	63.2 %
10 dB	9.0 %	45.8 %
5 dB	8.4 %	37.0 %

표 3. 잡음 제거 전과 제거 후의 인식률 비교

4. 결론

본 논문에서는 AWGN에 손상된 숫자음의 화자 독립 인식 실험에 웨이블릿 패킷을 이용한 전처리기를 제안하였고, 실험을 통해 성능을 살펴보았다.

SNR이 낮은 음성에서는 음성의 끝점 추출이 제대로 이루어지지 못하였는데, 이로 인해 인식률이 크게 떨어졌다. 하지만 잡음을 제거하고 인식함으로써 끝점추출이 가능해져 인식률의 차이가 커졌다. SNR이 약 20dB 정도의 음성에서는 잡음제거를 통해 인식률이 약 10% 향상되었다. 이러한 실험결과를 통해 제안한 전처리기가 AWGN의 제거, 인식에 효과가 있음을 확인하였고, 앞으로 음성의 특성을 더욱 잘 표현하는 basis에 대한 연구와 훈련과정에서 잡음을 고려한 코드북의 제작으로 더욱 인식률을 높이는 연구가 진행되어야 할 것이다.

5. 참고 문헌

[1] C. S. Burrus, R. A. Gopinath, H. Guo, *Introduction to Wavelets and Wavelet Transforms*, Prentice Hall, 1998.

[2] M. V. Wickerhauser, "*Adapted Wavelet Analysis from Theory to Software*", IEEE Press, 1994.

[3] D. Donoho, "*Denoising by Soft-Thresholding*", IEEE Trans Inform Theory 46:612-627, 1995.

[4] S. K. Nath et al. "*Wavelet based compression and denoising of optical tomography data*" Optics Communications , Vol.167 N.1-6 , pp.37-46 , 1999.

[5]P. E. Tikkanen, "*Nonlinear wavelet and wavelet packet denoising of electrocardiogram signal*", Biological Cybernetics , Vol.80 N.4 , pp.259-267 , 1999.

[6] B. Walczak, D. L. Massart, "*Noise Suppression and Signal Comppresion using the Wavelet Transform*" Chemometrics & Intelligent Laboratory Systems , Vol.36 N.2, 1997 .