

## 인간의 청각 메커니즘을 적용한 웨이블릿 분석을 통한 음성 향상에 대한 연구

이준석, 길세기, \*홍준표, \*\*홍승홍  
인하대학교 전자공학과  
전화 : 032-868-4691 / 핸드폰 : 019-280-1862

### A study of speech enhancement through wavelet analysis using auditory mechanism

JoonSuk Lee, SeKee Kil, JunPyo Hong, SungHong Hong  
School of Electrical and Computer Engineering, INHA UNIVERSITY  
E-mail : [ldslee@hotmail.com](mailto:ldslee@hotmail.com)

#### Abstract

This paper has been studied speech enhancement method in noisy environment. By mean of that we prefer human auditory mechanism which is perfect system and applied wavelet transform. Multi-resolution of wavelet transform make possible multiband spectrum analysis like human ears. This method was verified very effective way in noisy speech enhancement.

#### I. 서론

음성처리 및 인식 기술은 차세대 인터페이스(interface)로 가장 주목을 받고 있는 방법 중의 하나이다. 최근 음성인식 기능이 탑재된 친인간적인 시스템들이 쏟아져 나오고 있으며, 음성인식기의 상용화가 확산되고 있다. 그러나 아직도 실제 환경에서의 인식률은 많이 떨어지는 것으로 알려져 있다. 따라서 잡음 환경에서 강한 성능을 발휘하는 음성향상 방법이 연구되었다. 인간의 청각 모델은 가장 훌륭한 음성 인식 시스템이라 여겨지므로 본 논문에서는 인간의 청각의 메커니즘을 연구하고 적용시키는 방법을 찾아보았다. 인간의 귀는 음성에 대해 서로 다른 특성을 갖는 수많은 필터 बैं크에 의해 주파수 분석을 하는 것으로 알려져 있다. 이때 인간의 청각 기관이 갖는

한계에 의해서 특정 신호가 다른 신호에 가려지는 현상의 발생하며 전처리 과정을 통해 인지가 이루어진다.

인간의 청각시스템은 주파수 대역으로 해상도가 다른데 고주파 성분은 더 민감하게 반응하고 저주파 성분에는 덜 민감하다.[1] 본 논문에서는 이러한 청각 특성을 적용하여 음성신호를 주파수 대역별로 분해하여 분석하려고 시도했고, 그 방법으로 웨이블릿 변환을 사용하였다. 또한 Wavelet 변환을 통해 나온 Approximation 값은 원래의 신호의 De-noising을 한 효과가 있어 자체적으로 별도의 잡음 처리 과정이 필요하지 않다. [1] 따라서 이 방법은 잡음에 강한 음성인식에 좋은 아이디어가 되며 실제로 잡음 하에서 음성을 향상 시키는데 효과적인 방법이 될 것이다.

#### II. 인간의 청각 메커니즘

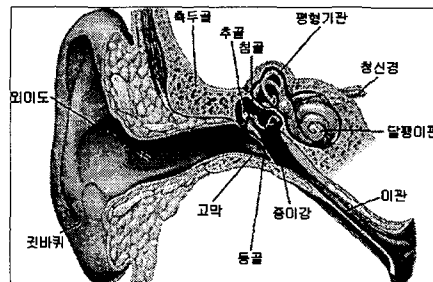


그림 1 귀의 구조

사람의 귀는 외이, 중이, 내이의 세 부분으로 나눌 수 있다. 소리의 진동은 귓바퀴에 모여져서 외이도를 따라 고막에 전달된다. 귓바퀴는 소리의 방향감을 살려주고 고주파음에 대해서 1차적인 증폭을 한다. 외이도는 한쪽이 고막에 의해서 막힌 일종의 공명관이라 볼 수 있으며, 말소리 이해에 중요한 정보를 가지는 2-3 KHz의 소리를 추가로 증폭하는 역할을 한다. 등골(Stapes)은 달팽이관(Cochlea)의 난원창(Oval window)에 접해 있고, 이소골의 진동이 등골의 족판(Footplate)을 통해 달팽이관에 전달되면 내부의 림프액(Lymph)이 움직이게 된다.

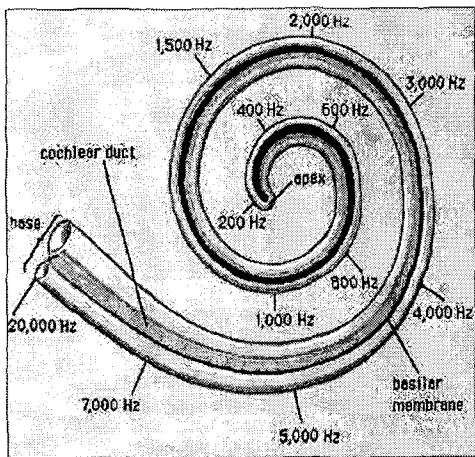


그림 2 달팽이관에서의 음성신호의 주파수 분석

그림 2에서처럼 달팽이관은 긴 관이 말려있는 형태로 펼쳤을 때 하부 (base)쪽에 고음(High frequency sound), 저음(Low frequency sound)은 상부(Apex)쪽에 관계된다. 소리 에너지는 하부에서 상부 쪽으로 이동하며 에너지의 이동에 따른 림프액의 움직임에 의해 가운데 층이 진동한다. 달팽이관의 가운데 층에 있는 수천개의 유모세포들이 림프액의 진동을 감지하여 그 자극을 전기적 신호로 바꾸어 청신경을 통해 뇌로 전달하면 우리가 실제로 소리들 듣게 된다. 청신경은 자극신호에 동기 되어 반응을 하게 되며, 이 발화 패턴은 신호의 주파수 특성을 반영한다. 신호의 주파수와 강도 정보는 청신경에 의해 신경 신호로 표현된다. 저주파 채널에서는 긴 신호가, 고주파 채널에서는 상대적으로 짧은 신호가 사용되므로 저주파 부분에서는 주파수 분해도가 좋아지는 반면 시간 분

해도가 나빠지며 고주파 부분에서는 이와 반대의 특성이 나타난다. 이는 실제 사람에게서 나타나는 특성이다.

### III. 웨이블릿 분석

웨이블릿은 기저 웨이블릿(mother wavelet)의 천이(shift)와 스케일링(scaling)으로 만들어 진다. 신호  $f(x)$ 의 웨이블릿 변환은 다음 식과 같이 정의 된다.[2,3]

$$Wf(x) = f(x) * \varphi_s(x) = \frac{1}{\sqrt{|s|}}$$

$$\int_{-\infty}^{\infty} f(t)\varphi\left(\frac{x-t}{s}\right)dx \quad \text{식 1}$$

여기서  $s$ 는 스케일이다.  $\varphi_s = \frac{1}{s}\varphi\left(\frac{x}{s}\right)$ 는 스케일  $s$ 에 의해 기저 웨이블릿  $\varphi(x)$ 를 확장(dilation)한 것이다. 임의의 신호  $f(x)$ 의 웨이블릿 변환은 식 1과 같이 임의의 신호와 확장된 기저 웨이블릿의 내적의 합으로 표시된다. 이에 따라 웨이블릿 변환에서는 주파수 대역이라는 용어 대신 스케일(scale)이라는 용어를 주로 사용하며, 입력 신호에 대한 웨이블릿 변환을 다른 말로 원신호의 시간, 스케일 공간표현이라 일컫는다.

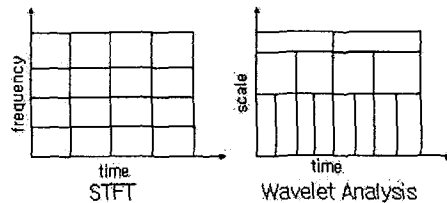


그림 3 신호해석 방법들

기저 웨이블릿의 종류에 따라 Harr, Daubechies, Morlet, Meyer 등의 웨이블릿이 있다. 웨이블릿 해석은 연속신호와 이산신호의 경우에 모두 적용될 수 있다. STFT (short time fourier transform)이 모든 주파수 대역에서 동일한 필터 윈도우 크기를 사용하는 반면에 웨이블릿 변환은 Multirate-filter Bank 하게 되며 Tree 구조의 Filter Bank를 사용하여 Non-uniform bandwidth를 제공한다.[4,5]

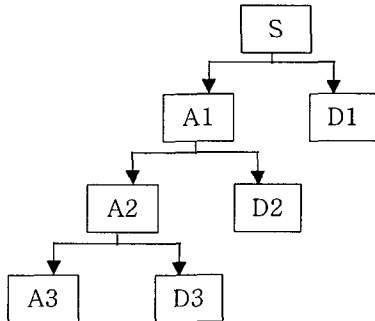


그림 4 Wavelet 변환의 Dynamic Tree 구조

입의 신호에 웨이블릿 변환을 가하면 저역 통과필터(LPF)  $g(n)$ 을 통과한 저주파 신호  $y(n)$ 과 고역통과필터(HPF)  $h(n)$ 을 통과한 고주파 신호  $y(n)$ 으로 나뉘어진다. 여기서,  $g(n)$ 과  $h(n)$ 을 분해 필터 (analysis filter) 라고 한다. 그림 4에서 보면 A1이 저주파, D1이 고주파 신호로 분리되었고 다시 A1에서 저주파 A2와 고주파 D2로 분리되어 나간다. [6]

결과적으로 사람의 청각기관과 같은 시스템으로 음성신호를 해석하게 된다. 음성신호는 주파수 영역에서 관찰할 때 에너지 분포가 고르지 못한 특성이 있다. 그러므로 신호를 분해하는 것이 필요하며 wavelet의 다중 해상도 해석을 통해서 이를 용이하게 할 수 있다.

#### IV. 음성 신호의 웨이블릿 분해 실험

16bit로 sampling 된 22050Hz 음성 wave를 이용해서 실험을 하였다. wavelet 기저함수는 음성신호의 특징을 잘 반영하여 널리 사용되는 Daubechies(db6)을 이용하고 level은 3으로 하였다.

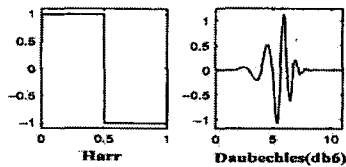


그림 5 기저 웨이블릿의 예

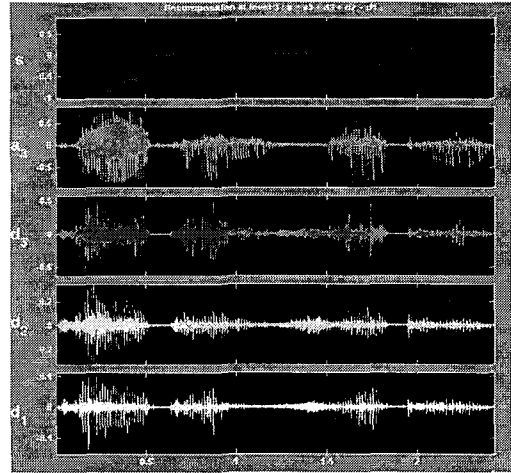


그림 6 '대단하다'의 wavelet 변환

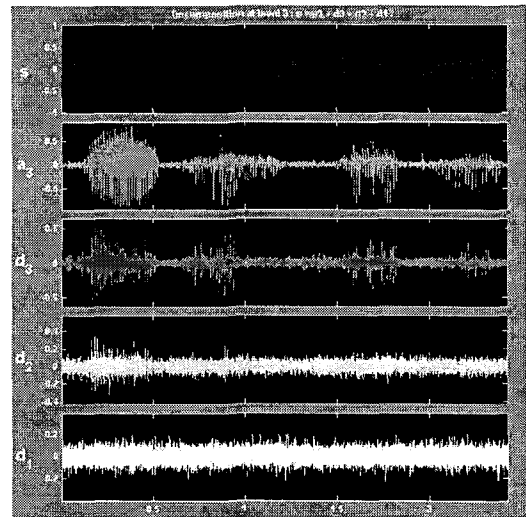


그림 7 30db 잡음 첨가 후 wavelet 변환

잡음을 첨가한 후에 d1,d2의 고주파 부분으로 분리 되어 a3 부분에는 중요한 정보가 그대로 분리됨을 보이고 있다. 웨이블릿을 이용하여 잡음을 분리할 경우 보통 d1,d2의 고주파 부분을 버리고 저주파 신호를 살리는 방법을 사용한다.

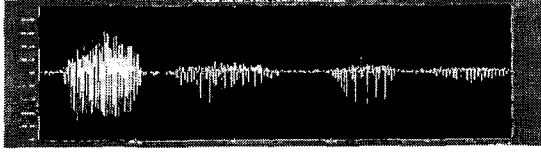


그림 8 필터 처리된 모습

그림 8은 웨이블릿을 이용한 노이즈 제거한 모습이다. 어둡게 표시된 부분이 잡음에 해당하는 부분이다.

## V. 결론

인간의 음성인식 방법을 적용시키기 위해 웨이블릿 변환을 사용해 보았다. 사람은 음성에 대해 서로 다른 특성을 갖는 수많은 필터 뱅크에 의해 주파수 분석을 할 수 있으며 인식의 바탕이 된다. 이렇게 웨이블릿 분석을 하면 잡음이 첨가된 경우에도 원하는 적절히 원하는 음성정보를 찾아 낼 수 있음을 알 수 있었다. 향후 더 연구하여 음성 신호의 분석과 합성을 하여 더욱 인간의 청각 기능에 더욱 가까운 시스템을 연구해 보려고 한다.

## 참고 문헌

- [1] 이흥기, 웨이블릿 변환을 이용한 한국어 음성 인식기, 인하대학교 공학석사학위 논문, 1998.
- [2] S.A.Horowitz, "Syntactic algorithm for peak detection in waveforms with application to cardiography communication," ACM18, 281, 1975.
- [3] 길세기, 웨이블릿 변환을 이용한 맥파의 인식에 관한 연구, 인하대학교 공학석사 학위 논문, 2000.
- [4] Ail N. Akansu and Richard A.Haddad, Multiresolution Signal Decomposition, ACADEMIC PRESS, INC, 1992
- [5] Gilbert Strang and Truong Nguyenr, Wavelet and Filter Banks, Wellesey~Cambridge Press, pp. 360-401, 1996
- [6] 이건상, 양성일, 권영현, 음성인식, 한양대학교 출판부, 2001