

웨이브렛 변환을 이용한 음성신호처리

배 전성, 석 종원
경북대학교 전자·전기 공학부

Speech Signal Processing Using Wavelet Transform

Keun-Sung Bae and Jong-Won Seok
School of Electronics and Electrical Eng., Kyungpook National University

요약

웨이브렛 이론은 응용수학에서 처음 소개된 후 다중해상도 표현 및 이산신호의 부대역 분해방법 등에 대한 단일화된 이론을 제공하고 있으며 최근 신호처리 전반에 걸쳐 널리 이용되고 있는 이론이다.

본 논문에서는 최근 들어 신호처리분야의 새로운 기법으로 제시된 웨이브렛 이론에 대한 소개와 더불어 이를 이용하여 음성개선, 유성음/무성음/목음 판별, 끝점검출, 피치 및 성문 폐쇄시점 검출 등의 음성신호처리에 적용한 예들을 소개한다.

I. 서론

신호를 다양한 해상도로 관측하고 여러 가지 스케일로 나누어서 처리하는 방법은 공학이나 물리학에서 독립적으로 많이 연구되어 왔다. 웨이브렛 변환은 이러한 독립적인 신호처리 분야, 즉, 컴퓨터 비전에서 사용되는 다중해상도(multiresolution) 신호처리, 음성이나 영상의 압축에 사용되는 부대역 부호화(subband coding) 등에 대한 단일화된 이론을 제공한다[1-5]. 웨이브렛 이론은 1910년 Harr가 처음으로 직교 웨이브렛을 만들었으며, 80년대 중반 Morlet 등이 처음으로 웨이브렛을 자진과 연구에 이용하였다. 이후 Daubechies[2-4]와 Mallat[6] 등에 의해 다중해상도 이론(multiresolution theory)이 발표된 후 웨이브렛 변환은 크게 발전하여 응용되고 있다. 이러한 웨이브렛 변환은 시간 및 주파수 국부성(localization)을 가지므로 통계적 특성을 모르거나 시간적으로 예측하기 힘든 신호해석에 널리 이용되고 있으며, 또한 웨이브렛 필터의 성질을 이용하여 신호의 압축 및 잡음제거 등에도 널리 이용되고 있다. 특히, 웨이브렛 변환이 가지는 장점인 시간 및 주파수의 국부성을 이용하여 수중 천이신호의 식별 등에도 연구되고 있으며, 최근에는 초 저속 전송속도 영상압축에도 널리 사용되고 있다[1,5].

한편, 최근 몇 년간 웨이브렛을 이용한 음성신호처리에 대한 연구가 활발히 진행되었는데 대표적인 것은 다음과 같다. Alessandro [7] 와 Ambikairajah [8] 등은 음성 분석 및 합성에 웨이브렛을 이용하였고, Wilde [9]는 dyadic 웨이브렛 변환을 이용하여 음성신호의 formant를 추출하였다. 그리고, Kadambé[10,11]

는 여러 가지 웨이브렛을 이용하여 피치검출에 이용하였으며 각 웨이브렛에 따른 피치검출의 성능을 비교하였다. 또한, Fineberg [12] 와 Favero [13]는 웨이브렛 영역에서 특정 파라미터를 추출하여 이를 음성인식에 이용하고자 하였다. 이와 같이, 웨이브렛 변환을 이용한 음성신호처리에 관한 연구가 앞에서 언급하였듯이 최근에 외국에서 활발히 진행되고 있는데 비해 국내에서의 연구활동은 상당히 미미한 수준이다.

본 논문에서는 저자의 연구실에서 수행한 웨이브렛을 이용한 음성신호처리의 예를 소개하고자 한다. 즉, 음성신호처리 분야에서 다른 주요 연구내용들에 웨이브렛 변환기법을 적용하여 웨이브렛 변환의 이용 가능성을 확인하고 이를 응용한 몇 가지 응용사례들을 소개하고자 한다.

II. 웨이브렛 변환

웨이브렛 변환은 FT(Fourier transform)과 같이 기저함수(basis function)들의 집합에 의한 신호의 분해로 이해될 수 있다. 이때 웨이브렛 변환에서의 각 기저함수를 웨이브렛이라 부르는데, FT의 기저함수들인 정현파와는 달리 유한한 길이를 가지는 기저함수로 정의되므로 웨이브렛 변환은 모 웨이브렛(mother wavelet)이라 부르는 원형 웨이브렛의 수축과 확장에 의해 얻어지는 웨이브렛들의 집합에 의해 얻어진다[2]. 웨이브렛 변환은 다음과 같이 정의된다.

$$(Wf)(a, b) = \langle \psi_{a,b}, f \rangle = \int f(t) \psi^*(a,b)(t) dt \quad (1)$$

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}} \psi\left(\frac{t-b}{a}\right) \quad (2)$$

식 (1)에서 $\psi(t)$ 는 원형(prototype) 웨이브렛이며, $\psi_{a,b}(t)$ 는 $\psi(t)$ 를 이동(shift)과 확장(dilation)하여 구해진다. 식 (2)를 살펴보면 scale factor a 가 증가할 수록 창 함수 $\psi\left(\frac{t-b}{a}\right)$ 의 길이가 길어짐을 알 수 있다. 따라서, 짧은 지속시간을 갖는 고주파신호에

대해서는 짧은 길이의 창함수를 사용하고 긴 지속시간을 갖는 저주파신호에 대해서는 긴 창함수를 이용하는 결과가 되어 주파수 영역에 따른 다중해상도를 갖게 된다. 식 (1)의 웨이브렛 변환은 시간과 scale factor가 연속인 값을 갖기 때문에 이를 실제 이용 가능한 이산적인 형태로 표시하면 다음과 같다.

$$\psi_{a,b}(t) = a_0^{-j/2} \psi(a_0^{-j}t - kT) \quad (3)$$

웨이브렛 변환을 이용한 계수 구현을 더욱 용이하게 하기 위해 $a_0 = 2$ 로 하는 dyadic 웨이브렛 변환은 식(4)와 같이 주어지고 식 (4)를 통상 이산 웨이브렛 변환(discrete wavelet transform)이라 한다.

$$(Wf)_d = 2^{-j/2} \int f(t) \psi^*(2^{-j}t - kT) dt \quad (4)$$

웨이브렛 변환에서는 주파수라는 용어 대신 스케일이라는 용어를 쓰는데, 하나의 웨이브렛을 통과한 신호를 그 스케일에서의 상세신호(detailed signal)라 한다. 이산신호(discrete-time signal)는 스케일의 유한개로 분리되고 첫번째 스케일의 상세신호가 원 신호에 포함되어 있는 가장 높은 주파수 대역의 성분을 나타내고 시간 해상도가 가장 높게 된다. 스케일이 증가할수록 시간 해상도가 떨어지고 점점 낮은 주파수 대역의 성분을 나타내게 된다.

그림 1은 이산 웨이브렛을 구현하기 위해 사용되는 트리(tree)형태의 필터뱅크를 나타내고 있다. 여기에서 H_1 과 H_0 은 각각 저역과 고역통과 필터를 나타내고 $\downarrow 2$ 는 2 배로 다운 샘플링(down sampling) 한다는 의미이다. 이산 웨이브렛 변환시 각 스케일은 각각 바로 위의 스케일에서 factor 2로 decimation 하여 구해지게 되므로 각 scale의 샘플수는 상위 스케일의 반이 된다. 입력신호가 저역통과 필터와 고역통과 필터를 거치게 되면 한 번의 웨이브렛 변환이 이루어지며, 이러한 과정을 반복적으로 수행하여 웨이브렛 변환된 신호를 얻을 수 있다.

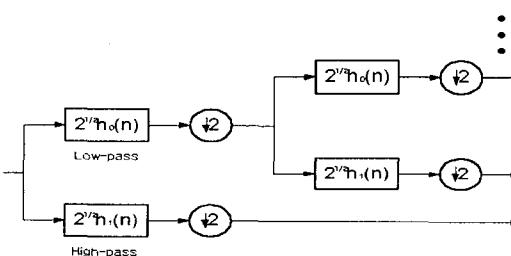


그림 1. 이산 웨이브렛 변환을 위한 트리형태의 필터뱅크 구조

그림 2는 이와 같은 과정을 거친 신호의 주파수 분해를 나타내고 있고 있는데, 이것이 바로 이산 웨이브렛 변환의 주파수 해상도이다. 그리고, 그림 2에서 옥타브 밴드로 나뉘어진 각각의 주파수 밴드를 다시 유한한 갯수의 밴드로 나누고 decimation 과정을 생략하게 되면 변환된 신호의 redundancy는 더욱

증가하게 되는데 이렇게 하여 구해진 웨이브렛 변환을 보통 연속 웨이브렛 변환(continuous wavelet transform)되었다고 한다. 그림 3은 안정된 유성음 한 구간을 연속 웨이브렛 변환하여 이를 2 차원 신호로 구성한 예를 보여주고 있다.

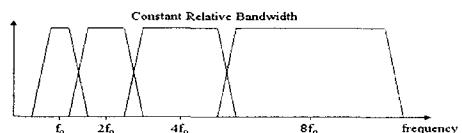


그림 2. 이산 웨이브렛 변환의 주파수 분해

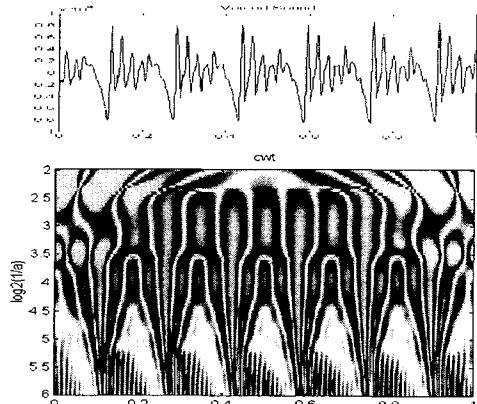


그림 3. 연속 웨이브렛 변환의 예

음성신호와 같이 저역과 고역의 주파수 성분을 함께 가지고 있으며, 특히 지속시간이 짧은 부성음의 고주파 성분이나 긴 지속시간을 갖는 유성음의 저주파 성분 또는 두 가지가 다 포함되어 있는 천이구간의 경우 앞에서 언급한 정현파를 기저함수(basis function)로 하여 고정된 시간-주파수 해상도를 갖는 STFT에 비해 더욱 효과적일 것이라는 것을 알 수 있다. 또한, 웨이브렛을 만들어내는 filter의 특성에 따라 각기 다른 응용에 이용될 수 있는 장점이 있다. Energy concentration 특성을 가지는 filter를 이용하면 데이터 압축, 잡음제거 그리고 신호검출 등에 이용될 수 있다. 그리고 Singularity 검출에 용이한 웨이브렛을 이용하면 영상신호의 edge, 음성신호의 glottal closure instant(GCI)의 검출에 이용될 수 있다. 그림 4 와 5는 실제 응용에 사용되는 smoothing 함수와 웨이브렛 함수를 보여 주고 있다. 그림 4는 Daubechies 4-tap 웨이브렛으로 Energy concentration 성질을 가지고 있어서 신호의 압축 및 잡음제거에 많이 사용된다. 그림 5는 Mallat 가 제안한 웨이브렛으로 신호에 존재하는 급격한 변화부분을 검출하는데 용이하여 영상신호의 edge나 음성신호의 GCI 검출에 널리 이용되고 있다.

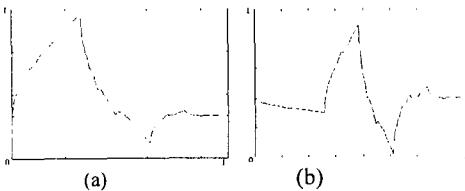


그림 4. Daubechies(4-tap) 웨이브렛의 예
(a) smoothing(scaling) 함수 (b) 웨이브렛 함수

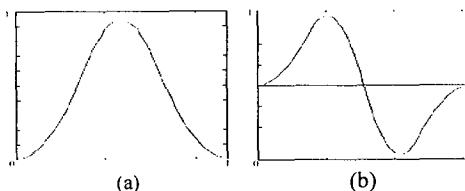


그림 5. Mallat 웨이브렛의 예
(a) smoothing(scaling) 함수 (b) 웨이브렛 함수

III. 웨이브렛 변환을 적용한 음성신호처리

3.1 잡음에 강인한 끝점검출[14-16]

음성신호의 끝점검출은 입력신호로부터 목음을 포함한 주변잡음과 음성을 구분하는 과정이다. 음성인식, 분석 등에서 음성신호의 시작점 및 끝점을 분리하여 정확하게 알아내는 일은 매우 중요하다[17]. 끝점검출을 사용하는 음성인식 시스템은 이러한 결과를 바탕으로 인식을 수행하기 때문에 인식율에 미치는 영향이 크다. 또한 정확한 끝점검출은 인식과정에서 계산량을 최소화 시킬 수 있다[18].

음성의 시작이나 끝부분에 존재하는 파열음이나 마찰음의 경우는 신호의 에너지가 유성음에 비해 작아서 잡음 환경하에서 검출하기가 용이하지 않으며 끝점검출 실패의 주요한 이유 중 하나가 된다[19-20]. 하지만 마찰음이나 파열음의 경우 비록 신호의 에너지는 유성음에 비해 상대적으로 작지만 주파수 영역에서 고주파 부분에 많은 에너지를 가지게 된다. 따라서, 주파수 영역에서의 에너지 분포를 고려함으로써 잡음 환경하에서도 음성을 검출할 수 있는 새로운 검출 파라미터를 웨이브렛 변환 영역에서 정의하였다. 그림 6은 /start/라는 음성을 웨이브렛 변환한 후 구한 첫 번째 detailed 와 세 번째 coarse 스케일에서의 웨이브렛 계수의 표준편차 값의 변화를 보여주고 있다. 그림 6에서도 볼 수 있듯이 세 번째 coarse 스케일은 낮은 주파수 영역에서의 에너지가 큰 유성음 구간을 대표한다고 볼 수 있으며 첫 번째 detailed 스케일은 마찰음이나 파열음의 존재 유무를 알려주게 된다. 이러한 성질을 이용하여 웨이브렛 영역에서 음성검출을 위한 새로운 검출 파라미터를 개발하였다. 즉 식 (5)와 같이 세 번째 coarse 와 가중치를 곱한 첫 번째 detailed 스케일의 합으로 검출 파라미터를 정의하였다.

$$PA = \sigma_{c3} + \lambda \sigma_{d1} \quad (5)$$

식 (5)에서 σ_{d1} 과 σ_{c3} 는 각각 첫 번째 detailed 스케

일의 표준편차와 세 번째 coarse 스케일의 표준편차를 나타내며 λ 는 weighting factor 이다. 그럼 7은 웨이브렛 영역에서 정의된 검출 파라미터를 이용하여 검출한 음성영역과 기존의 대표적인 끝점검출 파라미터인 에너지와 zero-crossing rate 를 이용하여 검출한 음성영역을 HMM 음성인식시스템을 이용하여 인식실험을 수행한 결과를 보여주고 있다.

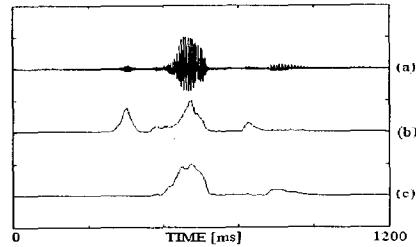


그림 6. 프레임별 웨이브렛 계수의 표준편차 분포
(a) 음성신호 /start/ (b) first detailed scale (c) third coarse scale

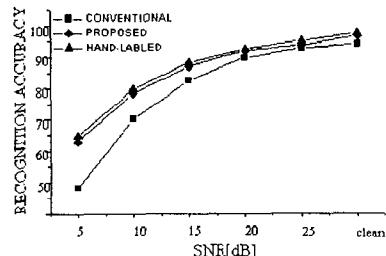


그림 7. 끝점검출 방식에 따른 인식율

3.2 음성개선[21-24]

잡음환경의 영향을 고려하지 않은 음성신호처리 시스템의 성능은 실제 환경에서 급격히 감소하게 된다. 따라서 음성신호에 포함된 잡음을 제거하는 음성개선(speech enhancement) 기술에 대한 연구 및 개발이 꾸준히 이루어지고 있다[25-27].

일반적으로 잡음이 첨가된 신호는 다음 식 (6)과 같이 나타낼 수 있다.

$$y = x + \varepsilon n \quad (6)$$

여기에서, x 는 입력되는 신호를 나타내고 y 는 잡음이 첨가된 신호를 나타낸다. 잡음신호는 평균이 0이고 표준편차가 ε 인 white Gaussian noise 이다. 식 (6)을 웨이브렛 변환 시키면 식 (7)과 같이 나타낼 수 있다.

$$Y = Wx + N \quad (7)$$

여기에서, 대문자는 웨이브렛 영역에서의 신호를 나타낸다. 즉, $Y = Wy$ 이다. 이때 orthogonal 웨이브렛 변환을 가정한다면 아래 식(8)을 만족하는 역 변환 matrix M 이 존재하게 되며.

$$MW = I \quad (8)$$

웨이브렛 영역에서 잡음이 첨가된 신호 Y 에서 원신호성분 X 의 추정값을 \hat{X} 라 하면 추정된 신호는 식(9)로 나타낼 수 있게 된다.

$$\hat{x} = M\hat{X} \quad (9)$$

여기서, \hat{X} 은 단순히 웨이브렛 계수를 유지(keeping)하거나 일정부분을 제거(killing)함으로써 추정될 수 있다. 웨이브렛 영역에서 추정된 잡음 값을 제거할 때 사용되는 thresholding 함수로는 식(10)에 정의되는 semi-soft 함수를 이용하였다.

$$THR_{semi} = \begin{cases} 0, & |Y| \leq \lambda_1, \\ \text{sgn}(Y) \left(\frac{\lambda_2 |Y| - \lambda_1}{\lambda_2 - \lambda_1} \right), & \lambda_1 < |Y| \leq \lambda_2, \\ Y, & |Y| > \lambda_2, \end{cases} \quad (10)$$

그림 8은 잡음신호(SNR=10dB)와 웨이브렛 영역에서 denoising[28] 기법을 적용하여 잡음을 제거한 후의 신호를 보여주고 있다.

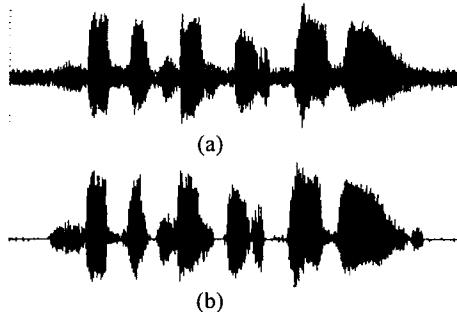


그림 8. 잡음신호와 잡음이 제거된 음성의 파형
(a) 잡음신호(SNR=10dB) (b) 잡음이 제거된 신호

3.3 성문폐쇄시점 검출 및 피치검출[29-32]

음성신호는 성대의 자발적인 운동에 의해 발생되는 공기의 흐름이 성도(vocal tract)를 지나면서 변조되어 공기압의 파동형태로 나타나는 것이다. 특히 모음과 같은 유성음을 발생할 경우 허파에서 방출되는 공기는 성문을 통과하면서 성대가 규칙적으로 진동하도록 만든다. 이처럼 유성음에는 성문이 닫혀져 있는 부분과 열려져 있는 부분이 존재하는데 성문이 닫히는 순간을 성문폐쇄시점(glottal closure instant : GCI)라 한다[33]. 음성신호에서 GCI의 정확한 검출은 피치검출 및 피치동기분석 등을 가능하게 한다.

Mallat는 웨이브렛을 smoothing 함수를 일차 미분한 합수로 선택할 때 웨이브렛 영역에서의 local maxima 가 원 신호에서 급격한 변화부분을 나타낸다는 사실을 증명하였다[6]. 이러한 원리는 음성신호에서 GCI의 검출에 응용될 수 있다. 일반적으로 유성음 발생시 성문의 폐쇄는 시간 영역에서 음성파형에 급격한 변화를 초래하게 되므로 음성신호를 웨이브

렛 변환한 신호에서 local maxima를 검출함으로써 GCI를 검출할 수 있다. GCI의 검출과정은 다음과 같다.

STEP 1: 음성을 200 샘플씩 윈도윙하여 detailed 스케일 3~5 까지 웨이브렛 변환한다. 이때 각 스케일에서의 필터링으로 인한 delay(각 스케일에서의 웨이브렛 길이의 1/2)를 고려해 주었다.

STEP 2: 분석구간이 유성음 구간인지를 판단하여 유성음 구간일 경우 스케일 3~5에 해당하는 신호들을 대상으로 파형을 관찰하면서 local maxima를 찾는다.

STEP 3: Step 2에서 검출한 local maxima들의 위치를 스케일별로 비교하여 스케일 3, 4, 5에 대하여 연속적으로 나타나는 local maxima들의 위치를 검출된 GCI로 한다.

STEP 4: 한 구간의 분석이 끝나면 Step 1로 돌아가서 위의 단계를 반복한다.

그림 9는 안정된 유성음 구간에서 GCI를 검출한 결과를 나타내고 있으며 그림 10은 검출된 GCI를 이용하여 구한 피치캐릭터를 나타내고 있다.

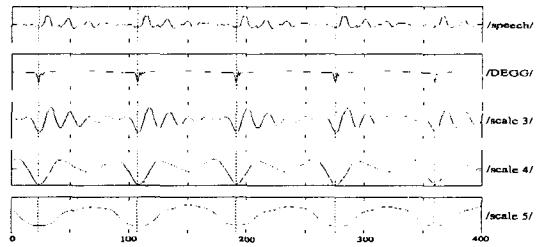


그림 9. 안정된 유성음 구간에서 각 스케일별 GCI 검출 결과

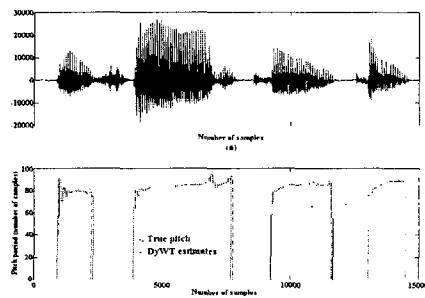


그림 10. 검출된 GCI를 이용하여 구한 피치캐릭터

3.4 유성음/무성음/목음 분류[34-35]

음성신호처리에서 유성음/무성음/목음의 정확한 분류는 피치검출, 음성검출, 음성합성, 음성 세그먼테이션 등의 음성분석에서 중요하게 다루어지는 문

제이다. 이러한 음성분류 알고리즘에서는 음성신호에서의 피치정보, 에너지, 영교차율 등이 파라미터로 널리 사용되고 있다[36,37]. 그러나 피치정보의 경우는 짧은 음성구간의 분류시에 어려움이 있으며 에너지나 영교차율의 경우는 다른 파라미터들과 상호 보완적으로만 사용되는 단점이 있다. 이 절에서는 웨이브렛 변환된 음성신호에서 스펙트럼상의 변화를 검출하기에 적합한 MLR(maximum likelihood ratio) [38,39] 값을 적용하여 음성신호에서 유성음/무성음/목음을 분류하였다. 스케일 s 에서의 MLR_s 은 식(10)과 같이 정의된다.

$$MLR_s(k) = \frac{1}{2} \left| \ln \frac{\sigma^2(k)}{\sigma^2_{\text{non-voicing}}} - \frac{\sigma^2(k)}{\sigma^2_{\text{non-voicing}}} \right| \quad (10)$$

식 (10)에서 k 는 k 번째 프레임을 가리킨다. 그림 9는 음성신호 /fish/에 대한 음성파형과 detailed 스케일 1과 4에서의 MLR 값을 프레임 단위로 나타낸 것으로 스케일 1에서는 상대적으로 무성을 /f/, /ʃ/ 가 스케일 4에서는 유성음 /i/가 잘 나타나고 있음을 볼 수 있다. 따라서, MLR의 이러한 특성을 이용하여 음성신호를 음성구간과 목음구간으로 분류한 뒤 검출된 음성구간을 유성음과 무성음으로 분류하였다. 그림 10은 연속적으로 발생된 문장에 대해 분류된 결과를 보여주고 있다.

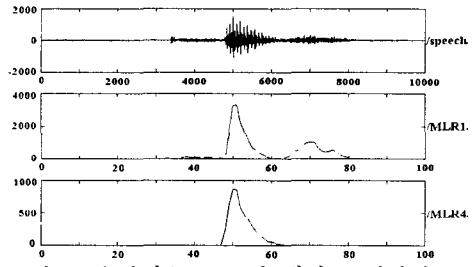


그림 9. 음성신호 /fish/에 대한 프레임별 MLR

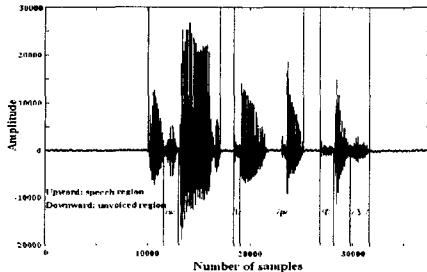


그림 10. 연속적으로 발생된 문장에 대한 유성음/무성음/목음 분류 결과

3.5 정현파 모델링[40]

정현파 모델링[41]은 음성이거나 오디오 신호의 부호화 및 분석/합성 등에 최근 들어 널리 이용되고 있다. 이 절에서는 기존의 정현파 모델링에서 고정된 길이의 window를 사용함으로써 발생되는 문제를 극복하고자 다해상도 신호해석에 따른 각각의 부대역 신호에 대해 각기 다른 길이의 window를 사용하였다.

다해상도 정현파 모델링에서는 nondecimated 웨이브렛을 이용하며 분석/합성 필터뱅크를 통과한 각 채널의 신호들은 독립적으로 정현파 모델링되어 다시 합성하게 된다. 정현파 모델링 시, 각 채널의 출력 신호는 각기 다른 길이의 window를 이용하여 분석하게 된다. 즉, 신호의 급격한 변화를 잘 반영하는 주파수가 높은 밴드의 신호일수록 짧은 window를 사용하고 반대로 신호의 전체적인 특성을 반영하는 낮은 밴드로 갈수록 긴 window를 사용하게 된다.

이와 같이 주파수에 따른 가변적 window를 이용함으로써 기존의 방법보다 더 작은 수의 정현파로 동일한 수준의 합성음을 재생할 수 있게 된다. 또한 신호의 onset time이 분석구간의 중간에 위치할 경우 분석구간 전체에 걸쳐 영향을 미치게 되어 합성음의 음질을 저하시키게 되는 pre-echo 왜곡현상도 감소시킬 수 있는 장점이 있다. 그림 11은 다해상도 정현파 모델링 과정을 나타내고 있다. 그림 12는 기존의 정현파 모델링 방식을 사용하여 합성한 파형과 다해상도 정현파 모델링을 사용하여 합성한 파형을 각각 보여주고 있다. 그림 12(c)와 같이 다해상도 신호해석을 사용할 경우 그림 12(b)와 경우 보다 약 20~30%의 정현파 수를 적게 사용하면서도 양호한 합성음을 재생시킬 수 있다.

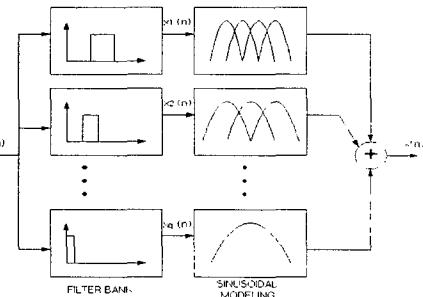


그림 10. 다해상도 정현파 모델링 과정

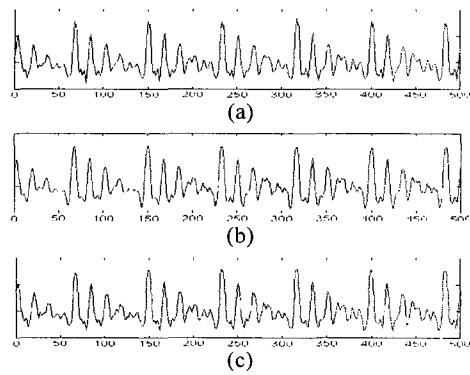


그림 4. 정현파 모델링을 이용한 음성합성
(a) 원신호 (b) 기존의 방식(30 개의 정현파 사용)
(c) 다해상도 정현파 모델링

IV. 결론

본 논문에서는 음성신호처리 분야에서 다루는 주요 연구내용들인 음성개선, 유성음/무성음/목음의 분류, 괴치 및 성문 폐쇄시점 검출, 끝점검출 등에 웨이브렛 변환기법을 적용한 예들을 소개하였다.

본 논문에서 얻은 연구결과는 아직 국내에서는 연구가 활발하지 못한 웨이브렛 변환을 이용한 음성 신호처리 분야에 많은 응용 및 활용 가능성을 제시했다고 할 수 있으며, 또한 본 연구를 통해 개발된 알고리듬들은 실제 음성신호처리 전반에 걸쳐 널리 활용될 수 있으리라 기대된다.

참고문헌

- [1] O.Riou and M. Vetterli. "Wavelet and Signal Processing", IEEE Signal Processing Magazine, pp. 14-38, Oct. 1991
- [2] I. Daubechies. Ten Lectures on Wavelets. SIAM. 1992.
- [3] I. Daubechies. "The Wavelet Transform Time Frequency Localization and Signal Analysis." IEEE Trans. on Info. Theory. Vol.36. No.5. pp. 961-1005, Sep. 1990
- [4] I. Daubechies. "Orthonormal Bases of Compactly Supported Wavelets." Comm. in Pure and Applied Math. Vol. 41. pp.909-996. 1988
- [5] M. Vetterli. "Multi-dimensional Sub-band Coding." Signal Processing. Vol. 6. No. 2. pp. 97-112. 1984.
- [6] S. Mallat and W. L. Hwang. "Singularity Detection and Processing with Wavelets." IEEE Trans Information Theory. vol.38. no.2. Mar. 1992.
- [7] Alessandro. C. "Speech analysis and synthesis using auditory-based wavelet representation." In ETRW:Signal Representations. Sheffield, England. pp. 31-38. 1992.
- [8] Ambikairajah, et al. "The application of wavelet transform for speech processing." In EUROSPEECH, pp. 151-154. 1993
- [9] Wilde S. and Curtis. K., "The wavelet transform for speech analysis." In ICSLP, pp.1621-1624. 1992.
- [10] Kadambé. S. and Bartels. G., "A comparison of wavelet functions for pitch detection of speech signals." In ICASSP. pp.449-452, 1991.
- [11] Kadambé. S. and Bartels. G., "Application of the Wavelet Transform for Pitch Detection of Speech Signals," IEEE Trans Information Theory. vol. 38, no. 2, Mar.1992.
- [12] Fineberg A. "A time-frequency analysis technique for speech recognition signal processing," In ICSLP, pp. 1615-1618. 1994.
- [13] Favero. R. and Gurgen F. "Using wavelet dyadic grids and neural networks for speech recognition." In ICSLP. pp.1539-1542. 1994.
- [14] Seok Jong Won and Bae Keun Sung. "A novel endpoint detection method for noisy speech signal." In Pro. ICA/ASA-98. vol. 103. no. 5. pp. 2818. 1998.
- [15] Seok Jong Won and Bae Keun Sung. "A novel endpoint detection using discrete wavelet transform." submitted to IEICE Trans. on Information and System. 1998
- [16] 석 종원, 배 건성. "웨이브렛 변환을 이용한 음성신호의 끝점검출". 한국음향학회 논문지 제출. 1999
- [17] L. R. Rabiner and M. R. Sambur. "An Algorithm for Determining the Endpoints of Isolated Utterances." Bell Syst. Tech. J. Vol. 54. No. 2. pp. 297-315. February 1975.
- [18] L.F. Lamel et. al., "An improved endpoint detector for isolated word recognition." IEEE Trans. ASSP. Vol. ASSP-29. No.4. pp. 777-785. 1981.
- [19] Jean-Claude Junqua, Brian Mak, and Ben Reave. "A Robust Algorithm for Word Boundary Detection in the Presence of Noise" IEEE Trans. Acoust., Speech, and Signal Processing. Vol 2. No.3. pp.406-412, 1994.
- [20] M. H. Savojo. "A Robust Algorithm for Accurate Endpointing of Speech Signals." Speech Communication. Vol. 8. No. 1. pp. 45-60. March 1989.
- [21] Seok Jong Won and Bae Keun Sung. "Speech enhancement with reduction of noise components in the wavelet domain." In Pro. IEEE ICASSP-97. vol. 2, pp.1323-1326. 1997
- [22] Seok Jong Won and Bae Keun Sung, "Removal of noise from speech using wavelet transform.", accepted in Journal of Electrical Engineering and Information Science 1999.
- [23] 한 미경, 석 종원, 배 건성, "다해상도 신호해석을 이용한 음성개선 방식 비교." 대한전자공학회 추계학술발표 대회. 제 21 권, pp. 1251-1254. 1998.
- [24] 석 종원, 한 미경, 배 건성, "다해상도 신호해석을 이용한 음성개선 방식 비교." 대한전자공학회 논문지 제출.
- [25] S. F. Boll. "Suppression of acoustical noise in speech using spectral subtraction," IEEE Trans. on ASSP. vol. 27. no. 2. 1979.
- [26] J. S. Lim and A. V. Oppenheim, "Enhancement and bandwidth compression of noisy speech." Proc. IEEE. vol. 67. no. 12. 1979.
- [27] J. Hansen and M. Clements. "Constrained iterative speech enhancement with to speech recognition.", IEEE Trans. on ASSP. vol. 39. no. 4. 1989
- [28] D. L.Donoho."Denosing by soft thresholding". IEEE Trans. on Info. Theory. 1994.
- [29] 손 영호, 석 종원, 배 건성, "EGG 신호와 웨이브렛 변환된 음성신호와의 Epoch 검출비교." 제 10 회 신호처리 학술대회, pp. 743-746. 1997.
- [30] 손 영호, 석 종원, 배 건성, "Quadratic Spline Wavelet 함수를 dldyd 한 괴치검출." 한국음향학회 학술발표대회. 제 16 권. pp. 119-122. 1997.
- [31] Seok Jong Won, Son Young Ho and Bae Keun Sung. "Comparison between DyWT- and EGG-based estimation of glottal closure instant for speech signal," In Proc. ICA/ASA-98. vol. 103. no. 5. pp. 2818. 1998.
- [32] Seok Jong Won and Bae Keun Sung, "Comparison of glottal closure instants obtained by using wavelet transform of speech signal and EGG signal." submitted to IEICE Trans. on Information and System. 1998
- [33] A. K. Krishnamurthy. D.G. Childers. " Two-Channel Speech Analysis." IEEE Trans. on ASSP. vol.34. No.4
- [34] 손 영호, 배 건성, "웨이브렛 변환을 이용한 음성신호의 유/무성음/목음 분류." 음성통신 및 신호처리 워크샵. 제 15 권. pp. 449-453. 1998.
- [35] 손 영호, 배 건성, "웨이브렛 변환을 이용한 음성신호의 유/무성음/목음 분류." 음성과학 논문지. 제 4 권. pp. 41-54. 1998.
- [36] Bishnu S.ATAL, Lawrence R. Rabiner "A Pattern Recognition Approach to Voiced, Unvoiced and Silence Classification with Applications to Speech Recognition" IEEE Trans ASSP. vol. 24. No. 3. Jun. 1976
- [37] Leah J. Siegel, "A Procedure for Using Pattern Classification Techniques to Obtain a Voiced/Unvoiced Classifier". IEEE Trans ASSP. vol. 27. No.1 Feb. 1979
- [38] B. T Tan, R.Lang, Heiko Schroder. Andrew Spray "Applying wavelet analysis to speech segmentation and classification" SPIE Wavelet Applications vol. 2242. 1994
- [39] V.Ralph Algazi, Kathy L. Brown, Michael J. Ready, David H. Irvine "Transform Representation of the Spectra of Acoustic Speech Segments with Applications-I". IEEE Trans Speech and Audio Processing. vol. 1. No.2 . Apr. 1993.
- [40] 석 종원, 배 건성, "다해상도 신호해석을 이용한 음성 신호의 정현과 모델링." 제 11 회 신호처리 학술발표 대회. pp. 305-308. 1998.
- [41] R. J. McAulay and T. F. Quatieri. "Speech analysis/ synthesis based on a sinusoidal representation." IEEE Trans. Acoustic, Speech, Signal Processing. Vol.34. pp 744-754. Aug.. 1986