

웨이브렛 변환을 이용한 음성신호의 성문폐쇄시점 검출

Detection of Glottal Closure Instant for Voiced Speech Using Wavelet Transform

배 건 성*
(Keun Sung Bae)

ABSTRACT

During the phonation of voiced sounds, instants exist where the glottis is opened or closed, due to the periodic vibration of the vocal cord. When closed, this is called the glottal closure instant(GCI) or epoch. The correct detection of the GCI is one of the important problems in speech processing for pitch detection, pitch synchronous analysis, and so on. Recently, it has been shown that the local maxima points of the wavelet transformed speech signal correspond to the GCIs of speech signal. In this paper, we investigate the accuracy of GCIs estimated from this wavelet transformed speech signal. For this purpose we compare them with the negative peak points of the differentiated EGG signal that represents the actual GCIs of speech signal.

Keywords : glottis, vocal tract, GCI, EGG, wavelet transform

I. 서 론

음성신호는 발성기관에서 2개의 얇은 막으로 구성된 성대의 자발적인 운동에 의해 발생하는 공기의 흐름이 성도(vocal tract)를 지나면서 변조되어 공기압의 파동형태로 나타나는 것이다. 이러한 음성신호는 성대를 통과한 공기흐름의 성질에 따라 크게 유성음(voiced sound)과 무성음(unvoiced sound)으로 나눌 수 있다. 특히 모음과 같은 유성음을 발성할 경우 허파에서 방출되는 공기는 닫혀진 성대에 의해 공기압이 점차 증가하여 성대가 떨어지기 시작하면서 좁은 공기 통로를 형성하게 되는데 이를 성문이라고 하며, 성문을 통과하는 공기는 서로 다른 두 힘의 상호작용에 의해 성대가 규칙적으로 진동을 하도록 만든다[1]. 이처럼 유성음에는 성문이 닫혀져 있는 부분과 열려져 있는 부분이 있게 되는데 성문이 닫히는 순간을 GCI 또는 epoch이라 한다. 이러한 성문폐쇄시점의 정확한 검출은 순시피치검출[2], 피치동기분석[3] 등에 필수적이므로 음성신호처리 분야에서 중요한 문제로 다루어지고 있다.

최근에는 음성신호의 웨이브렛 변환된 신호에서 성문폐쇄시점을 검출하여 이 결과를 음성신호의 피치주기 검출에 이용하려는 다양한 연구가 소개되었는데, Kadambe[4] 등은 음성

* 경북대학교 전자·전기공학부

신호를 웨이브렛 변환한 신호에서 국부 최대값(local maxima)을 검출하여 음성신호의 피치를 검출하는 방법을 제안하였다. 그러나 이러한 피치검출에 관한 연구들에서는 웨이브렛 변환을 이용하여 일차적으로 검출한 성문폐쇄시점이 실제의 성문폐쇄시점과 어느 정도 일치하는지에 대한 실험을 통한 분석없이 단지 성문폐쇄시점과 정확히 일치하는 것으로 간주하여 피치검출에 이용하고 있다. 웨이브렛 변환을 이용한 음성신호의 피치검출 방법이 보다 타당성을 가지고 또 효과적으로 이용되기 위해서는 웨이브렛 변환으로 검출한 성문폐쇄시점이 실제의 성문폐쇄시점과 비교하여 어느 정도 정확한가에 대한 분석이 필요하다. 따라서 본 연구에서는 음성신호의 웨이브렛 변환으로 검출한 성문폐쇄시점이 실제의 성문폐쇄시점과 어느 정도 잘 일치하는지를 성문폐쇄시점 검출에 있어서 기준신호로 적합한 미분된 EGG (DEGG: differentiated electroglottograph) 신호에서의 음의 극대값(negative peak) 위치와 비교·분석하였다. 음성신호의 웨이브렛 변환에서 기저함수로는 Mallat[5]가 제안한 quadratic spline 웨이브렛 함수를 이용하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 음성신호의 성문폐쇄시점과 EGG(electroglottograph) 신호에 대한 간략한 설명을 한다. 3장에서는 웨이브렛 변환에 대한 소개와 함께 quadratic spline 웨이브렛 함수를 이용한 음성신호의 성문폐쇄시점 검출에 대하여 설명하고, 4장에서는 웨이브렛 변환을 이용하여 실험한 결과를 DEGG 신호를 이용하여 얻은 결과와 비교·분석하여 5장에서 결론을 맺는다.

II. 음성신호의 성문폐쇄시점

일반적으로 모음과 같은 유성음 생성시에 허파에서 방출되는 공기는 닫혀진 성대에 의해 공기압이 증가하여 성대가 떨어지기 시작하면서 성문(glottis)이라 불리는 좁은 공기 통로를 형성하게 되며, 성문을 통과하는 공기는 서로 다른 두 힘의 상호작용에 의해 성대가 규칙적으로 진동하게 한다. 이때 두 힘은 각각 성문 아래에서의 압력(subglottal pressure)과 Bernoulli 효과이다[1]. 성문 아래에서의 압력은 닫혀진 성대를 떨어지도록 하여 성문이 열리도록 작용하며, Bernoulli 효과는 열린 상태의 성대를 서로 당기도록 하여 성문이 닫히도록 작용한다. 일반적으로 성문은 서서히 열리고 급격히 닫히는 특성을 지닌다. 특히 유성음의 경우는 성대가 급격히 닫히는 시점에서 최대의 에너지로 여기되고 성문이 닫혀 있는 동안에는 여기원이 없으므로 조음구조와 성도의 물리적 특성에 따른 자연스런 감쇠진동을 일으키게 된다. 그러나 성문이 서서히 열리는 시점에서는 열린 성문과 음원 신호에 의해 자연스런 감쇠진동은 방해를 받으므로 진동 주파수가 변화하고 더욱 급격한 감쇠 진동을 하다가 다시 성문이 급격히 닫히면서 위와 같은 개폐 과정이 반복되게 된다. 이처럼 유성음의 경우는 성문이 닫혀져 있는 시점, 즉 성문폐쇄시점이 존재하게 되는데, 음성신호 처리시에 이러한 성문폐쇄시점의 정확한 검출은 피치검출 및 피치동기분석 등을 가능하게 한다.

성문폐쇄시점을 검출하는 방법 중에 특히 많이 사용되는 방법으로 음성신호와 동시에 녹음하여 성문 진동을 관측할 수 있도록 고안되어진 EGG 신호를 이용하는 방법이 있다[6]. 이 방법은 녹음시에 부가적인 장비를 착용하여야 하는 한계가 있으나 검출이 용이하고 정확한

성문폐쇄시점을 검출할 수 있다는 점에서 성문폐쇄시점 검출시에 다른 방법들에 대한 기준으로 많이 사용되고 있다.

EGG 신호는 성대의 진동운동을 간접적으로 관측하기 위해 고안된 방법 중의 하나로 성대의 떨림을 임피던스 변화로 바꾸어 전기신호로 나타낸 것으로 laryngograph 신호라고도 하는데 여기신호와 성도 포먼트 사이의 상호작용에도 영향을 받지 않으면서 발성시의 주변 잡음에도 관계없이 성대의 떨림에 관한 정확한 정보를 준다. 특히, Childers[3, 6, 7] 등은 성대의 진동운동으로 인한 성문의 열린 정도를 초고속 촬영하여 그 결과를 그와 동기된 EGG 신호와 비교함으로써 EGG 신호가 아래쪽으로 최대편향이 급속히 일어나는 순간, 즉 DEGG 신호가 음의 극대값(negative peak)을 가지는 시점이 성대가 서로 가까워져 성문이 닫히는 순간과 일치하고 DEGG 신호의 양의 극대값(positive peak)이 생기는 시점은 성대가 서로 떨어져 성문이 최대로 열리는 순간과 대체적으로 일치한다는 것을 실험적으로 보였다. 그림 1은 남성화자가 발성한 문장내의 안정된 유성음 구간에서 음성신호, EGG 및 DEGG 신호를 나타내고 있다. 그림에서 수직방향의 점선은 실제의 성문폐쇄시점에 해당하는 DEGG 신호에서의 음의 극대값의 위치를 나타내고 있다. 따라서, 본 연구에서는 음성신호에 동기된 DEGG 신호를 웨이브렛 변환을 이용하여 구한 성문폐쇄시점과 어느 정도 일치하는지를 판단하는 기준신호로 이용하였다.

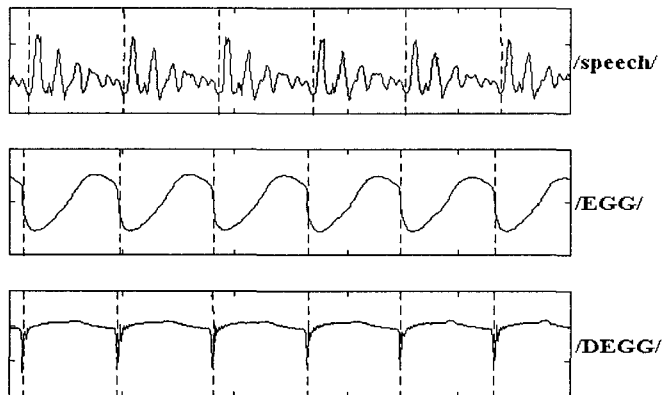


그림 1. 안정된 유성음 구간에서의 성문폐쇄시점과 EGG 신호와의 관계

III. 웨이브렛 변환을 이용한 성문폐쇄시점 검출

3.1 이산 웨이브렛 변환(Discrete Wavelet Transform)

웨이브렛 변환 이론은 응용수학에서 처음 소개된 후 최근 컴퓨터 비전 분야 등에서 연구되어 온 다중해상도(multiresolution) 신호처리 분야 및 음성이나 영상의 압축에 사용되는 부대역 부호화(subband coding) 등과 연관성이 있는 일종의 신호해석 방법이다[8]. 일반적으로 웨이브렛 변환은 모웨이브렛(mother wavelet)이라 부르는 원형 웨이브렛의 확장(dilation) 및 축소(contraction) 등의 과정을 통하여 특정 주파수 대역의 신호 성분을 분리하여 해석하

는 것을 가능하게 한다. 특히 고주파 대역에서는 축소된 웨이브렛을 이용하여 시간영역에서의 분해능을 개선시키고, 저주파 대역에서는 확장된 웨이브렛을 이용하여 주파수 영역에서의 분해능을 향상시킬 수 있다. 실제 응용에 이용되는 이산 웨이브렛 변환식은 식 (1), (2)로 주어진다. 특히 식 (2)의 이산 웨이브렛에서 계수 구현을 용이하게 하기 위하여 a_0 를 2로 한 dyadic 웨이브렛 변환(DyWT: dyadic wavelet transform)은 식 (3)과 같이 정의된다. 식 (1)~(3)에서 d 는 이산변환을 가리키며, j 는 스케일(scale), k 는 이동(translation), T 는 이동 간격을 나타낸다.

$$d_{j,k} = \int f(t) \varphi_{j,k}^*(t) dt \tag{1}$$

$$\varphi_{j,k}(t) = a_0^{-\frac{j}{2}} \varphi(a_0^{-j}t - kT) \tag{2}$$

$$d_{j,k} = \frac{1}{\sqrt{2^j}} \int f(t) \varphi^*\left(\frac{t}{2^j} - kT\right) dt \tag{3}$$

DyWT은 신호처리 관점에서 볼 때 constant-Q, octave band 특성을 갖는 대역통과 필터들의 출력으로 볼 수 있으며, DyWT을 이용하여 신호를 분석하는 과정은 그림 2에서와 같이 트리(tree) 형태로 구성된 필터뱅크 구조로 볼 수 있다. 그림 2의 블록도는 Mallat[5]가 제안한 웨이브렛 분해 필터뱅크 구조이다. 여기서 H_0 는 저역통과 필터이고, H_1 은 고역통과 필터이다. 입력신호가 저역통과 필터와 고역통과 필터를 차례로 거치게 되면 한번의 웨이브렛 변환이 이루어지며, 저역통과 필터를 통과한 신호에 대해 이러한 과정을 반복적으로 수행하여 웨이브렛 변환된 신호를 얻을 수 있다. 이 과정에서는 필터링된 신호의 크기를 그대로 유지하면서 다음 스케일 신호를 얻기 위해 필터링된 신호를 간축(decimation)하는 일반적인 부대역 분할 방법과 달리 필터 계수들 사이에 0을 삽입하여 신호와 컨볼루션함으로써 필터링된 신호를 간축하지 않는다. 이처럼 필터링된 신호가 같은 크기로 유지되어야 하는 이유는 신호에서 급격히 변화는 부분의 위치를 각 스케일에서 보존하기 위함이다. 한편 본 연구에서는 웨이브렛 변환시에 그림 2에서와 같이 웨이브렛 변환된 신호를 스케일 1에서부터 순차적으로 구하는 방법 대신에 각 스케일에 해당하는 웨이브렛 함수를 직접 구현하여 음성신호를 웨이브렛 변환함으로써 필요한 스케일의 신호를 직접 얻는 방법을 사용하였다.

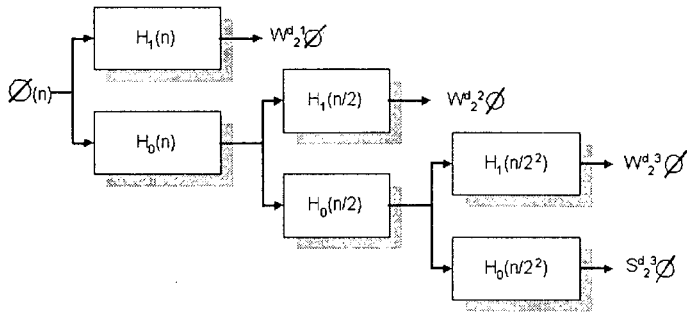


그림 2. 이산 웨이브렛 변환에 대한 dyadic 필터뱅크 구조

3.2 웨이브렛 변환을 이용한 성문폐쇄시점 검출

Mallat는 그림 3에서와 같은 스무딩(smoothing) 함수의 일차미분 형태로 웨이브렛 함수를 정의할 때 DyWT의 국부 최대값은 신호에 있어서 급격한 변화부분에 해당한다는 것을 입증하였다. 이때 스무딩 함수란 저주파 영역에 함수의 에너지가 집중되어 있는 함수를 가리킨다. 더욱이 Mallat는 실제 신호가 시간 $t=t_0$ 에서 갑작스런 변화를 하게 될 때 $t=t_0$ 에서 연속적인 스케일에 걸쳐서 DyWT한 값들은 국부 최대값을 지닌다는 사실도 입증하였다[5]. 본 연구에서는 웨이브렛 함수로 Mallat에 의해 제안된 quadratic spline 웨이브렛을 사용하였는데, 사용된 필터의 계수는 표 1과 같다. 표 1에서 H_0 는 저역통과 필터이고, H_1 은 고역통과 필터이다. 그림 3은 표 1에서의 필터계수를 이용하여 구현한 스무딩 함수와 웨이브렛 함수의 파형을 나타내고 있다. 앞에서 언급하였던 것처럼 (b)의 웨이브렛 함수는 (a)의 스무딩 함수를 일차 미분한 형태임을 확인할 수 있다.

표 1. Quadratic spline 웨이브렛의 필터 계수

n	H_0	H_1
-1	0.125	
0	0.375	-2.0
1	0.375	2.0
2	0.125	

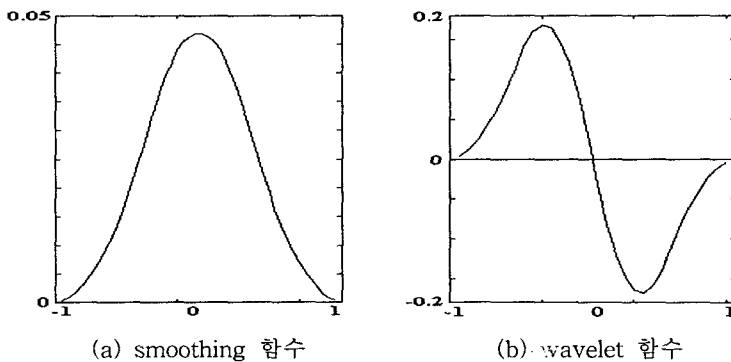


그림 3. Quadratic spline 웨이브렛의 스무딩 함수와 웨이브렛 함수

그림 4는 신호에서의 급격한 변화부분이 웨이브렛 변환시에 연속적인 스케일에 걸쳐서 국부 최대값으로 나타나는 예를 제시한 것으로 신호에서 급격히 증가하거나 감소하는 부분들이 웨이브렛 변환된 신호에서 국부 최대값들로 표현되는 것을 확인할 수 있다. 한편 스케일이 증가할수록 국부 최대값 성분들이 점차로 스무딩되어지는 현상을 볼 수 있는데 이러한 현상은 웨이브렛 변환시 스케일이 증가할수록 기저함수 웨이브렛의 주파수 성분이 저주파 대역으로 이동하기 때문이다.

웨이브렛 변환을 이용한 신호의 급격한 변화부분의 검출 원리는 음성신호에서 성문폐쇄시점의 검출에도 적용될 수 있다. 앞서 2장에서 언급한 것처럼 음성신호에서 유성음 발생시

에 존재하는 성문의 폐쇄는 시간 영역에서 음성파형에 급격한 변화를 초래하게 된다. 그러므로 성문폐쇄시점은 음성신호를 DyWT한 신호에서 국부 최대값으로 나타나게 되며 이러한 특징을 이용하여 음성신호를 웨이브렛 변환한 신호에서 국부 최대값을 검출함으로써 음성신호의 성문폐쇄시점을 검출할 수 있다.

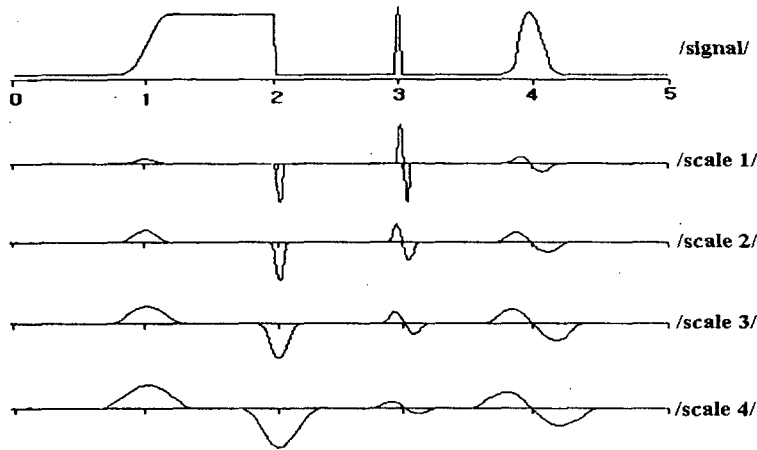


그림 4. 신호에서 급격한 변화 부분에 대한 웨이브렛 변환

웨이브렛 변환을 이용한 음성신호의 성문폐쇄시점의 검출시에 가장 핵심적인 부분은 웨이브렛 변환된 신호에서 국부 최대값을 정확하게 검출하는 것이다. 본 연구에서는 음성신호의 웨이브렛 변환된 신호 중에서 스케일 3~5의 신호만을 대상으로 국부 최대값을 검출하여 실제의 성문폐쇄시점과 비교하였다. 이처럼 웨이브렛 변환된 신호의 스케일을 제한한 것은 스케일이 상대적으로 낮을 경우는 웨이브렛 변환된 신호의 주파수 성분이 상대적으로 고주파 영역에 많이 존재하여 국부 최대값 외의 다른 고주파 성분들이 많이 나타나서 국부 최대값의 정확한 검출에 어려움이 있게 되며, 반대로 스케일이 과도하게 증가할 경우는 웨이브렛 변환된 신호의 주파수 성분이 저주파 영역으로 접근하여 신호 성분이 점차적으로 스무딩되어 나타남으로써 역시 웨이브렛 변환된 신호에서 국부 최대값을 정확하게 검출하기 어렵기 때문이다.

국부 최대값의 검출에 있어서는 자동으로 검출할 경우 문턱값 설정 등의 문제로 인하여 성문폐쇄시점에 해당하는 국부 최대값을 검출하지 못하게 되는 오류(GCI-deletion error)나 성문폐쇄시점이 아닌 국부 최대값을 검출하게 되는 오류(GCI-insertion error) 등이 발생하게 되어 정확한 비교를 할 수 없게 된다. 따라서 이러한 점을 고려하여 오류 없이 가능한 정확하게 국부 최대값을 검출하여 실제의 성문폐쇄시점과 비교·분석하기 위하여 유성음 구간을 대상으로 웨이브렛 변환된 신호를 음성파형과 동시에 관찰하면서 수동으로 국부 최대값을 검출하였다. 본 연구에서 quadratic spline 웨이브렛 함수를 이용한 음성신호의 성문폐쇄시점 검출 과정은 다음과 같다.

- STEP 1: 음성을 200샘플씩 윈도우하여 스케일 3~5까지 웨이브렛 변환한다. 이때 각 스케일에서의 필터링으로 인한 지연(각 스케일에서의 웨이브렛 길이의 1/2)을 보상해 준다.
- STEP 2: 분석구간이 유성음 구간인지를 판단하여 유성음 구간에 대하여 스케일 3~5에 해당하는 신호들을 대상으로 파형을 관찰하면서 국부 최대값을 찾는다.
- STEP 3: Step 2에서 검출한 국부 최대값들의 위치를 스케일별로 비교하여 스케일 3, 4, 5에 대하여 연속적으로 나타나는 국부 최대값들의 위치를 성문폐쇄 시점으로 한다.
- STEP 4: 한 구간의 분석이 끝나면 Step 1로 돌아가서 위의 단계를 반복하여 전체 음성에 대하여 분석한다.

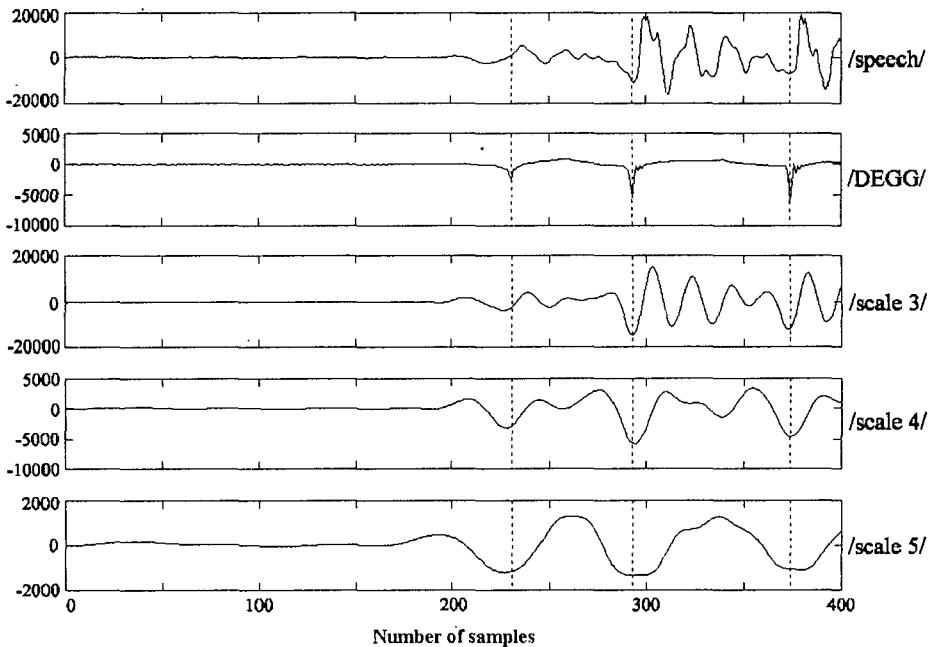
IV. 실험 및 고찰

본 연구에서는 웨이브렛 변환으로 검출된 성문폐쇄시점을 실제의 성문폐쇄시점과 비교하기 위하여 남자 2, 여자 2명의 화자로부터 4개의 문장을 대상으로 수집된 음성 데이터와 EGG 신호를 이용하였다. 한편 실험에 사용된 음성 데이터와 EGG 신호는 각각 10 kHz 샘플링하고 16비트로 양자화 하였으며, 특히 EGG 신호의 경우는 성문에서 검출되는데 비해 음성신호는 마이크에서 검출되므로 성문에서 입까지, 입에서 마이크까지 소리가 전달되는데 걸리는 시간만큼 음성신호는 EGG 신호보다 지연되어 나타나게 된다. 따라서 본 연구에서는 이러한 점을 고려하여 EGG 신호를 지연시켜 음성신호에 동기시켜 사용하였다.

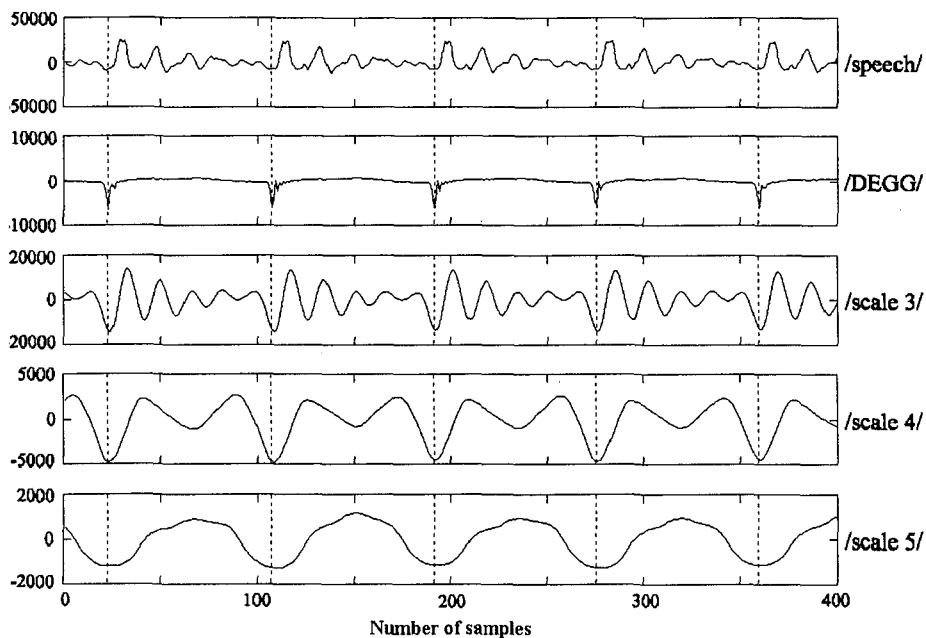
그림 5는 웨이브렛 변환으로 검출한 성문폐쇄시점과 실제의 성문폐쇄시점을 비교·분석하기 위하여 남성화자가 발성한 음성 데이터를 대상으로 음성신호, DEGG 신호, 그리고 스케일 3~5에 해당하는 웨이브렛 변환된 신호를 비교하여 나타내고 있다[9]. 이때 웨이브렛 변환과정에서 필터 계수의 극성이 반전되어 실험 결과에서 국부 최대값도 음의 극대값의 형태로 주어진다. 그림에서 수직방향의 점선은 DEGG 신호에서 음의 극대값의 위치를 나타내고 있다. 그림 5의 (a)는 유성음의 시작 구간을, (b)는 안정된 유성음 구간을, 그리고 (c)에서는 유성음이 끝나는 구간의 예를 각각 나타내고 있다. 이 결과에서 웨이브렛 변환으로 검출된 성문폐쇄시점은 실제의 성문폐쇄시점과 대체로 잘 일치함을 볼 수 있으며, 또 실제의 성문폐쇄시점에 해당하는 국부 최대값은 연속된 스케일에 걸쳐서 이전 스케일에서 국부 최대값 위치를 기준으로 몇 샘플 범위내의 위치에서 국부 최대값으로 나타남을 확인할 수 있다. 그러나 각 스케일에 걸쳐 웨이브렛 변환된 신호의 국부 최대값의 위치는 실제의 성문폐쇄시점과 정확히 일치하는 것이 아니라 성문폐쇄시점을 중심으로 좌우로 몇 샘플의 변화를 보이는 것을 볼 수 있다. 이러한 변화는 스케일이 증가할수록 크게 나타남을 볼 수 있다. 따라서 웨이브렛 변환을 이용한 성문폐쇄시점의 검출시에 스케일에 따른 국부 최대값의 위치

변화량에 대한 고려가 있어야 할 것이다. 그리고 유성음 구간의 위치에 따른 검출된 성문폐쇄시점의 정확성 측면에서는 안정된 유성음 구간에서 검출된 성문폐쇄시점이 유성음의 시작이나 끝구간에서 검출된 성문폐쇄시점에 비해 상대적으로 실제의 성문폐쇄시점과 더 잘 일치하는 것을 확인할 수 있다

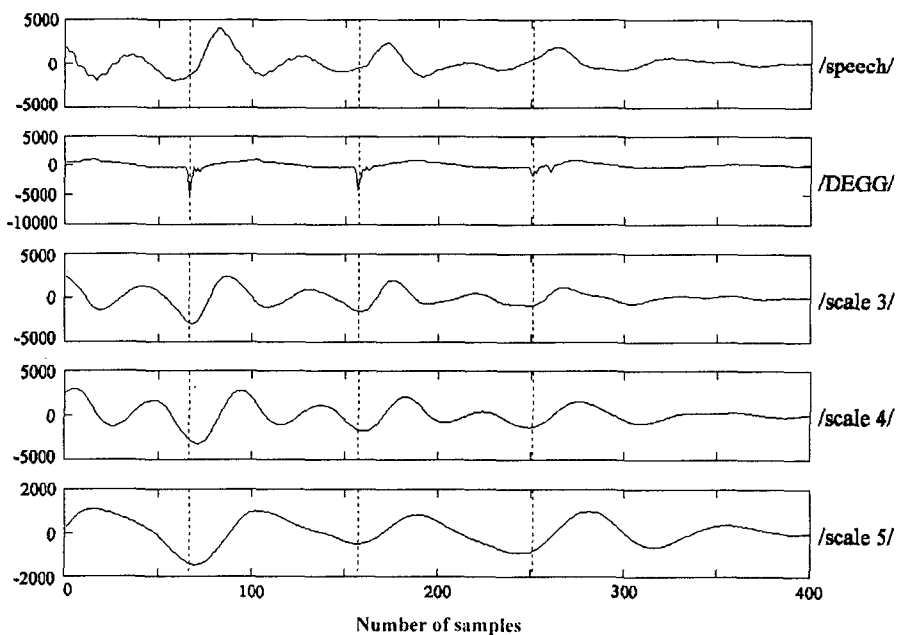
그림 6에서는 EGG 신호를 이용하여 검출한 성문폐쇄시점을 기준으로 DyWT으로 검출한 성문폐쇄시점의 분포를 남·여 화자별로 나타낸 것으로 화자에 관계없이 검출된 성문폐쇄시점들이 실제의 성문폐쇄시점을 중심으로 좌·우로 퍼져 있음을 볼 수 있다. 표 2, 3은 각각 남, 여 화자에 대하여 각 스케일에서 특정 샘플 범위 내에서 DyWT 기반 성문폐쇄시점의 평균 검출율을 제시하고 있다. 화자별 검출 성문폐쇄시점의 정확성에 있어서는 남자 화자의 경우가 다소 나은 결과를 보였으며, 스케일별 경향에 있어서는 스케일 3, 4에서 검출된 성문폐쇄시점의 96% 정도가 ± 5 샘플내의 분포를 보임을 확인할 수 있다. 한편 스케일 5에서 성문폐쇄시점의 검출율이 스케일 3, 4에 비해 다소 떨어지는 것은 그림 5에서 확인했던 것처럼 웨이브렛 변환의 특성상 스케일이 증가할수록 신호의 저주파 성분이 강조되어져 웨이브렛 변환된 신호가 상대적으로 많이 스무딩되어 성문폐쇄시점에 해당하는 국부 최대값의 위치를 정확히 찾지 못하기 때문이다. 따라서 웨이브렛 변환으로 검출한 성문폐쇄시점을 이용한 피치검출 등에 있어서 이러한 점을 고려하여 음성 데이터의 샘플링 주파수에 따라 적절한 스케일에서 검출한 성문폐쇄시점을 이용함으로써 검출 성능을 향상시킬 수 있을 것이다.



(a). voice onset 구간에서의 스케일별 성문폐쇄시점 검출

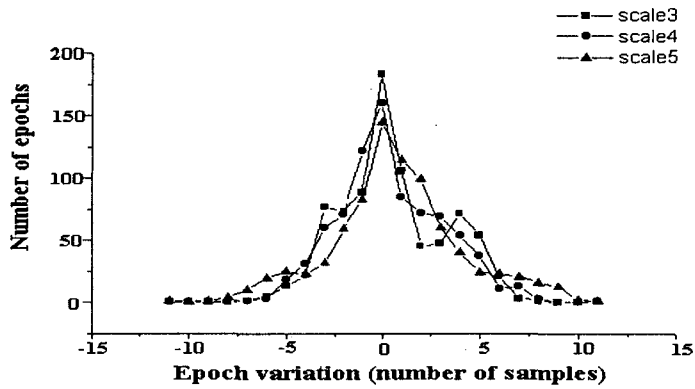


(b). 안정된 유성음 구간에서의 스케일별 성문폐쇄시점 검출

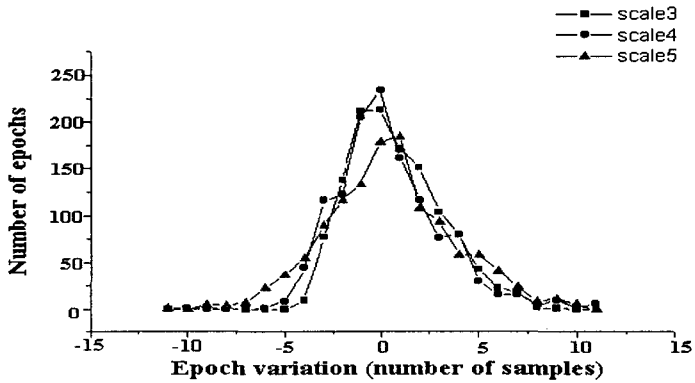


(c). voice offset 구간에서의 스케일별 성문폐쇄시점 검출

그림 5. 다양한 유성음 구간에 대한 성문폐쇄시점 검출



(a) 남자 화자



(b) 여자 화자

그림 6. DyWT으로 검출한 성문폐쇄시점의 화자별 분포

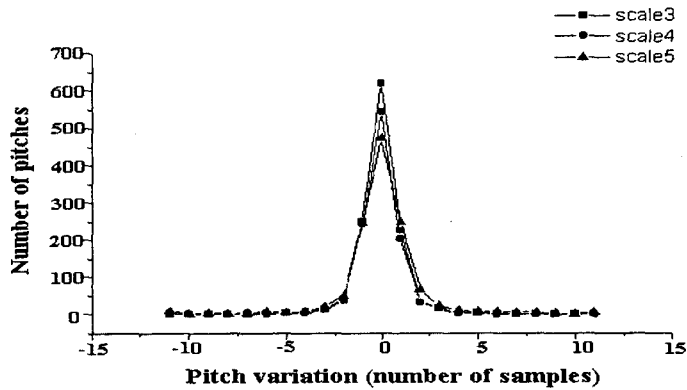
표 2. 각 스케일별 특정 샘플 범위 내에서의 성문폐쇄시점 검출률(남자 화자)

Tolerance \ Scale	±3 samples	±4 samples	±5 samples
Scale 3	85.7 (%)	92.9 (%)	96.4 (%)
Scale 4	82.4 (%)	92.4 (%)	95.5 (%)
Scale 5	72.8 (%)	81.8 (%)	89.5 (%)

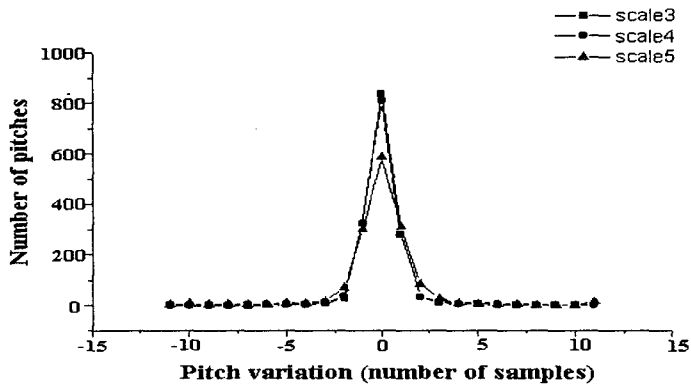
표 3. 각 스케일별 특정 샘플 범위 내에서의 성문폐쇄시점 검출률(여자 화자)

Tolerance \ Scale	±3 samples	±4 samples	±5 samples
Scale 3	76.3 (%)	87.8 (%)	96.1 (%)
Scale 4	78.7 (%)	89.2 (%)	95.9 (%)
Scale 5	72.8 (%)	80.5 (%)	86.4 (%)

그림 7에서는 EGG 신호를 이용하여 구한 피치주기와 DyWT으로 검출한 피치주기의 차이에 대한 분포를 화자별로 제시한 것이다. EGG 신호를 이용하여 구한 남자 화자들의 평균 피치주기는 84.5 샘플이고 여자 화자들의 평균 피치주기는 50 샘플이다. 이 결과로부터 DyWT으로 검출한 epoch을 이용하여 구한 음성신호의 피치주기는 음성신호의 실제 피치주기를 잘 반영하는 것으로 볼 수 있다[9].



(a) 남자 화자



(b) 여자 화자

그림 7. DyWT으로 검출한 피치와 실제 피치와의 차이

V. 결론

본 연구에서는 음성신호를 웨이브렛 변환한 신호에서 성문폐쇄시점에 해당하는 것으로 알려진 국부 최대값을 검출하여 그 결과를 EGG 신호를 이용하여 구한 실제의 성문폐쇄시점과 비교·분석하였다. 실험 결과를 통하여 음성신호의 웨이브렛 변환된 신호에서 검출한 국부 최대값은 DEGG 신호로부터 얻은 음성신호의 실제 성문폐쇄시점을 기준으로 좌·우로

퍼진 형태의 변화를 보이면서 음성신호의 성문폐쇄시점을 대체로 잘 반영함을 확인할 수 있었다. 웨이브렛 변환에 의해 검출된 성문폐쇄시점이 이러한 형태의 변화를 보이는 것은 음성신호 발생시 성도와 여기신호 사이의 coupling에 의한 것으로 생각된다. 특히 스케일 3, 4에서 검출된 성문폐쇄시점의 경우는 96% 정도가 ± 5 샘플내의 분포를 보였으며 이에 비해 스케일 5에서 검출된 성문폐쇄시점의 경우는 정확성이 다소 떨어져 나타났다. 이처럼 스케일이 증가함에 따라 성문폐쇄시점 검출의 정확성이 떨어지는 것은 웨이브렛 변환시 스케일 증가에 따른 신호의 스무딩 현상으로 성문폐쇄시점에 해당하는 국부 최대값의 위치를 정확히 검출하지 못했기 때문으로 판단된다.

웨이브렛 변환으로 검출한 성문폐쇄시점을 이용하여 구한 음성신호의 피치주기는 성문폐쇄시점 자체의 분포에 비해 상대적으로 안정적인 분포를 보이면서 실제 피치값을 잘 반영하는 것으로 나타났으며 특히 10 kHz로 샘플링된 음성신호의 경우 스케일 3에서 검출된 피치주기가 실제의 피치주기를 가장 잘 반영하는 것으로 확인되었다. 따라서 웨이브렛 변환을 이용하여 검출한 성문폐쇄시점을 음성신호의 피치검출에 이용할 경우 샘플링 주파수를 고려하여 적절한 스케일에서 검출되어진 성문폐쇄시점을 이용함으로써 피치검출의 성능을 향상시킬 수 있을 것으로 생각된다.

참 고 문 헌

- [1] Berg, J.W. 1958. van den, "Myoelastic-aerodynamic theory of voice production," *Speech and Hear. Res.*, vol. 1, 227-244.
- [2] 신무용, 김정철, 배건성. 1996. "Pitch Detection by the Analysis of Speech and EGG Signal" *한국음향학회지*, vol. 15, No. 5, 5-10.
- [3] A.K. Krishnamurthy, D.G. Childers. 1986. "Two-Channel Speech Analysis," *IEEE Trans. on ASSP*, vol. 34, No. 4, 730-743, Aug.
- [4] Kadambe, S. and Boudreaux-Bartels. 1992. "Application of the Wavelet Transform for Pitch Detection of Speech Signals," *IEEE Trans. on Information Theory*, Vol. 38, No. 2, 917-924, Mar.
- [5] Stephane Mallat and Sifen Zhong. 1992. "Characterization of Signals from Multiscale Edges," *IEEE Trans. on Pattern Analysis and Machine Intelligence*, vol. 14, No. 7, 710-732, July.
- [6] D.G. Childers and A.K. Krishnamurthy. 1991. "A critical review of electroglottography," *CRC Crit. Rev. Bioeng.*, vol. 12, No. 2, 136-164, Feb.
- [7] D.G. Childers, A.M. Smith, and G.P. Moore. 1989. "Relationships between electroglottography, speech and vocal cord contact," *Folida Phoniatica*, vol. 36, 105-108.
- [8] Olibier Rioul, Martin Vetterli. 1991. "Wavelets and Signal Processing", *IEEE SP Magazine*. 14-38, Oct.
- [9] 석종원, 손영호, 배건성. 1999. "웨이브렛 변환을 이용한 피치검출", *음성과학* 제5권, 제1호. 23-33.

접수일자: 2000. 7. 24.

게재결정: 2000. 9. 3.

▲ 배건성

대구광역시 북구 산격동 1370

경북대학교 전자·전기공학부 (우편번호: 702-701)

Tel : +82-53-950-5527, Fax : +82-53-950-5505

e-mail : ksbae@knu.ac.kr