웨이브렛 변환을 이용한 피치검출*

Pitch Detection Using Wavelet Transform

석 종 원** · 손 영 호** · 배 건 성**
(Jong Won Seok · Young Ho Son · Keun Sung Bae)

ABSTRACT

Mallat has shown that, with a proper choice of wavelet function, the local maxima of wavelet transformed signal indicate a sharp variation in the signal. Since the glottal closure causes sharp discontinuities in the speech signal, dyadic wavelet transform can be useful for detecting abrupt change in the voiced sounds, i.e., epochs. In this paper, we investigate the glottal closure instants obtained from the wavelet analysis of speech signal and compare them with those obtained from the EGG signal. Then, we detect pitch period of speech signal on the basis of these results. Experimental results demonstrated that local maxima of wavelet transformed signal give accurate estimation of epoch and pitch periods of voiced sound obtained by the proposed algorithm also correspond to those from EGG well.

Keyword : epoch, pitch detection, wavelet transform

I. 서 론

음성신호는 인간의 발성기관에서 2개의 얇은 막으로 구성된 성대(vocal folds)의 자발적인 운동에 의해 발생되는 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. 모음과 같은 유성음을 발성할 경우 허파에서 방출되는 공기는 닫혀진 성대에 의해 공기압이 점차 증가하여 성대가 떨어지기 시작하면서 좁은 공기 통로를 형성하게 되는데 이를 성문(glottis)이라고 하며, 성문을 통과하는 공기는 서로 다른 두 힘의 상호작용에 의해 성대가 규칙적으로 진동을 하도록 만든다. 이처럼 유성음에는 성문이 닫혀져 있는 부분과 열려져 있는 부분이 있게 되는데 성문이 닫히는 순간을 epoch이라 한다. 유성음을 발성할 때의 성대의

* 본 연구는 한국과학재단의 핵심전문연구비(과제번호: 971-0917-103-2) 지원으로 수행되었으며 지원에 감사드립니다.
** 경북대학교 전자 · 전기 공학부

단위시간당 진동횟수 즉, 기본주파수 또는 반복되는 진동운동의 기본주기 T_0를 피치라고 하는데, 음성합성, 음성인식, 화자인식, 화자검증, 피치동기 음성신호의 분석 및 합성 등 음성신호처리 분야에 있어서 매우 중요한 파라미터 중의 하나이다[1].

피치를 검출하는 방법에는 크게 event detection 방법과 non-event detection 방법이 있는데 이 때 event란 epoch을 의미한다. non-event detection 방법이 자기상관함수 또는 AMDF 등의 방법으로 평균적인 피치를 검출하는 방법인데 반해 event detection 방법은 먼저 epoch을 검출한 후 연속적인 epoch들 사이의 시간적 간격을 측정함으로써 피치를 검출하는 방법을 말한다. 최근에 음성신호를 웨이브렛 변환한 신호에서 local maxima가 실제 음성에서 급격한 변화를 나타내는 epoch에 해당된다는 결과가 발표되었으며 이를 이용한 피치 검출방법은 피치주기의 non- stationary한 변화부분에 대해서 뿐만 아니라 다양한 화자들 사이의 변화에 대하여서도 잘 대처할 수 있는 방법으로 알려져 있다[2-4]. 본 논문에서는 피치를 검출하는 방법으로서 Mallat이 제안한 quadratic spline 웨이브렛 함수를 이용하여 epoch을 검출하고 그 결과를 이용하여 음성의 피치를 구하였다. 이때 검출된 epoch은 epoch 검출에 있어서 기준신호로 적합한 것으로 알려진 EGG(Electroglottograph) 신호와 비교하여[6-8] 타당성을 조사하였으며, 그리고 이에 따른 피치검출 알고리즘을 연구하였다.

본 논문의 구성은 다음과 같다. 먼저 2장에서는 웨이브렛 변환에 대하여 설명하며 3장에서는 quadratic spline 웨이브렛 함수를 이용한 피치검출 방법에 대하여 설명한다. 그리고 4장에서는 3장에서의 방법으로 실험한 결과를 미분된 EGG(DEGG) 신호와 비교 설명하며 5장에서 결론을 맺는다.

II. 웨이브렛 변환

웨이브렛 변환은 응용수학에서 처음 소개된 후 최근 컴퓨터비젼 분야 등에서 연구되어 온 다중 해상도 표현과 연관성이 있음이 밝혀졌으며 이산 웨이브렛 변환 이론은 이산 신호의 subband 분할 방법과도 연관성이 존재한다. 그 중에서도 계수 구현을 더욱 용이하게 하는 dyadic 웨이브렛 변환(DyWT: Dyadic Wavelet Transform)은 식 (1)과 같고 이때 웨이브렛 함수는 식 (2)와 같이 정의된다.

$$W_2^{d,j} = \frac{1}{\sqrt{2^j}} \int f(t)\, \varphi^*\left(\frac{t}{2^j} - kT\right) dt \qquad (1)$$

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

식 (1)의 웨이브렛 변환 W_2^{dj}에서 d는 이산변환을 가리키며 j는 scale을 나타낸다.

웨이브렛 변환을 이용하여 신호를 분석하는 과정은 그림 1에서 보는 것과 같은 tree 형태의 필터뱅크로 생각할 수 있다. 여기서 H_0은 저역통과 필터이고, H_1은 고역통과 필터이다. 입력신호가 저역통과 필터와 고역통과 필터를 거치게 되면 한 번의 웨이브렛 변환이 이루어지며, 저역필터를 통과한 신호에 대해 이러한 과정을 반복적으로 수행하여 웨이

브렛 변환된 신호를 얻을 수 있다. 신호처리 관점에서 볼 때 DyWT는 constant-Q, octave band, 밴드패스 필터들의 뱅크 출력과 동일하다.

Mallat에 의해 제안된 quadratic spline 웨이브렛 함수는 식 (2)에서 정의된 웨이브렛 함수를 smoothing 함수 $\theta(t)$의 일차 미분으로 할 때 DyWT의 local maxima는 신호에 있어서 급격한 변화부분을 가리키고 local minima는 느린 변화를 나타낸다는 것을 입증하였다. 이 때 smoothing 함수란 어떤 함수의 Fourier 변환에서 저주파에 에너지가 몰려있는 함수를 말한다. 더욱이 Mallat은 시간 $t=t_0$에서 실제 신호가 갑작스런 변화를 하게될 때 $t=t_0$에서 연속적인 scale에 걸쳐서 DyWT한 값들은 local maxima를 가지게 된다는 것도 입증하였다[5].

본 논문에서 사용한 웨이브렛은 Mallat의 quadratic spline 웨이브렛으로서 사용된 필터의 계수는 표 1과 같다.

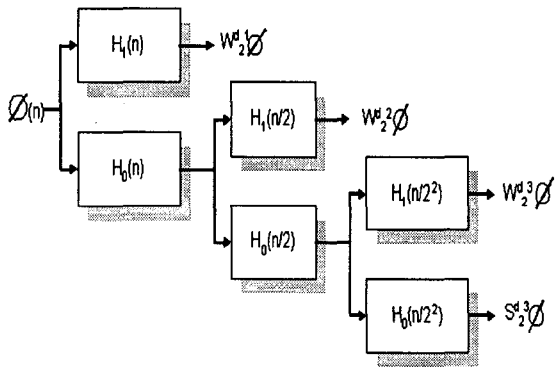


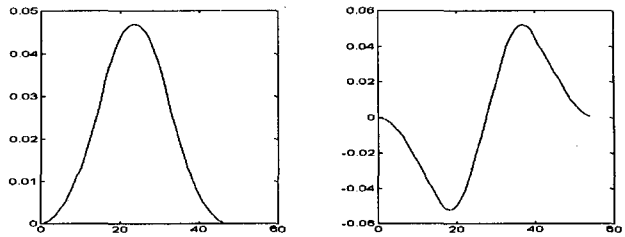
그림 1. 웨이브렛 분해 필터뱅크

표 1. quadratic spline의 필터 계수

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

그림 2는 smoothing 함수와 웨이브렛 함수의 관계를 보여주고 있다. 그림 2에서는 (a)의 smoothing 함수를 미분한 함수가 (b)의 웨이브렛 함수와 위상만이 반전된 동일한 함수임을 확인할 수 있다. 그러므로 epoch의 검출은 음의 값에서의 local maxima값이 검출되게 된다. 따라서,

본 논문에서의 local maxima와 global maxima는 모두 음의 방향의 값들이다.



(a) (b)

그림 2. 스케일 4에서 (a) smoothing 함수, (b) 웨이브렛 함수

III. Quadratic Spline을 이용한 피치 검출

Quadratic spline을 이용한 웨이브렛 변환에서는 일반적인 subband 분할 방식과는 달리 필터링된 신호의 크기를 그대로 유지하면서 다음 스케일 신호를 얻기 위해 필터링된 신호를 decimation하는 대신 필터계수들 사이에 0을 삽입하여 만든 웨이브렛 함수와 신호를 컨볼루션한다[5]. 앞에서 언급한 quadratic spline 웨이브렛 함수의 성질을 이용한 피치 검출 알고리즘은 아래와 같다.

STEP 1: 먼저 전체 음성을 스케일 3, 4, 5에서의 웨이브렛 함수를 가지고서 각각 웨이브렛 변환한다. 이때 웨이브렛 함수와 음성신호와의 convolution을 하게되는데, 각 스케일에서의 필터링으로 인한 delay(각 스케일에서의 웨이브렛 길이,W_L의 1/2)를 고려해주었다.

STEP 2: Step 1에서의 웨이브렛 변환한 신호를 분석하기 위하여 본 실험에서는 150 샘플씩 윈도윙하여 분석하도록 하였다. 먼저 해당 구간에 대하여 스케일 3에서의 global maxima를 구하여서 유성음과 무성음을 구분하는 threshold, T_u와 비교를 하여 작을 경우는 묵음이나 무성음 구간으로 간주하고 다음 구간으로 넘어간다. 이 때 threshold를 잡는 방법은 유성음 구간에서의 웨이브렛 변환한 값이 무성음이나 묵음 구간에 비해서 훨씬 크다는 것을 고려하여 적응적으로 잡도록 하였다.

STEP 3: Step 2에서 구간의 global maxima 값들이 T_u보다 클 경우에는 스케일 3과 4에서의 local maxima를 찾아서 스케일 3에서는 $T_3 \times$global maxima_3, 스케일 4에서는 $T_4 \times$global maxima_4를 넘어서는 local maxima값들을 찾아서 그 값들의 위치가 서로 같은지를 비교하고 스케일 5에서의 값도 epoch의 검출에 이용하였다. 이 때 위치의 어긋남이 10 sample 이내고 스케일5에서의 크기 조건을 만족하게 될 경우 일치하는 것으로 간주하고 값의 위치를 검출된 epoch으로 간주한다.

STEP 4: 한 프레임의 분석이 끝나면 다음 프레임으로 이동한다. 이 때 각 프레임은 10샘플 씩 겹치도록 하였다. 이것은 윈도우의 끝 부분에서 epoch에 대한 정보를 잃어버릴 경우의 영향을 보상해주기 위한 것이다. 그리고서는 다시 Step 1로 돌아가서 위의 단계를 반복한다.

STEP 5: Step 1-4에서 전 음성구간에 대해서 검출한 epoch을 가지고서 연속된 epoch들 사이의 시간적 간격을 측정함으로써 피치를 구한다.

유성음과 무성음을 구분하는 threshold, T_u는 알고리즘 자체에서 결정하도록 하였으며 T_3, T_4,는 각각 0.45와 0.5로 두었으며 epoch의 검출에서는 스케일 4에서의 신호를 기준으로 실험을 하였다.

IV. 실험 및 검토

웨이브렛 변환을 이용한 피치검출 실험을 위한 음성 데이터로는 남자 5, 여자 5명의 화자로부터 5종류의 문장을 대상으로 수집되어진 음성데이터를 이용하였다. 실험에 사용된 음성 데이터와 EGG 신호는 10 kHz 샘플링하고 16비트로 양자화하였으며, 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 3~5는 웨이브렛 변환으로 검출한 epoch과 실제의 epoch을 비교·분석하기 위하여 남성화자가 발성한 음성 데이터를 대상으로 음성신호, DEGG 신호, 그리고 스케일 3~5에 해당하는 웨이브렛 변환된 신호를 비교하여 나타내고 있다. 그림에서 수직방향의 점선은 DEGG 신호에서 음의 극대값의 위치를 나타내고 있다. 그림 3에서는 유성음의 시작 구간을, 그림 4에서는 안정된 유성음 구간을, 그리고 그림 5에서는 유성음이 끝나는 구간의 예를 각각 나타내고 있다. 이 결과에서 웨이브렛 변환으로 검출된 epoch은 실제의 epoch과 대체로 잘 일치하나 스케일이 증가할수록 변화가 커짐을 볼 수 있다. 그리고 안정된 유성음 구간에서 검출된 epoch이 유성음의 시작이나 끝 구간에서 검출된 epoch에 비해 상대적으로 실제의 epoch과 더 잘 일치하는 것을 확인할 수 있다.

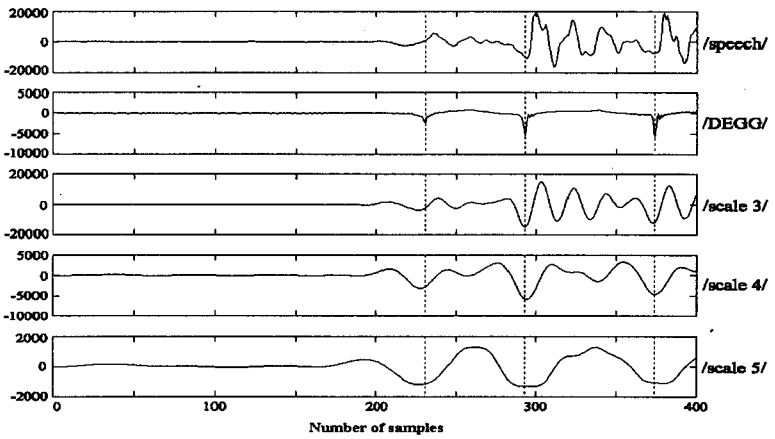


그림 3. Voice onset 구간에서의 스케일별 GCI 검출

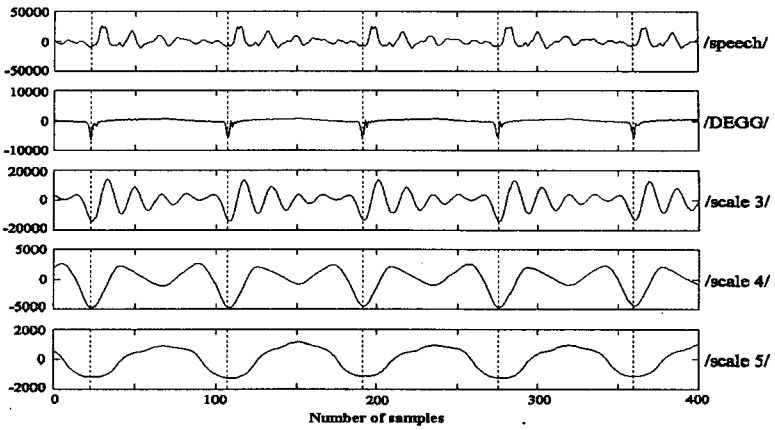


그림 4. 안정된 유성음 구간에서의 스케일별 GCI 검출

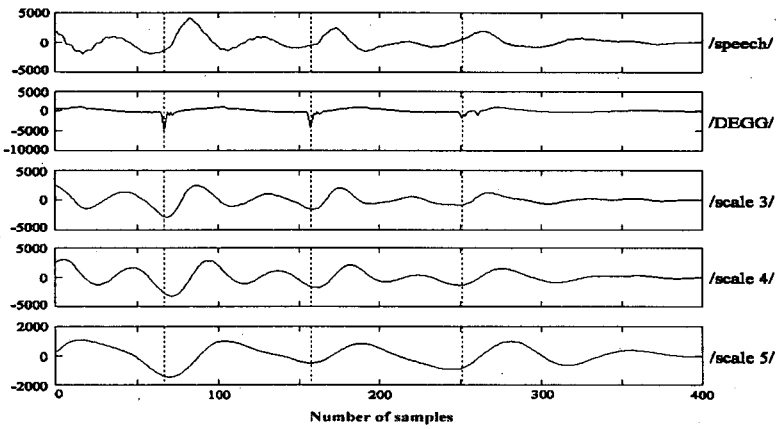


그림 5. Voice offset 구간에서의 스케일별 GCI 검출

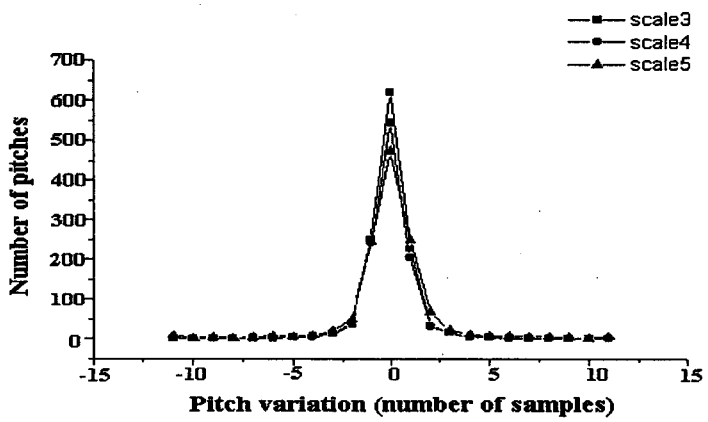
그림 6에서는 EGG 신호를 이용하여 구한 피치주기와 DyWT으로 검출한 피치주기의 차이에 대한 분포를 화자별로 제시한 것이다. 이 결과로부터 DyWT을 이용하여 구한 음성신호의 피치주기는 매우 안정된 결과를 보이고 있으며, 특히 화자에 따라 큰 변화없이 실제의 피치주기를 잘 반영함을 볼 수 있다. 표 2와 3에서는 EGG 신호를 이용하여 구한 피치주기를 기준으로 DyWT 방법으로 구한 피치주기의 차이를 특정 샘플값들에 대하여 화자별로 제시하고 있다. 이들 표에서는 피치주기의 차이를 절대적인 샘플값 외에도 화자들의 평균 피치주기에 대한 상대적인 양으로도 함께 제시하였다. 이때 EGG 신호를 이용하여 구한 남자 화자들의 평균 피치주기는 84.5 샘플이며, 여자 화자들의 평균 피치주기는 50 샘플이다. 이러한 결과들은 웨이브렛 변환을 이용한 피치주기의 검출시에 성문폐쇄 시점의 정확한 검출이 피치검출의 성능을 향상시킬 수 있음을 확인시켜준다.

그림 7은 음성신호에서 epoch이 non-stationary하게 변하는 구간의 예를 제시한 것이

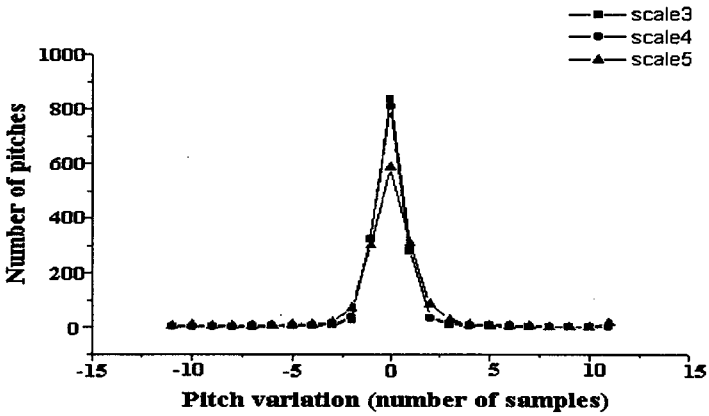
다. 웨이브렛 변환된 신호에서 국부 최대값은 이러한 구간에서도 음성신호의 성문폐쇄 시점을 잘 반영함을 볼 수 있다. 따라서 웨이브렛 변환을 이용하여 검출한 음성신호의 피치주기는 피치주기가 non-stationary하게 변하는 구간에서도 실제의 순시피치를 잘 반영할 수 있을 것으로 예상할 수 있다.

표 2. 특정 샘플 범위내에서의 피치 검출율(남자 화자)

Difference / Scale	±3 samples (0.36 %)	±4 samples (0.47 %)	±5 samples (0.59 %)
Scale 3	98.7 (%)	99.3 (%)	99.8 (%)
Scale 4	98.0 (%)	98.7 (%)	99.2 (%)
Scale 5	92.3 (%)	94.3 (%)	95.2 (%)



(a) 남자 화자



(b) 여자 화자

그림 6. DyWT으로 검출한 피치와 실제 피치와의 차이

표 3. 특정 샘플 범위내에서의 피치 검출율(여자 화자)

Difference / Scale	±3 samples (0.6 %)	±4 samples (0.8 %)	±5 samples (1 %)
Scale 3	97.9 (%)	98.6 (%)	99.2 (%)
Scale 4	97.7 (%)	98.4 (%)	99.0 (%)
Scale 5	93.4 (%)	94.6 (%)	95.7 (%)

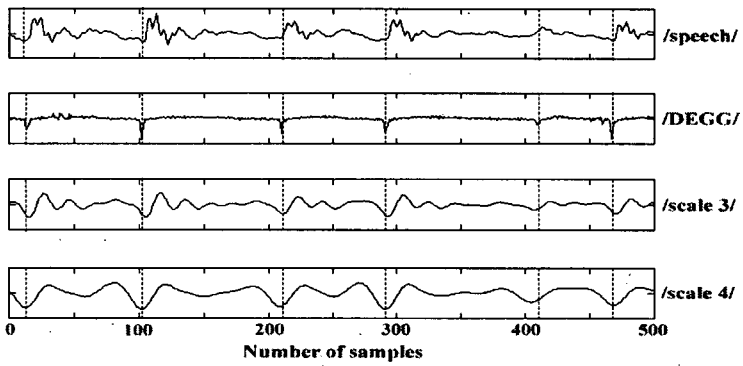


그림 7. 피치주기가 non-stationary한 구간의 예

그림 8은 남성화자가 발성한 음성신호를 DyWT한 파형을 관찰하여 검출한 epoch을 이용하여 구한 피치궤적과 EGG 신호를 이용하여 구한 실제의 피치궤적을 상호 비교하여 제시한 것이다. 이때 웨이브렛 변환으로 구한 피치궤적은 스케일 4에서 검출한 epoch을 이용하였다. 본 연구에서는 음성신호의 피치검출시에 스케일 4에서 검출한 epoch을 이용하였다. 이것은 피치주기의 정확성 측면에서는 10 kHz로 샘플링된 음성데이터의 경우 스케일 3에서의 피치주기가 가장 정확한 것으로 나타났으나 실제로 자동으로 검출할 경우 낮은 스케일에서 상대적으로 많이 발생하는 epoch의 검출 오류로 인한 성능의 저하를 고려하여 스케일 4에서 검출한 epoch을 이용하였다. 그림 9는 여성화자가 발성한 음성신호를 대상으로 본 연구에서 제안한 알고리즘을 이용하여 자동으로 검출한 피치궤적을 EGG 신호를 이용하여 구한 실제의 피치궤적과 비교하여 제시한 결과이다. 그림 8에서 DyWT을 이용하여 수동으로 epoch을 검출하여 피치궤적을 구한 경우와 마찬가지로 제안한 알고리즘을 이용하여 자동으로 검출하는 경우에도 실제의 피치궤적을 잘 반영함을 볼 수 있다. 그림 10에서는 문턱값 설정시에 대상으로 하지 않은 음성신호를 대상으로 제안한 알고리즘을 이용하여 자동으로 검출한 피치궤적을 제시한 것이다.

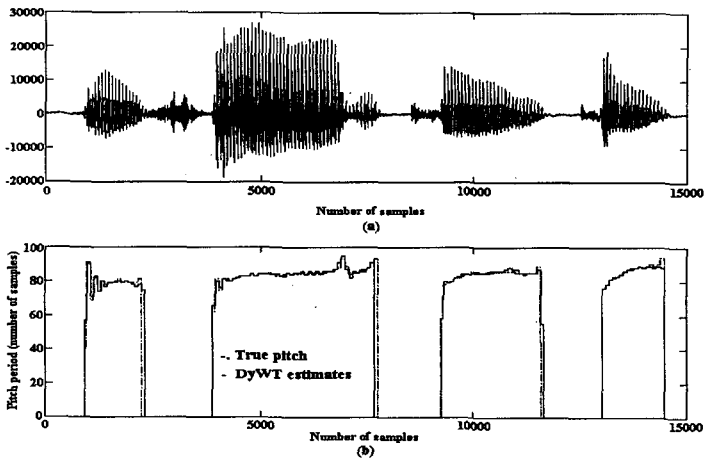


그림 8. DyWT 및 EGG 신호로 검출한 피치궤적
(a) 음성신호 /We saw the ten pink/, (b) 피치궤적

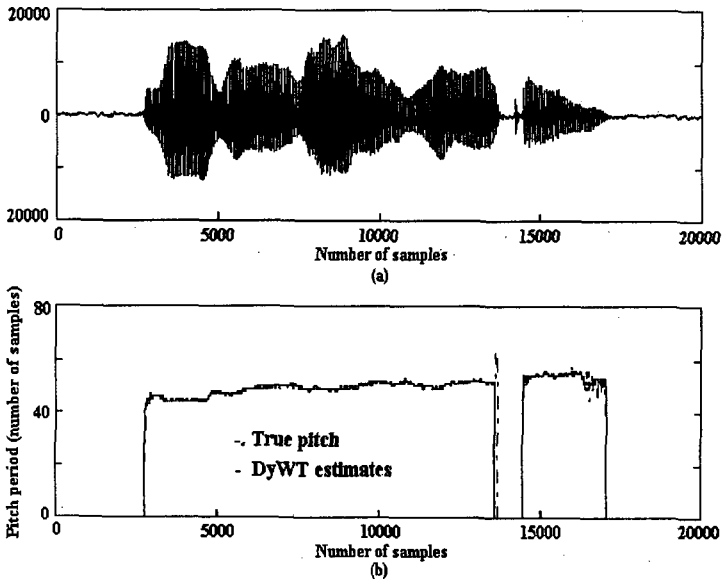


그림 9. DyWT을 이용하여 자동으로 검출한 피치궤적
(a) 음성신호 /We were away a year ago/, (b) 피치궤적

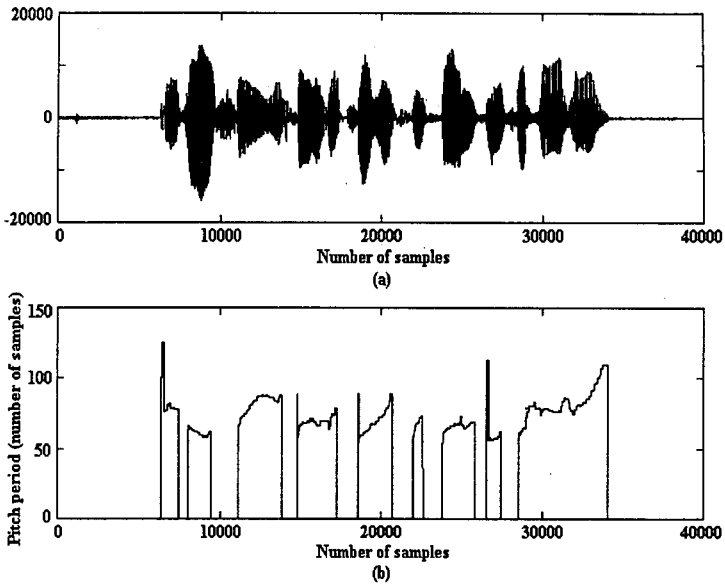


그림 10. DyWT을 이용하여 자동으로 검출한 피치궤적(계속)
(a) 음성신호 /대전시 유성 우체국 사서함 35호/, (b) 피치궤적

V. 결 론

본 논문에서는 Mallat의 quadratic spline 웨이브렛 함수를 이용한 음성신호의 웨이브렛 변환으로 epoch을 검출하고 검출된 epoch을 이용하여 피치검출 알고리즘을 제안하였다. 웨이브렛 변환으로 검출한 epoch을 이용하여 구한 음성신호의 피치주기는 안정적인 분포를 보이면서 실제 피치값을 잘 반영하는 것으로 나타났다. 따라서, 웨이브렛 변환을 이용한 음성신호의 피치주기 검출시에 샘플링 주파수에 따라서 적절한 스케일에서 검출되어진 epoch을 이용함으로써 피치검출 알고리즘의 성능을 향상시킬 수 있을 것으로 생각된다. 또 음성구간중 피치주기가 non-stationary하게 변하는 구간을 대상으로 한 경우나 여러 화자간의 변화 등에 있어서도 제안한 알고리즘이 잘 대처할 수 있음을 확인할 수 있었다.

웨이브렛 변환을 이용한 음성신호의 피치주기 검출에는 유성음 구간의 검출 및 국부 최대값의 검출시에 필요한 문턱값의 최적화 문제가 있다. 이것은 녹음 level이나 화자에 따른 변화 등으로 일정한 문턱값을 적용시키가 어렵기 때문이다. 따라서 정확한 피치주기를 검출하기 위하여 음성데이터에 따라 적응적으로 문턱값을 결정하는 방법과 또, 국부 최대값의 잘못된 검출로 인한 성문폐쇄 시점의 삽입이나 삭제 등의 경우에 오류를 보상해 주는 후처리방안에 대한 연구가 추가로 필요하다. 그리고 실제 음성신호 생성시 발생할 수 있는 다양한 잡음환경하에서도 정확한 피치검출을 위한 제안된 알고리즘의 최적화 및 성능평가 등에 대한 추가적인 연구가 필요할 것으로 본다.

참 고 문 헌

[1] 신무용, 김정철, 배건성 "2-채널(음성 및 EGG) 신호분석에 의한 피치검출." *한국음향학회*, Vol. 15(5), 5-10.

[2] Shubha Kadambe and G. Faye Boudreaux-Bartels, "Application of the Wavelet Transform for Pitch Detection of Speech Signals." *IEEE Trans Information Theory*, Vol. 38(2), 917-924.

[3] Glenn A. Shelby, "Tone detection using wavelet transforms." *SPIE*, Vol. 2491, 615-626.

[4] DU Limin, and HOU Ziqiang, "Determination of the Instants of Glottal Closure from Speech Wave Using Wavelet Transform." *ICSPAT*, vol.1

[5] S. Mallat and W. L. Hwang, "Singularity Detection and Processing with Wavelets." *IEEE Trans Information Theory*, Vol. 38(2), 617-643.

[6] D. G. Childers and A. K. Krishnamurthy, "A critical review of electroglottography." *CRC Crit. Rev. Bioeng.*, Vol. 12(2), 131-164.

[7] D. G. Childers, A. M. Smith, and G. P. Moore, "Relationship between electroglottography, speech and vocal cord." *Folia Phoniatrica*, Vol. 36, 105-108.

[8] A. M. Smith and D. G. Childers, "Laryngeal evalution using features from speech and the electrograph." *IEEE Trans. Biomed. Eng.*, Vol. 30, 755-759.

접수일자 : '99. 2. 11.
게재결정 : '99. 3. 15.

▲ 석 종 원
대구광역시 북구 산격동 1370
경북대학교 전자·전기공학부(우: 702-010)
Tel: (053) 940-8627, Fax: (053) 950-5505
e-mail: won@mmirl1.kyungpook.ac.kr

▲ 손 영 호
대구광역시 북구 산격동 1370
경북대학교 전자·전기공학부(우: 702-010)
Tel: (053) 940-8627, Fax: (053) 950-5505
e-mail: youngho@mmirl1.kyungpook.ac.kr

▲ 배 건 성
대구광역시 북구 산격동 1370
경북대학교 전자·전기공학부(우: 702-010)
Tel: (053) 950-5527, Fax: (053) 950-5505
e-mail: ksbae@ee.kyungpook.ac.kr