피치 검출을 위한 스펙트럼 평탄화 기법

김종국, 조왕래(*), 배명진

숭실대학교 정보통신공학과, 벽성대학(*)

Flattening Techniques for Pitch Detection

JongKuk KIM, WangRae JO(*), MyungJin BAE

Dept. of Information and Telecomm. Engr., Soongsil Univ. ByukSung College(*)

kokjk@hanmail.net

요 약

In speech signal processing, it is very important to detect the pitch exactly in speech recognition, synthesis and analysis. but, it is very difficult to pitch detection from speech signal because of formant and transition amplitude affect. therefore, in this paper, we proposed a pitch detection using the spectrum flattening techniques. Spectrum flattening is to eliminate the formant and transition amplitude affect. In time domain, positive center clipping is process in order to emphasize pitch period with a glottal component of removed vocal tract characteristic. And rough formant envelope is computed through peak-fitting spectrum of original speech signal in frequency domain. As a results, well get the flattened harmonics waveform with the algebra difference between spectrum of original speech signal and smoothed formant envelope. After all, we obtain residual signal which is removed vocal tract element. The performance was compared with LPC and Cepstrum, ACF. Owing to this algorithm, we have obtained the pitch information improved the accuracy of pitch detection and gross error rate is reduced in voice speech region and in transition region of changing the phoneme.

1. 서론

음성인식, 합성 및 분석과 같은 음성신호처리 분야에 있어서 기본주파수 즉, 피치를 정확히 검출하는 것은 중요하다. 만일 음성신호의 기본주파수를 정확히 검출할 수 있다면 음성인식에 있어서 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성합시에 자연성과 개성을 쉽게 변경하거나 유지할 수 있다. 또한 분석시 피치에 동기 시켜 분석하면 성문의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있다.

이러한 피치검출의 중요성 때문에 피치검출에 대한 방법들이 다양하게 제안되었으며 이러한 피치검출 방법은 시간영역법, 주파수영역법, 시간-주파수영역법으로 구분할 수 있다. 시간영역 검출법은 파형의 주기성을 강조한 후에 결정논리에 의해 피치를 검출하는 방법으로 병렬처리법, AMDF법, ACM법 등이 있다[1][2].

주파수영역의 피치검출법은 음성 스펙트럼의 고조파 간격을 측정하여 유성음의 기본주파수를 검출하는 방법으로 고조파분석법[3], Lifter법, Comb-filtering법 등이 제안되어 있다. 시간-주파수 혼성영역법은 시간영역법의 계산시간 절감과 피치의 정밀성, 그리고 주파수영역법의 배경잡음이나 음소변화에 대해서도 피치를 정확히 구할 수 있는 장점을 취한 것이다. 이러한 방법으로는 Cepstrum법, 스펙트럼비교법등이 있고, 이 방법은 시간과 주파수영역을 왕복할 때 오차가 가중되어 나타나므로 피치추출의 영향을 받을 수 있고, 또한 시간과 주파수영역을 동시에 적용하기 때문에 계산과정이 복잡하다는 단점이 있다[3][4].

따라서 본 논문에서는 정확한 피치검출을 위하여 음성신호의 주기성을 강조하기 위하여 시간 영역 처리와 주파수영역 처리를 통하여 스펙트럼 신호를 최대한 평탄화 시킴으로써 포만트의 영향을 제거하고 고조파 성분을 분리해 내는 새로운 피치검출 방법을 제안하고자 한다[8].

2. 피치검출시의 문제점

음성신호에서는 여파기 성분과 여기성분이 상호작용하기 때문에 피치검출이나 포만트 검출에 매우 어렵다. 특히 음성신호에 잡음이 부가될 경우에는 더욱 어려워진다. 따라서 낮은 SNR 조건하에서도 피치정보나 포만트 정보를 유지하는 스펙트럼 신호는 음성처리 분야에서 매우 중요하다고 할 수 있다. 그런데 스펙트럼은 고조파 성분과 포만트 성분이 함께 나타난다. 따라서 이를 잘 분리하는 것이 피치 검출이나 포만트 검출의 관건이라 하겠다. 음성신호의 피치주기를 정확하고 신뢰성있게 측정하는 것은 여러 가지 이유 때문에 아주 어렵다.

음성 파형에 대해 피치주기를 검출하려고 하면 성도 포만트에 따른 영향을 받게 된다. 성도포만트들은 문장을 구성하는 음소에 따라 변화하게 되고, 음소는 10ms 정도의 범위 내에서는 안정상태를 이룬다. 또한 피치검출시에 크게 영향을 주는 포만트들은 기본주파수에 근접한 제 1 포만트이고, 이 포만트의 에너지가 파형을 지배하기 때문에 이 성분을 적응적으로 제거 또는 억압시킬 필요가 있다. 주파수 영역에서 포만트의 영향을 제거하기 위해서는 스펙트럼에서 포만트 포락선을 제거하여 평탄화 시키면 된다.

3. 제안한 방법.

3.1). 시간영역처리

본 논문에서는 계산량을 줄이기 위해서 전처리과정으로 positive center clipping을 하였다. 피치검출과정을 수행하기 위해 먼저 중앙클리핑을 수행한다[1].

제안한 방법에서 먼저 시간영역처리에서 식(3.1)과 같이 음성파형의 기울기를 구하였다.

$$\text{기울기} = \frac{y_2 - y_1}{x_2 - x_1} \qquad (3.1)$$

x_1 = 첫 번째 프레임구간에서 최대값 위치

x_2 = 마지막 프레임구간에서 최대값 위치

y_1 = 첫 번째 프레임구간에서 샘플의 최대값

y_2 = 마지막 프레임구간에서 샘플의 최대값

센터 클리핑을 수행하는 이유는 성도성분을 제거하여 성문특성을 강조하여서 정확하게 피치를 검출하기 위해서 사용된다.

3.2). 주파수영역처리

주파수영역처리에서 두 신호의 하모닉스 스펙트럼을 peak-fitting을 하기 위하여 원래의 음성신호와 positive center clipping한 신호를 식(3.2)와 같이 DFT하였다.

$$S(k) = \sum_{n=0}^{N-1} s(n)\, e^{-j\frac{2\pi}{N}kn} \qquad 0 \le n \le N-1 \qquad (3.2)$$

평탄화된 센터 클리핑한 신호의 스펙트럼을 원 신호의 스펙트럼과 fitting하기 위하여 프레임별 구간에서 최대의 peak점과 그 점의 위치를 찾아서 피치주기를 결정하였다. 검색구간 검출방법은 프레임내의 피치간격 즉 처음 두 샘플간의 피크점을 밴드사이즈로 정하고 4샘플이나 5샘플의 간격으로 전 프레임에 대하여 검색한다. 또한 밴드사이즈의 반으로 중첩해서 피치간격을 검색하였다. 밴드사이즈의 두 샘플의 피크를 찾아서 그 점의 최대값과 위치를 결정한다.

피크점들은 샘플들의 기울기가 증가하면 +1, 감소하면 -1로 결정을 하고 그 기울기를 더해서 0이 되면 피크점으로 결정하였다. 결정된 피크값과 그 위치를 원 신호의 스펙트럼과 fitting을 한다. peak-fitting된 원 신호의 스펙트럼상에서의 peak점들을 가지고 전 프레임의 peak점들을 연결하고 원 신호의 포만트 포락선과 동일하게 만들기 위하여 선형 균등 인터폴레이션을 수행하였다. 결과적으로 스무딩된 포만트 포락선을 구할 수 있었다.

주파수영역에서 LPC 역필터링의 결과와 같이 원 신호의 스펙트럼과 스무딩된 포만트 포락선과의 대수차에 의하여 평탄화된 하모닉스 스펙트럼을 얻었다. 식(3.3)같이 역푸리에변환을 하면 잔차신호를 구할 수 있으며 이 신호는 식(3.4)을 이용하면 자기상관신호를 얻을 수 있다.

$$x(k) = \frac{1}{N}\sum_{n=-N/2}^{N/2-1} X(n) e^{j2\pi nk/N} \quad (k = k_0, \ldots, k_0 + N - 1) \qquad (3.3)$$

$$\phi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \qquad (3.4)$$

어떤 신호가 주기가 P인 주기신호라면 $\phi(k)$ 역시 주기를 갖고, $\phi(k)$의 최대값들은 $k = 0, \pm P, \pm 2P, \ldots$ 의 피치주기에서 발생한다. 본 논문에서 제안한 방법의 피치검출과정을 그림 1의 블럭도로 나타내었다. 결과적으로 본 논문에서 제안한 피치를 얻기 위하여 평탄화된 하모닉스 스펙트럼을 IFFT수행하여 잔차신호를 얻을 수 있었다. 제안한 방법의 피치 검출 과정을 그림 1의 블록도에 나타내었다.

4. 실험 및 결과

제안한 방법의 과정을 컴퓨터 시뮬레이션하기 위하여 사용한 장비는 IBM-PC(Pentium-Ⅲ) 시스템이며 여기에 음성신호를 입출력하기 위한 상용화된 16비트 AD/DA변환기를 인터페이스하여 아래의 문장들을 남녀 각 10명에게 발성시키면서 8kHz의 표본율로 데이터를 입력하였다. 제안한 방법을 구현하기 위해서 C-언어와 MATLAB으로 구현하여 수행하였다.

발성 1 : /인수네 꼬마는 천재소년을 좋아한다./
발성 2 : /예수님께서 천지창조의 교훈을 말씀하셨다./
발성 3 : /창공을 헤쳐 나가는 인간의 도전은 끝이없다./
발성 4 : /숭실대학교 정보통신과 음성통신 연구팀이다./

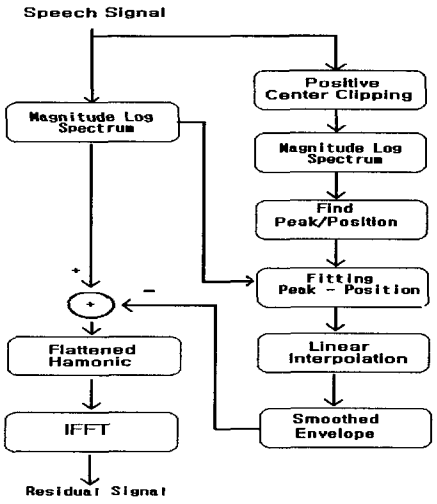

그림 1. 제안한 알고리즘의 블럭도

위의 각 음성시료에 대해 한 프레임의 길이를 256 샘플로 하여 3개의 Subframe 나누어 첫 번째와 마지막 Subframe의 최대값을 찾아 음성파형의 기울기를 구하고 주기성을 강조하기 위해 positive성분의 피크값을 찾아서 센터 클리핑을 수행하였다. 또한 센터 클리핑된 신호를 주파수 영역으로 변환 후 스펙트럼의 peak값과 위치를 찾은 다음 원 신호의 스펙트럼과의 peak-fitting을 통해 대략적인 포만트 포락선을 구했다. 그리고 대략적인 포만트 포락선을 선형 인터폴레이션을 수행하여 스무딩된 포만트 포락선을 구할 수 있었다[9]. 따라서 원 음성신호의 스펙트럼과 스무딩된 포만트 포락선의 대수차에 의해서 평탄화된 하모닉스 스펙트럼을 구할 수 있었으며 이 스펙트럼을 IFFT수행하여 성도성분이 제거된 잔차신호를 구할 수 있었다[10].

그림 2와 그림 3은 본 논문에서 제안한 유성음/전이구간에서의 출력파형을 보여 주고 있다. 그림 2와 그림 3의(b)는 그림 2와 그림 3의(d)를 평탄화된 하모닉스 스펙트럼을 IFFT수행하여 구한 잔차신호이며 피치구간별 최대 peak를 잘 나타내고 있다. 그림 2와 그림 3의(c)는 원 음성신호의 스펙트럼과 positive center clipping을 한 스펙트럼과의 peak-fitting을 하여 대략적인 포만트 포락선을 선형 인터폴레이션을 수행하여 구한 스무딩된 포만트 포락선이다. 또한 그림 2와 그림 3의(d)는 원 음성신호의 스펙트럼과 그림 2와 그림 3의 (c)와의 대수차에 의해 생성된 평탄화된 하모닉스 스펙트럼을 나타내고 있다[5].

그림 4는 본 논문에서 제안한 피치검출 성능을 평가하기 위하여 발성1에 대하여 LPC, Cepstrum과 비교하여 피치변화도를 나타내었다. 전체 음성신호의 각 프레임에 대하여 기준 피치를 정한 다음 제안한 방법과 LPC, Cepstrum와의 프레임별 피치변화도를 나타내었다.

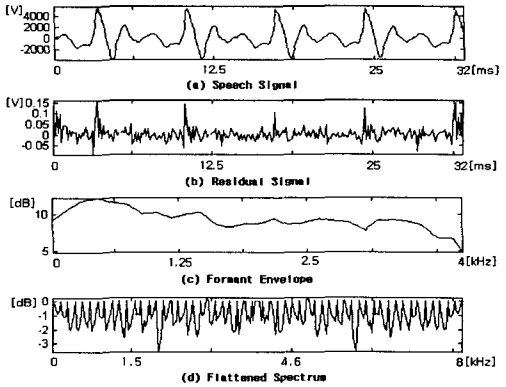

그림 2.. 제안한 알고리즘의 출력(유성음구간)

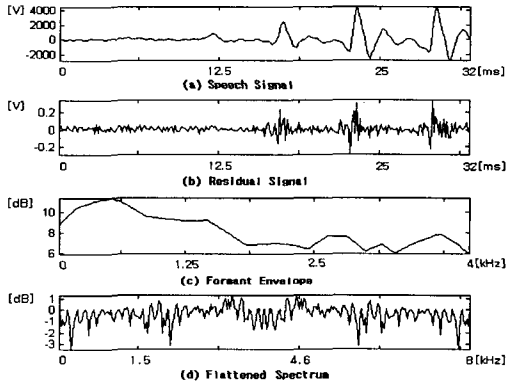

그림 3. 제안한 알고리즘의 출력(전이구간)

표 1은 발성1, 2에 대하여 제안한 방법과 LPC, Cepstrum과의 gross error rate를 나타내었다. gross error rate에서 음성신호는 clean speech이며 전체 음성프레임을 한 프레임씩의 피치에러를 구하여 기존의 피치검출 방법인 LPC, Cepstrum보다 성능이 향상되었음을 알 수 있다. 기존의 방법보다 gross error rate가 LPC는 1.54%, Cepstrum은 1.93%가 감소하였다.

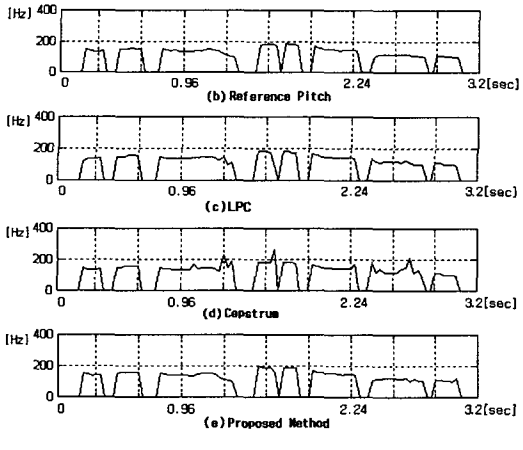

그림 4. LPC/Cepstrum에 대한 발성1의 피치변화도

표 2. LPC/Cepstrum에 대한 gross error rate 비교

음성 시료	분석 프레임수	gross error rate(%)		
		LPC	Cepstrum	Proposed
발성1	100	2.35	2.65	0.58
발성2	120	1.85	2.34	0.55
평균		2.1	2.49	0.56

5. 결 론

음성신호처리분야에서 피치를 정확히 검출하면 음성 인식시에 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성합성시에 자연성과 개성을 유지하거나 쉽게 변경할 수 있다. 또한 분석시 피치에 동기시켜 분석하면 성문의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있게 된다[1].

본 논문에서 제안한 피치검출방법은 기존의 방법들에 비하여 시간-주파수 변환에 따른 계산량 증가와 알고리즘의 복잡성이 있지만 피치성능을 비교하면 기존의 방법들보다 성능면에서는 향상되었다. 특히 음성이 변하는 전이구간에서 피치를 더 잘 찾을 수 있었다. 기존의 피치검출 알고리즘인 LPC, Cestrum과의 비교에서 gross error rate가 1.91%, LPC는 1.54%, Cestrum은 1.93%이 감소하였다. 결과적으로 본 논문에서의 주된 목적은 기존의 피치 검출 알고리즘에 관하여 처리시간과 복잡성보다는 피치검출 성능을 개선하고 향상하는데 목적을 두었다. 따라서 시간영역과 주파수 영역의 변환으로 인한 피치주기 검색 에러율을 최소화하였고 음성이 변하는 전이구간에서의 피치검출 방법을 개선하였다. 따라서 향후과제로는 또한 음성합성에서 중요한 피치시점 검출, 피치변경, 지속시간 변경에 적용할 수 있는 피치검출기를 구현해야 할 것이다.

5. 참고 문헌

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech signals*, Englewood Cliffs, Prentice-Hall, New Jersey, 1978.

[2] P. E. Paparnichalis, *Practical Speech Processing* Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1987.

[3] S. Seneff, "Real Time Harmonic Pitch Detection," IEEE Trans. Acoust. Speech, and Signal Processing, Vol. ASSP-26, pp. 358-365, Aug. 1978.

[4] M. Lee, C. Park, M. Bae, and S. Ann "The High Speed Pitch Extraction of Speech Signals Using the Area Comparison Method," KIEE, Korea, Vol. 22, No. 2, pp.13-17, March 1985.

[5] M. Lee, C. Park, M. Bae, and S. Ann "The High Speed Pitch Extraction of Speech Signals Using the Area Comparison Method," KIEE, Korea, Vol. 22, No. 2, pp.13-17, March 1985.

[6] M. Bae, J. Rheem, and S. Ann "A Study on Energy Using G-peak from the Speech Production Model," KIEE, Korea, Vol. 24, No. 3, pp. 381-386, May 1987.

[7] Hans Werner Strube , "Determination of the instant of glottal closure from the speech wave," J., Acoust., Soc., Am, Vol. 5, No. 5, pp. 1625-1629, November 1974.

[8] M. Bae, I. Chung, and S. Ann, "The Extraction of Nasal Sound Using G-peak in Continued Speech," KIEE, Korea, Vol. 24, No. 2 pp. 274-279, March 1987.

[9] 한진희, 장세현, 배명진, 김상룡, 김명제, "전처리된 가변 대역폭 LPF에 의한 피치검출법," 한국음향학회, 제 12 회 음성 통신 및 신호처리 워크샵논문집, Vol.SCAS-12, No.1, PP.221-224, 1995년 06월.

[10] 이을재, 박찬수, 배명진, 안수길, "시간-주파수 혼성형 피치검출에 관한 결정논리", 대한전자공학회, 추계학술발표논문집, Vol. 13, No. 2, 1992.