

시간-주파수영역에서의 새로운 피치검출 방법

김종국, 최호진, 배명진

승실대학교 정보통신공학과

A New Pitch Detection Method in time-Frequency Domain

JongKuk KIM, HoJin CHOI, MyungJin BAE

Dept. of Information and Telecomm. Engr., Soongsil Univ. Seoul 156-743, Korea

kokjk@hanmail.net

요 약

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, 음성합성 시에 자연성과 개성을 쉽게 변경하거나 유지할 수 있다. 또한 분석시 피치에 동기 시켜 분석하면 성문의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있다. 따라서 제안한 논문에서는 시간영역처리에서 시간영역파형에서 먼저 성도성분이 제거된 성문특성 즉 피치주기 성분을 강조하기 위해 음성신호의 기울기를 이용한 Positive 센터 클리핑을 수행하고 주파수 영역에서는 원 신호의 스펙트럼과 센터클립된 신호의 스펙트럼과의 Peak-Fitting을 수행하고 선형 인터폴레이션(스무딩)을 통해 평탄화된 스펙트럼을 얻었다. 결과적으로 유성음 구간과 음성이 변하는 전이구간에서 G-peak가 강조된 더욱 정확한 Pitch를 검출할 수 있었다.

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, 음성합성에

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 병렬 처리법, AMDF법, ACM법 등이 있다. 시간영역 검출법은 보통 시간영역에서 수행되므로 영역의 변환이 불필요하고 합, 차, 비교논리 등 간단한 연산만 필요하게 된다. 그러나, 음소가 천이구간에 걸쳐 있는 경우에는 프레임 내의 레벨변화가 심하고 피치주기가 변동하기 때문에 피치검출이 어렵다. 특히 잡음이 섞인 음성의 경우에는 피치검출을 위한 결정논리가 복잡해져서 검출 오류가 증가되는 단점이 있다[1][2].

주파수영역의 피치검출법은 음성 스펙트럼의 고조파 간격을 측정하여 유성음의 기본주파수를 검출하는 방법으로 고조파분석법[3], Lifter법, Comb-filtering법 등이 제안되어 있다. 일반적으로 스펙트럼은 한 프레임(20-40ms) 단위로 구해지므로, 이 구간에서 음소의 진이나 변동이 일어나거나 배경잡음이 발생하여도 평탄화되므로 그 영향을 적게 받는다. 그러나 처리 과정상 주파수영역으로의 변환과정이 필요함으로 계산이 복잡하며, 기본주파수의 정밀성을 높이기 위해 FFT의 포인트 수를 늘리면 그만큼 처리시간이 길어지고 변화특성에 분

해지게 된다.

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 누적되어 나타나므로 피치추출의 영향을 받을 수 있고, 또한 시간과 주파수영역을 동시에 적용하기 때문에 계산과정이 복잡하다는 단점이 있다[3][4].

따라서 본 논문에서는 시간영역과형에서 먼저 성도성분이 제거된 성분특성 즉 피치주기 성분을 강조하기 위해 음성신호의 기울기를 이용한 Positive 센터 클리핑을 수행하고 주파수 영역에서는 원신호의 스펙트럼과 센터 클리핑된 신호의 스펙트럼과의 Peak-Fitting을 수행하고 선형 인터폴레이션(스무딩)을 통해 평탄화된 스펙트럼을 얻었다. 결과적으로 유성을 구간과 음성이 변하는 전이 구간에서 G-peak가 강조된 더욱 정확한 피치를 검출할 수 있는 새로운 알고리즘을 제안하였다.

2. 피치검출시의 문제점

음성신호의 피치주기를 정확하고 신뢰성있게 측정하는 것은 여러 가지 이유 때문에 아주 어렵다. 첫번째 이유는 성문의 여기과형이 완전히 주기적인 파형이 아니라는 것이다. 한 주기내에 있는 파형의 세부구조와 함께 변화하는 음성파형의 주기를 측정하는 것은 어려운 작업이다. 두 번째는 피치주기를 측정하는데 있어서 어떤 경우에는 성도의 포먼트가 성분파형의 구조를 완전히 바꿀 수 있기 때문에 피치주기를 검출하기가 어렵게 된다. 일반적인 상호작용은 조음기관이 빠르게 변화하는 경우이며 피치검출시에는 방해요소로 나타나게 된다.

세 번째 문제점에는 유성음구간 동안 각 피치주기의 정확한 시작과 끝을 정의하기 어렵다는 것이다. 피치주기의 정확한 시작과 끝의 위치를 선택하는 것은 임의적일 수 있기 때문이다. 예를 들면, 음성파형에 근거하여 주기의 시작과 끝을 정의하기 위한 필요한 것은 각 피치주기의 시작과 끝점을 정의하기 위한 후보로는 주기내의 영교차 간격을 들 수 있다. 이러한 측정에서 필요한 것은 각 피치주기의 시작과 끝점의 정확한 위치를 정의하기 위해서는 주기에서 주기까지 일관성이 있어야 한다는 것이다. 이러한 일관성이 부족하면 잘못된 피치주기를 측정하게 된다. 음성파형의 주기성분을 영교차 단위로 검출하면 포먼트 구조에 민감하게 된다.

피치를 검출할 때 나타나는 네 번째 문제점은 무성음과 낮은 레벨의 유성음을 구별해야 하는 것이다. 많은 경우에 무성음 구간과 레벨이 낮은 유성음 구간의 변화는 별로 뚜렷하지 않기 때문에 정확한 위치를 찾기가 어렵다.

다섯 번째 문제점은 전화시스템을 통하여 전송되는

음성의 피치검출 문제에 직면할 때에 부가적인 복잡성이 발생하게 된다. 음성에 대한 전화시스템의 영향은 선형, 비선형처리를 포함하면서 음성신호에 잡음이 추가되는 경우이다. 전화시스템을 선형여파기로 간주하면 기본주파수와 고조파들의 성분을 약하게 만들 수 있는 내역통과 여파기로 볼 수 있다. 이 때문에 전화음질에 대한 신호의 주기성을 검출하기가 더 어렵게 된다. 음성신호에 대한 전화시스템의 비선형특성은 전송시스템의 채널에 따라 좌우된다.

3. 제안한 방법.

3.1. 시간영역처리

3.1.1) 센터 클리핑(Center Clipping)

피치검출과정을 수행하기 위해 먼저 아래같이 음성파형의 기울기를 구하였다.

$$\text{기울기} = \frac{y_2 - y_1}{x_2 - x_1} \quad (3-1)$$

x_1 = 첫 번째 프레임구간의 샘플 위치

x_2 = 마지막 프레임구간의 샘플 위치

y_1 = 첫 번째 프레임구간에서 샘플의 최대값

y_2 = 마지막 프레임구간에서 샘플의 최대값

센터 클리핑을 수행하는 이유는 성도성분을 제거하여 성분특성을 강조하여서 정확하게 피치를 검출하기 위해서 사용된다. 제안한 논문에서는 그림 3-1에서와 같이 먼저 계산량 감소와 피치주기를 용이하게 찾기 위하여 주파수 영역에서 좀 더 Flattened한 스펙트럼을 위해 음성신호의 기울기를 측정하여 Positive 센터 클리핑을 하였다.

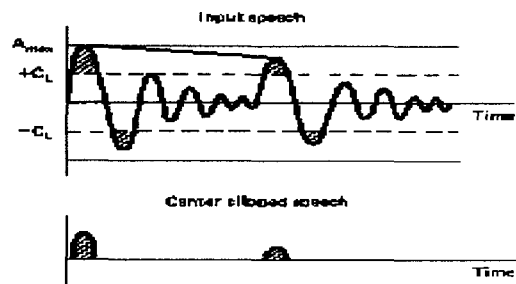


그림 3-1. 제안한 알고리즘의 센터 클리핑

3.2. 주파수영역처리

3.2.1) 하모닉스 Peak-Fitting

주파수영역에서 두신호의 스펙트럼을 Peak-Fitting을 하기 위하여 원래의 음성신호와 Positive 센터 클리

핑한 신호를 FFT하였다. Flattened된 센터클리핑된 신호의 스펙트럼을 원신호의 스펙트럼과 Fiting하기 위하여 프레임별 구간에서 최대의 Peak점과 그 점의 위치를 찾아서 피치주기를 결정한다. 결과적으로 샘플의 크기와 위치를 원 신호의 스펙트럼에 Fiting하였다.

3.2.2) G-peak 검출

Peak-Fiting된 원 신호의 스펙트럼상에서의 Peak점들을 가지고 전 프레임의 Peak점들을 연결하고 원 신호의 포먼트 포락선과 동일하게 만들기 위하여 선형 균등 인터플레이션을 수행하였다. 락선을 만들기 위하여 그 점들을 연결하여 선형 균등 인터플레이션을 수행하였다. 그 결과 스무딩된 포먼트 포락선을 구할 수 있었다. 그리고 원 신호의 스펙트럼과 스무딩된 포먼트 포락선의 대수차에 의하여 평탄화된 하모닉파형을 얻었다. 따라서 본 논문에서 제안한 피치를 얻기위하여 평탄화된 하모닉스파형을 IFFT수행하여 G-peak가 강조된 신호를 얻을 수 있었다.

제안한 방법의 피치검출과정을 그림3-2의 블록도로 나타내었다.

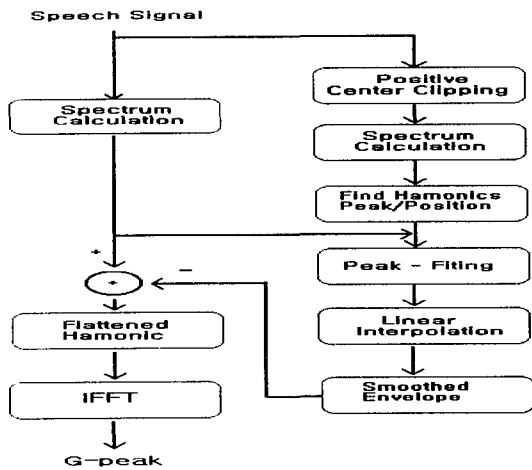


그림 3-2. 제안한 알고리즘의 블록도

4. 실험 및 결과

이상의 과정을 컴퓨터 시뮬레이션하기 위하여 IBM 펜티엄III(586)에 마이크가 부착된 16-비트 A/D변환기를 인터페이스시키고, 아래의 문장들을 남녀 각 3명에 게 발생시키면서 8kHz의 표본화 주파수로 표본화하여 저장한 다음에 시뮬레이션의 시료로 사용하였다:

- 발성1) "인수네 꼬마는 천재소년을 좋아한다."
- 발성2) "승실대 정보통신과 음성통신 연구팀이다."
- 발성3) "창공을 날으는 인간의 도전은 끝이없다."

위의 각 음성시료에 대해 한 프레임의 길이를 360 샘플로 하여 120샘플 단위로 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-Fiting을 통해 Rough한 포먼트 포락선을 구했다. 그리고 Rough한 포먼트 포락선을 선형 인터플레이션을 수행하여 스무딩된 포먼트 포락선을 구할 수 있었다. 따라서 원 신호의 스펙트럼과 스무딩된 포먼트 포락선의 대수차에 의하여 Flattened된 Harmonics파형을 구할 수 있었으며 이 파형을 IFFT수행하여 성도성분이 제거된 G-peak 성분을 구할 수 있었다. 또한 성도비교형가는 캡스טר엄방법과 LPC로 비교하였다

아래의 그림은 제안한 알고리즘의 출력파형을 그림 4-1과 4-2에 나타내었다

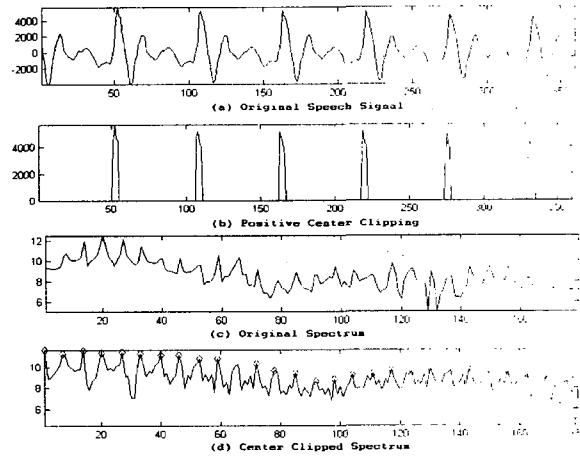


그림 4-1. 제안한 알고리즘 출력파형1

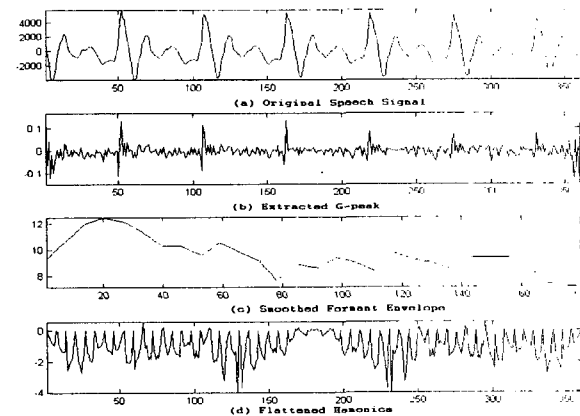


그림 4-2. 제안한 알고리즘 출력파형2

또한 기존의 피치검출방법과 새로이 제안한 알고리즘과의 피치검출 성능비교를 그림 4-3에 나타내었다.

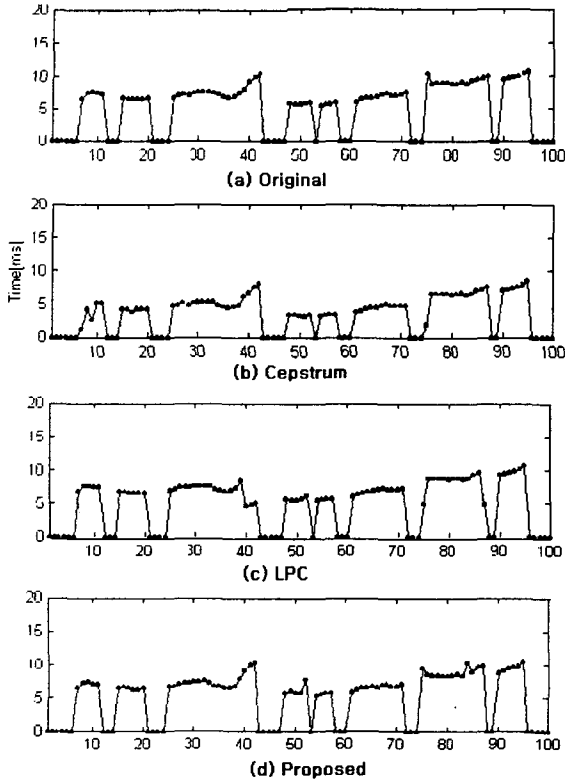


그림 4-3. 피치성능비교(contour)

5. 결 론

음성신호처리분야에서 피치를 정확히 검출하면 음성인식에 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.

제안한 방법은 시간영역처리에서 시간영역파형에서 먼저 성도성분이 제거된 성문특성 즉 피치주기 성분을 강조하기 위해 음성신호의 기울기를 이용한 Positive 센터 클리핑을 수행하고 주파수 영역에서는 원신호의 스펙트럼과 센터클리핑된 신호의 스펙트럼과의 Peak-Fitting을 수행하고 선형 인터플레이션(스무딩)을 통해 평탄화된 스펙트럼을 얻었다. 결과적으로 유성음 구간과 음성이 변하는 전이구간에서 G-peak가 강조된 더욱 정확한 Pitch를 검출할 수 있었다. 더불어 기존의 피치검출 알고리즘과 비교해서 계산량과 알고리즘의 복잡도가 단점이

지만 피치검출 성능평가에서 더 우수한 피치를 검출할 수 있었다.

6. 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] S. D. Stearns & R.A. David, *Signal Processing Algorithms*, Prentice-Hall, Inc, Englewood Cliffs, New-Jersey, 1988.
- [5] M. Bae, and S. Ann, "Fundamental Frequency Estimation of Noise Corrupted Speech Signals Using the Spectrum Comparison," *J., Acoust., Soc., Korea*, Vol. 8, No. 3, June 1989.
- [6] 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.
- [7] 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.
- [8] 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.
- [9] 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.
- [10] 한진희, 장세현, 배명진, 김상룡, 김병제, "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.