한국음향학회지 제21권 제3호 pp. 310~314 (2002)

개선된 피치검출을 위한 스펙트럼 평탄화 기법에 관한 연구

A Study on the Technique of Spectrum Flattening for Improved Pitch Detection

강 은 영*, 민 소 연**, 배 명 진*
(Eun-Young KANG*, So-Yeon MIN**, Myung-Jin BAE*)

*숭실대학교 정보통신공학과 대학원, **숭실대학교 전자공학과 대학원
(접수일자: 2001년 11월 29일; 채택일자: 2002년 2월 28일)

음성인식, 합성 및 분석과 같은 음성신호처리 분야에 있어서 기본주파수 즉, 피치를 정확히 검출하는 것은 중요하다. 그러나 포만트의 영향과 천이진폭의 영향 때문에 음성신호로부터 정확한 피치검출은 매우 어렵다. 따라서 본 논문에서는 음소의 천이나 변동의 영향이 적은 주파수 영역에서 스펙트럼을 평탄화 함으로써 포만트의 영향을 제거한 후 피치를 검출한다. 본 논문에서는 새로운 스펙트럼 평탄화 기법을 제안하고 기존의 방법인 LPC법, 켑스트럼법과 비교하여 어느 정도의 우수성을 보이는지 평가하였다. 또한 각각의 방법을 적용하여 기본주파수 (피치)를 검출한 결과는 제안한 방법이 우수함을 보여주고 있다.
핵심용어: 피치검출, 평탄화기법, 자기상관법
투고분야: 음성처리 분야 (2.4)

The exact pitch (fundamental frequency) extraction is important in speech signal processing like speech recognition, speech analysis and synthesis. However the exact pitch extraction from speech signal is very difficult due to the effect of formant and transitional amplitude. So in this paper, the pitch is detected after the elimination of formant ingredients by flattening the spectrum in frequency region. The effect of the transition and change of phoneme is low in frequency region. In this paper we proposed the new flattening method of log spectrum and the performance was compared with LPC method and Cepstrum method. The results show the proposed method is better than conventional method.
Keywords: *Pitch extraction, Flattening method, Autocorrelation method*
ASK subject classification: *Speech signal processing (2.4)*

I. 서 론

음성신호처리 분야에 있어서 기본주파수 즉, 피치정보는 매우 중요하다. 만일 음성신호의 기본주파수를 정확히 검출할 수 있다면 음성인식에 있어서 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성합성시에 자연성과 개성을 쉽게 변경하거나 유지할 수 있다. 또한 분석시 피치에 동기시켜 분석하면 성문의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있다. 이러한 피치검출의 중요성 때문에 피치검출에 대한 방법들이 다양하게 제안되었으며 그것은 시간영역법, 주파수영역법, 시간-주파수영역법으로 구분할 수 있다. 시간영역 검출법은 병렬처리법, AMDF (Average Magnitude Difference Function)법, ACM (Auto Correlation Method)법 등이 있는데 간단하지만 천이구간에서는 피치검출이 매우 어렵다. 주파수영역의 피치 검출법은 고조파분석법 [3], Lifter법, Comb-filtering법 등이 제안되어 있는데

책임저자: 강은영 (keyjsh@hanmail.net)
156-743 서울시 동작구 상도5동
숭실대학교 정보통신공학과 대학원 음성통신 연구실
(전화: 02-824-0906; 팩스: 02-820-0018)

음소의 천이나 변동에 영향을 적게 받는 반면 기본주파수의 정밀성을 높이기 위해 FFT (Fast Fourier Transform)의 포인터 수를 늘리면 그만큼 처리시간이 길어지고 변화특성에 둔해지게 된다. 시간-주파수 혼성영역법은 시간영역법의 장점과 주파수영역법의 장점을 취한 것이다. 이러한 방법으로는 켑스트럼법, 스펙트럼비교법 등이 있는데, 이 방법은 시간과 주파수영역을 동시에 적용하기 때문에 계산과정이 복잡하다는 단점이 있다[3,4].

본 논문에서는 주파수 영역에서 스펙트럼을 평탄화 시킴으로써 포만트의 영향을 제거하고 FFT의 포인터 수를 늘리지 않고도 주파수 해상도를 높여 피치 검출의 정확성을 높이는 피치 검출법을 제안하고자 한다. 제안하는 스펙트럼 평탄화 기법과 이를 이용한 피치검출은 제 2절과 3절에서 살펴보고 제 4절에서는 실험 및 결과에 대해 설명한 후에 5장에서 결론을 짓게 된다.

II. 스펙트럼 평탄화 과정

음성신호는 FFT변환을 통해 주파수 영역에서 스펙트럼 분석이 이루어진다. 그림 1은 본 논문에서 사용한 스펙트럼 평탄화 알고리즘의 블록도이다. 스펙트럼 신호로부터 포만트의 영향과 천이진폭의 영향을 제거하기 위한 첫 단계로서 주파수 대역을 몇 개의 서브밴드로 나눈다. 이때 서브밴드의 대역폭은 스펙트럼 평탄화에 많은 영향을 준다. 본 논문에서는 피치의 범위가 보통 2.5~25 ms인 것을 감안하여 300 Hz와 400 Hz를 서브밴드의 대역폭으로 사용하였다. 이는 입력음성에 따라 적응적으로 대처하기 위한 것이다. 다음 단계로 각각의 서브밴드에서 최대값을 취하여 프레임의 파라미터로 저장한다. 이 파라미터의 값은 8 KHz 샘플링을 했을 경우 10~13개가 된다. 이 값들은 직접 포만트 성분들을 반영하기 때문에 포만트 포락선을 잘 모델링한다고 할 수 있다. 다음은 구해진 파라미터들로 선형보간을 하여 대략적인 포만트 포락선을 얻은 후 스펙트럼 신호로부터 이를 빼주면 제 1차 스펙트럼 평탄화가 되는 것이다. 가장 이상적인 결과는 입력음성의 피치단위로 서브밴드의 대역폭이 결정된 경우에 나타난다. 따라서 제 1차 스펙트럼 평탄화의 결과를 보상하기 위해 평탄화된 신호를 가지고 다시 한번 위의 알고리즘을 거쳐 제 2차 스펙트럼 평탄화를 시킨다. 이때 서브밴드의 대역폭은 각각 3가지 경우의 대역폭을 사용했다. 제 1차 평탄화의 대역폭이 300 Hz였을 경우 200 Hz, 300 Hz, 400 Hz를 사용하고 400 Hz였을 경우에는 300 Hz, 400 Hz, 500 Hz를 사용했다.

각각의 결과에 대한 비교 평가 방법은 분산을 이용하였다. 분산을 계산하기 전에 각 결과신호들은 최대값이 영이 되도록 정규화시키고 평균이 영인 분산을 계산하여 분산값이 작은 것을 최종적인 결과로 사용하였다. 본 논문에서 사용한 분산은 다음과 같다.

$$Variance = \frac{2}{N} \sum_{k=1}^{N/2} (x(k) - m)^2 \qquad (1)$$

여기서 N은 FFT포인터 수이고 스펙트럼 신호가 Y축으로 대칭이기 때문에 분산은 N/2까지만 이루어진다. 또한 k

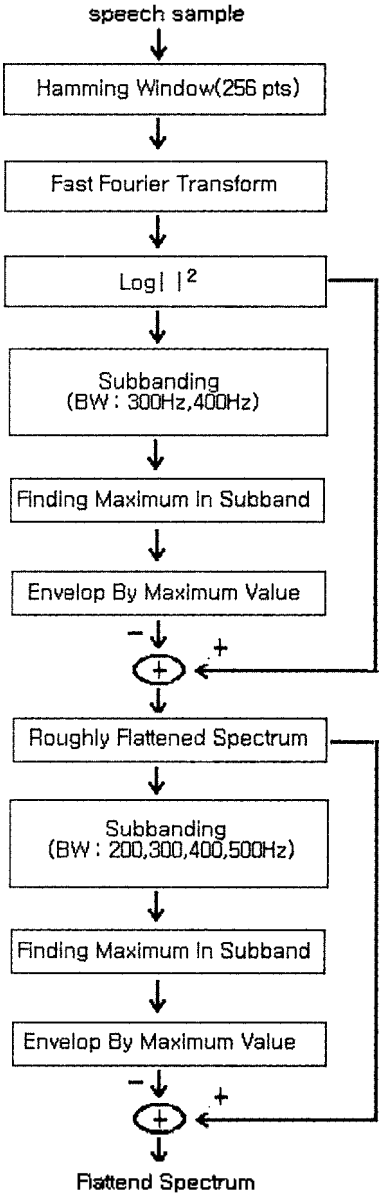

그림 1. 스펙트럼 평탄화 과정
Fig. 1. The flattening process of spectrum.

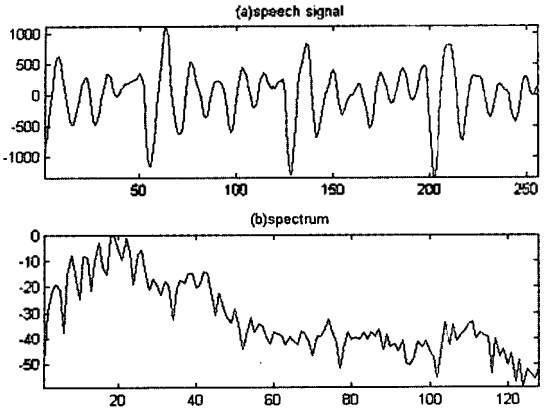

그림 2. 유성음 구간의 신호
(a) 시간영역신호 (b) 로그스펙트럼 신호
Fig. 2. Signal in voiced region.
(a) Time domain signal (b) Log spectrum signal

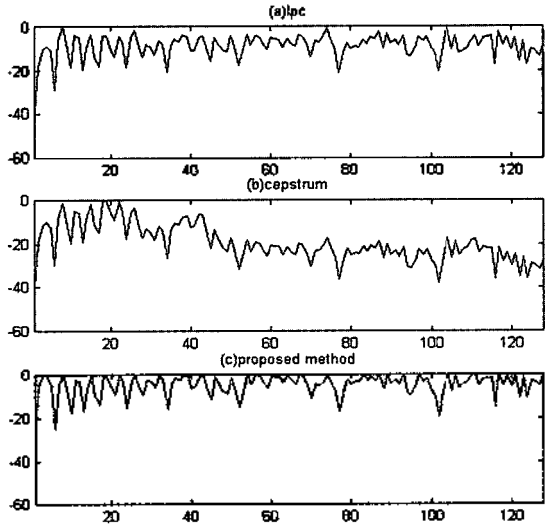

그림 3. 평탄화된 스펙트럼 신호(유성음)
Fig. 3. Flattened spectrum signal.
(a) LPC method (b) Cepstrum method
(c) Proposed method

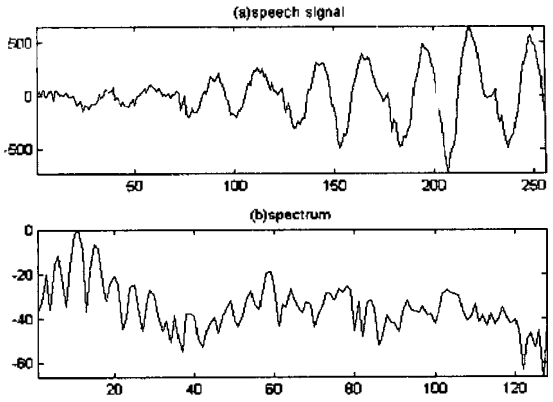

그림 4. 전이구간의 신호
(a) 시간영역신호 (b) 로그스펙트럼 신호
Fig. 4. Signal in transitional region.
(a) Time domain signal (b) Log spectrum signal

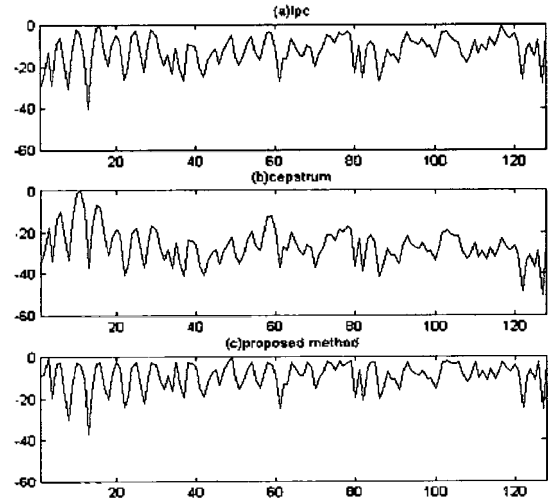

그림 5. 평탄화된 스펙트럼 신호(전이구간)
Fig. 5. Flattened Spectrum Signal.
(a) LPC method (b) Cepstrum method
(c) Proposed method

는 주파수 영역에서의 샘플 인덱스이고 m은 평균을 의미한다. 이때 m값은 0을 사용하여 0을 기준으로 평탄화의 정도를 평가하였다.

그림 2는 유성음 구간의 음성신호와 그 로그스펙트럼 신호이다. 이를 제안한 방법을 통해 평탄화시킨 결과가 그림 3에서 보여지고 있다. 그림 3의 (a)와 (b)는 대표적인 포만트 모델링 방법인 LPC (Linear Predictive Coding) 법과 켑스트럼법을 이용하여 스펙트럼을 평탄화한 예이다. 그림에서와 같이 제안한 평탄화 기법이 우수함을 알 수 있다. 그림 4는 전이구간의 신호와 그 로그스펙트럼이고 그림 5는 이를 평탄화시킨 결과이다. 마찬가지로 전이구간에서도 LPC법이나 켑스트럼법보다 스펙트럼 평탄화가 잘 이루어짐을 알 수 있다. 이런 우수한 성능 때문에 피치 검출을 할 때에 유성음 구간에서는 물론이고 전이구간에서도 정확한 피치검출이 가능한 것이다.

III. 피치검출과정

평탄화된 스펙트럼 신호로부터 기본주파수 (피치)를 구하기 위해 자기상관법을 사용하였다. P(k)가 로그 스펙트럼신호일 때 자기상관법은 다음과 같이 정의된다.

$$R(m)=\frac{2}{N}\sum_{k=0}^{N/2}P(k)P(k+m) \qquad (2)$$

여기서 m은 주파수 영역에서 지연된 샘플수를 나타낸다. P(k)는 좌우대칭이므로 자기상관은 FFT크기인 N의 1/2만 수행하여도 된다. 시간영역에서 효과적인 자기상관법을 사용하기 위해 전처리과정을 거치듯이 주파수 영역에서도 마찬가지로 전처리과정이 필요하다. 먼저 자기상관법은 안정구간에서 적용되어야 한다. 하지만 고주파영역에서 안정적이지 못한 고조파가 나타나므로 분석구간을 0-1 kHz까지 대역제한시킨다. 또한 주파수 해상도를 고려해야 한다. 주파수 해상도는 FFT포인트수에 비례하지만 그 길이는 항상 제한되어 있다. 따라서 주파수 해상도를 보상하기 위해 신호를 선형보간한다. 이는 더욱 정확한 피치검출을 할 수 있게 한다.

IV. 실험 및 결과

이상의 과정을 컴퓨터 시뮬레이션하기 위하여 IBM 펜

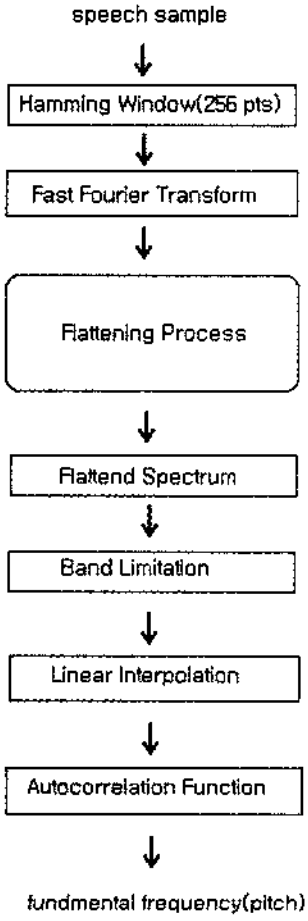

그림 6. 제안한 피치검출과정
Fig. 6. The process of pitch detection.

표 1. 남성화자의 분산값[dB]
Table 1. The variance of male speaker[dB].

	LPC	Cepstrum	New method
발성1	187.13	716.23	124.04
발성2	169.02	697.20	108.68
발성3	179.12	704.65	119.20
발성4	163.72	680.17	107.97
Average	174.74	699.56	114.97

티엄 (III)에 마이크가 부착된 16-비트 A/D변환기를 인터페이스시키고 아래의 문장들을 남녀 각 3명에게 발성시키면서 8 kHz의 표본화 주파수로 표본화하여 저장한 다음, 시뮬레이션의 시료로 사용하였다.

발성1) "인수네 꼬마는 천재소년을 좋아한다."
발성2) "예수님께서 천지창조의 교훈을 말씀하셨다."
발성3) "창공을 날으는 인간의 도전은 끝이없다."
발성4) "숭실대학교 음성통신 연구팀이다.""

표 1은 남성화자의 발성별 분산값을 보여주고 있다. 결과값에서 보여지듯이 켑스트럼법이 가장 큰 분산값을 나타내고 LPC법은 양호한 특성을 보이지만 제안한 방법보다 약 1.5배 큰 분산값을 보이고 있다.

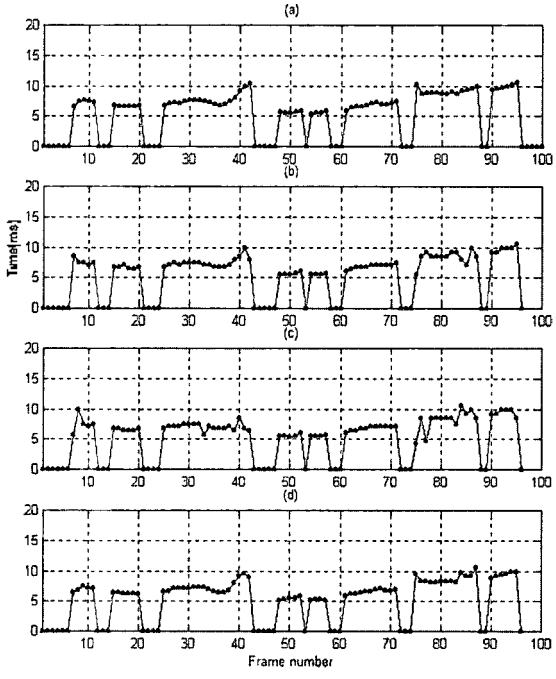

그림 7. 피치 변화도 (30dB)
Fig. 7. Pitch contour (30dB).
(a) Reference pitch
(b) LPC method
(c) Cepstrum method
(d) Proposed method

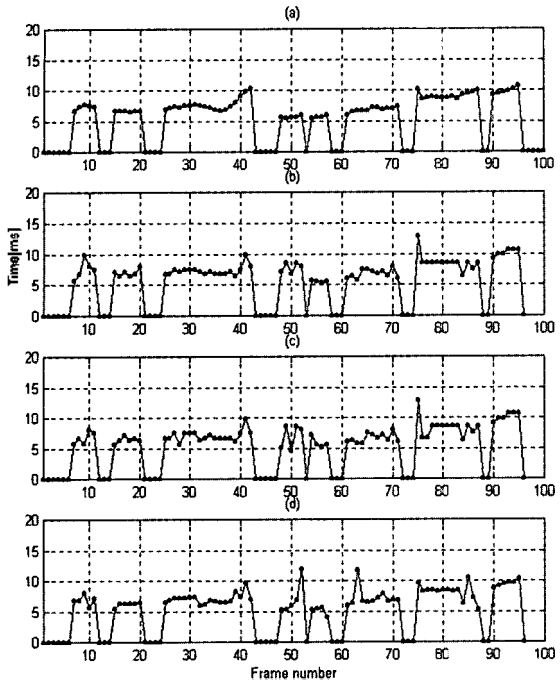

그림 8. 피치 변화도 (6dB)
Fig. 8. Pitch contour (6dB).
(a) Reference pitch
(b) LPC method
(c) Cepstrum Method
(d) Proposed Method

그림 7은 30 dB인 환경에서 발성 1에 대한 피치 변화도이다. 그림 7에서와 같이 LPC나 켑스트럼을 사용하는 것보다 제안한 알고리즘을 사용할 때 더 정확한 피치검출을 할 수 있었다.

그림 8은 6 dB인 환경에서의 결과이다. 마찬가지로 LPC나 켑스트럼을 사용하는 것보다는 제안한 알고리즘을 사용할 때 더 성능이 좋음을 알 수 있었다.

V. 결 론

본 논문은 주파수영역법 중에서 고조파 분석법을 개선한 것으로 스펙트럼을 평탄화시키고 이의 자기상관 신호를 구하여 기본주파수 (피치)를 구하는 것이다. 자기 상관 함수를 적용하기 전에 1 kH까지의 신호를 선형보간하여 주파수 해상도를 보상한다. 이는 FFT를 할 때 고려해야 할 주파수 해상도를 위해 포인터 수를 늘일 필요없어 처리시간을 고려하지 않아도 된다. 실험결과 특히 전이구간에서 좋은 성능 향상을 보였다.

참 고 문 헌

1. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech signals*, Englewood Cliffs, Prentice-Hall, New Jersey, 355-389, 1978.
2. A. M. Noll, "Cepstrum pitch determination," *J. Acoust. Soc. Am.*, vol. **41**, 293-309, February 1967.
3. S. Seneff, "Real time harmonic pitch detection," *IEEE Trans. Acoust. Speech, and Signal Processing*, vol. **ASSP-26**, 358-365, Aug. 1978.
4. J. D. Markel and A. H. Gray, Jr., *Linear Prediction of Speech*, Spriger-Verlag, New York, 1976.
5. R. L. Miller, "Performance characteristics of an experimental harmonic identification pitch extraction(HIPEX) systems," *J. Acoust. Soc. Amer.*, vol. **43**, 1593-1601, Dec. 1970.
6. M. Lee, C. Park, M. Bae, and S. Ann "The high speed pitch extraction of speech signals using the area comparison ethod," *KIEE*, Korea, **22** (2), 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, **24** (3), 381-386, May 1987.
8. Hans Werner Strube, "Determination of the instant of glottal closure from the speech wave," *J., Acoust., Soc., Am.* **5** (5), 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, **24** (2), 274-279, March 1987.

저자 약력

● 강 은 영 (Eun-Young KANG)

2000년 2월: 숭실대학교 정보통신공학과 졸업 (공학사)
2002년 2월: 숭실대학교 정보통신공학과 대학원 졸업 (공학석사)
※ 주관심분야: 음성코딩, 음성인식, 음성합성

● 민 소 연 (So-Yeon MIN)

1993년 2월 : 숭실대학교 전자공학과 졸업 (공학사)
1995년 2월: 숭실대학교 전자공학과 졸업 (공학석사)
2002년 2월: 숭실대학교 전자공학과 박사수료
※ 주관심분야: 음성코딩, 데이터통신

● 배 명 진 (Myung-Jin BAE)

1987년: 서울대학교 공과대학 전자공학과 졸업 (공학박사)
2002년 현재: 숭실대학교 정보통신전자공학부 교수
※ 주관심분야: 음성코딩, 음성인식, 음성합성, 데이터통신, 디지털 신호처리 등