

Harmonic 분산값 최소화 알고리즘에 의한 주파수 영역 평탄화 기법

민소연^{1*}, 김영규²

¹서일대학 정보통신과, ²(주) 팬텍 UMTS

The Technique of Spectrum Flattening by Algorithm for Minimized Harmonics Variance Value

So-Yeon Min^{1*} and Young-Kyu Kim²

¹Dept. of Information and Communication Engr., Seoil University

²PANTECH UMTS

요약 음성신호처리 분야에 있어서 정확한 기본주파수(피치)를 검출하는 것은 매우 중요하다. 그러나 포만트의 영향과 천이 진폭의 영향으로 인하여 음성신호로부터 정확한 피치를 검출하는 것은 매우 어렵다. 따라서 본 논문에서는 음소의 천이나 변동의 영향이 적은 주파수 영역에서의 하모닉스 분산값 최소화 알고리즘을 통해 스펙트럼을 평탄화 하여 피치를 검출하는 방법에 대하여 연구하였다. 실험결과에서는 제안한 방법이 기존의 방법인 LPC법, 켈스트럼 법과 비교하여 평탄화 특성이 어느 정도의 우수성을 보이는지를 평가하였다. 또한 각각의 방법을 적용하여 기본주파수를 검출한 결과를 비교함으로써 제안한 방법이 우수함을 입증하였다.

Abstract The exact fundamental frequency (pitch) extraction is important in speech signal processing. 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 flattening the spectrum in frequency region by proposed algorithm for minimized harmonics variance value. Experimental result showed the proposed method appeared an outstanding performance in compared with LPC, Cepstrum. Also, the results show the proposed method is better than conventional method.

Key Words : Fundamental frequency extraction, Flattening method

1. 서론

음성신호의 피치주기 검출과정은 다양한 음성신호처리 시스템에 필수적으로 사용되어지고 있다. 그러나 대부분의 피치 검출기는 잡음, 대역폭 제한, 주파수의 진폭특성 열화 등의 환경이나 주파수의 위상특성이 변질된 환경에서는 신뢰할 수 있는 결과를 얻지 못하고 있다.

피치주기 검출 과정은 다음과 같은 특징에 의해 정확하고 신뢰적인 측정이 어렵다. 첫째, 정확한 피치주기 검출은 성문의 여기파형이 완전히 주기적인 파형이 아니라 는 점에서 비롯된다. 둘째, 성도의 성문간의 상호작용에

의해 문제점이 발생한다. 셋째, 유성음구간 동안 각 피치 주기의 정확한 시작과 끝을 정의하기 어렵고 넷째, 무성음과 낮은 레벨의 유성음 구간의 변화가 뚜렷하지 않기 때문에 정확한 위치를 찾기 어렵다. 다섯째, 음성에 대한 전화시스템의 영향으로 선형, 비선형처리를 포함하면서 음성 신호에 잡음이 추가되는 경우를 들 수 있다[1,2].

그리고 피치검출의 중요성으로 인해 다양한 방법들이 연구되어졌으며 시간영역 피치검출법, 주파수영역 피치 검출법, 시간-주파수영역 피치검출법으로 구분할 수 있다. 시간영역 피치 검출법은 빠르고 분해능이 높으며 음소 변화에 따른 특성 검출 가능 등의 장점을 갖지만 천이

본 논문은 2009년도 서일대학 학술연구비에 의해 연구되었습니다.

*교신저자 : 민소연(symin@seoil.ac.kr)

접수일 10년 08월 03일

수정일 10년 09월 06일

게재확정일 10년 09월 08일

구간에서 피치 검출이 어렵다는 단점을 지니며 병렬처리법, AMDF (Average Magnitude Difference Function)법, 자기상관법(Auto Correlation Method) 등이 있다[2,3]. 주파수영역의 피치 검출법은 음소의 천이나 변동에 영향을 적게 받는 반면 기본주파수의 정밀성을 높이기 위해 FFT(Fast Fourier Transform)의 포인트 수를 늘리면 그만큼 처리시간이 길어지고 변화 특성에 둔한 특징을 갖는다. 주파수 영역에서 연구되어진 방법으로는 고조파분석법, Lifter법, Comb-filtering법 등이 있다. 시간-주파수 영역 피치 검출법은 시간영역에서의 계산시간 절감, 검출된 피치의 정밀성 등의 장점과 주파수영역에서 평균화 효과에 의해 배경잡음이나 음소간의 천이, 음소의 변동 등에서 나타나는 배주기와 반주기 문제를 극복할 수 있다는 장점을 갖는다. 지금까지 연구되어진 시간-주파수 영역 피치 검출법에는 켈스트럼법, 스펙트럼비교법 등이 있다 [1-3].

본 논문에서는 스펙트럼 평탄화 및 정확한 피치검출을 위해 고조파 분산값의 최소화 알고리즘을 제안한다. 2장에서는 기존의 스펙트럼 평탄화 방법에 대해 설명하고 3장에서는 스펙트럼 왜곡 최소화를 위한 새로운 스펙트럼 평탄화 알고리즘에 대해 설명한다. 4장에서는 실험결과를 분석하고 5장에서는 결론을 맺는다.

2. 기존의 스펙트럼 평탄화 방법

음성분석시 신호를 성도(vocal track) 여파기(filter)성분과 여기(excitation)성분으로 분류한다. 여기성분은 성문(vocal cord)을 통과하여 나오는 준 주기적인 공기의 흐름과 성도를 스치며 발생하는 넓은 대역의 잡음으로 분류한다. 성도여파기의 응답은 주파수 영역에서 스펙트럼 신호의 포먼트(formant) 포락선을 의미하며, 다음은 대표적인 두 가지 방법을 설명하고 있다[1,2,6].

2.1 선형예측분석(LPC)법

선형예측분석법은 음성스펙트럼이 가진 특성을 적은 수의 파라미터만으로 정확하게 표현할 수 있으며 현재의 음성샘플을 이전 p개의 샘플로부터 예측한다[1,7]. 식(1)에서 x_t 는 시간 t에서의 음성샘플을 의미하고 e_t 는 예측값과 실제값의 차이를 의미한다.

$$x_t = -\sum_{i=1}^p a_i x_{t-i} + e_t \quad (1)$$

식(2)에서는 음성샘플의 예측값(\hat{x}_t)과 잔차신호(e_t)를 정의하고 있다.

$$\hat{x}_t = -\sum_{i=1}^p a_i x_{t-i} \quad , \quad x_t - \hat{x}_t = e_t \quad (2)$$

식(2)를 각각 z 변환한 후, 계산과정을 거치면 식(3)을 도출할 수 있다.

$$X(z) = \frac{1}{A(z)}E(z) \quad , \quad A(z) = 1 + \sum_{i=1}^p a_i z^{-i} \quad (3)$$

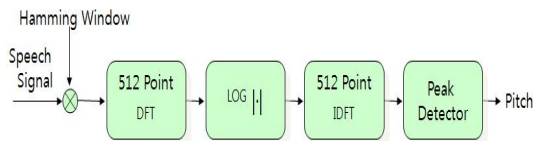
식(3)에서 $1/A(z)$ 를 선형예측필터(linear prediction filter)라고 하고, 음성신호가 가지고 있는 여러 가지 정보 중에서 $1/A(z)$ 의 주파수응답은 스펙트럼 포락선과 대응된다[5,7,11,12].

2.2 켈스트럼(Cepstrum)법

켈스트럼은 로그크기 스펙트럼의 역 푸리에 변환으로 정의하며, 식(4)와 같이 나타낼 수 있다[1,7].

$$c(\tau) = F^{-1} \log|X(w)| = F^{-1} \log|G(w)| + F^{-1} \log|H(w)| \quad (4)$$

식(4)에서 $X(w)$, $G(w)$, $H(w)$ 는 음성신호($x(t)$), 음원($g(t)$), 스펙트럼 포락선($h(t)$)을 푸리에 변환한 것이다. 켈스트럼은 음성이 갖는 정보에서 스펙트럼 포락선 정보와 세부 구조 정보를 분리하며, 켈스트럼의 낮은 시간대의 부분은 성도, 성문, 입술의 방사정보를 갖으며 높은 시간대의 부분은 여기 성분을 갖는다. 그림 1은 켈스트럼에 의한 피치검출과정을 나타내고 있다.



[그림 1] 켈스트럼에 의한 피치검출법

3. 제안한 방법

음성신호는 FFT변환을 통해 주파수 영역에서 스펙트럼 분석이 이루어진다. 그림 2는 본 논문에서 제안한 하모닉스 분산값 최소화 알고리즘에 의한 스펙트럼 평탄화

과정에 대해 나타내고 있고 이 과정에서 찾아진 peak 값과 valley 값의 총 개수에 대한 계산을 식(3)에서 정의하였다. 여기서 스펙트럼 신호가 Y축 대칭이기 때문에 frame/2 까지 계산되어 진다.

$$Peak(Mag_x(k)) \begin{cases} x(n) - x(n-1) > 0, 1 < n < \frac{frame}{2} + 1 \\ x(n+1) - x(n) < 0 \end{cases}$$

$$Valley(Mag_y(k)) \begin{cases} x(n) - x(n-1) < 0, 1 < n < \frac{frame}{2} + 1 \\ x(n+1) - x(n) > 0 \end{cases} \quad (3)$$

또한, 이 과정에서 한 칸 간격은 δ , 보간(interpolation)된 포락선(envelope)을 $Mag_k(j)$ 표기하고, 잔차 신호는 $residual(f_s)$, 분산 값을 $Variance$ 로 표기하여 계산한 값을 식(4)에서 나타내었다.

$$\delta = \frac{Mag_x(k+1) - Mag_x(k)}{x(k+1) - x(k)}$$

$$Mag_k(j) = Mag_x(k) + (\delta \cdot j), 1 \leq k \leq total, k \leq j \leq k+1$$

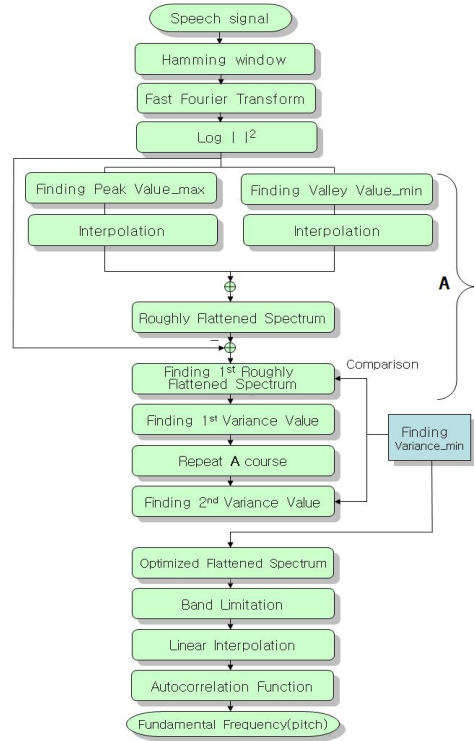
$$residual(f_s) = fft(s(n)) - Mag_k(j)$$

$$Variance = \frac{2}{FrameSize} \sum_{f_s=1}^{FrameSize/2} (Residual(f_s) - Residual_{max})^2 \quad (4)$$

개략적인 포맷트 포락선을 얻은 후 평탄화된 스펙트럼 신호에서 빼주면 1차 스펙트럼 평탄화 수행 및 분산 값이 계산되어지고, 같은 절차가 반복되어지는 2차 과정의 분산 값이 계산되어진 후 비교과정을 통해 고조파 분산을 최소화하는 최적화된 평탄화 스펙트럼을 얻을 수 있게 된다.

평탄화된 스펙트럼 신호에서 기본주파수(피치)를 구하기 위하여 식(5)에서 설명하고 있는 자기상관법(Auto-correlation)을 사용하였다[1,5,9]. 식(5)에서 m은 주파수영역에서의 지연(delay)된 샘플수를 나타낸다. 또한 Y(k)는 좌우대칭이므로 자기상관은 FFT 크기인 N의 1/2만 수행한다.

$$R(m) = \frac{2}{N} \sum_{k=0}^{N/2} Y(k)Y(k+m) \quad (5)$$



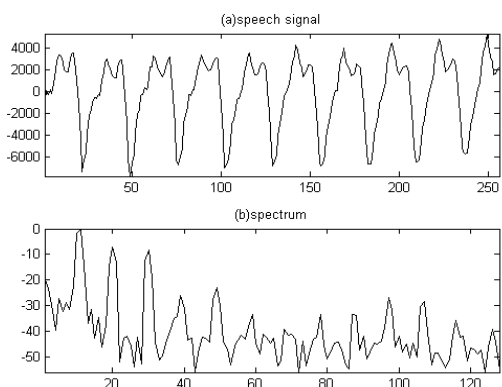
[그림 2] Harmonics 분산값 최소화 알고리즘

주파수영역에서 자기상관법을 사용하기 위하여 전처리 과정을 거치는데 고주파영역에서 안정적이지 못한 고조파가 나타나므로 분선구간을 0 ~ 1kHz까지 대역 제한시키고, 해상도 또한 고려하여야 한다. 주파수 해상도는 FFT 포인트 수에 비례하지만 그 길이는 항상 제한되어있다. 따라서 주파수 해상도를 보강하기 위해 신호를 선형 보간하여 더욱 정확한 피치를 검출할 수 있게 한다.

4. 실험 및 결과

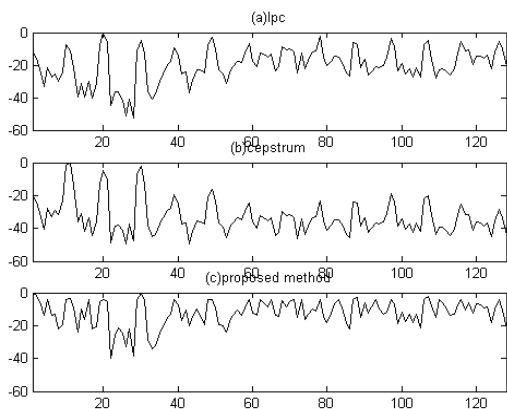
본 논문에서 제안한 방법을 실험하기 위하여 IBM PC(2GHz) 시스템을 사용하였고 여기에 음성신호의 입출력 인터페이스용으로 상용화된 16비트 A/D변환기를 사용하였다.

그림 3은 실험에서 사용한 다양한 음성신호들 중에서 유성음 구간에 해당하는 임의의 프레임에 대한 분석을 나타낸 결과로써 (a)는 각 샘플에 대하여 16비트 양자화를 취한 시간영역 신호를 나타내고 있고 (b)는 FFT 변환을 수행하여 각 포인트에 대해 데시벨(decibel)을 취한 주파수영역 특성 결과를 나타내고 있다.



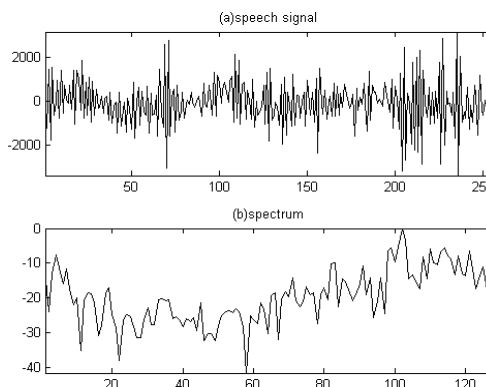
[그림 3] 유성음구간에 대한 분석 결과
(a) 시간영역 신호 (b) 주파수영역 분석결과

그림 4는 그림 3에서 나타난 임의의 유성음 구간에 대하여 기존의 LPC법, Cepstrum법과 제안한 방법에 대하여 FFT 변환을 수행한 결과이다. 즉, 각 포인트에 대해 데시벨(decibel)로 나타난 주파수 스펙트럼의 평탄화 정도를 비교한 결과를 나타내고 있다. 실험 결과를 통하여 제안한 방법이 기존의 방법들에 비교하여 스펙트럼 평탄화 결과가 우수함을 나타내고 있으며, 특히 캡스트럼법에 비해 고주파영역에서 20dB 이상 강조됨을 알 수가 있다.



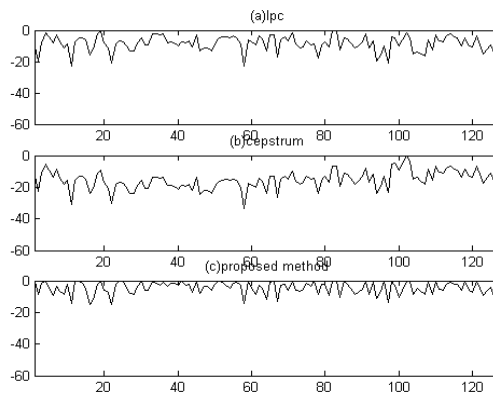
[그림 4] 주파수영역의 스펙트럼 분석
(a) LPC분석 (b)캡스트럼분석 (c)제안한 방법

그림 5는 다양한 음성시료 중, 무성음 구간에 해당하는 임의의 프레임 분석 결과로써 (a)는 각 샘플에 대하여 16비트 양자화를 위한 시간영역 신호를 나타내고 있고 (b)는 FFT를 수행한 각 포인트에 대해 데시벨(decibel)을 취한 결과를 보이고 있다.



[그림 5] 무성음구간에 대한 분석결과
(a) 시간영역 신호 (b) 주파수영역 분석결과

그림 6은 그림 5의 유성음 구간에 대하여 LPC법, Cepstrum법, 제안한 방법들의 FFT 변환을 수행한 결과이다. 즉, 각 포인트에 대해 데시벨(decibel)로 나타난 주파수 스펙트럼의 평탄화 정도를 비교한 결과를 나타내고 있다. 실험결과에서 보여지듯이 제안한 방법은 전 주파수 대역에 걸쳐 평탄화 정도가 우수함을 알 수가 있고 정확한 피치 주파수를 얻을 수 있음을 알 수가 있다.



[그림 6] 주파수영역의 스펙트럼 분석
(a) LPC분석 (b)캡스트럼분석 (c)제안한 방법

표 1에서는 실험에 사용한 전체 음성시료 들에 대하여 LPC법, 캡스트럼법, 제안한 방법 간의 분산값을 비교한 결과를 나타내고 있으며, 분석결과를 통해 제안한 방법의 분산값이 가장 작은 것을 알 수가 있다.

[표 1] 분산값 분석 결과 [dB]

	LPC	Cepstrum	Proposed Method
음성시료 1	270.57	994.34	208.24
음성시료 2	263.45	965.26	197.92
음성시료 3	364.87	742.18	286.38

5. 결론

본 논문은 피치검출을 위한 주파수 영역법 중에서 고조파 분석법을 개선한 것으로, 고조파 분산값 최소화를 통해 주파수 대역을 평탄화 하여 정확한 기본 주파수(피치) 값을 구하였다. 즉, 주파수 영역에서 포먼트 특성을 제거하여 스펙트럼을 효율적으로 평탄화하고 주파수 해상도를 위해 FFT 포인트 수를 늘리지 않고도 피치 검출시 정확도를 높일 수 있는 피치 검출법을 제안하였다.

실험 결과, 제안한 방법이 기존의 LPC법, 켈스트럼법에 비교하여 주파수 대역의 스펙트럼 평탄화 정도가 우수하고, 또한 고조파 분산값이 작음을 알 수가 있었다.

향후, 다양한 피치 변경율에 대한 결과와 잡음환경에서의 결과를 연구하여야 할 것이다.

참고문헌

- [1] 배명진, 디지털 음성합성, 동영출판사, pp. 225-227, 1998년.
- [2] 서지호, 김종국, 배명진, “고음질 음성합성을 위한 LSP를 이용한 피치검출 성능향상에 관한 연구”, 한국음향학회 논문지, 제29권 제1호, pp. 69-75, 1월, 2010년.
- [3] 강은영, 민소연, 배명진, “개선된 피치검출을 위한 스펙트럼 평탄화 기법에 관한 연구”, 한국음향학회 논문지, 제21권 제3호, pp. 310-314, 4월, 2002년.
- [4] 송한춘, 민소연, 이광형, “스펙트럼 왜곡 최소화를 위한 효율적인 피치 검출 알고리즘”, 한국통신학회 논문지, 제32권 제10호, pp. 351-355, 4월, 2007년.
- [5] A.M.Kondoz, Digital Seech, John & Sons Ltd, pp. 84-95, 1994.
- [6] 배명진, 이상호, 디지털 음성분석, 동영출판사, pp. 162-163. 1998년.
- [7] 오영환, 음성언어정보처리, 홍릉과학출판사, pp.24-25, 1998년.
- [8] 한진수, 음성신호처리, 오성미디어, pp. 52-61. 2003년.

- [9] L.R.Rabiner,R.W.Schafer, Digital Processing of Speech Signal, pp. 309-402, 1978.
- [10] 김영규, 서지호, 배명진, “잡음환경에서 서브밴드 스케일링 피치변경법에 관한 연구”, 한국음향학회 추계 학술발표대회논문집, 제22권 제2호, pp. 35-38. 2003년.
- [11] 민소연, 배명진, “검색구간 조절에 의한 LSP 변환시간 감소에 관한 연구”, 한국산학기술학회 추계학술발표대회논문집, 제7권 제2호, pp. 208-211. 2006년.
- [12] 민소연, “이동통신용 음성부호화기에서의 LSP 계산시간 감소에 관한 연구”, 한국산학기술학회, 제8권 제3호, pp. 563-568. 2007년.

민 소 연(So-Yeon Min)

[종신회원]



- 1994년 2월 : 송실대학교 전자공학과(공학사)
- 1996년 2월 : 송실대학교 일반대학원 전자공학과(공학석사)
- 2003년 2월 : 송실대학교 일반대학원 전자공학과(공학박사)
- 2005년 3월 ~ 현재 : 서일대학 정보통신과 교수

<관심분야>

통신 및 신호처리, 정보통신, RFID

김 영 규(Young-Kyu Kim)

[정회원]



- 2002년 2월 : 송실대학교 정보통신과 졸업(공학사)
- 2004년 2월 : 송실대학교 일반대학원 정보통신과 졸업(공학석사)
- 2005년 3월 ~ 현재 : (주) 팬택 단말기 HSDPA 개발 전임연구원

<관심분야>

통신 및 신호처리, 멀티미디어 통신