자기상관법을 이용한 제 1 포만트 검출법에 관한 연구

강 은 영, *민 소 연, 배 명 진
숭실대학교 정보통신공학과, *전자공학과
156-743 서울시 동작구 상도동 1-1

On a Study of Detecting First Formant Using Autocorrelation Method

EunYoung KANG, *SoYeon MIN, MyungJin BAE
Dept. of Information and telecommunication Engr.,Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
E-mail : keyjsh@hanmail.net

Abstract

In the speech analysis, to estimate formant center frequencies exactly is very important. If we know formant frequencies, we can expect which pronunciation is uttered. Generally, the magnitude of first formant frequency in voiced speech is 10dB more than other formant frequency. So, the shape of voice signal in time domain is affected by mainly first formant. Therefore we can get first formant frequency roughly by using ZCR(Zero Cross Rate).

In this paper, we proposed the improvement method to get first formant frequency by using ZCR. We did autocorrelation before getting ZCR. This procedure makes voice signal smooth so, first formant in voice signal is emphasized. As a result of this method, we got more exact ZCR and first formant frequency. Conventional method of formant estimate is done in frequency domain but proposed method is done in time domain. So, this is very simple.

1. 서 론

포만트 성분이 시간영역에서 어떻게 영향을 미치는지를 이해하는 것은 포만트 보코더, 음성인식, 포만트에 기초하여 합성하기 위한 음성분석등의 응용분야에서 매우 중요하다. 음성 파형의 공명 주파수를 알면 그 사람이 어떤 음성을 발음했는지를 알 수 있다. 일반적으로 모음에서는 제 1포만트와 제 2 포만트에 의해서 그 음소의 음운학적인 성질이 결정된다. 그 중에서도 제 1 포만트는 시간영역의 파형에 큰 영향을 미치게 된다. 또한, 일반적으로 제 1 포만트(F_1)는 성도의 수축과 두가지 관계가 있는데 첫째는 구강 앞쪽이 수축될수록 제 1 포만트가 낮아진다는 것이고 둘째는 인두가 수축할수록 제 1 포만트가 높아진다는 것이다. 본 논문에서는 자기상관법이 제 1 포만트 성분을 강조하는 특징을 이용하여 좀더 정확한 제 1 포만트를 구하였다. 한 피치구간에서 평균 영교차 간격의 역수는 $2F_1$의 주파수와 같기 때문에 자기 상관한 신호의 ZCR(영교차율)을 통하여 제 1 포만트를 구한다.

2. 기존의 포만트 추정법

포만트의 중심주파수와 그의 대역폭을 추정하는 일반적인 방법에는 단시간 퓨리에 변환, 필터뱅크법, LPC 파라미터법, 켑스트럼법등이 있으며 모두 스펙트럼영역에서 봉우리를 찾는 것이다.

2-1. 필터뱅크법

필터뱅크는 스펙트럼 분석방법중의 하나로 실시간에서 사용되고, 간단하며, 저렴하게 구현이 가능하여 널리 이용되고 있는 방법으로 입력음성의 주파수를 범위를 다르게 하여 각각의 스펙트럼을
분석한다. 상업적인 음성인식에 많이 사용되는 필터뱅크법은 DFT(Discrete Fourier Transform) 분석보다 자유롭게 다룰 수 있다. 왜냐하면 광대역 또는 협대역 분석의 경우에 고정된 것보다는 귀의 특성에 따라서 대역폭을 변화할 수 있기 때문이다. 8-12개의 대역통과 필터뱅크에서 진폭출력은 일반적으로 DFT보다 더 자세하고 효과적으로 스펙트럼을 표현한다. 일반적인 접근은 동일하고, 1kHz 이상의 고정된 대역폭 필터로 필터를 구성한 다음 log 단위로 필터의 대역폭을 증가시킨다.

2-2. LPC 파라미터법

음성의 선형예측분석은 몇 가지 유리한 점을 갖고 있지만 음성의 유성음 구간에서 포만트를 결정할때 문제점이 발생한다. 문제점을 해결하기 위한 두가지 방법이 있다. 그 중의 한 방법은 예측 파라미터로 포만트를 결정하는 것이다. 대부분 직접적인 방법으로 예측기 다항식을 인수분해하여 스펙트럼의 날카로운 극점들을 포만트로 결정하게 된다. 또 다른 방법은 스펙트럼을 구하고, 봉우리선택법(peak picking method)에 의하여 포만트를 결정한다. 선형예측법의 장점은 포만트의 중심주파수와 대역폭을 예측기 다항식을 인수분해함으로써 정확하게 결정할 수 있다는 것이다. 또한 포만트를 결정하는데 켑스트럼 평탄화(smoothing)된 스펙트럼을 얻는 방법들 보다 적은 수의 극점이 존재하기 때문에 LPC(Linear Predictive Coding)가 덜 복잡하다. 마지막으로 이질적인 극점들은 LPC분석에서 쉽게 분리된다. 왜냐하면 음성 포만트의 대역폭과 비교해 볼 때 극점들의 대역폭이 매우 크기 때문이다.

2-3. 켑스트럼법

켑스트럼의 높은 큐프런시(Quefrency) 대역에서는 스펙트럼의 주기성을 나타내는 정보인 기본주파수 즉 피치 정보를 얻을 수 있다. 그리고 낮은 큐프런시 대역에서는 로그가 취해진 스펙트럼 포락선 정보를 얻을 수 있다. 이때 낮은 큐프런시 대역에서 몇 개의 원소를 취하느냐에 따라서 스펙트럼 포락선의 정밀도가 결정된다. 즉, 더 많은 원소를 취할수록 스펙트럼 포락선이 원래의 스펙트럼 자체에 가까워지게 되며, 적은 원소를 취할수록 스펙트럼의 전체적인 모양을 반영하는 부드러운 형태를 띠게 된다.

리프터링(Liftering)을 통해 켑스트럼 계수들을 높은 큐프런시 계수와 낮은 큐프런시 계수로 구분하여 스펙트럼 포락선 정보를 얻을 수 있다. 로그 스펙트럼은 특정한 입력 음성 부분의 공명구조를 보여준다. 다시 말해서 스펙트럼의 봉우리들은 포만트 주파수와 일치한다. 이러한 사실을 통해서, 켑스트럼에 의해 스펙트럼에서 봉우리들의 위치를 찾아내어 포만트를 검출할 수 있다.

3. 제안한 알고리즘

본 논문에서 사용하는 제 1 포만트 추정법은 영교차율을 이용하는 것이다. ZCR(영교차율)에 대한 수식은 다음과 같다.

$$F_0 = (ZCR * F_s)/2 \qquad (1)$$

식 (1)에서 F_s는 신호의 표본화 주파수이고 F_0는 기본주파수이다. 일반적인 유성음 구간에서 F_1(제 1 포만트)의 에너지 봉우리는 다른 포만트들 보다 10dB 이상 높기 때문에 시간영역의 파형에서는 F_1의 영향이 주로 나타난다.

한 피치 구간에서 평균-영교차-간격 (Zero Crossing Interval : ZCI)의 역수는 $2F_1$의 주파수와 거의 같게 된다. 그리고 포만트들은 대역폭을 갖게 되므로 시간영역 파형의 한 피치구간에서는 감쇄진동을 하게 된다.

신호를 Autocorrelation(자기상관)하면 제 1 포만트 성분이 강조되는 것을 이용하여 신호의 ZCR을 구하기 전에 Autocorrelation을 한다. Autocorrelation은 스무딩 효과를 가져오기 때문에 제 1 포만트 성분을 강조한다. 따라서 좀더 정확한 ZCR을 얻을 수 있고 좀더 정확한 제 1 포만트를 추정할 수 있다. 그림 1은 신호를 Autocorrelation하면 제 1포만트 성분이 강조된다는 것을 보여주고 있다.

논문에서 사용한 Autocorrelation을 수식으로 표현하면 다음과 같다.

$$\Phi(k) = \sum_{m=-\infty}^{\infty} x(m)x(m+k) \qquad (2)$$

창함수를 통과한 음성신호에 대한 단구간 자기상관함수는 다음과 같이 표현된다.

(3)

$$R_n(k) = \sum_{m=-\infty}^{\infty} x(m)w(n-m)x(m-k)w(n-m+k)$$

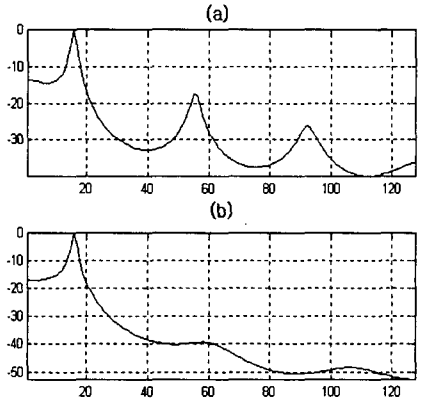


그림. 1 음성신호의 LPC 분석결과
(a) 음성신호 (b) Autocorrelation한 신호

4. 실험 및 결과

제안한 방법을 실험하기 위해서 먼저 IBM PC(333 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 20대 남성이 연구실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화 하여 사용하였다. 실험에 사용한 문장은 다음과 같다.

발성1) "인수네 꼬마는 천재 소년을 좋아한다."
발성2) "예수님께서 천지창조의 교훈을 말씀하셨다."
발성3) "청공을 날으는 인간의 도전은 끝이 없다."

그림 2는 음성신호를 프레임 단위로 자기상관한 결과이다. 이 자기상관(Autocorrelation)한 신호를 이용하여 평균적인 피치를 구한다. 평균적인 피치를 구하고 난 후에 피치구간에서 원 음성신호의 ZCR과 Autocorrelation한 신호의 ZCR을 구한다. 그리고 이 ZCR을 이용하여 각각의 제 1 포만트를 계산한다. 기준으로 사용할 제 1 포만트는 LPC를 이용하여 구한 스펙트럼 포락선에서 직접 계산한다.

표 1은 연속된 10프레임에 대해 각각의 제 1포만트를 계산한 것이다. 표 1에서 볼 수 있듯이 원음성 신호로부터 구한 제 1 포만트보다 자기상관법에 의한 신호로부터 구한 제 1 포만트가 기준 포만트에 가깝다는 것을 알 수 있다. 그림 3은 음성시료를 사용하여 프레임별 제 1 포만트 주파수를 비교한 것이다. 그림 4는 f1과 R_f1의 편차를 비교한것이다.

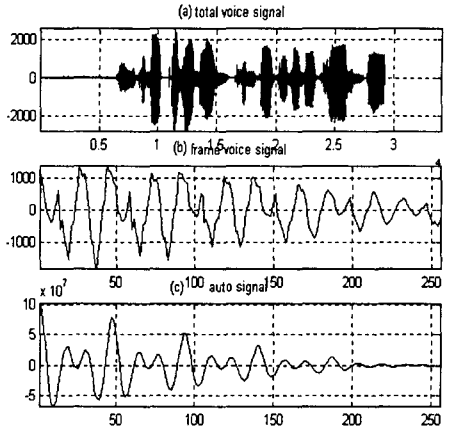


그림. 2 음성신호의 Autocorrelation
(a) 음성신호 (b) 프레임 처리된 음성신호
(c) 프레임 처리된 신호의 자기상관 결과

다음은 논문에서 사용된 약어에 대한 설명이다.

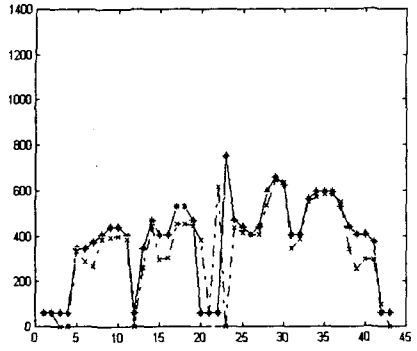
Specf1 : 신호의 LPC를 통해 얻은 스펙트럼 포락선으로부터 직접 구한 제 1 포만트

f1 : 음성신호의 한 피치 내에서 얻은 ZCR로 부터 구한 제 1 포만트

R_f1 : Autocorrelation한 신호에서 얻은 ZCR을 이용하여 구한 제 1 포만트

표1. 연속된 프레임의 제 1포만트 비교

specf1	f1	R_f1
625	507.9365	634.9206
406.2500	375	375
406.2500	258.0645	387.0968
562.5000	482.7586	551.7241
593.7500	357.1429	571.4286
593.7500	581.8182	581.8182
593.7500	436.3636	581.8182
531.2500	413.7931	551.7241
437.5000	417.9104	358.2090
406.2500	354.4304	405.0633

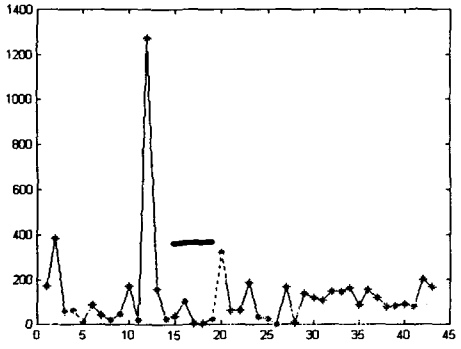


(specf1 : *, f1 : ---o , R _f1 : _._._.x)

그림. 3 프레임별 제 1포만트 비교

5. 결 론

음성신호를 처리할 때 원 음성신호를 조작하기 보다는 적절한 음성 파라미터를 조작하거나 더 효과적으로 저장하기 위해 에너지나 영교차율과 같은 간단한 전처리 과정을 거치게 되는데 음성의 시간영역 파형에서 주된 영향을 미치는 제 1 포만트의 정보를 미리 파악할 수 있다면 여러 응용분야에서 전처리 과정으로 사용될 수 있을 것이다.



(f1의 편차 : * , R_f1의 편차: --o)

그림. 4 f1과 R_f1의 편차 비교

본 논문에서는 신호를 Autocorrelation하는 것이 스무딩 효과를 가져오고 이는 제 1 포만트 성분을 강조한다는 사실을 바탕으로 좀더 정확한 ZCR을 얻을 수 있었고 이로부터 좀더 정확한 제 1 포만트를 구할 수 있었다. 이는 ZCR을 이용하여 대략적인 제 1 포만트 구하는 방법을 개선한 것이다. 기존의 포만트 추정법들은 주파수 영역에서의 스펙트럼 분석이 이루어진 반면 제안한 알고리즘은 시간영역에서 이루어지고 있다. 즉, 주파수 영역으로의 변환이 필요 없음을 의미한다.

참 고 문 헌

[1] Panos E.Papamichalis, *Practical Approaches to Speech Coding*, Prentice-hall inc., Englewood cliffs, N.J., 1987.

[2] L.R. Rabiner and R.W. Schafer, "Digital processing of Speech Signals Englewood Cliffs", New Jersey : Prentice-Hall, 1978.

[3] Thomas W. Parsons, Voice and Speech Processing, Mcgraw-Hill, 1987.

[4] A. M. Kondoz, *Digital Speech*, 1994.

[5] R.W.Schafer and L.R.Rabiner,"Digital representation of Speech Signal", Proc.IEEE, vol.63, pp.662-667, Apr. 1975.

[6] P.D.Welch, "The use of the Fast Fourier Transform for the Estimation of Power Spectra", IEEE Trans. Audio and Electro Acoust., vol. pp.70-73, Au-15, 1967.

[7] 배명진, 이상효, "디지탈 음성분석", 동영출판사, pp.90-162, 1998.