2001年度 第14回 信號處理合同學術大會 論文集 第14卷 1號

스펙트럼 평탄화 기법의 비교평가에 관한 연구

강 은 영, 한 상 일, 배 명 진
숭실대학교 정보통신공학과
156-743 서울시 동작구 상도동 1-1

A Study on the Comparison and Evaluation of Spectrum Flattening Techniques

EunYoung Kang, SangIl Han, MyungJin Bae
Dept. of Information and telecommunication Engr., Soongsil University
1-1 Sangdo-5Dong, Dongjak-Ku, Seoul 156-743, KOREA
keyjsh@hanmail.net

요 약

스펙트럼의 평탄화는 스펙트럼 신호로부터 포만트의 영향이나 천이진폭의 영향을 제거하는 것이다. 따라서 정확한 피치검출과 포만트검출에 적용할 수 있다.

본 논문에서는 새로운 스펙트럼 평탄화 기법을 제안하고 기존의 방법인 LPC법, Cepstrum법과 비교하여 어느 정도의 우수성을 보이는지 평가하였다. 평가 방법은 각각의 평탄화된 신호의 분산을 구하여 평탄화의 정도를 측정하였다. 이때 평탄화된 신호는 최고점이 영이 되도록 정규화 시키고 평균이 영인 분산을 계산하였다. 실험결과는 제안한 방법이 기존의 방법보다 우수함을 보여준다.

1. 서 론

음성인식, 합성 및 분석과 같은 음성신호처리 분야에 있어서 피치검출이나 포만트검출은 매우 중요하다. 하지만 음성신호에서는 여파기성분과 여기성분이 상호 작용하기 때문에 피치검출이나 포만트검출이 매우 어렵다. 특히 음성신호에 잡음이 부가될 경우에는 더욱 어려워진다. 따라서 낮은 SNR 조건에서도 피치정보나 포만트 정보를 유지하는 스펙트럼 신호는 음성처리 분야에서 매우 중요하다고 할 수 있다. 그런데 스펙트럼 신호에는 고조파 성분과 포만트 성분이 함께 나타난다. 따라서 이를 잘 분리하는 것이 피치검출이나 포만트검출의 관건이라 할 수 있다.

본 논문에서는 스펙트럼 신호를 최대한 평탄화 시킴으로써 포만트의 영향을 제거하고 고조파 성분을 분리해낸다. 기존의 분리방법에는 LPC법, Cepstrum법등이 있는데 LPC법은 포만트성분을 모델링한 것이고 Cepstrum법은 리프터링을 통해서 두가지 성분을 각각 얻을 수 있는 것이다.

기존의 방법들에 대해서는 2절에서 살펴보고 제 3절에서는 제안한 알고리즘을 설명한다. 실험 및 결과와 결론은 각각 제 4절과 5절에서 논의된다.

2. 기존의 스펙트럼 평탄화 방법

대부분의 음성분석에서는 신호가 시간적으로 변하는 성도 여파기성분과 여기성분으로 모델링될 수 있다고 가정한다. 여기성분은 성문을 통과하여 나오는 준주기적인 공기의 흐름과 성도를 스치며 발생되는 넓은 대역의 잡음으로 분류한다. 성도여파기의 응답은 일반적으로 천천히 변화하는데 주파수 영역에서 스펙트럼 신호의 포만트 포락선을 의미한다. 이는 지금까지 다음의 두가지 방법을 이용하여 구해지고 있다.

2.1 선형예측분석(LPC)법

인근한 음성 표본들은 높은 상관관계를 가지고 있다. 이러한 상관관계를 가정하여 간단한 선형예측을 다음과 같이 표현할 수 있다.

$$y_n \simeq \alpha_1 y_{n-1} + \alpha_2 y_{n-2} + \cdots + \alpha_p y_{n-p} \quad (1)$$

이 식에서 음성신호의 표본된 값(y_n)은 상수 α가 곱해진 과거의 p표본들에 의해서 예측할 수 있다는 가정을 보여주고 있다. 이 최소자승오차를 갖는 상수들을 선형예측계수라고 하고, 이 계수를 구하는 방법을 선형예측분석방법이라고 한다.

2.2 켑스트럼(Cepstrum)법

켑스트럼은 로그크기 스펙트럼의 역 푸리에 변환으로 정의되며 그 식은 다음과 같다.

$$c(\tau) = F^{-1}\log|X(w)| = F^{-1}\log|G(w)| + F^{-1}\log|H(w)| \quad (2)$$

여기서 x(t), g(t), h(t)를 각각 음성신호, 음원, 스펙트럼 포락함수라 할 때 이들을 푸리에 변환한 것이 X(w), G(w), H(w)이다. 켑스트럼은 주파수 영역의 함수를 역변환한 것이기 때문에 시간영역의 함수라고 할 수 있다. 켑스트럼이 가진 가장 큰 특징이라고 하면 음성이 갖는 정보에서 스펙트럼 포락정보와 세부 구조 정보를 분리해 낸다는 것이다. 켑스트럼의 낮은 시간대의 부분은 성도정보, 성문정보 그리고 입술의 방사정보를 가지고 있으며 높은 시간대의 부분은 여기성분을 갖는다.

3. 새로운 스펙트럼 평탄화 기법

음성신호는 FFT변환을 통해 주파수 영역에서 스펙트럼 분석이 이루어진다. 그림1은 본 논문에서 사용한 알고리즘의 블럭도이다. 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- 25ms인 것을 감안하여 300Hz와 400Hz를 서브밴드의 대역폭으로 사용하였다. 이는 입력음성에 따라 적응적으로 대처하기 위한 것이다. 다음 단계로 각각의 서브밴드에서 최대값을 취하여 프레임의 파라미터로 저장한다. 이 파라미터의 값은 8KHz 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 300Hz였을 경우 200Hz, 300Hz, 400Hz를 사용하고 400Hz였을 경우에는

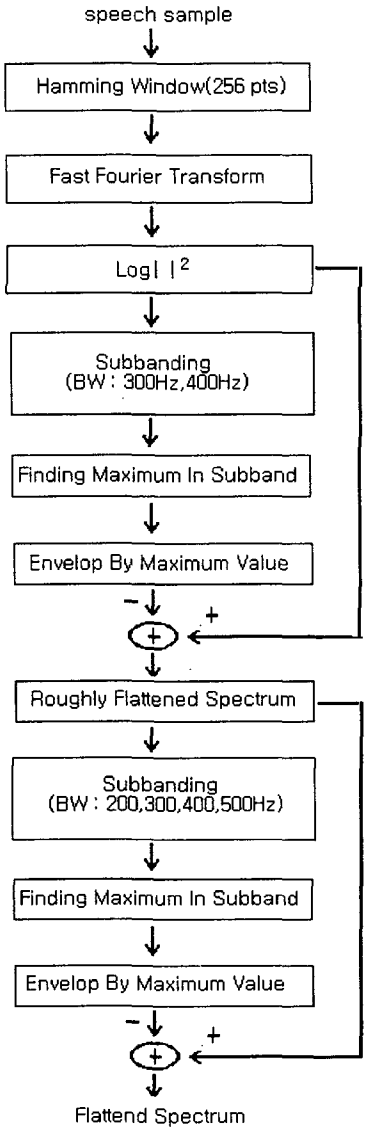

그림1. 제안한 알고리즘

300Hz, 400Hz, 500Hz를 사용했다.

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

$$Variance = \frac{2}{N}\sum_{k=1}^{N/2}(x(k)-m)^2 \quad (3)$$

여기서 N은 FFT포인터 수이고 스펙트럼 신호가 Y축으로 대칭이기 때문에 분산은 N/2까지만 이루어진다. 또한 k는 주파수 영역에서의 샘플인덱스이고 m은 평균을 의미한다. 이때 m값은 0을 사용하여 0을 기준으로 평탄화의 정도를 평가하였다.

그림 2는 남성화자의 음성신호와 그 스펙트럼 신호이다. 그림에서와 같이 일반적으로 남성화자의 음성신호는 시간영역에서 그 피치주기가 크고 주파수 영역의 스펙트럼 신호에서는 고조파 간격이 좁게 나타난다. 그림 3은 그림 2의 스펙트럼 신호를 평탄화한 결과이다. 그림에서 알 수 있듯이 LPC법과 Cepstrum법보다는 제안한

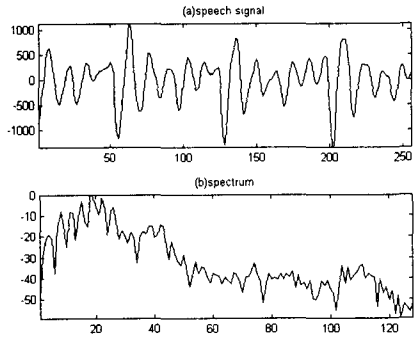

그림2. 남성의 음성신호
(a) 시간영역신호 (b) 로그스펙트럼 신호

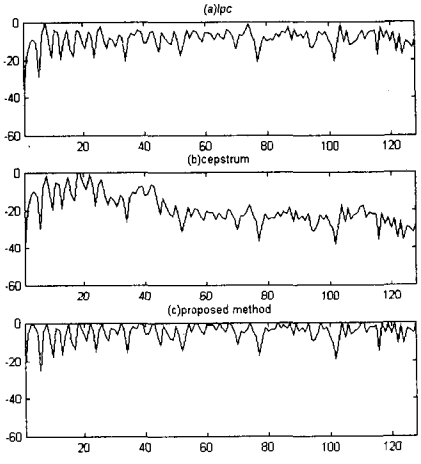

그림3. 평탄화된 스펙트럼 신호(남성)
(a) LPC method (b) Cepstrum method
(c) Proposed method

알고리즘이 훨씬 우수한 결과를 보이고 있다. 특히 주파수 영역의 양끝부분에서 좋은 성능을 보인다.

그림 4 는 여성화자의 음성신호와 그 스펙트럼 신호이다. 일반적으로 여성화자는 남성화자와는 반대로 음성신호의 피치주기가 작고 주파수 영역에서 스펙트럼 신호의 고조파 간격이 넓다. 그림 5는 그림 4의 스펙트럼 신호를 각각의 방법으로 평탄화한 결과들이며 역시 제안한 알고리즘이 가장 우수함을 알 수 있다.

4. 실험 및 결과

이상의 과정을 컴퓨터 시뮬레이션하기 위하여 IBM 펜티엄(III)에 마이크가 부착된 16-비트 A/D변환기를 인터페이스시키고, 아래의 문장들을 남녀 각 3명에게 발성시키면서 8kHz의 표본화 주파수로 표본화하여 저장한 다음에 시뮬레이션의 시료로 사용하였다:

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

표1은 남성화자의 발성별 분산값을 보여주고 있다. 결과값에서 보여지듯이 Cepstrum법이 가장 큰 분산값을 나타내고 LPC법은 양호한 특성을 보이지만 제안한 방법보다 약 1.5배 큰 분산값을 보이고 있다. 표 2는 여성화자의 발성별 분산값 비교인데 역시 제안한 방법이 가장 우수한 결과를 보여주고 있다.

표1. 남성화자의 분산값[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

표2. 여성화자의 분산값[dB]

	LPC	Cepstrum	New method
발성1	119.12	831.65	116.00
발성2	217.64	1099.38	139.73
발성3	206.46	1136.19	138.45
발성4	220.15	1253.43	144.09
Average	190.84	1080.16	134.57

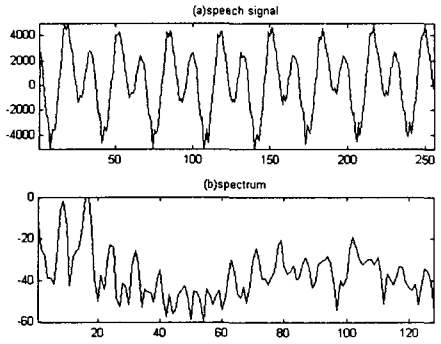

그림4. 여성의 음성신호
(a) 시간영역신호 (b) 로그스펙트럼 신호

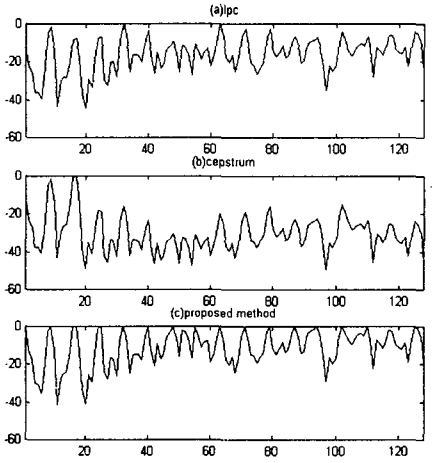

그림5. 평탄화된 스펙트럼 신호(여성)
(a) LPC method (b) Cepstrum method
(c) Proposed method

5. 결론

주파수 영역의 스펙트럼 신호는 잡음이 부가되는 경우에도 고조파정보와 포만트 포락선 정보를 유지하기 때문에 음성신호처리분야에서 매우 유용하다고 할 수 있다. 고조파 정보나 포만트 포락선 정보는 피치검출과 포만트 주파수 검출에 직접 이용되기 때문이다. 하지만 두 성분을 분리하는 방법에 따라 피치검출이나 포만트 주파수 검출에 영향을 미칠 수 있으므로 기존의 방법보다 두 성분을 더 잘 분리할 수 있는 방법이 필요한 것이다.

본 논문에서는 간단한 알고리즘으로 우수한 결과를 보이는 스펙트럼 평탄화 기법을 제안했다. 알고리즘은 주파수 대역을 서브밴드로 나누고 각 서브밴드에서 최대값을 취하여 파라미터화 하는 것인데 서브밴드의 대역폭과 선형 보간시 양 끝 부분의 보상법이 중요한 과정이다. 실험결과 제안한 알고리즘은 선형예측분석법보다 좋은 성능을 보였으며 정확한 피치검출과 포만트 검출에 적용될 수 있다.

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.