

LSP parameter의 위치정보를 이용한 성도특성 평탄화기법

김영규, *민소연, 배명진
승실대학교 정보통신공학과, *전자공학과
전화 : (02)824-0906, 팩스 : (02)820-0018

On the Flattening Techniques of Vocal track characteristics by using position information of the LSP (Line Spectrum Pairs)

YoungKyou Kim, *SoYeon MIN, MyungJin BAE
Dept. of Information and Telecommunication Engr, *Dept. of Electronics Engr,
Soongsil University
E-mail : skylimit@hanmail.net

요약문

음성신호는 성문특성으로 인해 고주파 특성이 약화되는 경향이 있다. 이를 보상하기 위해 Pre-emphasis filter를 사용한다. 수식으로 표현하면 $y(n)=s(n)-As(n-1)$ 와 같이 차분방정식으로 나타낼수 있다. 여기서 A값은 보통 0.9에서 1사이의 값을 주로 사용한다. 그러나 Pre-emphasis filter는 고주파 특성을 보상하는 과정에서 극점과 같이 영점도 왜곡된다. 본 논문에서는 음성특성에 따른 LSP(Line Spectrum Pairs) 분포특성을 이용하여 영점을 보존하고 vocoder 및 coding에 필연적인 고주파 특성 혹은 저주파 특성을 강조한다.

1. 서론

음성신호를 관찰했을 때 성문특성으로 인해 고주파쪽 특성이 약화되는 경향이 있다. 이를 보상하기 위해 Pre-emphasis 필터가 사용되어진다. Pre-emphasis 필터를 간단히 수식으로 표현하면 $y(n)=s(n)-As(n-1)$ 와 같이 차분방정식으로 나타낼 수가 있다[1][2]. 여기서 A 값은 보통 0.9에서 1사이의 값을 주로 사용하며 기존방법에는 $A = \frac{R_n(1)}{R_n(0)}$ 을 이용하여 보상한다. 본 논문에서는 주파수영역에서 포먼트의 위치를 찾아내어 대역폭을 결정하고 가중치를 주어 보상하는 것을 제안한다. 2장에서는 음성생성모델과 Pre-emphasis 필터에 대하여 설명

하고 3장에서는 본논문에 사용한 방법을 간단히 수식으로 살펴본다. 4장에서는 실험에서 도출한 결과와 5장에서는 결론을 맺는다.

II. 음성신호의 생성모델

2-1. 일반적인 음성생성모델

음성생성에 대한 선형모델은 50년대 후반 Fant에 의해 개발되었는데 그는 음성출력을 음원이 여파기를 통과하여 나오는 신호로 가정하고, 음원과 성도의 각 부분을 독립적인 것으로 간주하는 선형예측모델을 제시하였다. 음원에 대한 모델로 유성음의 음원은 준주기적인 펄스, 무성음의 음원은 백색잡음을 사용하였고, 성대에서 성문이 음원에 미치는 영향은 다음과 같은 성문모델로 모델링하였다[3][4].

$$G(z) = \frac{1}{(1 - e^{-cT} z^{-1})^2} \quad (2.1)$$

여기서, T는 준주기이고, cT는 감소인자이며 1보다 충분히 작다. 성문을 지난 신호는 성도를 거치면서 성도의 형태에 따라 몇 개의 공명주파수를 갖게 되는데 이 공명주파수와 대역폭을 2-극(pole) 여파기로 나타내면 성도에 대한 모델은 다음과 같이 구할 수 있다.

$$V(z) = \frac{1}{\prod^k [1 - 2e^{-cT} \cos(BiT)z^{-1} + e^{-2cT} z^{-2}]} \quad (2.2)$$

여기서, k 는 포먼트의 개수이고, c_i 와 B_i 는 포먼트 주파수와 대역폭을 결정하는 값으로 대역폭의 좁다는 가정에서 실제 포먼트의 공명주파수와 대역폭은 각각 $B_i/2\pi$, $c_i/2\pi$ 로 결정된다. 성도를 통과한 신호는 마지막으로 입술을 통과하는데 입술에서의 방사특성에 대한 모델은 다음과 같다.

$$L(z) = 1 - z^{-1} \quad (2.3)$$

음원에 대한 모델을 $E(z)$ 로 하면 음성출력 $X(z)$ 는 다음과 같다.

$$X(z) = E(z)G(z)V(z)L(z) \quad (2.4)$$

이를 전극(all-pole)형 합성모델로 다시 정의하면,

$$X(z) = E(z) \frac{1}{A(z)} \quad (2.5)$$

여기서,

$$A(z) = \frac{1}{G(z)V(z)L(z)} \quad (2.6)$$

이 $A(z)$ 는 다음 관계식에 의해 음성으로부터 성도특성을 제거하고 음원을 이끌어 내는 가역여파기임을 알 수 있다.

$$E(z) = X(z)A(z) \quad (2.7)$$

이상에서 살펴본 것과 같이 음성을 음원과 그 음원이 통과하는 성도여파기로 모델링하고 각각을 독립적으로 모델링함으로써 수학적으로 선형방정식의 해를 구할 수 있다[3][4].

2-2. PRE-EMPHASIS 필터의 특징

음성신호처리 분야에서 스펙트럼 경사를 평탄화해 줌으로써 신호의 동적범위를 억제하는 프리엠퍼시스 과정은 SNR을 높이는 데 유효한 것으로 알려져 있다[1][2]. 이 방법은 일반적으로 A/D 변환을 위한 저역통과 필터링에 앞서서 20dB/decade 정도의 고주파 영역을 강조하는 역할을 한다. 이 방법은 또한 A/D 변환 다음에도 수행되어질 수 있는데, 차분방정식이나 식 (2.8)과 같은 1차 디지털 필터링을 통하여 구현된다.

$$H(z) = 1 - az^{-1} \quad (2.8)$$

여기에서 a 값은 대략 1에 가까운 값을 갖는다. SNR을 가능한 높이기 위해서는 A/D 변환을 하기 앞서 프리엠퍼시스를 해야한다. 그리고 원신호의 스펙트럼 경사를 복구하기 위한 과정을 디엠퍼시스(de-emphasis)라고 한다.

III. 주파수영역에서의 시스템구현

3-1. LSP 파라미터 추출 및 구현

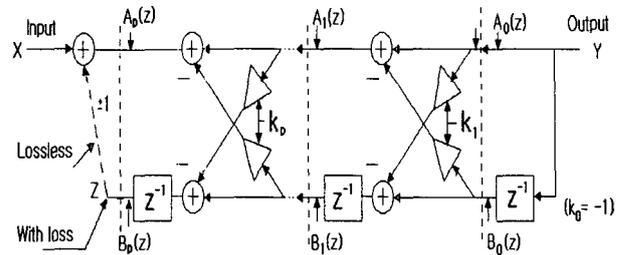


그림 3-1. PARCOR structure of LPC synthesis

LSP 파라미터를 추출하기 위해서 먼저 LPC 분석이 이루어 진다[5].

$$H(z) = 1/A_p(z) \quad (3.1)$$

$$\text{where } A_p(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (3.2)$$

$H(z)$ 는 LPC 필터이고 p 는 필터의 차수이다.

LSP 파라미터를 유도하기 위해서 PARCOR (Partial Correlation) 필터를 이용하여 식(3.1)과 식(3.2)를 표현하면 다음과 같다.

$$A_{p-1}(z) = A_p(z) + k_p B_{p-1}(z) \quad (3.3)$$

$$B_p(z) = z^{-1} [B_{p-1}(z) - k_p A_{p-1}(z)]$$

여기서 $A_0(z) = 1$, $B_0(z) = z^{-1}$ 이고

$$B_p(z) = z^{-(p-1)} A_p(z^{-1}) \quad (3.4)$$

그림 3-1의 PARCOR 구조는 손실이 없는 음파관에서 음파의 전달로 이해된다. 시스템은 단지 역방향(backward) 에너지 모양에서 Z 종점에서 손실이 있다. 이러한 음관은 Z 종점의 출력이 $k_{p-1} = \pm 1$ 의 경로를 통해 입력의 종점으로 귀환될 때 완전한 무손실이 된다. 각각의 공명 값인 Q 는 무한해지고 에너지 분포 스펙트럼은 몇 개의 선 스펙트럼에 집중된다. $k_{p-1} = -1$ 조건의 귀환은 입력종점에서 완전히 폐쇄되고 $k_{p-1} = +1$ 은 무한 자유공간상으로 개방된다[5][6].

그림 3-1에서 $k_{p-1} = \pm 1$ 인 전달함수를 $P_{p-1}(z)$ 와 $Q_{p-1}(z)$ 로 나타내면 다음과 같다.

$$k_{p-1} = 1 \text{ 일때, } P_{p-1}(z) = A_p(z) - B_p(z) \quad (3.5)$$

$$k_{p-1} = -1 \text{ 일때, } Q_{p-1}(z) = A_p(z) + B_p(z)$$

$$\Rightarrow A_p(z) = \frac{1}{2} [P_{p-1}(z) + Q_{p-1}(z)] \quad (3.6)$$

두 개의 근($k_{p-1} = \pm 1$)을 알고 있으므로 $P_{p-1}(z)$ 의

$Q_{p-1}(z)$ 의 차수를 줄일 수 있다. 즉,

$$P'(z) = \frac{P_{p-1}(z)}{(1-z)} = A_0 z^p + A_1 z^{(p-1)} + \dots + A_p \quad (3.7)$$

$$Q'(z) = \frac{Q_{p-1}(z)}{(1-z)} = B_0 z^p + B_1 z^{(p-1)} + \dots + B_p \quad (3.8)$$

$$\text{조건 : } A_0=1, B_0=1 \quad (3.9)$$

$$A_k = (a_k - a_{p-1-k}) + A_{k-1} \quad (3.10)$$

$$B_k = (a_k - a_{p-1-k}) - A_{k-1} \quad \text{for } k=1, \dots, p$$

LSP는 $0 \leq \omega_i \leq \pi$ 인 범위에서 $P'(z)$ 와 $Q'(z)$ 을 통해 얻어진 근의 각(angular) 위치를 나타낸다. LSP는 다음과 같은 두가지 성질을 지닌다.

첫째, $P'(z)$ 와 $Q'(z)$ 는 단위원 상에 놓여 있다.

둘째, $P'(z)$ 와 $Q'(z)$ 의 근들이 단위원 상에 번갈아 나타난다.

3-2. Real Root Method

$P'(z)$ 와 $Q'(z)$ 의 다차 방정식의 해를 구하는 방식에 따라 개발된 여러 가지 변환법 중 주로 사용되는 Real Root 방법의 특징은 다음과 같다[5].

$P'(z)$ 와 $Q'(z)$ 의 계수는 대칭적이기 때문에 식(3.7)의 차수는 $p/2$ 로 줄어든다.

$$P'(z) = A_0 z^p + A_1 z^{p-1} + \dots + A_1 z^1 + A_0 = z^{p/2} [A_0 (z^{p/2} + z^{-p/2}) + A_1 (z^{(p/2-1)} + z^{-(p/2-1)}) + \dots + A_{p/2}] \quad (3.11)$$

$$Q'(z) = B_0 z^p + B_1 z^{p-1} + \dots + B_1 z^1 + B_0 = z^{p/2} [B_0 (z^{p/2} + z^{-p/2}) + B_1 (z^{(p/2-1)} + z^{-(p/2-1)}) + \dots + B_{p/2}] \quad (3.12)$$

모든 근은 단위원 상에 있기 때문에, 아래와 같이 정의하면 단위원 상에서 식(3.11)의 값을 구할 수 있다.

$$\text{Let } z=e^{j\omega} \text{ then } z^1+z^{-1}=2 \cos(\omega) \quad (3.13)$$

$$P'(z) = 2e^{j\omega p/2} [A_0 \cos(\frac{p}{2}\omega) + A_1 \cos(\frac{p-2}{2}\omega) + \dots + \frac{1}{2} A_{p/2}] \quad (3.14)$$

$$Q'(z) = 2e^{j\omega p/2} [B_0 \cos(\frac{p}{2}\omega) + B_1 \cos(\frac{p-2}{2}\omega) + \dots + \frac{1}{2} B_{p/2}] \quad (3.15)$$

$x = \cos \omega$ 를 대입하면 식(3.14)와 식(3.15)을 x 에 대해서

풀 수 있다. $p=10$ 의 경우 다음식이 얻어진다.

$$P'_{10}(x) = 16A_0 x^5 + 8A_1 x^4 + (4A_2 - 20A_0)x^3 + (2A_3 - 8A_1)x^2 + (5A_0 - 3A_2 + A_4)x + (A_1 - A_3 + 0.5A_5) \quad (3.16)$$

$$Q'_{10}(x) = 16B_0 x^5 + 8B_1 x^4 + (4B_2 - 20B_0)x^3 + (2B_3 - 8B_1)x^2 + (5B_0 - 3B_2 + B_4)x + (B_1 - B_3 + 0.5B_5) \quad (3.17)$$

LSP는 식(3.18)에 의해서 구해진다.

$$LSP(i) = \frac{\cos^{-1}(x_i)}{2\pi T}, \quad \text{for } 1 \leq i \leq p \quad (3.18)$$

Real Root 방법은 다른 변환 방법보다 간단하지만 계산 시간을 예상할 수 없다는 단점을 갖는다[5]

3-3. 제안한 방법

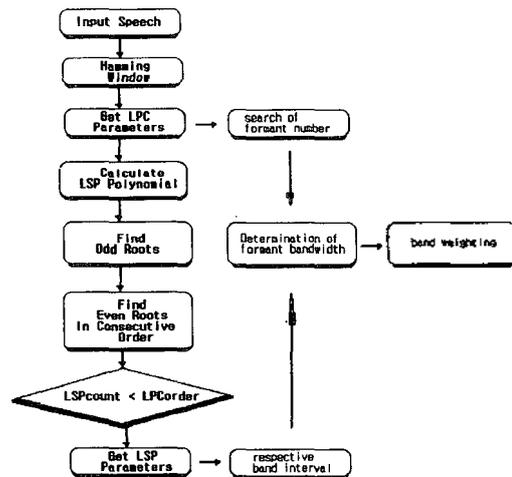


그림 3-2. 제안한방법의 블록도

LPC 파라미터의 특성을 이용하여 포먼트의 수와 크기를 저장하고 LSP 파라미터를 추출하여 파라미터의 위치정보를 이용하여 인근 파라미터의 거리를 식(3.19)과 같이 계산하여 포먼트 주파수의 위치의 대역폭을 찾아낸다.

$$D(p) = LSP(i+1) - LSP(i) \quad (3.19)$$

찾아낸 대역폭에서 LPC를 이용해 얻은 가장 높은 크기

를 가진 주파수에서의 크기(A)와 약화된 고주파 및 저주파크기(B)의 나중값(A/B)으로 대역 가중치를 주어 평탄화시킨다.

IV. 실험 및 결과

본 논문의 모의 실험하기 위해 IBM PC에 마이크 입력이 가능한 A/D 변환기를 인터페이스 시켰다. 남녀 화자가 각각 발성한 음성시료를 11KHz로 샘플링하고 16비트로 양자화하여 사용하였다. 한프레임의 길이는 256샘플이며 계수는 10차를 써서 특징벡터를 추출하였다.

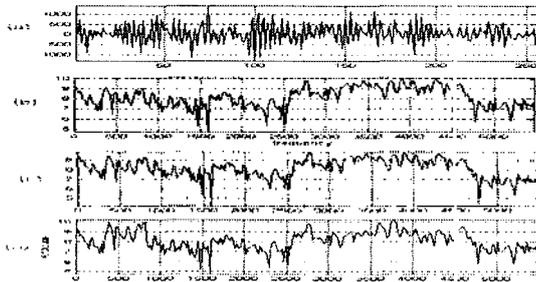


그림 4-1. 무성음의 스펙트럼 분석

- (a) 음성신호 (b) 원음성의 스펙트럼분석
- (c) 기존방법의 보상된 신호의 스펙트럼분석
- (d) 제안한 방법의 보상된 신호의 스펙트럼분석

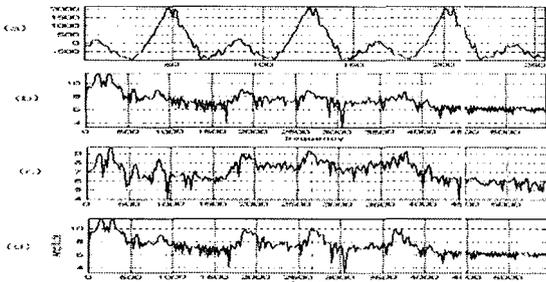


그림 4-2. 유성음의 스펙트럼 분석

- (a) 음성신호 (b) 원음성의 스펙트럼분석
- (c) 기존방법의 보상된 신호의 스펙트럼분석
- (d) 제안한 방법의 보상된 신호의 스펙트럼분석

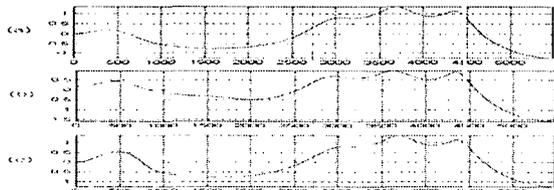


그림 4-3. 무성음의 LPC분석

- (a) 원음성의 LPC분석
- (b) 기존방법의 보상된 신호의 LPC분석
- (c) 제안한 방법의 보상된 신호의 LPC분석

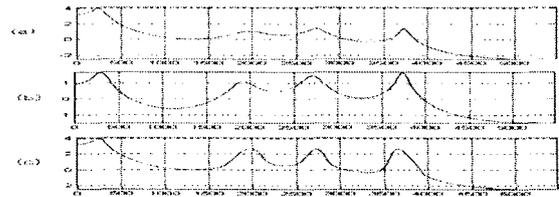


그림 4-4. 유성음의 LPC분석

- (a) 원음성의 LPC분석
- (b) 기존방법의 보상된 신호의 LPC분석
- (c) 제안한 방법의 보상된 신호의 LPC분석

그림 4-1과 그림 4-2는 무성음과 유성음의 음성신호와 기존 방법을 이용하여 보상한 신호와 제안한 방법을 이용한 것을 스펙트럼 비교 분석한 것이다.

그림 4-3과 그림 4-4는 무성음과 유성음의 음성신호와 기존 방법을 이용하여 보상한 신호와 제안한 방법을 이용한 것을 LPC 비교 분석한 것이다.

V. 결론

음성신호를 관찰했을 때 고주파 특성이 약화되는 경향이 있는데 이를 보상하기 위해 기존의 방법으로는 Pre-emphasis filter를 사용한다. 그러나 기존방법의 Pre-emphasis filter는 보상하는 과정에서 극점과 같이 영점도 왜곡된다. 이는 음성 복원과정에서 투박한 음질을 나타낸다. 본 논문에서는 음성의 프레임별 LSP 분포 특성을 이용하여 포먼트의 위치의 대역폭을 결정하고 가중치를 주어서 극점과 영점중 영점은 보존하고 음성 전달 및 인식에 중요한 극점을 강조하여 주파수 영역의 포락선 기울기를 평탄화 줌으로서 SNR을 높이고 음질 향상에 유효하다. 이 방법은 일반적으로 A/D 변환을 위한 저역통과 필터링에 앞서서 20dB/decade 정도의 고주파 영역을 강조하는 역할을 한다. 실험 결과, 각각의 프레임 구간 특성에 따라 고주파 특성, 혹은 저주파 특성을 강조함을 알수가 있는데 분석을 위해 LPC 분석을 한 결과, 좀더 성도특성을 제대로 tracking하고 있음을 반영한다.

참고문헌

- [1] 오영환, "음성언어정보처리", 홍릉과학출판사, 1998.
- [2] Douglas O. shaughnessy, "Speech Communication", IEEE Press, 1996.
- [3] L. R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signal", Prentice Hall, 1978.
- [4] 배명진, "디지털 음성분석", 동영출판사, 1998. 4.
- [5] A.M. Kondoz, Digital Speech, Jhon wiley & Sons, 1994.
- [6] L. R. Rabiner & Bing-Hwang Juang, Fundamentals Of Speech Recognition, Prentice-Hall AT&T, U.S.A, 1993.