2001年度 第14回 信號處理合同學術大會 論文集 第14卷 1號

Frequency Scaling을 통한 LSP 파라미터 Fitting에 관한 연구

*민소연, 배명진
*숭실대학교 전자공학과, 정보통신공학과

A Study on the Fitting of LSP(Line Spectrum Pairs) Parameter using Frequency Scaling

*SoYeon MIN, MyungJin BAE
Dept. of *Electronics, Information and Telecommunication Engr., Soongsil University
E-mail :pasternak@hanmail.net
mjbae@saint.ssu.ac.kr

요 약

LSP 파라미터는 음성코덱(codec)이나 인식기에서 음성 신호를 분석하여 전송형이나 저장형 파라미터로 변환되어, 주로 저전송률 음성부호화기에 사용된다. 그러나 LPC 계수를 LSP로 변환하는 방법이 복잡하여 계산시간이 많이 소요된다는 단점이 있다[1]. 기존의 LSP 변환 방법 중 음성 부호화기에서 주로 사용하는 real root 방법은 근을 구하기 위해 주파수 영역을 순차적으로 검색하기 때문에 계산시간이 많이 소요되는 단점을 갖는다. 본 논문에서 비교 평가한 알고리즘은 첫 번째, 기존의 real root 알고리즘, 두 번째는, LSP 파라미터의 분포 특성을 조사하여 이를 토대로 검색구간의 순서와 검색간격을 달리한 경우, 세 번째는 검색 시 mel scale을 사용한 알고리즘이다. 실험결과, 기존의 real root 방식에 비하여 두 가지 방식 모두가 변환시간의 40% 이상이 감소되는데 반하여 동일한 근을 찾음을 알 수가 있었고, 특히 분포특성을 이용하여 검색순서와 간격조절을 한 경우에 있어서, 기존의 방식보다 48%이상이 감소되었다.

I. 서 론

LSP 파라미터는 양자화 에러에 강하고 시스템의 안정성과 선형 보간성이 뛰어난 장점을 갖는다. 그러므로 음성 부호화기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환하는데 LSP 파라미터가 사용된다. LPC(Linear Predictive Coding) 계수를 LSP 파라미터로 변환하기 위한 다항식의 근을 찾는 방법에는 complex root, real root, ratio filter, chebyshev series, adaptive sequential LMS 방법 등이 있다. 이중에서 real root 방법이 간단하고 이해하기 쉽기 때문에 가장 많이 이용된다. real root 방법의 특징은 홀수번째 근을 먼저 찾은 후에, 찾아진 홀수번째 근 사이에서 짝수번째 근을 찾게 된다. 그러므로 전체적인 검색시간은 홀수번째 근을 찾는 시간에 의해 좌우가 된다[2].

LSP 홀수번째 계수들의 분포특성을 관찰한 결과 특정 주파수 영역에서 주로 나타나고 나머지 영역에서는 거의 나타나지 않았다. 본 논문에서는 LSP 근을 찾기 위한 몇 가지 방법들의 성능평가에 초점을 맞추어 보았다. 본 논문에서 고려한 첫 번째 방법은 분포특성을 이용하여 검색순서와 간격을 조절한 방법이고, mel scale을 이용하여 주파수 대역을 검색한 방법이 두 번째 방법에 해당된다. 본 논문의 구성은 다음과 같다. 2장에서는 LPC와 LSP 파라미터의 일반적인 특징에 대하여 살펴본다. 3장에서는 real root 방법과 본 논문에서 고려한 두 가지 방법의 특징에 대하여 설명한다. 그리고 4장에서는 실험 결과를 분석하고, 5장에서는

결론을 맺는다.

II. LPC 파라미터의 특징

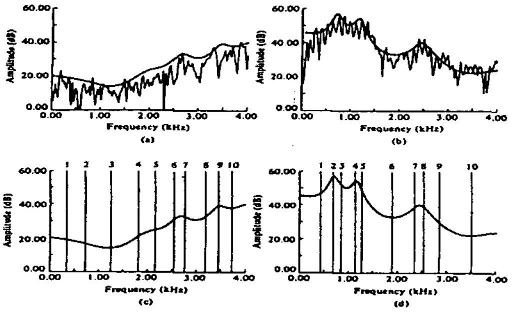

그림 1. LSP 파라미터
(a) 자음 /s/ (b) 모음 /a/
(c),(d) /s/와 /a/에 대한 LPC 분석과 LSP

LSP 파라미터는 음성 신호의 포만트 스펙트럼 정보를 반영한다. 유성음의 제 I 포만트(F_1)는 비교적 낮은 주파수 대역에서 좁은 대역폭을 갖는다. LSP 파라미터에서 선쌍(line pair)의 간격이 좁으면, 강한 공명이 일어난 것을 나타낸다. 즉, LSP는 p개의 불연속적인 주파수의 분포를 통하여 음성의 스펙트럼 포락선을 표현한다. 그림 1. (d)의 2, 3번째와 4, 5번째는 제 1포만트와 제 2포만트를 나타낸다. 따라서 포만트의 위치 및 분포 특성은 LSP 파라미터에 직접적인 영향을 미치게 된다[7].
또한 PARCOR 구조에서 $k_{p+1}=\pm 1$인 전달함수를 $P_{p+1}(z)$와 $Q_{p+1}(z)$로 나타내면 다음과 같다.

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

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

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,

$$P'(z)=\frac{P_{p+1}(z)}{(1-z)} = A_0 z^p + A_1 z^{(p-1)} + \cdots + A_p \tag{3}$$

$$Q'(z)=\frac{Q_{p+1}(z)}{(1-z)} = B_0 z^p + B_1 z^{(p-1)} + \cdots + B_p \tag{4}$$

$$\text{조건 :} \quad A_0=1 \ , \ B_0=1 \tag{5}$$

$$A_k=(\alpha_k-\alpha_{p+1-k})+A_{k-1} \tag{6}$$
$$B_k=(\alpha_k-\alpha_{p+1-k})-A_{k-1} \quad \text{for} \quad k=1,\cdots p$$

LSP는 $0\le\omega_i\le\pi$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.

III. LSP 파라미터의 Fitting법

III-1. Real Root 방법의 특징

$P'(z)$와 $Q'(z)$의 계수는 대칭적이기 때문에 식(6)의 차수는 $p/2$로 줄어든다. 그리고, 모든 근이 단위원 상에 있기 때문에, 아래와 같이 정의하면 단위원 상에서 값을 구할 수 있다.

$$Let \quad z=e^{j\omega} \quad then \quad z^1+z^{-1}=2\cos(\omega) \tag{7}$$

$$P'(z)=2e^{jp\omega/2}[A_0\cos(\frac{p}{2}\omega)+ A_1\cos(\frac{p-2}{2}\omega)+\cdots+\frac{1}{2}A_{p/2}] \tag{8}$$

$$Q'(z)=2e^{jp\omega/2}[B_0\cos(\frac{p}{2}\omega)+ B_1\cos(\frac{p-2}{2}\omega)+\cdots+\frac{1}{2}B_{p/2}] \tag{9}$$

$x=\cos\omega$를 대입하면 식(8)과 식(9)를 x에 대해서 풀 수 있다. $p=10$의 경우 다음식이 얻어진다[4][7].

$$P'_{10}(x)=16A_0x^5+8A_1x^4+(4A_2-20A_0)x^3 +(2A_3-8A_1)x^2+(5A_0-3A_2+A_4)x +(A_1-A_3+0.5A_5) \tag{10}$$

$$Q'_{10}(x)=16B_0x^5+8B_1x^4+(4B_2-20B_0)x^3 +(2B_3-8B_1)x^2+(5B_0-3B_2+B_4)x +(B_1-B_3+0.5B_5) \tag{11}$$

LSP는 식(12)에 의해서 구해진다.

$$LSP(i)=\frac{\cos^{-1}(x_i)}{2\pi T}, \quad \text{for} \quad 1\le i\le p \tag{12}$$

III-2. LSP 파라미터의 분포특성을 이용하여 검색 간격을 조절한 방법

앞에서 설명한 real root 방법에서는 홀수번째 계수를 먼저 찾은 후 짝수번째 계수는 이미 찾아진 홀수번째 계수 사이에서 찾는다. 따라서 홀수번째 계수를 찾는데 걸리는 시간이 변환시간의 대부분을 차지하게 된다.
그림 2는 음성신호에서 LSP 파라미터의 홀수 번째 계수의 분포도를 나타낸 것으로, 8kHz 표본화율에 10차의

선형예측계수를 사용하였다. 그림 2에서 LSP 홀수번째 계수들은 특정 주파수 영역에서 주로 나타나고 나머지 영역에서는 거의 나타나지 않음을 보인다. real root 방법에서는 홀수 번째의 계수를 찾을 때 주파수대역을 순차적으로 검색한다. 그러나 그림 2의(c), (d), (e)에서는 분포도가 순차적이지 않음을 보인다. 따라서 검색 순서를 파라미터가 많이 나타나는 주파수 대역부터 검색한다면 검색시간을 단축할 수 있다.

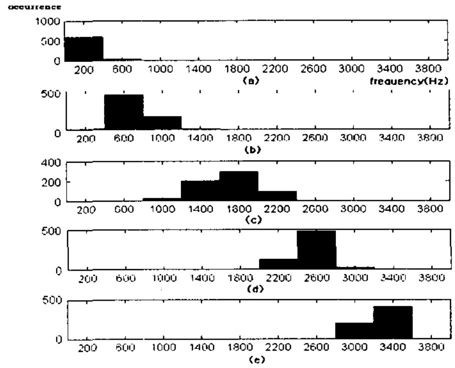

그림 2. LSP 홀수번째 계수의 분포도
(a) 1 번째 LSP 계수 (b) 3 번째 LSP 계수
(c) 5 번째 LSP 계수 (d) 7번째 LSP 계수
(e) 9 번째 LSP 계수

표 1. 주파수 대역별 검색순서와 검색간격

검색순위 (Hz) / LSP계수	1	2	3	4	5
LSP 1	0-400 (5)	400-800 (5)	otherwise		
LSP 3	400-800 (5)	800-1200 (5)	otherwise		
LSP 5	1600-2000 (5)	1200-1600 (10)	800-1200 (20)	2000-2400 (40)	otherwise
LSP 7	2400-2800 (5)	2000-2400 (10)	1600-2000 (20)	otherwise	
LSP 9	2800-3200 (5)	3200-3600 (10)	otherwise		

본 논문에서 LSP 파라미터의 fitting을 위해 고려한 방식 중에서 두 번째 방식은 다음과 같은 특징을 이용하였다. LSP 파라미터의 분포특성은 순차적이지가 않다. 즉, 어느 특정 주파수 대역에 편중되어서 나타난다. 그러므로 분포특성을 이용하여 분포도가 높은 곳에서는 조밀한 간격으로 검색하고 낮은 분포특성을 보이는 주파수 대역에서는 해상도를 낮추어 검색하는 방법을 취한다면 LSP 파라미터의 근을 구하는데 있어서 계산시간을 줄이게 된다. 표 1에서는 주파수 대역별 검색순서와 검색간격을 나타내었다. 표 1에서 보여지듯이 다섯번째 LSP 파라미터를 찾을때, 분포도가 가장 높은 주파수 대역인 1600~2000Hz 대역은 5Hz 단위로 검색하고 1200~1600 Hz에서는 10Hz 단위로 검색한다. 그리고 분포도가 3, 4번째인 주파수 대역에서는 검색간격을 각각 20Hz, 40Hz로 조정한다.

III-3. Mel Scale을 이용한 LSP 변환 알고리즘의 성능 개선

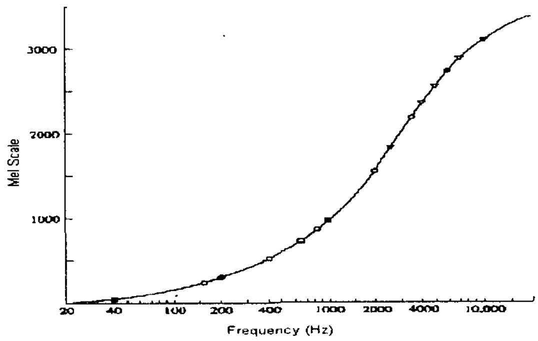

그림 3. 본 논문에서 적용한 Mel Scale

mel scale이란 인간의 청각특성을 고려한 주파수 척도이며 1000Hz에 1000Mel을 대응하고 이로부터 실험적으로 값을 결정한다. 일반적으로 그림 3과 같이 mel 척도는 1000Hz 이하에서는 주파수와 선형적으로 비례하며 1000Hz 이상에서는 log scale이다. 본 논문에서는 Fant에 의하여 근사된 식을 이용하였고 식 (13)에서 나타내었다. 식 (13)은 mel scale과 주파수 사이의 관계를 나타내고 있다[2][5][8].

$$F_{mel} = k * \log[1 + \frac{f}{1000}] \qquad (13)$$

여기서,

$$k = 1000/\log 2$$

식 (13)에서 나타낸 mel scale과 주파수의 관계를 그림 3에서 도시하고 있다.

IV. 실험 및 결과

실험에 사용된 음성시료는 연구실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화하여 사용하였다. 실험에 사용한 발성한 음성시료는 다음과 같다.

발성1: "인수네 꼬마는 천재소년을 좋아한다."

발성2: "창공을 날으는 인간의 도전은 끝이 없다."

발성3: "예수님께서 천지창조의 교훈을 말씀하셨다."

발성4: "숭실대학교 음성통신 연구팀이다"

본 논문에 사용한 real root 알고리즘은 C-언어로 구현된 CELP 부호화기에서 발췌하여 사용하였다[4].

표 2. Real Root Vs. 검색간격 조절 방법

비교 / 시료	Real Root 방법 (단위:sec)	비교평가 방법 1 (단위:sec)	감소율 (단위:%)
발성 1	80.09	40.35	49.62
발성 2	93.94	46.83	50.15
발성 3	104.27	53.44	48.75
발성 4	85.3	45.82	46.28

표 3. Real Root Vs. Mel scale을 이용한 방법

비교 / 시료	Real Root 방법 (단위:sec)	비교평가 방법 2 (단위:sec)	감소율 (단위:%)
발성 1	80.09	41.56	48.11
발성 2	93.94	47.62	49.31
발성 3	104.27	55.84	46.45
발성 4	85.3	47.09	44.79

표 2는 real root방법과 LSP 분포특성을 이용하여 검색순서와 간격을 조절한 방법의 계산시간 비교결과 이고 표3은 real root 방법과 mel scale을 이용하여 계산시간을 단축한 방법의 비교 결과이다. 표 2의 경우에 있어서 제안한 방법 1이 약 48.7%의 계산시간이 단축되었고, 표 3의 경우에 있어서 제안한 방법 2가 평균 47.17% 정도의 계산시간이 단축되었음을 보인다. 그림 4의 (c), (d), (e)는 각각, real root 방식으로 찾은 LSP 파라미터, 비교 평가한 첫 번째 방법, 두 번째 방법에 해당한다. 그림 4에서 보듯이 세 가지 방법에서 모두 동일한 근을 찾음을 알 수가 있다.

V. 결 론

저전송률 음성부호화기에서의 음성신호 선형예측 방법 중 LSP 파라미터를 이용하는 방법이 가장 많이 사용되어진다. 이것은 LSP 파라미터의 전송형 특징 중 낮은 전송률에서도 왜곡이 적고 선형보간 특성이 뛰어나기 때문이다. 하지만 LPC 계수를 LSP 파라미터로 변환하기 위해서는 많은 계산시간이 소요된다. 본 논문에서는 음성부호화기에서 주로 사용되는 real root 방법과 제안한 두가지 방법 간에 계산시간 단축에 관하여 관찰하였다. 즉, 첫 번째 제안한 방법은 LSP 분포특성을 이용하여 검색순서와 간격 조절을 통하여 평균 48.7%의 계산시간을 단축하였고, 두 번째 mel scale을 이용한 경우에 있어서 평균 47.17%가 단축되었다. 그러나 세 가지 방법 모두가 동일한 LSP 파라미터를 찾음을 알 수가 있었다. 향후, 제안한 알고리즘을 음성부호화기에 적용하여 부호화기의 전체적 성능향상에 대한 연구가 이루어져야겠다.

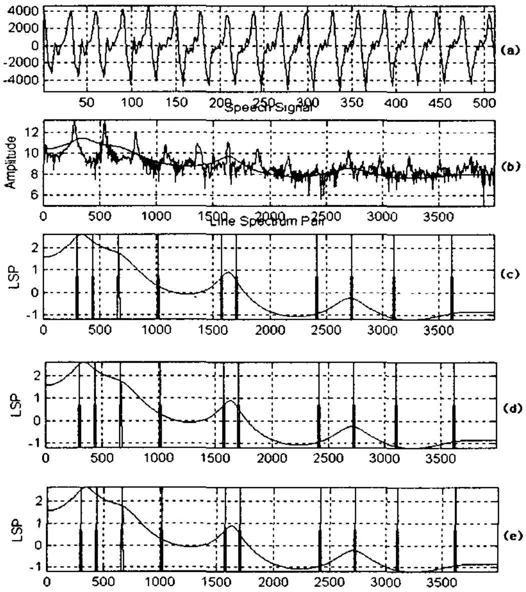

그림 4. 실험에서 얻어진 LSP 파라미터의 분포
(a) 음성 파형 (b) 음성신호의 스펙트럼
(c) Real Root 방법
(d) 분포특성과 검색간격 조절을 이용한 방법
(e) Mel Scale을 이용한 방법

참고문헌

[1] L R. Rabiner, R.W Schafer, " Digital Processing of Speech Signal", Prentice Hall, 1978.

[2] 배명진,"디지탈 음성분석", 동영출판사, 1998. 4.

[3] Oppenheim, Schafer,"Discrete Time Signal Processing", Prentice Hall, 1989.

[4] Emanuel C. Ifeachor,"Discrete Time Signal Processing", Addison Wesley,1993.

[5] 오영환, "음성언어정보처리",홍릉과학출판사,1998.

[6] Douglas O, shaughnessy,"Speech Communication",IEEE Press,1996.

[7] A. M. Kondoz,"Digital Speech", John Wiley & Sons Ltd, 1994.

[8] 민소연, 정찬중, 배명진, "주파수 대역의 적응적 조절을 통한 LSP 변환 알고리즘의 성능 개선에 관한 연구", 한국통신학회, 하계종합학술 발표회 논문집, 2001년 7월.