

LSP와 포먼트 분포특성을 이용한 음성신호의 LSP 변환 알고리즘의 성능개선

이 회 원, 나 덕 수, 김 만 기, 배 명 진
 숭실대학교 정보통신공학과
 전화 : (02) 824-0906 / 팩스 : (02) 820-0018

An Improved LSP Transformation Algorithm Using the LSP and Formant Distribution in a Speech Signal

HeeWon LEE, DuckSu NA, ManKi KIM, MyungJin BAE
 Dept. Information and Telecommunication Engr., Soongsil University
 E-mail : hwlee@assp.soongsil.ac.kr, mjbbae@saint.soongsil.ac.kr

Abstract

LSP 파라미터는 일정한 스펙트럼 민감도와 낮은 스펙트럼 왜곡을 보이고 선형보간이 용이하여 주로 저전송률 음성부호화에 사용된다. 그러나 LPC 계수를 LSP로 변환하는 방법이 복잡하여 계산시간이 많이 소요된다는 단점이 있다[1].

본 논문에서는 LSP 변환 방법 중 음성 부호화기에서 주로 사용하는 real root 방법의 계산 시간 단축방법을 제안한다. real root 방법은 다항식의 근을 구하여 LSP로 변환하는 방법이다. 그러나 이 방법은 근을 구하기 위해 주파수 영역을 순차적으로 검색하기 때문에 계산시간이 많이 소요된다. 제안하는 알고리즘은 LSP 파라미터의 분포도에 따라 검색구간의 순서를 주파수 별로 달리 한다. 그리고 제 1포먼트와 제 2포먼트의 연관성을 고려하여 검색구간을 조절한다. 기존의 real root 방법과 제안한 방법을 비교한 결과 평균 48% 이상의 검색시간이 단축되었다.

I. 서 론

현재 사용되는 음성부호화기나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환하는데 LSP 파라미터가 사용된다. LSP 파라미터는 양자화 에러에 강하고 시스템의 안정성이 보장된다.

LPC 계수를 LSP 파라미터로 변환하려면 먼저 다항식의 근을 찾는 과정이 필요하다. 기존의 방법으로는 complex root, real root, ratio filter, chebyshev series, adaptive sequential LMS 방법 등이 있다[3].

그 중에서 가장 많이 사용되는 것이 real root 방법이다. real root 방법은 LPC를 LSP로 변환하는 방법 중 비교적 간단하고 이해하기 쉬워 주로 사용된다[3]. 그러

나 이 방법도 계산이 복잡하여 변환시간이 많이 소요되는 단점을 가지고 있다[1].

본 논문에서는 음성신호의 LSP의 분포와 포먼트의 특징을 이용하여 real root 방법의 계산 시간을 단축하는 방법을 제안한다.

먼저 II장에서는 LSP 파라미터의 추출 방법과 근을 찾는 방법에 대해 살펴보고, III장에서는 음성신호의 LSP 분포와 유성음에서 포먼트의 위치 및 분포특성에 따른 제안한 변환 방법을 설명한다. 그리고 IV장에서는 제안한 방법의 실험 및 결과를 설명하고, V장에서는 결론을 맺겠다.

III. LSP 파라미터 추출

LSP 파라미터를 추출하기 위해서 먼저 LPC(Linear Predictive Coding) 분석이 이루어져야 한다[3].

$$H(z) = 1/A_p(z) \quad (2.1)$$

$$\text{where } A_p(z) = 1 + \sum_{k=1}^p a_k z^{-k} \quad (2.2)$$

$H(z)$ 는 LPC 필터이고 p 는 필터의 차수이다. LSP 파라미터를 유도하기 위해서 PARCOR(Partial Correlation) 필터를 이용해서 식(2.1)과 식(2.2)를 표현하면 다음과 같다.

$$A_{p-1}(z) = A_p(z) + k_p B_{p-1}(z) \quad (2.3)$$

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

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

$$B_p(z) = z^{-(p+1)} A_p(z^{-1}) \quad (2.4)$$

그림 2-1에 보이는 것처럼 PARCOR 구조는 손실이 없는 음파관에서 음파의 전달로 이해된다. 시스템은 단

지 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 Q는 무한해지고 에너지 분포 스펙트럼은 몇 개의 선 스펙트럼에 집중된다[5][6].

$k_{p+1} = -1$ 조건의 귀환은 입력종점에서 완전히 폐쇄되고 $k_{p+1} = +1$ 은 무한 자유공간상으로 개방된다[1][3].

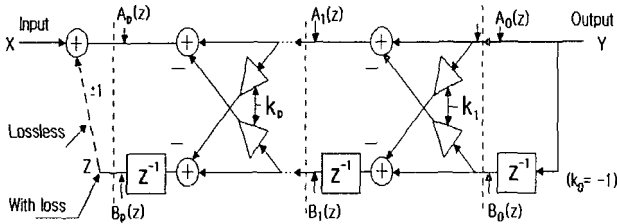


그림 2-1. PARCOR structure of LPC synthesis

그림 2-1.에서 $k_{p+1} = \pm 1$ 인 전달함수를 $P_{p+1}(z)$ 와 $Q_{p+1}(z)$ 로 나타내면:

$$\begin{aligned} k_{p+1} = 1 \text{ 일때, } & P_{p+1}(z) = A_p(z) - B_p(z) \\ k_{p+1} = -1 \text{ 일때, } & P_{p+1}(z) = A_p(z) + B_p(z) \end{aligned} \quad (2.5)$$

$$\Rightarrow A_p(z) = \frac{1}{2} [P_{p+1}(z) + Q_{p+1}(z)] \quad (2.6)$$

두 개의 근($k_{p+1} = \pm 1$)을 알고 있으므로 $P_{p+1}(z)$ 와 $Q_{p+1}(z)$ 의 차수를 줄일 수 있다. 즉,

$$\begin{aligned} P(z) &= \frac{P_{p+1}(z)}{(1-z)} \\ &= A_0 z^p + A_1 z^{(p-1)} + \dots + A_p \end{aligned} \quad (2.7)$$

그리고

$$\begin{aligned} Q(z) &= \frac{Q_{p+1}(z)}{(1-z)} \\ &= B_0 z^p + B_1 z^{(p-1)} + \dots + B_p \end{aligned} \quad (2.8)$$

$$\text{조건 : } A_0 = 1, B_0 = 1 \quad (2.9)$$

$$A_k = (a_k - a_{p+1-k}) + A_{k-1} \quad (2.10)$$

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

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 가지고 있다.

첫째, $P(z)$ 와 $Q(z)$ 는 단위원 상에 놓여 있다.

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

Real root method

일반적으로 $P(z)$ 와 $Q(z)$ 의 다차 방정식의 해를 구

하는 방식에 따라 여러 가지 변환법이 개발되었다. 그러나 이러한 변환법 중 real root 방법이 비교적 간단하고 이해하기 쉬워 주로 사용되어지고 있다[3].

$P(z)$ 와 $Q(z)$ 의 계수는 대칭적이기 때문에 식(2.7)의 차수는 $p/2$ 로 줄어든다.

$$\begin{aligned} P(z) &= A_0 z^p + A_1 z^{p-1} + \dots + A_1 z^1 + A_0 \\ &= z^{p/2} [A_0 (z^{p/2} + z^{-p/2}) + \\ &A_1 (z^{(p/2-1)} + z^{-(p/2-1)}) + \dots + A_{p/2}] \end{aligned} \quad (2.11)$$

$$\begin{aligned} Q(z) &= B_0 z^p + B_1 z^{p-1} + \dots + B_1 z^1 + B_0 \\ &= z^{p/2} [B_0 (z^{p/2} + z^{-p/2}) + \\ &B_1 (z^{(p/2-1)} + z^{-(p/2-1)}) + \dots + B_{p/2}] \end{aligned} \quad (2.12)$$

모든 근이 단위원 상에 있기 때문에, 단지 아래와 같이 정의하고 단위원 상에서 식(2.11)의 값을 구할 수 있다.

$$\text{Let } z = e^{j\omega} \text{ then } z^1 + z^{-1} = 2 \cos(\omega) \quad (2.13)$$

$$\begin{aligned} P(z) &= 2e^{j\omega p/2} [A_0 \cos(\frac{p}{2} \omega) + \\ &A_1 \cos(\frac{p-2}{2} \omega) + \dots + \frac{1}{2} A_{p/2}] \end{aligned} \quad (2.14)$$

$$\begin{aligned} Q(z) &= 2e^{j\omega p/2} [B_0 \cos(\frac{p}{2} \omega) + \\ &B_1 \cos(\frac{p-2}{2} \omega) + \dots + \frac{1}{2} B_{p/2}] \end{aligned} \quad (2.15)$$

$x = \cos \omega$ 를 대입해서 식(2.14)와 식(2.15)을 x 에 대해서 풀 수 있다. 예를 들어서 $p = 10$ 이면 다음과 같이 얻어진다.

$$\begin{aligned} P_{10}(x) &= 16A_0 x^5 + 8A_1 x^4 + (4A_2 - 20A_0) x^3 \\ &+ (2A_3 - 8A_1) x^2 + (5A_0 - 3A_2 + A_4) x \\ &+ (A_1 - A_3 + 0.5A_5) \end{aligned} \quad (2.16)$$

유사하게,

$$\begin{aligned} Q_{10}(x) &= 16B_0 x^5 + 8B_1 x^4 + (4B_2 - 20B_0) x^3 \\ &+ (2B_3 - 8B_1) x^2 + (5B_0 - 3B_2 + B_4) x \\ &+ (B_1 - B_3 + 0.5B_5) \end{aligned} \quad (2.17)$$

LSP는 식(2.18)에 의해서 구해진다.

$$LSP(i) = \frac{\cos^{-1}(x_i)}{2\pi T}, \text{ for } 1 \leq i \leq p \quad (2.18)$$

이 방법은 다른 변환 방법보다 비교적 간단하나 계산 시간이 어느 정도 걸릴지 예상할 수 없다는 단점이 있다 [3].

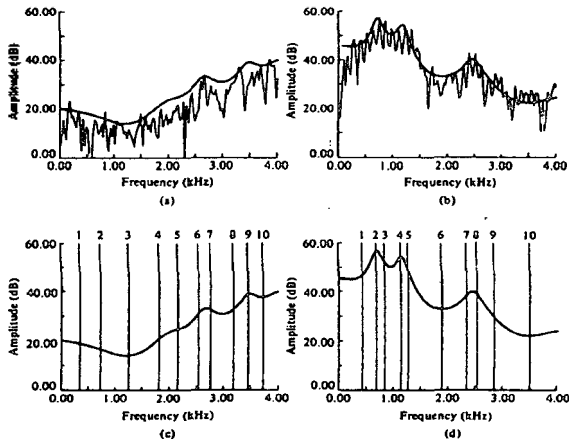


그림 2-2. 선형스펙트럼상의 예
 (a) 자음 /s/ (b) 모음 /a/
 (c),(d) /s/와 /a/에 대한 LPC 분석과 LSP

III. 음성신호의 LSP 분포와 포먼트 위치 및 분포 특성

real root LSP 변환 방법에서 먼저 홀수 번째의 계수를 먼저 찾은 후 짝수 번째의 계수는 먼저 찾은 것의 사이에서 찾게된다. 따라서 홀수 번째의 계수를 찾는데 걸리는 시간이 변환시간의 대부분을 차지하게 된다. 그림 3-1은 음성신호에서 LSP 파라미터의 홀수 번째 계수의 분포도를 나타낸 것이다. 8kHz 표본화율에 10차의 선형예측계수를 사용하였다. 가로축은 주파수대역이고 세로축은 발생빈도를 나타낸다.

LSP 홀수 번째 계수들은 특정 주파수 영역에서 주로 나타나고 나머지 영역에서는 거의 나타나지 않는 것을 볼 수 있다. real root 방법은 이러한 홀수 번째의 계수를 찾을 때 순차적으로 주파수대역을 검색한다. 그러나 그림 (c), (d), (e)를 살펴보면 분포도가 순차적이지 않음을 알 수 있다. 따라서 검색 순서를 파라미터가 많이 나타나는 주파수 대역부터 검색한다면 검색시간을 단축할 수 있다.

또한, LSP 파라미터는 음성신호의 포먼트 스펙트럼 정보를 반영한다. 유성음의 제 1 포먼트(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 표현하는 방법이다[2]. 그림 2-2의 (d)의 2, 3번째와 4, 5번째는 제 1포먼트와 제 2포먼트를 나타낸다. 따라서 포먼트의 위치 및 분포 특성은 LSP 파라미터에 직접적인 영향을 미치게 된다.

그림 3-2에 모음에 대한 F_1 에서 F_3 까지의 평균포먼트의 위치가 표시되어 있다. 남성 33명의 발음에서 얻은 평균 포먼트 주파수의 위치와 크기가 나타나 있다. /i/에서 /a/까지의 모음의 포먼트 위치에서 F_1 과 F_2 의 위치가 점점 가까워지는 것을 볼 수 있다[1][2]. 이러한 특성을 real root 방법의 근 검색에 적용하면 검색 시간을 단축 할 수

있다. 1, 2, 3번째 LSP 파라미터의 간격을 검색하여 F_1 의 존재를 결정하고, 그 위치에 따라 나머지 LSP 파라미터를 검색하는 구간을 조금씩 조절하면 보다 음성신호에 적응적으로 처리 할 수 있다.

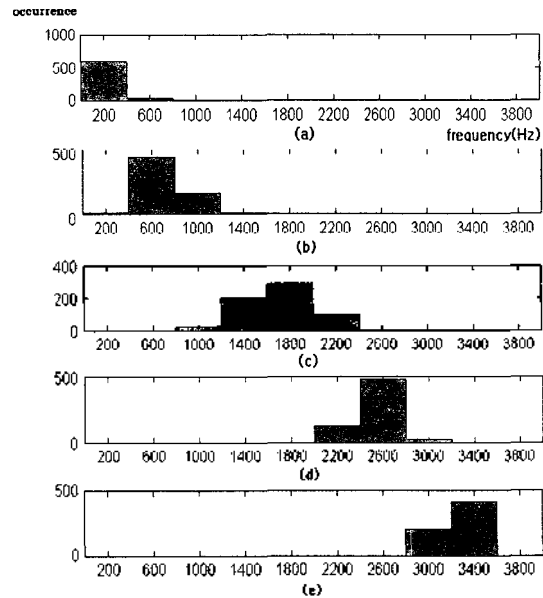


그림 3-1. LSP 홀수 번째 계수의 분포도

- (a) 1 번째 LSP 계수
- (b) 3 번째 LSP 계수
- (c) 5 번째 LSP 계수
- (d) 7 번째 LSP 계수
- (e) 9 번째 LSP 계수

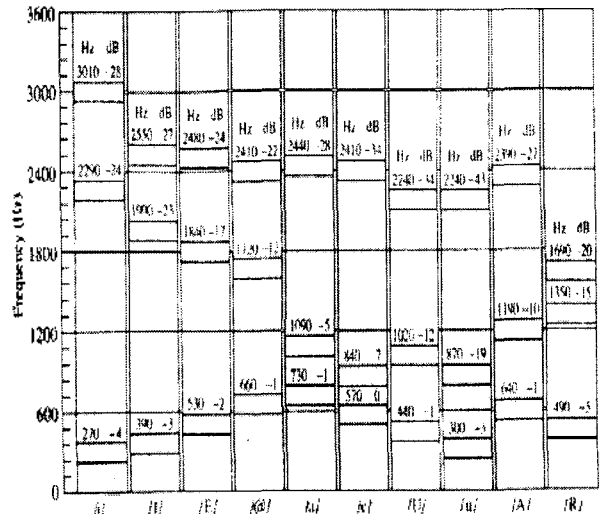


그림 3-2. 모음의 평균 포먼트 위치

표 3-1은 홀수 번째 LSP 파라미터의 주파수 대역별 검색 순서로서 분포도에 의해 제일 많이 나타나는 구간을 제일 먼저 검색하게 된다. 그리고 1~2, 2~3번째 LSP 파라미터의 간격을 검색하여 문턱값 이하의 간격이 나타나면 제 1 포먼트로 간주한다. 제 1 포먼트가 존재하면 표

3-2와 같이 검색구간 조절한다. '이동' 주파수만큼 표 3-1의 검색 구간을 바꾼다.

표 3-1. 주파수 대역별 검색순서

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

표 3-2. 제 1 포맷트 위치에 따른 검색구간 조절

	위치	이동(Hz)
제1포맷트	200이하	+300
	200-300	+200
	300-400	+100
	400-500	0
	500-600	-100
	600-700	-200
	700-800	-300

IV. 실험 및 결과

제안한 방법을 실험하기 위해서 먼저 IBM PC(233 MHz)에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 남자와 여자가 연구실 환경(30dB의 SNR)에서 발성한 음성을 8kHz로 표본화하고 16bit로 양자화하여 사용하였다. 발성한 문장은 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]. 먼저 C-언어로 구현된 real root 방법의 LSP 변환함수 부분을 Matlab 언어로 바꾸어 구현하고 이것을 기반으로 하여 제안한 알고리즘을 구현하였다.

표 4-1은 제안한 알고리즘과 기존의 real root 알고리즘의 LSP 변환방법의 계산시간을 비교한 것이다. 제안한 알고리즘의 LSP 변환 방법이 기존의 방법 보다 평균 48%이상 빠르다는 것을 보여주고 있다. 그러나, 변환된 LSP 계수를 비교해 보면 기존의 것과 동일하였다. 즉, LSP 계수의 값은 변하지 않고 계산시간만 48% 감소하였다.

표 4-1. 기존의 방법과 제안한 방법의 계산시간 비교

	기존의 방법 (단위:초)	제안한 방법 (단위:초)	감소율 (단위:%)
발성(1)	120.15	62.39	48.07
발성(2)	74.04	38.48	48.03
발성(3)	73.22	37.68	48.54
발성(4)	119.90	62.78	47.63

IV. 결 론

저전송률 음성부호화기에서는 음성신호의 선형예측 방법 중 LSP 파라미터를 이용하는 방법이 많이 사용되어지고 있다. 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 알고리즘의 LSP 변환방법의 계산시간을 단축하였다. real root 알고리즘에서는 LSP 파라미터를 구하기 위해 주파수 대역을 순차적으로 검색한다. 그러나 제안한 알고리즘에서는 주파수 대역을 LSP파라미터의 분포도에 따라 검색 구간의 순서를 정하여 검색하였다. 그리고 제 1 포맷트 위치와 제 2 포맷트 위치의 관계를 반영하여 검색구간을 조절하여 보다 음성신호에 적용적으로 처리할 수 있게 하였다. 이렇게 함으로써 평균 48%의 변환시간을 단축할 수 있었다.

앞으로 제안한 알고리즘을 음성부호화기에 적용하여 부호화기의 전체적인 성능향상에 관한 연구를 진행해야 할 것이다.

V. 참고 문헌

- [1] 배명진, "디지털 음성분석", pp.95-120, 동영출판사, 1998. 4
- [2] L. R. Rabiner, R.W. Schafer, "Digital Processing of Speech Signal", pp.38-115, Prentice Hall, 1978
- [3] A. M. Kondoz, "Digital Speech", pp. 84-92, John Wiley & Sons Ltd, 1994.
- [4] ITU-T Recommendation G.723.1, March, 1996
- [5] John R. Deller, Jr., John G. Proakis, John H.L. Hansen, "Discrete-Time Processing of Speech Signals", pp.124-125, Maxwell Macmillan International, 1993.
- [6] Sadaoki Furui, "Digital Speech Processing, Synthesis, and Recognition", pp129, MARCEL DEKKER, INC. 1991.