

# 해상도 조절과 검색순서 조절을 통한 음성부호화기용 복잡도 감소 알고리즘

민소연<sup>1\*</sup>, 이광형<sup>2</sup>, 배명진<sup>3</sup>

## The Reduction Algorithm of Complexity using Adjustment of Resolution and Search Sequence for Vocoder

So-Yeon Min<sup>1\*</sup>, Kwang-Hyoung Lee<sup>2</sup> and Myung-Jin Bae<sup>3</sup>

**요 약** 일정한 스펙트럼 민감도와 낮은 스펙트럼 왜곡을 보이고 선형 보간이 용이하다는 장점을 갖는 LSP 파라미터는 음성코덱(codec)이나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환되어, 주로 저전송률 음성부호화에 사용된다. 그러나 LPC 계수를 LSP로 변환하는 방법이 복잡하여 계산시간이 많이 소요된다는 단점이 있다.

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 제안하는 알고리즘은 LSP 분포 특성에 따라 검색구간의 순서와 검색간격을 달리하며, 제1 포먼트와 제2 포먼트의 연관성을 고려하여 검색구간을 조절한다. 기존의 real root 방법과 제안한 방법을 비교한 결과 검색시간이 평균 48.13% 단축되었다.

**Abstract** We propose the complexity reduction algorithm of real root method that is mainly used in the Vocoder. The real root method is that if polynomial equations have the real roots, we are able to find those and transform them into LSP(Line Spectrum Pairs). However, this method takes much time to compute, because the root searching is processed sequentially in frequency region. The important characteristic of LSP is that most of coefficients are occurred in specific frequency region. So, the searching frequency region is ordered and adjusted by each coefficient's distribution in this paper.

Transformation time can be reduced by proposed algorithm than the sequential searching method in frequency region. When we compare this proposed method with the conventional real root method, the experimental result is that the searching time was reduced about 48% in average.

**Key Words** : LSP(Line Spectrum Pairs), 음성부호화기(Vocoder), 복잡도(Complexity)

### 1. 서론

음성부호화기나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환하는데 LSP 파라미터가 사용된다. LSP 파라미터는 양자화 에러에 강하고 시스템의 안정성과 선형 보간성이 뛰어난 장점을 갖는다.

LPC(Linear Predictive Coding)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 있고 가장 많이 사용되는 것이 real root 방법이다[1,2,3]. 왜냐하면 real root 방법이 LPC를 LSP로 변환하는 방법 중 비교적 간단하고 이해하기 쉽기 때문이다. 그러나 real root 방법도 계산이 복잡하여 변환시간이 많이 소요되는 단점을 갖고 있다.

본 논문에서는 LSP 분포특성에 따라 검색구간의 순서와 간격을 조절함으로써 순차적으로 검색하는 real root 방법의 계산 시간을 단축하는 방법을 제안한다. 또한 제

본 연구는 서울시 산학협력사업으로 구축된 서울 미래형콘텐츠컨버전스 클러스터 지원으로 수행되었습니다.

<sup>1</sup>서울대학 정보통신과

<sup>2</sup>서울대학 인터넷정보과

<sup>3</sup>숭실대학교 정보통신 전자공학부

\*교신저자: 민소연(symin@seoil.ac.kr)

안한 방법은 음성신호의 특징을 반영하였는데, 즉 시간영역 정보를 이용하여 유무성음 구간과 묵음 구간을 판별하고, 찾아낸 LSP 정보를 이용하여 검색 대역을 조절함으로써 복잡도를 감소하는 특징을 갖고 있다. 본 논문의 구성은 다음과 같다. 2장에서는 LSP 파라미터의 추출 방법과 근을 찾는 방법에 대해 살펴본다.

3장에서는 LSP 분포특성에 따른 제안한 변환 방법을 설명한다. 그리고 4장에서는 실험 및 결과를 설명하고, 5장에서는 결론을 맺는다.

## 2. LSP 파라미터의 추출

PARCOR 필터 구조에서  $k_{p+1} = \pm 1$ 의 쌍을 가진 귀환 시스템 전달 함수는 각각  $P_p(z)$ 와  $Q_p(z)$ 로 표현된다.  $k_{p+1} = 1$ 일 때 식 (1)의 차분 필터로 표현 가능하다[1,2,3].

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

또한  $k_{p+1} = -1$ 일 때는 식(2)와 같이 합 필터로 표현되어진다.

$$Q_p(z) = A_p(z) - B_p(z) = A_p(z) - Z^{(p+1)}A_p(1/z) \quad (2)$$

식 (1),(2)을 통해 다음의 식이 유도되어진다.

$$A_p(z) = \frac{1}{2}[P_p(z) + Q_p(z)] \quad (3)$$

위에서 설명한 과정을 거쳐 다음과 같은 특성이 얻어진다. 즉,  $A_p(z)$ 와  $B_p(z)$ 의 p번째 다항식이 순환적 관계를 만족하고  $A_p(z) = 0$ 의 모든 근이  $|z|=1$ 의 단위 원내에 있다면  $P_p(z) = 0$   $Q_p(z) = 0$  근은 모두 단위 원 상에 존재하고 각각이 나뉘어져 있다. 다음 정리로부터  $P_p(Z) = 0$ 과  $Q_p(Z) = 0$ 의 모든 근들은 단위 원안에 존재한다. 그리고  $P_p(Z)$ 과  $Q_p(Z)$ 들은 아래 방법으로 표현된다. p가 짝수일 때 식(4),(5)와 같이 표현되어진다.

$$P_p(z) = (1-z) \prod_{i=2(even)}^p (1-2\cos\omega_i z + z^2) \quad (4)$$

$$Q_p(z) = (1+z) \prod_{i=1(odd)}^p (1-2\cos\omega_i z + z^2) \quad (5)$$

또한 p가 홀수일 때는 식(6),(7)과 같이 나타낸다.

$$P_p(z) = (1-z^2) \prod_{i=2(even)}^p (1-2\cos\omega_i z + z^2) \quad (6)$$

$$Q_p(z) = \prod_{i=1(odd)}^p (1-2\cos\omega_i z + z^2) \quad (7)$$

여기서,  $A_p(z)$ 는  $[P_p(z) + Q_p(z)]/2$ 로 주어진다.  $A_p(z)$ 는 식(8),(9)에 의해 재표현 되어진다. p가 짝수이고,  $c_i = -2\cos\omega_i$ ,  $c_i = -2\cos\omega_i$ ,  $c_i = -2\cos\omega_i$ ,  $c_0 = -z$ ,  $c_1 = -z$ 일 때 다음과 같이 표현가능하다[1,2,3].

$$A_p(z) - 1 = \frac{z}{2} \left[ \sum_{i=2}^b (c_i + z) \prod_{j=0}^{i-2} (1 + c_j z + z^2) - \prod_{j=2}^b (1 + c_j z + z^2) + \sum_{i=1}^{b-1} (c_i + z) \prod_{j=1}^{i-2} (1 + c_j z + z^2) + \prod_{j=1}^{b-1} (1 + c_j z + z^2) \right] \quad (8)$$

p가 홀수일 때는 식(9)로 표현 가능하다.

$$A_p(z) - 1 = \frac{z}{2} \left[ \sum_{i=2}^{b-1} (c_i + z) \prod_{j=0}^{i-2} (1 + c_j z + z^2) - z \prod_{j=2}^b (1 + c_j z + z^2) + \sum_{i=1}^b (c_i + z) \prod_{j=1}^{i-2} (1 + c_j z + z^2) \right] \quad (9)$$

## 3. 음성신호의 LSP 분포와 포먼트 위치 및 분포 특성을 이용한 검색시간 단축방법

real root LSP 변환 방법에서는 홀수 번째 계수를 먼저 찾은 후 짝수 번째 계수는 홀수 번째 사이에서 찾는다. 따라서 홀수 번째 계수를 찾는데 걸리는 시간이 변환시간의 대부분을 차지하게 된다.

그럼 1은 음성신호에서 LSP 파라미터의 홀수 번째 계

수의 분포도를 나타낸 것이다. 8kHz 표본화율에 10차의 선형예측계수를 사용하였다.

그림 1에서 LSP 홀수 번째 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 (c), (d), (e)에서는 분포도가 순차적이지 않음을 보인다. 따라서 검색 순서를 파라미터가 많이 나타나는 주파수 대역부터 검색한다면 검색시간을 단축할 수 있다 [7][8][9].

그리고 LSP 파라미터는 음성신호의 포먼트 스펙트럼 정보를 반영함을 알 수가 있다. 유성음의 제 1 포먼트(F1)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[2][8][9].

그림 2에 모음에 대한 F1에서 F3까지의 평균 포먼트의 위치가 표시되어 있다. 남성 33명의 발음에서 얻은 평균 포먼트 주파수의 위치와 크기가 나타나 있다. /i/에서 /a/까지의 모음의 포먼트 위치에서 F1과 F2의 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 F1의 존재를 결정하고, 그 위치에 따라 나머지 LSP 파라미터를 검색하는 구간을 조절하면 음성신호를 적응적으로 처리 할 수 있다[7][8].

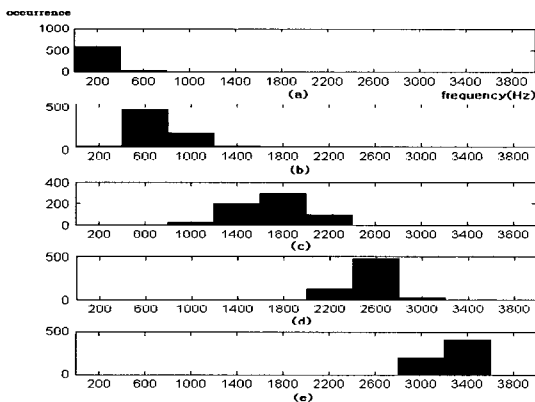


그림 1. LSP 홀수 번째 계수의 분포도  
 (a) 1 번째 LSP 계수 (b) 3 번째 LSP 계수  
 (c) 5 번째 LSP 계수 (d) 7 번째 LSP 계수  
 (e) 9 번째 LSP 계수

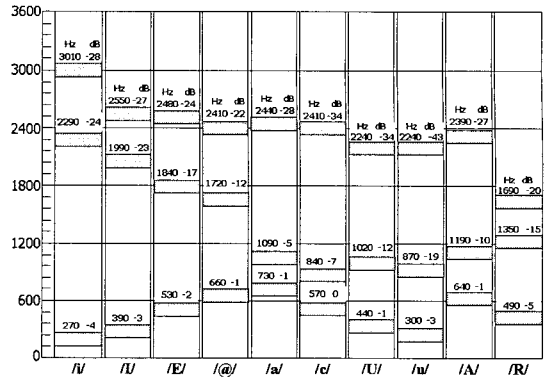


그림 2. 모음의 평균 포먼트 위치

표 1은 홀수 번째 LSP 파라미터의 주파수 대역별 검색 순서와 검색간격을 나타낸다. LSP 분포도에 따라 검색간격을 표 1의 값으로 조절하였다. 그리고 1~2, 2~3 번째 LSP 파라미터의 간격을 검색하여 문턱 값 이하의 간격이 나타나면 제1 포먼트로 간주한다. 제1 포먼트가 존재하면 표 2와 같이 검색구간을 조절한다. 즉, 표 2의 '이동' 주파수만큼 표 1의 검색 구간을 바꾼다. 그림 3은 제한한 음성신호의 LSP 분포와 포먼트 위치 및 분포 특성을 이용한 검색시간 단축 알고리즘의 전체 블록도를 나타내고 있다. 그림 3의 제한한 방법에서는 에너지를 이용하여 유무성음 구간과 묵음 구간을 판별하여 음성신호 구간에 맞는 처리를 하였다. 우선, 유무성음 구간이라고 판별이 되면 LSP 정보를 이용하게 된다. LSP 간격 정보를 이용하여 모음이라고 판별되면 검색 주파수 대역을 통계적인 데이터 값에 의해 조절하고 모음이 아니라고 판별이 되면 홀수 번째 LSP 파라미터의 분포특성에 의해 주파수 대역을 검색한다. 또한 묵음 구간은 LSP 파라미터의 분포특성이 비교적 일정하기 때문에, 순차적으로 주파수 대역을 검색하게 된다. 즉, 음성구간 특성에 맞게 적응적으로 처리를 하여 검색시간을 줄일 수 있게 된다.

표 1. 주파수 대역별 검색순서와 검색간격(단위:Hz)

검색순위 LSP계수 (Hz)	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 0 (40)	otherwise
LSP 7	2400-2800 (5)	2000-2400 (10)	1600-2000 (20)	otherwise	
LSP 9	2800-3200 (5)	3200-3600 (10)	otherwise		

표 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

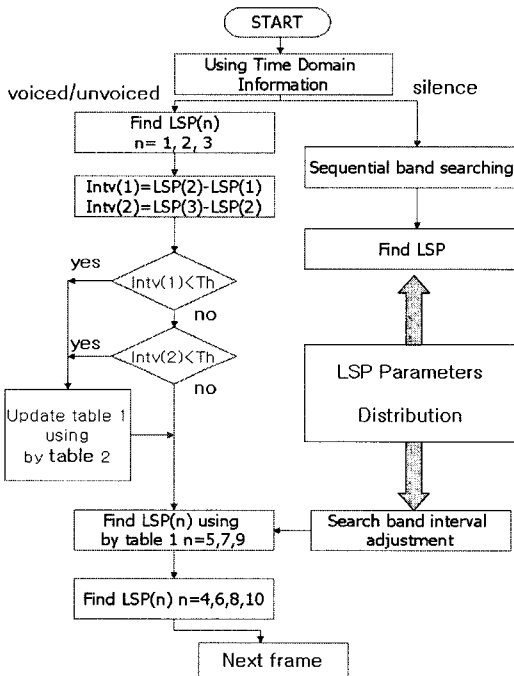


그림 3. 제안한 알고리즘의 순서도

### 4. 실험 및 결과

실험을 위해 IBM PC에 마이크 입력이 가능한 A/D 변환기를 인터페이스 하였다. 음성시료는 남성과 여성이 실험실 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: “서일대학 음성통신 연구팀이다”

본 논문에서 사용한 real root 알고리즘은 C-언어로 구현된 CELP 부호화기에서 발췌하여 사용하였다[4]. 먼저 C-언어로 구현된 real root 방법의 LSP 변환함수 부분을 Matlab 언어로 구현하고 이것을 기반으로 하여 제안한 알고리즘을 구현하였다.

표 3은 제안한 알고리즘과 기존의 real root 알고리즘의 LSP 변환방법의 계산시간을 비교한 결과로써, 제안한 알고리즘이 평균 48%이상 빠르게 나타남을 알 수 있다. 그러나, 변환된 LSP 계수를 비교해 보면 기존의 것과 동일하였다. 즉, LSP 계수의 값은 변하지 않고 계산시간만 약 48% 감소하였다.

표 3. 기존의 방법과 제안한 방법의 계산시간 비교

비교	순차적 검색 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
발성 1	87.13	44.64	48.76
발성 2	143.99	73.01	49.29
발성 3	106.21	53.76	49.38
발성 4	87.10	47.82	45.09

표 4에서는 제안한 방법의 객관적 음질 평가 방법인 SNR(Signal to Noise Ratio)에 대한 실험결과를 나타내었다. 실험과정에서 신호 대 잡음비를 계산하기 위하여 사용한 식은 (10),(11)과 같이 나타낼 수가 있다. 식에서 S(n)은 원 음성신호에 해당하고 S1(n)은 부호화기를 통과한 신호에 해당한다. 식(10)에서 잡음 에너지는 원 신호와 G.723.1을 통과한 신호간의 차이를 제공하여 구하였고, 한 프레임은 240 표본으로 하여 처리하였다. 식(11)은 전체 음성의 평균 신호 대 잡음비를 나타내며, 여기서 N은 총 프레임 수에 해당한다. 표 4의 결과를 통해 제안한 방법이 기존 방법과 유사한 음질을 얻을 수 있음을 알 수가 있다.

$$SNR_1(i) = 10\log(s^2(n)/n_1^2(n))$$

$$n_1^2(n) = \sum_{n=0}^{Frame-1} (s(n) - s_1(n))^2 \tag{10}$$

$$s^2(n) = \sum_{n=0}^{Frame-1} (s(n))^2$$

$$Ave_{SNR_1} = \frac{1}{N} * \sum_{i=1}^N SNR_1(i) \tag{11}$$

표 4. 부호화기를 통과한 신호의 신호 대 잡음비(단위:dB)

시료 \ 비교	기존방법	제안한 방법
발성 1	13.06	13.20
발성 2	11.74	11.21
발성 3	10.68	10.88
발성 4	10.18	10.04

표 5에서는 제안한 방법의 성능 평가를 위해 주관적 음질 평가방법인 MOS(Mean Opinion Score) test를 측정 한 결과를 나타내고 있다. MOS test에서는 부호화기를 통과하기 전의 원 음성신호, 기존의 방법을 이용한 합성음, 제안한 방법을 이용한 합성음에 대한 실험결과를 나타내었고, 제안한 방법은 기존방법과 비교하였을때, 기존 방법과 유사한 MOS test 결과를 얻을 수 있었다.

표 5. 부호화기를 통과한 신호들의 MOS test 결과

시료 \ 비교	원음성	기존방법	제안한방법
발성 1	3.8	3.7	3.7
발성 2	3.9	3.8	3.8
발성 3	3.8	3.7	3.7
발성 4	3.7	3.6	3.6

## 5. 결론

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 알고리즘의 LSP 변환방법의 계산시간을 단축하였다. real root 알고리즘에서는 LSP 파라미터를 구하기 위해 주파수 대역을 순차적으로 검색한다. 그러나 제안한 알고리즘에서는 LSP 파라미터의 분포도에 따라 검색구간의 순서와 검색구간의 간격을 조절하였다. 또한 제1 포먼트 위치와 제2 포먼트 위치의 관계를 반영하여 검색구간을 조절하여 보다 음성신호를 적응적으로 처리할 수 있게 하였다. 결과적으로 평균 48%의 변환시간을 단축할 수 있었다. 또한 SNR과 MOS test 결과에서는 기존 방법과 유사한 결과를 얻을 수 있었다.

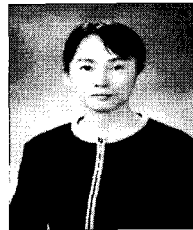
향후, 제안한 알고리즘을 음성부호화기에 적용하여 부호화기의 전체적 성능향상에 대한 연구와 보다 효과적인 검색구간의 순서와 간격 조절에 대한 연구가 이루어져야 한다.

## 6. 참고 문헌

- [1] 배명진, "디지털 음성분석", pp.95-120, 동영출판사, 1998. 4.
- [2] L. R. Rabiner, R.W. Schaffer, " 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.
- [7] 이희원,나덕수,정찬중,배명진, "LSP파라미터를 이용한 음성신호성분분리에 대한 연구", 하계 전자공학회 학술대회 논문집, 1999년 6월.
- [8] 이희원,나덕수,김만기, "LSP와 포먼트 분포특성을 이용한 음성신호의 LSP변환 알고리즘", 추계 음향학회 학술대회 논문집, 1999년 11월.
- [9] 민소연,이희원,나덕수,배명진, "Time Domain 정보와 LSP 분포 특성을 이용한 LSP 계산량 감소", 하계 통신학회 학술대회 논문집, 2000년 7월.

민 소 연(So-Yeon Min)

[종신회원]



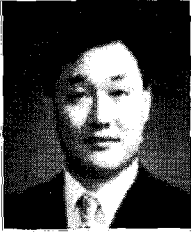
- 1994년 2월 : 숭실대학교 전자공학과 (공학사)
- 1996년 2월 : 숭실대학교 전자공학과 (공학석사)
- 2003년 2월 : 숭실대학교 전자공학과(공학박사)
- 2005년 3월 ~ 현재 : 숭실대학교 전자공학과 조교수

<관심분야>

통신 및 신호처리, 정보통신, 임베디드시스템

이 광 형(Kwang-Hyoung Lee)

[종신회원]



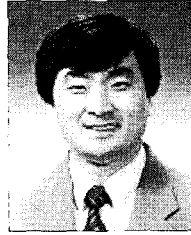
- 1998년 2월 : 광주대학교 컴퓨터 공학과 (공학사)
- 2002년 2월 : 송실대학교 컴퓨터 공학과 (공학석사)
- 2005년 2월 : 송실대학교 컴퓨터 공학과(공학박사)
- 2005년 3월 ~ 현재 : 서일대학 인터넷정보과 조교수

<관심분야>

멀티미디어 통신, 영상처리, 데이터베이스

배 명 진(Myung-Jin Bae)

[정회원]



- 1987년 2월 : 서울대학교 전자공학과 (공학박사)
- 2007년 10월 현재 : 송실대학교 정보통신전자공학부 교수

<관심분야>

음성코딩, 음성인식, 음성합성, 데이터통신 디지털신호처리 등