

음성특성을 이용한 LSP 변환시간 단축에 관한 연구

강은영, *민소연, 배명진
*숭실대학교 전자공학과, 정보통신공학과

A Study on the Reduction of LSP(Line Spectrum Pair) Transformation Time Using the Voice Characteristic

EunYoung KANG, *SoYeon MIN, MyungJin BAE
Dept. of *Electronics, Information and Telecommunication Engr.,
Soongsil University
E-mail : keyjsh@hanmail.net

요약문

LSP 파라미터는 일정한 스펙트럼 민감도와 낮은 스펙트럼 왜곡을 보이고 선형보간이 용이하다는 장점을 갖는다. 그러나 LPC 계수를 LSP 파라미터로 변환하는 방법이 복잡하여 계산시간이 많이 소요된다. 기존의 LSP 변환 방법 중 음성 부호화기에 주로 사용되는 Real Root 방법은 근을 구하기 위해 주파수 영역을 순차적으로 검색하기 때문에 계산시간이 많이 소요된다.

본 논문에서 제안하는 방법은 음성 특성을 이용하는 것으로, 목음의 경우는 목음 구간에서 일정하게 나타나는 LSP 파라미터의 분포 특성을 이용하여 검색하고 유/무성음에 대해서는 LSP 파라미터의 분포도에 따라 검색구간의 순서와 검색간격을 달리한다. 또한, 모음에 대해서는 제1 포먼트와 제2 포먼트의 연관성을 고려하여 검색구간을 조절한다. 기존의 Real Root 방법과 제안한 방법을 비교한 결과 검색시간이 평균 46.5% 단축되었다.

1. 서론

음성부호화기나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환 시, LSP 파라미터를 사용한다. 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[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[1].

본 논문에서는 음성특성에 따른 LSP 분포특성을

이용하여 검색순서와 검색간격을 조절함으로써 기존의 Real Root 방법의 계산 시간을 단축하는 방법을 제안한다.

2장에서는 LSP 파라미터의 추출 방법과 근을 찾는 방법에 대해 살펴보고, 3장에서는 제안한 변환 방법을 설명한다. 그리고 4장에서는 실험 및 결과를 설명하고, 5장에서는 결론을 맺는다.

2. LSP 파라미터의 추출

LSP 파라미터를 추출하기 위해서 먼저 LPC 분석이 이루어 진다[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 이용하여 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 $A_0(z) = 1$, $B_0(z) = z^{-1}$ 이고

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

그림 2-1의 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[5][6]. $k_{p+1} = -1$ 조건의 귀환은, 입력종점에서

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

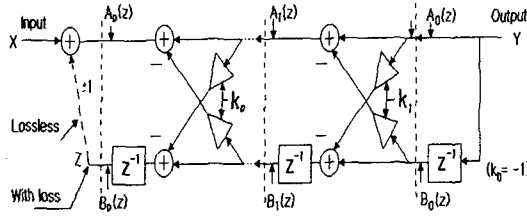


그림 2-1. PARCOR structure of LPC synthesis

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

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

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

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

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,

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

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

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

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

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

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 지닌다.

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

$P(z)$ 와 $Q(z)$ 의 다차 방정식의 해를 구하는 방식에 따라 개발된 여러 가지 변환법 중 주로 사용되는 Real Root 방법의 특징은 다음과 같다[3].

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

$$P(z) = A_0 z^p + A_1 z^{p-1} + \dots + A_1 z^1 + A_0$$

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

$$Q(z) = B_0 z^p + B_1 z^{p-1} + \dots + B_1 z^1 + B_0$$

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

모든 근은 단위원 상에 있기 때문에, 아래와 같이 정의하면 단위원 상에서 식(2.11)의 값을 구할 수 있다.

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

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

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

$x = \cos \omega$ 를 대입하면 식(2.14)와 식(2.15)을 x 에 대해서 풀 수 있다. $p=10$ 의 경우 다음식이 얻어진다.

$$P_{10}(x) = 16A_0 x^5 + 8A_1 x^4 + (4A_2 - 20A_0) x^3 + (2A_3 - 8A_1) x^2 + (5A_0 - 3A_2 + A_4) x + (A_1 - A_3 + 0.5A_5)$$

$$Q_{10}(x) = 16B_0 x^5 + 8B_1 x^4 + (4B_2 - 20B_0) x^3 + (2B_3 - 8B_1) x^2 + (5B_0 - 3B_2 + B_4) x + (B_1 - B_3 + 0.5B_5) \quad (2-17)$$

LSP는 식(2.18)에 의해서 구해진다.

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

Real Root 방법은 다른 변환 방법보다 간단하지만 계산시간을 예상할 수 없다는 단점을 갖는다[3].

3. 음성특성을 이용한 제안한 방법

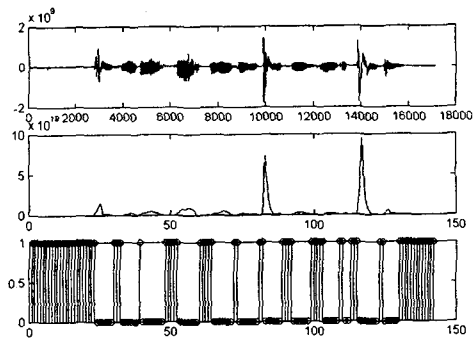


그림 3-1. 시간영역 정보를 이용한 목음구간 검출 (a)음성신호 (b)에너지 (c)검출된 목음구간

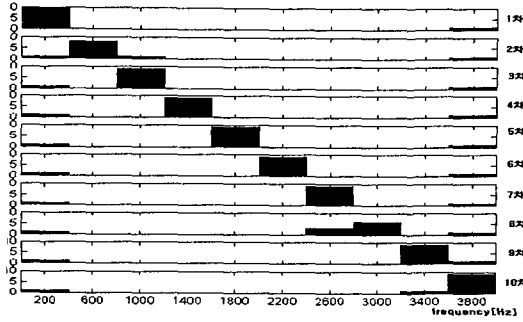


그림 3-2. 목음 구간에서의 LSP 계수의 분포도

Real Root LSP 변환 방법에서는 홀수번째 계수를 먼저 찾은 후 짝수번째 계수는 이미 찾아진 홀수번째 계수사이에서 찾는다. 따라서 홀수번째 계수를 찾는 데 걸리는 시간이 변환시간의 대부분을 차지하게 된다.

먼저 검색순서를 정하기 위해 목음구간과 음성구간의 LSP의 분포도를 구한다. 그림 3-1에서는 시간영역 정보를 이용하여 목음구간을 검출한 결과를 나타낸다. 이 과정에서 얻어진 목음 구간에 대한 LSP의 분포도를 살펴보면 그림 3-2와 같이 일정한 분포특성을 가짐을 알 수가 있다[8].

그림 3-3은 음성신호에서 LSP 파라미터의 홀수 번째 계수의 분포도를 나타낸 것이다. 8kHz 표본화율에 10차의 선형예측계수를 사용하였다. 이 그림에서 LSP 홀수번째 계수들이 특정 주파수 영역에서 나타나는 것을 알 수 있다. 특히 그림 (c), (d), (e)에서는 분포도가 순차적이지 않으므로 파라미터가 많이 나타나는 대역부터 검색하여 검색시간을 단축한다[7][8][9]. 분포도에 따른 검색순서는 표 3-2와 같다.

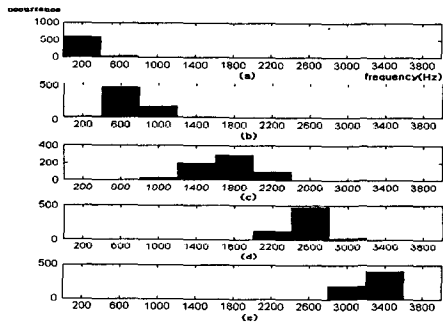


그림 3-3. LSP 홀수 번째 계수의 분포도
(a) 1 번째 LSP 계수 (b) 3 번째 LSP 계수
(c) 5 번째 LSP 계수 (d) 7 번째 LSP 계수
(e) 9 번째 LSP 계수

표 3-1. LSP 분포도에 따른 검색간격(단위:Hz)

분포도	1	2	3	4
간격	5	10	20	40

표 3-2. 주파수 대역별 검색순서

순서	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	2000-2400	otherwise	
LSP 7	2400-2800	2000-2400	2800-3200	otherwise	
LSP 9	3200-3600	2800-3200	otherwise		

또한, 그림 3-4에서 보여지듯 포먼트의 위치 및 분포 특성이 LSP 파라미터에 직접적인 영향을 미치므로 검색 방법에 제 1 포먼트(F_1)와 제 2 포먼트(F_2)의 관계를 이용한다. 그림 3-5에 모음에 대한 F_1 에서 F_3 까지의 평균 포먼트 위치와 크기가 표시되어 있다. 이는 남성 33명의 발성에서 얻은 것이다. 그림으로부터 /i/에서 /a/까지의 모음의 F_1 과 F_2 의 위치가 점점 가까워지는 것을 볼 수 있다[1][2]. 이러한 특성을 이용하여 좀더 정확한 검색 구간을 설정하면 검색 시간을 줄일 수 있다. 1, 2, 3번째 LSP 파라미터의 간격을 검색하여 문턱값 이하의 간격이 나타나면 제 1 포먼트로 간주하고, 그림 3-5를 참고하여 표 3-2의 LSP 파라미터의 검색 구간을 표 3-3에 따라 조절한다[7][8]. 또한 효율적 검색을 위해 사용한 LSP 분포도에 따른 검색 간격 조절 범위를 표 3-1에 나타내었다.

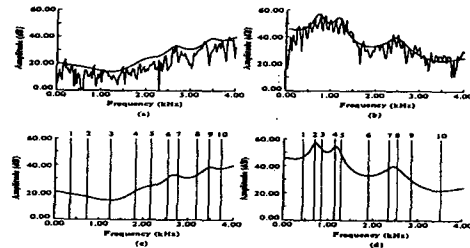


그림 3-4. 선형스펙트럼상의 예

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

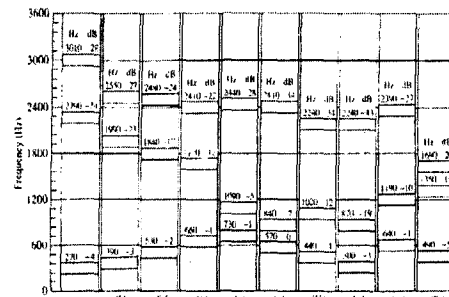


그림 3-5. 모음의 평균 포먼트 위치

표 3-3. 제1포먼트 위치에 따른 검색구간 조절

	위치	이동(Hz)
제1포먼트	200이하	+300
	200-300	+200
	300-400	+100
	400-500	0
	500-600	-100
	600-700	-200
	700-800	-300

4. 실험 및 결과

실험을 위해 IBM PC(333MHz)에 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 음성시료는 다음과 같다.

- 발성1: "인수네 꼬마는 천재소녀를 좋아한다."
- 발성2: "창공을 날으는 인간의 도전은 끝이 없다."
- 발성3: "예수님께서 천지창조의 교훈을 말씀하셨다."
- 발성4: "숭실대학교 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 4-1은 제안한 알고리즘과 기존의 Real Root 알고리즘의 LSP 변환방법에 대해 계산시간을 비교한 결과로써, 제안한 알고리즘이 평균 46%이상 빠르게 나타남을 알수있다. 그러나, 변환된 LSP 계수를 비교해 보면 기존의 것과 동일하였다. 즉, LSP 계수의 값은 변하지 않고 변환시간만 약 46% 감소하였다.

표 4-1. 기존의 방법과 제안한 방법의 변환시간 비교

비교 시료	순차적 검색 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
발성 1	84.8	45.19	46.71
발성 2	140.89	73.06	48.14
발성 3	103.22	53.82	47.86
발성 4	84.53	47.93	43.29

5. 결론

저전송률 음성부호화기에서 음성신호 선형예측 방법으로 LSP 파라미터를 많이 사용하는데 이는 LSP 파라미터가 왜곡이 적고 선형보간이 용이하기 때문이다. 하지만 LPC 계수에서 LSP 파라미터로의 변환이 복잡하여 계산시간이 많이 소요된다.

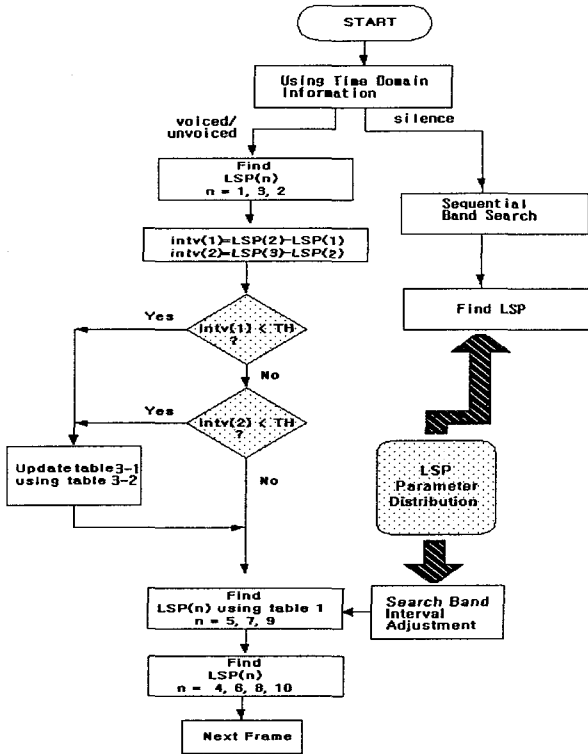


그림 4-1. 제안한 알고리즘의 순서도

본 논문에서는 음성의 특성을 이용하여 음성 부호화기에서 주로 사용되는 Real Root 알고리즘의 LSP 변환시간을 단축하였다. Real Root 알고리즘에서는 LSP 파라미터를 구하기 위해 주파수 대역을 순차적으로 검색한다. 그러나 제안한 알고리즘에서는 LSP 파라미터의 분포도에 따라 검색구간의 순서를 정하고, 제1 포먼트 위치와 제2 포먼트 위치의 관계를 반영하여 검색구간을 조절하여 보다 음성신호를 적응적으로 처리할 수 있게 하였다. 즉, 분포도에 따라 검색간격을 달리하여 좀더 정확하고 효율적인 검색을 수행하였다. 모의실험 결과 평균 46.5%의 변환시간을 단축할 수 있었다.

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", pp.129, MARCEL DEKKER, INC, 1991.
- [7] 이희원, 나덕수, 정찬중, 배명진, "LSP파라미터를 이용한 음성신호성분분리에 대한 연구", 학제 전자공학회 학술대회 논문집, 1999년 6월.
- [8] 이희원, 나덕수, 김민기, "LSP와 포먼트 분포특성을 이용한 음성신호의 LSP 변환 알고리즘", 추계 음향학회 학술대회 논문집, 1999년 11월.
- [9] 민소연, 이희원, 나덕수, 배명진, "Time Domain 정보와 LSP 분포 특성을 이용한 LSP 계산량 감소", 학제 통신학회 학술대회 논문집, 2000년 7월.