

LSP 변환 알고리즘들의 비교 평가에 관한 연구

임지선*

*(주)에이치씨티

e-mail:jisun722@nate.com

A Study on the Parameter Extraction for Performance Comparison of LSP transformation Time

Ji-Sun Lim*

*HCT. Co., Ltd.

요 약

LPC 계수를 LSP 변환하는 방법에는 복소근, 실근, 비율 필터, 체비셰프 급수, 적응적 순차형 최소제곱 평균 방법(adaptive sequential LMS) 등이 있다. 이 방법들 중 음성 부호화기에서 주로 사용하는 실근 방법은 근을 구하기 위해 주파수 영역을 순차적으로 검색하기 때문에 계산시간이 많이 소요되는 단점을 갖는다. 본 논문에서는 LPC에서 LSP로 변환하는 4가지 고속 알고리즘을 제안한다. 첫 번째 방식에서는 검색간격에 멜 스케일을 적용하였고, 두 번째는 홀수번째 LSP 파라미터의 분포도를 이용하여 검색순서를 조정하는 방법이다. 세 번째 방식과 네 번째 방식에서는 각각, 모음 특성, LSP 분포특성과 해상도를 이용하여 계산시간을 단축하였다. LSP 변환시간은 4가지 방법 모두 35~50% 단축되었다. 또한 실험결과에서는 각 알고리즘의 고유한 특성에 대하여 분석한다.

1. 서론

LPC(Linear Predictive Coding) 계수는 음성신호를 모델링한 all pole 모델 다항식의 계수이며 이는 상당히 큰 dynamic range 값을 갖는다. 이 특성은 LPC 계수를 양자화 하는데 있어서 많은 비트 수를 요구하고, 작은 양자화 에러에 의해서 복원시에 스펙트럼 왜곡이 생기며 pole이 단위원 밖으로 벗어나서 발산하는 경우가 생기게 된다. 즉 LPC 계수를 직접 전송했을 때 스펙트럼의 왜곡과 불안정한 문제점을 가지고 있다. 이 두 문제를 해결하는 방안으로 두가지 방법이 제시되고 있다. 첫 번째 PARCOR 파라미터의 전송방식에서는 LPC 분석 과정 중에서 Durbin 알고리즘의 반복 과정에서 중간 변수로 생성되는 반사계수를 사용한다. 이 계수는 사람의 성도를 p개의 무손실 관으로 모델링 할 때 각 section과 다음 section의 면적 차이에 따른 반사계수이다[1,2,3].

이 경우에 있어서 dynamic range 문제를 야기하지 않고 양자화 오류에 대한 안정성 문제도 발생하지 않는다. PARCOR 계수를 해당 비트 별로 양자화하고 복원한 경우에 있어서 LPC 분석과 비교하면 스펙트럼 왜곡이 작고 안정하다. 두 번째 방식인 LSP(Linear Spectrum Pairs)는 dynamic range가 작은 부분으로 제한되고 양자화 에러에 강하다. 또한 프레임간 선형보간 특성이 PARCOR에 비해

서 우수하고, 스펙트럼 민감도가 작아서 같은 양자화 비트 수에 대하여 스펙트럼 왜곡이 작다는 장점을 지닌다.

그러므로 음성부호화기나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환하는데 LSP 파라미터가 사용된다[1,2,3]. LPC(Linear Predictive Coding)계수를 LSP 파라미터로 변환하려면 먼저 다항식의 근을 찾는 과정이 필요하다. 기존의 방법중에서 가장 많이 사용되는 방식은 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. 본 논문에서는 LPC에서 LSP로 변환하는 4가지 고속 알고리즘을 제안한다. 첫 번째 방식에서는 검색간격에 멜 스케일을 적용하였고, 두 번째는 홀수 번째 LSP 파라미터의 분포도를 이용하여 검색순서를 조정하는 방법이다. 세 번째 방식과 네 번째 방식에서는 각각, 모음 특성, LSP 분포특성과 해상도를 이용하여 계산시간을 단축하였다. 실험결과에서는 위의 변환시간 단축 알고리즘의 계산시간 결과와 여러 경우에 있어서 각 알고리즘의 적합성에 대하여 나타낸다.

2. LSP 파라미터의 추출

PACORR 필터에서 $k_{p+1} = \pm 1$ 의 쌍을 가진 귀환 시스템

전달 함수는 각각 $P_p(z)$, $Q_p(z)$ 로 표현된다.

$$k_{p+1} = 1 \text{ 일 때}$$

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

$$k_{p+1} = -1 \text{ 일 때}$$

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

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

p가 짝수일 때,

$$P_p(z) = (1-z) \prod_{i=2(\text{even})}^p (1 - 2 \cos \omega_i z + Z^2) \quad (3)$$

$$Q_p(z) = (1+z) \prod_{i=1(\text{odd})}^p (1 - 2 \cos \omega_i z + Z^2) \quad (4)$$

p가 홀수일 때,

$$P_p(z) = (1-z^2) \prod_{i=2(\text{even})}^p (1 - 2 \cos \omega_i z + Z^2) \quad (5)$$

$$Q_p(z) = \prod_{i=1(\text{odd})}^p (1 - 2 \cos \omega_i z + Z^2) \quad (6)$$

여기서, $A_p(z)$ 는 $[P_p(z)+Q_p(z)]/2$ 로 주어진다. p가 짝수이고, $C_i = -2\cos\omega_i$, $C_0 = -z$, $C_1 = -z$ 일 때,

$$A_p(z) - 1 = \frac{z}{2} \left[\sum_{j=2}^p (c_j + z) \prod_{j=0}^{j-2} (1 + c_j z + z^2) - \prod_{j=2}^p (1 + c_j z + z^2) + \sum_{j=1}^{p-1} (c_j + z) \prod_{j=1}^{j-1} (1 + c_j z + z^2) + \prod_{j=1}^{p-1} (1 + c_j z + z^2) \right] \quad (7)$$

p가 홀수일 때는 식 (8)로 표현된다.

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

(c_i, c_{i+1}) , $c_i = \cos\omega_i$, $i=1,3,5,\dots,(p+1)/2$ 또는, (ω_i, ω_{i+1}) 일 경우 구현되어진다. ω_i 와 ω_{i+1} 은 하나의 주파수쌍으로 표현되는데 이것이 선형 스펙트럼쌍(LSP)이다[1,2,3].

선형예측계수 a_i ($i=1,2,\dots,p$), PARCOR 계수 k_i ($i=1,2,\dots,p$) 선스펙트럼표현 (y_i, m_i) , $1,3,\dots,p-1$ 과 선형스펙트럼쌍 표현 (ω_i, ω_{i+1}) ($i=1,3,\dots,p-1$) 들은 모두 전극스펙트럼의 표현이다[1,2,3].

3. 제안한 LSP 변환 고속 알고리즘

3-1. 멜 스케일을 적용한 검색간격 조절법

멜 스케일은 인간의 청각특성을 고려한 주파수 척도이며 1000Hz에 1000Mel을 대응하고 이로부터 실험적으로 값을 결정한다. 일반적으로 Mel 척도는 1000Hz 이하에서는 주파수와 선형적으로 비례하며 1000Hz 이상에서는 로그 스케일이다. 본 논문에서는 Fant에 의하여 근사된 식을 이용하였고 식 (9)에서 나타내었다. 식 (9)는 멜 스케일과 주파수 사이의 관계를 나타내고 있다[6,9].

$$F_{mel} = k * \log \left[1 + \frac{f}{1000} \right] \quad (9)$$

여기서,

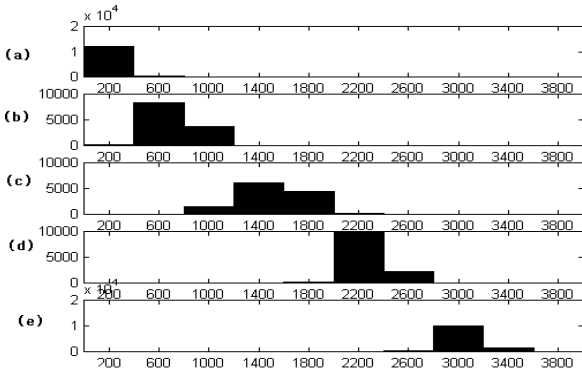
$$k = 1000 / \log 2$$

3-1-1. 제안한 방법(I)의 분석

음성신호가 들어오면 해밍 윈도우를 취하고 LPC 계수를 추출한 후 이것으로 다차 방정식을 만든다. 그리고 다음 과정에서는 멜 스케일을 적용하여 홀수 번째 근을 찾고 찾아진 홀수 번째 근 사이에서 짝수 번째 근을 찾기 시작한다. 위의 과정은 10개의 근을 찾을 때까지 반복된다.

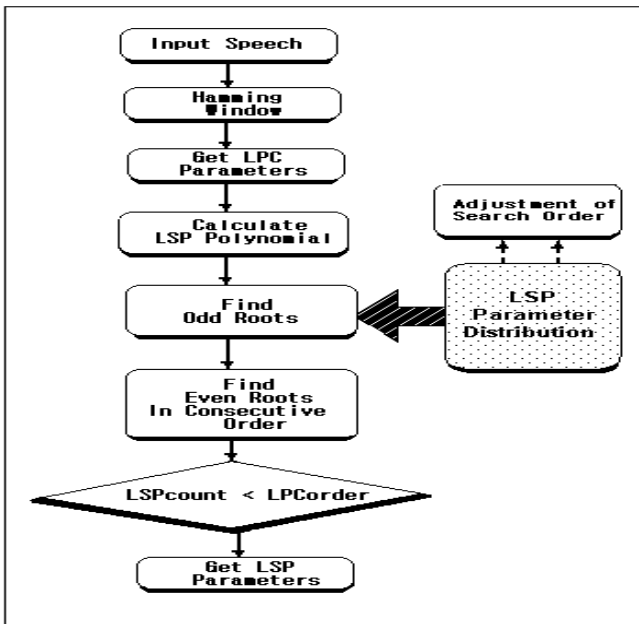
3-2. 홀수번째 LSP 파라미터 분포특성을 이용한 검색순서 조절법

음성부호화기에서 많이 사용되어지는 실근 방법에서는 10차 LSP 파라미터를 구하는데 있어서 홀수 번째 근을 찾기 위하여 주파수 대역을 순차적으로 검색한다. 이후 이미 찾아진 홀수 번째 근 사이에서 짝수 번째 근을 찾기 위해 주파수 대역을 순차적으로 검색한다. 그러므로 홀수 번째 근을 찾기 위하여 주파수 대역을 검색하는 시간이 전체 변환시간의 대부분을 차지하게 된다. 제안한 방법 두 번째에서는 LSP 파라미터의 분포도가 어느 특정 주파수 대역에 주로 분포한다는 특성을 이용하여 근을 찾을 때 분포특성이 높은 대역을 먼저 검색하여 변환시간의 단축을 도모하였다. 음성신호의 LSP 분포도의 확률적 특성을 알기 위하여 20분 정도의 긴 음성의 분포도를 조사하였다. 음성시료는 8kHz 표본화 율에 10차의 선형예측계수를 사용하였다. 그림 2는 동화, 아기돼지 삼 형제를 음성시료로 사용한 경우에 해당되는 LSP 파라미터의 분포특성이다. 발성시료에 따라 홀수 번째 LSP 파라미터의 분포도가 약간의 변동이 있지만 대부분의 음성시료에 있어서 거의 유사한 분포특성을 지님을 알 수가 있었다.



[그림 2] 홀수 번째 LSP 파라미터의 분포특성
 (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

3-2-1. 제안한 방법 (II)의 분석



[그림 3] 제안한 방법(II)의 순서도

예를 들어 그림 2의 (b)의 경우에는 3번째 LSP 파라미터가 주로 400~800Hz 사이에 근이 가장 많이 분포한다. 그러므로 3번째 근을 찾을 때에는 분포도가 가장 높은 영역을 먼저 검색하고 근을 못 찾은 경우에 있어서 그 다음으로 분포도가 높은 800~1200Hz 사이를 검색하게 된다. 그림 3은 제안한 방법 (II)의 순서도 이다.

3-3. 시간 영역 정보와 제 1 포먼트 위치에 따른 검색간격 조절법

LSP 파라미터는 음성신호의 포먼트 스펙트럼 정보를 반영한다. 유성음의 제 1 포먼트는 비교적 낮은 주파수 대역에서 좁은 대역폭을 갖는다. LSP 파라미터에서 선쌍의 간격이 좁으면, 강한 공명이 일어난 것을 나타낸다. 제안한 방법(III)에서 제1 포먼트와 제2 포먼트의 관계를 살펴보기 위

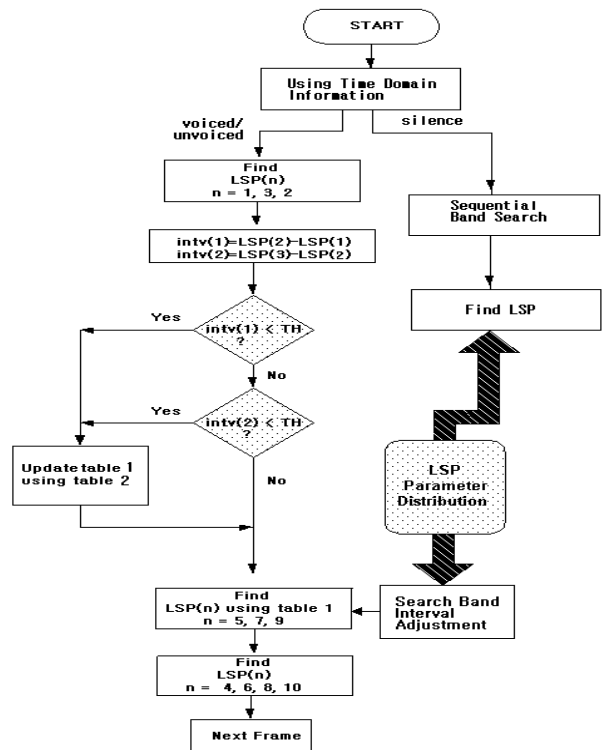
하여 모음에 대한 F₁에서 F₃까지의 평균 포먼트의 위치를 조사하였다. 남성 33명의 발음에서 얻은 평균 포먼트 주파수의 위치와 크기를 조사한 결과, /i/에서 /a/까지의 모음의 포먼트 위치에서 F₁과 F₂의 위치가 점점 가까워지는 것을 볼 수 있다[5,6,7]. 이러한 특성을 real root 방법의 근 검색에 적용하면 검색 시간 단축을 가져온다. 1, 2, 3번째 LSP 파라미터의 간격을 검색하여 F₁의 존재를 결정하고, 그 위치에 따라 나머지 LSP 파라미터를 검색하는 구간을 조절하면 음성신호를 적응적으로 처리 할 수 있다[7][8].

3-3-1. 제안한 방법(III)의 분석

그림 4에서 1~2, 2~3번째 LSP 파라미터의 간격을 검색하여 문턱값 이하의 간격이 나타나면 제1 포먼트로 간주한다. 제1 포먼트가 존재하면 표 1과 같이 검색구간을 조절한다. 즉, 표 1의 '이동' 주파수만큼 그림 2의 검색 구간을 바꾼다.

3-4. 검색순서와 검색 간격 조절법

제안한 방법(IV)에서는 LSP 파라미터가 많이 분포하는 주파수 대역을 먼저 검색하되 검색 간격을 조밀하게 검색하고, 역으로 분포도가 낮은 대역을 검색할 때는 해상도를 낮추어 검색하여 계산 시간을 단축한다. 표 1에서 나타낸 것은 5번째 근을 찾을 때 1200~1600Hz를 먼저 검색하되 해상도는 5Hz로 하여 검색하고 1600~2000Hz를 검색할 때는 10Hz의 해상도를 설정하여 검색함을 나타내고 있다[7,8].



[그림 4] 제안한 방법(III)의 순서도

[표 1] 제 1포맷트에 의한 검색구간 조절표

	위치	이동(Hz)
제1 포맷트	200이하	+300
	200-300	+200
	300-400	+100
	400-500	0
	500-600	-100
	600-700	-200
	700-800	-300

[표 2] 제 1포맷트에 의한 검색구간 조절표

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

4. 실험결과 및 결론

본 논문에서 사용한 real root 알고리즘은 C-언어로 구현된 CELP 부호화기에서 발췌하여 사용하였다[1,4,5]. 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 알고리즘 3가지 모두, 평균 35~50% 이상 빠르게 나타남을 알 수가 있다. 또한, 변환된 LSP 계수를 비교해 보면 기존의 것과 동일하였다. 표 4는 잡음, 모음, 무성음, 목음이 인가되었을 때 3가지 제안한 방식이 어느 정도 적합한가를 나타낸 것이다. 예를 들어 방법 3의 경우 잡음이 인가된 경우 시간 영역 정보인 에너지를 가지고 무성음과 목음구간의 판단이 불가능하므로 부적합하다는 결론이 얻어진다. 향후 3가지 방식을 비교하기 위하여 정확한 파라미터의 추출이 연구되어야 한다.

[표 3] 제안한 방법의 LSP 변환시간 단축결과
음성시료:/인수네 꼬마는 천재소녀를 좋아한다/

비교 알고리즘	순차적 검색 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
방법 1	90.48	47.56	47.36
방법 2		44.27	51.07
방법 3		59.27	34.49
방법 4		42.23	53.33

[표 4] 주어진 특성을 반영하기 위한 적합도
(◎:매우 우수, ○: 우수, △:보통, ×:부적합)

비교 알고리즘	잡음환경	모음	무성음	목음
방법 1	○	○	△	○
방법 2	◎	○	◎	○
방법 3	×	◎	△	◎

참고문헌

- [1] 한진수, "음성신호처리", 34-61, 오성미디어, 2000.
- [2] 배명진, "디지털 음성분석", pp.95-120, 동영출판사, 1998.
- [3] L R. Rabiner, R.W Schafer, " Digital Processing of Speech Signal", pp.38-115, Prentice Hall, 1978.
- [4] A. M. Kondoz, "Digital Speech", pp. 84-92, John Wiley & Sons Ltd, 1994.
- [5] ITU-T Recommendation G.723.1, March, 1996.
- [6] John R. Deller, Jr., John G. Proakis, John H.L. Hansen, "Discrete-Time Processing of Speech Signals", pp.124-125, Maxwell Macmillan International, 1993.