

불균등 간격조절과 선형 스펙트럼 쌍 분포특성을 이용한 계산량 단축 알고리즘

주상규*
*(주)주성시스템
e-mail:dkmadl@naver.com

A Reduction Algorithm of Computational Amount using Adjustment the Not Uniform Interval and Distribution Characteristic of LSP

Sang-Gyu Ju*
*JuSung System Co., Ltd.

요 약

Fast algorithm is proposed by using mel scale and the distribution characteristic of LSP parameters, and is to reduce the computational amount. Computational amount means the calculating times of transformation from LPC coefficients to LSP parameters. Among conventional methods, the real root method is considerably simpler than other, but nevertheless, it still suffer from its indeterministic computational time. Because the root searching is processed sequentially in frequency region. In this paper, the searching interval is arranged by using mel scale but not it is uniform and searching order is arranged by the distribution characteristic of LSP parameters that is most LSP parameters are occurred in specific frequency region. In experimental results, computational amount of the proposed algorithm is reduced about 48.95% in average, but the transformed LSP parameters of the proposed method were the same as those of real root method.

1. 서 론

CELP (Code Excited Linear Prediction) 방식은 4.8kbps 내외의 전송률에서 양호한 음질을 얻을수 있으며 ITU-T, TIA/EIA 등 여러 국제 표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. CELP 계열 부호화기들 중에서 G.723.1은 멀티미디어 통신 환경하의 음성 전송 표준 부호화기로 개발되었다. G.723.1은 5.3kbps/ 6.3kbps의 이중 전송률을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷폰과 그 외의 이동통신용 부호화기로 사용되고 있으며 낮은 전송률에 비해서 우수한 음질을 제공하고 있다. 더욱이 최적의 전송환경을 위하여 두개의 전송률을 사용하기 때문에 다른 부호화기 표준안들에 비해서 응용성이 높다. 현재 음성 코덱이나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환되는 LSP 파라미터는 일정한 스펙트럼 민감도와 낮은 스펙트럼 왜곡을 보이고 선형 보간이 용이하다. LSP 변환 방법 중 음성 부호화기에서

주로 사용하는 실근 방법은 근을 구하기 위해 주파수 영역을 순차적으로 검색하기 때문에 계산시간이 많이 소요된다. 본 논문에서 제안하는 알고리즘은 멜 스케일을 이용하여 검색간격 정보를 불균등하게 조절하고, 동시에 홀수 번째 LSP 파라미터의 분포특성을 이용하여 검색순서를 조절하여 계산량을 감소시킨다. 실험과정에서는 LPC에서 LSP 파라미터로 변환시 LSP 파라미터의 변동사항과 계산량 감소 정도를 측정하였다.

2. LSP 파라미터 추출

LSP 파라미터를 추출하기 위해서 먼저 LPC(Linear Predictive Coding)분석이 이루어져야 한다. LPC 분석은 고역 여파된 신호에 대해서 실행되어지며 10차의 선형 예측 분석이 수행된다. 180표본의 해밍 윈도우(Hamming window)가 각각의 부 프레임 중앙에 위치하게 되며 11개의 자기 상관계수가 윈도우 처리된 신호로부터 계산되어진다. 상관계수 $R[0]$ 는 1025/1024의 백색 잡음 상관 인

자 처리가 되며 다른 10개의 상관계수는 이항 윈도우(binomial window) 계수 테이블에 의해 곱해진다. 선형 예측 계수(LPC)는 Levinson Durbin recursion을 사용하여 계산되어지며, 모든 입력 프레임, 각 부 프레임마다 하나씩 계산되어 네 집합의 LPC 계수가 계산되어진다. 이 LPC 계수는 단구간 인지가중화 필터를 만들기 위해 사용되는데, 이때 LPC 합성 필터는 다음과 같이 정의된다[1~4].

$$A_i(z) = \frac{1}{1 - \sum_{j=1}^{10} a_{ij}z^{-j}}, \quad 0 \leq i \leq 3 \quad (1)$$

여기서 i 는 0과 3사이에서 정의되는 부 프레임 인덱스이며 j 는 차수를 나타낸다. LPC(Linear Predictive Coefficients) 양자화를 하기 위하여, 먼저 7.5Hz의 대역폭 확장(bandwidth expansion)이 수행된다. 그 결과로 $A_3(z)$ LP 필터는 예측 분할 벡터 양자화기를 이용하여 양자화 되어진다. 양자화 과정은 LP 계수 $\{a_{ij}\}_{j=1..10}$ 가 단위 원과 영 교차에 대한 보간 과정의 검색에 의해 LSP 계수 $\{p'_j\}_{j=1..10}$ 로 변환된다.

3. 제안한 알고리즘

멜 스케일은 인간의 청각특성을 고려한 주파수 척도이며 1000Hz에 1000Mel을 대응하고 1000Hz 이하에서는 주파수와 선형적으로 비례하며 1000Hz 이상에서는 로그 스케일이다. 본 논문에서는 Fant에 의하여 근사화된 식을 이용하였고 이를 이용하여 제안한 알고리즘에서의 검색간격을 조절하기 위해 사용한 식을 표현하면 다음과 같이 표현 가능하다[5,6].

$$f_n = (10^{point(n)/(1000/\log 2)} - 1) * 1000 * 0.5 \quad (2)$$

여기서 $f_0 = 0$, $point(n) = index * n$, for $1 \leq n \leq 400$, $index = F_{mel}/400$ 을 의미한다. 또한, 식(2)에서 $F_{mel} = 1000 / \log 2 * \log[1 + FS/1000]$, $FS = 8000$ 이다. 식(2)를 통해서 LSP 파라미터를 구하기 위한 주파수 대역은 0Hz부터 4kHz까지 식(2)의 멜 스케일로 400개의 검색 간격으로 나뉘어지게 된다. 그림 1은 여성화자가 발성한 동화 아기돼지 삼 형제를 요약한 5분 30초 짜리 음성에 대한 홀수 번째 LSP 파라미터의 분포특성을 나타낸 실험 결과이다. 그림 1에서와 같이 홀수 번째 LSP 파라미터의 분포특성은 어느 특정 주파수 대역에 주로 분포함을 알 수가 있

다. 그림 1의 분포도에 의하여 멜 스케일로 조정되어진 검색대역의 순서를 수식으로 나타내면 식(3),(4)로 표현 가능하다. 즉, 식(3),(4)는 검색간격 양 끝점에 해당하는 위치의 다항식의 값이 되고 홀수 번째 근을 찾기 위해서 적용되어지는 식이다. 예를 들어 식(3)에서 5번째 근을 검색할 시에는 멜 스케일로 조정되어진 포인트 값 221~262까지 검색되어지고 이 범위는 약 1200Hz에서 1600Hz 사이가 된다. 만일 이 범위에서 근을 찾지 못한 경우, 멜 스케일로 조정되어진 262~294 사이에서 근을 검색하게 되고 이 범위는 1600~2000Hz에 해당한다. 또한 식(5),(6)은 짝수 번째 근을 찾을 때 적용되어지는 식이고, 이미 찾아진 홀수 번째 근 사이에서 근을 찾기 위해 검색되어진다. 제안한 방법에서 고려한 검색 순서와 조절된 포인트 값은 표 1에 나타나있다. 예를 들어 9번째 LSP 파라미터를 검색할 때는 먼저, 멜 포인트 344~364를 검색하는데 이는 주파수 2800~3200Hz에 해당됨을 의미하고 이 구간에서 근을 찾지 못한 경우는 364~383을 검색하는데, 이는 주파수 3200~3600Hz에 해당한다.

(3)

$$Q_n'(f_n) = [A_0 \cos(2\pi f_n * \frac{p}{2}) + A_1 \cos(2\pi f_n * \frac{p-2}{2}) + \dots + \frac{1}{2} A_{p/2}]$$

- 1st $f_n = 1, \dots, 108, 108, \dots, 175$
- 3rd $f_n = 108, \dots, 175, 175, \dots, 224, 1, \dots, 108$
- 5th $f_n = 221, \dots, 262, 262, \dots, 294, 175, \dots, 224, 294, \dots, 321$
- 7th $f_n = 294, \dots, 321, 321, \dots, 344, 261, \dots, 293$
- 9th $f_n = 344, \dots, 364, 364, \dots, 383, 321, \dots, 344$

$$Q_i'(f_i) = [A_0 \cos(2\pi f_i * \frac{p}{2}) + A_1 \cos(2\pi f_i * \frac{p-2}{2}) + \dots + \frac{1}{2} A_{p/2}]$$

$$f_i = f_n + 1 \quad (4)$$

$$P_n'(f_n) = [A_0 \cos(2\pi f_n * \frac{p}{2}) + A_1 \cos(2\pi f_n * \frac{p-2}{2}) + \dots + \frac{1}{2} A_{p/2}]$$

$$f_n = LSP(\psi_m), \quad m = 1, 3, 5, 7, 9 \quad (5)$$

$$P_i'(f_i) = [A_0 \cos(2\pi f_i * \frac{p}{2}) + A_1 \cos(2\pi f_i * \frac{p-2}{2}) + \dots + \frac{1}{2} A_{p/2}]$$

$$f_i = LSP(\psi_{m+2}), \quad m = 1, 3, 5, 7, 9 \quad (6)$$

만일, $Q_i'(f_i) = 0$ 인 경우에 있어서 찾고자 하는

LSP 파라미터 값이 된다.

또한, $Q'_n(f_n) * Q'_i(f_i) < 0$ 인 경우는 근이 존재한다고 판단을 하고 다음의 과정을 거치면서 LSP 값이 구해지게 된다.

$$f_{mid} = (f_n + f_i) / 2$$

$$\text{if } f_{mid} - f_n < intv_{thres} \quad (7)$$

$$LSP(\psi) = f_n + |Q'_n(f_n)| / |Q'_i(f_i) - Q'_n(f_n)| * (f_i - f_n)$$

그리고 $f_{mid} - f_n > intv_{thres}$ 의 경우는 검색 간격을 좁혀나가면서 근을 검색하게 된다. 즉, 이 경우에 있어서도 $Q'(f_{mid}) = 0$ 이 되면 찾고자 하는 LSP 파라미터 값이 된다. 또한 $Q'(f_{mid}) * Q'_n(f_n) < 0$ 인 경우에 있어서 $f_i = f_{mid}$, $Q'_i(f_i) = Q'(f_{mid})$ 로 설정이 바뀌고 $Q'(f_{mid}) * Q'_i(f_i) < 0$ 의 경우라면 $f_n = f_{mid}$, $Q'_n(f_n) = Q'(f_{mid})$ 로 설정이 바뀐다. 그 다음은 식(7)의 과정을 거치면서 근을 검색하게 된다.

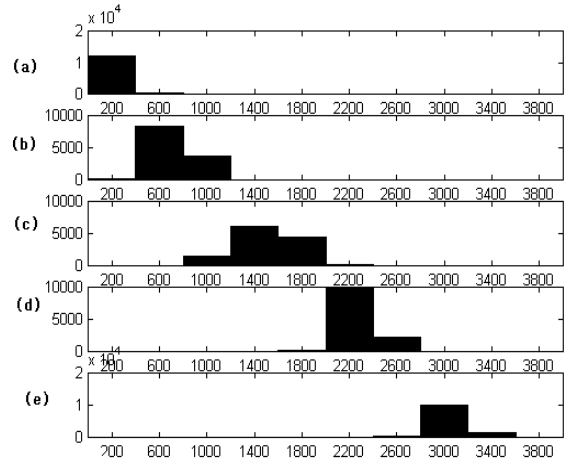
4. 실험 및 결과

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

- 발성1: “인수네 꼬마는 천재소녀를 좋아한다.”
- 발성2: “창공을 날으는 인간의 도전은 끝이 없다.”
- 발성3: “예수님께서 천지창조의 교훈을 말씀하셨다.”
- 발성4: “OO대학교 음성통신 연구팀이다.”

본 논문에 사용한 실근 알고리즘은 C-언어로 구현된 CELP 부호화기에서 발췌하여 사용하였다. 먼저 C-언어로 구현된 실근 방법의 LSP 변환부분에 멜스케일과 LSP 파라미터의 분포특성을 이용하여 제안한 알고리즘을 구현하였다. 모의 실험의 첫 번째 단계에서는 LPC 계수에서 LSP 파라미터로 변환시 계산량과 LSP 파라미터 값의 변화를 측정하였다. 그림 2는 기존의 방법과 제안한 방법을 사용하여 찾아진 LSP 파라미터의 분포를 나타낸 실험 결과이다. 그림 2의 (a)는 음성시료 1을 사용한 경우 임의의 무성음 프레임에 대한 음성 파형이고, (b)는 스펙트럼 분석을 나타낸다. (c)는 실근 방법을 이용한 경우 찾아진 LSP 파라미터의 분포를 나타내고 (d)는

제안한 알고리즘을 적용하여 변환된 LSP 파라미터의 분포특성이다. 그림 2에서 알 수 있듯이 기존의 방법과 제안한 방법 간의 LSP 파라미터의 값에는 차이가 없다. 표 2, 3은 음성시료 4가지에 대한 계산량 측정결과와 기존의 방법에 비하여 계산량 감소율을 나타낸 결과이다. 4가지 음성시료에 대하여 제안한 방법을 사용한 경우 계산량 감소는 각각, 50.23%, 48.72%, 49.26%, 47.58%로 나타난다.



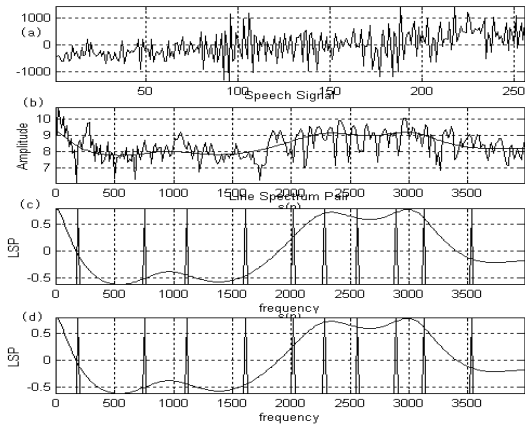
[그림 1] 홀수 번째 LSP 파라미터의 분포특성
 (a) 1 번째 LSP 파라미터
 (b) 3 번째 LSP 파라미터
 (c) 5 번째 LSP 파라미터
 (d) 7 번째 LSP 파라미터
 (e) 9 번째 LSP 파라미터

[표 1] 본 논문에서 고려한 검색간격과 순서조절

검색순서 LSP계수	1	2	3	4	5
LSP 1	1/108	108/175	otherwise		
	0/400	400/800			
LSP 3	108/175	175/224	otherwise		
	400/800	800/1200			
LSP 5	221/262	262/294	175/224	294/321	otherwise
	1200/1600	1600/2000	800/1200	2000/2400	
LSP 7	294/321	321/344	261/293	otherwise	
	2000/2400	2400/2800	1600/2000		
LSP 9	344/364	364/383	otherwise		
	2800/3200	3200/3600			

5. 결론

저 전송률 음성부호화기에서의 음성신호 선형예측 방법 중 LSP 파라미터를 이용하는 방법이 가장 많이 사용되어진다. 왜냐하면 LSP 파라미터가 낮은 전송률에서도 왜곡이 적고 선형 보간 특성이 뛰어나기 때문이다. 본 연구에서는 LPC 계수에서 LSP 파라미터로 변환하는 과정에서 근을 찾기 위한 검색대역에 멜 스케일과 홀수 번째 LSP 파라미터의 분포특성을 이용하여 변환과정의 계산량 감소를 도모하였다. 4 가지 음성시료에 대한 실험결과 계산량은 각각, 50.23%, 48.72%, 49.26%, 47.58% 정도 감소하였지만 찾아진 LSP 파라미터의 값에는 변동이 없었다.



[그림 2] 실험에서 얻어진 LSP 파라미터의 분포
 (a)음성파형
 (b)음성신호의 스펙트럼 분석
 (c)실근 방법에 의해 얻어진 LSP 파라미터
 (d)제안한 방법에 의해 얻어진 LSP 파라미터

[표 2] 제안한 알고리즘의 계산량과 감소율(1)

		실근 방법	제안한 방법	감소율 (단위:%)
음성시료 1	+	1,195,566	716,538	50.26
	-	872,592	500,352	50.23
	*	3,567,588	1,867,560	50.38
	/	217,653	111,789	50.04
음성시료 2	+	1,394,704	835,838	49.04
	-	1,017,942	583,662	49.06
	*	4,161,858	2,178,492	47.95
	/	253,908	130,400	48.83

[표 3] 제안한 알고리즘의 계산량과 감소율(2)

		실근 방법	제안한 방법	감소율 (단위:%)
음성시료 3	+	1,566,650	938,833	49.29
	-	1,143,447	655,587	49.27
	*	4,674,993	2,446,926	49.40
	/	285,213	146,467	49.08
음성시료 4	+	1,285,966	770,648	47.61
	-	938,582	538,142	47.59
	*	3,837,398	2,008,580	47.70
	/	234,113	120,229	47.40

참고문헌

[1] 배명진, “디지털 음성분석”, 동영출판사, 1998.4.
 [2] L. R. Rabiner, R.W. Schafer, “Digital Processing of Speech signal”, Prentice Hall, 1978.
 [3] A. M. Kondoz, “Digital Speech”, John Wiley & Sons Ltd, 1994.
 [4] ITU-T Recommendation G.723.1, March, 1996.