음성 부호화기에서 불균등 간격조절을 통한 계산량 단축법

전우진*
*(주)디아트그래픽스
e-mail:woojin@naver.com

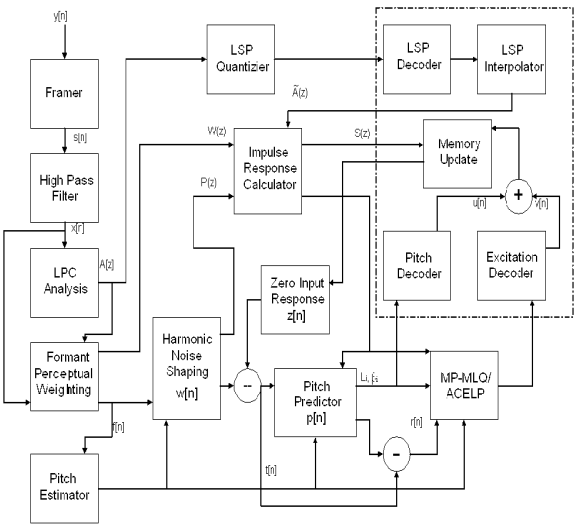
A Reduction Method of Computational Complexity through Adjustment the Non-Uniform Interval in the Vocoder

Woo-Jin Jun*
*D Art Graphics Co., Ltd.

LSP(Line Spectrum Pairs) Parameter is used for speech analysis in vocoders or recognizers since it has advantages of constant spectrum sensitivity, low spectrum distortion and easy linear interpolation. However the method of transforming LPC(Linear Predictive Coding) into LSP is so complex that it takes much time to compute. Among conventional methods, the real root method is considerably simpler than others, but nevertheless, it still suffers from its indeterministic computation time because the root searching is processed sequentially in frequency region. We suggest a method of reducing the LSP transformation time using voice characteristics.

1. 서 론

현재까지 발표된 음성부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 CELP (Code Excited Linear Prediction) 구조이다. 이 방식은 4.8kbps 내외의 전송률에서 양호한 음질을 얻을 수 있으며 ITU-T, TIA/EIA 등 여러 국제 표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다. CELP 계열 보코더들 중에서 G.723.1은 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었다. (그림 1)의 블록도에서 나타난 G.723.1은 5.3kbps/6.3kbps의 이중 전송률을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷폰과 그 외의 이동통신용 보코더로 사용되고 있으며 낮은 전송률에 비해서 우수한 음질을 제공하고 있다. 더욱이 최적의 전송환경을 위하여 두 개의 전송률을 사용하기 때문에 다른 보코더 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 real root 방법은 근을 구하기 위해 주파수 영역을 순차적으로 검색하기 때문에 계산시간이 많이 소요된다[4].



[그림 1] G.723.1의 블록도

본 논문에서 제안하는 알고리즘은 LSP 파라미터의 분포도와 이에 따라 검색간격을 조절하여 계산시간

감소를 목적으로 한다. 실험과정에서는 LPC에서 LSP 파라미터로 변환시 계산시간 감소량와 G.723.1 보코더에서의 총 처리시간 단축에 관하여 결과를 보인다.

2. LSP 파라미터 추출

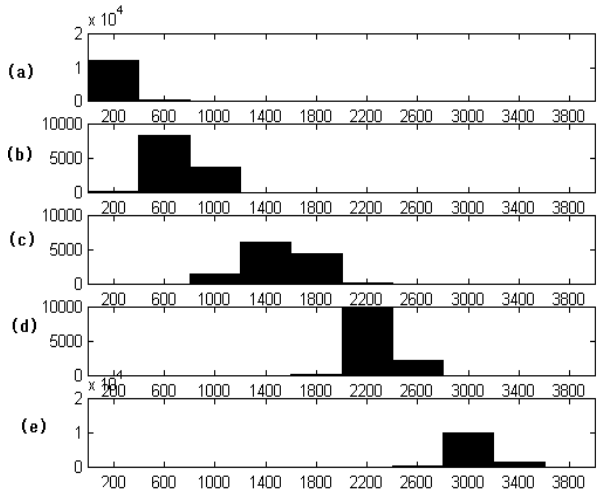
LSP 파라미터를 추출하기 위해서 먼저 LPC(Linear Predictive Coding)분석이 이루어져야 한다. LPC 분석은 고역 여파된 신호에 대해서 실행되어지며 10차의 선형 예측 분석이 수행된다. 180표본의 해밍 윈도우(Hamming window)가 각각의 부 프레임 중앙에 위치하게 되며 11개의 자기상관계수가 윈도우 처리된 신호로부터 계산되어진다. 상관계수 $R[0]$는 1025/1024의 백색 잡음 상관인자 처리가 되며 다른 10개의 상관계수는 이항 윈도우(binomial window) 계수 테이블에 의해 곱해진다. 선형 예측 계수(LPC)는 Levinson Durbin recursion을 사용하여 계산되어지며, 모든 입력 프레임, 각 부 프레임 마다 하나씩 계산되어 네 집합의 LPC 계수가 계산되어진다. 이 LPC 계수는 단구간 인지 가중화 필터를 만들기 위해 사용되는데, 이때 LPC 합성 필터는 다음과 같이 정의된다[6~8].

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

여기서 i는 0과 3사이에서 정의되는 부 프레임 인덱스이며 j는 차수를 나타낸다. LPC(Linear Predictive Coding) 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 $\{\alpha_j\}_{j=1..10}$가 단위원과 영 교차에 대한 보간 과정의 검색에 의해 LSP 계수 $\{p'_j\}_{j=1..10}$로 변환된다[6~8].

3. 검색순서와 검색간격을 조절한 방법

음성신호의 LSP 분포도의 확률적 특성을 알기 위하여 25분 정도의 긴 음성의 분포도를 조사하였다. 음성시료는 8kHz 표본화 율에 10차의 선형예측계수를 사용하였다. (그림 2)는 동화, 아기돼지 삼형제를 음성시료로 사용한 경우에 해당되는 LSP 파라미터의 분포특성이다[1~3,5].



[그림 2] LSP 홀수번째 파라미터의 분포도

[표 1] 음성시료 1 :
검색간격 조절에 따른 LSP 파라미터의 변화(1)

주파수 / LSP	검색간격(해상도)					
	5Hz	10Hz	20Hz	40Hz	80Hz	200Hz
LSP(1)	0.0198	0.0198	0.0198	에러발생		
LSP(2)	0.0827	0.0827	0.0827			
LSP(3)	0.1273	0.1273	0.1273			
LSP(4)	0.1783	0.1783	0.1783			
LSP(5)	0.2647	0.2647	0.2647			
LSP(6)	0.2910	0.2910	0.2910			
LSP(7)	0.3362	0.3362	0.3362			
LSP(8)	0.3807	0.3807	**0.3809**			
LSP(9)	0.4138	0.4138	0.4138			
LSP(10)	0.4412	0.4412	0.4412			

[표 2] 음성시료 2 :
검색간격 조절에 따른 LSP 파라미터의 변화(2)

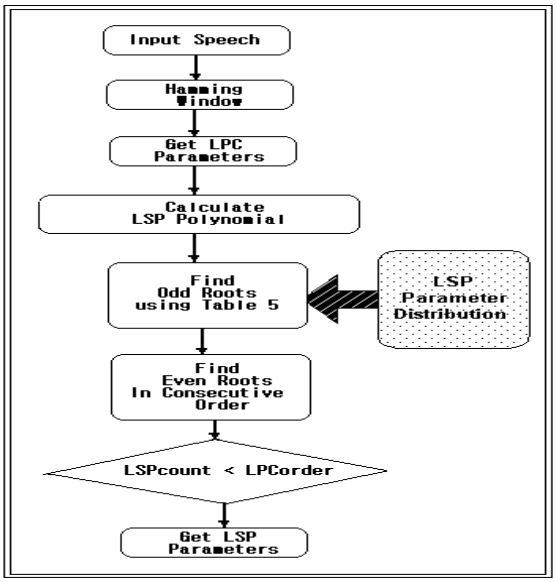
주파수 / LSP	검색간격(해상도)					
	5Hz	10Hz	20Hz	40Hz	80Hz	200Hz
LSP(1)	0.0278	0.0278	0.0279	0.0279	0.0279	**0.0316**
LSP(2)	0.0408	0.0408	0.0408	0.0408	**0.0417**	**0.0413**
LSP(3)	0.0565	0.0565	0.0565	**0.0563**	**0.0553**	**0.0527**
LSP(4)	0.0915	0.0915	**0.0911**	**0.0903**	**0.0874**	**0.0594**
LSP(5)	0.2053	0.2053	0.2053	0.2053	**0.2052**	**0.2050**
LSP(6)	0.2950	0.2950	0.2951	0.2951	**0.2953**	**0.3054**
LSP(7)	0.3245	0.3245	0.3245	0.3245	0.3245	0.3245
LSP(8)	0.3895	0.3895	0.3895	**0.3896**	**0.3915**	**0.4026**
LSP(9)	0.4279	0.4279	0.4279	**0.4278**	**0.4278**	**0.4277**
LSP(10)	0.4520	0.4520	**0.4519**	**0.4518**	**0.4494**	**0.4531**

G.723.1 보코더에서 LSP 다항식의 근을 찾기 위해 순차적으로 영교차점을 체크하여 근을 구한다. 그러나 (그림 2)를 살펴보면 분포도가 어느 특정 주파수 대역에 주로 분포함을 알 수가 있다. (표 1~4)는 음성시료 4가지에 대하여 검색간격(해상도)을 달리하였을 때 얻어지는 LSP 파라미터의 변화

결과에 해당한다. (그림 2)와 (표 1~2)에서 보여 지는 결과를 토대로 하여 LSP 파라미터를 얻기 위하여 G.723.1 보코더에 적용한 검색순서와 검색간격을 나타내었다.

[표 5] LSP 파라미터 분포특성에 따른 검색순서와 검색간격(해상도)

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	otherwise	
LSP 7	2000-2400	2400-2800	1600-2000	otherwise	
LSP 9	2800-3200	3200-3600			



[그림 3] 제안한 알고리즘의 블록도

4. 실험결과

실험을 위해 IBM PC(566 MHz)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 발성한 음성시료는 다음과 같다.

발성1: "○○대학 음성통신 연구팀이다."
발성2: "인수네 꼬마는 천재소년을 좋아한다."
발성3: "예수님께서 천지창조의 교훈을 말씀하셨다."
발성4: "창공을헤쳐 나가는 인간의도전은 끝이없다."

제안한 방법을 사용하여 LSP 파라미터 계산량 감소를 보이기 위하여 두 단계 실험을 하였다. 첫 번째는 CELP 부호화기에서 real root 알고리즘을 발췌하여 본 논문에서 제안한 알고리즘을 적용하여 LPC에서 LSP 변환시간을 측정하였다. 두 번째는 첫 번째의 과정을 G.723.1 보코더에 적용하여 부호화기 전체 성능에 어느 정도의 계산량이 단축되는지를 실험하였다. 제안한 알고리즘은 C-언어로 구현하여 수행하였다. (표 6)은 G.723.1 보코더에서 발췌한 real root 알고리즘과 제안한 방법 간의 LPC에서 LSP 파라미터로 변환하는 시간을 측정한 결과를 나타내고 평균 46%의 수치로 나타났다. (표 7)은 G.723.1 Annex A를 통과한 경우와 제안한 알고리즘을 통과한 경우를 비교한 처리시간 비교결과이다. 실험결과 보코더의 처리시간이 평균 5.27% 단축됨을 알 수가 있다.

[표 6] LPC에서 LSP로의 변환시간 (단위: sec)

비교 / 시료	순차적 검색 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
발성 1	1.95	1.07	45.13
발성 2	1.59	0.84	47.17
발성 3	2.35	1.21	48.5
발성 4	2.2	1.22	44.55

[표 7] G.723.1 보코더의 총 처리시간(단위: sec)

비교 / 시료	순차적 검색 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
발성 1	6.860	6.512	5.07
발성 2	6.530	6.250	4.28
발성 3	7.580	7.210	4.88
발성 4	7.910	7.370	6.83

5. 결 론

본 논문에서는 G.723.1 보코더에서 LSP 변환시 다항식의 근을 찾는 순서를 음성신호의 LSP 분포특성에 맞게 조정함으로써 계산시간을 단축하였다. 즉, 비교적 긴 음성시료들의 LSP 파라미터 누적분포도를 통계적으로 분석하여, 순차적으로 근을 검색하지 않고 파라미터의 분포특성을 이용하여 검색순서와 해상도를 조정하는 새로운 알고리즘을 제안하였다. 실험결과 검색시간은 단축되었고 파라미터

값은 기존의 방법과 변화가 없었다. 또한 G,723.1 음성부호화기에 제안한 알고리즘을 적용한 결과, 계산시간은 단축되면서 음질의 열하는 발생하지 않았다. 제안한 알고리즘은 real root 방법을 사용하는 CELP 계열 부호화기에 광범위하게 적용되어질 수 있다. 향후 보다 효과적이고 일정한 보코더의 계산량을 감소하기 위하여 외국어 음성시료의 LSP 파라미터 분포특성이 검토되어져야한다.

참고문헌

[1] 한진수, “음성신호처리”, 34-61, 오성미디어, 2000.
[2] 배명진, “디지탈 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.