

음성부호화기에서 가변 대역폭에 의한 계산시간 단축방법에 관한 연구

민소연*, 이광형**, 김정재***

*서일대학 정보통신과

e-mail:symin@seoil.ac.kr

A Study on the Reduction Algorithm of Transformation Time by Variable Bandwidth in Vocoder

So-Yeon Min*, Kwang-Hyoung Lee**, Jung-Jae Kim***

*Dept of Information&Communication Seoil University

**Dept of Internet Information Seoil University

요 약

CELP 계열 음성부호화기는 4.8kbps 내외의 전송률에서 양호한 음질을 얻을수 있으며 ITU-T, TIA/EIA 등 여러 국제 표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어졌다. 이중에서 G.723.1은 멀티미디어 통신 환경하의 음성 전송 표준 보코더로 개발되었으며 G.723.1은 5.3kbps/6.3kbps의 이중 전송률을 갖는 구조로 현재 별정 통신으로 상용화되는 인터넷폰과 그 외의 이동 통신용 보코더로 사용되고 있다. G.723.1에서 사용되는 LSP 파라미터는 일정한 스펙트럼 민감도와 낮은 스펙트럼 왜곡을 보인다. 본 논문에서 제안하는 알고리즘은 음성신호의 에너지가 대개 낮은 주파수에 분포한다는 특징을 이용하여 대역폭을 가변으로 조정하였다. 실험결과, 부호화기 전체 계산시간은 약 6%의 감소하였다.

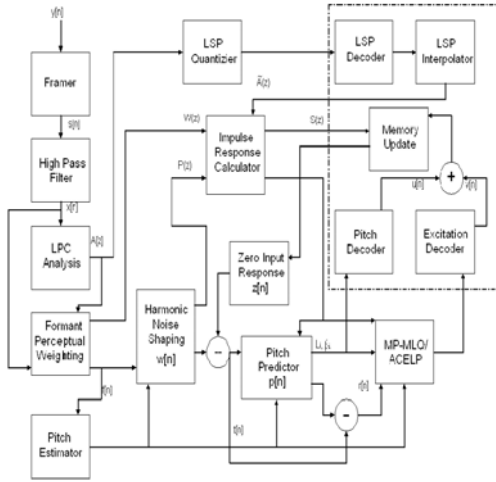
1. 서 론

음성부호화란 음성 신호를 디지털 부호로 변환시키는 것을 말하며 음성의 전송 또는 저장 등에 사용된다. 음성부호화의 주된 기능은 음성 신호를 분석하여 잉여성분을 제거한 후에 제거되지 않는 나머지 성분들을 타당한 방법을 이용하여 부호화 하는 것이다[1]. 표준 음성 부호화기는 1972년에 ITU-T recommendation G.711로 채택된 64kbps PCM (Pulse Code Modulation) 방식으로부터 출발하여, 32kbps의 ADPCM (adaptive DPCM), 16kbps의 LD-CELP(Low-Delay CELP) 방식으로 표준화되었다. 현재 ITU-T에서는 PCS, IMT-2000등에서 사용할 수 있는 8kbps 음성부호화기에 대한 표준화 작업으로 1996년에 CS-CELP(conjugated structure algebraic CELP)를 G.729로, 그리고 인터넷폰 및 화상 통신용 음성 부호화기로 ACELP/MP-MLQ (algebraic CELP/Multipulse Maximum Likelihood

Quantization)의 5.3/6.3kbps dual rate를 G.723.1 draft recommendation으로 선정하였다[1]. 저전송율 CELP 부호화기에서 LSP(Line Spectrum Pairs) 파라미터의 계산시간은 codebook을 검색하는데 걸리는 시간 다음으로 부호화기의 많은 계산시간을 차지한다. 현재 음성 코덱이나 인식기에서 음성신호를 분석하여 전송형이나 저장형 파라미터로 변환되는 LSP 파라미터는 일정한 스펙트럼 민감도와 낮은 스펙트럼 왜곡을 보이고 선형 보간이 용이하다[2~6]. 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 음성신호의 에너지가 대개 낮은 주파수에 분포한다는 특징을 이용하여 검색대역에 멜 스케일을 이용하였다. 멜 스케일은 1000Hz 이하에서는 주파수와 선형적으로 비례하며 1000Hz 이상에서는 로그 스케일이다. 제안한 방법의 성능평가를 위해 두 단계의 실험을 하였다.

즉, 제안한 방법을 사용하여 LPC에서 LSP 파라미터로 변환시 계산시간 감소량과 G.723.1 보코더에서의 총 처리시간 단축에 관하여 결과를 보인다.

2. G.723.1 음성 부호화기



[그림 1] G.723.1의 블록도

부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있고 인지 가중화된 오차 신호의 최소화를 시도한다. 부호화기는 240샘플 프레임마다 처리한다. 8kHz 샘플링율에서 30ms이다. 각 프레임은 DC 성분을 제거하기 위해 하이 패스 필터를 통과하고, 60샘플의 4개의 부프레임으로 나누어진다. 모든 부프레임에서 10차의 선형 예측 부호화기 필터는 입력신호를 사용하여 계산되어진다. 마지막 부프레임을 위한 선형 예측 부호화기 필터는 예측 분할 벡터 양자화를 사용하여 양자화된다. 양자화 되지 않은 선형 예측 부호화기의 계수는 인지 가중화된 음성 신호를 얻기 위한 단구간 인지 가중화 필터를 구축하기 위해 사용되고, 전체 프레임을 필터링하기 위해 사용하고 인지 가중화된 음성신호를 얻기 위해 사용된다. 모든 두 개의 부 프레임 (120 samples)을 처리하기 위한 개회로 피치 주기 예측은 인지 가중화된 음성 신호를 사용하여 계산되어진다. 피치 주기는 18에서 142표본의 구간에서 찾으며 피치 예측은 120표본의 블록 단위로 실행되어진다. 이전에 계산되어 예측된 피치 주기를 사용하여 하모닉 노이즈 셰이핑 필터가 만들어지며 LPC 합성 필터, 포먼트 인지가중화 필터, 하모닉 노이즈 셰이핑 필터의 결합은 임펄스 응답을 생성하기 위해 사용된다. 피치 주기 예측 값과 임펄스 응답을 이용

하여 폐회로 피치 예측기 값이 계산되어진다. 5차 피치 예측기가 사용되는데 피치 주기는 개회로 피치 예측값 주위의 차이 값이 작은 구간에서 계산되어진다. 피치 예측기는 초기 목적벡터로부터 값을 뺄 때 이용되는데 피치 주기와 차이 값 둘 모두 복호화기로 전송되어진다. 마지막으로 여기신호의 비주기 요소가 근사화 되어진다..

3. LPC 분석과 LSP 파라미터의 추출

LPC 분석은 고역 여파된 신호에 대해서 실행되어지며 10차의 선형 예측 분석(Linear predictive analysis)이 수행된다. 180표본의 해밍 윈도우가 각 부프레임의 중앙에 위치하게 되며 11개의 자기 상관 계수가 윈도우 처리된 신호로부터 계산되어진다. 상관계수 $R[0]$ 는 1025/1024의 백색 잡음 상관 인자 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 프레임의 각 부프레임마다 하나씩 계산되어 네 집합의 LPC 계수가 계산되어진다. 이 LPC 계수는 단구간 인지가중화 필터를 만들기 위해 사용되는데, 이때 LPC 합성 필터는 다음과 같이 정의된다[1][2][3][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 확장이 수행된다. 그 결과로 $A_3(z)$ LP 필터는 예측 분할벡터 양자화를 이용하여 양자화 되어진다. 양자화 과정은 LP 계수 $\{a_j\}_{j=1..10}$ 가 단위원과 영교차에 대한 인터플레이션 과정의 검색에 의해 LSP 계수 $\{p'_j\}_{j=1..10}$ 로 변환된다. DC 성분 p_{DC} 가 LSP 계수 p' 로부터 제거되면 LSP 벡터 p 가 얻어진다. 1차 고정 예측기 b(12/32)는 프레임에서 DC 성분이 제거된 예측 LSP 벡터 \bar{p}_n 과 잔여 LSP 에러 벡터 e_n 값을 얻기 위해서 이전에 복호화된 LSP 벡터 \tilde{p}_{n-1} 에 적용된다. 과정은 다음의 식에 의해 나타난다.

$$p^T_n = [p_{1,n}, p_{2,n} \dots p_{10,n}] \quad (2)$$

$$\bar{p}^T = [\bar{p}_{1,n}, \bar{p}_{2,n} \dots \bar{p}_{10,n}] \quad (3)$$

$$\bar{p}_n = b[\tilde{p}_{n-1} - p_{DC}] \quad (4)$$

$$e_n = p_n - \bar{p}_n \quad (5)$$

양자화 되지 않은 LSP 벡터 p'_m , 양자화된 LSP 벡터 \tilde{p}_n , 잔여 LSP 에러 벡터 e_n 은 각각 3, 3, 4의 크기를 갖는 3개의 부프레임으로 분리된다. 각 m번째 부벡터는 8비트 코드북을 사용하여 양자화된 벡터이며 에러 표준값 $E_{l,m}$ 이 최소가 되는 적당한 부벡터 코드북 기록의 인덱스 l 가 선택되어진다.

$$p'^T_m = [p'_{1+3m}, p'_{2+3m} \dots p'_{K_m+3m}], \quad K_m = \begin{cases} 3, & m=0 \\ 3, & m=1 \\ 4, & m=2 \end{cases}$$

$$\tilde{p}^T_{l,m} = [\tilde{p}_{1,l,m}, \tilde{p}_{2,l,m} \dots \tilde{p}_{K_m,l,m}], \quad 0 \leq m \leq 2, 1 \leq l \leq 256$$

$$p' = p + p_{DC} \quad (6,7,8)$$

$$\tilde{p}_{l,m} = \bar{p}_m + p_{DC_m} + e_{l,m}, \quad 0 \leq m \leq 2, 1 \leq l \leq 256 \quad (9)$$

$$E_{l,m} = (p'_m - \tilde{p}_{l,m})^T W_m (p'_m - \tilde{p}_{l,m}), \quad 0 \leq m \leq 2, 1 \leq l \leq 256 \quad (10)$$

여기서 $e_{l,m}$ 은 m번째로 분할된 잔여 LSP 코드북의 l번째 기록이고 W_m 은 직교 가중 행렬이다. 이 행렬은 양자화 되지 않은 LSP 계수 벡터 p' 로부터 결정된다. 여기서 선택된 인덱스들이 채널로 전송된다. LSP 계수들의 복호화 과정은 세 개의 부벡터 $\{e_{m,n}\}_{m=0..2}$ 들이 10차의 벡터 \tilde{e}_n 으로 복호화되면서 시작된다. 예측벡터 \bar{p}_n 는 복호화된 벡터 \tilde{e}_n 와 DC 벡터 p_{DC} 와 합해져 복호화된 LSP 벡터 \tilde{p}_n 을 만든다. 안정성 확인은 다음 식에 따라 복호화된 LSP 벡터를 확인하기 위해 복호화된 LSP 벡터 \tilde{p}_n 에서 수행된다.

$$\tilde{p}_{j+1,n} - \tilde{p}_{j,n} \geq \Delta_{\min}, 1 \leq j \leq 9 \quad (11)$$

여기서 Δ_{\min} 은 31.25Hz이다. 만약 안정성 확인이 \tilde{p}_j 와 \tilde{p}_{j+1} 동안 실패하면, \tilde{p}_j 와 \tilde{p}_{j+1} 는 다음과 같이 수정된다.

$$\tilde{p}_{avg} = (\tilde{p}_j + \tilde{p}_{j+1})/2, \quad \tilde{p}_j = \tilde{p}_{avg} - \Delta_{\min}/2 \quad (12)$$

$$\tilde{p}_{j+1} = \tilde{p}_{avg} + \Delta_{\min}/2 \quad (13)$$

수정은 안정성이 만족될 때까지 수행된다. 만약 10번을 반복하여도 안정 조건이 만족하지 않으면 이전 LSP 벡터가 사용된다. 선형 보간은 복호화된 LSP 벡터 \tilde{p}_n 와 이전 LSP 벡터 \tilde{p}_{n-1} 사이에서 각 부 프레임마다 수행된다. 네개의 인터폴레이션된 LSP 벡터 $\{\tilde{p}_i\}_{i=0..3}$ 는 LPC벡터 $\{\tilde{a}_i\}_{i=0..3}$ 로 변환된다.

$$\tilde{p}_{ni} = \begin{cases} 0.75 \tilde{p}_{n-1} + 0.25 \tilde{p}_n, & i=0 \\ 0.5 \tilde{p}_{n-1} + 0.5 \tilde{p}_n, & i=1 \\ 0.25 \tilde{p}_{n-1} + 0.75 \tilde{p}_n, & i=2 \\ \tilde{p}_n, & i=3 \end{cases} \quad (14)$$

$$\tilde{a}_i^T = [\tilde{a}_{i1}, \tilde{a}_{i2}, \dots, \tilde{a}_{i10}]^T, \quad 0 \leq i \leq 3 \quad (15)$$

4. 실근 방법과 멜 스케일

실근 LSP 변환 방법에서는 홀수 번째의 계수를 먼저 찾은 후 짝수 번째의 계수는 이미 찾아진 홀수 번째 계수 사이에서 찾는다. 그리고 실근 방법에서는 홀수 번째의 계수를 찾을 때 주파수대역을 순차적으로 검색한다. 따라서 홀수 번째의 계수를 찾는 데 걸리는 시간이 변환시간의 대부분을 차지하게 된다. 그리고 멜 스케일이란 인간의 청각특성을 고려한 주파수 척도이며 1000Hz에 1000Mel을 대응하고 이로부터 실험적으로 값을 결정한다. 일반적인 의미의 멜 스케일은 1000Hz 이하에서는 주파수와 선형적으로 비례하며 1000Hz 이상에서는 로그 스케일이다. 본 논문에서는 Fant에 의하여 근사화된 식을 이용하였고 식 (16)에서 나타내었다. 식 (16)은 멜 스케일과 주파수 사이의 관계를 나타내고 있다[7~9].

$$(16)$$

$$F_{mel} = k * \log [1 + \frac{f}{1000}]$$

여기서, $k=1000/\log 2$

실험에서 실근 방법과 멜 스케일을 이용하여 주파수 대역을 검색한 경우를 비교를 위하여 두 단계의 실험을 하였다. 첫 번째는 LPC에서 LSP로의 변환과정에 있어서, 제안한 방법의 계산시간 감소를 측정하였고, 두 번째는 부호화기를 통과하였을 때 총 계산시간의 단축정도를 측정하였다. 본 논문에서 제안한 알고리즘의 블록도를 그림 2에서 나타낸다. 즉, 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 사이

에서 짝수 번째 근을 찾기 시작한다. 위의 과정은 10개의 근을 찾을 때까지 반복된다. 제안한 알고리즘이 시뮬레이션은 C-언어로 구현하여 수행하였다. 표 1은 G.723.1 보코더에서 발췌한 실근 알고리즘과 제안한 방법간의 LPC에서 LSP 파라미터로 변환하는 시간을 측정된 결과를 나타내고 평균 48%의 수치로 나타났다. 표 2는 G.723.1 Annex A를 통과한 경우와 제안한 알고리즘을 통과한 경우를 비교한 처리시간 비교결과이다. 실험결과 보코더의 처리시간이 평균 6% 단축됨을 알 수가 있었다.

5. 결론

저전송률 음성부호화기에서의 음성신호 선형예측 방법 중 LSP 파라미터를 이용하는 방법이 가장 많이 사용되어진다. 제안한 방법에서는 음성부호화기에서 주로 사용되는 실근 방법에 멜 스케일을 사용하여 계산시간을 단축하였다. 즉, 음성신호는 1kHz의 범위에 중요한 성분을 많이 가지고 있다는 성질을 이용하여 1kHz 까지는 검색간격을 선형적으로 그 이상의 주파수 대역에서는 로그 스케일로 검색하였다. 향후, 제안한 알고리즘이 음질에 미치는 연구가 이루어져야겠다.

[표 1] LPC에서 LSP로의 변환시간 (단위: sec)

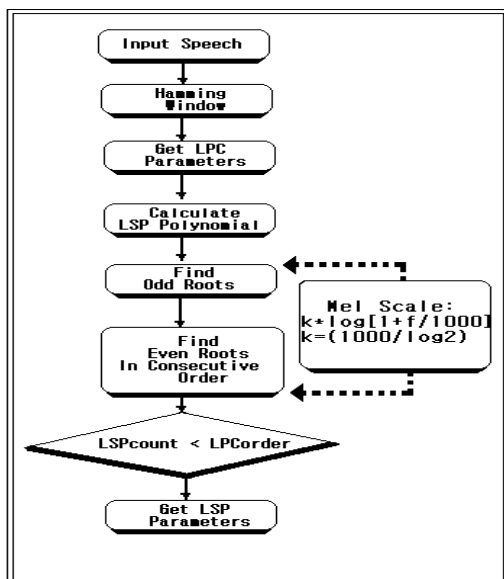
비교 시료	Real Root 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
발성 1	1.58	0.82	48.10
발성 2	1.96	1.02	47.95
발성 3	2.35	1.19	49.36
발성 4	2.20	1.20	45.45

참고 문헌

[표 2] G.723.1에서의 총 계산시간 (단위: sec)

비교 시료	Real Root 방법 (단위:sec)	제안한 방법 (단위:sec)	감소율 (단위:%)
발성 1	6.12	5.8	5.5
발성 2	6.52	6.2	6.5
발성 3	7.57	6.1	5.92
발성 4	7.90	6.2	6.81

- [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", pp129, MARCEL DEKKER, INC. 1991.
- [7] Thomas Parson, "Voice and Speech Processing", pp71-73, Mac Graw Hill.
- [8] 민소연, 정찬중, 배명진, “주파수 대역의 적응적 조절을 통한 LSP 변환 알고리즘의 성능 개선에 관한 연구”, 한국통신학회, 하계종합학술 발표회 논문집, 2001년 7월 6일.
- [9] 민소연, 배명진. “LSP 파라미터 분포특성을 이용한 주파수 대역 조절법에 관한 연구”, 음향학회지, 제21권, 제3호, 2002년 4월.



[그림 2] 제안한 알고리즘의 블록도