

G.723.1 MP-MLQ 고정 코드북 검색 시간 단축에 관한 연구

김정진, 장경아, 목진덕, 배명진, 홍성훈*, 성유나*

승실대학교 정보통신공학과

*C&S Technology 음성처리개발팀

mjbae@saint.soongsil.ac.kr

The Research of Reducing the Fixed Codebook Search Time of G.723.1 MP-MLQ

JeongJin KIM, GyungA JANG, JinDuck MOK, MyungJin BAE,
SeongHoon HONG*, YooNa SUNG*

Dept. of Telecomm. Engr., Soongsil Univ.

*C&S Technology Speech Processing Team

본 연구는 C&S Technology의 지원비에 의해 연구되었습니다.

Abstract

In general CELP type vocoders provide good speech quality around 4.8kbps. Among them, G.723.1 developed for Internet Phone and videoconferencing includes two vocoders, 5.3kbps ACELP and 6.3kbps MO-MLQ. Since 6.3kbps MP-MLQ requires large amount of computation for fixed codebook search, it is difficult to realize real time processing.

In order to improve the problem this paper proposes the new method that reduces the processing time up to about 50% of codebook search time. We first decide the grid bit, then search the codebook. Grid bit is selected by comparison between synthetic speech, which is synthesized with only odd or even pulses of target vector, and DC removed original speech.

As a result, we reduced the total processing time of G.723.1 MP-MLQ up to about 26.08%. In objective quality test 11.19dB of segSNR was obtained, and in subjective quality test there was almost no speech degradation.

I. 서론

현재까지 발표된 음성 부호화기 중 가장 많은 연구가 이루어지고 있는 방식은 CELP(Code Excited Linear Prediction)구조이다. 이 방식은 4.8kbps내외의 전송율에서 양호한 음질을 얻을 수 있으며 ITU-T, TIA/EIA 등

여러 국제표준화 기구를 통해 다양한 응용분야에서 표준화가 이루어지고 있다[1].

이들은 인터넷 폰이나 화상회의, voice mail system, voice-pager 등에 응용이 가능하며 현재 상용버전으로 true speech나 G.723.1[2] 보코더가 나와 사용되고 있다. 이 중 그림 1에 도시되어 있는 G.723.1은 5.3/6.3kbps의 이중 전송율을 갖는 구조로 현재 별정 통신으로 상용화 되는 인터넷 폰과 그 외의 이동 통신용 보코더로 사용되어지고 있으며 낮은 전송율에 비해서 우수한 음질을 제공하고 있다. 더불어 최적의 전송 환경을 위하여 두 개의 전송율을 사용하기 때문에 다른 보코더 표준안들에 비해서 더욱 응용성이 높다. 그러나 G.723.1 역시 음성신호를 성분 분리하여 합성하는 방식인 CELP 보코더 계열의 합성에 의한 분석방법을 사용하기 때문에 많은 계산량으로 인한 처리 시간의 소모를 피할 수 없다는 문제점을 갖고 있다[3]-[5]. 또한 G.723.1 Dual Bit rate speech codec은 두개의 서로 다른 보코더를 포함하고 있어 DSP칩으로 구현시 많은 내부 메모리와 계산량을 필요로 하며 특히 MP-MLQ(Multi-Pulse Maximum Likelihood Quantization) 모드는 ACELP(Algebraic CELP)에 비해 많은 계산량을 요구한다. 따라서 PC를 기반으로 하는 인터넷 폰에는 저가의 DSP를 사용 가능하게 할 수 있는 적은 알고리즘 계산량을 요구하는 보코더 알고리즘이 적합하다고 볼 수 있다.

따라서 본 논문에서는 음성부호화기를 실시간으로 구현하는 경우에 발생하는 문제점들을 극복하여 저전송율을 환경하에서 계산량을 감소시키는 기법을 제안하고자 한다. 이를 위하여 고음질 저전송 보코더인 G.723.1에 관

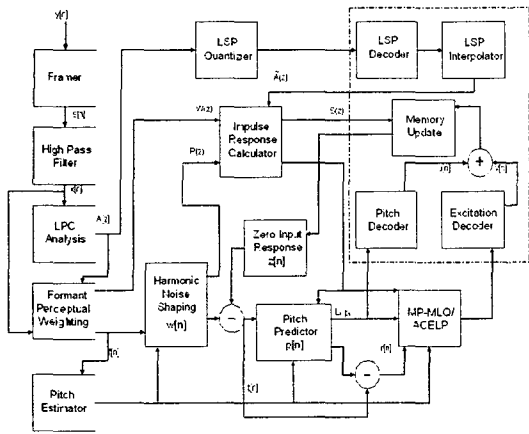


그림 1. G.723.1 CELP 보코더에 대한 블록도

기반 기술을 확보하는데 있다. 본 논문에서는 G.723.1 MP-MLQ의 최적 펄스 검색시 Grid 비트의 사전 결정을 통해 전체적으로 작/홀수에 동일하게 소요되는 검색을 절반으로 줄이는 방법을 제한한다.

II. G.723.1 6.3kbps MP-MLQ[2]

간차 신호 $\{r[n]\}_{n=0-59}$ 는 새로운 목적 벡터로써 MP-MLQ 블록에 전송된다. 이 블록은 이 벡터에 대한 양자화를 수행한다. 양자화 과정은 목적 벡터 $r[n]$ 을 $r'[n]$ 을 이용하여 근사화하는 과정이다.

$$r'[n] = \sum_{j=0}^n h[j] \cdot v[n-j], \quad 0 \leq n \leq 59 \quad (1)$$

여기서 $v[n]$ 은 임펄스 응답 $h[n]$ 을 갖는 조합필터 $S(z)$ 에 대한 여기원이고 다음과 같이 정의된다.

$$v[n] = G \sum_{k=0}^{M-1} \alpha_k \delta[n-m_k], \quad 0 \leq n \leq 59 \quad (2)$$

G 는 이득, $\delta[n]$ 은 Dirac 함수, $\{\alpha_k\}_{k=0-M-1}$ 와 $\{m_k\}_{k=0-M-1}$ 는 각각 Dirac 함수의 부호(± 1)와 위치를 나타낸다. 그리고 M 은 펄스의 개수를 나타내는데 짝수 번째 부프레임에서는 6을, 홀수번째 부프레임에서는 5의 값을 갖는다. 펄스의 위치를 선정하는데는 제한이 있다. 펄스의 위치는 모두가 짝수번째 펄스이거나 홀수번째 펄스이어야 한다. 이것을 나타내는 것이 Grid bit이다. 그러므로 여기에서 문제점은 예러 신호 $err[n]$ 의 MSE

를 최소화시키는, 알 수 없는 파라미터 G , $\{\alpha_k\}_{k=0-M-1}$, $\{m_k\}_{k=0-M-1}$ 를 찾는 것이다 :

$$err[n] = r[n] - r'[n] = r[n] - G \sum_{k=0}^{M-1} \alpha_k h[n-m_k] \quad (3)$$

위의 파라미터에 대한 추정과 양자화는 합성에 의한 분석법에 의해 수행된다. 파라미터 G_{max} 는 다음과 같은 과정에 의해 추정되고 양자화된다. 먼저, 임펄스 응답 $h[n]$ 과 새로운 목적벡터 $r[n]$ 과의 상호상관 함수를 계산한다.

$$d[j] = \sum_{n=j}^{59} r[n] \cdot h[n-j], \quad 0 \leq j \leq 59 \quad (4)$$

추정된 이득은 다음과 같다.

$$G_{max} = \frac{\max\{|d[j]|\}_{j=0-59}}{\sum_{n=0}^{59} h[n] \cdot h[n]} \quad (5)$$

그런 다음, 추정된 이득 G_{max} 는 logarithmic 양자화에 의해 양자화된다. 이런 스칼라 이득 양자화기는 ACELP와 MP-MLQ 모두에 쓰이며 각각의 레벨이 3.2 dB인 24 단계로 구성되어 있다. 이렇게 양자화된 \hat{G}_{max} 주변으로 부가적인 이득 값들이 $[\hat{G}_{max} - 3.2,$

$\hat{G}_{max} + 6.4]$ 의 범위 내에서 선택된다. 이런 이득 값들 각각에 대해서 펄스들의 위치와 부호들은 순차적으로 최적화된다. 이러한 과정은 짝수나 홀수번째 Grid에 대해 반복된다. 마지막으로 $err[n]$ 의 MSE를 최소화하는 양자화된 파라미터들의 조합이 선택된다. 이렇게 얻어진 펄스 위치와 이득의 최적 조합이 전송되어진다. 펄스 위치를 전송하기 위해서는 $\binom{30}{M}$ combinatorial 코딩이 사용된다. 또한 고정 코드북의 코드워드 수가 2의 멱급수 형태가 아니기 때문에 부가적인 3비트는 13-비트 인덱스를 형성하기 위한 4개의 부프레임에 대해 combinatorial 코드들의 최상위 4 비트의 조합에 의해 절약될 수 있다.

피치 구간이 짧은 음성신호의 음질을 향상하기 위해 다음의 부가적인 과정이 사용된다. 만약 부프레임 0과 1의 피치 지연 L_0 가 58보다 작거나 부프레임 2과 3의 피치 지연 L_1 이 58보다 작다면, L_0 나 L_1 의 피치 주기 인덱스를 갖는 Dirac 함수의 펄스열은 위의 양자화 과정에서 사용된 단일 Dirac 함수 대신에 각 펄스 위치 m_k 에 대해 사용된다. 간차 신호를 나타내기 위해 Dirac 함수

의 펄스열을 사용할 것인지 단일 Dirac 함수를 사용할 것인지에 대한 선택은 MSE 계산에 의해 이루어진다. 즉, MSE를 최소화하는 파라미터들이 선택되고 그 값이 전송되게 된다.

III. 제안한 방법

본 논문에서는 먼저 DC 성분이 제거된 원 음성과 짝수와 홀수번째 펄스를 이용한 합성음 사이의 에러값 중에서 그 값이 적은 펄스의 Grid를 사전에 선택함으로써 최적의 펄스 검색 시간을 단축하는 방법을 제안한다.

III.1 목적 벡터 생성

다음의 수식을 이용하여 짝/홀수번째 펄스를 갖는 목적 벡터를 생성한다.

$$v_i[2 \times n + i] = \sum_{n=0}^{L/2-1} r[2 \times n + i], \quad i=0, 1 \quad (6)$$

여기서 L는 부프레임의 길이이다. 이렇게 얻어진 목적 벡터의 진폭은 G.723.1에서 수행하는 방법과 유사하게 다음과 같이 변환시킨다.

$$v_i[n] = \begin{cases} +1, & \text{if } v_i[n] > 0 \\ -1, & \text{if } v_i[n] < 0 \\ 0, & \text{otherwise} \end{cases} \quad (7)$$

III.2 Grid 비트 결정

위에서 얻은 목적 벡터와 조합필터, $S(z)$ 의 임펄스 응답 $h(n)$ 과 콘볼루션을 통해 합성음을 생성한다.

$$s'_i[m] = \sum_{n=m}^{59} v_i[n] \cdot h[n-m], \quad i=0, 1 \quad (8)$$

이렇게 얻어진 신호는 DC가 제거된 원음성과 비교되어지고 에러가 최소가 되는 i 가 Grid 비트로 결정된다.

$$err0 = \sum_{n=0}^{59} |s[n] - s'_0[n]| \quad (9)$$

$$err1 = \sum_{n=0}^{59} |s[n] - s'_1[n]|$$

$$Grid = \begin{cases} 0, & \text{if } err0 < err1 \\ 1, & \text{if } err1 < err2 \end{cases} \quad (10)$$

만약 위의 조건을 만족하지 않는 경우 G.723.1 MP-MLQ의 경우처럼 전체 짝/홀수 펄스에 대해 검색을 수행한다.

IV. 실험 및 결과

본 논문에서 제안한 알고리즘을 평가하기 위하여 컴퓨터 시뮬레이션을 실행하였다. 이용된 장비로는 IBM Pentium(166MHz)에 상용화된 AD/DA 컨버터를 인터페이스하여 사용하였다. 음성 시료는 8kHz로 샘플링하고 16 비트로 양자화하였다. 한 프레임의 길이는 240 샘플로 하고 부프레임의 길이는 60샘플로 하였다. 처리결과 성능을 측정하기 위해 라디오 뉴스를 통해서 입력받은 다음의 5문장에 대해 실험하였다. 시료 1의 경우 남성과 여성의 앵커 목소리가 포함된 문장이며, 시료 2에서 4까지는 남성 앵커 그리고 시료 5의 음성은 여성 앵커에서 추출하였다.

시료 1 : “아 서울지방의 현재 기온은 십오점 구도입니다. 오늘 예상되는 비의 양은 서울 경기 지방이 오일 리 내외 그 밖의 지방은 오에서 이십밀리 사이의 적은 양이 되겠습니다.”

시료 2 : “주택건설 촉진을 위한 합동대책회의를 열고 이러한 내용을 담은 개선안을 최종 확정했습니다.”

시료 3 : “청약 저축에 가입해서 이년이 지나면 일순위 자격으로 다시 주택을 청약할 수 있게 됩니다.”

시료 4 : “고유가 시대에 대비해서 에너지 다소비 구조를 개선하기 위한 강력한 에너지 절약 정책이 추진됩니다.”

시료 5 : “광명시는 광명 사거리 중심으로 여러 곳에서 지하철 공사가 진행되고 있습니다.”

음질 평가는 객관적 음질 평가와 주관적 음질 평가를 수행하였다. 객관적 음질 평가의 경우 segSNR을 사용하여 측정하였으며, 주관적 음질 평가의 경우 MOS test를 사용하였다. 실험 결과 객관적 평가의 경우 평균 약 11.193 dB를 얻을 수 있었으며 주관적 음질 평가의 경우 평균 약 3.884정도의 MOS score를 얻을 수 있었다. 표 1에서는 객관적 음질 평가를 표 2에서는 주관적 음질 평가에 대한 결과를 나타내고 있다. 객관적 평가에 사용한 수식은 다음과 같다.

$$SNR(dB) = 10 \log_{10} \left(\frac{E_s^2}{E_e^2} \right)$$

$$segSNR = \frac{1}{M} \sum_{m=1}^M SNR(m) \quad (dB)$$

처리 시간을 측정하기 위해 C 언어에서 제공하는 clock 함수를 사용하였다. 측정은 각각의 시료에 대해 G.723.1 알고리즘과 제안한 알고리즘의 MP-MLQ 블록에 대한 처리 시간을 측정하였으며 각 시료에 대해 7번씩 수행한 다음 최대와 최소치를 뺀 나머지 5개에 대해 평균을 측정하였다. 측정 결과 평균 26.08%의 처리 시간 감소를 얻을 수 있었다. 측정한 결과는 표 3에 나타내었다. 그림 2는 시료 1에 대해 G.723.1을 통과시킨 파형과 제안한 알고리즘을 통과시킨 파형을 나타낸 그림이다.

V. 결론

인터넷 폰과 화상회의를 목적으로 ITU-T에서 표준화된 G.723.1은 일반적인 CELP 계열 보코더와 같이 음성 신호를 성분 분리하여 합성하는 방식인 합성에 의한 분석방법을 사용하기 때문에 많은 계산량으로 인한 처리 시간의 소모를 피할 수 없다는 문제점을 갖고 있다. 따라서 본 논문에서는 G.723.1 MP-MLQ 알고리즘의 Grid 비트를 DC 성분이 제거된 원 음성과 짝/홀수번째 펄스를 이용하여 합성한 합성음 사이의 에러를 비교하여 Grid 비트를 결정하는 다음 최적펄스를 검색하는 방법을 제안하였다. 실험결과 객관적 음질평가의 경우 약 11.19dB를 얻었으며 처리시간은 26.08%를 감소할 수 있었다. 주관적 음질 평가의 경우 음질 열하는 거의 없었다.

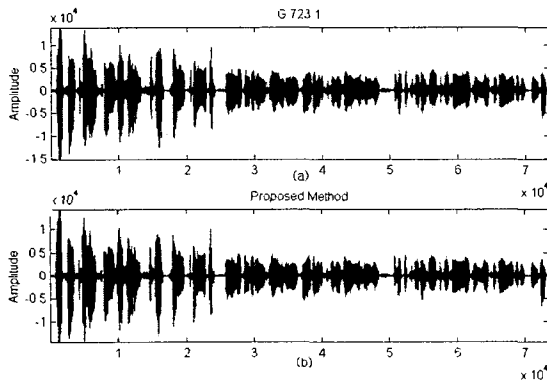


그림 2. G.723.1 알고리즘과 제안한 알고리즘을 통과한 음성 파형의 비교
(a) G.723.1 알고리즘 (b) 제안한 알고리즘

표 1. 객관적 음질 평가 결과 (segSNR : dB)

	시료 1	시료 2	시료 3	시료 4	시료 5	평균
segSNR	13.44	9.55	10.62	11.12	11.24	11.19

표 2. 주관적 음질 평가 결과 (MOS score)

	G.723.1	제안한 알고리즘
시료 1	3.89	3.88
시료 2	3.91	3.89
시료 3	3.92	3.92
시료 4	3.88	3.88
시료 5	3.87	3.85
평균	3.894	3.884

표 3. 처리시간 결과 비교 (sec)

	G.723.1	제안한 알고리즘	감소율(%)
시료 1	12.086	8.858	26.7
시료 2	5.908	4.878	17.4
시료 3	5.458	4.076	25.3
시료 4	6.416	4.720	26.4
시료 5	7.740	5.060	34.6
평균			26.08

참고 문헌

- [1] A. M. Kondoz, "Digital Speech - Coding for Low Bit Rate Communication System", John Willy & Sons, 1994.
- [2] ITU-T Recommendation G.723.1, March, 1996
- [3] N. S. Jayant and P. Noll, Digital Coding of Waveform-Principles and Applications to Speech and Video, pp.220-221, Prentice-Hall, 1978.
- [4] B. Atal. "Efficient coding of LPC parameter by temporal decomposition", Proc. of ICASSP, pp.81-85, 1983.
- [5] W. B. Kleijn et al., "Fast Methods for the CELP Speech Coding Algorithm", IEEE Trans., Acoustics, Speech and Signal Processing, Vol.38, No.8, pp.1330-1341, August 1990.
- [6] 홍성훈, 김정진, 박영호, 배명진, "다중레벨 진폭코드북을 이용한 음성 부호화에 관한 연구," 대한전자공학회, 대한전자공학회 추계종합학술대회 논문집, Vol.21, No.2, p.1219-1222, 1998년 11월.
- [7] D. KIM, M. BAE, J. KIM, K. BYUN, K. HAN, H.YOO, "On a Reduction of Pitch Searching Time by Preprocessing in the CELP Vocoder", J. Acoust. Soc, Korea, Vol.13, No.3, pp.33-40, June 1994.