1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 2(s) 호

CELP 음성 부호화기의 성능 향상 방법

박 호 종
광운대학교 전자공학과

Performance Improvement of CELP Speech Coder

Hochong Park
Dept. of Electronics Engineering
Kwangwoon University
e-mail : hcpark@daisy.kwangwoon.ac.kr

요약

본 논문에서는 CELP 음성 부호화기의 성능을 향상시키는 방법을 제안한다. 제안된 방법은 최적 코드북 검색 과정에서 추가적인 알고리듬의 지연 없이 미래 정보를 이용하고 두 인접한 코드북 부프레임 사이의 동시 최적화를 통하여 음성 부호화기의 성능을 향상시킨다. 또한, 제안된 코드북 검색 과정의 계산량을 조절하기 위한 방법도 제공된다. 제안된 방법의 성능을 검증하기 위하여 IS-96A QCELP 음성 부호화기를 이용하여 합성음의 스펙트럼과 Segmental SNR로 성능을 측정하는 모의실험을 실시하였으며, 제안된 방법을 적용한 QCELP 음성 부호화기가 기존의 QCELP에 비하여 향상된 성능을 보여주었다.

1. 서론

1985년 CELP 방식의 음성 부호화기가 처음 개발된 이래[1], 여러 전송률을 가지는 많은 표준 CELP 음성 부호화기가 개발되었고, 다양한 디지털 통신 분야에 성공적으로 적용되었다[2 - 8]. 그러나, 아직 8kbps 및 그 이하의 전송률에서는 CELP 음성 부호화기의 성능이 아직 원하는 수준에 이르지 못하고 있어 이러한 음성 부호화기를 사용하는 디지털 통신 시스템의 품질이 떨어지고 있다. 예로 IS-96A QCELP 음성 부호화기를 사용하는 국내의 CDMA 셀룰라 시스템의 통화 품질이 사용자의 만족을 얻지 못하고 있고[4, 9], 따라서 그 이후에 개발된 PCS에서는 통신 시스템의 음질을 향상시키기 위하여 고전송률의 13.3kbps QCELP 음성 부호화기를 채택하였다[5]. PCS의 고품질 음성 통신에 대응하기 위하여 CDMA 시스템은 새로 개발된 8.55kbps IS-127 EVRC 음성 부호화기를 사용하는 새로운 서비스를 시작하였다[8]. 그러나, CDMA에 사용되는 QCELP와 EVRC 사이에 호환성이 없고, 초기 QCELP만 채용한 단말기가 이미 널리 사용되고 있어 이 단말기를 사용하는 가입자들은 새로운 음성 부호화기의 혜택을 받지 못한다. 따라서 QCELP와 호환성을 가지면서 QCELP의 성능을 향상시키는 방법의 연구가 필요하다. 또한 성능 향상 방법은 기존의 QCELP에 필요한 계산량과 큰 차이가 없어 기존의 Hardware System에 Software 변경만을 통하여 향상된 성능을 구현할 수 있어야 한다.

CELP 음성 부호화기의 성능은 크게 다음의 3가지 항목에 의하여 결정된다 - 음성 모델, 모델 파라미터 검색, 파라미터 양자화. 음성 모델과 선택된 모델 파라미터의 양자화는 음성 부호화기의 표준에 의하여 정하여지는 것이며, 성능 향상의 목적으로 임의로 변경될 수 없다. 그러나 파라미터 검색은 임의의 방법을 사용하여 가장 우수한 음질을 합성할 수 있는 파라미터를 찾는 것이 필요하며, 따라서 새로운 검색 방법을 사용하여 호환성을 유지하면서 성능을 향상시킬 수 있다. 단, 이때 계산량이 너무 많지 않아야 하는 조건이 있다. 본 논문에서는 합성음의 음질을 향상시키는 보다 좋은 파라미터를 찾는 새로운 검색 방법을 제안한다. 이것 이외에 가중치 함수, 최적 조건 등의 다른 성능 향상 방법에 대하여서는 다루지 않는다.

CELP 음성 부호화기에서는, 합성음과 원음의 청각적 특성을 고려한 오차의 에너지를 최소로 하는 파라

미터를 Analysis-by-Synthesis 방법으로 찾는다. 이때 최적화 과정은 짧은 구간의 주어진 부프레임(Subframe) 안에서 실행되며, 이전의 모든 정보는 필터의 초기값의 형태로 주어진다. 그러나 현재의 부프레임 이후에 들어오는 어떠한 미래 정보도 최적화 과정에 이용되지 못한다. 이로 인하여 특히 천이 구간에서 최적의 파라미터 검색이 이루어지지 못하는 경향이 있다. 또한, 짧은 부프레임 단위로 최적화 과정이 실행되므로 부프레임 경계에서의 신호 특성이 최적화에 충분히 고려되지 못하는 경계 문제가 발생한다. 특히 CELP에서는 코드북의 검색 단위, 즉 코드북 부프레임의 길이가 가장 짧으므로 최적의 코드북을 찾지 못하게 된다.

이와 같은 부프레임 단위의 검색 과정의 문제점을 해결하기 위하여 새로운 코드북 검색 방법을 제안한다. 이 방법에서는 인접한 두 코드북 부프레임에 걸쳐 최적화가 이루어지고, 따라서 현 부프레임 이후의 정보를 이용할 수 있어 보다 우수한 코드북을 찾을 수 있다. 또한 인접 부프레임 경계를 제거하여 성능의 향상을 가져온다. 이 방법은 Delayed-Decision 방법에 기초를 두지만 추가적인 부호화기의 지연이 없이 성능을 향상시킬 수 있다. 두 부프레임을 동시에 고려하여 최적화를 하므로 계산량이 증가하며, 계산량 줄이는 방법과 계산량을 조절하는 방법도 제시한다.

2. 새로운 코드북 검색 방법

제안된 코드북 검색 방법은 부프레임의 구조에 따라 변경되며 본 논문에서는 한 개의 피치 부프레임이 동일한 길이의 두 개의 코드북 부프레임으로 나누어지는 구조를 가정한다. IS-96A QCELP는 이 구조를 가진다.

주어진 피치 부프레임에서 피치 주기와 피치 이득을 기존의 방법으로 구한다. 다음 현 피치 부프레임 안의 두 코드북 부프레임에 대하여 다음과 같은 3 단계 과정으로 코드북 검색을 한다 - (i) 첫 코드북 부프레임 검색, (ii) 둘째 코드북 부프레임 검색, (iii) 두 코드북 부프레임에서의 전체 최적화 과정을 통한 검색.

(1) 첫 코드북 부프레임 검색

코드북 검색 과정 전에 이미 구하여진 피치 필터와 포먼트 필터를 이용하여 먼저 코드북 검색을 위한 목표 신호 $x(n)$을 구한다. 이때 $x(n)$의 길이는 피치 부프레임의 길이 Lp 이다. 다음 첫 코드북 부프레임에 해당하는 위치에서 K 개의 최적 코드북을 Analysis-by-Synthesis 방법으로 찾는다. 일반적으로 코드북 검색을 할 때 양자화된 이득을 이용하지만, 제안된 방법에서는 양자화 되지 않은 코드북 이득을 이용하여 코드북을 검색한다. 코드북 이득의 양자화는 나중에 실행된다.

(2) 둘째 코드북 부프레임 검색

둘째 코드북 부프레임 위치에서 최적의 코드북을 찾는다. 그러나 이때 앞 코드북 부프레임의 영향이 전달되므로 이것을 고려하여 최적의 코드북을 찾아야 한다. 즉, 앞 부프레임의 영향과 둘째 부프레임의 코드북의 합으로 목표 신호를 표현하면 되고, 따라서 코드북 입장에서 본 실제 목표 신호가 변경된다.

둘째 부프레임에 대한 앞 부프레임의 영향은 다음과 같이 구한다. 앞 코드북 부프레임에서 구한 K 개의 최적 코드북을 $v_k(n)$, k = 1, ..., K, n = 0, ..., $Lp/2$-1 이라 하고, 이 신호의 n = $Lp/2$ 부터 n = Lp-1 까지 0으로 채워 길이 Lp인 신호 $v_k(n)$을 얻는다. 이를 피치 필터와 가중치가 고려된 포먼트 필터를 통과시켜 $y_k(n)$을 얻는다. 그러면, 첫 부프레임의 k 번째 코드북에 대한 새로운 목표 신호 $d_k(n)$는 $d_k(n) = x(n) - G_k\, y_k(n)$, n = $Lp/2$, ... , Lp-1 이 된다. 이때 G_k는 각 코드북 $v_k(n)$에 대한 양자화 되지 않은 코드북 이득이다.

다음, 각 k 에 대하여 목표신호 $d_k(n)$ 을 이용하여 둘째 부프레임에서 L 개의 최적 코드북을 찾고, 이를 $u_{kj}(n)$, j = 1, ..., L, n = $Lp/2$, ... , Lp-1 이라 한다. 물론 이 때도 양자화 되지 않은 코드북 이득을 이용한다. 이와 같이 하면 $K \times L$ 개의 코드북 검색 결과가 나온다. 다음, $u_{kj}(n)$을 피치 필터와 가중치가 고려된 포먼트 필터를 통과시켜 $z_{kj}(n)$, n = $Lp/2$, ..., Lp-1을 얻는다. 이때 n = 0, .., $Lp/2$-1에서 $z_{kj}(n)$ = 0 이다.

(3) 전체 최적화

두 코드북 부프레임에 걸쳐 복수로 선택된 많은 코드북 후보 중에서 최적의 코드북과 그것의 이득을 구한다. 이때 이득의 양자화 오차를 동시에 고려하여 성능의 향상을 꾀한다.

각 코드북 부프레임의 후보 $v_k(n)$ 과 $u_{kj}(n)$, $k \in \{1, K\}$, $j \in \{1, L\}$,을 이용하여 합성된 결과는 아래의 식이 된다. 여기서 Gq_a 와 Gq_b 는 양자화된 코드북 이득이고 a와 b는 양자화 인덱스이며, N bit 양자화를 하면 $a \in \{1, 2^N\}$, $b \in \{1, 2^N\}$ 가 된다.

$$Gq_a\, y_k(n) + Gq_b\, z_{kj}(n)$$

따라서 아래의 값을 최소로 하는 k, j, a, b 을 찾으면 두 부프레임에 걸친 최적 코드북과 이득을 구한 것이 된다.

$$\| x(n) - Gq_a\, y_k(n) - Gq_b\, z_{kj}(n) \|^2 \qquad n = 0, 1, \ldots, Lp-1$$

이 과정에서 하나 중요한 사실은 두 코드북 부프레임의 이득의 양자화가 동시에 실시된다는 것이다. 즉, 각 부프레임별의 양자화가 필요 없으므로 이득의 양자화를 벡터 양자화 방법을 사용할 수 있다. 물론 표준에서 정하는 양자화 방법을 그대로 사용하여야 기존의 부호화기와 호환성을 가지나, 만일 새로운 부호화기를 설계할 경우 코드북 이득을 벡터 양자화 하면 성능을 향상시킬 수 있다.

제안된 방법은 두 부프레임에 걸쳐 후보를 다수 정하고 이로부터 최적의 코드북을 선택하므로 기존의 방법보다 계산량이 많다. 이를 줄이기 위하여 다음과 같은 방법을 이용한다. 각 부프레임에서 코드북 검색을 할 때 이미 최적의 코드북 이득 G_k 와 G_{kj} 가 구하여져 있다. 이 정보를 이용하여, 이득의 양자화에서 이것과 동일한 부호(Sign)를 가지는 것만 검색한다. 즉, 위 식에서 모든 a 와 b 에 대하여 계산을 하여 크기를 비교하지 않고, Gq_a와 Gq_b 중에서 각각 G_k 와 G_{kj} 와 동일한 부호를 가지는 것만 검색을 한다. 이렇게 하면 검색 계산량이 1/4로 줄어든다.

제안된 검색 방법의 전체 계산량은 K 와 L 의 값에 따라 정하여진다. 이 값들은 매 피치 부프레임마다 변경 가능하고, 따라서 전체 시스템 자원 중에서 매 피치 부프레임마다 음성 부호화기에 할당된 계산량에 따라 K 와 L 값을 가변적으로 하여 실시간 동작을 가능하게 할 수 있다. 물론 이 값들이 크면 성능은 향상된다.

3. 모의 실험

제안된 코드북 검색 방법을 8kbps IS-96A QCELP에 적용하여 기존의 QCELP와의 성능을 비교하였다. QCELP는 가변 전송률을 가지고 따라서 부프레임 구조가 가변적이다. 그러나 음질은 주로 8kbps 전송률에서의 성능에 좌우되므로 모든 프레임을 8kbps로 동작시키는 고정 전송률로 하고 실험하였다. 이 경우 피치 부프레임의 길이는 40 샘플이고 코드북 부프레임의 길이는 20 샘플이다.

모의 실험에 사용한 음성 데이터는 6분 동안의 영어이다. 먼저 객관적인 음질을 Segmental SNR (SNRseg)과 스펙트럼 비교를 통하여 측정하였고, 여러 K와 L 에 대한 SNRseg 결과가 표 1에 정리되어 있다. 기존의 QCELP로 동작시킬 때의 SNRseg은 10.87dB 이며 K 와 L 값이 증가함에 따라 성능이 향상되는 것을 알 수 있다. 특히 각 한 개씩의 코드북 후보를 가지는 $K = L = 1$ 인 경우에도 성능이 향상되었는데, 이는 단지 코드북 이득을 두 부프레임에 걸쳐 동시에 최적화

(K, L)	(1,1)	(1,2)	(1,3)	(2,1)	(2,2)
SNRseg(dB)	10.90	10.92	10.94	10.97	11.00

(K, L)	(2,3)	(3,1)	(3,2)	(3,3)	
SNRseg(dB)	11.02	11.04	11.06	11.08	

표 1. 제안된 검색 방법을 사용한 QCELP의 SNRseg 값. (기존의 QCELP에 의한 SNRseg 는 10.87dB 이다.)

하였기 때문에 얻은 이득이다.

원음과 합성음의 스펙트럼 비교를 통한 음질 비교 결과가 그림 1에 정리되어 있다. 여기서는 여성 유성음 구간에 대하여 $K = L = 3$ 인 경우를 이용하였다. (a)는 기존의 QCELP에 의한 결과로서 $\omega = 0.22$ 근처에서 피치 하모닉 구조가 불규칙하고, $\omega = 0.2$ 근처와 $\omega = 0.4$ 이상에서 Envelope의 차이가 매우 심하다. (b)는 제안된 검색 방법을 사용한 QCELP의 결과를 보여준다. (a)에 비하여 피치 하모닉의 구조가 매우 규칙적이고, 전체적으로 Envelope이 원음과 비슷한 모양을 가진다.

주관적인 음질을 간단한 청취 테스트를 통하여 측정하였다. 대부분의 경우 제안된 방법을 사용한 QCELP의 음질이 기존의 QCELP 보다 우수한 결과를 얻었다.

4. 결론

본 논문에서는 CELP 음성 부호화기의 성능을 향상시키기 위한 새로운 코드북 검색 방법을 제안하였다. 이 방법은 두 코드북 부프레임을 동시에 최적화 하여 부프레임 경계에서의 문제점을 해결하고 미래의 정보를 미리 이용하여 코드북을 검색하므로 성능이 향상된다. 제안된 방법은 코드북을 검색하는 방법만을 개선하였으므로 기존의 방법과 호환성을 가진다. 따라서 간단한 Software의 변경을 통하여 음성 부호화기의 성능을 향상시킬 수 있고, 나아가 이를 사용하는 통신 시스템의 품질을 개선 할 수 있다. 모의 실험을 통하여 객관적, 주관적 음질을 측정한 결과 제안된 검색 방법이 QCELP 음성 부호화기의 성능을 향상시키는 것을 확인

하였다.

5. 참고 문헌

[1] M. R. Schroeder and B. S. Atal, "Code-excited linear prediction: High quality speech at very low rates," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.937-940, 1985.

[2] I. Gerson and M. Jansiuk, "Vector sum excited linear prediction(VSELP) speech coding at 8kb/s," *Proc. IEEE International Conference on Acoustics, Speech and Signal Processing*, pp.461-464, 1990.

[3] ITU Recommendation G.728, *Coding of speech at 16kbit/s using Low-Delay Code Excited Linear Prediction*, 1992.

[4] TIA/EIA/IS-96A, *Speech Service Option Standard for Wideband Spectrum Digital Cellular* System, 1994.

[5] CDG-27, *High Rate Speech Service Option for Wideband Spread Spectrum Communication Systems*, 1995.

[6] GSM 06.60, *Enhanced Full Rate (EFR) Speech Transcoding*, 1996.

[7] ITU Recommendation G.729, *Coding of speech at 8kbit/s using Conjugate-Structure Algebraic -Code-Excited Linear Predictive(CS-ACELP)*, 1996.

[8] TIA/EIA/IS-127, *Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems*, 1997.

[9] TIA/EIA/IS-95, *Mobile Station-Base Station Compatibility Standard for Dual-Mode Wideband Spread Spectrum Cellular System*, July 1993.

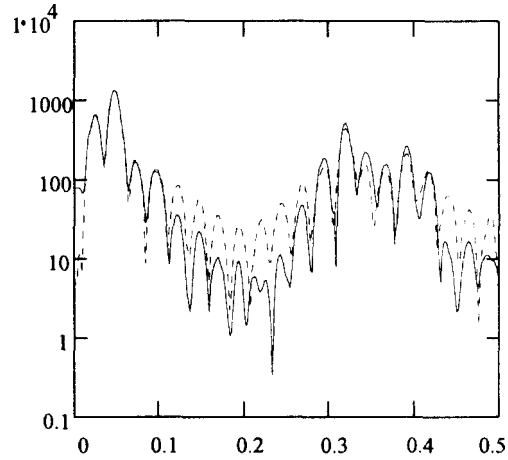

그림 1(a)

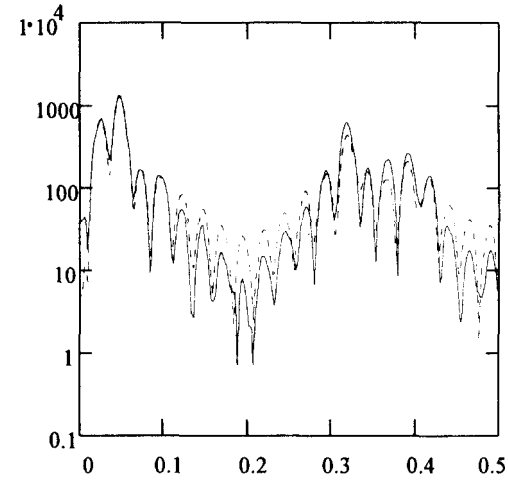

그림 1(b)

그림 1. 여성 모음을 이용한 원음(점선)과 합성음(실선)의 스펙트럼 비교. (a) 기존의 QCELP를 사용한 경우의 스펙트럼 비교. (b) 제안된 코드북 검색 방법을 이용한 QCELP를 사용한 경우의 스펙트럼 비교.(K = L = 3 을 사용하였음)