論文 96-21-6-8

프레임 분류와 합성필터의 변형을 이용한 적은 지연을 갖는 음성 부호화기의 성능 개선

正會員 林 銀 熙*, 李 周 鎬**, 金 炯 明***

Improving LD-CELP using Frame Classification and Modified Synthesis Filter

Eun Hee Rhim*, Ju Ho Lee**, Hyung Myung Kim*** *Regular Members*

요 약

중간 주파수 대역(8kbps) 이하에서 적은 지연을 갖는 벡터여기 선형예측 음성 부호화기(LD-CELP)에 대하여 고려한다. 합성필터를 입력 프레임의 종류에 따라 변화시켜 음성 부호화기의 성능을 향상시키고자 한다. 먼저 프레임을 유성음과 무성음 그리고 개시 프레임으로 분류한다. 유성음과 무성음 프레임에서는 합성필터의 스펙트럼 포락을 음운의 특성에 적합하도록 변화시킨다. 개시 프레임에서는 합성필터의 성격을 바꾸어주기 위하여 바이어스 필터를 이용한다. 제안된 부호화기는 다른 적은 지연을 갖는 벡터여기 선형예측 음성 부호화기들에 비하여 비슷한 지연 시간을 갖으면서 더 나은 음질을 제공하였다.

ABSTRACT

A low delay code excited linear predictive speech coder(LD-CELP) at bit rates under 8kbps is considered. We try to improve the performance of speech coder with frame type dependent modification of synthesis filter. We first classify frames into 3 groups: voiced, unvoiced and onset. For voiced and unvoiced frame, the spectral envelope of the synthesis filter is adapted to the phonetic characteristics. For transition frame from unvoiced to voiced, the synthesis filter which has been interpolated with the bias filter is used. The proposed vocoder produced more clear sound with similar delay level than other pre-existing LD-CELP vocoders.

I. 서 론

음성 부호화란 통신 채널을 통하여 음성을 전달하는 방식이다. 이는 다양한 방법으로 이루어 질 수 있고 그만큼 다양한 음성 부호화기가 존재한다. 음성 부호화기의 성능은 합성된 신호의 음질, 지연(delay) 시간, 계산상의 복잡도, 전송율 등으로 가늠될 수 있다. 델타 변조(Delta Modulation)나 펄스 부호 변조(Pulse Coded

*한국통신 연구개발본부 교환기술연구소 연구원
**한국과학기술원 전기 및 전자공학과 박사과정
***한국과학기술원 전기 및 전자공학과 부교수
論文番號:96027-0125
接受日字:1996年 1月 25日

Modulation)처럼 샘플 단위의 부호화 과정을 거치는 부호화 방법은 toll quality의 음질을 제공하지만 전송율이 높아지는 단점을 갖는다. 이와는 달리 낮은 전송율을 얻기 위해 분석 및 합성(Analysis-by-Synthesis) 방법을 사용하면 통신을 하기에 지장이 없을 정도의 음질을 얻을 수 있다. 실제로 이 방법을 사용하는 음성 부호화기가 4kbps 대에서까지 제안되었다[8][12]. 그러나 전송율만으로 부호화기의 성능을 대표하는 데에는 무리가 있다. 분석 및 합성 방법을 사용할 경우에는 프레임의 길이가 길어지게 되고 따라서 지연 시간과 복잡도가 증가하게 되며 이는 부호화기를 실제 환경에서 운영하거나 구현하는데 있어서 바람직하지 못한 제한을 줄 수 있다. 여기서 지연은 부호화기와 복호기가 서로 연결되어 있다고 가정할 때 하나의 음성 신호가 부호화기의 입력으로 들어가서 복호기의 출력으로 나올 때까지 걸리는 시간을 의미하는 것으로 다른 장비나 통신상의 거리와 관계없이 오직 부호화 방법에만 의존한다[1]. 유선 전화망의 경우 지연이 길어지면 반향 제거기(echo canceler)가 존재하더라도 화자에게 반향이 들리게 된다. 더구나 반향 제거기의 복잡도는 지연 시간에 비례하며, 이러한 현상은 반향 제거기의 차수가 정해진 무선망의 다중화기(multiplexer)에서도 중요한 문제점이 된다. 또 셀룰라 시스템에서는 음성 부호화와 채널 부호화에 소요되는 지연의 최대 허용치를 정한 바 있다[3][4]. 그러므로 기존의 부호화기에 비해 성능이 떨어지지 않으면서 지연을 줄여주는 부호화기가 필요하다.

LD-CELP 방식의 음성 부호화기는 음질과 지연 시간 면에서의 성능 향상을 주된 목적으로 하여 제안된 것이다. Chen이나 Kabal은 16kbps에서 프레임의 크기를 매우 작게 구성하여 지연이 1.2ms 이내가 되면서 훌륭한 음질을 얻을 수 있다는 것을 보여주었다[1][2]. 중간 전송율에서도 다양한 LD-CELP 방식의 부호화기들이 제안되었다. 그들은 주로 짧은 프레임을 사용한 후 여기 신호를 잘 묘사할 수 있는 방법들을 제시한 것이라고 볼 수 있다[3-6]. 그러나 중간 전송율에서의 부호화기들은 통신 레벨의 음질과 적은 지연 수준을 제공하지만 성능의 근본적인 한계가 있으며 더구나 낮은 전송율에서는 기존의 CELP 방식만큼의 성능을 얻지 못한다. 이는 16kbps에서 8kbps로 전송율을 내리게 되면 프레임의 길이가 길어지고 여기신호가 정확하게 묘사되지 못하므로 음성 부호화기의 성능이 저하된 것으로 보인다. 본 논문에서는 관점을 달리하여 여기신호의 한계를 합성필터의 변형으로 극복하여 부호화기의 성능을 향상시키려는 시도가 제안되었다.

우리는 부호화기의 합성필터를 프레임의 음운학적 특성을 고려하여 변형시킨다. 그렇게 하기 위해 먼저 프레임을 유성음, 무성음, 개시의 3가지 종류로 분류한다. 여기신호는 프레임의 종류에 따라 다중 탭을 사용하는 가변 코드북이나 통계학적 코드북을 이용하여 근사한다. 본론에서는 제안된 부호화기의 부호화 과정을 살펴보고 부호화기의 성능에 대하여 토론한다. 그리고 뒤이어 간략한 결론을 내린다.

Ⅱ. 본 론

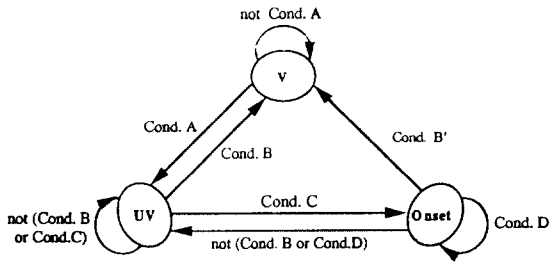
CELP 방식처럼 LD-CELP에서도 각 프레임 당 고정된 비트들을 사용하여 입력된 음성 신호를 부호화한다. 이러한 부호화 방법은 입력 신호의 음운학적 특성을 무시하게 되어 청각적으로 중요한 부분에 비효율적인 비트들을 할당하게 된다. 잘 알려진 대로 입력 프레임의 성격이 유성음일 때에는 여기신호는 가변 코드 벡터(Long Term Predictor)와 주로 상관성을 갖는다. 반면에 무성음 성격의 프레임에 해당하는 여기신호는 통계학적 코드 벡터(Short Term Predictor)만으로도 묘사 가능하다. 물론 분명하고 자연스러운 합성음을 얻기 위해서는 하나의 프레임 당 2개의 코드북이 사용되어야 하지만 사용할 비트들이 제한된 경우에는 2개의 코드북 중 하나만을 선택하는 방법이 사용 가능한 비트들을 효율적으로 할당하는 방법이 된다.

Chen은 3 탭의 LTP를 무성음 구간 또는 문장이나 음절 사이의 침묵(silence) 구간에 사용하지 않았지만, 이렇게 하는 것이 합성음의 음질에 영향을 끼치지 않는다[4]. Soheili 등은 입력 프레임을 유성음과 무성음으로 구분하고 각각 하나의 코드북을 할당하였다. 그러나 앞서 말했듯이 간단하게 하나씩의 코드북을 할당하는 것으로는 좋은 음질의 부호화기를 얻는 데에는 부족한 면이 있었고, 이는 Soheili가 지연된 후처리 필터를 사용하여 음질을 높이려고 시도한 점에서도 확인할 수 있다. 그러한 한계는 LD-CELP 방식의 고

유한 주파수 특성 때문이라고도 할 수 있다. 합성된 신호들의 파워 스펙트럼은 입력 신호들의 파워 스펙트럼에 비해 상대적으로 강한 벨리 부분을 가지고 있는 경향이 있다. 즉 유성음의 포먼트(formant)들이 벨리 부분에 가려 부각되지 못한다. 또, 전반적으로 높은 주파수 성분들이 줄어드는 현상을 볼 수 있다. Kataoka 등은 스펙트럼상의 문제를 해결하기 위해 학습시킨 통계학적 코드북을 사용하였고, 합성필터의 파라미터들을 구할 때 부분적으로나마 합성된 현재 프레임의 출력을 고려하였다[3]. 그 결과 비록 복잡도가 많이 증가하기는 하였으나 출력 신호의 스펙트럼에 직접 영향을 미치는 합성필터가 프레임의 성격에 적합하도록 조절되어야 함을 내포하고 있다. LD-CELP 방식의 부호화기들의 합성필터는 역방향 예측(backward prediction) 방식을 사용하여 구한다. 즉, 현재 프레임에 적용할 합성필터의 파라미터들을 현재 이전의 프레임까지만 고려된 과거 합성신호 집합에서 구하기 때문에 본질적으로 정확한 합성필터를 사용하는 것이 아니다. 본 논문에서는 합성필터의 스펙트럼의 벨리 성분을 감소시키기 위해 포먼트 즉 피크 부분을 강조하였다. 유성음 성격의 프레임에서는 음성 신호의 낮은 주파수대에 있는 포먼트들이 음색을 결정하므로, 합성필터의 스펙트럼 상 낮은 주파수 영역에 높은 주파수 영역보다 큰 가중치를 주었다. 또한 무성음 성격의 프레임에서는 음성 신호가 주로 높은 주파수 성분들로 나타내어지므로, 합성필터의 스펙트럼 상 높은 주파수 영역에 가중치를 가하여 높은 주파수 성분을 보강하였다. 프레임의 성격이 무성음에서 유성음으로 변하는 부분은 부호화기에서 잘 표현할 수 없는 부분이다. 유성음으로 변할 때 새롭게 등장하는 포먼트들을 부호화기에서 찾아내는 데 수 프레임의 시간이 소요되므로 천이 상태의 신호가 수 프레임 동안 잘못 묘사되기 쉽다. 제안된 부호화기에서는 유성음의 형태를 갖는 바이어스 필터를 도입하였다. 그리고,역방향으로 예측된 합성필터와 바이어스 필터를 보간하여 새로운 피치를 짧은 시간 내에 따라가도록 하였다. 다음의 각 부분에서는 입력 프레임을 분류하는 방법과 각 프레임에서의 음성 합성 방법에 대하여 살펴본다.

1. 프레임 종류의 분류

프레임의 종류는 입력 프레임의 제로점 교차 수와 정규화된 최대 자기 상관값, 이전 프레임과 현재 프레임의 전력 비를 바탕으로 분류된다. 프레임의 종류는 유성음, 무성음 그리고 무성음에서 유성음으로 변해 가는 개시로 구성된다. 유성음에서 무성음으로 변해 가는 프레임은 특별한 구분 없이도 부호화기에서 양호하게 묘사되므로 다른 종류로 구분하지 않는다. 각 종류를 하나의 스테이트로 구성한 후, 현재의 프레임의 종류가 이전 프레임의 종류에 의하여 결정되도록 스테이트 간의 천이를 형성한다. 신호의 특성상 유성음일 경우 높은 자기 상관값을 가지면서 제로 교차점의 수가 적고, 무성음일 경우에는 자기 상관값이 낮아지면서 제로 교차점의 수가 크므로 표본 조사를 통하여 두 종류의 임계값들을 정하였다. 개시에 해당되는 프레임은 이전 프레임에 비해 현재 프레임의 전력이 크게 증가하면서 유성음에 가까운 특성이 나타나는지를 확인한 후에 결정하였다. 그리고 유성음에서 개시로의 변화는 개시 프레임의 설정 취지에 어긋나므로 고려하지 않는다. 기존의 유성음, 무성음으로의 분류 방법보다 더 세분화된 제안된 분류 방법은 프레임의 현재 상태를 결정하는데 있어서 이전 프레임의 상태를 고려하여 오판할 가능성을 줄일 수 있다. 그림 1은 제안된 부호화기에서 사용된 프레임 분



Cond.A : zr > THRESH_zr and co < THRESH_co

Cond.B : (zr < THRESH_zr and co > THRESH_co) and pr < THRESH-pr

Cond.B' : (zr < THRESH_zr and co > THRESH_co) and pr < THRESH_pr*2/3

Cond.C : (zr < THRESH_zr*3/2 or co > THRESH_co*2/3) and pr > THRESH_pr

Cond.D : (zr < THRESH_zr*3/2 or co > THRESH_co*2/3) and pr > THRESH_pr*2/3

zr, co, pr : 실험값 (제로 교차점수, 최대 자기 상관값, 전력비)

THRESH_* : 각 실험값에 대한 기준치

그림 1. 상태 천이 다이어그램
Fig. 1 State transition diagram

류 방법을 나타낸다. 여기서 Cond.A~Cond.D는 실험적으로 결정된 각각의 천이 조건을 나타낸다.

2. 유성음 프레임

유성음으로 분류된 프레임은 가변 코드 벡터로 표현되는 여기신호를 합성필터에 통과시킨 신호로 근사한다. LTP의 차수를 증가시켜 통계학적 코드 벡터의 역할을 흡수하고자 하되 전송율의 제약이 있으므로 3차의 LTP를 이용한다. CELP 방식을 사용하는 부호화기에서 n번째 프레임에서 사용하는 역 LPC 필터는 일반적으로 10차 이상의 올폴(all-pole) 구조를 이용하며, 이때 각 계수 값은, 입력된 수 프레임들로 구성된 집합 $\{s_{w,n-m-1}: s_{w,n-m+1}: \cdots : s_{w,n}\}$으로부터 구할 수 있다. 여기서 입력 신호 집합에 윈도우를 가한 신호는 s_k에 아래 첨자 w를 더하여 나타내었다. LD-CELP 방식을 사용하는 부호화기에서는 현재의 프레임을 제외한 이전 프레임 신호로 구성된 집합을 이용한다. 그리고 부호화기와 복호기에서 같은 역 LPC 필터(또는 합성 필터)를 구해야 하므로 합성된 신호 $\hat{s}_k$ 등으로 구성된 집합 $\{s_{w,n-m}, s_{w,n-m+1}, \cdots, s_{w,n-1}\}$을 이용하며, 이로부터 얻어진 합성필터는 식 (1)처럼 나타내어진다.

$$H_s(z) = \frac{1}{1 - \sum_{i=1}^{10} a_i z^{-i}} \tag{1}$$

여기서 $\{a_i\}_{i=1, \ldots, 10}$는 합성필터 또는 역 LPC 필터의 계수이다.

제안된 부호화기에서 사용될 합성필터, $\tilde{H}_s(z)$는 식 (2)와 같이 나타내어진다.

$$\tilde{H}_s(z) = \frac{1}{1 - \sum_{i=1}^{10} a_i^v z^{-i}} \tag{2}$$

여기서 $\{a_i^v\}_{i=1, \ldots, 10}$는 윈도우가 가해진 합성신호 집합의 자기 상관 수열과 이 수열의 포먼트 성분에 가중치를 가한 수열의 합으로부터 구해진 필터계수이다. 과거 m개의 합성 신호 프레임들로 이루어진 집합의 자기 상관 수열을 $R_n(i)$라 하자. LPC 필터의 차수가 10이므로 자기 상관값의 시간 차(time lag)는 10까지만 고려한다. 여기에 포먼트를 강조하는 가중치 수열 $w_v(i)$을 가하여 얻어진 새로운 수열은 식 (3)과 같다.

$$R_{w,n}(i) = w_v(i) \cdot R_n(i) \qquad i = 0, \ldots, 10 \tag{3}$$

여기서 $w_v(i)$는 시간 차가 적은 부분이 시간 차가 큰 부분보다 큰 값을 갖는 가우시안 모양의 가중치 수열로 그림 2에 나타나있다.

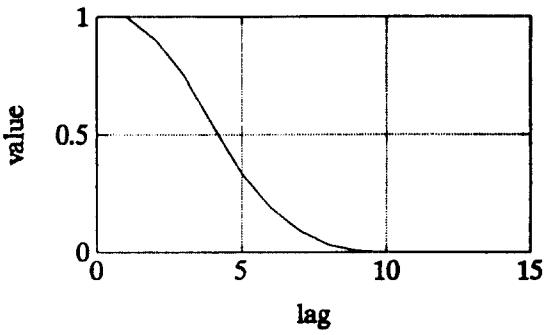


그림 2. 유성음 프레임에서의 가중치 수열
Fig. 2. Weighting sequence for voiced frame

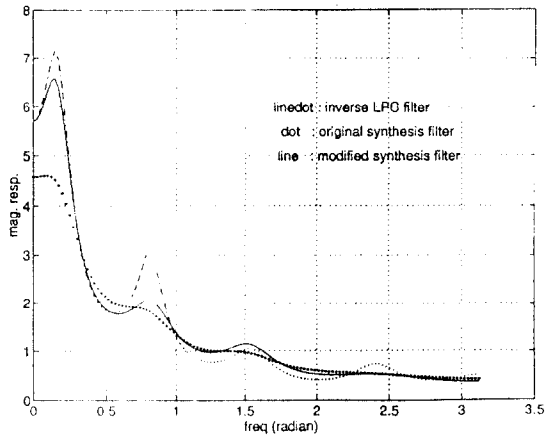


그림 3. 유성음 프레임에서 합성필터의 주파수 특성(크기) 비교
Fig. 3. Mag. resp. comparison of synthesis filters for voiced frame

$$\tilde{R}_n(i)=\frac{R_n(i)+R_{w,n}(i)\cdot\delta_1}{1+\delta_1} \qquad 0\le\delta_1<1 \qquad (4)$$

식 (4)에 의해 얻어진 새로운 자기 상관 수열 $\tilde{R}_n(i)$에 Levinson-Durbin 알고리즘을 적용하여 $\{a_i^v\}_{i=1,\ldots,10}$를 구한다. 그림 3은 포먼트가 강조된 합성필터, 원래의 합성필터, 그리고 입력 신호 집합으로부터 얻어진 역 LPC 필터의 주파수 특성을 비교한 것이다. 포먼트에 비해 스펙트럼의 벨리 성분들이 상대적으로 감쇄되었음을 확인할 수 있다.

새로운 합성필터의 형태가 확정된 뒤 적합한 피치와 LTP의 3-탭(tap) 이득(gain)들은 기존의 방법과 마찬가지로 사람귀의 Masking 효과가 고려된 MSE를 최소화시키는 계산 과정을 통해 얻어진다. 특히 피치는 사용 가능한 비트의 한계를 고려해 델타 탐색(delta search) 방법으로 구하였다.

3. 무성음 프레임

무성음으로 구분된 프레임은 통계학적 코드 벡터를 합성필터에 통과시켜 얻어지는 신호로 근사한다. 합성필터는 유성음의 경우와 비슷한 과정으로 높은 주파수 성분이 보강된 스펙트럼 포락을 갖도록 조정된다. 변형된 합성필터, $\tilde{H}_s(z)$는 식 (5)와 같이 나타내어진다.

$$\tilde{H}_s(z)=\frac{1}{1-\sum_{i=1}^{10} a_i^{uv} z^{-i}} \qquad (5)$$

여기서 $\{a_i^{uv}\}_{i=1,\ldots,10}$는 다음과 같은 과정으로 구해지는 $\tilde{R}_n(i)$에 Levinson-Durbin 알고리즘을 적용하여 얻는다.

$$R_{w,n}(i)=w_{uv}(i)\cdot R_n(i) \qquad i=0,\ldots,10 \qquad (6)$$

$$\tilde{R}_n(i)=\frac{R_n(i)+R_{w,n}(i)\cdot\delta_2}{1+\delta_2} \qquad 0\le\delta_2<1 \qquad (7)$$

이 때 $w_{uv}(i)$는 $w_v(i)$의 역순으로 시간 차가 큰 부분에 높은 가중치가 적용되도록 만들어진 수열이다. 그림 4은 변형된 합성필터의 스펙트럼 포락을 원래의 합성필터와 입력 신호 집합으로부터 얻어진 역 LPC 필터와 비교한 것으로, 높은 주파수 성분이 보강된 것을 확인할 수 있다.

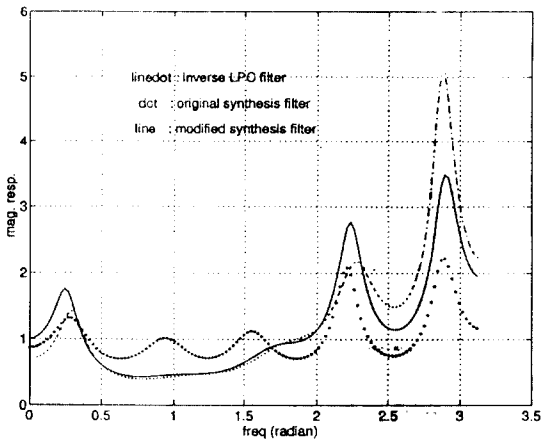


그림 4. 무성음 프레임에서 합성필터의 주파수 특성(크기) 비교

Fig. 4. Mag. resp. comparison of synthesis filters for unvoiced frame

통계학적 코드 벡터와 해당되는 이득은 기존의 방법과 같은 방법으로 구해진다.

4. 개시 프레임

기본적으로 개시 프레임의 여기신호 모델은 무성음 프레임에서와 같으나 합성필터의 형태는 다음과 같다.

$$\tilde{H}_s(z)=\frac{1}{1-\sum_{i=1}^{10} c_i z^{-i}} \qquad (8)$$

의 계수 $\{c_i\}$는 예측된 계수 $\{a_i\}$와 스펙트럼의 포락이 유성음의 특징을 띄는 바이어스 필터의 계수 $\{b_i\}$를 식 (9)처럼 보간하여 얻어진 것이다.

$$c_i=a_i\cdot(1-\delta_3)+b_i\cdot\delta_3 \qquad (0<\delta_3<1,\ i=1,\ldots,10) \qquad (9)$$

바이어스 필터를 통해 새로 도입된 포먼트들은 실제 입력 프레임의 포먼트의 값들과 다를 경우가 많다. 그러나 정확한 피치는 이어지는 짧은 프레임 내에서 구해진다. 다만 정확한 피치를 찾기 위해서는 합성필터의 스펙트럼 포락이 유성음의 특성을 갖는다는 선결 조건이 만족되어야하므로, 개시 프레임에서는 이

조건을 만족시키도록 하여 입력 신호의 특성이 급변하더라도 이에 잘 대응하도록 한 것이다.

이 밖에도 제안된 부호화기에서는 합성 신호의 청각적인 선명도를 높이기 위해 후처리 필터가 사용되었다. 식 (10)에 나타난 이 필터는 Chen이 제안한 것을 바탕으로 tuning factor들을 재조정한 것이다[1]. 사용한 factor 값은 $\zeta_1=0.65$, $\zeta_2=0.75$, $\zeta_3=0.15$이다.

$$H_p(z)=\frac{1-\sum a_i \zeta_1^i z^{-i}}{1-\sum a_i \zeta_2^i z^{-i}} \cdot (1+\zeta_3 z^{-1}) \tag{10}$$

Ⅲ. 실험 및 결과 고찰

1. 실험 조건

제안된 부호화기의 성능을 8kbps에서 2가지의 다른 LD-CELP 방식들과 비교하도록 한다. Chen등이 제안한 LD-CELP 부호화기를 CHEN이라 표기하고, Soheili 등이 제안한 것을 SUR라고 나타낸다[4][5]. 제안된 방법은 MSLD로 나타내기로 한다. 8kbps에서 CELP 방식의 한가지인 VSELP와 제안된 부호화기의 성능을 비교하여 적은 지연을 갖는 부호화기의 약점인 역방향 예측된 합성필터의 성능이 순방향 예측된 VSELP와 비슷한지를 비교한다. 8kbps에서는 객관적인 지표로 SPSEG(Signal to nonPitch predicted portion Ratio)를 측정한다[8]. n-번째 프레임에서 입력신호를 $\vec{x}_n$이라 하고, 필터의 제로 입력 응답을 $\vec{y}_o$라고 하자. 그리고 adaptive codebook의 여기신호를 $\vec{ex}_a$, stochastic codebook의 여기신호를 $\vec{ex}_s$라고 하고, 각 여기신호로부터 얻은 합성신호를 각각 $\vec{y}_a$, $\vec{y}_s$라고 한다면 SP는 다음과 같다.

$$SP=\frac{\|\vec{x}_n\|}{\|\vec{x}_n-(\vec{y}_o+\vec{y}_a)\|} \tag{11}$$

제안된 부호화기의 특성상 SPSEG는 유성음 프레임에서 SNRSEG와 같은 값을 갖게 된다.

4.8kbps에서 제안된 부호화기는 순방향 예측 방식을 사용하는 DoD-CELP와 주관적 평가를 통하여 비교된다. 4.8kbps 대에서 제안된 다른 LD-CELP 형식의 부호화기가 존재하지 않으므로 부득이 지연 시간이 훨씬 긴 CELP 형식을 고르게 되었다. 주관적인 평가에서 총 40여초로 구성된 테스트 문장들을 제안된 방법과 비교하는 방법으로 합성한 후 임의의 순서대로 2번씩 들려주었다. 청취자는 7명이었다.

표 1은 각 LD-CELP 부호화기들의 비트 할당을 보인 것이다. 피치를 구할 때 델타 탐색을 한 것은(D)로 역방향 예측을 한 것은(B)로 나타내었다. 또 3차 LTP를 사용한 것은 (3)으로 나타내었다. 제안된 부호화기에서 사용한 Stochastic codebook은 random sequence를 학습시킨 것을 사용하였다. 그리고 $\{\beta_i\}_{i=-1,0,1}$는 LBG 알고리즘에 따라 7 비트로 벡터 양자화하였다. 그밖에 이득 g는 비균등 스칼라 양자화한 것을 사용하였다.

표 1. 비트 할당
Table 1. Bit allocation

	CHEN	SUR		MSLD(8k/4.8k)	
frame size (sample)	20	11		14/23	
flag (bit)	1	1		2	
pitch (″)	4(D)	3(B)		5(D)	
beta (″)	5	7(3)		7(3)	
code index (″)	7		7		8
gain (″)	3		3		4
total (″)	20	11		14	

2. 결과 및 고찰

8kbps에서 사용하는 측정 데이터는 각각 30초 정도의 남자와 여자가 말하는 한국어와 영어로 구성된 문장들이다. MSLD는 CHEN, SUR, VSELP보다 높은 SNRSEG를 얻었다. 이는 더 짧은 프레임을 사용하는 LD-CELP 방식이 입력 파형을 용이하게 따라가기 때문인 것으로 보인다. 또 SPSEG는 표 2에서처럼 비슷하게 프레임 분류 방법을 사용한 SUR보다 MSLD가 높은 값을 얻었다. 이는 LTP의 피치를 구하는 방법에 있어서 순방향 예측(forward prediction)을 하는 것이 역방향 예측을 하는 것보다 더 효과적임을 확인시켜 준 것이다.

합성필터에 어떠한 변형도 하지 않았을 때보다 제안된 방법대로 변형을 할 경우에 SNRSEG와 SPSEG

표 2. SPSEG 비교
Table 2. SPSEG comparison (dB)

	CHEN	SUR	MSLD
male	3.87	5.89	8.72
female	5.54	6.62	9.43

가 증가한 것을 확인할 수 있었다.

4.8kbps에서의 선호도 조사 결과가 표 3에 있다. 사용한 데이터는 8kbps에서와 비슷한 방법으로 구성된 문장들로 남, 여 각각 20초 정도의 문장들이다.

표 3. 선호도 비교
Table 3. Preference ratio comparison (%)

	MSLD	DoD-CELP
male	67.7	32.3
female	54.2	45.8

제안된 부호화기의 지연이 6ms(4.8kbps에서 10ms)이므로 기존의 CELP 방식의 부호화기의 지연인 50ms~60ms에 비하면 약 85% 정도의 지연 시간 감축이 이루어졌다.

Ⅳ. 결 론

중간 및 낮은 전송율을 갖는 LD-CELP 방식의 음성 부호화기에 대하여 고려하였다. 사용 가능한 비트들을 효과적으로 쓰기 위해 입력 프레임을 종류별로 분류한 후 각각 하나의 코드북을 할당하였다. 여기신호의 비정확성을 합성필터를 개선시키므로써 보상하려 시도했다. 객관적인 결과들로 8kbps에서 향상된 부호화기의 성능을 확인할 수 있었다. 또, 4.8kbps에서도 주관적인 결과로 향상된 성능을 확인할 수 있었다. 제안된 부호화기는 합성 필터를 순방향으로 적응시키는 기존의 CELP 방식보다 85% 정도의 지연 시간 감축을 달성하면서도 비슷하거나 우수한 성능을 얻었다. 그러나 낮은 전송율로 갈수록 합성 신호의 톤(tone)이 약간씩 내려가는 경향이 관찰되었다. 이를 극복할 수 있는 후 처리 기술을 제안하는 것이 다음에 풀어야 할 문제로 남아있다.

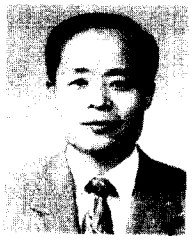
참 고 문 헌

1. J. H. Chen, R. V. Cox, Y. C. Lin, N. Jayant, and M. J. Melchner, "A Low Delay CELP Coder for the CCITT 16kbps Speech Coding Standard," in IEEE J. Select. Areas In Commun., vol. 10, no. 5, pp. 830-849, June 1992.
2. V. Iyenger, P. Kabal, "A Low Delay 16kb/s Speech Coder," in IEEE Trans. Signal Proc., vol. 39, no. 5, pp. 1049-1057, May 1991.
3. A. Kataoka, T. Moriya, "A Backward Adaptive 8kbit/s Speech Coder using Conditional Pitch Prediction," in Proc. IEEE Global Commun. Conf., pp. 1889-1893, 1991.
4. J. H. Chen, M. S. Rauchwerk, "An 8kB/s Low Delay CELP Speech Coder," in Proc. IEEE Global Commun. Conf., pp. 1894-1898, 1991.
5. R. Soheili, A. M. Kondoz, and B. G. Evans, "An 8 kB/s LD-CELP with Improved Excitation and Perceptual Modelling," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc., pp. 616-619, 1993.
6. J. H. Yao, J. J. Shynk, and A. Gersho, "A Low-Delay Vector Excitation Coding of Speech at 8kbit/s," in Proc. IEEE Global Commun. Conf., pp. 695-699, 1991.
7. J. H. Chen, A. Gersho, "Gain Adaptive Vector Quantization with Application to Speech Coding," in IEEE Trans. Commun., vol. COM-35, no. 9, pp. 918-930, Sept. 1987.
8. Y. Shoham, "Constrained-Stochastic Excitation Coding of Speech at 4.8Kb/s," in Advances in Speech Coding, ed. B. S. Atal, V. Cuperman, and A. Gersho, pp. 339-348, Kluwer Academic Publishers, Boston, MA, 1991.
9. J. H. Yao, J. J. Shynk, and A. Gersho, "Low-Delay Speech Coding with Adaptive Interframe Pitch Tracking," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc., pp. 410-414, 1993.
10. W. B. Kleijn, R. P. Ramachandran, and R. Kroon, "Generalized Analysis-by-Synthesis Coding and its Adaptation to Pitch Prediction," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc., pp. 337-340, 1992.
11. M. Foodeei, P. Kabal, "Backward Adaptive Prediction: High-Order Prediction and Formant-Pitch Configurations," in Proc. IEEE Int. Conf. Acoust., Speech, Signal Proc., pp. 2405-2408, 1991.
12. C. H. Kwon, "On Improving the Excitation Sig-

nal in Low-Rate CELP Coding," in Ph. D. Dissertation, KAIST, June 1994.
13. P. E. Papamichalis, Practical Approaches to speech coding, Englewood Cliffs, NJ : Prentice-Hall, Inc., pp. 137-142, 1987.

林 銀 熙(Eun Hee Rhim) 정회원
1971년 9월 10일생
1994년 2월 : 한국과학기술원 전기 및 전자공학과 학사
1996년 2월 : 한국과학기술원 전기 및 전자공학과 석사
1996년 2월~현재 : 한국통신 연구개발 본부 교환기술 연구소 연구원
※주관심분야 : 이동통신, 통신신호처리, 음성코딩

李 周 鎬(Ju Ho Lee) 정회원
1971년 12월 3일생
1993년 2월 : 한국과학기술원 전기 및 전자공학과 학사
1995년 2월 : 한국과학기술원 전기 및 전자공학과 석사
1995년 3월~현재 : 한국과학기술원 전기 및 전자공학과 박사과정
※주관심분야 : 이동통신, 통신신호처리, 채널코딩



金 炯 明(Hyung Myung Kim)정회원
1952년 10월 24일생
1974년 2월 : 서울대학교 공학사
1982년 4월 : 미국 Pittsburgh대학 전기공학과 석사
1985년 12월 : 미국 Pittsburgh대학 전기공학과 공학박사
1986년 4월~1992년 8월 : 한국과학기술원 전기 및 전자공학과 조교수
1992년 9월~현재 : 한국과학기술원 전기 및 전자공학과 부교수
※주관심분야 : 디지탈 신호와 영상처리, 다차원시스템 이론, 비디오신호 전송통신 이론, 이동 통신 기술 분야