
다중 코드북 구조를 이용한 CELP형 음성부호화기

박규정*, 한승조**

CELP speech coder by the structure of multi-codebook

Kyu-Joung Park, Seung-Jo Han

요 약

본 논문에서는 다중 코드북 구조를 이용하여 CELP 부호화기의 계산량의 증가 없이 고품질의 음성을 합성할 수 있는 코드북 구조를 제안한다. 또 제안된 코드북 구조를 가지는 4.8kbps CELP형 음성부호화기를 설계한다. 제안된 다중 코드북 구조는 기본 코드북과 두 종류의 스펙트럼과 피치 보강을 위한 코드북으로 구성되어 각각의 코드북에서 검색하여, 두 개의 코드북의 합으로 여기신호를 표현하고 각각의 서로 다른 코드북 이득을 사용하기 때문에 정확한 이득을 표현할 수 있어 기존의 개선 알고리즘보다 더 나은 음질을 제공할 수 있다. 본 코드북 구조를 갖는 4.8kbps CELP형 부호화기를 설계하여 컴퓨터 모의 실험한 결과, 같은 전송률을 갖는 DoD CELP 부호화기보다 segSNR가 0.81dB 더 높게 나타났다.

ABSTRACT

In this paper we propose a multi-codebook structure which can synthesize high quality speech without increasing of CELP coder's computation. We also design a 4.8kbps CELP speech coder with the proposed codebook structure. The proposed multi-codebook structure is made up of basic codebook and the other codebook which is formed for strengthen spectrum and pitch. Multi-codebook structure can represent accurate gains since it represents excitation signals as summation of two kinds of codebooks and uses different codebook gains respectively. Therefore it can provide better speech quality than other conventional structures. In computer simulation of the 4.8kbps CELP coder designed with the proposed codebook structure its segSNR was 0.81dB more high than the DoD CELP coder of same transmission rates.

* 송원대학 정보통신과

** 조선대학교 전자정보공과대학 전자·정보통신공학부

접수일자: 2000. 10. 22

I. 서 론

최근 이동통신의 수요가 급증하면서 유선통신의 수요를 넘어 서고 있다. 이에 따라 제한된 채널을 효율적으로 이용하기 위하여서도 높은 음성전송품질(toll quality)을 제공하기 위한 많은 연구가 계속되고 있다. 또한 멀티미디어의 발달과 함께 정보전달과 저장에 있어서 인간에게 가장 친숙한 음성의 데이터 처리에 대한 연구는 매우 중요하다.

이러한 음성신호처리기술은 1938년 R. Dudley에 의해 한국 박람회에서 음성신호를 음원과 여파기로 분리해서 분석하고 이것을 다시 복원하는 음성부호화기(VOCODER)의 이론이 제안된 후로부터 음성부호화에 관한 연구가 계속되어왔다. 음성부호화법에는 크게 파형 부호화법(waveform coding), 신호원 부호화법(source coding), 두 기법을 혼합한 혼성 부호화법(hybrid coding) 등이 있다. 이러한 부호화법 중에서 혼성 부호화법은 비교적 낮은 전송율에서도 높은 음질의 음성을 합성할 수 있다. 혼성 부호화법은 여기신호를 어떻게 모델링하느냐에 따라 RELP(Residual Excited Linear Prediction), MPLPC(Multi-Pulse Excited Linear Prediction Coding), RPLPC(Regular Pulse Excited LPC), CELP(Code Excited Linear Prediction), VSELP(Vector Sum Excited Linear Prediction)로 구분된다. 이 중 CELP 음성부호화기[1]는 16kbps 이하의 전송률에서도 높은 음질을 합성할 수 있어, 많은 음성부호화 방법의 기본이 되고 많은 음성부호화기에 응용이 된다. 그 예로 ITU의 표준 음성부호화기인 LD-CELP[2]와 CS-ACELP[3], VSELP[4], EVRC, QCELP 등의 CELP 방식의 부호화기가 개발되었다.

CELP 음성부호화기는 입력 음성을 분석하여 선형 예측계수와 피치 정보를 분석한 다음, 이를 이용 코드북내에 저장된 여기신호열과 합성하여 입력 음성과의 오차를 최소로 하는 여기신호열을 선택하는 합성에 의한 분석법(analysis-by-synthesis)을 사용한다. 이러한 방법은 모든 코드북내의 여기신호열을 합성하여 비교 분석하기 때문에 높은 음질의 음성을 합성할 수 있지만, 그로 인하여 방대한 계산량을 필요로 한다. 또한 입력 음성에서 선형예측계수와 피치 정보 분석 후 잔류신호는 백색잡음형태로 나타나야 하나, 완벽한 분석이 이루어지지 못하고 잔류신호에

약간의 선형예측계수와 피치 정보가 남아 있게 되어 음질을 떨어뜨리는 원인이 된다. 그리고 코드북의 크기와 구조에 따라 부호화기의 성능과 계산량이 달라지기 때문에 코드북 구조 개선에 관한 연구가 계속되고 있다.

본 논문에서는 codebook I과 codebook II-1, 그리고 codebook II-2로 이루어진 다중 코드북을 이용하여 여기신호열을 합성하는 알고리즘을 제안한다. codebook I과 codebook II의 합으로 여기신호열을 합성하고, 각각 두 번 검색한다. 또한 codebook II-1, 그리고 codebook II-2는 유·무성음에서 스펙트럼 정보와 피치 정보를 보강하기 위하여 사용된다.

2장에서는 CELP 음성부호화기의 기본구조에 대해서 설명하고, 3장에서는 다중 코드북 구조에 대해서 기술한다. 4장에서는 제안된 다중 코드북 구조를 이용하여 4.8kbps CELP형 음성부호화기를 설계하여 모의 실험 및 성능 평가를 5장에서는 결론을 맺는다.

II. CELP 음성부호화기

CELP(Code Excited Linear Prediction) 음성부호화기는 합성에 의한 분석법을 사용하는 혼성부호화기 중에서 가장 널리 쓰이고 있다. CELP 음성부호화기는 1982년에 Atal에 의해서 처음 제안된 후 최근까지도 변형된 알고리즘이 계속 제안되고 있다.

CELP 음성부호화기의 부호화 과정은 선형예측기의 필터계수와 피치 필터계수를 구하는 과정과 여기신호를 결정하는 코드북 탐색과정으로 나뉜다. 선형예측기의 예측 필터계수는 단구간 상관관계(short-term correlation)를 나타내는 LPC 분석을 함으로써 구해지고, 피치 필터계수는 장구간 상관관계(long-term correlation)를 이용하여 구한다[5].

이러한 피치 필터 $1/P(z)$ 와 예측 필터 $1/A(z)$ 는 다음과 같이 표현될 수 있다.

$$\frac{1}{P(z)} = \frac{1}{1 - bz^{-M}} \dots\dots\dots (1)$$

$$\frac{1}{A(z)} = \frac{1}{1 - a_1z^{-1} - a_2z^{-2} \dots - a_mz^{-m}} \dots (2)$$

식 (1)의 피치 필터 파라미터 b 와 M 은 개회로(open loop) 구조나 폐회로(close loop) 구조 중의 하나를 이용하여서 얻어질 수 있는데, 폐회로 구조를 이용하여

구한 경우의 음질이 훨씬 우수하므로 일반적으로 폐회로 구조를 사용한다. 식 (2)의 LPC 계수는 음성 압축시 양자화에 부적절하다고 알려져 있어, LPC 계수화 같은 스펙트럼에 관한 정보를 가지며, 안정성이 양자화시에도 보장되고 생동폭이 매우 좁은 특성을 가진 LSP(Line Spectrum Pair) 계수로 변환하여 사용한다[6].

이러한 피치 정보와 스펙트럼 정보를 원 음성신호에서 제거하고 남아 있는 신호가 백색잡음 형태로 나타난다. 이러한 신호를 잔류신호(여기신호)라고 하는데 이는 음성신호보다 더 낮은 분산을 가지고 있어 더 쉽게 양자화할 수 있다. 그러나 저전송율을 필요로 하는 부호화기에서는 이 잔류신호를 직접 양자화하기 어렵기 때문에 코드북을 이용하여 양자화 한다. 그러나 잔류신호와 코드북내의 여기신호열과의 오차

를 최소화시킨다고 해서 입력 음성신호와 합성신호의 오차가 최소화된다고 말할 수 없다.[5] 그래서, CELP 음성부호화기는 코드북내의 여기신호열과 입력 음성신호로부터 구한 파라미터를 이용하여 합성한 음성신호와 입력 음성신호와 비교하여 오차를 최소화하는 여기신호열로 잔류신호를 양자화하는 합성에 의한 분석(analysis-by-synthesis)법을 사용한다. 그림 1은 CELP 음성부호화기의 기본 구조를 나타낸 것이다.

그림 1에서 $s(n)$ 은 입력 음성신호이고 $x(n)$ 은 코드북내에 저장되어 있는 여기신호열이다. 입력 음성신호를 이용하여 피치 정보와 스펙트럼 정보를 구한 후, 코드북 내에 있는 여기신호열 $x(n)$ 에 이득 b_k 을 조절한 후 피치 필터 $[1/P(z)]$ 와 스펙트럼 필터 $[1/A(z)]$ 를 통과시켜 합성신호 $\hat{s}(n)$ 를 구한다. 그 후 합성신호 $\hat{s}(n)$ 와 입력 음성신호 $s(n)$ 과의 차이를 오

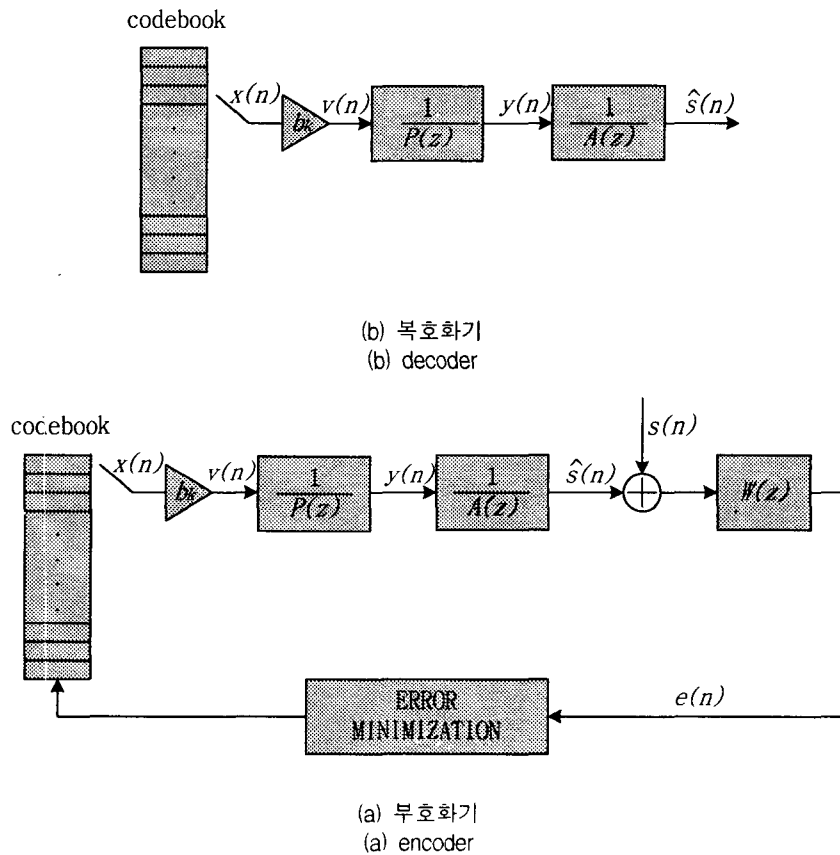


그림 1. CELP 음성부호화기의 기본 구조
Fig. 1 Basic structure of the CELP speech encoder

(weighting filter) $W(z)$ 를 통과시켜 오차신호 $e(n)$ 을 구한다. 이 오차신호 $e(n)$ 를 평균 제곱오차를 이용하여 비교함으로써 가장 적은 오차를 가지는 여기신호열 $x(n)$ 을 구한다.

이러한 방법은 코드북 내의 모든 여기신호열을 비교 검색하기 때문에 높은 음질의 음성을 합성할 수 있으나, 코드북의 크기가 클수록 연산량이 많아지게 된다. 그러나 코드북 크기의 단순한 축소는 심각한 음질저하를 가져올 수 있다. 그래서 많은 알고리즘이 개발되었는데 그 대표적인 예가 희박 중첩형 코드북(sparse and overlapped codebook)이다. 희박 중첩형 코드북이란 코드북의 여기신호열에 0의 값을 많이 부여하여 계산량을 줄이고, 인접 여기신호열간에 상관관계를 부여함으로써 검색시간을 단축시켰다. 그러나 이러한 방법은 인접 여기신호열 간의 의존성 때문에 랜덤 코드북에 비해 음질 저하를 초래할 수 있다.

III. 다중 코드북 구조

1. 다중 코드북 구조

단구간 예측기, 즉 스펙트럼 분석으로는 입접 샘플간의 상관관계는 제거할 수 있지만, 유성음 구간에서는 준주기성을 제거하지는 못한다.[1] 이러한 준주기성, 즉 피치 정보를 제거하기 위해 장구간 예측기를 사용한다. 입력 음성신호를 분석하여 피치 정보와 스펙트럼 정보를 원 음성신호에서 제거하고 남아 있는 신호는 이론적으로는 백색잡음 형태로 나타난다. 그러나 실제 부호화기에서는 완벽한 분석이 이루어지지 못하고, 또 양자화 잡음 때문에 잔류신호에 약간의 피치성분이나 스펙트럼 정보가 남아 있게 된다. 이런 상태로 코드북을 이용하여 다시 양자화하기 때문에 원 음성신호의 정보가 손실되어 음질을 저하시키는 원인이 된다. 또한 피치는 음색을 결정하는 중요한 정보이기 때문에 피치 정보는 음질에 많은 영향을 미친다고 할 수 있다.

본 논문에서는 3개의 코드북을 사용하여 최적의 여기신호열을 검색하는 다중 코드북 구조를 제안한다. 하나의 코드북을 사용하는 기존의 방법과는 달리 3개의 코드북을 사용하여 두 개의 선형 회귀 필터, 즉 피치와 스펙트럼 정보를 보강하여 준다.

다중 코드북 구조는 *codebook I*과 *codebook II-1*, 그리고 *codebook II-2*로 이루어진다. *codebook*

*I*는 -1, 1과 70% 이상의 0의 값을 갖는 희박성 코드북이고, *codebook II-1*은 -1, 0, 1의 값만을 가지는 랜덤 코드북이다. *codebook II-2*의 구조는 다른 코드북에 비해 약간 다르다. *codebook II-2*는 $\pm 1, \pm 1.5, 2.5$ 과 80% 이상의 0의 값을 갖는 희박 중첩형 코드북이다. 0, $\pm 1, \pm 1.5, 2.5$ 의 값들이 한 점을 중심으로 여기신호열 크기의 1/5내로 분포되어 있고, 인접 여기신호열간에 상관관계를 갖는 중첩성을 가지고 있다. 이러한 구조를 가지는 *codebook II-2*는 유성음의 피치 정보를 보완하기 위한 코드북이다.

2. 다중 코드북 부호화기 구조

그림 2는 제안된 다중 코드북 구조를 나타낸 것이다. 다중 코드북의 검색 방법은 다음과 같다. 입력 음성신호의 피치 정보와 스펙트럼정보를 구한 후, *codebook I*내의 여기신호열 $x_1(n)$ 을 이득을 적절히 조절하여 두 개의 선형 회귀 필터 $1/P(z), 1/A(z)$ 를 통과시켜 합성 음성 $\hat{s}I(n)$ 을 구한다. 이 합성 음성 $\hat{s}I(n)$ 과 $s(n)$ 과의 차이를 오차 가중 필터 $W(z)$ 에 통과시키면 오차신호 $e(n)$ 을 얻는다. 이 $e(n)$ 을 평균 제곱 오차를 이용하여 비교함으로써 가장 적은 오차신호를 나타내는 $x_1(n)$ 을 구한다. 이렇게 구한 여기신호열 $x_1(n)$ 의 이득을 조절하여 만든 여기신호 $v_1(n)$ 과, *codebook II-1, II-2*내의 여기신호열 $x_{2-1}(n), x_{2-2}(n)$ 의 이득을 조절하여 만든 여기신호 $v_2(n)$ 을 합성하여 합성 여기신호 $v(n)$ 을 구한다. 이 여기신호열을 다시 시변 선형 회귀 필터를 통과시켜 합성 음성신호 $\hat{s}(n)$ 을 구한다. 그 후 합성 음성신호 $\hat{s}(n)$ 과 입력 음성 $s(n)$ 과의 차를 오차 가중 필터 $W(z)$ 에 통과시켜 얻은 오차신호 $e(n)$ 을 평균 제곱오차를 이용하여 비교함으로써 가장 적은 오차신호를 나타내는 여기신호 $x_2(n)$ 하나를 구한다. 즉 *codebook II-1, II-2*에서 각각의 여기신호열을 구하는 것이 아니라 두 개의 코드북에서 단 하나의 최적 여기신호열을 구한다.

잡음성분이 많은 무성음에서는 주로 *codebook II-1*의 여기신호열이, 주기적이고 피치성분이 많은 유성음에서는 주로 *codebook II-2*의 여기신호열이 최적의 여기신호열이 된다. 또한 유·무성음 천이구간에서는 *codebook II* 전체구간 중 최소의 오차를 나타내는 최적의 여기신호열이 선택된다. 두 개의 코드북의 여기신호열을 합성함으로써 여러 가지의 여기신호를 만들어 낼 수 있다. *codebook I*의 크기가 K

이고, *codebook II-1*의 크기가 N , *codebook II-2*의 크기가 M 이라면, 이론적으로 $K \times (N+M)$ 개의 여기 신호를 합성해 낼 수 있다. 또 이득 또한 코드북 각각 구하기 때문에 더욱 더 다양한 여기신호를 합성해 낼 수 있다. 그리고 *codebook II-1*에서는 무성음 성분을 *codebook II-2*에서는 유성음 성분의 피치를 보장하여 음질을 향상시켜준다. 그러나 두 개의 코드북 사용으로 인한 코드북에 대한 비트할당이 늘어나게 된다는 단점이 있다.

3. 다중 코드북 복호화기 구조

다음 그림 3은 제안된 다중 코드북의 복호화기의 구조를 나타낸 것이다. 복호화기에서는 전송된 두 개의 코드북 여기신호열을 이용하여 합성여기신호를 만든다. 즉, 전송된 *codebook I*의 여기신호열에 전송된 이득 b_k 를 조절한 여기신호 $v_1(n)$ 과 *codebook II*의 여기신호열에 이득을 조절한 여기신호 $v_2(n)$ 과의 합성여기신호를 만든다. 그 후, 두 개의 선형 회귀 필터 $1/P(z)$, $1/A(z)$ 를 통과시켜 음성을 합성한다.

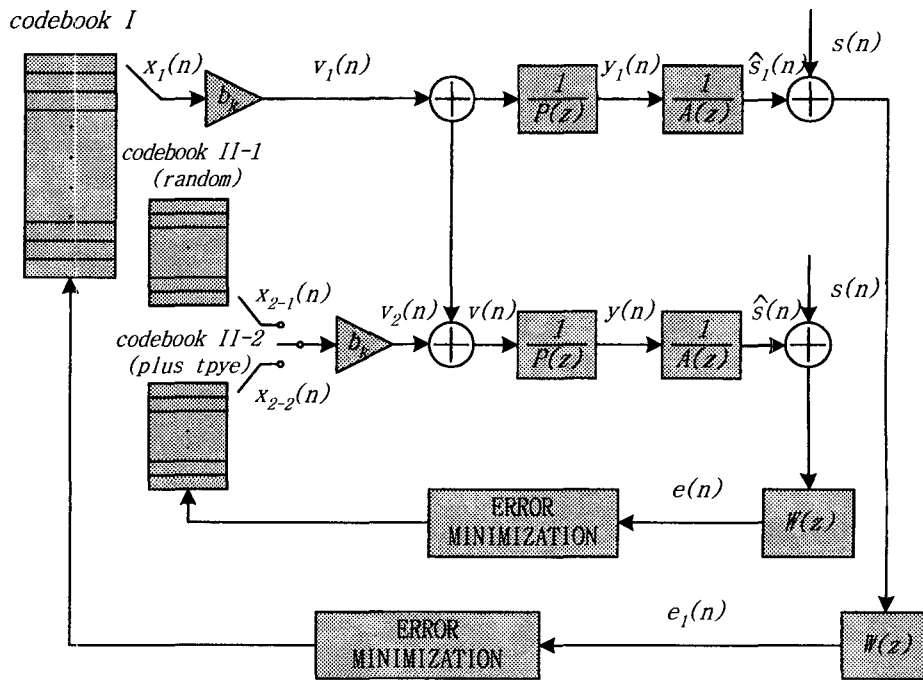


그림 2. 다중 코드북 부호화기 구조
Fig. 2 Structure of the multi-codebook encoder

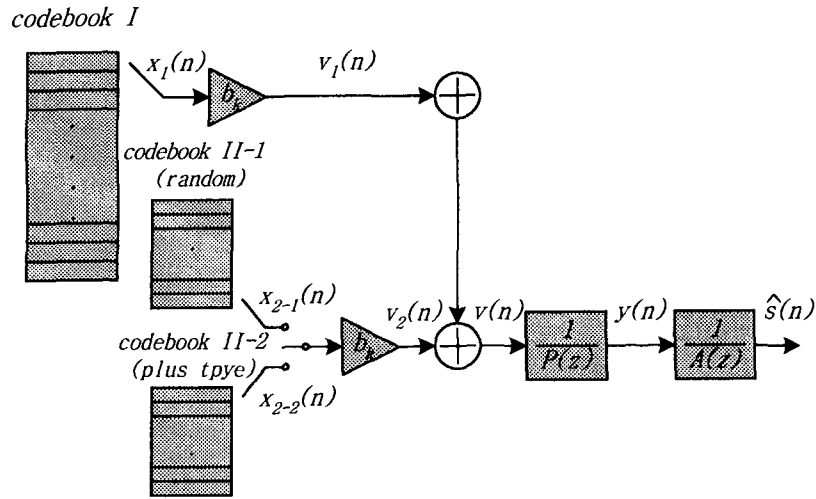


그림 3. 다중 코드북 복호화기 구조
Fig. 3 Structure of the multi-codebook decoder

IV. 실험 및 성능 평가

1. 다중 코드북 구조의 4.8kbps CELP형 음성 부호화기

제안된 방법의 성능평가를 위해서 제안된 코드북 구조를 갖는 음성부호화기를 C언어로 프로그래밍하여 컴퓨터 모의 실험을 하였다.

음성 분석을 위해 각각의 프레임(frame)으로 나누었다. 프레임의 길이에 따라 음성분석 방법과 음질이 달라진다. 인간은 30ms 이하의 지연에서는 인지를 잘하지 못하기 때문에 본 논문에서는 4kbps대의 음성부호화기에서 일반적으로 사용하는 30ms를 한 프레임으로 잡았다. 그리고 한 프레임을 각각 7.5ms씩 4개의 부프레임으로 나누어 분석하였다.

LPC 분석을 위해 37.5ms의 Hamming 창함수를 7.5ms씩 중첩하여 사용하였다. 선형예측계수를 양자화하는 방법에는 스칼라 양자화(Scalar Quantization : SQ)법과 벡터 양자화(Vector Quantization : VQ)법이 있다. LPC 계수는 전송을 위해 LSP로 변환하였다. 본 논문에서는 다중 코드북 사용으로 인한 비트류의 증가를 막기위하여 비교적 적은 비트률에서도 만족할 만한 성능을 내는 양자화하기 위하여 벡터 양자화법을 사용하였다. 벡터 양자화법은 10차의 LSP 계수를 3~4개로 묶어 코드벡터로 양자화하는 방법

이다. 이러한 방법으로는 분리 벡터 양자화(Split Vector Quantization : SVQ)[7], 다단 벡터 양자화(Multi-Stage Vector Quantization : MSVQ)[8] 등이 있다. 이러한 벡터 양자화기는 24~30비트에서 만족할 만한 성능을 유지한다고 알려져 있다.[7][8] 이러한 벡터양자화 방법 중 비교적 구현이 용이하고, 연산량이 적은 SVQ를 사용한다. 본 연구에서는 10개의 LSP계수를 낮은 주파수부터 차례로 4, 3, 3개씩 묶는 4-3-3 SVQ를 사용하여 총 26비트로 양자화 하였다.

장구간 상관관계에 해당하는 피치 검색은 적절한 피치 조건을 만족하는 값을 반복적인 비교를 통하여 구하는 폐회로 구조에 의해 수행되기 때문에 코드북 검색 다음으로 계산량이 많다. 이러한 계산량을 줄이기 위해 개발된 알고리즘중 가장 대표적인 것이 적응 코드북(adaptive codebook)과 델타 피치 검색(delta pitch search)[9][10]이다. 이 중 본 논문에서는 음질의 저하 없이 계산량과 비트수를 줄일 수 있는 델타 검색법을 사용한다. 30ms의 한 프레임 중 첫 번째/세 번째 부프레임에서는 20에서 147사이의 모든 범위에서 피치 검색을 수행하고, 두 번째/네 번째 부프레임에서는 바로 앞 첫 번째/세 번째 부프레임에서 찾은 피치 값을 중심으로 32범위 안에 있는 피치 값들에 대해서만 검색을 수행한다. 첫 번째/세 번째 부프레임의 피치값은 7비트씩, 두 번째/네 번째 부프레임의

피치값은 5비트씩으로 부호화하였다. 그리고 피치 이득은 -1과 +2 사이의 값을 16개의 값으로 나누어 4비트로 부호화하였다.

코드북은 3장에서의 다중 코드북을 사용한다. *codebook I*은 -1, 1과 70% 이상의 0의 값을 갖는 희박성 코드북으로 128의 크기를 가지고, *codebook II*는 64의 크기를 가진다. *codebook II*는 2개의 코드북으로 나뉘는데, *codebook II-1*은 -1, 0, 1의 값만을 가지는 랜덤 코드북으로 32의 크기를 가진다. *codebook II-2*는 ± 1 , ± 1.5 , 2.5과 80% 이상의 0의 값을 갖는 희박 중첩형 코드북으로 32의 크기를 갖는다. *codebook II-2*의 여기신호열은 입력신호에서 구한 피치에 따라 조정이 되기 때문에 항상 변한다. *codebook II-2*의 여기신호열의 중심값이 피치값을 중심으로 위치가 조정된다. 즉 0, ± 1 , ± 1.5 , 2.5의 값들이 한 점을 중심으로 여기신호열 크기의 1/5내로 분포되어 있고, 그 중심점이 입력신호에서 구한 피치값을 중심으로 조정된다. 이는 코드북에 할당되는 비트수를 줄이기 위한 방법으로, 또한 *codebook II-2*는 피치를 보강하기 위한 것으로 피치값에 따라 여기신호열의 위치가 가변함으로 코드북 성능의 저하 없이 비트수를 낮출 수 있다.

2. 실험 방법

제안된 4.8kbps CELP형 음성부호화기 규격을 다시 나타내면 다음 표 1, 2와 같다.

표 1. 4.8kbps CELP형 음성부호화기의 규격
Table 1. Specification of the the 4.8kbps CELP speech encoder

항 목	값
Data Rate	4800 bit/sec
Sampling Rate	8kHz
Frame Size	240 samples(30ms)
Pitch Subframe Size	60 samples(7.5ms)
Codebook Subframe Size	60 samples(7.5ms)
LPC Filter Order	10
Pitch Predictor Order	1
Codebook Order	128×64 =8192

표 2. 4.8kbps CELP형 음성부호화기의 Bit 할당
Table 2. Bit assignment of the 4.8kbps CELP speech encoder

Parameter	bit/7.5ms subframe	bit/30ms frame
LPC(LSP)		28
Pitch	7, 5, 7, 5	24
Pitch gain	4, 4, 4, 4	16
Codebook	7+6, 7+6, 7+6, 7+6	52
Codebook gain	3+3, 3+3, 3+3, 3+3	24
Total		144

평가 시료로는 20~50대 남녀 화자가 각각 연령대 별로 6명(남자 3명, 여자 3명)씩 총 24명이 발음한 음성을 사용하였다. 녹음 환경은 소음이 적은 조용한 실내에서 컴퓨터 마이크로폰을 이용하여 직접 녹음하였다. 음성 데이터는 8kHz로 샘플링(sampling)하였고, 16비트로 양자화하였다. 실험에 사용된 문장시료는 다음 표 3과 같다.

표 3. 평가 시료
Table 3. Speech database

문 장	내 용
문장 1	영, 일, 이, 삼, 사, 오, 육, 칠, 팔, 구, 십
문장 2	여보세요. 안녕하십니까?
문장 3	지금은 전화를 받을 수 없으니, 다음에 다시 걸어 주십시오.
문장 4	여보세요. 안녕하십니까? 김철수의 전화입니다.
문장 5	바람과 햇님이 힘이 더 세다고 다투고 있을 때
문장 6	한 나그네가 따뜻한 외투를 입고 걸어 왔습니다.
문장 7	북풍은 힘껏 불었으나 불면 불수록 나그네는 외투를 단단히 여몄습니다.
문장 8	북풍은 햇님이 들 중에 힘이 더 세다고 인정하지 않을 수 없었습니다.

3. 성능 평가

제안된 다중 코드북 구조를 가지는 4.8kbps CELP형 음성부호화기의 성능을 평가하기 위하여 비교 기준 부호화기로 4.8kbps DoD CELP 음성부호화기[11]을 선정하였다. 성능 평가에는 객관적 평가 방법인 *segSNR*(segmental SNR)을 사용하였다.

실험 결과는 다음 표 4와 같다. 표 4에서와 같이 제안된 다중 코드북 구조의 4.8kbps CELP형 음성부호화기의 *segSNR*는 13.45dB로 같은 전송률을 갖는 DoD CELP의 12.64dB보다 0.81dB 높게 나타났다.

그림 4는 원 입력 음성의 파형과 DoD CELP 음성부호화기와 제안된 코드북 구조를 갖는 음성부호화기의 합성 파형을 나타낸 것이다. 그림에서 볼 수 있듯이 제안된 코드북 구조를 갖는 음성부호화기가 DoD CELP 음성부호화기보다 더 원 입력 음성신호를 잘 표현하였음을 알 수 있다. 즉 제안된 다중 코드북 구조의 CELP형 음성부호화기는 같은 전송율을 가지는 DoD CELP 음성부호화기보다 더 우수한 성능을 보였다.

기존의 코드북보다 복잡한 여기신호를 잘 표현할 수 있어 높은 음질의 음성을 합성할 수 있었다.

다중 코드북 구조는 -1, 1과 70% 이상의 0의 값을 갖는 희박성 코드북인 *codebook I*과 -1, 0, 1의 값을 갖는 랜덤 코드북 *codebook II-1*, 그리고 피치를 보강하기 위한 *codebook II-2*로 구성하였다. *codebook I*에서 최적 여기신호열 검색 후, 이 여기신호와 다시 *codebook II*의 여기신호열의 합으로 최적 여기신호열을 검색하여, 두 번 검색하고 두 개의 여기신호열과 각각의 이득으로 여기신호를 표현하기 때문에 다양한 음성을 합성할 수 있다. 또한 *codebook II*에서 *codebook II-1*는 주로 무성음을, *codebook II-2*에서는 주로 유성음의 피치성분을 보강을 하여 더 나은 음질의 음성을 합성할 수 있었다.

제안된 다중 코드북을 이용하여 CELP형 음성부호화기를 설계하여 같은 전송율을 갖는 DoD CELP 음성부호화기와 비교한 결과 *segSNR*이 0.81dB 더 높게 나타났다.

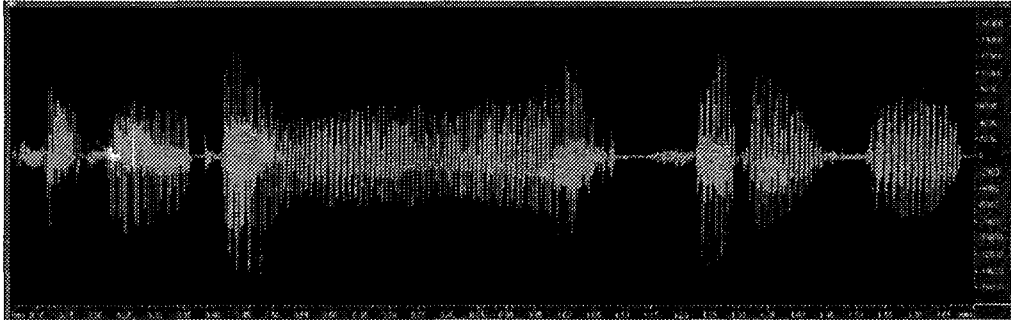
제안된 다중 코드북 구조를 가지는 음성부호화기는 실시간 구현이 가능한 연산량을 가지고 성능을 분석한 결과 같은 조건에서의 4.8kbps DoD CELP 음성부호화기에 비해 성능이 우수함을 확인하였다.

표 4. *segSNR* 비교표 (단위 : dB)
Table 4. *segSNR* comparison

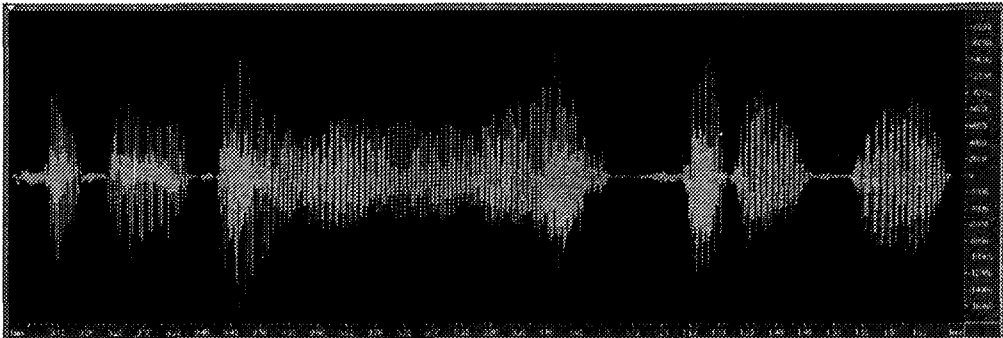
문 장	DoD CELP	다중 코드북 구조 CELP
문장 1	12.02	13.54
문장 2	11.96	12.05
문장 3	13.24	13.57
문장 4	12.27	14.01
문장 5	12.89	13.43
문장 6	12.91	13.04
문장 7	13.02	13.95
문장 8	12.83	14.02
평 균	12.64	13.45

V. 결론

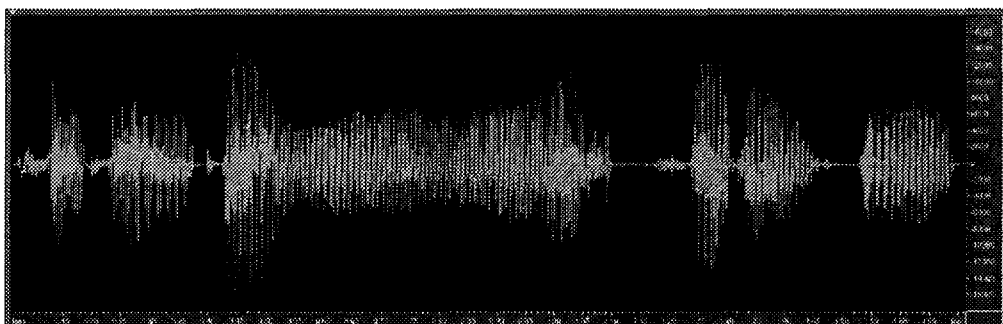
본 논문에서는 다중 코드북 구조를 이용하여 높은 음질의 음성을 합성할 수 있는 알고리즘을 제안하였다. 3개의 코드북을 이용하여 음성을 합성함으로써



(a) 입력 음성 파형
(a) Input speech waveform



(b) DoD CELP 음성부호화기의 합성 음성 파형
(b) Synthesized speech waveform of the DoD CELP speech encoder



(c) 다중 코드북 CELP형 음성부호화기의 합성 음성 파형
(c) Synthesized speech waveform of the multi code book CELP speech encoder

그림 4. 입력 음성과 합성 음성의 파형비교
Fig. 4 Waveform comparison between input speech and synthesized speech

참고문헌

- [1] M. R. Schroeder, B. S. Atal, "Code-Excited Linear Prediction(CELP) High-Quality Speech at Very Low Bit Rates," IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc., pp. 937-940, 1985.
- [2] ITU, "ITU Recommendation G.728 : Coding of speech at 16kbit/s using Low-Delay Code Excited Linear Prediction," International Telecommunication Union, 1992
- [3] ITU, "ITU Recommendation G.729 : Coding of speech at 8kbit/s using Conjugate_Structure Algebraic-Code-Excited Linear Predictive(CS-ACELP)," International Telecommunication Union on 1996.
- [4] I. A. Gerson, M. A. Jasiuk, "Vector-Sum Excited Linear Prediction(VSEL) Speech Coding at 8Kbps," IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc., pp. 641-644, 1990.
- [5] Peter Kroon and Ed F. Deprettere, "Class of Analysis-by-Synthesis Predictive Coders for high quality speech coding at rates between 4.8 and 16Kbps," IEEE J. SAC., vol. SAC-4, pp. 39-48, 1986.
- [6] F. K. Soong, B. H. Juang, "Line Spectrum Pair and Speech Data Compression", IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc., pp. 1.10.1-1.10.4, 1984.
- [7] K. W. Law, and C. F. Chan, "A Novel Split Residual Vector Quantization Scheme for Low Bit Rate Speech Coding," IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc., pp. I-493-496, 1994.
- [8] K. K. Paliwal, and B. S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 bits/frame," IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc., pp. 661-664, 1991.
- [9] T. Fingscheidt, T. Wiechers and E. Delfs, "Implementation Aspects of the GSM Half-rate Speech Codec", EUROSPEECH., pp. 723-726, Sep. 1995.
- [10] I. A. Gerson, M. A. Jasiuk, "Techniques for Improving the Performance of CELP-Type Speech Coders," IEEE journal on Selected Areas in Communications., vol. 10, no. 5, Jun., 1992.
- [11] J. P. Campbell, T. E. Tremain, and V. C. Welch, "An Expandable Error-protected 4800bps CELP Coder," IEEE ICASSP., pp. 735-738, 1989.
- [12] L. Torres and J. Huguet, "An omprovement on codebook search for vector quantization," IEEE Trans. Comm., COM-42, pp.208-210, 1994.
- [13] K. Mano et al, "Design of a pitch synchronous innovation CELP coder for mobile communications," IEEE Journal on Selected Areas in communications., vol. 13, pp. 31-41, Jan. 1995.
- [14] S. Miki, K. Mano, H. Ohmuro, T. Moriya, "Pitch Synchronous Innovation CELP (PSI-CELP)," EUROSPEECH., pp. 261-264, Sep. 1993.



박규정(Kyu-joung Park)

1976년 2월 명지대학교 전자
공학과 학사

1982년 2월 명지대학교 전자
공학과 석사

1998년 8월~현재 조선대학교
대학원 전자공학과 박사과정

1979년~1984년 2월 제주전문대학 전임강사

1984년 송원대학 정보통신과 교수

※주관심분야: 음성신호처리, 통신보안



한승조(Seung-jo Han)

1980년 2월 조선대학교 전자
공학과 학사

1982년 2월 동대학원 전자공
학과 석사

1994년 2월 충북대학교 대학
원 전자계산학과 박사

1986년 5월~1987년 3월 Univ. of New Okeans 객원
교수

1995년 2월~1996년 1월 Univ. of Texes 객원교수

1997년~현재 조선대학교 전자·정보통신공학부 교수

※주관심분야: 통신보안시스템설계, 네트워크 보안,
ASIC 설계, 음성신호처리