
한국어 음성에 있어서 저전송률을 갖는 개선된 VSELP코드북 설계

김형종*, 한승조**

Design of Low Bit Rate VSELP Codebook for the Korean Speech

Hyung-Jong Kim*, Seung-Jo Han**

요 약

본 논문에서는 제한된 대역에서 낮은 전송률로 좋은 품질을 유지하도록 하는 개선된 4.8kbps VSELP를 제안한다. 그러나 대부분의 경우에 있어서 낮은 전송률에는 좋은 품질을 유지하지 못하는 실정이다. 이러한 문제점을 해결하기 위해 많은 방법들이 제안되어 왔으나 대부분 외국어를 기준으로 맞추어져 우리 언어 구조에 적합하지 않다. 본 실험은 잡음이 없는 실험실에서 녹취한 데이터를 가지고 수행되었다. 본 논문은 저전송률을 가지며 한국어 음성에 적합한 코드북을 설계하고 8kbps의 VSELP와 4.8kbps의 VSELP를 SEGWSNR(Segmentally Weighted SNR) 평가와 MOS(Mean Opinion Score) 평가를 수행하고 주관적 평가에 있어서 4.8kbps의 VSELP의 우수함을 보인다.

Abstract

This paper proposed an improved 4.8kbps VSELP in order to keep the good quality in band-limited channel. In the most cases, it is difficult to keep the good quality at the low bit rate. In order to solve the problems, many methods are proposed, but they are not suitable to the Korean language structure because they are designed for being suitable to the foreign language structure. In experiment, we use the noiseless Korean voice data. We show that the proposed 4.8kbps VSELP is not excellent to the 8kbps VSELP in SEGWSNR(Segmentally Weighted SNR), but it is the superior to the 8kbps VSELP in the MOS(Mean Opinion Score) test.

* 조선대학교 대학원 전자공학과 박사과정

** 조선대학교 전자·정보통신공학부 교수

접수일자 : 1999년 9월 6일

I. 서 론

제한된 통신 대역을 효율적으로 이용하고 양질의 서비스를 하기 위해 대부분의 통신 방식이 디지털화 되고 있다. 사용할 수 있는 주파수 대역폭이 제한되어 있기 때문에 가능한 낮은 전송 속도로 음성 데이터를 압축하면서 음질의 손상을 최소화할 수 있는 디지털 음성 부호화기에 대한 연구가 필요하다[1]. 이러한 시점에서 대부분의 음성 코드북은 한국어에 적합한 구조로 설계되지 못하고 외국에 의존하고 있는 실정이다. 음성이라고 하는 것은 국가, 개인, 연령, 성별 등에 따라서는 달라지는 것은 물론이고 동일인의 경우에도 시간, 감정에 따라 특성이 달라진다. 현재 상용되고 있는 CDMA기술도 한국어의 발음 특성과 외국어의 발음 특성과의 차이점, 한국인과 외국인의 발성구조 및 발성습관의 차이 등을 고려하여 설계되어 있지 않다. 이렇듯 외국의 보코더 기술로 우리말을 효과적으로 전송한다는 것은 매우 어려운 일이다. 국내의 음성 코딩 기술확보와 국제화 측면에서도 한국어의 음운 구조에 적합한 보코더는 반드시 우리가 개발해야 한다.

음성 부호화 연구의 비약적 발전을 가져온 CELP(Code Excited Linear Prediction)는 half-rate 정도의 저전송 비트율에서도 우수한 음질을 유지하려고 분석-합성(Analysis By Synthesis) 코딩 방법으로서 성능 개선 연구와 함께 계산량을 줄이기 위한 연구가 진행되어왔다. 또한 VSELP(Vector Sum Excited Linear Prediction)는 CELP형 부호화기의 일종으로서 기존의 CELP방식에 비해 아주 적은 양의 계산을 필요로 하고, 또한 음질이 우수하여 북미 및 일본의 디지털 이동통신의 full-rate(8kbps, 6.7kbps) 표준안으로 채택되어있고 최근에는 GSM(Global Systems for Mobile Communication)의 half-rate 표준안으로 채택되어있다[2].

VSELP는 데이터의 압축을 위해 이산 벡터들의 시퀀스를 디지털 시퀀스로 매핑하는 벡터 양자화 코딩 기술을 사용하고 구조화된 코드북을 사용하여 CELP의 계산량을 줄이고, 코드북 훈련 과정을 통해 음질을 향상시킨 방법으로 채널 잡음에도 강한 CELP형 부호화기이다. IS-54/IS-136 표준의 음

성 압축 방식인 VSELP(Vector Sum Excited Linear Prediction)는 A/D로 변환되어 들어오는 64kbps의 음성 신호를 7.95kbps로 압축하여 송신단으로 보내고, 수신단으로부터 받아들여진 압축된 음성신호를 복원해주는 기능을 한다. VSELP 알고리즘은 사람의 음성신호를 해석하여 특징적인 음성 파라미터를 추출하고 추출된 파라미터를 이용하여 효율적으로 음성신호를 압축하는 방식이며 음성 해석 과정에서 추출된 파라미터를 이용하여 음성을 합성함으로써 적은 비트율로서 효과적인 음성신호를 만들어 낼 수 있다.

본 논문에서는 이러한 특성을 갖는 VSELP 부호기를 한국어 음성에 적합하고 저전송률을 가지며 음질이 우수한 개선된 저전송률 VSELP 알고리즘을 제안한다.

II. VSELP의 이론적 배경

제한된 통신 대역을 효율적으로 이용하기 위하여 신호를 압축하여 전송하는 방법이 다각적으로 연구 되어왔다. 기본적으로 시간적 중복성을 감소 제거하는 방법, 공간적 중복성을 감소 제거하는 방법, 통계적 중복성을 제거하는 방법 등이 있다. 디지털 음성 부호화 방법은 크게 파형 부호화기(waveform coder), 음원 부호화기(source coder)와 두 부호기의 장점을 이용한 복합 부호화기(hybrid coder)로 나눌 수 있다[3]. 파형 부호화기는 고음질의 원음을 재생할 수 있다는 장점은 있으나 데이터량이 많아 16kbps 이상의 전송속도를 가져야하며 음원 부호화기는 4.8kbps 이하의 전송 속도로 음성신호를 부호화 할 수 있으나 합성된 재생음은 질이 떨어진다는 단점이 있다. 이 중 코드북을 이용한 방법은 코드북을 완벽히 검색해야 하기 때문에 광범위한 계산을 요구하고, 또한 부하를 줄이기 위해 고속 코드북 검색이 필요하다. 코드북 검색과정은 코드워드(i)를 찾아야하는데 식(1)의 최소값으로 정한다.

$$E_{k,i} = \sum_{n=0}^{N-1} (t(n) - r_{kf,i}(n))^2 \dots\dots\dots (1)$$

식(1)에서 첫 번째 코드북의 $k=1$, 두 번째 코드북

의 $k=2$ 이고 r_k 는 각각의 코드벡터 i 의 최적이득이며 코드북 검색과정이 같기 때문에 코드북을 구분하는 k 는 생략된다.

코드북의 벡터는 코드북 검색의 이론을 기초로 하여 코드북 벡터는 식(2)을 최대로 하는 값으로 선택된다.

$$r_i = \frac{C_i^2}{G_i} \dots\dots\dots (2)$$

A/D 변환된 입력 음성 신호의 직류 성분과 60Hz Hum을 제거하기 위해 음성 처리하기 전에 4차 Chebyshev II 고역필터를 사용한다[11].

$$H_{hp}(z) = \frac{\sum_{i=0}^4 a_i z^{-i}}{1 - \sum_{i=1}^4 b_i z^{-i}} \dots\dots\dots (3)$$

이때 계수는 다음과 같다.

- $a_0 = 0.898025036$
- $a_1 = -3.59010601$
- $a_2 = 5.38416243$
- $a_3 = -3.59010601$
- $a_4 = 0.898024917$
- $b_1 = 3.78284979$
- $b_2 = -5.3739122$
- $b_3 = 3.39733505$
- $b_4 = -0.806448996$

전처리된 음성신호에서

$$\phi(i, k) = \sum_{n=N_p}^{N_p+i} s(n-i)s(n-k) \dots\dots\dots (4)$$

을 이용하여 공분산을 계산하고 이 공분산을 실효 대역폭이 80Hz인 이항 윈도우에 의해 평활한다. 이러한 SST(Spectral Smoothing Technique)는 포먼트의 대역폭을 넓혀주고 반사계수 양자화기의 성능을 개선시키는 역할을 한다[4][5].

VSELP 부호화기에서는 여러 개의 기저 벡터 선형 조합으로 구성되는 여기 신호로부터 최적의 여기 신호 벡터를 결정하고, 구해진 벡터를 피치 합성 필터 및 포먼트 합성 필터에 통과시켜 출력 신호를 발생시킨 후 스펙트럼을 보상하여 최종 합성음을 얻는다. 부호화기에서는 입력 신호와 합성된 신호의 인지 가중오차 에너지(perceptually weighted)가 최소가 되도록 전송파라미터를 구한다. 부호화기에서는 부호화기에서 전송된 코드북 관련 파라미터로부

터 여기 신호를 생성한 후, 선형 예측 계수를 사용하여 구성된 합성 필터를 통과시켜서 합성음을 재 생성한다. 그리고 합성음의 주관적 음질을 향상시키기 위해 피치 선필터링(pitch pre-filtering)과 적응 후필터링(adaptive post-filtering)을 수행한다[6].

부호기로부터 전송된 프레임에 대한 정보들을 복호화한 후에 코드북 look-up 방법으로 여기 신호를 구한다. 이 여기 신호는 합성필터 $1/A(z)$ 의 입력 신호가 되며 장기 필터의 상태를 새로이 하기 위해 사용되고 복호기에는 부호기에 없는 적응 spectral 후단필터(adaptive spectral postfilter)가 사용되는데 이는 ADPCM 음성 부호화기에서 합성음의 인지 음질을 향상시킬 목적으로 제안된 것으로 복호기에서도 같은 목적으로 사용되고 전달함수는 다음과 같다[7][8].

$$\hat{H}(z) = \frac{1 - \sum_{i=1}^{N_p} \eta_i z^{-i}}{1 - \sum_{i=1}^{N_p} \alpha_i \nu z^{-i}} \dots\dots\dots (5)$$

여기서 α_i 는 보간된 예측계수이고 ν 는 0.8의 값을 갖는다. η_i 는 식(5)의 분모에 대응하는 all-pole 필터의 임펄스 응답된 윈도우 자기 상관값 ($\{R(i), i = 0, \dots, N_p\}$)으로부터 구해진다. Levinson recursion을 거쳐 윈도우된 자기 상관값이 적응 spectral 후단 필터의 분모에 있는 계수 $\{\eta_i, i = 1, \dots, N_p\}$ 로 결정된다. 적응 spectral 후단 필터의 응답 특성을 보정해 주기 위해서 일차 필터 $\hat{H}(z)$ 가 적응 후단 필터의 일부로 사용되는 전달함수는 식(6)과 같다.

$$\hat{H}(z) = 1 - uz^{-1} \dots\dots\dots (6)$$

식(6)에서 u 값은 0.4를 사용한다. 한편 후단 필터의 출력은 원래보다 커지므로 적응 spectral 후단 필터의 입력과 출력의 전력비가 1이 되도록 하기 위해서 스케일링이 필요하다. 적응 spectral 후단 필터의 입력을 $\hat{s}(n)$ 이라고 하고 출력을 $\hat{s}_p(n)$ 이라고 하면 이들 스케일 인자(gain scale factor) S_{scale} 은

$$S_{scale} = \sqrt{\sum_{n=0}^{N-1} \hat{s}^2(n) / \sum_{n=0}^{N-1} \hat{s}_p^2(n)} \dots\dots\dots (7)$$

식(7)과 같이 입력 신호 에너지와 출력 신호 에너지의 비의 제곱근(square root)으로 정의되는데 이 S_{scale} 이 스케일링에 직접 사용되지 않고

$$S'_{scale}(n) = 0.9875 \cdot S'_{scale}(n-1) + 0.125 \cdot S_{scale} \dots\dots\dots (8)$$

식(8)과 같이 1차 저역 필터를 거친 후에 $\hat{s}_p(n)$ 에 곱해진다. 이와 같이 합성 필터와 적응 spectral 후단 필터를 거친 후 $S'_{scale}(n)$ 에 의해서 스케일링된 여기 신호가 최종적으로 복원된 음성 신호가 된다[9].

Ⅲ. 개선된 VSELP 여기 코드북 구조

개선된 VSELP와 여기된 다른 보코더 사이에는 차이가 있는데 그 차이점은 VSELP에는 두 개의 여기 코드북을 가지고있으며 두 개의 코드북은 각

각 M 차의 기본 벡터를 가지고 있다. 그리고 각각의 기저벡터는 N 개의 성분을 가지고 있다. 제안된 VSELP 부호화기는 식(9)를 이용하여 코드벡터를 M 개의 기저벡터 선형 결합으로 나타낸다.

$$u_j(n) = \sum_{m=1}^M \theta_{jm} v_m(n) \dots\dots\dots (9)$$

여기서, m 번째 기저벡터에 곱해지는 이득은 코드워드 m 번째 비트 값에 의해 결정되므로 M 개의 기저벡터만 가지면 2^M 개의 코드벡터를 생성해 낼 수 있다. 선형 결합은 θ_{jm} 에 의해 이루어지고 θ_{jm} 은 코드워드 j의 m 번째 비트가 1일 때는 +1의 값을 갖고, m 번째 비트가 0일 때는 -1의 값을 갖는다. j가 3 = 00000112 이고 M = 7 이라 하면

$$u_3 = -v_7(n) - v_6(n) - v_5(n) - v_4(n) - v_3(n) + v_2(n) + v_1(n) \dots\dots\dots (10)$$

이 된다. 이와 같은 구조로 되어 있어 VSELP는 1 비트 에러가 있을 때 M 개의 기저벡터 중 한 개의 기저벡터에만 영향을 미치기 때문에 1비트에

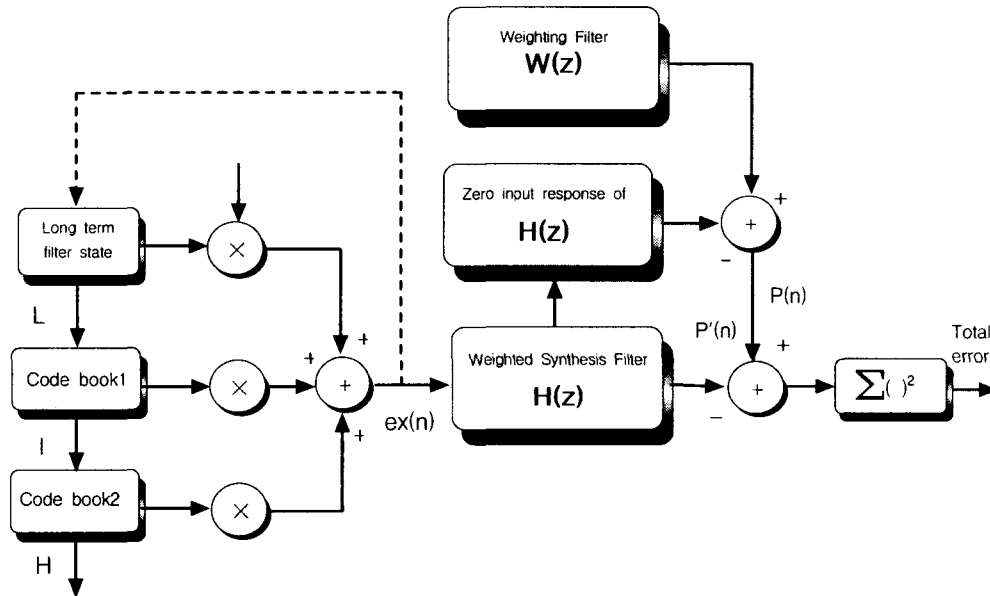


그림 1. VSELP 코드북 구조
Fig. 1 VSELP Codebook Structure

러로 전혀 다른 코드벡터를 이끌어내는 CELP 부호기에 비해 채널 오차에 강하다.

IV. 기저벡터의 최적화

훈련 데이터베이스(training sequence)가 K 개의 입력 음성 벡터($\{S_i; i=1, \dots, K\}$)로 구성되고 각 입력 음성 벡터의 크기는 부프레임의 크기와 일치한다고 하자. 그리고, 부프레임에서 i 번째 입력 음성 벡터를 부호화 했을 때의 각 신호 및 파라미터를 다음과 같이 나타낸다.

- P_i : 가장 필터링된 입력 음성 벡터
- L_i : 적응 코드북 탐색에서 선택된 피치 주기
- I_i : 첫 번째 코드북에서 선택된 코드워드
- H_i : 두 번째 코드북 탐색에서 선택된 코드워드
- b_{L_i} : L_i 에 의해 결정되는 장기 예측 필터 상태 벡터
- u_1, I_i : 코드워드 I_i 에 대응되는 코드벡터
- u_2, H_i : 코드워드 H_i 에 대응되는 코드벡터
- β_i : b_{L_i} 에 대한 여기 이득
- $\gamma_{1,i}$: $u_1 I_i$ 에 대한 여기 이득
- $\gamma_{2,i}$: $u_2 H_i$ 에 대한 여기 이득
- ex_i : 여기 신호 벡터
- e_i : 가중 오차 벡터

기저 벡터를 최적화하는 기준이 되는 평균 가중 오차는 다음과 같다.

$$E = \frac{1}{K} \sum_{i=1}^K \|e_i\|_2 \quad \dots\dots\dots (11)$$

가중 필터링된 입력 음성 벡터와 가중 합성필터링된 여기 신호 벡터의 차로 정의되는 가중오차 e_i 는 다음과 같다.

$$\begin{aligned} e_i &= p_i - H_i ex_i \\ &= p_i - H_i (\beta_i b_{L_i} + \gamma_{1,i} u_{1,i} + \gamma_{2,i} u_{2,i}) \\ &= p_i - H_i (\beta_i b_{L_i} + \gamma_{1,i} \sum_{m=1}^M \theta_{L_i m} v_{1,m}^T + \gamma_{2,i} \sum_{m=1}^M \theta_{H_i m} v_{2,m}^T) \end{aligned} \quad \dots\dots\dots (12)$$

여기서 H_i 는 i 번째 부프레임에서 가중 합성 필

터 $H_i(z)$ 의 임펄스 응답 행렬로 lower triangular Toeplitz 행렬이다.

평균가중오차를 최소화하는 최적의 기저 벡터를 얻기 위해서 평균가중오차를 다음 식과 같이 각각의 기저 벡터에 대해 편미분한 후 그 결과를 0으로 하는 값들을 구한다.

$$\frac{\partial E}{\partial v_{1,k}^T} = \frac{1}{K} \sum_{i=1}^K \frac{\partial \|e_i\|^2}{\partial v_{1,k}^T} = 0 \quad \dots\dots\dots (13)$$

$$\frac{\partial \|e_i\|^2}{\partial v_{1,k}^T} = 2 e_i^T \frac{\partial c_i}{\partial v_{1,k}^T} = -2 \gamma_{1,i} \theta_{L_i k} c_i^T H_i \quad \dots\dots\dots (14)$$

$$\frac{\partial \|e_i\|^2}{\partial v_{2,k}^T} = 2 c_i^T \frac{\partial c_i}{\partial v_{2,k}^T} = -2 \gamma_{2,i} \theta_{H_i k} c_i^T H_i \quad \dots\dots\dots (15)$$

$$\begin{aligned} v_{j,k} &= [v_{j,k}(0) \dots v_{j,k}(N-1)] \\ j &= 1, 2 \quad k = 1, \dots, 7 \end{aligned} \quad \dots\dots\dots (16)$$

식(14)와 식(15)를 사용하여 식(13)을 풀면

$$\left[\sum_{i=1}^K A_i \right] \hat{v} = \sum_{i=1}^K B_i (H_i^T)^{-1} \quad \dots\dots\dots (17)$$

식(17)이 얻어진다. 이 식을 만족하는 최적의 기저 벡터(\hat{v})를 얻기 위해서는 코드북을 훈련할 때 마다 $2 \times M \times N$ (코드북 개수 \times 코드북당 기저 벡터의 개수 \times 부프레임당 샘플 개수)개의 선형 연립 방정식을 풀어서 총 인지 가중 오차 에너지를 최소로 하는 기저벡터 코드북을 구할 수 있다.

V. 코드북 훈련과정

VSELP 부호화기는 부호화 과정 특성상, 가장 필터링된 입력 음성이 현재의 여기 코드북에 대해 의존되어있고 부호화기의 부프레임이 독립적이지 않고 이전 부프레임의 영향을 받게 되므로 코드북 훈련은 다음과 같이 폐루프(closed-loop) 방식으로 이루어진다[10].

비트율 감소를 위해서는 벡터양자화가 효과적이고 LSP(Line Spectrum Pair) 주파수를 사용하여 비트수를 줄이는 것이 적합하다. 기존의 연구 결과에

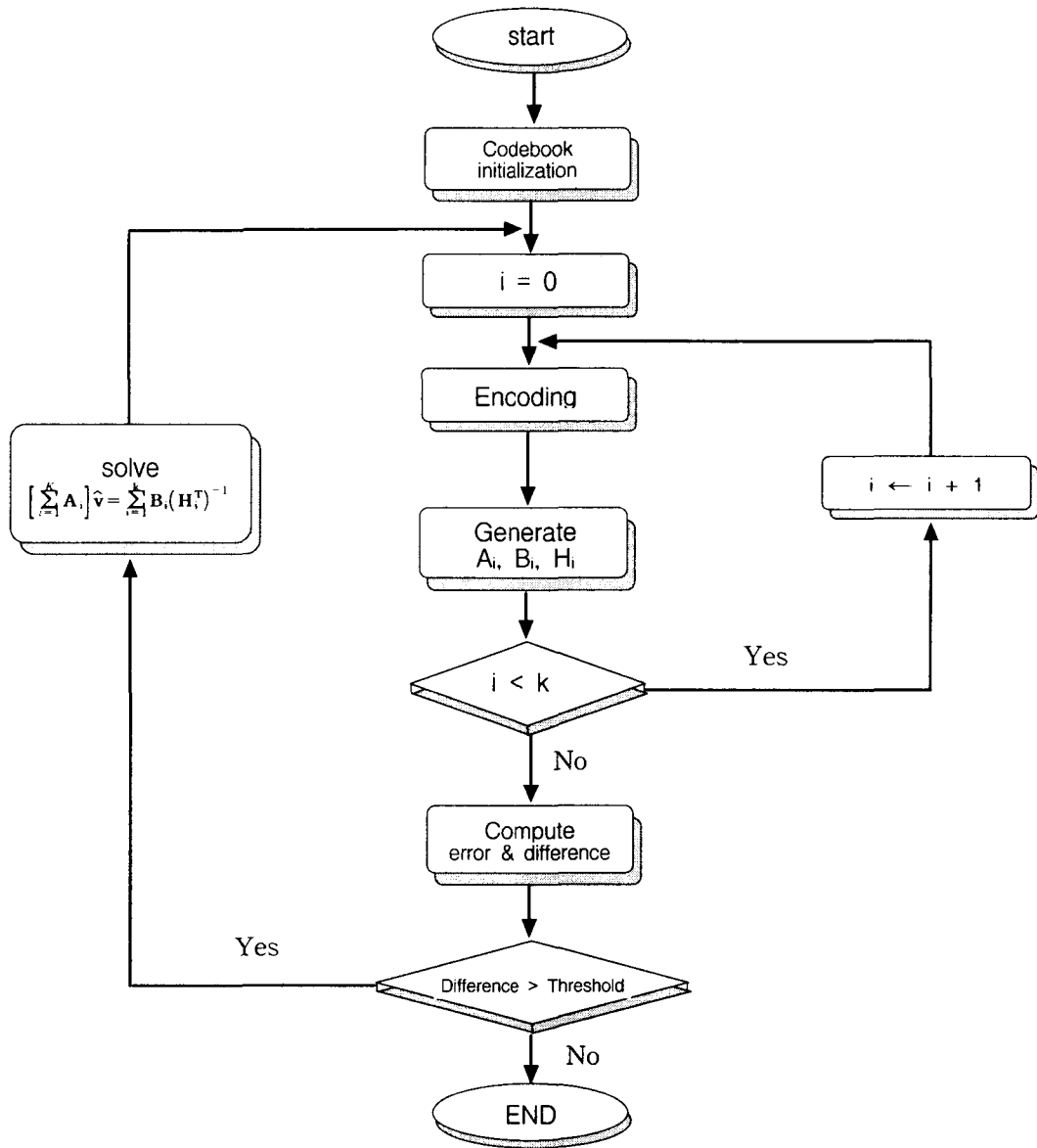


그림 2. 페루프 방식의 코드북 디자인
 Fig. 2 Codebook Design by Closed loop

의하면 20ms 프레임 길이에서 벡터 양자화를 사용할 경우에는 24~30 비트 정도에서 만족할만한 성능을 갖을 수 있다.

VSELP 부호화기에서 적응 코드북으로부터 장구

간 예측을 할 때, 피치 검색은 20~146 샘플 내에서 수행되므로 연산량이 많으나 연산량을 줄이기 위해 유성음 구간에서 천천히 변화하는 피치를 이웃하는 프레임의 피치 추정치로부터 현재 프레임

의 피치를 추정하는 델타 검색을 수행한다[2][11].

음성의 준주기성을 나타내는데 있어서 CELP 부호화기는 장구간 예측기 또는 적응 코드북으로 표현할 수 있다. 낮은 전송률을 갖기 위해 4kbps 정도의 전송률에서는 프레임의 길이가 긴 경우 유성음 구간에서는 여러 개의 피치가 존재하게 되어 장구간 예측을 실행한 뒤에도 피치 성분을 완전히 제거 할 수 없다[12]. 장구간 예측을 실행한 후에도 남아있는 피치 성분을 제거하기 위해서는 통계적 코드북에서 가져온 코드 벡터를 피치 주기에 따라 재구성하면 나머지 주기 성분을 예측하는데 편리하다[12]. 이 방법을 이용하면 M 개의 기저벡터만으로 피치 주기를 다시 구성하여 통계적 코드북을 생성할 수 있어 계산량이 적어져 전송시 데이터량을 감소시킬 수 있다.

개선된 VSELP는 순차적인 코드북 탐색과정을 거치게 되는데 먼저 장기 예측 지연 계산을 하기 위해 적응 코드북을 사용한다. 장기 예측 지연(long term predictor lag) L을 결정하는데는 일반적으로 개루프방법(open loop method)과 폐루프방법(closed-loop method) 두 가지 방법이 있는데 개루프 방법은 L값을 입력 신호 또는 잔여 신호로부터 직접 구한다. 폐루프 방법은 L값을 현재의 입력 신호와 장기 필터(long term filter)로부터 얻어지는 과거 신호로부터 결정한다. 개루프방법과 폐루프방법에 있어서 성능은 후자의 것이 훨씬 우수하지만 피치 주기를 얻기 위해서 비선형식을 풀어야하는 단점이 있다. 개선된 VSELP 부호화기에서는 폐루프 방식의 단점을 해결하기 위해 장기 예측 필터의 출력이 부프레임 시작시기의 장기 예측 필터 상태에만 의존하도록 장기 예측 필터를 다음과 같은 식으로 변형한다.

$$B_n(z) = \frac{1}{1 - \beta z^{-1} \lfloor \frac{n+L}{L} \rfloor} \quad n = 0, \dots, N-1$$

..... (18)

여기서 $\lfloor \frac{n+L}{L} \rfloor$ 는 floor 함수이고 β 는 장기 예측 필터 상태에 대응하는 이득이다. 식(18)에 의해 피치 주기가 L 일 때 장기 예측 필터의 출력 b_L 은 다음과 같다.

$$b_L(n) = r(n - \lfloor \frac{n+L}{L} \rfloor L) \quad n = 0, \dots, N-1$$

..... (19)

식(19)에서 $r(n)$ 은 장기 예측 필터의 이전 상태를 나타내는 신호로 n이 0보다 작은 경우에만 존재한다. 이때 지연값(lag)은 2.5ms(400Hz)에서 18.25ms(55Hz)의 범위를 갖는다. 피치 주기 L은 20에서 146 사이의 127개가 분포하고 부프레임 크기는 $N = 40$ 이므로 $\lfloor \frac{n+L}{L} \rfloor$ 는 1 또는 2의 값만 가지게 되고 $L \geq N$ 이면

$$b_L(n) = r(n-L) \quad 0 \leq n \leq N-1 \quad \text{이고,}$$

$$L < N \text{이면}$$

$$b_L = \begin{cases} r(n-L) & 0 \leq n \leq L \\ r(n-2L) & L \leq n \leq 1 \end{cases} \quad \text{..... (20)}$$

이므로, 장기 예측 필터의 출력 $b_L(n)$ 은 부프레임 출발시에 존재하는 장기 예측 필터 상태 $r(n)$ 에만 의존하게 된다. 장기 예측 필터의 출력 b_L 에 대한 가중 합성 필터 $H(z)$ 의 정상상태 응답을 $b'_L(n)$ 이라 하면 총 가중 오차를 최소화시키는 장기 예측 지연 L은 다음 식이 최대값을 가질 때 결정되어야 한다.

$$C_L^2 / G_L \quad \text{..... (21)}$$

$$C_L = \sum_{n=0}^{N-1} b'_L(n) p(n)$$

$$G_L = \sum_{n=0}^{N-1} \{b'_L(n)\}^2$$

..... (22)

이 때 이전의 계산결과를 이용하는 순환 알고리즘을 사용하여 계산량을 줄일 수 있다. 효율적인 코드북 탐색을 위해서 $b'_L(n)$ 이 효과적으로 계산되어야 하며 다음 식과 같이 나타낸다.

$$b'_L(n) = \sum_{i=0}^n z_L(n-iL) \quad \text{..... (23)}$$

여기서

$$z_L(n) = \sum_{i=0}^{\min(n/L-1)} r(i-L) h(n-i) \quad \text{..... (24)}$$

이고 이전 피치 주기는

$$z_{L-1}(n) = \sum_{i=0}^{\min(n, L-2)} r(i-L)h(n-i) \dots\dots\dots (25)$$

로 나타낸다.

$z_L(n)$ 은 이전 피치 주기에서 구한 $z_{L-1}(n)$ 으로부터 순환적으로부터 다음과 같은 식으로 나타낼 수 있다.

$$z_L(n) = \begin{cases} r(-L)h(n) + z_{L-1}(n-1) & 1 \leq n \leq N-1 \\ r(-L)h(0) & n=0 \end{cases} \dots\dots\dots (26)$$

$p(n)$ 과 $b'_L(n)$ 의 상호상관 C_L 이 0보다 작은 경우는 피치가 없는 것으로 간주한다. 피치가 없을 때를 나타내는 경우까지 포함하여 피치 주기 L 은 128개의 값을 가지게 되고 이것을 7bit로 부호화한다.

비트율이 4.8kbps인 경우 기저 벡터 코드북이 한 개만 사용되기 때문에 부프레임에서 인지 가중된 오차 신호는 다음과 같이 나타낼 수 있다.

$$e_l = S_u - \hat{S}_u \dots\dots\dots (27)$$

VI. 실험 및 고찰

훈련 데이터베이스가 여성 음성인 경우 DB_f , 남성인 경우 DB_m 으로 나타낸다. 본 실험에서 코드북 훈련 횟수는 14회로 하였다. 코드북 디자인 반복 횟수에 따른 SEGWSNR의 변화를 도표로 나타내었고, 다음 표는 여성음과 남성음을 각각 비교한 것이다.

객관적 평가를 위해 가중분할신호대 잡음비 SEGWSNR(Segmentally Weighted SNR)를 보인다. SEGWSNR은 단지 상대적인 값만이 의미가 있으며 다음과 같은 식으로 정의 될 수 있다[13].

$$SEGWSNR = 10 \log_{10} \frac{\sum_{k=1}^K \|S_k\|^2}{\sum_{k=1}^K \|S_u - \hat{S}_u\|} \dots\dots\dots (28)$$

식(28)에서 S 는 입력 음성 벡터이고 \hat{S} 는 합성된 음성 벡터이고 S_u 와 \hat{S}_u 은 $W(z)$ 에 의해 가중 필터링된 신호이다.

주관적 평가를 위해 MOS(Mean Opinion Score) 테스트를 실시한다. 이는 15명의 평가자를 구성하여 다음의 음성 시료를 이용하여 제안된 4.8kbps와 8kbps의 전송률을 갖는 VSELP 음성부호화기로 성능을 평가하였다.

다음은 실험에 사용된 음성 시료이다.

1. 바람과 햇님이 서로 힘이 더 세다고 다투고 있을 때
2. 한 나그네가 따뜻한 외투를 입고
3. 뜨거운 햇빛을 가만히 내리 쬐니
4. 인정하지 않을 수 없었습니다

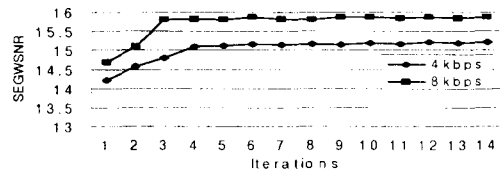


그림 3. 반복 횟수에 따른 성능향상
Fig. 3 Improvement by iterations

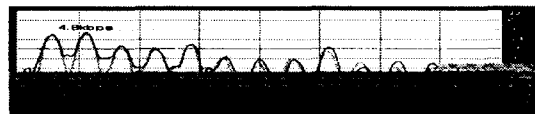


그림 4. 개선된 4.8kbps VSELP와 8kbps VSELP의 유성음 주파수 비교
Fig. 4*Comparing the improved 4.8kbps VSELP with 8kbps VSELP in Voiced Frequency

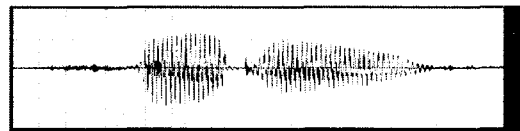


그림 5. 샘플 입력 파형
Fig. 5 Sampled input wave form

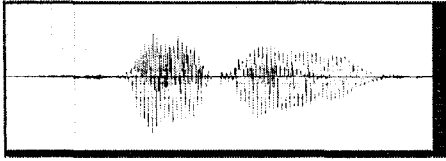


그림 6. 개선된 4.8kbps VSELP에서의 음성 파형
Fig. 6 Speech wave form in the improved 4.8kbps VSELP

표 1. 여성음에서 VSELP의 SEGWSNR
Table 1 SEGWSNR of VSELP with Woman Voice

VSELP	Woman Voice			
SAMPLE	1	2	3	4
4.8kbps	14.41	13.42	16.57	13.36
8kbps	16.16	15.94	18.36	15.90

표 2. 남성음에서 VSELP의 SEGWSNR
Table 2 SEGWSNR of VSELP with Man Voice

VSELP	Man Voice			
SAMPLE	1	2	3	4
4.8kbps	11.56	12.95	15.53	12.27
8kbps	13.64	15.5	17.73	14.36

표 3. 여성음 샘플 데이터의 MOS 평가
Table 3 MOS Test of Woman Voice

WOMAN	MOS			
SAMPLE	1	2	3	4
4.8kbps	3.45	3.36	3.52	3.41
8kbps	3.52	3.34	3.58	3.42

표 4. 남성음 샘플의 MOS 평가
Table 4 MOS Test of Man Voice

MAN	MOS			
SAMPLE	1	2	3	4
4.8kbps	3.36	3.38	3.41	3.32
8kbps	3.42	3.40	3.38	3.36

VII. 결 론

본 논문에서 우리는 제안된 4.8kbps VSELP 알고리즘이 객관적인 평가 방법인 SEGWSNR에서 8kbps VSELP 보다 2.18 dB 낮고, 주관적인 평가 방법인 MOS 평가에서 각각 평균 3.4와 3.42의 값을 갖는다는 것을 알 수 있다. 여기서 주관적인 평가인 MOS결과에서 여성음의 경우 음성 시료 2에서 남성음의 경우 음성 시료 3에서 제안된 방법이 기존의 방법 보다 우수함을 보였다. 제안된 알고리즘은 기저 벡터를 이용하여 코드북을 생성하고 코드북 탐색시에는 Gray 코드를 이용하여 코드북이 차지하는 메모리와 코드북 탐색 시간을 줄일 수 있었다. 제안된 4.8kbps 알고리즘은 유성음 구간의 분석 프레임 내부에 한 개 이상의 피치가 잔류하고 있어 장구간 예측을 실시한 후에도 주기적인 피치 성분을 완전히 제거하지 못하였고, 8kbps VSELP에 비해 계산량이 많아 프로그램이 다소 복잡하다는 문제점을 가지고 있다. 그러나 제안된 알고리즘은 전송률을 낮추어 제한된 주파수 대역에 효율적으로 사용할 수 있다는 장점을 갖는다. 추후 연구 방향은 낮은 복잡성을 갖는 저전송률 음성부호화기 개발에 초점을 두고자 한다.

VIII. 참고문헌

- [1] W.B. KLEIJN, K.K PALIWAL, *SPEECH CODING AND SYNTHESIS*, ELSEVIER, pp. 3-9, 1995.
- [2] T. Fingsheidt, T. Weishers and E. Deifs, "Implementation Aspect of the GSM Half-rate Speech Codec," EURO-SPEECH, pp.723-726, Sep. 1995.
- [3] B.S. Atal and L.R. Rabiner "Speech research directions", AT&T Technical Journal, Vol.65, issue 5, Sept./Oct., 1986.
- [4] B.S. Atal, et. al., "Advanced in Speech Coding", Kluwer Academic Pub., 1992
- [5] A.M. Kondoz, "Digital Speech," Jhon Wiley & Sons, New York, 1994.
- [6] B.S. Atal, et. al., "Speech and Audio Coding

for Wireless and Network Applications," Kluwer Academic Pub., 1994.

- [7] I. A. Gerson and M. A. Jasiuk, "Vector Sum Excited Linear Prediction(VSELP) Speech Coding at 8Kbps," Proc. ICASSP, pp. 461-464, Apr. 1990.
- [8] N. S. Jayant and V. Ramoorthy, "Adaptive Postfiltering of 16kb/s ADPCM Speech," Proc. ICASSP," pp.829-832, Apr. 1985.
- [9] EIA/TIA/IS54-A, Cellular System Dual-Mode Mobile Station-Base Station Compatibility, Standard, Electronic Industry Association, U.S.A., May 1991.
- [10] G. Davidson, M. Yong, and Gersho, "Real-Time Vector Excitation Coding of Speech at 4800BPS," Proc. ICASSP, pp.2189-2192, Apr. 1987.
- [11] A. H. Gray, Jr., R. M. Gray and J. D. Markel, "Comparison of Optional Quantizations of Speech Reflection Coefficients," IEEE Trans. on Acoust. Speech and Signal Proc., Vol. ASSP-25, pp.9-23, 1977.
- [12] S. Miki, K.Mno, H. Ohmuro, T. Moriya, "Pitch Synchronous Innovation CELP(PSI-CELP)," EUROSPEECH, pp. 261-264, Sep. 1993.
- [13] C. Wheddon and R. Linggard, *Speech and Language Processing*, Chapman and Hall, 1990.

김 형 종(Hyung-Jong Kim)

1995년 2월 조선대학교 공과대학 전자공학과 학사
1997년 2월 조선대학교 일반대학원 전자공학과 석사
1997년 ~ 현재 조선대학교 일반대학원 전자공학과 박사과정

* 주관심분야 : 음성신호처리, 통신보안



한 승 조(Seung-Jo Han)

1980년 조선대학교 전자공학과 학사

1982년 조선대학교 대학원 전자공학과 석사

1994년 충북대학교 대학원 전자계산학과 박사

1997년 조선대학교 전자·정보통신공학부 교수

1986년 6월~1987년 3월 Univ. of New Orleans 객원교수

1995년 2월~1996년 1월 Univ. of Texas 객원교수

*주관심분야 통신보안시스템설계, 네트워크 보안, ASIC설계, 음성합성