

4 kbps PSI-VSELP 음성 부호화 알고리즘

A 4 kbps PSI-VSELP Speech Coding Algorithm

최 용 수*, 강 홍 구**, 박 상 옥*, 윤 대 회*

(Yong-Soo Choi*, Hong-Goo Kang**, Sang-Wook Park*, Dae-Hee Youn*)

※본 연구는 한국학술진흥재단의 연구비 지원에 의한 것입니다.

요 약

본 논문에서는 기존의 4.8 kbps VSELP에 상응하는 음질을 갖는 4 kbps PSI-VSELP(Pitch Synchronous Innovation-Vector Sum Excited Linear Prediction) 음성 부호화기를 제안한다. 'Half-rate'는 4 kbps 내외에서 지역에 따라 다소 차이가 있으므로 기존의 half-rate 부호화기를 사용하기 위해서는 비트율을 감소시킬 필요가 생긴다. 이때, 비트율 감소에 따른 성능 저하를 최소화하기 위해서는 전송 변수가 부호화기의 성능에 미치는 영향을 고려하여 비트 할당을 결정하는 것이 바람직하다. 본 논문에서는 이러한 접근 방식으로 비트율 감소 연구를 수행한 후, 제안된 부호화기에 대한 4 kbps 비트 할당을 결정한다. VSELP 부호화기의 음질 향상을 위해서, 성능에 가장 중요한 영향을 미치는 기저 벡터는 반복 폐회로 훈련 과정을 통해 최적화하며, PSI 기법을 VSELP 부호화기에 도입한다. 제안된 음성 부호화기의 성능을 평가하기 위해 배경 잡음과 채널 에러가 없는 환경에서 실험을 하였다. 실험 결과, 제안된 4 kbps PSI-VSELP는 4.8 kbps VSELP에 비해 객관적 음질은 낮았지만, 주관적 음질은 더 높게 나타났다.

ABSTRACT

This paper proposes a 4 kbps PSI-VSELP(Pitch Synchronous Innovation-Vector Sum Excited Linear Prediction) speech coder which produces speech equivalent to that of the conventional 4.8 kbps VSELP. Since the 'half-rate' is differently defined from country to country, there may be a need to reduce the bit rate of conventional half-rate coder. To minimize the degradation of speech quality caused by bit-rate reduction, it is desirable to perform bit-allocation based on the careful consideration of the effect of various transmission parameters. This paper adopts this analytical approach for bit-allocation at 4 kbps. To improve the quality of the VSELP coder at 4 kbps, basis vectors which play the most important role in the performance, are optimized by an iterative closed-loop training process and the PSI technique is employed in the VSELP coder. To demonstrate the performance of the proposed speech coder, we performed experiments under the noiseless and error free conditions. From experimental results, even though the proposed 4 kbps PSI-VSELP coder showed lower scores in the objective measure, higher scores in subjective measure was obtained compared with those of the conventional 4.8 kbps VSELP.

I. 서 론

디지털 셀룰라와 같은 이동 통신 시스템에 대한 수요가 급증함에 따라 제한된 채널을 보다 효율적으로 사용하기 위한 연구가 필요하게 되었다. 이러한 연구는 기존의 full-rate 이동 통신 시스템의 비트율의 절반에 해당하는 half-rate에서도 우수한 성능을 유지하는 음성 부호화

기 개발이 주류를 이루고 있다. 그런데 half-rate라고 하는 비트율은 4 kbps 내외에서 지역마다 다소 차이가 있으므로, 이미 제안되어 있는 half-rate 부호화기를 사용하기 위해서는 비트율을 감소시킬 필요가 발생한다. 이때, 단순한 비트율의 감소는 심각한 성능 저하를 가져오므로 전송 변수들이 부호화기의 성능에 미치는 영향을 고려한 후, 기존 부호화기의 성능 저하를 최소로 하면서 비트율을 감소시키는 방법이 바람직하다.

낮은 비트율 음성 부호화 연구의 비약적 발전을 가져온 CELP(Code Excited Linear Prediction)[1]는 half-rate 정도의 낮은 비트율에서도 우수한 음질을 갖는 분석-합

*연세대학교 전자공학과

**AT&T Research Lab.

접수일자: 1996년 9월 13일

성(Analysis By Synthesis) 부호화 방법으로서 지난 십여 년간 성능 개선 연구와 함께 계산량을 줄이기 위한 연구가 진행되어 왔다. VSEL P(Vector Sum Excited Linear Prediction)[2-4]는 CELP형 부호화기의 하나로써 기존의 CELP [1]에 비해 매우 적은 계산량을 필요로 하며, 채널 에러에 대해 강인하고 음질도 우수하여 북미 및 일본 디지털 셀룰라의 full-rate(각각 8 kbps, 6.7 kbps) 표준안으로 채택된 상황이며, 최근에는 GSM(Global Systems for Mobile Communication)의 half-rate 표준안으로 채택되었다[5].

본 논문에서는 이미 제안되어 있는 4.8 kbps VSEL P[4]를 기반으로 하여 비트율 감소 및 성능 향상 연구를 통해 4.8 kbps VSEL P에 상응하는 음질을 갖는 4 kbps PSI-VSEL P(Pitch Synchronous Innovation-VSEL P) 부호화기를 제안한다. 비트율 감소를 위해서 전송할 각 파라미터가 부호화기의 성능에 미치는 영향을 고려하여, 최종적인 4 kbps 비트 할당을 결정한다. 4 kbps VSEL P 부호화기의 음질 향상을 위해서 코드북을 구성하는 기저 벡터(basis vector)와 양자화 표를 한국어 음성 데이터 베이스를 사용하여 훈련 과정을 통해 설계한다. 또한 피치동기화(Pitch Synchronous Innovation)[6] 기법을 VSEL P에 도입하여 주관적 음질을 크게 향상시킬 수 있는 방법을 제안한다.

제안된 음성 부호화기의 성능을 평가하기 위해 배경 잡음과 채널 에러가 없는 환경에서 실험을 하였다. 실험 결과, 제안된 4 kbps PSI-VSEL P는 객관적 음질(SNR) 평가에서는 기존의 4.8 kbps VSEL P에 비해 낮은 점수를 보였지만 주관적 음질(MOS) 평가에서는 더 높은 점수를 얻었다.

본 논문의 구성은 다음과 같다. II 장에서는 연구 대상으로 선정한 4.8 kbps VSEL P[4] 부호화기에 대해 간단히 설명하고, III 장에서는 비트율 감소의 기본 전략에 대해 살펴본다. IV 장에서는 음질 향상을 위한 기저 벡터의 최적화 방정식을 유도한 후, 반복 폐회로 훈련 과정(iterative closed-loop training process)의 결과를 살펴 보고, 주관적 음질 향상을 위한 PSI 기법의 VSEL P 적용 방법에 대해 설명한다. V 장에서는 4 kbps PSI-VSEL P를 위한 비트 할당과 합성음에 대한 주/객관적 음질 평가 및 결과를 고찰하고, 마지막으로 VI 장에서 결론을 맺는다.

II. 기존의 4.8 kbps VSEL P 부호화기

VSEL P[2-4]는 구조화된 코드북을 사용하여 CELP의 계산량을 줄이고, 코드북 훈련 과정을 통해 음질을 향상 시킨 방법으로 채널 잡음에도 강한 우수한 CELP형 부호화기이다.

VSEL P 부호화에서는 여러 개의 기저 벡터의 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. 그림 1은 4.8 kbps

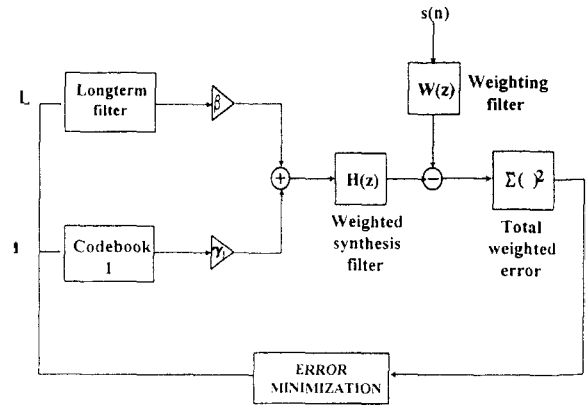


그림 1. 4.8 kbps VSEL P 부호화기.

표 1. 4.8 kbps VSEL P의 비트 할당.

Parameter	bits/7.5ms subframe	bits/30ms frame
LPC coefficients		38
frame energy		5
excitation code-1	11	44
LTP lag	7	28
gain(GS, P0)	7	28
(unused)		1
Total	25	144

VSEL P[4] 부호화기이다.

부호화기에서는 입력 신호와 합성된 신호의 인지 가중된(perceptually weighted) 오차 에너지가 최소가 되도록 표 1에 요약된 전송 파라미터들을 구한다.

복호화기에서는 부호화기에서 전송된 코드북 관련 파라미터들로부터 여기 신호를 생성한 후, 선형 예측 계수를 사용하여 구성된 합성 필터를 통과시켜 합성음을 재 생성한다. 또한 합성음의 주관적 음질을 향상시키기 위해 피치 프리필터링(pitch prefiltering)과 적응 포스트필터링(adaptive postfiltering)을 행한다[2-4].

III. 비트율 감소의 기본 전략

A. 선형 예측 계수

VSEL P는 선형 예측(LPC) 계수 전송을 위해 반사 계수(reflection coefficient)[7]를 스칼라 양자화한다. 그러나, 비트율을 낮추기 위해서는 스칼라 양자화(scalar quantization: SQ)보다 벡터 양자화(vector quantization: VQ)가 효과적이며, 반사 계수보다는 LSP(line spectrum pair) 주파수를 사용하는 것이 비트 수를 줄이는 데 적합하다[8-10]. 기존의 연구 결과에 의하면 20 ms 프레임 길이에서 스칼라 양자화기를 사용하여 LSP 계수를 양자화할 경우에는 32~40비트, 벡터 양자화기를 사용할 경우에는 24~30비트 정도에서 만족할만한 성능을 유지한다고 알려져 있다 [9][10]. 벡터 양자화의 경우에는 10 차의 LSP 주파수를

하나의 코드 벡터로 하여 양자화 하면 계산량이 너무 많으므로 3~4 개를 하나의 코드 벡터로 사용하는 분리 벡터 양자화(split vector quantization: SVQ)[9] 방법이나 다단벡터 양자화(multi-stage vector quantization: MSVQ)[10] 방법이 주로 이용되고 있다.

본 논문에서는 벡터 양자화 방법 중에 비교적 구현이 용이하고, 연산량이 적은 SVQ를 사용한다. SVQ 방법 중에서 10 차의 LSP 벡터를 묶는 방법에 따라 성능에 차이가 있다. 본 연구에서는 비교적 성능이 우수하다고 알려져 있는 낮은 주파수부터 차례로 4, 3, 3 개씩 묶는 4-3-3 SVQ[11]를 사용한다.

B. 코드북 이득

VSELP에서는 적은 비트에서도 효과적인 양자화를 위해 적응 코드북 이득과 통계적 코드북 이득을 바로 양자화하지 않고 새로운 변수로 변환하여 벡터 양자화를 한다. 본 논문에서는 음성 데이터 베이스를 부호화하여 구한 코드북 이득 변수들을 이득 양자화에 할당된 비트 수에 따라 LBG 알고리즘을 사용하여 훈련 과정을 통해 이득 양자화 코드북을 설계하고, 비트 변화에 따른 성능을 평가한다. 그림 2는 기저 벡터 수와 코드북 이득에 할당된 비트를 변화시켜가며 얻은 결과이다.

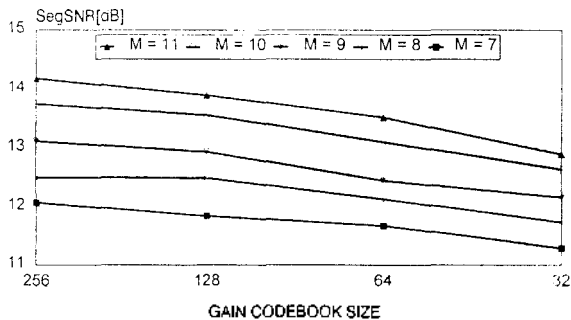


그림 2 이득 코드북 크기에 따른 segSNR.

C. 피치

VSELP 부호화기에서 적응 코드북으로부터 장구간 예측을 할 때, 피치 검색은 20~146 샘플 내에서 수행되므로 연산량이 매우 많다. 일반적으로 유성음 구간에서 피치는 천천히 변화하므로 이웃하는 프레임의 피치 추정치로부터 현재 프레임의 피치를 추정하는 '델타 검색(delta search)'[5][13]을 수행하면 연산량을 줄일 수 있다. 이러한 유성음 구간에서의 피치 특성과 유/무성음 천이 구간에서의 피치 변화를 고려하여, 본 논문에서는 첫 번째/세 번째 부프레임에서는 전체 범위에서 피치 검색을 수행하고, 두 번째/네 번째 부프레임에서는 첫 번째/세 번째 부프레임에서 찾은 피치 앞뒤로 몇 샘플 내에서 피치 검색을 수행하는 델타 검색을 수행한다. 첫 번째/세 번째 부

프레임에서 폐회로(closed-loop) 장구간 예측 피치는 개회로(open-loop) 피치 검색 방법[13]으로 구해진 초기 피치 값을 중심으로 몇 샘플 이내에서 찾음으로써 큰 성능 저하 없이 계산량을 크게 줄일 수 있다.

D. 기저 벡터

CELP에서 여기 신호를 구성하기 위해 사용되는 코드북은 재생 신호의 음질에 있어서 매우 중요한 요소이다. CELP 여기 신호는 통계적 코드북으로부터 선택된 최적 코드 벡터와 적응 코드북으로부터 선택된 장구간 코드 벡터의 합으로 이루어지지만, 장구간 여기 신호는 과거 통계적 코드북으로부터 선택된 여기 신호가 스케일링된 상태이므로, 결국 통계적 코드북의 크기나 특성은 재생 신호 음질을 결정하는 주된 요소가 된다.

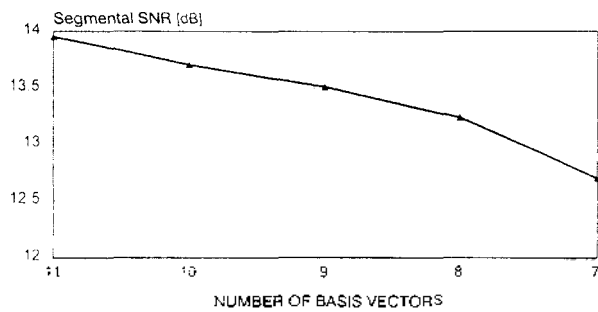


그림 3 기저 벡터 수에 따른 segSNR.

그림 3은 4.8 kbps VSELP에서 다른 모든 파라메타는 고정하고, 기저 벡터 수만 감소시켜 가며 구한 segSNR이다. 기저 벡터 수가 한 개씩 줄어들 때마다 약 0.2~0.3 dB 정도 성능 저하가 발생함을 알 수 있다.

IV. VSELP 부호화기의 성능 향상

A. 기저 벡터의 최적화

VSELP 부호화기의 음질에 가장 중요한 역할을 하는 기저 벡터는 사용 언어에 따라 성능 차이가 있으므로 사용 언어를 훈련 데이터 베이스로 하여 최적화 과정을 수행하여야 한다. 또, 기저 벡터의 수가 달라질 때마다 최적화가 필요하다. VSELP 코드북 최적화 과정은 주어진 랜덤 코드 벡터로부터 기저 벡터 코드북을 결정하는 방법으로 수행한다[4]. 이 방법은 입력 음성 신호와 합성된 음성 신호 간의 가중 오차 에너지를 최소화하는 기저 벡터 군을 구하는 것이다. 각 부프레임의 오차 에너지가 총 훈련 데이터를 통해서 누적된 전체 오차 에너지를 기저 벡터의 각 샘플들에 대하여 편미분을 취하고 이 값을 '0'으로 놓으면 각 샘플 값들을 미지수로 하는 $N \times M$ (N 개의 샘플로 구성된 M 개의 기저 벡터)의 연립 방정식을 전개할 수 있다. 이 연립 방정식을 풀면 총 인자 가중 오차에

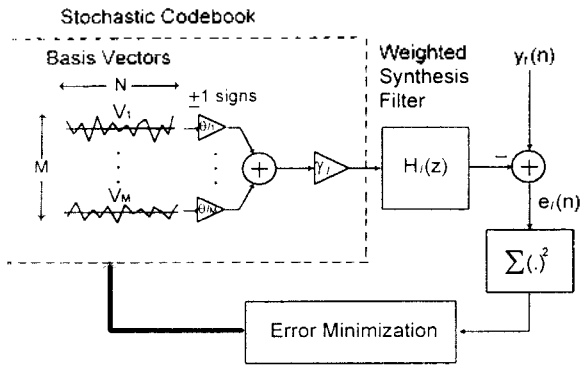


그림 4. 기저 벡터 최적화 과정.

너지를 최소화 하는 기저 벡터 코드북을 구할 수 있다. 비트율이 4.8 kbps인 경우에는 기저 벡터 코드북을 하나만 사용하므로, 부프레임에서의 인지 가중된 오차 신호는 (1)과 같은 형태로 표현된다. 기저 벡터 코드북을 \$I\$라 하고, \$y_i, H_i, \gamma_{ii}, x_{ii}\$은 각각 \$l\$ 번째 부프레임에서 장구간 예측 후의 목적 신호, 인지 가중 합성 필터의 임펄스 응답 행렬, 통계적 코드북 이득, 기저 벡터로부터 구성된 여기 신호라 할 때 인지 가중 오차 신호 \$e_i\$은 다음과 같이 표현된다[2][3].

$$e_i = y_i - \gamma_{ii} H_i x_{ii} \quad (1)$$

여기서, \$y_i = [y_i(0) \ y_i(1) \ \dots \ y_i(N-1)]^T\$, \$x_{ii} = \sum_{m=1}^M \theta_{im} v_{im}\$, \$0 \le i \le 2^M - 1\$, \$v_{im} = [v_{im}(0) \ v_{im}(1) \ \dots \ v_{im}(N-1)]^T\$, \$1 \le m \le M\$, \$e_i = [e_i(0) \ e_i(1) \ \dots \ e_i(N-1)]^T\$이며, \$N\$은 부프레임의 길이, \$M\$은 기저 벡터 갯수이다.

(1)을 다시 각 샘플 \$e_i(n)\$로 표현한 후 그 에너지 \$\xi_i\$를 구하면 다음과 같다.

$$e_i(n) = y_i(n) - \gamma_{ii} h_i(n) * x_{ii}(n) \quad (2)$$

$$\xi_i = \|e_i\|^2 = \sum_{n=0}^{N-1} e_i^2(n) \quad (3)$$

기저 벡터의 훈련 과정(training process)에 사용되는 전체 시험 데이터에 대한 누적 심리 가중 오차 에너지는 다음과 같다.

$$\Xi = \sum_{i=1}^L \xi_i = \sum_{i=1}^L \sum_{n=0}^{N-1} e_i^2(n) \quad (4)$$

여기서 \$L\$은 전체 시험 데이터의 부프레임 수이다. 최적 기저 벡터 \$v_{im}\$을 구하기 위해서 (4)를 각 기저 벡터의 샘플로 편미분하여 그 값을 '0'으로 놓는다.

$$\frac{\partial \Xi}{\partial v_i(k)} = 0 \quad (5)$$

$$\Leftrightarrow 2 \sum_i \sum_n e_i(n) \frac{\partial e_i(n)}{\partial v_i(k)} = 0, \quad 1 \leq i \leq M, \quad 0 \leq k \leq N-1 \quad (6)$$

$$\Leftrightarrow \sum_i \sum_n \gamma_{ii}^2 \sum_m \theta_{im} \theta_{ii} \sum_j h_i(n-k) h_i(n-j) v_{im}(j) = \sum_i \sum_n \gamma_{ii} \theta_{ii} y_i(n) h_i(n-k) \quad (7)$$

방정식 (7)을 풀면 최적 기저 벡터 \$v_{im}\$을 구할 수 있다. 그림 4는 기저 벡터의 최적화 과정을 나타낸다.

VSELP 부호화기는 부호화 과정의 특성상 각 부프레임 들은 상호 독립적이지 못하고 이전 부프레임의 영향을 받게 되므로 위에 기술한 최적화 과정은 그림 5와 같이 폐회로(closed loop) 형식으로 동일한 훈련 데이터를 통하여 반복 수행되어야 한다.

코드북 최적화에는 FM 라디오 뉴스 방송에서 남성 화자 3명, 여성 화자 3명이 발음한 약 5분 길이의 한국어 음성 데이터를 훈련 데이터로 사용하였다. 이러한 길이의 음성 데이터 베이스를 사용하였을 경우, 많은 계산량을 필요로 하는 실험으로서 15회 반복 과정에 소요되는 시간은 100 MHz 586급 개인용 컴퓨터를 사용하여 대략 1 개월 정도이다. 또한 코드북 이득의 양자화표를 설계하는 과정과 상호 보완적으로 수행되어야 하므로 실제로는 이보다 두 배 이상의 시간이 소요된다.

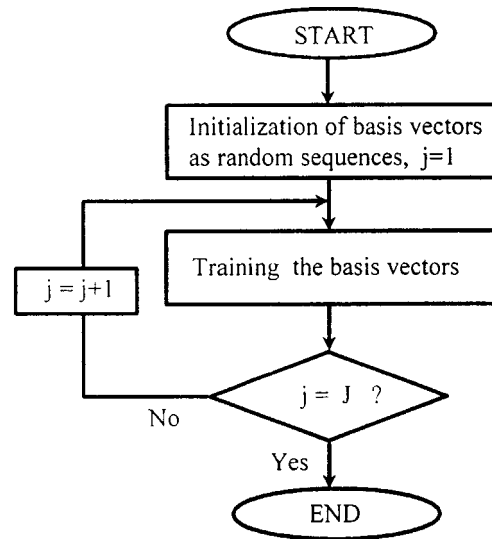


그림 5. 반복 폐회로 훈련 과정.

그림 6은 반복 훈련 과정에서 사용된 한국어 음성 데이터에 대한 wsegSNR(weighted segmental SNR)의 향상을 보여 준다. wsegSNR은 단지 상대적인 값만이 의미 있으며 다음과 같이 정의된다[14].

$$wsegSNR = \frac{1}{L} \sum_{i=1}^L 10 \log_{10} \frac{\|x_i\|^2}{\|y_i - \gamma_{ii} H_i V \theta_i\|^2} \quad (8)$$

여기서, x_l 는 l 번째 부프레임에서 인지 가중된(perceptually weighted) 입력 신호이며 L 은 전체 부프레임 수이다.

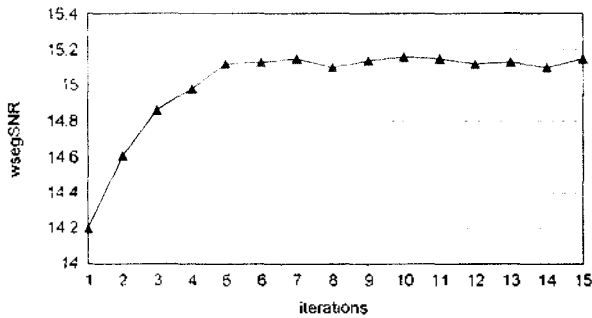


그림 6. 반복 회수에 따른 성능 향상.

그림 7과 8은 각각 초기 랜덤 기저 벡터와 훈련 과정을 거쳐 최적화된 기저 벡터의 스펙트럼으로서 훈련 과정을 통해 장구간 예측 후의 목적 신호의 통계적 특성을 반영함을 알 수 있다.

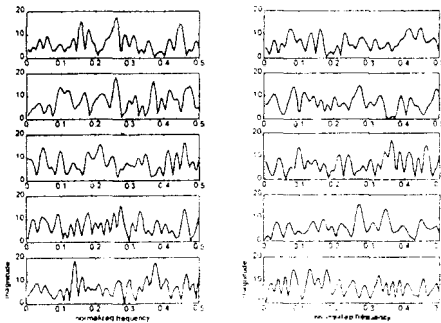


그림 7. 랜덤 기저 벡터 스펙트럼.

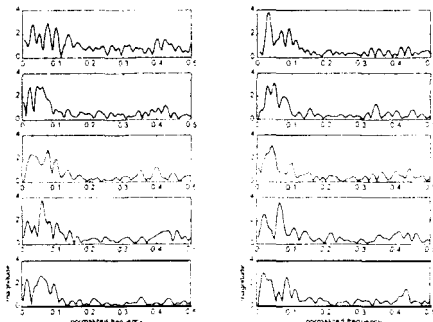


그림 8. 최적화된 기저 벡터 스펙트럼.

B. 피치동기화 VSELP

일반적인 CELP 부호화기에서 음성의 준주기성(quasi-periodicity)은 장구간 예측기 혹은 적응 코드북으로 표현된다. 그러나, 4 kbps 내외의 낮은 전송율을 유지하기 위

해 프레임 길이가 길어지는 경우, 유성음 구간에서 분석 프레임 내에 한 개 이상의 피치가 존재하게 되어 장구간 예측 후에도 추가적인 피치 성분이 완전히 제거되지 않는다[6]. 이를 극복하기 위해 통계적 코드북으로부터 취한 코드 벡터를 피치 주기에 따라 재구성하는 피치동기화(Pitch Synchronous Innovation: PSI)[6] 방법은 장구간 예측 후에도 여전히 남아 있는 주기적 성분들을 예측하는데 효과적이다.

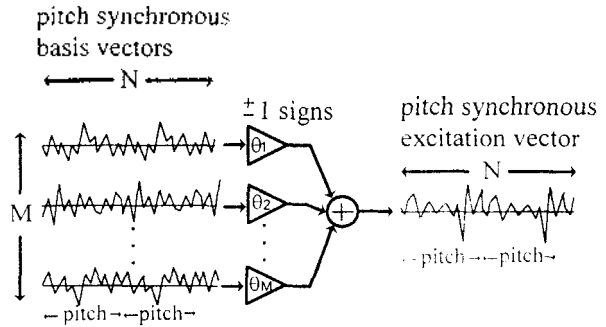


그림 9. 피치동기화 VSELP 여기 신호 구조.

피치동기화 과정을 VSELP에 적용하면 그림 9에서처럼 M 개의 기저 벡터만을 각각 피치 주기로 재구성하면 되므로 추가적인 계산량이 매우 적고, 피치 검색 정보만으로 피치동기화된 통계적 코드북을 구성할 수 있으므로 전송시 추가적인 비트가 필요 없다[6]. 이러한 피치동기화는 유성음 구간에서만 수행하는 것이 바람직하므로 이를 위한 모드 구분 방법으로 프레임 단위 개회로(open-loop) 장구간 예측 기여도를 이용한 유/무성음 판별(V/UV decision) 방법[13]이 제안되었다. 그러나 이 방법은 추가적인 계산과 비트가 필요하다는 단점이 있다. 따라서, 본 논문에서는 이러한 문제점을 해결하기 위해 프레임 에너지를 이용하여 간단하고 부가 비트가 필요 없는 유/무성음 판별 방법을 사용한다. 부프레임에서 양자화되고 보간된 프레임 에너지 $R_q'(o)$ [3]로부터 다음과 같이 부프레임 단위로 유/무성음 판별을 한다.

$$voicing = \begin{cases} 1 & \text{if } R_q'(o) > \eta \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

여기서, η 는 실험을 통해 얻어진다.

그림 10은 [13]의 방법과 (8)의 방법을 사용하여 유/무성음을 판별한 예로서 제안된 방법이 비교적 좋은 성능을 보임을 확인할 수 있다. 유/무성음 [13]의 방법에서 사용된 모드-0(무성음 또는 묵음), 모드-1(천이구간), 모드-2(유성음), 모드-3(완전 무성음)을 모드-0은 무성음으로 모드-1, 2, 3은 유성음으로 다시 분류하여 도식한 것이다.

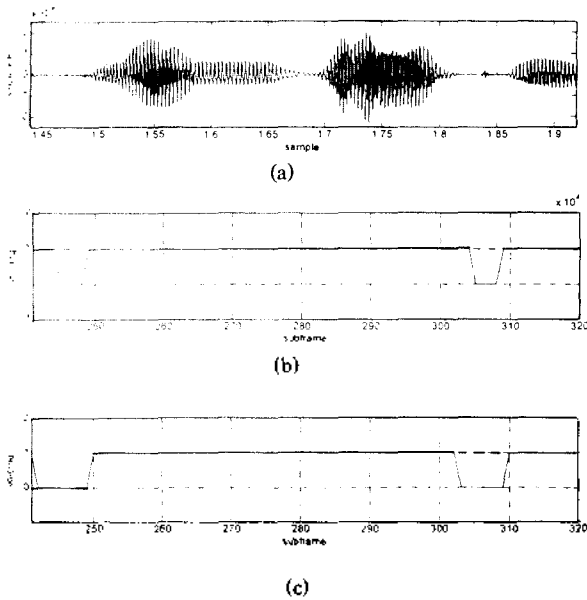


그림 10. 유/무성음 판별 예.
 a) 음성 신호.
 b) [13]의 방법을 사용한 유/무성음 판별.
 c) 본 논문에서 제안된 유/무성음 판별.

V. 실험 및 결과 고찰

III, IV 장의 실험 결과를 바탕으로 하여, 4 kbps VSELP의 비트 할당을 하였다. 표 1과 표 2를 비교할 때, 비트 감소에 따른 성능 저하가 가장 적은 LPC 계수는 LSP 주파수를 SVQ를 사용하여 10 비트를, 기저 벡터 수는 한 개를 줄여 한 프레임에 대해 4 비트를, 델타 피치 검색을 이용하여 프레임 당 10 비트를 감소시켰다. 또한 4 kbps VSELP의 성능을 최대화하기 위해서 반복적인 폐회로 훈련 과정을 거쳐 기저 벡터를 최적화하였으며, PSI 기법을 도입하여 주관적 음질을 향상시켰다. 그림 10은 최종적으로 제안된 4 kbps PSI-VSELP의 부호화기이다.

제안된 4 kbps PSI-VSELP의 성능 평가를 위해 배경 잡음과 채널 에러가 없는 환경에서 컴퓨터 모의 실험을 하였다. 객관적 평가 방법으로 segSNR(segmental SNR)을 사용하였으며, 주관적 평가 방법으로는 MOS(Mean Opinion Score)를 사용하였다. MOS 평가는 연구실에 있는 남녀 20 명을 청취 대상으로 하였고, 평가 시료로는 20 대 남녀 화자 각각 2명이 발음한 약 7초 길이의 총 8 개의 한국어 문장을 사용하였다. 표 3은 평균 segSNR과 평균 MOS를 나타내었다.

표 3에서 알 수 있듯이, 제안된 4 kbps PSI-VSELP는 4.8 kbps VSELP에 비해 비트율 감소로 인하여 객관적 음질(segSNR)은 약간 낮았지만 주관적 음질(MOS)에서는 오히려 더 높았다. 이것은 제안된 방법이 피치동기화에 의해 주관적인 음질에 중요한 요소인 피치를 잘 재생하였기 때문이다.

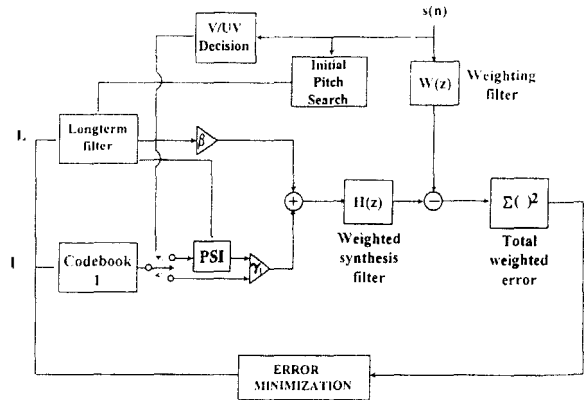


그림 11. 4 kbps PSI-VSELP 부호화기.

표 2. 4 kbps PSI-VSELP의 비트 할당.

Parameter	bits/7.5ms subframe	bits/30ms frame
LPC coefficients		28
frame energy		5
excitation code-1	10	40
LTP lag	7(2)	22
gain(GS, P0)	7	24
(unused)		1
Total		120

표 3. 주/객관적 음질 평가.

	segSNR(dB)		MOS	
	male	female	male	female
4.8 kbps VSELP	11.59	14.42	3.48	3.23
4 kbps PSI-MVSELP	10.45	12.47	3.56	3.41

VI. 결 론

본 논문은 4.8 kbps VSELP를 기준 부호화기로 선정 한 후, 4 kbps로의 비트율 감소 연구를 수행하고, 이를 바탕으로 4 kbps PSI-VSELP를 제안하였다. CELP형 부호화기의 음질에 가장 중요한 역할을 하는 코드북은 전체 인 지 가중 오차 에너지가 최소화되도록 폐회로 반복 훈련 과정을 거쳐 최적화하였다. 또, VSELP에 피치동기화 기법을 도입하여 계산량의 큰 증가 없이 음질을 향상시켰다. 제안된 4 kbps PSI-VSELP는 4.8 kbps VSELP에 비해 객관적 음질은 약간 낮았지만 주관적 음질은 오히려 높았다. 프레임 길이 변화에 따른 비트율 감소 연구가 추후 연구 과제가 될 것이다.

참 고 문 헌

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. I. A. Gerson, M. A. Jasuik, "Vector-Sum Excited Linear Prediction(VSELP) Speech Coding at 8Kbps," *IEEE Proc. Int. Conf. Acoust. Speech and Signal Proc.*, pp. 641-644, 1990.
3. EIA/TIA-PN2759 Dual-Mode Mobile Station-Base Station Compatibility Standard, Motorola, Aug., 1991.
4. I. A. Gerson, M. A. Jasuik, "Vector-Sum Excited Linear Prediction(VSELP)," *Advances in Speech Coding*, Kluwer Academic Publisher, 1991.
5. T. Fingscheidt, T. Wiechers and E. Delfs, "Implementation Aspects of the GSM Half-rate Speech Codec," *EUROSPEECH*, pp. 723-726, Sep. 1995.
6. S. Miki, K. Mano, H. Ohmuro, T. Moriya, "Pitch Synchronous Innovation CELP(PSI-CELP)," *EUROSPEECH*, pp. 261-264, Sep. 1993.
7. A. H. Gray, Jr., R. M. Gray and J. D. Markel, "Comparison of Optimal Quantizations of Speech Reflection Coefficients" *IEEE Trans. on Acoust. Speech and Signal Proc.*, Vol. ASSP-25, pp. 9-23, 1977.
8. 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.
9. 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. 1-493-496, 1994.
10. 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.
11. 연세대학교, "Half-rate(4 kbps) VOCODER의 비교 분석에 관한 연구", 한국전자통신연구소, 최종보고서, 1995.
12. T. Moriya, S. Miki, K. Mano and H. Ohmuro, "Training Method of the Excitation Codebook for CELP," *Proc. EUROSPEECH*, pp. 1155-1158, 1993.
13. I. A. Gerson, M. A. Jasuik, "Techniques for Improving the Performance of CELP-Type Speech Coders," *IEEE Journal on Selected Areas in Communications*, Vol. 10, No. 5, Jun., 1992.

▲최 용 수(Yong-Soo Choi)

한국음향학회지 vol. 14, No. 4, 1995. 참조.

▲강 흥 구(Hong-Goo Kang)

한국음향학회지 vol. 14, No. 4, 1995. 참조.

1995년 8월:연세대학교 대학원 전자공학과 졸업(공학박사)

1996년 3월~현재:미국 AT&T Murray Hill 연구소 연구원.

▲박 상 욱(Sang-Wook Park)



1996년 2월:연세대학교 공과대학 전자공학과 졸업(공학사)

1996년 3월~현재:연세대학교 대학원 전자공학과 석사과정

※주관심분야:음성 부호화, 음질 평가, 실시간 시스템

▲윤 대 희(Dae-Hee Youn)

한국음향학회지 vol. 13, No. 1E, 1994. 참조.