

한국어에 의한 EVRC LSP 코드북 설계

Design of EVRC LSP Codebooks with Korean

이진걸*
(Jingeol Lee*)

*배재대학교 정보통신공학부

(접수일자: 2001년 8월 13일; 수정일자: 2002년 1월 3일; 채택일자: 2002년 1월 10일)

음성부호화 알고리즘인 EVRC (Enhanced Variable Rate Codec)는 현재 북미 및 한국 CDMA 디지털 셀룰러 시스템에 사용되고 있다. EVRC 음성부호화기에서 음성의 주파수영역에서의 에너지 분포와 관련되어 있는 LSP (Line Spectral Pairs)값은 가중분할 벡터 양자화 (Weighted Split Vector Quantization)에 의해 코딩된다. 이러한 코딩 과정에 사용되는 LSP 코드북이 개발국 언어 혹은 영어로 설계되었음을 감안하면 한국어통화에 대해서는 한국어로 설계된 LSP 코드북에 의해 향상된 성능을 기대할 수 있다. 본 논문에서는 한국어로 EVRC의 LSP 코드북을 LBG 알고리즘을 기반으로 한 벡터 양자화기법으로 설계하였으며 이 코드북에 의한 벡터양자화 성능향상 및 그에 따른 음질향상을 각각 SD (Spectral Distortion) 및 신호대 잡음비 (SNR), SegSNR 측정으로 입증하였다.

핵심용어: EVRC, LSP, 코드북, Spectral Distortion

투고분야: 음성처리 분야 (2.2)

The EVRC (Enhanced Variable Rate Codec) is currently in service as a speech codec in digital cellular systems in North America and Korea. In the EVRC, the LSP (Line Spectral Pairs) related to energy distribution of speech signals in the frequency domain are coded by weighted split vector quantization. Considering that the LSP codebooks might be trained with the language of the develop country of the codebooks or English, it is expected that codebooks trained with Korean provide the performance improvements in the communication in Korean. In this paper, the EVRC LSP codebooks are designed with Korean adopting the LBG algorithm based vector quantization, and the performance improvement of the vector quantization and the accompanying speech quality improvement are demonstrated by spectral distortion, SNR and SegSNR measurements, respectively.

Keywords: EVRC, LSP, Codebook, Spectral distortion

ASK subject classification: Speech signal processing (2.2)

I. 서론

음성부호화 알고리즘인 EVRC (Enhanced Variable Rate Codec)는 현재 북미 및 한국 CDMA 디지털 셀룰러 시스템에 사용되고 있으며 GSM AMR 부호화기와 함께 차세대 이동통신인 IMT-2000의 음성부호화 표준안으로 알

책임저자: 이진걸 (jingeol@mail.paichai.ac.kr)
302-735 대전광역시 서구 도마2동 439-6
배재대학교 정보통신공학부
(전화: 042-520-5707; 팩스: 042-520-5773)

려져 있다. EVRC는 RCELP (Relaxation Code-Excited Linear Prediction)를 기반으로 하여 구현됨에 따라 일반 CELP (Code-Excited Linear Prediction)와 같이 부 프레임 (Subframe) 별로 피치 (Pitch)값을 구하지 않고 프레임 별로 피치 값을 구하여 선형보간 (Linear Interpolation)된 피치 값에 따라 시간축 변화된 (Time-warped) 잔차신호 (Residual signal)를 이용한다. 따라서 여기 (Excitation) 신호 및 채널코딩에 더 많은 비트를 할당함에 따라 고품질의 음성품질을 유지하면서 채널잡음에 강한 장점이

표 1. LSP 분할
Table 1. LSP parameter splits.

코드북 번호 k	LSP 값 $\Omega(m)$ 의 분할	LSP 분할 시작 인덱스 $B(k)$	코드벡터 크기 $n_{sub}(k)$	코드북 크기 $n_{code}(k)$
전송률 1				
1	$\Omega(i); i \in (1, 2)$	1	2	64
2	$\Omega(i); i \in (3, 4)$	3	2	64
3	$\Omega(i); i \in (5, 6, 7)$	5	3	512
4	$\Omega(i); i \in (8, 9, 10)$	8	3	128
전송률 1/2				
1	$\Omega(i); i \in (1, 2, 3)$	1	3	128
2	$\Omega(i); i \in (4, 5, 6)$	4	3	128
3	$\Omega(i); i \in (7, \dots, 10)$	7	4	256
전송률 1/8				
1	$\Omega(i); i \in (1, \dots, 5)$	1	5	16
2	$\Omega(i); i \in (6, \dots, 10)$	6	5	16

있다[1-3]. EVRC의 전송률은 전송률결정 알고리즘 (Rate Determination Algorithm)에 의해 전송률 1 (8550 bits/sec), 1/2 (4000 bits/sec), 1/8 (800 bits/sec)이 결정되는데 음성구간은 전송률 1 혹은 1/2로 전송되고 주변잡음구간은 전송률 1/8로 전송된다. 또한 프레임(길이는 20 msec로 입력음성의 샘플링 주파수, 8 KHz를 고려하면 한 프레임에는 160개의 음성샘플이 존재)별로 음성의 모델변수 (Model Parameter)인 LPC (Linear Predictive Coding) 계수를 추출한다. LPC 계수는 LSP (Line Spectral Pairs)값으로 변환되고 이를 가중분할 벡터 양자화 (Weighted Split Vector Quantization)에 의해 양자화되어 전송된다.

LSP 값은 음성의 주파수 영역에서의 에너지 분포와 관련되어 있으므로 언어학적인 관점에서 언어별로 고유한 특성이 있을 것이다. 따라서, LSP 벡터 양자화 과정에 사용되는 LSP 코드북이 개발국 언어 혹은 영어로 설계되었음을 감안하면 한국어통화에 대해서는 한국어로 설계된 LSP 코드북에 의해 향상된 음성품질을 기대할 수 있다. 본 논문에서는 한국어로 EVRC의 LSP 코드북을 LBG 알고리즘을 기반으로 한 벡터 양자화기법으로 설계하였으며 이 코드북에 의한 성능향상을 SD (Spectral Distortion) 및 신호대 잡음비, SegSNR 측정으로 입증하였다.

본 논문의 구성은 2장에서는 EVRC에서 LSP값의 양자화 방법을 소개하였으며 3장에서는 LBG 알고리즘을 요약하였고 4장에서는 LBG 알고리즘을 이용하여 한국어로 LSP 코드북을 설계하여 음성의 품질을 평가하였으며 5장에서 결론을 기술하였다

표 2. LSP 코드북의 개수
Table 2. Number of LSP codebooks.

전송률	코드북의 개수 k_{num}
1	4
1/2	3
1/8	2

II. EVRC에서의 LSP값의 양자화

양자화기는 가중분할 벡터 LSP 양자화기 (Weighted Split Vector LSP Quantizer)의 형태를 추하고 있으며 입출력은 다음과 같다[1].

입력: 현재 프레임의 양자화되지 않은 LSP, $\Omega(i)$
 현재 프레임의 전송률
 출력: 현재 프레임의 양자화된 LSP, $\Omega_q(i)$

주어진 전송률에 대한 양자화기 분할은 표 1에 나타내었으며 각 전송률에서의 코드북의 개수, k_{num} 은 표 2에 나타내었다.

2.1. 가중치 계산

LSP 주파수에 대한 가중치 $w(i)$ 는 다음과 같이 정의된다.

$$w(i) = \begin{cases} \frac{50}{2\pi} + 1; & \Delta\Omega(i) = 0, \quad 1 \leq i \leq 10 \\ \frac{0.5}{2\pi\Delta\Omega(i)} + 1; & \text{그외 경우,} \end{cases} \quad (1)$$

$$d_{\alpha}(i) = \begin{cases} \alpha(i+1) - \alpha(i); & i = 1, \\ \min((\alpha(i) - \alpha(i-1)), (\alpha(i+1) - \alpha(i))); & 2 \leq i \leq 9, \\ \alpha(i) - \alpha(i-1); & i = 10. \end{cases}$$

2.2. 오차 행렬 계산

각 LSP 주파수 집합 k 에 대해 오차 행렬 $e(k)$ 를 계산한다.

$$e(k, j) = \sum_{i=1}^{n_{sub}(k)} w(B(k) + i - 1)(\alpha(B(k) + i - 1) - q_{rate}(k, i, j))^2; \quad 1 \leq j \leq n_{size}(k), \quad 1 \leq k \leq k_{num}. \quad (2)$$

분할 코드북별 오차 행렬의 값이 최소가 되는 인덱스 j 에 해당하는 양자화된 LSP값이 출력된다. 여기서, i 는 각 LSP 인덱스 (Index), j 는 코드북의 인덱스, $q_{rate}(k, i, j)$ 는 코드북 값이다.

LPC 모델에서 포먼트 (Formant) 주파수 및 대역폭은 이웃한 LSP 값의 유사도에 의해 결정되므로 식 (1)과 (2)는 이러한 성질을 고려하여 포먼트 영역에 더 큰 가중치를 주었으며 낮은 포먼트 영역보다 높은 포먼트 영역에 더 큰 가중치를 주었음을 알 수 있다[4].

III. LBG 알고리즘

3.1. 벡터 양자화

N 레벨, k 차원 벡터 양자화는 각각의 입력 벡터 $x = (x_0, \dots, x_{k-1})$ 에 대해 코드북 $\hat{A} = \{y_i; i = 1, \dots, M\}$ 으로부터 출력 벡터 $\hat{x} = q(x)$ 를 출력한다. 벡터 양자화는 코드북 \hat{A} 와 입력 벡터 공간을 집합 $S_i = \{x \mid q(x) = y_i\}$ 로 할당하는 분할 $S = \{S_i; i = 1, \dots, M\}$ 로 특징지어진다[5,6].

3.2. 왜곡 측정

입력 벡터 x 를 출력 벡터 \hat{x} 로 표현하는 과정에서 발생하는 왜곡을 $d(x, \hat{x})$ 로 표시한다. 여러 가지 왜곡 측정 방법이 제안되었으나 가장 일반적인 방법은 아래 식과 같이 주어지는 오차자승왜곡 (Squared Error Distortion) 및

$$d(x, \hat{x}) = \sum_{i=0}^{k-1} |x_i - \hat{x}_i|^2 \quad (3)$$

식 (2)와 같은 형태의 가중자승왜곡 (Weighted-Squares Distortion)이 있다.

$$d(x, \hat{x}) = \sum_{i=0}^{k-1} w_i |x_i - \hat{x}_i|^2 \quad (4)$$

3.3. 알고리즘

- 1) 초기화: 주어진 훈련벡터 $\{x_j; j = 0, \dots, n-1\}$ 에 대해 N -레벨의 개수 (Codebook Size), 왜곡 문턱치 $\epsilon \geq 0$, 초기 N 레벨 코드북 값 \hat{A}_0 을 정한다. $m = 0$, $D_{-1} = \infty$ 로 초기화 한다.
- 2) 주어진 코드북 $\hat{A}_m = \{y_i; i = 1, \dots, M\}$ 을 이용하여 훈련벡터에 대해 최소왜곡분할 $P(\hat{A}_m) = \{S_i; i = 1, \dots, M\}$ 을 구한다.

$$x \in S_i \text{ if } d(x, y_i) \leq d(x, y_j) \text{ for all } j \quad (5)$$

결과로서 주어지는 평균왜곡을 구한다.

$$D_m = D(\{\hat{A}_m, P(\hat{A}_m)\}) = n^{-1} \sum_{j=0}^{n-1} \min_{y \in \hat{A}_m} d(x_j, y) \quad (6)$$

- 3) $(D_{m-1} - D_m) / D_m \leq \epsilon$ 이면 \hat{A}_m 을 최종 코드북으로 하여 계산을 중지하고 아니면 계속한다.
- 4) 분할 $P(\hat{A}_m)$ 에 대해 최적의 코드북 $\hat{x}(P(\hat{A}_m)) = \{\hat{x}(S_i); i = 1, \dots, N\}$ 을 구한다. $\hat{A}_{m+1} = \hat{x}(P(\hat{A}_m))$, $m = m + 1$ 으로 한 후 2)로 간다.

최적의 코드북은 평균왜곡이 최소가 되는 코드북 값에 해당한다. 즉, $d D_m / d \hat{x} = 0$ 를 만족하는 코드북 값은 오차자승왜곡 혹은 가중자승왜곡측정에 의해 분할된 훈련 벡터들의 중심이 된다.

3.4. 분할 (Splitting)에 의한 초기 코드북 값 \hat{A}_0 선정

M 레벨 양자화기 ($M = 2^R, R = 0, 1, \dots$)로 아래와 같이 N 레벨 양자화기의 초기 코드북 값을 구할 수 있다.

- 1) 초기화: $M = 1$ 로 하고, $\hat{A}_0(1) = \hat{x}(A)$, 즉 훈련 벡터들의 중심으로 한다.
- 2) 주어진 M 개의 벡터 $\{y_i; i = 1, \dots, M\}$ 로 구성된 초기 코드북 $\hat{A}_0(M)$ 에 대해 각각의 벡터 y_i 를 두 개의 벡터 $y_i + \epsilon, y_i - \epsilon$ 로 분할한다. ϵ 은 고정된 교란벡터 (Perturbation Vector)이다. 결과로 $2M$ 개의 벡터 $\{y_i + \epsilon, y_i - \epsilon; i = 1, \dots, M\}$ 으로 구성된 코드북 \hat{A} 를 구한다. M 을 $2M$ 으로 교체한다.

3) $M=N$ 이면 $\hat{A}_0 = \bar{A}(M)$ 으로 하여 \hat{A}_0 가 N 레벨 양자화 알고리즘에 초기 코드북이 된다. 그렇지 않으면 코드북 $\hat{A}_0(M)$ 을 생성하기 위해 $\bar{A}(M)$ 에 대해 M 레벨 양자화 알고리즘을 적용한 후 2)로 간다.

상기 분할 알고리즘을 이용하여 벡터 양자화는 훈련 벡터들의 중심인 한 레벨의 양자화로부터 시작한다. 이 한 벡터는 두 개의 벡터로 분할되어 3.3절의 알고리즘에 따라 두 레벨 양자화가 되고 이러한 방법을 반복하여 1, 2, 4, 8, ..., N 레벨 양자화가 이루어진다. 분할 알고리즘에서 교란벡터 ϵ 의 값을 정하는 방법으로 난수발생에 의해 임의로 정하거나, 교란벡터의 i 번째 원소가 훈련 벡터의 i 번째 원소의 표준편차에 비례하게 정하거나, 또는 주성분 분석 (Principal Component Analysis)에 의해 훈련 벡터의 가장 큰 고유치에 해당하는 고유 벡터에 비례하게 정할 수 있다.

IV. 실험결과

EVRC의 LSP 코드북의 양자화 성능을 SD 측정으로 평가하였다. i 번째 프레임의 SD, D_i (dB)는 다음의 식으로 정의된다.

$$D_i = \frac{1}{F_i} \int_0^{F_i} [10 \log_{10}(P_i(f)) - 10 \log_{10}(\hat{P}_i(f))]^2 df \quad (7)$$

F_i 는 샘플링 주파수이고 $P_i(f)$ 와 $\hat{P}_i(f)$ 는 다음의 식으로 정의되는 i 번째 프레임의 LPC 전력 스펙트럼이다.

$$\begin{aligned} P_i(f) &= 1 / |A_i(\exp(j2\pi f F_i))|^2 \\ \hat{P}_i(f) &= 1 / |\hat{A}_i(\exp(j2\pi f F_i))|^2 \end{aligned} \quad (8)$$

$A_i(z)$ 와 $\hat{A}_i(z)$ 는 각각 i 번째 프레임의 양자화 전과 후의 LPC 다항식이다. LPC 다항식은 다음의 식으로 정의된다.

$$A(z) = 1 + a_1 z^{-1} + \dots + a_M z^{-M} \quad (9)$$

M 은 LPC 분석의 차수로 EVRC에서는 10이며 $\{a_i\}$ 는 LPC 계수이다. 평균 SD와 함께 음질에 영향을 있는 큰 SD 값을 갖는 프레임의 비율을 측정하였다. 대역폭, 0 - 4 KHz 범위에서 SD를 측정하였으며 평균 SD, SD가 2 - 4 dB 범위에 있는 비율 및 4 dB 이상인 비율을 측정하

표 3. 원래 코드북에 의한 SD 측정값

Table 3. SD measurements using the original codebook.

	평균 SD (dB)	2-4 dB (%)	>4 dB (%)
영어음성1 (여)	1.40	8.04	2.68
영어음성2 (여)	1.38	8.80	0.80
영어음성3 (남)	1.30	2.68	0.00
영어음성4 (남)	1.29	3.77	0.00
한국어음성1 (여)	1.63	16.13	2.42
한국어음성2 (여)	1.53	10.63	5.00
한국어음성3 (남)	1.69	19.23	3.85
한국어음성4 (남)	1.80	21.52	5.70

였다[4]. SD 측정은 FFT를 이용하여 다음의 식으로 정의되는 근사식을 이용하였다.

$$D_i \approx \frac{1}{N} \sum_{j=1}^N [10 \log_{10}(P_i(F, j/M)) - 10 \log_{10}(\hat{P}_i(F, j/M))]^2 \quad (10)$$

FFT의 길이, $N=1024$ 로 하였다.

LSP 값은 음성의 주파수 영역에서의 에너지 분포와 관련되어 있으므로 언어학적인 관점에서 언어별로 고유한 특성이 있을 것이다. 이러한 점을 확인하기 위하여 EVRC의 원래 LSP 코드북의 양자화 성능을 8 문장의 시험음성 (여자영어음성 2 문장, 남자영어음성 2 문장, 여자한국어음성 2 문장, 남자한국어음성 2 문장)에 대해 SD를 측정하여 표 3에 정리하였다. 측정결과에서와 같이 EVRC의 원래 LSP 코드북은 한국어음성보다 영어음성에 대해 평균 SD 및 큰 SD 값을 갖는 프레임의 비율 측면에서 더 낮은 측정치를 보임에 따라 한국어보다 영어음성에 더 적합하게 설계되었음을 알 수 있다.

한국어로 LSP 코드북을 설계하기 위하여 8개의 FM 방송으로부터 음성 데이터베이스를 구축하였다. 총 6분 분량으로 여자음성 186초 (12 사람), 남자음성 170초 (13 사람)로 구성하였으며 시험음성은 데이터베이스에 포함시키지 않았다. FM 방송음성은 EVRC의 입력형태와 같이 샘플당 16 비트, 샘플링 주파수 8 KHz로 디지털 음성 데이터베이스를 구축하였다. 이 음성 데이터베이스에 대한 프레임별 LSP 계수는 EVRC와 같은 신호처리 방식으로 추출하였으며 표 1과 2와 같이 전송률별로 분할된 LSP 계수로 구성된 총 9개의 훈련자료 (Training Data)를 구하였다. 이러한 각각의 훈련자료에 LBG 알고리즘을 적용한 결과 (표 1에 따라 개별 코드벡터 크기 및 코드북 크기 적용)를 조합하여 한국어로 훈련된 코드북을 설계하였다. EVRC와 같은 벡터양자화 조건을 설정하기 위하여 LBG 알고리즘에 사용된 왜곡측정은 식 (2)의 가중자승왜

표 4. 평균왜곡
Table 4. Average distortions.

코드북 번호	난수발생	주성분분석
전송률 1		
1	0.000032	0.000031
2	0.000103	0.000102
3	0.000133	0.000132
4	0.000192	0.000188
전송률 1/2		
1	0.000080	0.000080
2	0.000141	0.000133
3	0.000095	0.000082
전송률 1/8		
1	0.001007	0.000992
2	0.000851	0.000848

곡을 사용하였다. 한편, 초기 코드북 관련 교란벡터의 값은 난수(Random Number)발생 및 주성분분석(Principal Component Analysis)에 의해 발생시켰으며 이러한 방법들에 의한 성능을 코드북 설계시 훈련자료에 대한 평균왜곡 및 시험음성에 대한 SD 측정으로 비교하였다. 주성분분석에 의한 교란벡터 발생은 임의의 코드벡터로 표현되는 훈련벡터의 부분집합에 대해 주성분분석을 하여 가장 큰 고유치(Eigenvalue)에 해당하는 고유벡터(Eigenvector)를 구하고 해당 훈련벡터들에 대하여 이 고유벡터 성분들을 구하여 이 값들 중 최대값과 최소값의 차이의 0.25배의 크기로 코드벡터를 중심으로 이 고유벡터 방향 양쪽으로 생성시켰다. 난수발생 및 주성분분석에 의한 코드북 설계시(동일한 왜곡문턱치 적용, $\epsilon=0.01$) 훈련자료에 대한 평균왜곡을 표 4에 정리하였다. 표 4에서와 같이 주성분분석에 의한 교란 벡터 발생이 난수에 의한 방법보다 코드북 설계시 평균왜곡이 더 적음을 알 수 있다. 이는

표 5. 한국어로 설계된 코드북에 의한 SD 측정값(난수/주성분분석)
Table 5. SD measurements using the codebook designed by korean (Random number/Principal component analysis).

	평균 SD (dB)	2-4 dB (%)	>4 dB (%)
영어음성1 (여)	1.60/1.58	23.21/23.21	0.89/0.89
영어음성2 (여)	1.45/1.38	15.20/11.20	0.00/0.00
영어음성3 (남)	1.49/1.47	12.50/10.71	0.00/0.00
영어음성4 (남)	1.64/1.63	22.64/24.53	0.00/0.00
한국어음성1 (여)	1.29/1.25	8.87/7.26	0.00/0.00
한국어음성2 (여)	1.15/1.18	6.88/8.75	0.00/0.00
한국어음성3 (남)	1.26/1.26	5.38/9.23	0.00/0.00
한국어음성4 (남)	1.29/1.29	9.49/8.86	0.63/0.63

난수에 의해 교란벡터를 구할 경우 최종 코드북 값들이 양자화 평균왜곡의 지역최소(Local Minima)에 해당하는 값으로 출력되는 반면 주성분 분석에 의한 초기 코드북 값은 난수에 의한 방법보다 양자화 왜곡의 전역최소(Global Minimum)에 해당하는 최종 코드북 값과 더 유사하므로 LBG 알고리즘에 의해 이 최종 코드북 값으로 수렴할 확률이 더 크기 때문이다. 이상과 같이 난수 및 주성분 분석에 의해 한국어로 설계된 각각의 코드북으로 동일한 시험음성에 대해 SD를 측정하여 그 결과를 표 5에 정리하였다. 표 3 및 표 5의 측정결과로부터 한국어 음성에 대해서는 한국어로 설계된 두 개의 코드북이 평균 SD 및 큰 SD 값을 갖는 프레임의 비율 측면에서 원래 코드북보다 더 낮은 측정값을 출력하는 반면 영어음성에 대해서는 원래 코드북보다 큰 SD 값을 출력함을 알 수 있다. 또한, 평균 SD 및 SD 값이 2-4 dB인 프레임의 비율의 분포는 표 3 및 표 5에서 영어음성과 한국어음성에 대해 유사한 값으로 반전되어 있으며, 특히 한국어음성에 대해서는 SD 값이 4 dB 이상인 프레임의 비율의 분포 측면에서 한국어로 설계한 코드북이 우수함을 보이고 있다. 이러한 결과는 한국어로 LSP 코드북을 설계함으로써 영어보다 한국어의 언어적 특성에 더 적합한 코드북이 생성되었음을 보여준다. 또한, 주성분분석에 의한 교란벡터 발생은 표 4에서와 같이 코드북 설계시 평균왜곡이 난수발생방법보다 작았으나 한국어음성에 대해서 SD 측정값이 난수발생방법과 비슷한 범위에 있음을 알 수 있다. 이는 두 방법에 따른 훈련자료에 대한 평균왜곡의 차가 근소하고 한국어음성이 훈련자료와 특성이 다소 다름에 기인한다. 그러나 주성분분석에 의한 교란벡터 발생은 영어음성에 대해서는 난수에 의한 방법보다 더 낮은 SD 값을 출력함을 보여준다. 상대적으로 영어음성이 한국어음성보다 한국어 훈련자료와 특성이 더 상이하므로 주성분분석에 의한 방법은 훈련자료와 특성이 다른 시험음성에 더 강인함을 알 수 있다. 이상의 측정결과로부터 주성분분석에 의한 교란벡터 발생방법이 난수에 의한 방법보다 우수함을 확인하였다.

한국어로 설계된 두 개의 코드북에 의한 음질을 시험하기 위하여 시험음성을 원래의 코드북 및 한국어로 설계된 코드북을 이용하여 EVRC 알고리즘의 코딩과 디코딩 과정을 거쳐 출력된 음성에 대해 신호대 잡음비 및 SegSNR을 측정하여 표 6에 정리하였다[7]. 표 6에서와 같이 영어음성은 원래 코드북이 한국어음성은 한국어로 설계된 코드북이 신호대 잡음비 및 SegSNR 측정면에서 대부분의 시험음성에 대해 1 dB이하의 우수한 경향이 있음을 보여

표 6. 신호대 잡음비 및 SegSNR 측정(원래 코드북/난수/주성분 분석)

Table 6. SNR and SegSNR measurements (Original codebook /Random number/Principal component analysis).

	SNR (dB)	SegSNR (dB)
영어음성1 (여)	18.39/17.92/17.91	15.96/15.89/15.79
영어음성2 (여)	17.89/17.80/17.31	15.06/15.17/14.86
영어음성3 (남)	14.85/14.87/15.17	12.33/12.36/12.28
영어음성4 (남)	17.50/16.91/16.96	15.15/14.69/14.82
한국어음성1 (여)	22.18/22.13/22.33	16.69/16.63/16.80
한국어음성2 (여)	22.11/22.60/22.00	18.03/18.09/17.93
한국어음성3 (남)	15.41/15.64/15.62	14.41/14.61/14.46
한국어음성4 (남)	14.77/14.86/14.89	12.15/12.23/12.13

주나 영어음성 3 및 한국어음성 1과 같이 표 3 및 표 5의 SD 측정값의 차가 신호대 잡음비 및 SegSNR 측정값에 반영되지 않는 경우도 있다. 또한, 한국어 음성에 대해 원래 코드북과 한국어로 설계된 코드북 (난수에 의한 교란 벡터 발생)을 사용하여 EVRC 알고리즘의 코딩과 디코딩과정을 거친 출력음성을 비교청취시험해 본 결과 1 dB 이하의 신호대 잡음비 및 SegSNR 차로 예상되듯 음질의 차이를 구별 할 수 없었다. 이는 EVRC 알고리즘 내에서 이웃한 프레임의 LSP값이 보간되어 사용되므로 한국어로 설계된 LSP 코드북에 의한 LSP 벡터양자화의 성능향상의 정도가 EVRC의 한국어 통화음질향상에 그대로 반영되지 않기 때문으로 추정된다.

V. 결론

본 논문에서는 8 문장의 시험음성 (영어음성 4 문장, 한국어음성 4 문장)에 대하여 EVRC의 원래 LSP 코드북과 FM 방송으로 구축한 한국어음성 데이터베이스에 LBG 알고리즘을 적용하여 설계한 코드북의 성능을 SD 측정, 신호대 잡음비 및 SegSNR으로 평가하였다. 시험결과 SD 측정에서 원래 코드북은 영어음성에 더 우수한 반면 한국어에 의해 설계된 코드북은 한국어에 더 우수한 성능을 보임에 따라 한국어의 언어적 특성에 더 적합한 코드북이 생성되었음을 확인하였다. 초기 코드북 관련 주성분분석에 의한 교란벡터 발생은 코드북 설계시 평균왜곡이 난수 발생방법보다 작았으나 한국어음성에 대해서 SD 측정값이 난수발생방법과 비슷한 범위에 있었다. 그러나 영어음성에 대해서는 난수발생방법보다 더 낮은 SD 값을 출

력을 확인함으로써 주성분분석에 의한 방법은 훈련자료와 특성이 다소 다른 시험 음성에 더 강인함을 알 수 있었다. 신호대 잡음비 및 SegSNR 측정으로 영어음성은 원래 코드북이 한국어음성은 한국어로 설계된 코드북이 대부분의 시험음성에 대해 1 dB이하의 우수한 경향이 있음을 확인하였다.

감사의 글

본 논문은 2000학년도 배재대학교 교내학술연구비 지원에 의하여 수행된 것임.

참고 문헌

1. TIA/EIA/IS-127, "Enhanced Variable Rate Codec, Speech Service Option 3 for Wideband Spread Spectrum Digital Systems".
2. W. B. Kleijn, P. Kroon, and D. Nahumi, "The RCELP Speech-Coding Algorithm," *European Transactions on Telecommunications*, vol. 5, pp. 573-582, Sep./Oct. 1994.
3. D. Nahumi and W. B. Kleijn, "An Improved 8 kb/s RCELP Coder," *IEEE Workshop on Speech Coding*, 1995.
4. Kuldip K. Paliwal and Bishnu S. Atal, "Efficient Vector Quantization of LPC Parameters at 24 Bits/Frame," *IEEE Transactions on Speech and Audio Processing*, vol. 1, pp. 3-14, Jan. 1993.
5. Yoseph Linde, Andres Buzo and Robert M. Gray, "An Algorithm for Vector Quantizer Design," *IEEE Transactions on Communications*, vol. Com-28, pp. 84-95, Jan. 1980.
6. Allen Gersho, Robert M. Gray, *Vector Quantization and Signal Compression*, Kluwer Academic Publishers, 1992.
7. Deller, John R., Hansen, John H. L., and Proakis, John G., *Discrete-Time Processing of Speech Signals*, John Wiley & Sons, 1993.

저자 약력

● 이진걸 (Jingeol Lee)



1981년 2월: 고려대학교 전자공학과 (공학사)
 1985년 8월: 고려대학교 전자공학과 (공학석사)
 1994년 12월: University of Florida 전기공학과 (공학박사)
 1982년 9월 ~ 1990년 12월: 국방과학연구소 연구원
 1995년 6월 ~ 1997년 2월: 삼성전자주식회사 수석연구원
 1997년 3월 ~ 현재: 배재대학교 정보통신공학부 조교수
 ※ 주관심분야: Voice CODEC, Watermark