

# W-CDMA 시스템을 위한 가변율 음성코덱 설계

정우성<sup>1</sup>, 강상원<sup>2</sup>, 성호상<sup>2</sup>, 이인성<sup>3</sup>, 김재원<sup>4</sup>, 최송인<sup>4</sup>  
<sup>1</sup>LG정보통신주, <sup>2</sup>한양대학교, <sup>3</sup>충북대학교, <sup>4</sup>한국전자통신연구원

## Design of a variable rate speech codec for the W-CDMA system

Woo-Sung Chung<sup>1</sup>, Sang-Won Kang<sup>2</sup>, Ho-Sang Sung<sup>2</sup>, In-Sung Lee<sup>3</sup>, Jae-Won Kim<sup>4</sup>, Song-In Choi<sup>4</sup>  
<sup>1</sup>LG Information & Communications Ltd., <sup>2</sup>Hanyang University, <sup>3</sup>Chungbuk University, <sup>4</sup>ETRI

swkang@selab.hanyang.ac.kr

### ABSTRACT

Recently, 8 kb/s CS-ACELP coder of G.729 is standardized by ITU-T SG15 and it has been reported that the speech quality of G.729 is better than or equal to that of 32kb/s ADPCM(G.726). However G.729 is the fixed rate speech coder, and it does not consider the property of voice activity in mutual conversation. If we use the voice activity, we can reduce the average bit rate in half without any degradations of the speech quality.

In this paper, we propose an efficient variable rate algorithm for G.729. The variable rate algorithm consists of two main subjects, the rate determination algorithm and the design of sub rate coders. For the robust VAD algorithm, we combine the energy-thresholding method, the phonetic segmentation method by integration of various feature parameters obtained through the analysis procedure, and the variable hangover period method. Through the analysis of noise features, the 1 kb/s sub rate coder is designed for coding the background noise signal. Also, we design the 4 kb/s sub rate coder for the unvoiced parts.

The performance of the variable rate algorithm is evaluated by the comparison of speech quality and average bit rate with G.729. Subjective quality test is also done by MOS test. Conclusively, it is verified that the proposed variable rate CS-ACELP coder produces the same speech quality as G.729, at the average bit rate of 4.4 kb/s.

### 1. 서 론

국제 통신 표준화기구인 ITU-T의 SG15에서는 차세대 멀티미디어 및 개인 통신 서비스를 위한 음성 부호화 방식으로 8 kb/s의 전송 속도로 유선망에서의 음성 품질을 재창하는 conjugate structure algebraic CELP(CS-ACELP)[1]를 제안하여 이를 표준안으로 채택하였다. G.729로 명명된 8 kb/s CS-ACELP는 잡음 및 비잡음 환경에서 32 kb/s ADPCM(G.726)[2]과 비교할 때 그 이상의 음질을 나타내는 것으로 알려져 있다[3]. 이후 digital

simultaneous voice over data(DSVD) 관련 분야에 응용할 목적으로 CS-ACELP의 계산량 감소를 실현한 G.729 Annex A[4]가 개발되었다. DSVD와 같이 음성 부호화기의 상대적 저 전송율이 요구되는 분야에 있어서 목음 구간 압축 기술을 이용한 가변율 알고리즘 개발은 매우 유용하다 할 수 있다. 즉, 쌍방향 대화에서 한쪽의 실제 대화시간을 반영하는 voice activity factor(VAF)가 약 0.4라는 특성을 이용하여 음성을 실제 음성이 존재하는 구간과 목음 구간으로 구분하고 각각의 경우에 대해 서로 다른 전송률을 가지는 가변율 음성 부호화기를 시스템에 사용함으로써 평균 전송률을 크게 줄일 수 있는 것이다[5]. G.729 Annex B는 이러한 목음 구간 압축 기술을 사용하여 G.729 및 G.729 Annex A 모두를 사용할 수 있도록 설계된 ITU-T의 표준안이다[6].

그러나 현재 개발된 G.729 Annex B는 discontinuous transmission 구조를 가지고 있으므로 IS-95를 기반으로 한 IMT-2000 시스템을 구축하고 여기에 CS-ACELP를 사용하기 위해서는 적합한 continuous transmission 구조의 가변율 알고리즘 개발이 필수적이다.

본 논문에서는 CS-ACELP를 위한 continuous transmission 구조의 효율적인 가변율 알고리즘을 제안하고 음질 및 평균 전송률의 측면에서 G.729와의 성능 비교를 수행한다.

본 논문의 구성은 다음과 같다. 2 장에서는 CS-ACELP가 갖는 기본 구조와 특징에 대해 살펴보고, 3 장에서는 설계된 가변율 알고리즘에 대하여 설명하며, 4 장에서는 가변율 알고리즘에 대한 실험 및 결과에 대하여 언급하고, 마지막으로 5 장에서 결론을 맺는다.

### 2. CS-ACELP 음성 부호화기

CS-ACELP는 기본적으로 분석/합성 방식의 CELP 구조 [7]를 이용하며 전체 구조는 그림 1과 같다.

입력된 신호는 8 kHz로 샘플링되며 10 ms의 프레임과 5 ms의 부프레임을 갖는다. LP 계수 추출을 위해 총 240 샘플중 40 샘플을 이후의 프레임에서 취하는 비대칭 창을 사용하므로써 15 ms의 총 알고리즘 지연시간을 갖는다.

10 차의 linear prediction(LP) 해석필터를 통해 얻어진

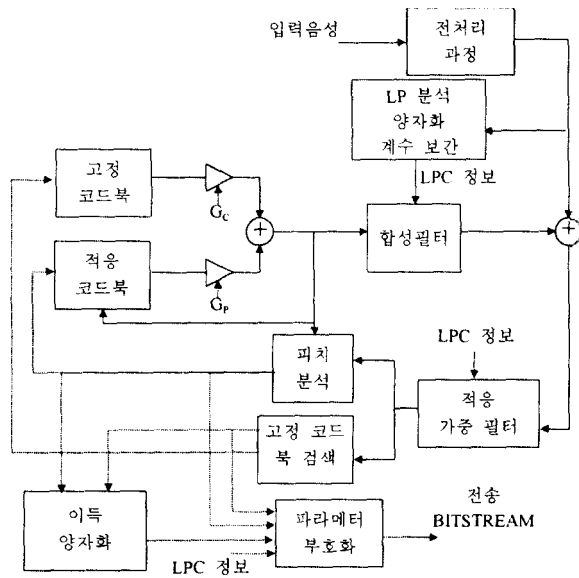


그림 1. CS-ACELP 음성 부호화기의 구조

LP 계수는 안정도 특성을 가지는 line spectrum pair(LSP) 계수[8]로 변환되며, 이 값과 4 차의 moving average(MA) 예측기를 통해 예측된 값과의 차이는 각각 128 개와 32 개의 요소값을 갖는 2 단계의 VQ 및 split VQ[9]를 통하여 총 18 비트의 코드워드로 코딩된다. 이 때 서로 다른 계수값을 가지는 2 종류의 MA 예측기가 사용되며 그 중 각각의 예측기에 의해 구해진 예측값과 LSF 계수값 사이의 weighted mean squared error(WMSE)가 최소화되는 것을 선택한다[10]. 이렇게 양자화된 LSP 계수는 두번째 부프레임을 위한 것이며 첫번째 부프레임의 LSP 계수는 이전 프레임에서 얻어진 LSP 계수와의 보간을 통해 얻어질 수 있다. LP 합성필터는 다음식에 의해 정의된다.

$$\frac{1}{\hat{A}(z)} = \frac{1}{1 - \sum_{i=1}^{10} \hat{a}_i z^{-i}} \quad (1)$$

여기서  $\hat{a}_i$ 는 양자화된 LP 계수이며,  $i=1, \dots, 10$  이다.

피치정보를 위한 적응코드북과 여기신호를 얻기 위한 고정코드북 파라미터들은 각 부프레임마다 전송된다. 최적 피치지연값을 찾기 위한 절차는 다음과 같다. 먼저 개루프를 통한 피치검색은 적응 가중필터를 통과한 음성신호에 대하여 매 프레임당 한번씩 수행된다. 각 부프레임에서는 페루프 검색을 수행하여 미리 지정된 개루프 피치지연값을 기준으로 1/3의 정밀도를 갖는 최적 피치지연값을 얻는다. 두 번째 부프레임에서는 검색의 효율을 높이기 위해 음성의 안정도 특성을 고려하여 첫번째 부프레임에서 구해진 피치지연값 주변의 제한된 범위내에서만 검색을 수행한다. 최적 피치지연값을 찾아내면 음성의 여기신호 추출을 위한 고정코드북 검색이 이루어진다. 고정코드북 검색을 위한 target 신호는 피치검색에서 사용된 target 신호  $x(m)$ 에서 적응 코드북에 의한 기여분을 제거함으로써 얻어진다.

표 1은 최적 여기신호 검출을 위해 CS-ACELP에서 채택한 고정코드북의 구조를 나타낸 것이다. 기존의 음성 부호화방식과는 달리 음성의 여기신호백터의 요소값으로서 각 부프레임당 4 개의 펄스만이 지정된 위치중에서 +1 또는 -1의 값을 갖게 되어있다. 이러한 대수적

(algebraic) 코드북 구조에서는 4 개의 검색루프를 통해 연쇄적으로 최적 펄스의 위치를 찾게 되므로 계산량의 감소와 함께 우수한 음질을 나타낼 수 있게 한다[11].

표 1. 8kb/s CS-ACELP 음성 부호화기의 고정 코드북 구조

펄스	부호	위 치
$i_0$	$s_0 : \pm 1$	$m_0 : 0, 5, 10, 15, 20, 25, 30, 35$
$i_1$	$s_1 : \pm 1$	$m_1 : 1, 6, 11, 16, 21, 26, 31, 36$
$i_2$	$s_2 : \pm 1$	$m_2 : 2, 7, 12, 17, 22, 27, 32, 37$
$i_3$	$s_3 : \pm 1$	$m_3 : 3, 8, 13, 18, 23, 28, 33, 38$ $4, 9, 14, 19, 24, 29, 34, 39$

적용 코드북과 고정코드북의 이득값 양자화에는 검색의 효율을 위하여 conjugate 구조를 갖는 2 개의 코드북을 사용한다[12]. 각각의 코드북은 8 개와 16 개의 요소값을 가지며 구해진 이득값들에 의해 미리 선택된 4 개와 8 개의 요소값들에 대해서만 검색이 이루어진다. 결국  $4 \times 8 = 32$  가지의 조합된 경우들에 대해서만 검색을 수행하므로 세 적응 및 고정코드북에 대한 최적 이득값들을 찾게 되며, 이러한 방법은 각각 8 개와 16 개의 요소값들을 가지는 코드북의 검색과 비교할 때 상당한 계산량의 감소를 얻게 한다. 고정코드북 이득의 양자화를 위해서는 4 차의 MA 예측기를 사용한다.

부호화의 마지막 단계에서는 다음 프레임의 target 신호를 구하기 위해 합성 및 가중필터의 메모리 갱신을 수행한다. CS-ACELP에서는 이러한 필터의 메모리 갱신을 위해 필요한 연산량의 감소를 위해 이전의 부호화 절차에서 이미 얻어진 중간 값들을 이용한다. 다음의 표 2는 각 파라미터에 할당된 비트를 나타낸다.

표 2. 8 kb/s CS-ACELP 음성 부호화기의 비트 할당

파라미터	부프레임 1	부프레임 2	프레임당 비트수
LSPs	L0[1], L1[7], L2[5], L3[5]		18
Adaptive-codebook delay	P1[8]	P2[5]	13
Pitch-delay parity	Po[1]		1
Fixed-codebook index	C1[13]	C2[13]	26
Fixed-codebook sign	S1[4]	S2[4]	8
Codebook gains (stage 1)	GA1[3]	GA2[3]	6
Codebook gains (stage 2)	GB1[4]	GB2[4]	8
Total			80

### 3. 가변율 알고리즘 설계

가변 전송률 방식의 음성 부호화기 설계를 위해서는 전송률 결정을 위한 VAD 알고리즘과 각 전송률에서 사용하기 위한 부음성 부호화기의 개발이 이루어져야 한다. 본 논문에서는 입력 음성 프레임을 VAD 알고리즘에 의하여 유성음, 무성음, 묵음 구간으로 구분하고 각각을 8, 4, 1 kb/s의 전송률로 부호화하도록 한다. 가변율 CS-ACELP 음성 부호화기의 구조는 그림 2에 나타나 있다.

#### 3.1 VAD 알고리즘

본 논문에서의 전송률 결정 절차는 4 단계로 구성된다. 첫번째 단계는 에너지 비교법[13]을 통한 입력 음성 프레임의 전송률 결정 과정이다. 일반적인 VAD 알고리즘

## W-CDMA 시스템을 위한 가변율 음성코덱 설계

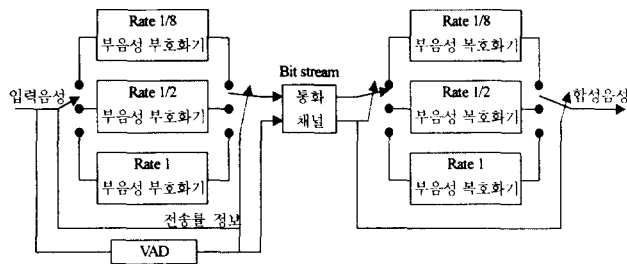


그림 2. 제안된 가변 전송률 CS-ACELP 음성 부호화기의 구조

에 사용되는 에너지 비교법은 높은 signal-to-noise ratio(SNR) 환경, 즉 주변소음에 비해 음성의 에너지가 큰 환경에 있어서 비교적 만족스러운 성능을 나타내지만, 낮은 SNR 환경하에서는 그다지 우수한 성능을 갖지 못한다. 따라서 이동통신 환경하의 주변소음 변화특성을 고려할 때 신뢰성 있는 VAD 알고리즘을 위해서는 추가적인 음성 활성화 측정법이 도입되어야 한다. 이를 위하여 본 논문에서는 입력 음성 프레임의 분석 과정에서 얻어진 몇가지 특성 파라미터들을 음성학적 구분에 따른 문턱값들과 비교하여 에너지 비교법에 의한 초기 전송률 결정을 확인하고 정정하는 절차와 낮은 SNR 환경을 갖게 되는 음성 끝단에서의 잘못된 전송률 결정을 방지하기 위한 가변 hangover 구간법을 사용하였다. 8 kb/s 와 1 kb/s 부음성 부호화기들 간의 근본적인 성능 차이로 인한 부자연성 억제하기 위한 VAD decision smoothing 과정을 통해 최종 전송률 결정을 수행하도록 하였다. 그림 3은 한 프레임에 대한 VAD 동작의 전체적인 흐름도를 나타낸 것이다.

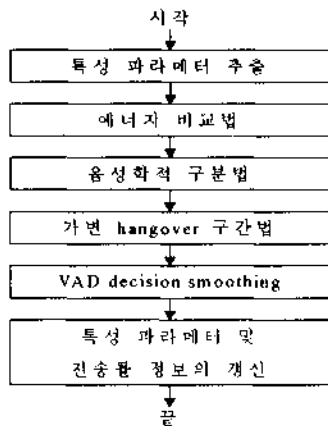
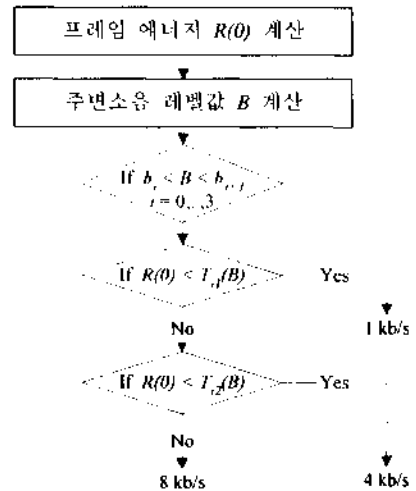


그림 3. VAD 알고리즘의 흐름도

1) **에너지 비교법** : 제안된 가변율 알고리즘에서는 초기 음성활성도 측정을 위해 프레임 에너지를 몇 개의 적응 문턱값들과 비교하여 해당 전송률을 결정하는 에너지 비교법을 사용한다. 그림 4는 에너지 비교법의 흐름도이다. 각 문턱값들은 주변소음 레벨값을 통하여 계산되며, 매 프레임마다 적응적으로 갱신된다. 주변소음 레벨값  $B$ 는 과거 프레임의 에너지와 과거 프레임의 주변소음 레벨값으로 정의된다[13].

적응 문턱값 결정에 사용되는 함수는 주변소음 레벨값  $B$ 에 따른 프레임 에너지의 통계적 분포를 기초로 설계하였다. 함수의 설계 절차는 다음과 같다.



여기서,  $B_i$ 는 전체  $B$  영역의 각 부분들  
 $T_{ij}(B)$ 는  $B_i$  영역에 적용되는 적응문턱함수

그림 4. 에너지 비교법의 알고리즘 흐름도

- 유성음, 무성음, 묵음의 각 구간들에 대하여 주변소음 레벨값  $B$ 를 변화시키면서 프레임 에너지  $R(i)$  값을 추출한다.
- 얻어진  $R(i)$ 의 분포에 근거한 least-square regression 방식을 통하여 문턱값 결정 함수를 결정한다.

주변소음 레벨값  $B$ 의 범위는 4개의 구간으로 구분되며, 위의 절차를 통하여 얻어진 2차 함수가 각 구간들에서 사용된다. 적응 문턱값 결정 함수는 다음과 같다.

- (a)  $0 \leq B < 25000$ 
    - $T_1(B) = -(1.015142 \times 10^{-4}) \times B^2 + 5.665344 \times B + 32925$
    - $T_2(B) = -(1.21 \times 10^{-4}) \times B^2 + 6.798413 \times B + 39510$
  - (b)  $25000 \leq B < 112251$ 
    - $T_1(B) = -(7.643810 \times 10^{-6}) \times B^2 + 2.726486 \times B + 47727$
    - $T_2(B) = -(9.172571 \times 10^{-6}) \times B^2 + 3.271783 \times B + 57173$
  - (c)  $112251 \leq B < 470000$ 
    - $T_1(B) = -(7.806122 \times 10^{-7}) \times B^2 + 2.579286 \times B - 22228$
    - $T_2(B) = -(9.367347 \times 10^{-7}) \times B^2 + 3.095143 \times B - 26673$
  - (d)  $470000 \leq B < 3794733$ 
    - $T_1(B) = -(1.964687 \times 10^{-7}) \times B^2 + 2.828598 \times B - 268442$
    - $T_2(B) = -(2.357625 \times 10^{-7}) \times B^2 + 3.394318 \times B - 322130$
- .....(2)

2) **음성학적 구분법** : 에너지 비교법에 의해 결정된 프레임의 전송률을 검증하기 위하여 본 논문에서는 음성학적 특성을 나타내는 파라미터로서 영교차율[14]과 개루프 피치지연 변화값을 사용한다. 개루프 피치지연 변화값은 다음과 같이 정의한다.

$$Pitch_{diff} = Pitch_{current} - Pitch_{prev} \quad (3)$$

영교차율은 무성음 구간에서 높은 값의 분포를 가지며, 현재 프레임과 과거 프레임 간의 개루프 피치지연 변화값은 유성음 구간에서 '0' 주위에 분포하는 것이 실험에 의해 확인되었다. 따라서, 이러한 특성 파라미터들을 유성음, 무성음, 묵음 구간에 대한 분석으로부터 얻어진 문턱값들과 비교하여 에너지 비교법에 의해 결정된 전송률

을 확인하고 수정할 수 있다. 그림 5는 각각의 전송률에 적용되는 제안된 음성학적 구분법의 순서도를 나타낸 것이다.

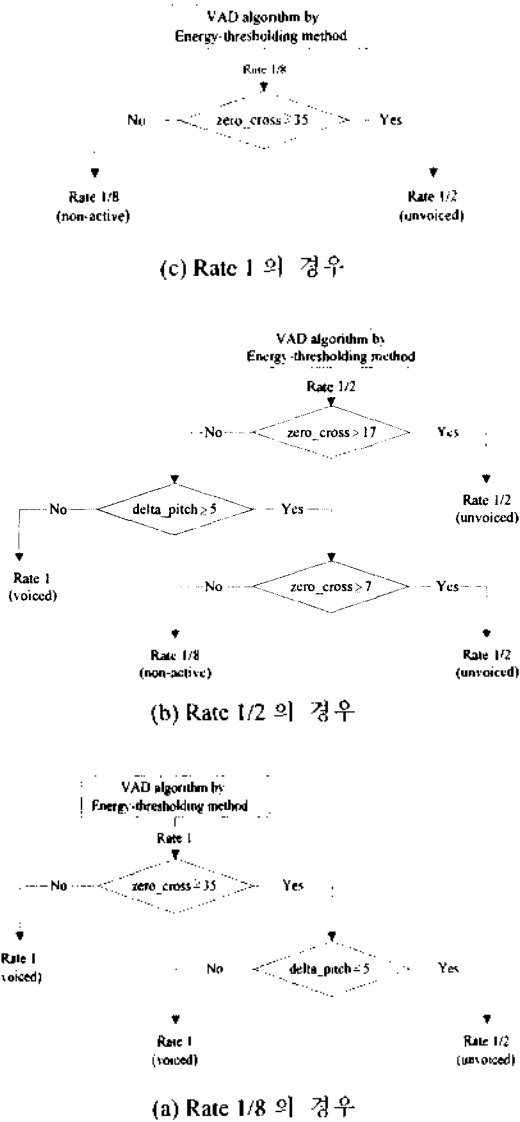


그림 5. 각 전송률에 대한 추가적인 음성학적 구분법의 알고리즘 흐름도

3) **가변 hangover 구간법**: Hangover 구간의 적용은 낮은 SNR 환경을 갖는 음성 끝단의 변환구간을 대상으로 한다. Hangover 구간법은 음성구간의 판단에서 묵음구간으로의 판단이 너무 빨리 이루어짐으로써 생길 수 있는 음질 저하를 방지하기 위한 것으로, 음성 구간에서 묵음 구간으로의 결정을 지연시키는 역할을 한다. 고정된 개수의 hangover 프레임 개수를 주변소음 정도에 관계없이 일률적으로 적용하는 경우, 주변소음이 작은 경우에는 전송률의 낭비가 생기고 주변소음이 큰 경우에는 hangover 프레임 개수가 충분치 못할 수 있으므로 hangover 구간의 길이는 SNR에 반비례하게 변화 시키는 것이 일반적이다[15]. 본 논문에서는 좀 더 효율적인 가변 hangover 프레임 수의 길정을 위해 다음의 새로운 방식을 제안하였다. 즉, SNR이 일정 수준

이하인 경우, 과거 프레임이 rate 1(8 kb/s)으로 부호화되고 현재 프레임이 rate 1/8(1 kb/s)로 결정 되더라도 음성학적 구분법에서 사용한 특성 파라미터들의 비교 절차를 통해 묵음 구간임이 판명될 때 까지 현재 프레임을 포함한 이후의 몇 프레임들을 rate 1으로 부호화 하도록 하는 것이다. 여기에 hangover 프레임의 최대 및 최소 개수에 제한을 두어 음질의 저하와 전송률의 낭비를 방지하도록 하였다. 주변 소음에 대한 음성 신호의 상대적 크기를 나타내는  $SNR_n^*$  은 다음과 같이 정의한다.

$$SNR_n^* = 10 \log_{10} \left( \frac{R(0)}{B} \right) \quad (4)$$

여기서,  $B$  는 주변소음 레벨값이다.

4) **VAD decision smoothing**: 과거 프레임이 rate 1으로 부호화되고 현재 프레임이 rate 1/8로 부호화 되도록 결정된 경우, 현재 프레임을 포함한 이후의 2 프레임들 rate 1/2(4 kb/s)로 부호화 하도록 하므로써 rate 1 및 rate 1/8 부음성 부호화기 간의 기본적인 성능 차이로 인한 단절음 효과를 억제하도록 하였다[15].

3.2 4 kb/s 부음성 부호화기

무성음 및 전송률 변환 구간은 4 kb/s 부음성 부호화기를 통하여 부호화 된다. 4 kb/s 부음성 부호화기는 부 프레임의 길이를 10 msec로 늘이고 전체 바이트 할당을 full-rate 부호화기의 절반으로 줄임으로써 설계하였다.

LSF 계수의 양자화를 위해서는 2 stage VQ 중 1st-stage VQ 만을 사용하였으며, 프레임간 예측에 사용하기 위한 2 종류의 MA 예측기와 Lloyd 알고리즘[16]을 통하여 128 개의 요소 벡터값을 가지는 코드북을 설계하였다. 피치 분석은 프레임 당 한 번 수행되며, 정수 피치 지연값을 위해 7 비트가 할당된다.

고정 코드북 검색에는 grid index 법이 사용되었다[17]. Grid index 법은 그 파라미터로서 최적 펄스 위치와 각 펄스의 부호, 그리고 최적 grid index를 사용한다. 최적 펄스 위치와 부호값은 가중 필터를 통과한 입력 음성과 합성음 간의 MSE를 최소화하는 값으로 결정된다. Focused 검색 방식은 검색 절차에 있어 계산량을 감소시키며 이때 사용하는 고정 코드북의 구조를 표 3에 나타내었다.

표 3. 4kb/s 부음성 부호화기의 고정 코드북 구조

펄스	부호	위 치
$i_0$	$s_0 : \pm 1$	$m_0 : 0, 10, 20, 35, 40, 50, 60, 70$
$i_1$	$s_1 : \pm 1$	$m_1 : 2, 12, 22, 32, 42, 52, 62, 72$
$i_2$	$s_2 : \pm 1$	$m_2 : 4, 14, 24, 34, 44, 54, 64, 74$
$i_3$	$s_3 : \pm 1$	$m_3 : 6, 16, 26, 36, 46, 56, 66, 76$ $8, 18, 28, 38, 48, 58, 68, 78$

각 펄스의 위치는 3 또는 4 비트로 부호화되며 부호에는 1 비트씩이 할당된다. 모든 펄스들의 위치는 홀수 자리를 위해 동시에 1 만큼 shift 되며, shift 여부를 나타내는 최적 grid index를 위해 1 비트를 할당하므로써 고정 코드북을 위하여 총 18 비트를 할당하였다. 표 4는 4 kb/s 부음성 부호화기의 파라미터들에 대한 비트 할당을 나타낸 것이다.

## W-CDMA 시스템을 위한 가변율 음성코덱 설계

표 4. 4 kb/s 부음성 부호화기의 비트 할당

파라미터	코드워드	프레임당 비트수
LSPs	L0[1], L1[7]	8
Adaptive-codebook delay	P[7]	7
Grid fixed-codebook index	G[1]	1
Fixed-codebook index	C[13]	13
Fixed-codebook sign	S[4]	4
Codebook gains (stage 1)	GA[3]	3
Codebook gains (stage 2)	GB[4]	4
<b>Total</b>		<b>80</b>

### 3.3 1 kb/s 부음성 부호화기

목음으로 판단된 음성 프레임은 1 kb/s 부음성 부호화기를 통해 부호화되며, 부호화를 위한 파라미터로는 LSF 계수, 여기신호 이득값, seed 값을 사용한다. LSF 파라미터를 양자화하기 위해 2 stage VQ 중 1st-stage VQ 만을 사용하여, 프레임간 예측을 위해서는 full-rate 부호화기에서 사용되는 2개의 MA 예측기 중 실험을 통해 3:1의 비율로 목음 구간에서 자주 선택되는 것으로 고정시켰다.

목음 구간의 신호는 피치 성분이 거의 없고, normalize된 여기 신호는  $-\sqrt{3}$  과  $\sqrt{3}$  사이의 uniform 분포를 갖는 random 신호로 가정할 수 있으므로, 여기 신호의 코드백터는 의사난수 코드백터로 대체할 수 있다. 이렇게 함으로써 코드북 인덱스는 전송할 필요가 없게된다. 의사난수 코드백터는 송신측과 수신측에 공동으로 갖고 있는 의사난수 발생기를 통하여 만들어지며, 의사난수 발생기에서의 초기값 발생을 위해 10비트의 패킷이 사용된다. 이 패킷은 LSF 코드워드 7비트, 과거 프레임의 여기신호 이득값 2비트, 그리고 seed 값 1비트로 이루어진다. Seed 값은 의사난수 발생기의 불규칙성을 확보하기 위해 패킷의 MSB 위치에 사용한다. 그림 6은 random 신호 발생을 위한 비트 패킷의 구성이다.

Seed	LSF [0]	LSF [1]	LSF [2]	LSF [3]	LSF [4]	LSF [5]	LSF [6]	Gprv [0]	Gprv [1]		
MSB										LSB	

그림 6. Random 신호 발생을 위한 비트 패킷의 구성

Random 여기신호의 발생은 그림 7과 같이 이루어지며, 여기신호 이득값은 다음의 식으로 계산된다.

$$G = \sqrt{\frac{\sigma_r^2}{\sigma_e^2}} \quad (5)$$

여기서  $\sigma_r^2$  과  $\sigma_e^2$  는 LP 잔여신호  $r(n)$ 과 발생된 random signal  $e(n)$ 의 분산이다. 표 5는 1 kb/s 부음성 부호화기의 비트 할당을 나타낸다.

표 5. 1 kb/s 부음성 부호화기의 비트 할당

파라미터	코드워드	프레임당 비트수
LSPs	L[7]	7
Codebook gain	G[2]	2
Seed	S[1]	1
<b>Total</b>		<b>10</b>

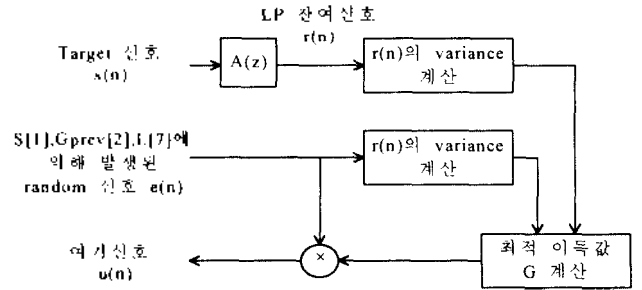


그림 7. Random 여기 신호의 발생

## 4. 실험 및 결과

4 장에서는 개발된 가변율 음성 부호화기의 성능을 평균 전송률 및 음질 측면에서 기술한다. 가변율 음성 부호화기는 기본적으로 full-rate 음성 부호화기에 비하여 음질 저하가 없고 저 전송률을 실현하는 것을 목적으로 한다.

전송률 측면에서의 성능 측정을 위해서는 이동통신 환경의 잡음을 포함한 VAF 0.47 정도의 장문들을 사용하였다. 표 6은 사용된 음성 샘플들에 대한 설명과 측정된 평균 전송률을 나타낸 것이며 평균적으로 약 4.4kb/s의 전송률로 부호화를 수행하는 것으로 나타났다. 그림 8은 원본 음성샘플의 파형과 가변율 알고리즘에 의해 결정된 전송률의 변화 예를 나타낸 것이다.

표 6. 사용된 음성 샘플 및 평균 전송률

잡음환경	잡음정도 [dB]	샘플수	VAF	가변율 CS-ACELP의 평균 전송률
자동차 내부	24	207,097 (25.9초)	0.47	4.17 kb/s
거리	26	207,136 (25.9초)	0.47	4.31 kb/s
지하철 내부	32	206,737 (25.8초)	0.47	4.71 kb/s

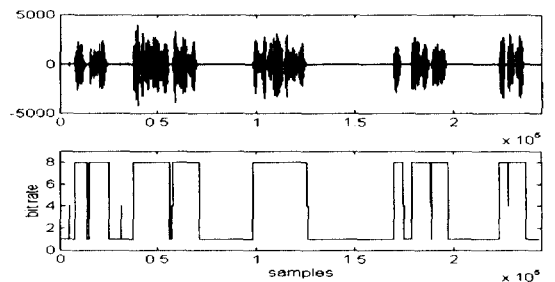


그림 8. 가변율 CS-ACELP에 의한 전송률 결정 예

다음으로, 음질 측면의 성능 평가를 위하여 채널에러 환경하에서 20명의 청취자를 대상으로 주관적 음질 평가 방법의 하나인 mean opinion score(MOS) 시험을 수행하였다. 서로 다른 bit error rate(BER) 조건하의 MOS 시험 결과는 그림 9와 같다.

그림 9에서 보는 바와 같이 에러가 없는 경우와 0.1% BER 인 경우에는 G.729가 약간 우수한 음질을 보이며, 3% BER의 경우에는 가변율 음성 부호화기가 더 우수한

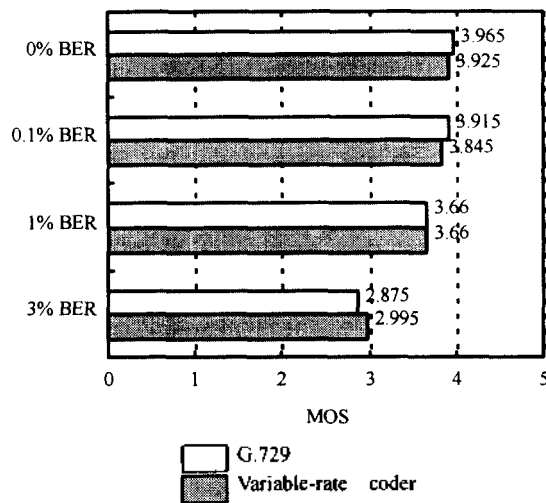


그림 9. 가변율 CS-ACELP의 MOS 시험 결과

음질을 갖는 것으로 나타났다. 그러나, 이러한 성능차는 시험 오차의 한계를 고려할 때 무시될 수 있는 수준의 것이며, 따라서 제안된 가변율 음성 부호화기는 G.729와 같은 수준의 음질을 갖는 것으로 판단된다.

### 5. 결 론

본 논문에서는 W-CDMA 시스템을 기반으로 하는 IMT-2000에서의 사용을 목적으로 CS-ACELP 음성 부호화기를 위한 효율적인 가변율 알고리즘을 설계하였다. 입력 음성 프레임은 VAD 알고리즘에 의하여 유성음, 무성음, 묵음 구간으로 구분되며 각각은 8, 4, 1 kb/s의 전송률로 부호화 된다. 4 kb/s 부음성 부호화기에는 full-rate 부호화기의 구조를 기반으로 Grid index 방식을 채택하여 무성음 및 변환 구간을 부호화하도록 하였으며, 1 kb/s 부음성 부호화기는 묵음 구간에서의 주변소음 부호화를 위해 설계 되었다. 주변소음의 영향에 대해 안정적인 VAD 알고리즘을 위하여 에너지 비교법과 음성학적 구분법, 가변 hangover 구간법을 병행하여 사용하였다. 제안된 알고리즘의 성능 평가를 위하여 몇 개의 음성 샘플에 대한 주관적 음질 평가시험 및 평균 전송률 측정을 수행한 결과, 음질저하 없이 full rate 부호화기의 절반 수준인 약 4.4kb/s의 평균 전송률로 음성 부호화를 수행하는 것으로 나타났다.

### 참 고 문 헌

- [1] ITU-T Recommendation G.729, 'Coding of speech at 8 kb/s using conjugate-structure algebraic code-excited linear-prediction (CS-ACELP),' June 1995.
- [2] CCITT Recommendation G.726, '40-, 32-, 24-, and 16-kb/s adaptive differential pulse code modulation,' 1990.
- [3] P. Usai, A. Takahashi, D. Pascal, J. Sotscheck, and G. Schröder: 'Subjective performance of the proposed ITU-T 8 kb/s speech coding standard,' *IEEE Speech Coding Workshop*, Annapolis, Sept. 1995, pp. 5-6.
- [4] R. Salami, C. Laflamme, B. Bessette, and J. P. Adoul: 'ITU-T G.729 Annex A: reduced complexity 8 kb/s CS-ACELP codec for digital simultaneous voice and data,' *IEEE Commun. Mag.*, Sept. 1997, Vol. 35, pp. 56-63.

- [5] E. Paksoy, K. Srinivasan and A. Gersho: 'Variable rate speech coding with phonetic segmentation,' in *Proc. ICASSP*, 1993, pp. 155-158.
- [6] A. Benyassine, E. Shlomot and Huan-Yu Su: 'ITU-T recommendation G.729 Annex B: a silence compression scheme for use with G.729 optimized for V.70 digital simultaneous voice and data applications,' *IEEE Commun. Mag.*, Sept. 1997, vol. 35, pp. 64-73.
- [7] B. S. Atal and M. R. Schroeder: 'Stochastic coding of speech signals at very low bit rates,' *Conf. Rec. Int. Conf. Commun.*, May 1984, pp. 1610-1613.
- [8] N. Sugamura and N. Farvardin: 'Quantizer design in LSP speech analysis-synthesis,' *IEEE J. Select. Areas in Commun.*, Feb. 1988, Vol.6, pp. 432-440.
- [9] K. K. Paliwal and B. S. Atal: 'Efficient vector quantization of LPC parameters at 24 bits/frame,' *IEEE Trans. speech and audio proc.*, Jan. 1993, Vol. 1, pp. 3-14.
- [10] A. Kataoka, T. Moriya, J. Ikedo, and S. Hayashi: 'LSP and gain quantization for the proposed ITU-T 8 kb/s speech coding standard,' *IEEE Speech Coding Workshop*, Annapolis, Sept. 1995.
- [11] C. Laflamme, J. P. Adoul, R. Salami, S. Morissette, and P. Mabilieu: '16 kb/s wideband speech coding technique based on algebraic CELP,' in *Proc. ICASSP'91*, Toronto, Ont., Canada, May 14-17, 1991.
- [12] A. Kataoka, T. Moriya, and S. Hayashi: 'Conjugate structure CELP coder for the CCITT 8 kb/s standardization candidate,' in *Proc. IEEE Workshop on Speech Coding for Telecomm.*, October 13-15, 1993.
- [13] Telecomm. Industry Association, 'Speech service option standard for wideband spread spectrum digital cellular system,' April 1994.
- [14] L. R. Rabiner and R. W. Schafer: 'Digital Processing of Speech signals,' (Prentice Hall, New Jersey, 1978), pp. 127-130.
- [15] Qualcomm Inc., 'High rate speech service option for wideband spread spectrum communications system,' Feb. 1995.
- [16] S. P. Lloyd: 'Least squares quantization in PCM,' *IEEE Trans. Inform. Theory*, 1982, Vol. IT-28, pp. 129-137.
- [17] International Telecomm. Union, 'Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3 kb/s,' March 1996.