한국음향학회지 제24권 제6호 pp. 358-364 (2005)

G.723.1 기반 비트율 scalable 음성 코덱 개발

Design of a Bitrate Scalable Speech Codec Based on G.723.1

이 준 석*, 강 상 원*, 이 강 은**, 박 동 원***
(Joonseok Lee, Sangwon Kang, Kangeun Lee, Dongwon Park*)

*한양대학교 전자컴퓨터공학부, **삼성종합기술원 디지털 연구소, ***배재대학교 IT공학부
(접수일자: 2004년 4월 11일; 수정일자:2005년 3월 7일; 채택일자: 2005년 8월 4일)

본 논문에서는 ITU-T 표준으로 채택된 G.723.1을 기본 계층으로 하고 G.723.1의 합성 에러 신호를 추가적 인 부호화 과정을 통하여 부호화 하는 비트율 scalable 코덱을 제안하였다. 그리고 제안된 scalable 음성 코덱을 ITU-T 표준 음질 측정 소프트웨어인 P.862 (PESQ)를 이용하여 성능 분석을 하였다.
제안된 비트율 scalable 코덱을 적용함으로써 G.723.1 5.3kbps와 개선 계층 6.7kbps가 함께 동작할 경우 G.723.1 5.3kbps 보다 MOS값이 0.372 향상되었으며, G.723.1 6.3kbps와 개선 계층 5.7kbps가 함께 동작할 경우 G.723.1 6.3kbps 보다 0.267 향상되었다.

핵심용어: 음성 부호화, 비트율 scalable 코덱, ACELP, LPC 양자화
투고분야: 음성처리 분야 (2.2)

In this paper, we present a bitrate scalable speech codec which uses an ITU-T G.723.1 as the baseline coder and encodes the synthesis error signal in an enhancement coder. ITU-T P.862 (PESQ) is used to evaluate the performance of the bitrate scalable coder.
Experiments show that 6.7kbps enhancement layer based on G.723.1 5.3kbps produces the increase of 0.39 in MOS and 5.7kbps enhancement layer based on G.723.1 6.3kbps produces the increase of 0.267 in MOS.

Keywords: *Speech coding, Bitrate scalable codec, ACELP, LPC quantization*
ASK subject classification: *Speech Signal Processing (2.2)*

I. 서 론

현재 패킷 교환망을 기반으로 한 응용 서비스의 범위는 매우 다양하다. 이러한 다양한 응용에는 서로 다른 전송 특성이 사용되며, 기저 네트워크에 따라 요구되는 서비스가 다르다. 또한 서비스의 종류에 따라 음성 코덱에서 요구되는 전송률, 주파 수 대역 및 복잡도가 다르다. Scalable 음성 코덱은 부호화된 전체 비트 스트림으로부터 일부의 비트 스트림 만을 받아 복호화 할 수 있는 방식이며, 음질의 향상은 개선 계층을 통해 추가되는 비트에 의해 얻어질 수 있다. 일반적으로 scalable 음성 코딩 방식은 SNR 비트율 scalability와 bandwidth 비트율scalability 두 가지 방식으로 분류 될 수 있다.

책임저자: 이 준 석 (junseok9191@hanmail.net)
경기도 안산시 사1동 1271번지 한양대학교 컴퓨터공학부
(전화: 031-400-4035; 팩스: 031-418-8476)

SNR 비트율 scalable 음성 코덱의 부호화기는 기본 계층과 개선 계층으로 구성되어 있고, 기본 계층은 기본 음질을 갖는 음성을 복원 할 수 있도록 기본 비트로 음성을 부호화한다. 개선 계층에서는 복호화기에서 기본 계층의 음질을 더 향상시킬 수 있는 음성 특성 파라미터에 추가 비트를 할당하여 부호화한다. 복호화단에서는 기본 계층과 개선 계층 에서 전송된 전체 비트 중에서 네트워크나 채널의 환경, 원하는 전송률에 따라 선택적으로 비트를 받아 음성을 복원한다.

Bandwidth 비트율 scalable 음성 코덱의 경우 협대역 코덱과 광대역 확장 코덱으로 구성되어 있으며, 부호화기로 입력된 광대역 음성은 저대역 부분 과 고대역으로 나뉘어 대역 통과된 후, 저대역 음성 은 협대역 코덱에서 기본 비트를 사용하여 부호화되 고 고대역 음성은 광대역 코덱에서 추가 비트를 사용하여 부호화 된다. 복호화단에서는 협대역 및 광대역 코덱의 부호화단에서 전송

된 전체 바트를 받아 복원할 경우 광대역의 음성을 복원할 수 있으며, 기본 비트만을 받아 복원 할 경우 협대역의 음성을 복원 할 수 있다.

본 논문에서는 기본 코덱 정보와 함께 부가적인 협대역 복원 정보를 비트 스트림으로 전송하여 기본 코덱보다 더 나은 음질을 제공하는 SNR 협대역 비트 율 scalable 음성 코덱을 설계하였다. 또한 기본 음질을 제공하기 위해 ITU-T에서 협대역 음성 코덱으로 표준화된 G.723.1[1]을 기본 계층으로 하고, 기본 계층보다 더 양질의 음성을 복원할 수 있도록 부가적인 협대역 복원 정보를 전달할 수 있는 개선 계층을 설계하였다. 그리고 설계된 개선 계층 유동 소수점 부호화기의 메모리 요구량과 계산량을 측정 하였고, 비트율 scalable 음성 코덱에 의해서 복원 되는 음질의 성능 을 P.862[5]로 평가를 하였다.

II. G.723.1 기반 비트율 scalable 음성 코덱

2.1. 비트율 scalable 음성 코덱 기본구조

본 논문에서는 협대역 비트율 scalable 음성 코덱의 총 비트율이 12kbps가 되도록 설계 하였으 며 그 구조를 그림 1에 나타내었다.

제시된 구조를 살펴보면 기본 계층인 G.723.1 로부터 합성된 음성 신호와 원 음성 신호간의 차 이값을 개선 계층으로 입력해서 합성한다. 최종적 으로 개선 계층의 출력과 기본 계층의 출력을 합하 여 향상된 음질을 출력한다. 본 논문에서 제시한 개 선 계층은 독립된 음성 코덱으로서 기본 계층의 구 조에 상관없이 삽입될 수 있으므로 광범위한 호환성 을 갖는 embedded 시스템이다. 설계된 개선 계층은 기본 계층의 모드에 따라 두 가지 전송률을 갖는다. 즉, 기본 계층이 G.723.1 5.3kbps 모드로 동작한다 면 개선 계층은 표 1과 같이 프레임당 총 197비트를 사용하 여 6.7kbps전송률로 동작하며, 기본 계층이 G.723.1 6.3kbps 모드로 동작한다면 개선 계층

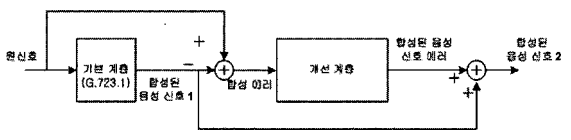

그림 1. G.723.1기반 비트율 scalable 음성 코덱의 구조도
Fig. 1. Block diagram of a bitrate scalable speech codec based on G.723.1

표 1. G.723.1 5.3kbps 기반 6.7kbps 개선 계층 비트 할당
Table 1. Bit allocation of the 6.7kbps enhancement layer based on G.723.1 5.3kbps.

파라미터	부-프레임 0	부-프레임 1	부-프레임 2	부-프레임 3	합계
LSP 양자화				27	27
적응 코드북	7	2	7	2	18
적응 및 고정 코드북 이득	12	12	12	12	48
고정 코드북 위치	20	20	20	20	80
고정 코드북 부호	6	6	6	6	24
합계: 197					

표 2. G.723.1 6.3kbps 기반 5.7kbps 개선 계층 비트 할당
Table 2. Bit allocation of the 5.7kbps enhancement layer based on G.723.1 6.3kbps.

파라미터	부-프레임 0	부-프레임 1	부-프레임 2	부-프레임 3	합계
LSP 양자화				24	24
적응 코드북	7	2	7	2	18
적응 및 고정 코드북 이득	12	12	12	12	48
고정 코드북 위치	16	16	16	16	64
고정 코드북 부호	4	4	4	4	16
합계: 170					

은 표 2와 같이 프레임당 총 170 비트를 사용하여5.7kbps 전송률로 동작한다. 개선 계층은 기본적으로 G.723.1 5.3kbps구조를 사용하였으며 LPC 모듈, 고정 코드북 모듈, 그리고 고정 코드북 이득 값 양자화 모듈을 기본 계층의 합성 에러 신호에 맞게 다른 구조를 사용하여 설계하였다.

기본 계층에서 합성된 음성 샘플은 원 음성 샘 플 보다 7.5ms (60 샘플) 지연이 발생한다. 따라서 합성 에러 신호를 구하기 위하여 두 신호간의 시간 맞춤이 필요하다. 두 신호간의 시간 맞춤 후 구해진 한 프레임의 합성 에러 신호 e 는 식 1과 같다.

$$e(n) = s_{org}(n) - s_{synth}(n+60), \quad n = 0,\ldots,239 \tag{1}$$

여기서 s_{org} 는 원 음성 신호를, s_{synth} 는 기본 계층의 합성된 음성 신호를 나타낸다. 복호화기에서 최종적 으로 합성되는 신호 s_{enh} 는 식 2에 의해서 구해진다.

$$s_{enh}(n) = s_{synth}(n) + e_{synth}(n), \quad n = 0,\ldots,239 \tag{2}$$

여기서 e_{synth} 는 개선 계층에서 복원된 합성 에러 신호이다.

설계된 비트율 scalable 음성 코덱은 프레임내 입력신호에 대해 기본 계층과 개선 계층이 연속적으로 동작하므로 전체 알고리즘 지연은 기본 계층의 알고리즘 지연인 37.5ms와 같다.

2.2. LPC 분석

개선 계층의 부호화기로 입력되는 합성 에러 신호는 전처리 과정을 거치지 않은 원 음성신호와 합성된 음성신호간의 차이 값이므로 DC 성분을 포함한 저주파 성분을 제거하기 위해 전처리 과정으로 식 3과 같은 고대역 필터를 통과한다.

$$H(z)=\frac{1-z^{-1}}{1-127/128\,z^{-1}} \tag{3}$$

전처리 과정을 거친 신호에 대해 LPC 계수 분석이 이루어 지는데, 먼저 그림 2와 같이 300 샘플 길이의 Hamming-cosine 윈도우를 사용하여 네 번째 부-프레임이 중앙에 위치하도록 윈도우를 취한다. 식 4는 Hamming-cosine 윈도우를 나타낸다.

$$W(n)=\begin{cases}0.54-0.46\cos(\dfrac{2\pi n}{2L_1-1}), & n=0,\ldots,L_1-1\\ \cos(\dfrac{2\pi(n-L_1)}{4L_2-1}), & n=L_1,\ldots,L_1+L_2-1\end{cases} \tag{4}$$

여기서, L_1=278 이며, L_2=22 이다.

윈도우가 취하여진 신호를 이용하여 11개의 자기 상관계수를 구한 후 Levinson-Durbin 알고리즘을 이용하여 네 번째 부-프레임에 사용되는 10개의 LPC 계수를 추출한다. 구하여진 LPC 계수는 단구간 지각 가중 필터와 합성 필터의 구성을 위해 사용되는데, LPC 합성 필터는 식 5와 같이 정의된다.

$$A_i(z)=\frac{1}{1-\sum_{j=1}^{10}a_{ij}z^{-j}},\quad 0\le i\le 3 \tag{5}$$

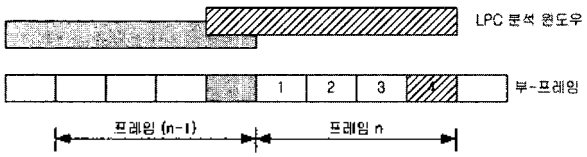

그림 2. 개선 계층 windowing 과정
Fig. 2. Windowing of enhancement layer

표 3. 5.7kbps 개선 계층의 LSP 양자화에 대한 비트할당
Table 3. Bit allocation of the LSP quantization in 5.7kbps enhancement layer.

파라미터	비트할당
패스 정보	10
코드워드 정보	2x4 (for stage 1~4) 1x6 (for stage 6~10)
합계	24

표 4. 6.7kbps 개선 계층의 LSP 양자화에 대한 비트할당
Table 4. Bit allocation of the LSP quantization in 6.7kbps enhancement layer.

파라미터	비트할당
패스 정보	10
코드워드 정보	2x7 (for stage 1~7) 1x3 (for stage 8~10)
합계	27

2.3. LSP 양자화기

추출된 LPC 계수는 높은 양자화 성능을 얻기 위해 양자화 특성이 LPC 합성 필터에 영향이 적고 LPC 계수와 수학적으로 등가인 line spectral pair (LSP)[3]계수로 변환 된다. 변환된 LSP 계수 양자화를 위해 프레임내 상관도를 이용한 1차 AR 필터 예측 구조를 적용하며, 예측 에러 값을 block-constrained trellis coded quantization (BC-TCQ)[2]로 양자화하는 그림 3과 같은 구조를 갖는다.

먼저 식 6과 같이 LSP 벡터 $\underline{p}'$에서 LSP 벡터의 DC 성분을 제거한 후 프레임내 요소간 상관도를 이용해서 식 7과 같이 1차 AR 필터 예측이 이루어진다. 구해진 예측 에러 값 t_j 는 BC-TCQ 양자화기를 사용하여 양자화가 이루어진다.

$$\underline{p}(n)=\underline{p}'(n)-\underline{p}_{DC}(n) \tag{6}$$

$$t_j(n)=p_j(n)-\rho_j\times p_{j|j-1}(n) \tag{7}$$

여기서 $\underline{p}'(n)$은 LSP 벡터이고, ρ_j는 프레임내 요소간 상관도를 나타낸다.

LSP 벡터 양자화에 사용되는 비트 수는 아래 표 3및 표 4와 같다. 개선 계층이 5.7kbps로 동작할 경우 표 3

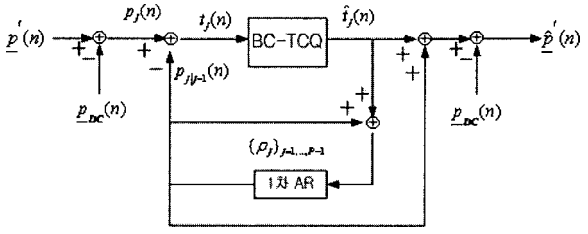

그림 3. AR intra-predictive BC-TCQ 인코딩 과정
Fig. 3. LSP quantizer using AR intra-predictive BC-TCQ.

과 같이 패스 정보에 10비트, 코드워드 정보에 14비트를 할당하여 총 24비트로 양자화가 이루어 지고, 개선 계층이 6.7kbps로 동작할 경우 표 4와 같이 패스 정보에 10비트, 코드워드 정보에 17비트를 할당하여 총 27비트로 양자화가 이루어 진다. 여기서 패스 정보는 trellis 구조에서 viterbi 알고리즘에 의해 결정되는 최적 경로 정보이고, 코드워드 정보는 각 branch에 할당된 부-코드북의 코드워드 정보이다[2].

2.4. LSP 선형보간

네 번째 부-프레임의 LSP 정보만이 복호화기로 전달되기 때문에 나머지 세개 부-프레임들의 LSP 벡터는 이전 프레임의 네 번째 부-프레임 LSP 벡터와 현재 프레임의 네 번째 부-프레임 LSP 벡터를 이용하여 식 8과 같이 보간된다.

$$\hat{p}'_{ni} = \begin{cases} 0.65\,\hat{p}'_{n-1} + 0.35\,\hat{p}'_n, & i=0 \\ 0.32\,\hat{p}'_{n-1} + 0.68\,\hat{p}'_n, & i=1 \\ \hat{p}'_n, & i=2 \\ \hat{p}'_n, & i=3 \end{cases} \tag{8}$$

여기서 i 는 부-프레임 인덱스이다. 위와 같은 보간 과정을 통해 구해진 LSP 벡터는 LPC 계수 벡터로 변환된 후 인지적 가중 필터의 계수 값으로 사용된다.

선형 예측 분석이 끝난 잔여 신호는 2절에서 구한 양자화 되지 않은 LPC 계수를 바탕으로 식 9로 표현되는 인지적 가중 필터를 통과하게 된다.

$$W_i(z) = \frac{1 - \sum_{j=1}^{10} a_{ij} z^{-j} \gamma_1^j}{1 - \sum_{j=1}^{10} a_{ij} z^{-j} \gamma_2^j}, \quad 0 \le i \le 3 \tag{9}$$

여기서 γ_1=0.9, γ_2=0.4의 값을 가지며 주파수적 가중치를 결정한다. 인지적 가중 필터를 통과한 잔여신호는 G.723.1과 같은 적응 코드북 탐색 과정을 거치게 된다.

2.5. 고정 코드북 탐색

적응 코드북의 기여도가 제거된 신호에 대해 고정 코드북 탐색과정이 이루어진다. 개선 계층의 고정 코드북은 G.723.1 5.3kbps 모드의 구조를 바 탕으로 ACELP 방식을 사용하였고 G.723.1 5.3kbps 모드와 달리 grid 비트에 의한 펄스 위치 제약을 받지 않도록 설계하였다.

고정 코드북 탐색의 기본적 개념은 식10에 주 어진 목표 신호 $\mathbf{r}$과 가중된 합성 음성간의 MSE를 최소화 하는 것이다.

$$E_\zeta = \left\| \mathbf{r} - G\mathbf{H}\mathbf{v}_\zeta \right\|^2, \tag{10}$$

여기서 $\mathbf{r}$은 가중 음성으로부터 가중 합성 필터의 영입력 응답과 피치 기여값을 제거한 벡터이고, G는 코드북 이득값, $\mathbf{v}_\zeta$ 는 ζ 번째 대수 코드 벡터 이다. 행렬H 는 가중 합성 필터의 임펄스 응답인 h(n) 으로 구성되는데, 대각선 성분은 h(0), 나머지 성분은 h(1) ,…, h(L-1)을 사용하여 계산된 대칭적 행렬로 구성된다. 식 10을 미분하면 식 11과 같이 정리되고 식11을 최대화하는 최적의 코드 벡터를 찾아야 한다.

$$\tau_\xi = \frac{\mathbf{C}_\xi^2}{\epsilon_\xi} = \frac{(\mathbf{d}^T \mathbf{v}_\xi)^2}{\mathbf{v}_\xi^T \Phi \mathbf{v}_\xi} \tag{11}$$

여기서 d 는 대상 벡터 신호 $\mathbf{r}$과 임펄스 응답 h 사 이의 상관 행렬이고, Φ는 임펄스 응답의 공분산 행렬이다. 벡터 d와 행렬 Φ는 탐색에 필요한 요소 만 순서에 맞춰 미리 계산함으로써 매우 빠른 탐색 이 가능하다.

2.5.1. 5.7kbps 개선 계층의 고정 코드북

5.7kbps 개선 계층은 표 5와 같은 ACELP 구조의 고정 코드북을 사용하였다. 각 부-프레임을 기준으로 고정 코드북 탐색이 이루어지며, 고정 코드북은 4개의 트랙으로 구성되며 각 트랙에서 하나의 펄스가 찾아진다. 하나의 트랙에서 펄스가 존재할 수 있는 위치는 표 5와 같이 15개이고, 찾아지는 펄스의 크기는 ±1이다. 따라서 고정 코드북에 할당되는 비트는 부-프레임당 위치 정보 16비트와 부호 정보 4비트이다. 고정 코드북 탐색은 식 11의 값을 최대화 하는 코드북을 찾는 방식으로 4개의 겹쳐진 루프를 통하여 위치 및 부호가 결정된다.

2.5.2. 6.7kbps 개선 계층의 고정 코드북

6.7kbps 개선 계층은 표 6과 같은 ACELP 구조의 고정 코드북을 사용하였다.

각 부-프레임을 기준으로 고정 코드북 탐색이 이루 어지며, 6개의 트랙이 존재하고 각 트랙마다 하나의 펄스가 찾아진다. 하나의 트랙에서 펄스가 존재할 수 있는

표 5. 5.7kbps 개선 계층의 ACELP 구조
Table 5. ACELP codebook of 5.7kbps enhancement layer.

부호	가능한 위치	비트할당	
		위치	부호
+,-1	0,6,12,18,24,30,36,42,48,54	10	1
+,-1	1,7,13,19,25,31,37,43,49,55		1
+,-1	2,8,14,20,26,32,38,44,50,56		1
+,-1	3,9,15,21,27,33,39,45,51,57	10	1
+,-1	4,10,16,22,28,34,40,46,52,58		1
+,-1	5,11,17,23,29,35,41,47,53,59		1

위치는 표 6에서 보여지는 것과 같이 10개 이고, 찾아지는 펄스의 크기는 1이다. 3개의 트 랙에서 트랙 당 하나의 펀스가 위치할 수 있는 경우 는 1000 (=10x10x10) 가지가 존재하며 10비트로 표현 이 가능하다. 따라서 고정 코드북에 할당되는 비트 는 부-프레임당 위치 정보 20 (=10x2)비트와 부호 정 보 6비트이다. 최적의 펄스위치 검색 과정은 다음과 같다.

1) 각 트랙에서 식 11의 값을 최대로 하는 4개 의 펄스 위치를 찾는다.

2) 단계 1에서 구해진 트랙별 4개 펄스 위치들에 한해, 첫 번째, 두 번째, 세 번째 트랙을 이용한 3개의 겹쳐진 루프를 통하여 식 11의 값을 최대로 하는 3개 펄스들의 최적 위치를 찾는다.

3) 단계 2에서 구해진 3개 펄스의 최적 위치를 가정하고 단계 1에서 구해진 트랙별 4개 펄스 위치들에 한해 네 번째, 다섯 번째, 여섯 번째 트랙을 이용한 3개의 겹쳐진 루프를 통하여 식 11의 값을 최대로 하는 3개 펄스들의 최적 위치를 찾는다.

4) 단계 3에서 찾아진 3개 펄스를 가정하고 단계 2와 단계 3을 한번 더 수행한다.

식 11의 값을 구할 때 사용되는 C 와 $\mathcal{E}$ 의 값은 식 12 및 식 13에 의해 각각 정의된다.

$$\mathbf{C} = d'(m_0)+d'(m_1)+d'(m_2)+d'(m_3)+d'(m_4)+d'(m_5) \quad (12)$$

$$\begin{aligned}\varepsilon &= \Phi'(m_0,m_0)+\Phi'(m_1,m_1)+2\Phi'(m_0,m_1)\\ &+\Phi'(m_2,m_2)+2[\textstyle\sum_{i=0}^{1}\Phi'(m_i,m_2)]\\ &+\Phi'(m_3,m_3)+2[\textstyle\sum_{i=0}^{2}\Phi'(m_i,m_3)]\\ &+\Phi'(m_4,m_4)+2[\textstyle\sum_{i=0}^{3}\Phi'(m_i,m_4)]\\ &+\Phi'(m_5,m_5)+2[\textstyle\sum_{i=0}^{4}\Phi'(m_i,m_5)]\end{aligned} \quad (13)$$

여기서 는 번째 펄스 위치이고, 와 $\Phi'(i,j)$ 는 번째 펄스 위치에서 상관 행렬 d의 부호를 나타 내는 $s(k)$를 이용하여 식14와 같이 표현된다.

표 6. 6.7kbps 개선 계층의 ACELP 구조
Table 6. ACELP codebook of 6.7kbps enhancement layer.

부호	가능한 위치	비트	
		위치	부호
+,-1	0,4,8,12,16,20,24,28,32,36,40,44,48,52,56	4	1
+,-1	1,5,9,13,17,21,25,29,33,37,41,45,49,53,57	4	1
+,-1	2,6,10,14,18,22,26,30,34,38,42,46,50,54,58	4	1
+,-1	3,7,11,15,19,23,27,31,35,39,43,47,51,55,59	4	1

$$d'(j) = d(j)s(j), \quad \Phi'(i,j) = s(i)s(j)\Phi(i,j) \quad (14)$$

2.6. 고정 코드북 이득값 양자화

개선 계층에서 구해지는 고정 코드북 이득값은 인접 부-프레임간에 상관도가 매우 높다. 따라서 주 어진 비트에서 더 높은 양자화 성능을 얻기 위해 부-프레임간 상관도를 이용한 예측 구조를 고정 코드북 이득값 양자화기에 적용하였다. 또한 프레임 에러가 발생했을 경우 초래되는 예측 애러 누적 현상을 최소화 하기 위해 각 프레임내 첫 번째 부-프레임에서는 로그 변환한 이득값을 균일(uniform) 양자화 방식을 사용해 양자화를 수행하였고, 나머지 부-프레임에서는 로그 변환한 이득값에 대해 부-프레임간 상관도를 이용한 1차 AR 필터 예측을 한 후, 구해지는 예측 에러 값을 균일 양자화 방식을 사용해 양자화를 수행하였다. 그리고 양자화 성능을 보다 더 좋게하기 위하여 위의 과정을 통해 양자화 된 이득값을 중심으로 폐-루프 최적화 과정을 한번 더 수행하였다. 구체적인 이득값 양자화 과정은 아 래와 같다. 고정 코드북 이득값 는 식15에 의해 서 구해진다.

$$y(n) = h(n) * c(n), \quad g = \frac{\sum r(n)\cdot y(n)}{\sum y(n)\cdot y(n)} \quad (15)$$

여기서, *는 신호간의 컨볼루션이고, $y(n)$ 은 고정 코드북의 기여분이고, $r(n)$ 은 고정 코드북 목표 신호이다.

부-프레임 0에서의 고정 코드북 이득값 양자 화는 그림 4와 같이 이루어진다. 먼저 각 프레임내 부-프레임 0에서 구해진 고정 코드북 이득값을 로 그 변환하고 변환된 값을 24레벨을 갖는 균일 양자 화기로 양자화한다. 그리고 양자화된 이득값을 중심 으로 8개 레벨 범위 안에서 폐-루프 최적화 과정을 수행한다.

부-프레임 1,2,3에서의 고정 코드북 이득값 양자화는

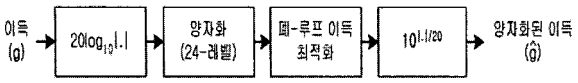

그림 4. 부-프레임0에 대한 고정 코드북 이득값 양자화
Fig. 4. Quantization of fixed codebook gain for subframe 0.

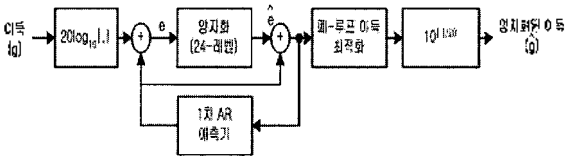

그림 5. 부-프레임1,2,3에 대한 고정 코드북 이득값 양자화
Fig. 5. Quantization of fixed codebook gain for subframes 1,2,3.

그림 5와 같이 이루어진다. 각 부-프레임에서 구해진 이득값을 로그 변환한 후, 변환된 값 $g_{\log}$에 대해 식 16과 같이 1차 AR 필터 예측을 적용하여 예측 에러 e를 구한다. 구해진 예측 에러값은 24 레벨을 갖는 균일 양자화기로 양자화가 이루어진다. 마지막으로 양자화된 이득값을 중심으로 8개 레벨 범위 안에서 폐-루프 최적화 과정을 수행한다.

$$e^j = g^j_{\log} - \rho^j \cdot g^{j-1}_{\log}, \quad j=1,2,3 \tag{16}$$

여기서, j 는 부-프레임의 인덱스를 나타내고, ρ는 부-프레임간 상관도이다.

III. 실험 결과

BC-TCQ방식을 이용한 개선 계층의 LSP 양자화 기 성능이 G.723.1에서 사용되는 split VQ방식과 SD값을 통해 분석되었다. 성능 평가를 위해 20,376 프레임의 음성 샘플을 사용하여 BC-TCQ를 훈련하였 으며, 5,654프레임의 음성 샘플을 사용하여 테스트 를 하였다. 훈련용 및 테스트용 음성 샘플은 한국어 남자, 여자, 영어 남자, 여자의 음성 순으로 반복 구성되어 있다. 표 7에서 보여지는 것과 같이 split VQ 방식과 BC-TCQ방식의 SD 성능을 비교해 보면, 프레임당 24 비트에서 BC-TCQ방식의 2~4dB outlier 값은 split VQ방식보다 2.762% 더 작다. 그리고 BC-TCQ방식의 평균 SD값은 split VQ방식보다 0.14dB 작다.

설계된 개선 계층에서 요구되는 메모리량과 WMOPS (Weighted Million Operation Per Sec) 계산량을 표 8 및 9에 나타내었다. 개선 계층이 6.7kbps로 부호화가 이

표 7. 개선 계층의 LSP 양자화기 SD 성능
Table 7. SD performance of the LSP quantizer in enhancement layer.

	Split VQ (24비트)	BC-TCQ (24 비트)	BC-TCQ (27 비트)
Ave. SD(dB)	1.297	1.157	0.950
2dB~4dB(%)	5.389	2.627	0.758
>4dB(%)	0.0	0.0	0.0

표 8. 개선 계층의 ROM 및 RAM 크기
Table 8. ROM and RAM size of the enhancement layer.

	6.7kbps 개선 계층	5.7kbps 개선 계층
ROM 크기 (words)	609	561
RAM 크기 (words)	1232	1186

표 9. 개선 계층의 WMOPS 계산량
Table 9. WMOPS complexity of the enhancement layer.

6.7kbps 개선 계층 (WMOPS)			5.7kbps 개선 계층 (WMOPS)		
최소	최대	평균	최소	최대	평균
14.713	14.961	14.811	19.919	25.044	21.879

표 10. G.723.1을 기본 계층으로 하는 비트율 scalable 코덱의 PESQ 성능
Table 10. PESQ performance of a bitrate scalable codec based on G.723.1.

코 덱	PESQ
G.723.1 (5.3kbps)	3.480
G.723.1 (6.3kbps)	3.612
G.723.1 (5.3kbps) + 개선계층(6.7kbps)	3.852
G.723.1 (6.3kbps) + 개선계층(5.7kbps)	3.879
G.729 E (11.8 kbps)	3.946

루어질 경우 609 및 1232 words의 ROM 및 RAM 메모리를 필요로 하고, 평균 WMOPS 계산량은 14.811을 나타낸다. 개선 계층이 5.7kbps로 부호화 가 이루어질 경우 561 및 1186 words의 ROM 및 RAM 메모리를 필요로 하고, 평균 WMOPS 계산량은 21.879 를 나타낸다.

표 10에서는 G.723.1에 적용한 비트율 scalable 코덱의 성능을 G.723.1 및 G.729 annex E[4]와 평균 PESQ 값으로 비교하였다. 실험에 사용된 음성은 ITU-T에서 음성 코덱의 음질 평가를 위해 제공하는 잡음없는 한국어 음성으로서 8kHz의 샘플링 주파수 를 갖는다. 음성 샘플은 G.191[6]을 사용하여 3개 신호레벨들(-16dB, -26dB, -36dB)로 나누었으며, 각 레벨 별로 남자 음성 96개와 여자 음성 96개로 구성되어 있다.

구체적으로 살펴보면, G.723.1 5.3kbps 모드를 기반으로 설계된 비트율scalable 음성 코덱은 G.723.1 5.3kbps 보다 PESQ 값이 0.372 더 높고, G.723.1 6.3kbps 모드를 기반으로 설계된 비트율 scalable 음성 코덱은 G.723.1 6.3kbps 보다 PESQ 값이 0.267 더 높다. 또한 G.723.1을 기본 계층으 로 하는 비트율 scalable 코

덱은 embedded 시스템 형태로 설계되어야 하는 구조적 제한을 받기 때문에 개선 계층이 6.7 및 5.7kbps로 동작할 때 고정된 전송률로 설계된 G.729 annex E 보다 PESQ 성능이 각각 0.094 및 0.067 떨어진다.

IV. 결 론

본 논문에서는 협대역 신호에 대해 G.723.1 5.3kbps 및 6.3kbps 두 개 모드를 기본 코덱으로 해서 전체 12kbps로 동작하는 비트율 scalable 음 성 코덱을 제안하였다. G.723.1을 기반으로 하는 비트율 scalable 음성 코덱은 기본 계층에서 합성 한 신호와 원 신호간의 차이 값인 합성 에러 신호를 개선 계층에서 합성하고, 기본 계층과 개선 계층에 서 생성되는 합성 신호들의 합을 최종 합성 신호로 출력하는 구조를 갖는다. 설계된 개선 계층은 기본 계층의 구조 내부에 삽입되는 것이 아니라 독립적으 로 동작하므로 기본 코덱의 부호화 방식에 구애를 받지 않는다. 따라서 어떤 종류의 기본 계층 음성 부호화기에도 적용이 가능한 넓은 호환성을 갖는다. 개선 계층은 두 가지 전송률을 갖는다. 기본 계층 이 G.723.1 5.3kbps 모드로 동작하면 개선 계층은 6.7kbps로 동작하고, 기본 계층이G.723.1 6.3kbps 모드로 동작하면 개선 계층은 5.7kbps로 동작한다. 설계된 개선 계층은 기본적으로 G.723.1 5.3kbps 모 드의 구조를 사용하였으며 LPC 모듈, 고정 코드북 모듈, 그리고 고정 코드북 이득값 양자화 모듈들을 합성 에러 신호에 맞게 새로이 설계 하였다.

G.723.1 기반 비트율 scalable 코덱의 전체적인 성능 평가는 ITU-T 표준 음질 측정 소프트웨어인 P.862 (PESQ)를 사용하였다. G.723.1 5.3kbps모드와 개선 계층 6.7kbps가 동작할 경우 G.723.1 5.3kbps 모드 보다 PESQ 수치가 0.372 더 높았고, G.723.1 6.3kbps모드와 개선 계층 5.7kbps가 동작할 경우 G.723.1 6.3kbps모드 보다 0.267 더 높았다.

감사의 글

본 연구는 삼성전자㈜ 삼성종합기술원의 지원을 받아 수행 되었습니다.

참고문헌

1. "Dual rate speech coder for multimedia communications transmitting at 5.3 and 6.3kbit/s,"CCITT Recommendation G.723.1, March 1996.
2. S. W. Kang, Y. W. Shin and T. R. Ficher, "Low complexity predictive trellis-coded quantization of speech line spectral frequencies," IEEE Trans. on Signal Processing, 52 (7), 2070-2079, July 2004.
3. K. K. Paliwal and B. S. Atal, "Efficient vector quantization of LPC parameter at 24bits/frame," IEEE Trans. Speech Audio Processing, 1 (1), 3-14, Jan., 1993.
4. "Coding of speech at 8kbit/s using conjugate structure algebraic code excited linear prediction (CS-ACELP),"CCITT Recommendation G.729, March 1996.
5. "Perceptual evaluation of speech quality an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,"ITU-T Recommendation P.862, Feb., 2001.
6. "ITU-T Software Tool Library 2000 User's Manual," ITU-T Users'Group on Software Tools 7, Dec., 2000.

저자 약력

• **이 준 석 (Joonseok Lee)**

2003년 2월: 한양대학교 전자컴퓨터 공학부 졸업 (학사)
2003년 3월~ 현재: 한양대학교 전자전기제어계측공학과 대학원 과정
※주관심 분야: 음성 신호처리, 임베디드 시스템

• **강 상 원 (Sangwon Kang)**
한국음향학회지 제20권 제4호 참조

• **이 강 은 (Kangeun Lee)**
한국음향학회지 제23권 제4호 참조

• **박 동 원 (Dongwon Park)**
한국음향학회지 제23권 제5호 참조