

가변율 half rate 음성 부호화기의 설계

정회원 성 호 상*, 강 상 원*

Design of a variable half rate speech codec

Hosang Sung*, Sangwon Kang* *Regular Members*

요 약

본 논문에서는 다양한 멀티미디어 서비스를 위한 가변율 half rate 음성 부호화기를 설계하였다. 음성의 특성 및 활성도를 이용한 효율적인 가변율 알고리즘을 위해, 프레임 에너지와 음성 파라미터들을 이용한 voicing 결정 방식을 제시하였다. 유성음을 위한 4 kb/s 음성 부호화기는 저속에서 좋은 특성을 보이는 generalized analysis-by-synthesis (AbS) 구조를 이용하고, LP 계수는 line spectrum pair (LSP) 계수로 변환한 후 예측 2단 VQ를 통해서 양자화하며, 여기 신호는 음질저하를 최소화하며 복잡도를 감소시킨 shift 방식의 대수적 고정 코드북 구조를 사용하였다. 무성음을 위해 별도의 4 kb/s 부호화기를 설계하였는데, 피치 탐색방법을 제외한 나머지 부분은 유성음을 위한 부호화기와 동일하다. 또한 묵음과 주변소음 구간을 위해 1 kb/s 부음성 부호화기를 설계하였다. 본 논문에서 제시된 가변율 half rate 부호화기는 voice activity factor (VAF)가 0.47인 시험음성에서 약 2.6 kb/s의 평균 전송률을 보였으며, mean opinion score(MOS) 3.69의 성능을 보였다.

ABSTRACT

In this paper, we combine a generalized analysis-by-synthesis (AbS) structure and an algebraic excitation scheme to propose a variable half rate speech codec. Also we introduce an efficient voicing detection algorithm using both frame energy and acoustic parameters to detect voiced, unvoiced and silence region. 4 kb/s speech codec for the voiced region uses the generalized AbS structure which is efficient for the low bit rate speech codec. LP coefficients are converted to line spectrum pairs (LSP) and quantized using a predictive 2-stage VQ, and a low complexity algebraic codebook which uses shifting method is used for the fixed codebook excitation. 4 kb/s speech codec for the unvoiced region is same as that for the voiced region, except that the latter uses an open loop pitch lag in the whole frame. Also 1 kb/s subrate codec is designed for the silence and background noise region. Experiments show that the variable half rate codec provides the average bit rate of about 2.6 kb/s for the test sentences whose voice activity factor (VAF) is 0.47. Average mean opinion score (MOS) of the proposed variable half rate codec is 3.69.

1. 서 론

국제 통신 표준화 기구인 ITU-T의 SG-15에서는 차세대 멀티미디어 및 이동 통신 서비스를 위한 음성 부호화 방식으로서 8 kb/s의 전송속도로 유선망

에서의 음성품질을 제공하는 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]. 이 후 ITU-T에서

* 한양대학교 제어계측공학과 (swkang@selab.hanyang.ac.kr)

논문번호 : 99060-0218, 접수일자 : 1999년 7월 31일

※ 본 연구는 정보통신부에서 지원하는 대학기초연구지원사업(과제번호: C1-98-5011)으로 수행되었습니다.

는 4 kb/s 음성 부호화기의 표준을 진행 중이지만 아직 요구사항을 완전히 만족시키는 부호화기가 결정이 되지 못한 상태이다. 현재 저 전송률 음성 부호화기는 크게 CELP형의 analysis-by-synthesis (AbS)방법^[4]과 하모닉 부호화 방법^[5]으로 나눌 수 있는데, 어떤 방법이 4 kb/s 전송속도에서 더 나은 성능을 가지는지 아직 결론을 내리지 못하고 있는 상태이다. 본 논문에서는 CELP형의 AbS 방법을 사용하였으며, 이 방법이 4 kb/s의 전송률에서 성능이 많이 저하되는 단점을 generalized AbS구조^[6]를 이용하여 보완하였다. Generalized AbS구조는 피치를 파라미터화하여 부호화하며 이때 발생하는 오류를 원 신호의 warping을 통해서 방지하는 방식으로, 음질 저하없이 피치 신호에 적은 비트를 할당할 수 있으며, 개루프로 분석이 이루어지므로 전송률과 복잡도의 측면에서 많은 장점이 있다. 그러나 피치 신호를 보간할 때 피치주기의 변화가 적은 유성음 구간에서는 문제가 없지만 피치신호가 불규칙적인 무성음 구간에서는 성능이 급격히 감소한다. 이러한 성능 감퇴를 막기 위해 무성음 구간을 위한 별도의 음성 부호화기가 반드시 필요하게 된다. 그래서 본 논문에서는 유성음 구간용 4 kb/s 음성 부호화기 뿐만 아니라 무성음 구간에 해당하는 음성 신호의 부호화를 위해 무성음 구간용 4 kb/s 음성 부호화기를 설계하였다. 또한 유, 무성음, 목음 및 주변소음 구별을 위한 voicing 결정 알고리즘과 목음 및 주변소음을 위한 1 kb/s 음성 부호화기도 설계하였다.

본 논문의 구성은 다음과 같다. II장에서는 generalized AbS 구조에 대해 설명하며, III장에서는 설계된 half rate 음성 부호화기에 대해 설명하고, IV장에서는 voicing 결정 알고리즘과 목음 구간을 위한 1 kb/s 음성 부호화기를 다루었으며, V장에서는 설계된 알고리즘에 대한 실험 및 결과를 분석하며, 마지막으로 VI장에서 결론을 맺는다.

II. Generalized AbS 구조

Generalized AbS 구조는 relaxed CELP (RCELP)^[7]에서 사용되었으며 피치 주기의 선형적인 보간에 의해 많은 비트를 아낄 수 있는 장점이 있으나 보간된 부분에서 발생하는 피치주기의 작은 차이가 일반적인 AbS 메카니즘의 성능에 심각한 영향을 줄 수 있는 단점이 있다. 그래서 generalized

AbS 구조는 피치 주기를 파라미터화 하여 선형보간을 하며 이 보간된 피치 주기에 따라 원음이 손상받지 않는 범위 내에서 원 신호의 시간변형을 통한 왜곡을 수행한다. 이 왜곡된 신호를 이용하여 AbS 메카니즘을 사용하면 피치 보간에 의해 생기는 작은 불일치를 없앨 수 있다. Generalized AbS 구조는 다음의 세 단계로 이루어진다. 첫째 피치 주기 contour를 결정하며, 둘째 과거의 합성신호를 현재로 mapping하며, 셋째 원 신호의 warping을 수행한다.

1. 피치주기 contour의 결정

피치주기는 계산량을 줄인 피치 탐색 알고리즘에 의해 20 ms마다 구해진다. 피치 탐색을 위해 프레임 경계(boundary)에 window 중심을 맞추어 피치 주기를 구하고, 구해진 피치주기를 사용하여 선형보간 되어진다. j 가 현재 부프레임을 나타낼 때, t_j 는 현재 부프레임의 시작 위치이며 이전 부프레임 피치 주기의 마지막 값을 $d(t_j)$ 라 하면 선형보간 되어진 순간적인 주기 $\tau_c(n)$ 은 식 (1)과 같이 나타낸다.

$$\tau_c(n) = d(t_j) + a(n - t_j), \quad t_j < n < t_{j+1} \quad (1)$$

여기서 a 는 부프레임의 크기에 따라 정해지는 상수이다.

보간은 기본주기의 두 배나 세 배에 해당하는 주기성이 발생하는 프레임에 대해서는 수행되지 않으며, 이 경우 이전 프레임 경계의 피치주기가 적당한 정수값에 의해 곱해지거나 나누어진다.

2. 과거 합성 신호의 mapping

RCELP에서 피치주기는 샘플 단위로 선형보간한다. 계산적인 이점을 위해 각 샘플에서의 피치주기는 샘플링 주기의 1/8로 반올림한다. 그리고 과거의 재생 여기신호로부터 현재의 여기신호 샘플을 계산하기 위해 적당한 polyphase 보간 필터를 사용한다. 식 (2)는 polyphase 보간 필터를 사용한 적응 코드북 기여값을 나타낸다.

$$u(n) = \sum_{i=0}^{2L} u(i+n-T_E(n)-f_i)I_E(i+(2f_i+1)T_{IE}(n)), \quad 0 \leq n < (L+10) \quad (2)$$

여기서 L 은 현재 부프레임 크기이며, $T_E(n)$ 과 $T_{IE}(n)$ 은 식 (3)에 의해 구해진다. 그리고 피치주기 contour가 8배로 oversampling되었을 때 oversampli-

ing factor R 은 8이 되며, $I_E(n)$ 은 polyphase 보간 필터 계수의 집합이며, f 은 보간 필터 tap의 갯수이다.

$$T_E(n) = \begin{cases} \text{round}\{\tau_c(n)\}, & \tau_c(n) > 0 \\ -\text{round}\{-\tau_c(n)\}, & \tau_c(n) \leq 0 \end{cases}$$

$$T_{IR}(n) = \text{trunc}\{(T_E(n) - \tau_c(n) + 0.5)R + 0.5\} \quad (3)$$

3. 원신호의 warping

원 신호의 warping 목적은 여기신호에 대한 고정 코드북 기여값(contribution)을 결정할 때 linear-prediction based analysis-by-synthesis (LPAS) 과정을 사용하기 위해서이다. 그래서 RCELP 부호화기의 여기신호에 대한 피치예측기 기여값은 변형된 잔여신호와 정합되어야 하며 시간변형은 원 신호의 변화가 음질에 거의 영향을 주지 않는 범위 내에서 실시한다. 원 신호의 피치주기 contour를 변형하는 가장 좋은 방법은 잔여신호에 대해 연속적인 time-warp 작용을 수행하는 것이다. 이 방식을 사용하면 변형된 신호에 불연속점이 없어진다. 정확한 정합을 위해 원신호의 변형은 식 (4)의 정합기준(matching criterion)을 이용해야 한다.

$$\epsilon' = \frac{((x_c(l) + q(l))^T H^T H u(l))^2}{(x_c(l) + q(l))^T H^T H (x_c(l) + q(l)) u(l)^T H^T H u(l)} \quad (4)$$

여기서, $x_c(l)$ 은 시간 변형된 LP 잔여신호이며, H 는 spectral weighting matrix이며, 현재 프레임의 여기신호 $q(i)$ 는 다음 식과 같이 정의된다.

$$q(i) = \begin{cases} 0, & i < \hat{i} \\ z(i) - \sum_{n=1}^{\hat{i}-i} \gamma^n a_n z(i-n), & \hat{i} \leq i < \hat{i} + N \\ 0, & i \geq \hat{i} + N \end{cases} \quad (5)$$

여기서, \hat{i} 은 부프레임의 첫 번째 샘플을 가리키며, $z(i)$ 는 합성필터의 zero input 응답을 나타낸다. γ 는 weighting factor이며, a_n 은 LP 계수를 나타내며, N 은 LP 계수의 차수인데 8 kHz 샘플링 주기에서 보통 10을 사용한다. 피치예측기 기여값 사이의 정합 정밀도를 위해 고정 코드북 기여값을 더하기 전에 부프레임 단위로 warping 작용을 수행해야 한다. 이 과정에서 잔여신호는 delay나 advance에 대해 작은 segment들로 나누어지는데 그것을 shift라 하며 상수 값을 갖는다. 그러한 segment들의 경계

에서 shift가 바뀔 때마다 잔여신호의 어떤 부분이 생략되거나 반복된다. 이러한 사실 때문에 shift segment 경계는 중요한 특징값(예, 피치 펄스)을 포함하지 않은 부분으로 설정되어야 한다.

III. Half rate 음성 부호화기 설계

유성음을 위한 4 kb/s 음성 부호화기의 전체적인 구조는 그림 1과 같다. 그림 1에 의하면 원 신호의 시간 변형 블록을 제외하면 일반적인 CELP 구조^[8]와 유사함을 알 수 있다. 여기서 시간변형 부분은 II장에서 설명한 방식을 사용하였다.

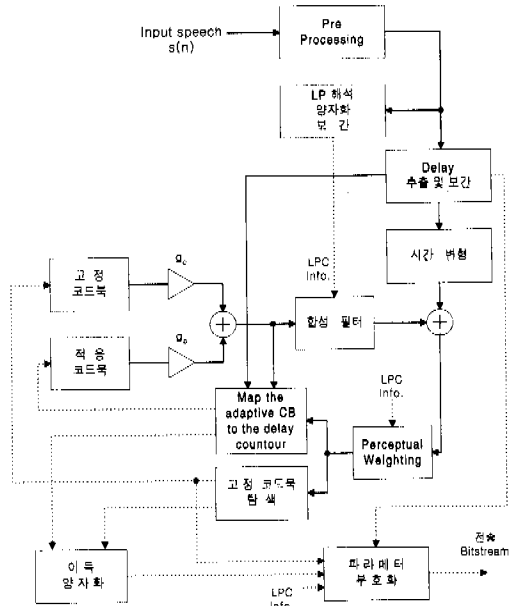


그림 1. Generalized Abs 방식을 이용한 4 kb/s 음성 부호화기의 구조

1. 프레임 구조 및 비트 할당

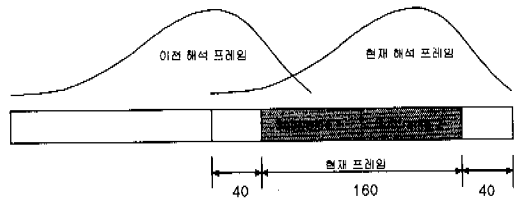


그림 2. 프레임 구조 및 windowing

입력된 신호는 8 kHz로 샘플링되고 그림 2에서와

같이 20 ms의 프레임과 10 ms의 부프레임을 가지며 5 ms의 lookahead를 가지므로 총 알고리즘 지연은 25 ms가 된다. LP 계수 추출을 위해, 과거 프레임의 40 샘플과 현재 프레임의 160 샘플 그리고 이후 프레임의 40 샘플로 총 240 샘플을 이용하며 비대칭의 hamming window를 사용한다. 표 1은 제안된 4 kb/s 음성 부호화기의 비트 할당을 나타낸다.

표 1. Generalized AbS 구조를 이용한 4 kb/s 음성 부호화기의 비트 할당

파라미터	부프레임 1	부프레임 2	프레임당 비트수
LSPs	23		23
적용 코드북 delay	7		7
고정 코드북 grid index	1	1	2
고정 코드북 위치	13	13	26
고정 코드북 부호	4	4	8
코드북 이득 (stage 1)	3	3	6
코드북 이득 (stage 2)	4	4	8
Total			80

2. LPC 계수의 양자화

10차의 linear prediction (LP) 해석 필터를 통해 얻어진 LP 계수는 양자화와 안정도 측면에서 유리한 LSP계수^[9]로 변환되며, 이 값과 4차의 moving average (MA) 예측기를 통해 예측된 값과의 차이 신호를 각각 128개와 32개의 코드워드값을 갖는 2 단계의 VQ와 3 split VQ^[10]를 통하여 총 23 비트로 부호화 한다. 첫 번째 단계의 VQ는 7 비트를 할당하였으며 10차의 벡터 크기를 갖는다. 두 번째 단계의 VQ는 총 15비트(=5+5+5)를 할당한 3 split VQ를 사용하였으며 각각은 3,3,4의 차원을 갖는다. 부프레임 2는 현재 프레임에서 구해진 LP 계수를 그대로 사용하며 부프레임 1은 다음 식과 같이 이전 프레임과 현재 프레임에서 구해진 두 LP 계수를 각각 25% 및 75%의 비율로 보간하여 사용한다.

$$\omega_i = 0.25\omega_i^{(previous)} + 0.75\omega_i^{(current)}, \quad i = 1, \dots, 10 \quad (6)$$

3. 적은 계산량을 갖는 algebraic 코드북의 설계

일반적인 algebraic 구조의 코드북^[11]은 좋은 성능을 가지는 데 반해 고정 코드북 부프레임의 크기가 클 경우 계산량 측면과 사용 비트 측면에서 불리하다. 특히 위치정보가 80가지인 10 ms의 부프레임에서 일반적인 algebraic 구조는 복잡도 측면에서 비효율적이다. 적은 계산량을 갖는 algebraic 코드북은 grid 인덱스를 사용하는 algebraic 구조로서 적은 계산량으로 우수한 성능을 나타낸다. 고정 코드북 부프레임(10 ms: 80 샘플)의 여기신호 벡터는 4개의 영이 아닌 펄스만 제외하고 모두 영의 성분들로 구성되어지며 그 구조는 표 2에 나타나 있다. 부호화 시 첫 세 펄스들의 위치정보에 각각 3비트를 할당하고, 네 번째 펄스의 위치에 4비트를 할당하며, 부호정보에 각각 1비트씩을 할당함으로써 총 17비트가 할당된다. 모든 펄스들의 위치는 표 2에서와 같이 동시에 0 또는 1 샘플만큼 shift를 가능하게 하였다. 이러한 두 가지 shift값 중 합성여러를 최소화하는 최적 shift값을 구하고, 이렇게 구해진 shift값을 전송하기 위해 1비트를 사용한다. 그래서 전체 고정 코드북에 18비트를 할당하였다.

표 2. 적은계산량을 갖는 algebraic 코드북의 구조

(a) Grid index = 0 일 때

펄스	부 호	위 치
i_0	$s_0 : \pm 1$	$m_0 : 0, 10, 20, 30, 40, 50, 60, 70$
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	$s_3 : \pm 1$	$m_3 : 6, 16, 26, 36, 46, 56, 66, 76$ $8, 18, 28, 38, 48, 58, 68, 78$

(b) Grid index = 1 일 때

펄스	부 호	위 치
i_0	$s_0 : \pm 1$	$m_0 : 1, 11, 21, 31, 41, 51, 61, 71$
i_1	$s_1 : \pm 1$	$m_1 : 3, 13, 23, 33, 43, 53, 63, 73$
i_2	$s_2 : \pm 1$	$m_2 : 5, 15, 25, 35, 45, 55, 65, 75$
i_3	$s_3 : \pm 1$	$m_3 : 7, 17, 27, 37, 47, 57, 67, 77$ $9, 19, 29, 39, 49, 59, 69, 79$

4. 이득값의 양자화

적응 코드북과 고정 코드북의 이득값 양자화는 G.729에서 사용된 방식을 이용하였다. 즉 검색의 효율을 높이기 위해 conjugate 구조를 갖는 2개의 코드북을 사용하였다^[12]. 각각의 코드북은 8개와 16개의 요소값을 가지며, 구해진 이득값들에 의해 미리 선택된 4개와 8개의 요소값들에 대해서만 검색이 이루어진다. 결국 적응 및 고정코드북에 대한 최적 이득값들을 32가지의 조합된 경우들에 대해서만 검색을 수행함으로써 각각 8개와 16개의 요소값들을 가지는 코드북의 검색과 비교할 때 상당한 계산량의 감소를 얻게 된다.

5. 무성음을 위한 4 kb/s 음성 부호화기 설계

무성음을 위한 4 kb/s 음성 부호화기는 프레임 크기 및 지연시간 등 대부분은 유성음을 위한 4 kb/s 음성 부호화기와 동일하지만 generalized AbS 구조를 사용할 수 없으므로 적응코드북 검색부분을 무성음구간에 맞게 설계하였다. 무성음 구간은 피치 주기가 일정치 않아서 정확한 피치탐색의 필요성이 없으므로, 복잡도와 구현상의 문제를 고려하여 전체 프레임에서 개루프 피치 탐색을 1번만 실시하며 페루프 피치 탐색은 하지 않는다. 그리고 구해진 정수의 피치 lag는 7비트(128개 lag값)로 부호화 된다. 그리고 무성음을 위한 음성 부호화기에서 가장 필요한 것은 유성음을 위한 음성 부호화기와 메모리 hangover 부분을 구분하여 사용해야 한다는 것이다. 나머지 LP 계수의 양자화, 고정 코드북 탐색, 이득 양자화 부분은 유성음을 위한 4 kb/s 음성 부호화기와 공동으로 사용하며, 비트 할당도 동일하다.

6. Postfiltering

그림 3은 복호화 과정을 나타낸 것이며, 복호화기에서는 전송된 비트 스트림을 이용해 전송된 파라미터를 복원한다. 그리고 이 파라미터를 이용하여 여기신호와 이득값, 피치 지연값과 이득값 그리고 LP 계수를 결정하며 이를 이용하여 합성음성을 복원한다. 부호화기에서 각 계수는 weighted domain에서 추출된다. 이 방식은 auditory masking 특성을 이용했기 때문에 복원된 음성에 대해 noise 감소 효과가 있지만, 4.8 kb/s 이하 저속의 부호화기에서는 효과가 적다. 그래서 추가적인 noise 감소 기술이 필요하다. CELP는 근본적으로 waveform 형태의 부호화기이므로 합성음성에는 Gaussian 형태의 noise가 포함되어 있어서 추가적인 음질 향상기술이

필요하다. 이 문제를 해결하기 위해 postfiltering 과정을 거친다. postfilter의 계수는 매 부프레임마다 갱신되며, postfilter는 long-term postfilter, short-term postfilter, tilt compensation filter로 이루어져 있다. Postfilter의 계수는 filtering의 양을 결정하는데 이 값이 커지면 noise 제거 효과는 뛰어나나 합성음성에서 자연스러움이 없어지는 단점이 있다. 그러므로 청취실험에 의해 가장 적절한 계수를 결정하였다.

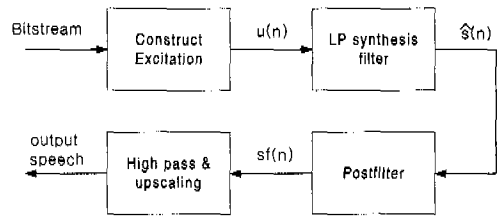


그림 3. 복호화 과정

IV. Voicing 결정 알고리즘 및 1 kb/s 음성 부호화기 설계

본 논문에서는 이미 제안된 효과적인 voicing 결정 알고리즘 및 1 kb/s 음성 부호화기^[13]를 20 ms의 프레임 구조를 갖는 본 부호화기에 맞게 새로이 설계하였다.

1. Voicing 결정 알고리즘

Voicing 결정 알고리즘은 가변율 음성 부호화기의 평균 전송률을 결정짓는 부분으로서 고정율 방식과 비교하여 음질의 저하가 없는 범위 내에서 최소의 평균 전송률을 가지도록 음성구간을 구분하는 것이 그 목적이다. 본 논문에서 설계된 voicing 결정 알고리즘은 음성의 활성화 측정을 위해 음성 프레임의 에너지를 이용하여 주변의 소음 정도에 관계없이 최적의 전송률 결정이 가능하도록 주변소음 레벨에 따라 적응적으로 변화하는 문턱값 함수를 설계하고 이와 함께 음성학적 구분법과 signal-to-noise ratio (SNR)에 따른 가변 hangover 방식을 병행하였다. 그리고 4 kb/s와 1 kb/s 부음성 부호화기들 간의 근본적인 성능 차이로 인한 부자연성을 억제하기 위한 voicing decision smoothing 과정의 수행을 통해 최종 전송률 결정을 하였다. 그림 4는 voicing 결정 시 실제 음성 부분에서 voicing이 선택되는 형태를 나타낸 것이며 bit rate 축에서 4보

다 약간 위쪽으로 선택된 부분은 유성음 구간을 의미하며, 4보다 약간 아래쪽으로 선택된 부분은 무성음 구간을 그리고 1은 목음 및 주변소음구간을 의미하는데 그림을 보면 무성음, 목음 및 주변소음 구간을 잘 찾아내고 있는 것을 알 수 있다.

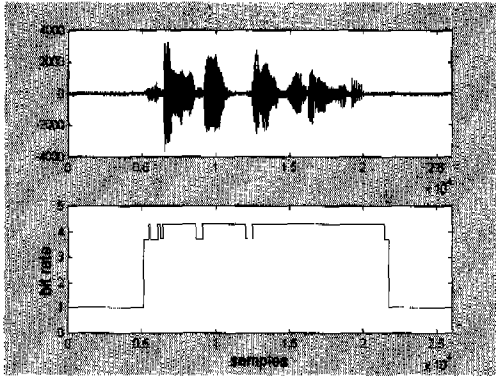


그림 4. 제한된 코래에 의한 voicing 결정 결과

2. 1 kb/s 부음성 부호화기 설계

본 연구에서는 1 kb/s의 전송률로 부호화될 주변소음의 범위를 피치 성분이 미약한 주변소음들로 제한하고 이에 최적인 부음성 부호화기를 설계하였다. 주변소음의 부호화에 사용되는 파라미터들은 단구간 상관도, 즉 소리의 파형(shape)을 결정짓는 LP 계수와 여기 신호의 발생에 사용되는 seed 값, 그리고 여기 신호의 크기를 나타내는 이득값으로 구성된다. 1 kb/s 부음성 부호화기를 통해 부호화된 파라미터들의 비트할당은 표 3과 같다.

표 3. 1 kbps 부음성 부호화기의 비트 할당

파라미터	프레임당 비트수
LSP	7
여기신호 이득	12 (= 3 + 3 + 3 + 3)
Seed	1
Total	20

V. 실험 및 결과

먼저 설계된 LSP 양자화기의 성능을 평가하기 위해 spectral distortion (SD) 값을 구하였다. 시험음성은 8 kHz로 샘플링된 남성 및 여성화자의 약 1분 40초(802,427 샘플)동안의 한국어 음성을 사용하였으며 LSP 양자화기에는 총 23비트를 할당하였다.

표 4는 사용된 LSP 양자화기의 성능을 나타낸 것으로 outlier가 기준을 약간 벗어나지만 평균 SD는 transparent한 성능을 제공함을 알 수 있다.

표 4. LSP 양자화기의 성능

평균 SD(dB)	% of Outlier	
	2~4 dB (%)	> 4 dB (%)
1.00	2.73	0

본 논문에서 설계된 가변을 half rate 부호화기의 성능평가를 위한 음질 측정방법으로서 mean opinion score(MOS) 시험을 이용한 주관적 평가를 실시하였다. MOS 시험은 8명의 청취자를 대상으로 수행되었으며 reference 음성 부호화기의 MOS는 기준에 알려진 음성 부호화기인 8 kb/s QCELP를 MOS score 3.5로 설정하고, 2.4 kb/s LPC 10e를 MOS 2.5로 설정하였다. 그리고 음질 평가 시 채널 잡음 환경하의 성능 분석을 위해 bit error rate (BER) 시험을 함께 수행하였다. 음질측정에 사용된 음성 샘플은 이동 통신 잡음 환경에서 녹음된 6개의 단문들로 구성하였다. 주변 소음의 채취는 자동차 내부, 차도가 있는 거리, 지하철 내부, 역 대합실, 에어컨 소음 등 이동통신 환경하에서 가능한 일반적인 소음들 위주로 이루어졌으며, 이 중 대표적인 주변소음 3가지와 잡음이 없는 음성 샘플들을 혼합하여 잡음 환경의 음성 샘플들을 만들었다. Error free인 경우의 MOS score는 평균 3.69로 나타났는데 이 수치는 전송률이 2배인 가변율 8 kb/s QCELP보다 나은 성능임을 알 수 있다. 그림 5는 MOS 시험 결과를 그래프로 나타낸 것이다. 다음으로 전송률 측면에서의 성능 측정을 위해서는 소음이 포함되지 않은 장문과 이동통신 환경의 잡음을 포함한 VAF 0.47정도의 장문들을 사용하였다.

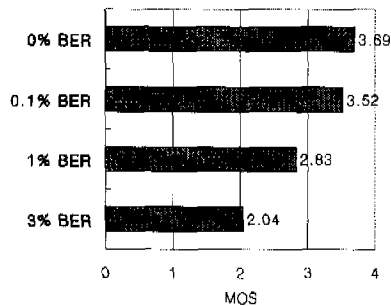


그림 5. MOS 시험 결과

표 5는 설계된 가변율 half rate 음성 부호화기와 가변율 8 kb/s QCELP 부호화기의 잡음 환경 및 비잡음 환경에서 시험한 평균전송률을 비교한 것이며 평균적으로 약 2.6 kb/s의 전송률로 부호화를 수행하는 것으로 나타났다. 설계된 가변율 half rate 음성 부호화기는 8 kb/s QCELP 부호화기와 비교할 때 나은 음질을 가지는 데 반해 약 57%의 전송률을 가짐을 알 수 있다.

표 5. 설계된 가변율 half rate 음성 부호화기와 8 kb/s QCELP 부호화기의 평균 전송률 비교

잡음 환경	잡음정도 [dB]	샘플수	전송률	
			가변율 8 k QCELP 부호화기	가변율 half rate 부호화기
Clean speech	0	206,966 (25.8 초)	4.29 kb/s	2.56 kb/s
자동차 내부	24	207,097 (25.9 초)	3.86 kb/s	2.53 kb/s
거리	26	207,136 (25.9 초)	4.31 kb/s	2.70 kb/s
지하철 내부	32	206,737 (25.8 초)	4.75 kb/s	2.74 kb/s
평균 전송률			4.30 kb/s	2.63 kb/s

VI. 결론

본 논문에서는 generalized Abs 개념을 algebraic CELP 부호화기에 도입한 새로운 가변율 half rate 음성 부호화기를 설계하였다. 20ms의 프레임 크기를 가지며 5 ms의 lookahead를 고려해서 총 25 ms의 알고리즘 전송지연을 갖는다. 전체적인 구조는 G.729를 부분적으로 이용하였으며 LSP 양자화기와 적응 코드북 및 여기 코드북을 4 kb/s 전송 속도에 맞게 새로이 설계하였다. Generalized Abs 구조는 피치를 파라미터화 하여 부호화하므로 피치 신호에 적은 비트를 할당할 수 있으므로 여분의 비트를 LSP 양자화기에 사용하였다. 그리고 무성음에서 성능이 저하되는 단점을 없애기 위해 voicing 결정 알고리즘을 사용하여 유, 무성음과 복음으로 음성을 구분하였고 유성음과 묵음을 자기 다른 알고리즘으로 부호화 하였다. 무성음을 위한 음성 부호화기는 유성음을 위한 음성 부호화기와 피치 lag 부호화 부분만 다르게 설계되었으며 나머지 부분은 동일하다.

그리고 묵음 및 주변소음을 위해 1 kb/s의 알고리즘을 설계하였다. 주관적인 음질 평가를 위한 MOS 시험결과 평균 3.69로 나타났으며, VAF 0.47인 음성에 대해 2.6 kb/s의 평균전송률을 가지는 것으로 나타났다.

참고 문헌

- [1] ITU-T Recommendation G.729, "Coding of speech at 8kb/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] B.S. Atal "Predictive coding of speech signals at low bit rates," *IEEE Trans. Comm.* Vol. 30, No.4, 1982, pp. 600-614.
- [5] R. J. McAulay and T. F. Quatieri, "Speech Analysis /Synthesis Based on a Sinusoidal Representation," *IEEE Trans. on Acoust. Speech and Signal Proc.*, Vol. ASSP-34, No. 4, Aug. 1986, pp. 744-754.
- [6] W. B. Kleijn, R. P. Ramachandran, and P. Kroon, "Generalized analysis-by-synthesis coding and its application to pitch prediction," in *Proc Int. Conf. Acoust., Speech Signal Processing (San Francisco)*, 1992, pp. 1337-1340.
- [7] W. B. Kleijn, P. Kroon, and D. Nahumi, "The RCELP speech coding algorithm," *European Trans. on Telecom.*, Vol. 4, No. 5, 1994, pp. 573-582.
- [8] 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.
- [9] N. Sugamura and N. Farvardin "Quantizer design in LSP speech analysis-synthesis,"

IEEE J. Select. Areas in Commun., Vol.6, Feb. 1988, pp. 432-440.

<주관심 분야> 음성코딩, 이동통신, 신호처리

- [10] K. K. Paliwal and B.S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans Speech and Audio Proc*, Vol. 1, Jan. 1993, pp. 3-14.
- [11] C. Laflamme, J. P. Adoul, R. Salami, S. Morissette, and P. Mabillean, "16 kb/s wideband speech coding technique based on algebraic CELP," in *Proc. ICASSP '91*, Toronto, 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] 정우성, 강상원, 성호상, 이인성, 김재원, 최송인, "Design of a variable rate speech codec for the W-CDMA system," *KSCSP '98*, Vol.15, No.1, pp.142-147.

성 호 상(Hosang Sung)



1996년 2월 : 한양대학교
제어계측공학과
졸업(공학사)

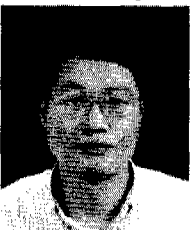
1999년 2월 : 한양대학교
제어계측공학과
졸업(공학석사)

1999년5월 현재 : 한국전자통신
연구원

<주관심 분야> 음성신호처리

강 상 원(Sang-won Kang)

정회원



1958년 3월 20일생

1980년 2월 : 한양대학교
전자공학과 졸업
(공학사)

1982년 2월 : 서울대학교
대학원 전자공학과 졸업
(공학석사)

1990년 7월 : Texas A&M 대학교 전기공학과 졸업
(공학박사)

1982년12월 ~ 1994년2월: 한국전자통신연구소

1994년 3월 ~ 현재: 한양대학교 제어계측공학과
부교수