1998년도 한국음향학회 학술발표대회 논문집 제 17 권 제 2(s) 호

가변율 half rate 음성 부호화기의 설계

성호상, 강상원

한양대학교 제어계측공학과

Design of a variable half rate speech codec

Ho-sang Sung, Sang-won Kang

Dept. of Control & Instrm. Eng., Hanyang Univ.

E-mail) swkang@selab.hanyang.ac.kr

요 약 문

본 논문에서는 다양한 멀티미디어 서비스를 위해 가변율 half rate 음성 부호화기를 설계하였다. 유, 무성음과 묵음의 구분을 위해 본 논문에서는 프레임 에너지와 음성 파라메터들을 이용한 효과적인 voicing 결정 알고리즘을 사용하였다. 유성음을 위한 half rae 음성 부호화기는 시작에서 좋은 특성을 보이는 generalized AbS구조를 이용하였다. LPC 계수는 LSP 계수로 변환한 후 predictive 2-stage VQ를 통해서 양자화하며, 여기 신호는 음질저하를 최소화하며 복잡도를 감소시킨 shift 방식의 대수적 고정 코드북 구조를 사용하고, 적응코드북과 여기코드북의 이득은 VQ로 양자화 하였다. 무성음을 위한 부호화기는 대부분이 유성음을 위한 부호화기와 유사하지만, 무성음에서는 피치간 상관도가 매우 낮으므로 피치 보간 방법을 사용하지 않고 개루프로 피치 lag를 찾은 후 전체 프레임에 사용한다. 1 kb/s 부호화기는 묵음 구간과 주변소음 구간에 사용되며 이 구간의 신호를 피치 성분이 미약한 주변소음들로 제한하고 이에 최적인 부음성 부호화기를 설계하였다.

최종적으로 완성된 가변율 half rate 부호화기는 voice activity factor(VAF)가 0.47인 시험음성에서 약 2.6 kb/s의 평균 전송률을 보였다. 주관적 음질 평가의 일환으로 IS-96 표준 코덱인 가변율 8 kb/s QCELP와 A-B preference 시험을 실시하였나 시험 결과 평균전송률이 약 2배인 가변율 8 kb/s QCELP보다 우수한 음질 성능을 보였다.

I. 서 론

낮은 전송률 음성 부호화기는 크게 CELP형의 분석-합성 방법[1]과 하모닉 부호화 방법[2]으로 나눌 수 있는데, 본 논문에서는 generalized AbS구조[3]를 이용한 CELP형의 분석-합성 방법을 사용하였다. 본 논문의 구성은 다음과 같다. II장에서는 generalized AbS 구조에 대해 설명하며, III장에서는 설계된 half rate 음성 부호화기에 대해 설명하며 IV장에서는 Voicing 결정 알고리즘과 묵음 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 RCELP[4]에서 사용되었으며 피치 주기의 선형적인 보간에 의해 많은 비트를 아낄 수 있는 장점이 있으나 보간된 부분에서 발생하는 피치 주기의 작은 차이가 일반적인 AbS 메카니즘의 성능에 심각한 영향을 줄 수 있는 단점이 있다. 이를 보완하기 위해 Generalized AbS 구조는 다음의 세 단계로 이루어진다. 첫째, 피치 주기 contour를 결정하며, 둘째, 과거의 합성신호를 현재로 mapping하며, 셋째, 원신호의 warping을 수행한다.

1 피치주기 contour의 결정

피치주기는 정확한 피치 탐색 알고리즘에 의해 20 ms마다 구해진다. 피치 탐색을 위해 프레임 경계(boundary)에 window 중심을 맞추어 피치주기를 구하고, 구해진 피치주기를 사용하여 선형보간 되어진다. 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. 과거합성신호의 mapping

RCELP에서 피치주기는 샘플 단위로 선형보간 된다. 계산적인 이점을 위해 각 샘플에서의 피치주기는 샘플링 주기의 1/8로 반올림 되어진다. 그리고 과거의 재생

여기신호로부터 현재의 여기신호 샘플을 계산하기 위해 적당한 polyphase 보간 필터를 사용한다.

3. 원신호의 warping

원신호의 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 정합 되어져야 하며 시간변형은 원신호의 변화가 음질에 거의 영향을 주지 않는 범위내에서 실시한다. 원신호의 피치주기 contour를 변형하는 가장 좋은 방법은 잔여신호에 대해 연속적인 time-warp 작용을 수행하는 것이다. 이 방식을 사용하면 변형된 신호에 불연속점이 없어진다. 그러나 변형된 잔여신호와 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 경계에서 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.

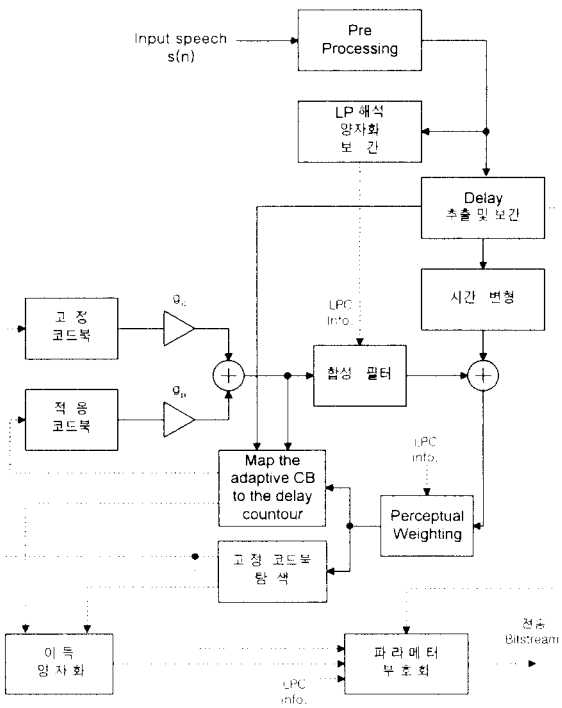

그림 1. Generalized AbS 구조를 이용한 4 kb/s 음성 부호화기의 구조

III. Half rate 음성 부호화기 설계

유성음을 위한 4 kb/s 음성 부호화기의 전체적인 구조는 그림 1과 같다. 그림 1에 의하면 원신호의 시간변형 블럭을 제외하면 일반적인 CELP 구조[5]와 유사함을 알 수 있다. 여기서 시간변형 부분은 II장에서 설명한 방식을 사용하였다.

1. 비트 할당 및 프레임 구조

입력된 신호는 8 kHz로 샘플링되며 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

파라메터	부프레임 1	부프레임 2	프레임당 비트수
LSPs	23		23
적응 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[6]로 변환되며, 이 값과 4차의 moving average (MA) 예측기를 통해 예측된 값과의 차이신호를 각각 128개와 32개의 코드워드값을 갖는 2단계의 VQ와 3 split VQ[7]를 통하여 총 23 비트로 부호화한다. 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 dimension을 갖는다. 부프레임 2은 현재 프레임에서 구해진 LP 계수를 그대로 사용하며 부프레임 1은 이전 프레임과 현재 프레임에서 구해진 두 LP 계수를 각각 25% 및 75%의 비율로 보간하여 사용한다.

$$\omega_i = 0.25\omega_i^{(previous)} + 0.75\omega_i^{(current)}, \quad i = 1,\ldots,10 \quad (15)$$

3. Low complexity algebraic 코드북의 설계

고정 코드북으로 사용되는 Low complexity 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 구성되어 지며 그 구조는 표 2에 나타나 있다.

부호화 시 첫 네 펄스들의 위치정보에 각각 3비트를 할당하고, 네 번째 펄스의 위치에 4비트를 할당하며, 부호정보에 각각 1비트씩을 할당하므로써 총 17비트가 할당된다. 모든 펄스들의 위치는 동시에 0 또는 1 샘플만큼 shift를 가능하게 하였다. 이러한 두 개의 shift값 중 합성에러를 최소화하는 최적 shift값을 구하고, 이렇게 구해진 shift값을 전송하기 위해 1비트를 사용한다.

표 2. Grid index = 0 일 때의 low complexity algebraic 코드북의 구조

펄스	부호	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

4. 이득값의 양자화

적응 코드북과 고정 코드북의 이득값 양자화는 G.729에서 사용된 방식을 이용하였다. 수 검색의 효율을 높이기 위해 conjugate 구조를 갖는 2개의 코드북을 사용하였다[8]. 각각의 코드북은 8개와 16개의 요소값을 가지며, 구해진 이득값들에 의해 마지막 4개와 8개의 요소값들에 대해서만 검색이 이루어진다.

5. 무성음을 위한 4 kb/s 음성 부호화기 설계

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

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[9]를 이용해 20 ms의 프레임 구조를 가지는 본 알고리즘에 맞게 새로이 설계하였다.

1. Voicing 결정 알고리즘

Voicing 결정 알고리즘은 가변율 음성 부호화기의 평균 전송률을 결정하는 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 활성도 측정을 위해 음성 프레임의 에너지를 이용하여 주변의 소음정도에 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 구분적인 성능 차이로 인한 부자연성을 억제하기 위한 VAD decision smoothing 과정의 수행을 통해 최종 전송률 결정을 하였다.

2. 1 kb/s 부음성 부호화기 설계

본 연구에서는 1 kbps의 전송률로 부호화될 주변소음의 범위를 피치 성분이 미약한 주변소음들로 제한하고 이에 최적인 부음성 부호화기를 설계하였다. 주변소음의 부호화에 사용되는 파라메터들은 단구간 상관도, 주파수의 파형(shape)을 결정짓는 LP 계수와 여기 신호의 만들에 사용되는 seed 값, 그리고 여기 신호의 크기를 나타내는 이득값으로 구성된다.

V. 실험 및 결과

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

표 3. LSP 양자화기의 성능

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

표 3은 사용된 LSP 양자화기의 성능으로서 outlier가

기준을 약간 벗어나지만 평균 SD는 transparent한 성능을 제공함을 알 수 있다. 그리고 제안된 가변율 half rate 음성 부호화기의 음질 성능을 평가하기 위해 잘 알려진 표준 코덱과 A-B preference 시험을 실시하였다. 이 시험에 사용된 표준 코덱으로는 현재 CDMA 디지털 이동통신 시스템에 사용중인 가변율 8 kb/s QCELP를 이용하였다. 대상 음성 샘플은 이동통신 환경에서 채취한 주변소음을 심각한 남성 및 여성 화자의 한국어 음성 샘플을 각각 3개씩 사용하였다. 10명의 청취자를 대상으로 시험을 실시하였으며, 표 4는 A-B preference 시험 결과를 보였다. 시험 결과를 보면 모든 경우에서 설계된 가변율 half rate 음성 부호화기가 더 나은 성능을 가짐을 알 수 있다. 그리고 표 5는 두 부호화기의 잡음 환경 및 비잡음 환경에서 시험한 평균전송률을 비교한 것이다. 설계된 가변율 half rate 음성 부호화기는 8 kb/s QCELP 부호화기와 비교할 때 더 나은 음질을 가지는데 반해 약 61%의 전송률을 보였다.

표 4. 설계된 가변율 half rate 음성 부호화기와 8 kb/s QCELP 부호화기의 A-B preference 시험 결과

	가변율 8 kb/s QCELP 부호화기 Bette	설계된 가변율 half rate 부호화기 Better	Same
남성화자	16.7 %	76.7 %	6.7 %
여성화자	23.3 %	63.3 %	13.3 %
전 체	20.0 %	70.0 %	10.0 %

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

잡음 환경	잡음정도[dB]	샘플수	전송률	
			가변율 8k 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
시리	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
평 균 전송률			4.30 kb/s	2.63 kb/s

VI. 결 론

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 파라메터화 하여 부호화하므로 피치 신호에 적은 비트를 할당할 수 있으므로 여분의 비트를 LSP 양자화기에 사용하였다. 이로 인하여 LSP 양자화기는 상당한 성능의 증가가 있었다. 그리고 무성음에서 성능이 저하되는 단점을 없애기 위해 voicing 결정 알고리즘을 사용하여 유, 무성음과 묵음으로 음성을 구분하였고 유성음과 묵음을 각기 다른 알고리즘으로 코딩하였다. 그리고 묵음을 위해 1 kb/s의 알고리즘을 설계하였다. 시험결과 전체적인 음질성능은 가변율 8 kb/s QCELP보다 우수하며 약 2.6 kb/s의 평균 전송률을 가지는 것으로 나타났다.

참 고 문 헌

[1] B.S. Atal "Predictive coding of speech signals at low bit rates," *IEEE Trans. Comm.* Vol. 30, No.4, 1982, pp. 600-614.

[2] 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.

[3] 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.

[4] 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.

[5] 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.

[6] 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.

[7] 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.

[8] 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.

[9] 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.