

PVQ를 이용한 RPE 구동시퀀스 양자화 연구

강상원, 조성호, 김해진
한양대학교

A Study on Excitation Sequence Quantization in RPE Speech Coding

Sangwon Kang, Sungho Cho and Haejin Kim
Hanyang Univ.

요약

본 논문에서는 RPE 음성부호화기에서 합성필터로 인한 구동벡터 양자화잡음의 증폭효과를 분석하고 regular pulse 시퀀스의 양자화로 인한 성능감소를 줄이기 위해 pyramid vector 양자화방식을 도입하였다. 제안된 방식의 성능평가는 구동시퀀스 양자화를 위해 adaptive PCM을 이용하는 GSM 표준 RPE 방식과의 객관적 및 주관적 성능비교를 통해 수행하였다. 성능비교 결과 제안된 방식은 대략 1 dB의 SNR 및 segmental SNR 값 증가를 가져왔고 또한 비공식 청취시험결과 평요도의 증가를 느낄 수 있었다.

1. 서론

다자를 이동통신시스템에서 기존의 음성전송품질(toll quality)을 그대로 유지하면서 시스템의 용량을 증대시키기 위한 핵심기술로서 음성부호화기는 매우 중요하다. 현재 표준화된 디지털이동통신용 음성부호화기의 종류로는 유럽의 GSM 표준인 13 kbps RPE-LTP(regular pulse excitation with long-term prediction)[1], 북미의 IS-54 TDMA 표준인 8 kbps VSELP(vector sum excited linear prediction)[2] 및 IS-96 CDMA 용 음성부호화기 표준인 QCELP(Qualcomm code excited linear prediction)[3]로 대변할 수 있다. 이러한 방식들에 대한 성능개선이 음질개선 및 시스템 용량개선 측면에서 그리고 도래할 개인통신시스템에의 적용을 위해 현재 선진각국에서 이루어지고 있다.

이동통신용 음성부호화기의 구조는 저속에서 고음질의 음성부호화를 위해 전형적인 파형부호화방식이나 보코더방식 대신 분석/합성구조를 갖는 혼합부호화방식이 사용된다. 대표적인 혼합부호화방

식은 code-excited linear prediction(CELP)[4], multi-pulse excitation(MPE)[5] 및 regular-pulse excitation(RPE)[6]들이 있다. 이러한 방식들은 음성신호의 short-term 상관도를 제거하기 위해 linear prediction coding(LPC)방식을 사용하며, 결과적인 잔여(residual) 신호를 분석/합성 방식을 이용한 지연결정(delayed decision)방식에 따라 양자화한다.

RPE 부호화기는 시변 디지털필터를 구동하기 위해 균등간격의 펄스시퀀스를 사용하는 방식으로 구동(excitation) 벡터의 양자화계수로는 벡터내 첫 펄스의 위치와 벡터내 모든 펄스의 크기값이 사용된다. 그래서 RPE 방식의 성능에 대한 한가지 중요한 제약조건은 구동펄스의 양자화이다. 본 논문에서는 구동시퀀스 양자화의 효과를 분석하고 합성필터의 큰 이득값이 양자화 잡음의 증폭을 야기시켜 전체 부호화기의 성능열화가 초래됨을 살펴본다. 그래서 이러한 성능열화를 줄이기위해 구동시퀀스를 효율적으로 양자화하는 pyramid vector 양자화기법을 도입한다. 음질에 대한 본 논문의 효과를 분석하기 위해 SNR/SEGSNR 값 비교와 비공식 청취시험을 이용하였다.

II. RPE 기본 시스템

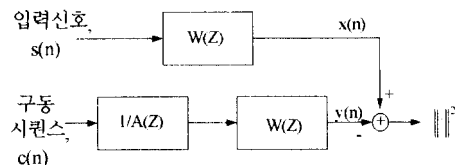


그림 1. RPE 부호화기의 기본구조도

울적으로 부호화함으로써 RPE 부호기의 양자화잡음을 크게 줄일 수 있다. 그런데 벡터크기가 13이며 벡터내 펄스당 3비트를 할당하는 RPE-LTP 방식에서 해당하는 벡터양자화기를 일반적인 Lloyd 알고리즘으로 설계하는 것은 복잡도 때문에 불가능하다. 그래서 본 논문에서는 매우 큰 부호화성능을 제공하면서도 낮은 복잡도를 갖는 벡터양자화방식인 pyramid vector quantization(PVQ)[9]을 사용하였다. Memoryless 소스에서 PVQ 부호화방식은 스칼라 양자화방식보다 훨씬 우수한 성능을 제공한다. 표 2는 memoryless Laplacian 소스에 대한 Lloyd-Max 스칼라 양자화기(SQ)와 PVQ의 SNR 성능을 나타내었다. RPE 구동시퀀스의 크기값은 흔히 Laplacian으로 모델화됨으로 PVQ는 RPE 구동펄스 크기를 부호화하는 효율적인 양자화기로서 양자화잡음을 크게 줄일 수 있을 것으로 사료된다.

표 2. Memoryless Laplacian 소스에 대한 최적스칼라 양자화기(SQ)와 PVQ의 SNR 성능

Rate(비트)	Lloyd-Max SQ (dB)	PVQ(dB)(dim=13)
1	3.01	5.07
2	7.54	9.93
3	12.64	15.27

PVQ는 pyramid의 표면에 존재하는 큐빅적자점들을 codeword로 사용하는 일종의 격자 양자화방식이다. 인코딩 및 디코딩 알고리즘이 쉽게 구현되며, 인코딩의 복잡도가 벡터 dimension에 선형으로만 증가하는 장점이 있다. Memoryless Laplacian 소스에 대한, PVQ 부호화기의 성능은 dimension이 클 경우 엔트로피 제한 스칼라 양자화기[10]와 비슷한 성능을 나타낸다. 이와같이 PVQ는 흔히 Laplacian으로 모델화되는 RPE 구동시퀀스 크기벡터값을 위한 효율적인 양자화방식으로 사용될 수 있다. PVQ 구현을 위한 복잡도가 그리 크지 않으므로 13 정도의 벡터 dimension은 펄스당 부호화율이 클 경우에도 쉽게 구현될 수 있다.

RPE 구동펄스의 벡터크기 L_p 는 흔히 10 이상이므로, 반정파라미터의 상대적인 편차는 무시될 수 있다. 그래서 다음과 같이 표현되는 하나의 단일 pyramid 상에 최적 codeword들을 설계해도 무관하다.

$$S(L_p, L_p/\lambda) = \{X : \sum_{i=1}^{L_p} |x_i| = L_p/\lambda\}, \quad (4)$$

여기서 $\lambda = 1/E\{|X|\}$. K 를 양의 정수라 할때 pyramid $S(L_p, K)$ 는 그것의 표면에 (codeword에 해당하는) 많은 큐빅적자점들을 가진다. 만약

$S(L_p, K)$ 상의 codeword 수를 $N(L_p, K)$ 라 할때 K 는 다음 조건을 만족하는 가장 큰 정수로 선택되어야 한다.

$$N(L_p, K) \leq 2^{m_r}$$

여기서 R 은 구동펄스당 평균비트수이다. 구동시퀀스 크기벡터 \underline{l} 에 대한 PVQ 부호화 알고리즘은 다음과 같다

1. 각 크기벡터 \underline{l} 에 가장 가까운 $\hat{\underline{l}} \in S(L_p, L_p/\lambda)$ 를 선택한다.
2. $\hat{\underline{l}}$ 를 $K \cdot \|\hat{\underline{l}}\|$ 로 스케일링하고 결과적인 벡터 $\hat{\underline{v}}$ 가 $S(L_p, K)$ 상에 있도록 한다.
3. $\hat{\underline{v}}$ 에 가장 가까운 정수점 codeword $\hat{\underline{c}} \in S(L_p, K)$ 를 다음과 같이 구한다.
 - a) $\hat{\underline{v}}$ 의 각 요소값을 가장 가까운 정수로 반올림해서 결과적인 벡터 $\hat{\underline{c}}$ 을 구한다.
 - b) $\|\hat{\underline{c}}\|$ 을 계산한다. 만약 $\|\hat{\underline{c}}\| = K$ 이면 step 4로 가라. 만약 $\|\hat{\underline{c}}\| > K$ 이면 가장 큰 에러를 야기하면서 앞서 위로 반올림된 $\hat{\underline{c}}$ 의 $(\|\hat{\underline{c}}\| - K)$ 개 nonzero 요소들을 하나씩 감소한다. 그리고 만약 $\|\hat{\underline{c}}\| < K$ 이면 가장 큰 에러를 야기하면서 앞서 아래로 반올림된 $\hat{\underline{c}}$ 의 $(K - \|\hat{\underline{c}}\|)$ 개의 요소들을 하나씩 증가시킨다.
4. $\hat{\underline{c}}$ 를 $\|\hat{\underline{c}}\|/K$ 로 스케일 한 후 PVQ 출력벡터 $\underline{\hat{c}}$ 를 구한다.

V. 실험결과

RPE 부호화기에서의 PVQ의 효율성이 시뮬레이션으로 평가되었다. 표 3은 성능비교에 이용된 5개의 한국어 문장들이다. 각 음성 문장들은 3.2 kHz로 bandlimited 되어진후 13비트 resolution으로 8 kHz로 샘플링되었다. 13 kbps로 동작하는 RPE-LTP 부호화기에 대한 파라미터 값들이 표 4에 보여졌다. 구동시퀀스와 long-term prediction(LTP)을 위한 부프레임 길이는 5ms이며, 각 구동시퀀스는 13개의 regular pulse로 구성된다. 이러한 regular pulse 크기는 해당 부프레임내 최대 펄스값으로 normalized 된 후 adaptive PCM(APCM)방식으로 양자화된다. 본 논문에서는 GSM 표준안에서 regular pulse 양자화를 위해 사용되는 APCM 방식대신 앞서 제안한 PVQ 방식을 적용하여 성능을 비교해 보았다. 성능은 SNR/SEGSR에 의한 객관적인 방식과 비공식 청취시험을 통해 수행하였다.

RPE 음성부호화기의 기본구조가 그림 1에 보여졌다. P 차 LPC 계수값 $\{a_k, k=1, 2, \dots, p\}$ 들이 먼저 구해진후 $\{a_k, k=1, 2, \dots, p\}$ 으로 부호화되어 전송된다. 음성의 short-term 상관도를 모델하는데 사용되는 합성필터는 $1/A(Z)$ 이다. 여기서 $A(Z)$ 는 $1 - \sum_{k=1}^p a_k z^{-k}$ 로 표시된다. [7]에서 언급된 바와같이 weighting 필터 $W(Z)$ 는 사람 귀의 주파수 masking 효과를 이용함으로써 양자화잡음의 loudness를 줄이는데 사용된다. 이러한 필터는 다음과 같이 선택된다.

$$W(Z) = \frac{A(Z)}{A(Z/\gamma)} = \frac{1 - \sum_{k=1}^p a_k z^{-k}}{1 - \sum_{k=1}^p a_k \gamma^k z^{-k}} \quad (1)$$

여기서 γ 는 0과 1 사이의 인치 weighting 요소이다. [6]에서와 같이 RPE 시스템은 음성을 L 샘플의 프레임으로 나눈후 각 프레임에 대해 regular pulse 패턴을 선택한다. 허용되는 펄스패턴은 각 펄스 사이에 NS-1개의 '0' 값을 갖는다. 각 프레임내 첫번째 펄스의 위치는 첫 NS개의 위치중 하나이므로 NS개의 regular pulse 패턴이 존재한다. RPE 시스템의 인코딩과정을 통해 프레임내 첫번째 펄스의 위치와 각 펄스의 크기가 결정된다. 만약 weighting 된 입력 신호를 $x(n)$ 이라 하고 weighting 된 합성신호를 $y(n)$ 이라 할때 $x(n)$ 과 $y(n)$ 과의 차의 자승값을 구동시퀀스를 찾기위한 distortion measure로 사용한다. NS개의 regular pulse 패턴에 대해, 구동선택 알고리즘의 왜곡(distortion)을 최소화하는 최적 벡터값을 계산한다. 그때 최소 왜곡을 제공하는 펄스패턴(\mathbf{b} 로 표시)을 선택하고 부호화된 offset과 양자화된 펄스의 크기들이 전송 혹은 저장된다. 디코딩은 양자화된 구동시퀀스로 합성필터를 필터링하는 과정을 포함한다.

III. 양자화 잡음의 증폭

그림 1에 보여진 기본구조에서 스펙트럼 weighted 합성신호는 다음과 같이 계산된다.

$$y(n) = \sum_{k=1}^p a_k \gamma^k y(n-k) + c(n) = \hat{y}(n|n-1) + c(n) \quad (2)$$

여기서 $\hat{y}(n|n-1)$ 는 필터출력의 예측값이다. 이때 다음과 같이 L 샘플의 한 프레임에 대한 mean squared error(MSE)를 최소화하는 것이 요구된다.

$$\begin{aligned} & \sum_{n=n_0}^{n_0+M-1} (x(n) - y(n))^2 \\ &= \sum_{n=n_0}^{n_0+M-1} (x(n) - \hat{y}(n|n-1) - c(n))^2 \\ &= \sum_{n=n_0}^{n_0+M-1} (\hat{x}(n|n-1) - c(n))^2 \quad (3) \end{aligned}$$

여기서 $\hat{x}(n|n-1) (= x(n) - \hat{y}(n|n-1))$ 는 예측에러에 해당한다. 만약 구동시퀀스 $c(n)$ 이 RPE 포맷(벡터 \mathbf{c} 로 표시)으로 제한될때, $\mathbf{c} = \mathbf{b}$ 는 식(3)을 최소화하는 최적시퀀스이다. 이 방식이 갖는 한가지 문제점은 \mathbf{b} (혹은 \mathbf{c})의 양자화로 인한 양자화잡음이 디코더에서의 합성필터에 의해 증폭이 된다는 점이다. 그래서 비록 RPE 시퀀스 크기값들이 자승에러값을 최소화하도록 최적으로 선택된다 할지라도 이러한 크기값들은 양자화로 인한 에러로 suboptimum 한 값이 되며, 이러한 suboptimality는 합성필터의 예측이득으로 합성음성의 음질에 크게 영향을 끼치게 된다. 일반적으로 합성필터의 예측이득은 매우크다 (특히 유성음 부분).

표 1은 구동 시퀀스 양자화 잡음에 대한 평균 합성필터이득을 보였다. 표에서 보여진 평균이득값을 구할때 일반화된 Lloyd 알고리즘[8]을 이용하여 설계된 스칼라 양자화기가 구동시퀀스 크기를 양자화하는데 사용되었다. 대략 12.5 dB의 양자화 잡음이득은 합성음성의 품질에 심각한 열화를 야기시킨다.

표 1 양자화 잡음의 평균 합성필터이득

문장	평균이득(dB)
1	12.7
2	13.1
3	11.8
4	11.5
5	13.2

합성필터 계수는 입력신호의 선형예측 해석에 의해 결정됨으로 합성필터의 이득은 음성의 특성이다. 따라서 합성음성의 양자화 잡음을 줄이기위해 구동시퀀스의 양자화를 개선할 필요가 있다. 개선된 구동시퀀스 양자화방식을 위해 pyramid vector 방식을 다음절에 소개한다.

IV. Pyramid Vector Quantization

현재 표준화 되어있는 13 kbps RPE-LTP에서는 구동시퀀스의 양자화를 위해 펄스당 3비트의 adaptive PCM(APCM)방식을 사용하고 있지만 벡터 양자화를 이용해서 구동시퀀스의 크기벡터값을 효

표 3. 성능분석을 위해 사용된 한국어 문장들

1	미는 피부 한 겹질 차이입니다	(여자)
2	지나친 흡연은 건강을 해칩니다.	(여자)
3	이번 겨울은 예년과 달리 포근합니다.	(남자)
4	과학기술은 경제발전의 원동력이다.	(남자)
5	일에서 십까지의 함은 오십오입니다	(남자)

표 4. 13 kbps RPE-LTP 의 파라미터 값

파라미터	비트
o 구동시퀀스	47 비트/5ms
- grid 위치	(2)
- block 크기	(6)
- 펄스 크기	(39)
o Reflection 계수	36 비트/20ms
o LTP 계수	9 비트/5ms
- lag	(7)
- gain	(2)

PVQ 에 의한 구동시퀀스 양자화를 도입한 13 kbps RPE-LTP 의 SNR/SEGSNR 성능이 표 5 에 보여졌다 비교를 위해 APCM 을 이용하는 전형적인 13 kbps RPE-LTP 의 성능도 나타내었다. 구동시퀀스만 양자화하지 않을 때의 성능도 구동시퀀스 부호화에 대한 최대한치라는 의미로 포함하였다

표 5. PVQ 에 의한 구동시퀀스 양자화를 도입한 13 kbps RPE-LTP 의 성능(SNR/SEGSNR (dB))

문장	APCM	PVQ	Unquantized
1	18.24/17.59	19.33/18.43	21.45/20.75
2	17.90/17.21	18.67/18.31	21.13/20.82
3	19.31/18.53	20.13/19.70	22.67/22.12
4	16.27/15.65	17.15/16.74	19.87/19.44
5	17.32/16.95	18.29/17.75	20.33/19.84
평균	17.81/17.19	18.71/18.18	21.09/20.59

표 5 에서와 같이 PVQ 구동시퀀스 부호화방식을 도입함으로써 SNR 및 SEGSNR 값이 대략 1 dB 증가하였으며, 이러한 성능향상은 비공식 청취시험을 통해서도 느낄 수 있었다.

VI. 결론

RPE-LTP 방식은 현재 이동통신 등에서 사용되고 있는 주요한 음성부호화기중 하나이다. 그런데

RPE 방식에서는 구동시퀀스를 부호화 함으로서 야기되는 양자화 잡음이 합성필터에 의해 증폭된다. 이러한 증폭에 의한 성능열화를 줄이기 위해 구동시퀀스를 효율적으로 양자화하는 다차원 부호화기법으로 PVQ 방식을 제안하였다. 제안된 방식의 성능은 구동시퀀스 양자화기로서 APCM 을 이용하는 전형적인 13 kbps RPE-LTP 와의 객관적 및 주관적 성능비교를 통해 수행하였다. 성능비교결과 PVQ 방식은 대략 1 dB 의 SNR 및 SEGSNR 값 향상을 가져왔고 또한 비공식 청취시험결과 명료도의 증가를 느낄 수 있었다.

참고문헌

- [1] ETSI, "GSM full-rate speech transcoding," September 1989
- [2] EIA/TIA, "Cellular system dual-mode mobile station-base station compatibility standard," March 1991.
- [3] EIA/TIA, "Speech service option standard for wide-band spread spectrum digital cellular system," April 1994
- [4] M. R. Schroeder and B.S. Atal, "Code-excited linear prediction(CELP): high-quality speech at very low bit rates," in Proc IEEE Int. Conf. Acoust., Speech, Signal Processing, vol.3, pp.937-940, March, 1985.
- [5] B. S. Atal and J. R. Remde, "A new model of LPC excitation for producing natural-sounding speech at very low bit rates," in Proc IEEE Int. Conf. Acoust., Speech, Signal Processing, pp. 614-617, Apr. 1982.
- [6] P. Kroon, E. F. Deprettere, and R. J. Sluyter, "Regular-pulse excitation: A novel approach to effective and efficient multipulse coding of speech," IEEE Trans. Acoust., Speech, Signal Processing, vol. ASSP-34, pp 1054-1063, Oct. 1986
- [7] B. S. Atal, "Predictive coding of speech at low bit rates," IEEE Trans. Commun. vol. COM-30, pp.600-614, April 1982
- [8] Y. L. Linde, A. Buzo, and R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans. Commun., vol COM-28, pp 84-95, Jan. 1980.
- [9] T. R. Fischer, "A pyramid vector quantizer," IEEE Trans Inform Theory, vol IT-32, pp.568-583, July 1986
- [10] W. A. Finamore and W. A. Pearlman, "Optimal encoding of discrete-time continuous-amplitude memoryless source with finite output alphabets," IEEE Trans. Inform Theory, vol. IT-26, pp 144-155, Mar 1980