

격자벡터양자화기를 이용한 음성신호의 LSP 주파수 양자화

正會員 강 정 원*, 정 재 호**, 정 대 권***

Quantization of Line Spectrum Pair Frequencies using Lattice Vector Quantizers

Jung-Won Kang*, Jae-Ho Chung**, Dae-Gwon Jeong*** *Regular Members*

요 약

이 논문에서는 선 스펙트럼 계수(LSP: Line Spectrum Pair)를 양자화하기 위하여 2가지 종류의 양자화기를 설계하였다. 첫번째로 설계한 양자화기는 벡터양자화기와 격자벡터양자화기를 결합한 2단 벡터양자화기로 4가지 종류의 격자벡터양자화기를 적용하여 각각의 격자벡터양자화기의 성능을 비교하였다. 그리고 두번째로 설계한 양자화기는 VDPCM 구조에서 예측계수가 시간축상과 주파수축상에서의 예측오차를 최소로 하도록 설계된 양자화기로서, 낮은 복잡도로 인하여 실시간 구현이 가능한 격자벡터양자화기를 사용하여 선 스펙트럼 계수를 양자화하였다. 이 구조의 양자화기는 QCELP에 사용된 스케일러 양자화기에 비해 프레임당 10 bit 정도 부호화율을 줄일 수 있음을 보였다. 한편, 2단 벡터양자화기는 36 bits/frame 이하에서는 구형의 의사균일 격자벡터양자화기의 성능이 전체적으로 좋았고, 37 bits/frame 이상에서는 피라미드형의 의사균일 격자벡터양자화기가 평균 주파수 왜곡 0.97 dB로 가장 좋은 성능을 나타내었다.

ABSTRACT

Two different low rate speech coders using one of four types of lattice vector quantizers(LVQ's) with fairly low complexity were investigated for an application to mobile communications. More specifically, two-stage vector

*미국 조지아 전기 및 컴퓨터공학과
Dept. of Electronical & Computer Eng., Georgia Institute of Technology.

**인하대학교 전자공학과
Dept. of Electronic Eng., Inha Univ.

***한국항공대학교 항공전자공학과
Dept. of Avionics Eng., Han-kuk Aviation Univ.
論文番號: 95213-0614

接受日字: 1995年 6月 14日

quantizer-lattice vector quantizer(VQ-LVQ) systems and vector differential pulse code modulation(VDPCM) systems with lattice vector quantizers were simulated to encode the line spectrum frequencies of various sentences at the rate 22 to 39 bits per 20 msec frame.

The simulation results showed that the VDPCM system with the lattice VQ can save up to 10 bits/frame compared to the quantization scheme used in QCELP system. For the VQ-LVQ system, the spherical quasi-uniform LVQ below 36 bits/frame outperformed the other 3 types of LVQ's and the pyramidal quasi-uniform LVQ at 37 bits/frame outperformed the other 3 types of LVQ's with the spectral distortion 0.97.

I. 서 론

최근에 저전송속도 음성부호화 방식 가운데 하나인 분석 및 합성 방법에 관한 연구가 활발히 진행되고 있다. 이러한 방식에서 가장 많이 사용되는 계수인 선형예측 계수(LPC: Linear Predictive Coding)에 대한 연구 또한 관심이 집중되고 있다. 선형예측 계수는 음성신호의 포먼트 성분을 예측하므로 포먼트 예측 계수라고도 하는 것으로, 일반적으로 이를 직접 양자화할 경우 왜곡이 크기때문에 직접 양자화하지 않고, 선형예측 계수와 등가적의미를 가지나 양자화 성능이 우수한 선 스펙트럼 계수로 변환하여 양자화한다.

선 스펙트럼 계수의 스케일러 양자화에 관해서는 많은 연구결과가 이미 발표되어, 프레임당 32-40 bit에서 1 dB 이하의 평균 주파수 왜곡을 얻을 수 있음을 보이고 있다⁽¹⁾⁽²⁾. 벡터양자화도 활발히 연구되어 하이브리드 벡터-스케일러 양자화를 사용하여 30-32 bit에서 1 dB 정도⁽³⁾, 캐스케이드형 벡터양자화로 26 bit에서 1.1 dB의 성능을 보이고 있으며⁽⁴⁾, 분리형 벡터양자화와 다단계 벡터양자화를 통해서도 거의 유사한 성능을 보이고 있다⁽⁵⁾. 한편, 근래에 발표된 Fischer의 논문에서는 벡터양자화기와 구형 균일 격자 벡터양자화기를 결합한 2단 벡터양자화기로 24 bit에서 1 dB이하의 평균 주파수 왜곡을 보이고 있다⁽⁶⁾.

선 스펙트럼 계수의 양자화 성능을 나타내는 주파수 왜곡(SD: Spectral Distortion)은 다음과 같이 정의된다.

$$SD = \left\{ \frac{1}{N_f} \sum_{n=1}^{N_f} \left(\frac{100}{\pi} \int_0^{\pi} [\log_{10} |A_n(e^{j\omega})|^2 - \log_{10} |\hat{A}_n(e^{j\omega})|^2]^2 d\omega \right)^{1/2} \right\} \quad (1)$$

여기서 N_f 는 전체 프레임 수를, $A_n(e^{j\omega})$ 와 $\hat{A}_n(e^{j\omega})$ 는 각각 n 번째 프레임을 양자화하기전과 양자화한 뒤의 스펙트럼을 나타낸다. 일반적으로 위와 같이 표현되는 평균 주파수왜곡이 1 dB이고, 2~4 dB가 2% 미만, 4 dB이상일 없을 때 원음과 같은 수준의 합성음을 얻을 수 있다고 알려져있다.

이 논문에서는 두가지 종류의 양자화기를 설계하였다. 하나는 벡터양자화기와 격자벡터양자화기를 결합한 2단 벡터양자화기로 벡터양자화 오차를 격자벡터양자화하는 구조이고, 다른 하나는 기본적으로 VDPCM 구조를 이용해서 예측계수를 시간축상과 주파수축상에서의 예측오차를 최소화 하도록 설계된 양자화기로, 예측오차를 격자벡터양자화하는 구조이다. 사용된 격자벡터양자화기는 구형 균일 격자벡터양자화기, 피라미드형 균일 격자벡터양자화기, 구형 의사균일 격자벡터양자화기, 피라미드형 의사균일 격자벡터양자화기 4가지로 각각의 경우에 있어서 한글 문장에 대해 양자화하여 비교하였다.

이 논문은 다음과 같은 구조로 되어 있다. 제 II 절에서 선형예측 계수와 선 스펙트럼 계수에 관해 고찰하였고, 제 III 절에서 모의실험에 사용된 벡터양자화기와 4가지 격자벡터양자화기에 대해서, 그리고 제 IV 절에서는 이 논문에서 설계한 두가지 양자화기의 구조에 대해 살펴보고, 제 V 절에서 모의실험에 관해 기술하였으며, 끝으로 제 VI 절에서 결론을 맺고 있다.

II. 선형예측 계수와 선 스펙트럼 계수

음성 신호 모델중 선형예측 부호화 방식은 많은 연산량 없이 음성 스펙트럼을 매우 효과적으로 나타내 준다. 선형예측 부호화 방식은 유한개의 계수로 성도

응답의 정보를 나타내며, 극점만을 사용하는 AR 모델로 전달함수는 식 (2)와 같다⁽¹⁾⁽²⁾.

$$H(z) = \frac{G}{A_p(z)} = \frac{G}{1 - \sum_{i=0}^p \alpha_i z^{-i}} \quad (2)$$

여기서 p 는 예측차수이며, $\{\alpha_i\}$ 는 선형예측 계수이다.

그러나 선형예측 계수는 음성신호를 효과적으로 나타내 주지만 각각의 계수, $\alpha_i (1 \leq i < p)$ 를 양자화하는데 10 bit미만이 사용될 경우 주파수 왜곡이 상당히 커지고 시스템이 종종 불안정해진다. 이를 방지하기 위해 선형예측 계수를 선 스펙트럼 계수로 변환한 뒤, 선 스펙트럼 계수를 양자화한다. 그것은 선 스펙트럼 계수가 양자화로 인한 주파수 왜곡이 적으며, 순차성이 유지되는 한 시스템의 안정성을 보장받을 수 있다는 특징 때문이다. $A_p(z)$ 를 예측필터라 할때 식 (3)과 (4)에서 $\omega_i (1 \leq i < p)$ 가 선 스펙트럼 계수이다. 이 선 스펙트럼 계수가 식 (5)의 순차성을 만족시키면 예측 필터 $A_p(z)$ 는 안정하다.

$$P(z) = A_p(z) - z^{-(p+1)} A_p(z^{-1}) \\ = (1 - z^{-1}) \prod_{i=1, \text{even}}^p (1 - 2z^{-1} \cos \omega_i + z^{-2}) \quad (3)$$

$$Q(z) = A_p(z) + z^{-(p+1)} A_p(z^{-1}) \\ = (1 + z^{-1}) \prod_{i=1, \text{odd}}^p (1 - 2z^{-1} \cos \omega_i + z^{-2}) \quad (4)$$

$$0 = \omega_0 < \omega_1 < \omega_2 \dots < \omega_{p-1} < \omega_p < \omega_{p+1} = \pi \quad (5)$$

III. 벡터양자화기와 격자벡터양자화기

III-1 벡터양자화기

일반적으로 L 레벨 k 차원 벡터양자화기 Q 는 유클리드 공간 R^k 로부터 R^k 상의 유한 부분 집합 C 로의 사상으로 정의된다.

$$Q: R^k \rightarrow C$$

여기서 흔히 부호책이라 부르는 $C = \{Y_i | 1 \leq i \leq L\}$ 는 출력벡터 Y_i 의 집합이며, L 은 출력벡터의 수 또는 부

호의 수이다⁽⁸⁾. 입력벡터 $X = \{x_0, x_1, x_2, \dots, x_{k-1}\}$ 에서 Y 로의 사상 곧, 양자화의 성능을 평가하는데는 여러가지 왜곡 척도가 쓰이고 있으나, 가장 많이 쓰이며 가장 단순한 왜곡척도는 제곱 오차로서 다음과 같이 주어진다.

$$d(X, Y) = \|X, Y\|^2 = \sum_{i=0}^{k-1} (x_i - y_i)^2 \quad (6)$$

2단 벡터양자화기의 첫단의 VQ설계에 사용된 왜곡 척도는 음성부호기에 흔히 사용되는 가중 제곱 오차인 IHM(Inverse Harmonics Mean) 오차를 사용하였으며, 구체적인 표현식은 다음과 같다⁽⁹⁾.

$$d(X, Y) = \sum_{i=0}^{k-1} w_i (x_i - y_i)^2 \quad (7)$$

여기서 $w_i = \left\{ \frac{1}{x_i - x_{i-1}} + \frac{1}{x_{i+1} - x_i} \right\}^2$ 으로 정의되며, 인접한 선 스펙트럼일수록 스펙트럼의 피크가 나타나기 때문에 가중치가 증가한다.

이 논문에서는 LBG 알고리즘을 사용하여 국부적으로 최적화된 부호책을 만들었으며, 초기 부호책은 분산 알고리즘을 적용하였다⁽¹⁴⁾.

III-2. 격자벡터양자화기

벡터양자화기가 부호책을 저장해야 하는 반면 격자벡터양자화기는 단지 부호책의 크기 정보만 저장하므로 부호화시에 모든 가능한 출력부호와 비교하지 않고 손쉽게 출력벡터를 계산해 낼 수 있다.

일반적으로 격자는 식 (8)과 같이 정의되는 n -차원 벡터들의 집합이다.

$$A = \{x | x = c_1 a_1 + c_2 a_2 + \dots + c_n a_n\} \quad (8)$$

여기서 a_i 는 격자의 기본벡터, c_i 는 정수, n 은 차원을 나타낸다.

이 논문에서 사용된 격자벡터양자화기는 정수벡터를 기본벡터로 하는 정수격자를 사용하였으며, 이러한 무한개의 격자중 원하는 비트수와 차원에 의해 일정한 수의 격자를 출력벡터로 선택하여 각각의 출력벡터 사이의 경계를 설정하였다.

격자벡터양자화기의 형태는 가우스 신호에 적합한

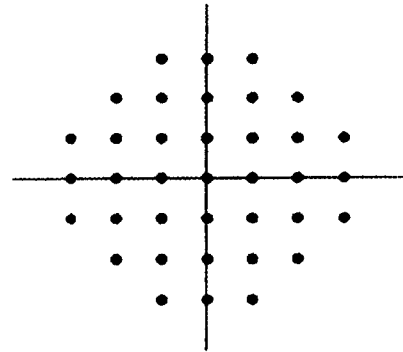
구형 벡터양자화기와 라플라스 신호에 적합한 피라미드형 벡터양자화기가 있으며, 격자의 균일성에 있어서는 전체적으로 균일한 격자를 갖는 균일 격자벡터양자화기와 국부적으로는 균일격자를 갖지만 전체적으로는 비균일격자를 갖는 의사균일 격자벡터양자화기가 있다⁽¹⁰⁾.

1. 균일 격자벡터양자화기

1.1 정수 부호책 설계

평균 0, 분산 1이고, 독립적이며 동일한 분포를 갖는 라플라스 신호의 벡터는 동일한 확률 밀도함수의 궤적이 피라미드 형태를 나타내므로 라플라스 입력원에 적합한 부호책은 피라미드 형태를 가져야 한다. 부호책의 크기는 피라미드의 크기를 의미하며, 이 피라미드 내부와 표면에 존재하는 모든 정수격자가 부호로 정해진다. 곧, 차원당 비트수와 벡터양자화기의 차원에 따라 필요로 하는 전체 부호 수가 결정되고, 이러한 정수격자를 포함하는 피라미드를 부호책으로 정하는 것이다⁽¹⁰⁾.

평균 0, 분산 1이고, 독립적이며 동일한 분포를 갖는 가우스 신호의 벡터는 동일한 확률 밀도함수의 궤적이 구형을 나타내므로 가우스 입력원에 적합한 부호책은 구형을 가져야 한다. 구형 정수 부호책 내의 정수벡터의 갯수는 수론에 의해 설명되는데, 원점으로부터의 거리의 제곱이 임의의 정수가 되는 정수벡터의 갯수가 오래전부터 연구되어왔다. 이 논문에서는 이를 이용하여 주어진 부호화율과 벡터의 차원에 따라 구형 부호책을 키워나가는 방식으로 부호책을 설



(b)

그림 1. (a) 2차원 피라미드형 격자벡터양자화기

(b) 2차원 구형 격자벡터양자화기

Fig. 1 (a) Laplacian Z^2 lattice codebook.

(b) Gaussian Z^2 lattice codebook

계하였다⁽¹¹⁾.

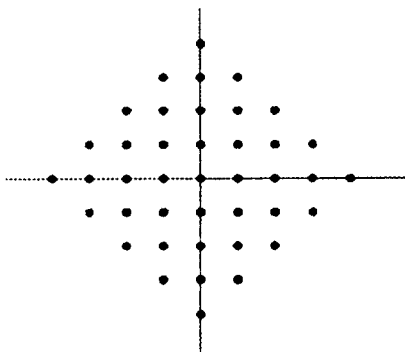
그림 1은 2차원 피라미드형 격자부호책과 구형 격자부호책의 보기이다.

1.2 스케일 계수 최적화

최적의 부호책 곧, 주어진 입력신호를 최소의 왜곡으로 부호화하는 부호책을 설계하기 위해서는 입력원의 통계적 특성에 따라 부호거리의 거리가 최적이 되도록 정수격자부호책을 조정하여야 한다.

2. 의사균일 격자벡터양자화기

의사균일 격자벡터양자화기의 부호책은 원점을 중심으로 여러개의 균일격자 벡터양자화 부호책이 겹겹이 싸여 있는 형태로 이루어져 있다. 각각의 균일 격자부호책내의 격자거리는 일정하고, 부호책이 원점에서 멀어 질수록 격자거리가 늘어난다. 이 논문에서는 인접하는 부호책의 격자거리가 특정한 계수의 곱에 의해서 증가한다고 가정하였다. 곧, 전체부호책은 격자거리가 c_0, sc_0, s^2c_0, \dots 와 같이 증가하는 부호책들로 이루어진다. 여기서 c_0 는 원점을 포함하고 있는 부호책의 최적 격자거리이며, s 는 압신계수로서 격자거리의 증가율에 해당되는 것으로서 원점에서 멀어질수록 격자거리가 증가하므로 $s > 1$ 이 된다. 그림 2는 2차원 피라미드형 의사균일 격자벡터양자화



(a)

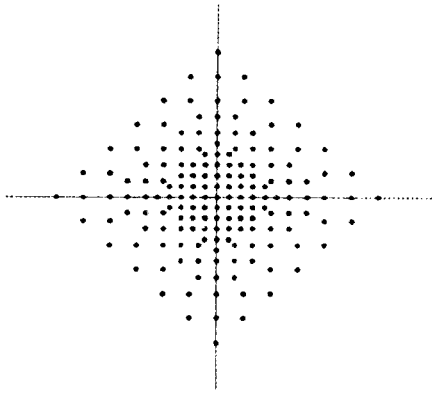


그림 2. 2차원 피라미드형 의사균일 격자벡터양자화기
Fig. 2 Laplacian quasi-uniform Z^2 lattice codebook

기의 보기이다.

다음 표는 최소왜곡을 갖기 위한 차원당 R bit의 부호화율을 갖는 피라미드형 n -차원 의사균일 격자벡

터양자화기와 구형 n -차원 의사균일 격자벡터양자화기의 부호책 파라미터들을 보여주고 있다⁽¹⁰⁾.

IV. 선 스펙트럼 계수의 양자화 방법

IV-1. 2단 벡터양자화

설계된 2단 벡터양자화기의 구조는 그림 3과 같다. 2단 벡터양자화기에서 얻고자하는 출력은 n 번째 프레임의 선 스펙트럼 계수 $w(n)$ 이 벡터양자화기를 거쳐 양자화된 $\tilde{w}(n)$ 과 첫단의 벡터양자화 오차 $e(n)$ 을 격자벡터양자화한 $\hat{e}(n)$ 이다. 수신단에서 합성할 때는 두 양자화된 값들을 서로 더한 값으로 합성하게 된다.

Fischer의 경우 같은 VQ-LVQ의 구조를 가지나, 선형예측 계수를 대역폭 확장시키고, 벡터양자화 오차에 적절한 조정 계수를 가하는 것에 의해 프레임당 24 bit에서 1 dB의 평균 주파수 왜곡을 얻었다. 그러나 이 논문에서 사용된 2단 벡터양자화기는 대역폭

표 1. n -차원 피라미드형 의사균일 격자벡터양자화기의 부호책 파라미터

Table 1. Codebook parameters for n -dimensional Pyramidal quasi-uniform LVQ

부호책 파라미터	
1번째 부호책의 최적격자거리	$c_0 = 2^{-R+0.5}(n+2) \log s \left[\frac{(s^n-1) \sum_{k=0}^n (ks^{-k})^n}{n!} \right]^{1/n}$
k 번째 부호책의 최적격자거리	$s^k c_0$
k 번째 부호책의 외각반경(l_1^*)	$r_k = \left[\frac{(n+2) \log s}{\sqrt{2}} \right] k$

* l_1 는 절대거리를 의미한다.

표 2. n -차원 구형 의사균일 격자벡터양자화기의 부호책 파라미터

Table 2. Codebook parameters for n -dimensional Spherical quasi-uniform LVQ

부호책 파라미터	
1번째 부호책의 최적격자거리	$c_0 = 2^{-R+0.5} \sqrt{\pi(n+2)} \log s \left[\frac{(s^n-1) \sum_{k=0}^n (k^{1/2} s^{-k})^n}{\Gamma(n/2+1)} \right]^{1/n}$
k 번째 부호책의 최적격자거리	$s^k c_0$
k 번째 부호책의 외각반경(l_2^*)	$r_k = \sqrt{(n+2) \log s} s \sqrt{k}$

* l_2 는 유클리드 거리를 의미한다.

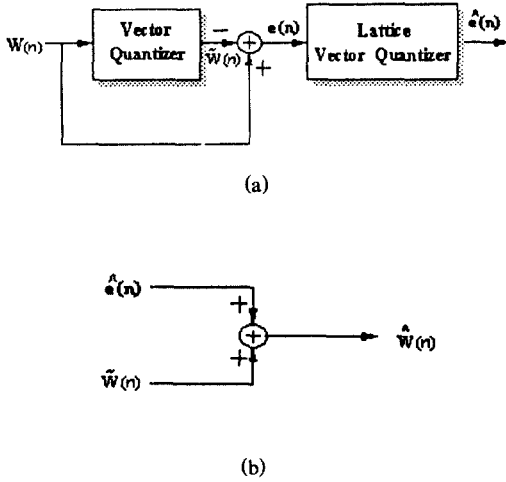


그림 3. 2단 벡터양자화기의 블럭도 (a) 부호기 (b) 복호기
Fig. 3 Block diagram of two-stage vector quantizer (a) encoder (b) decoder

확장이나 조정 계수와 같은 것은 고려하지 않았다

IV-2. VDPCM 오차의 격자벡터양자화

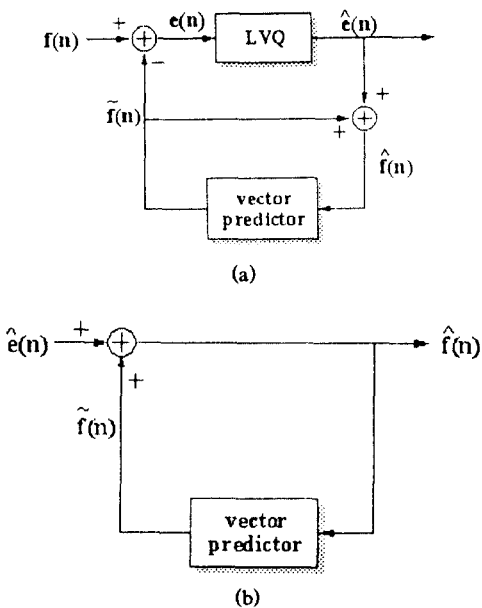


그림 4. VDPCM 양자화기의 구조 (a) 부호기 (b) 복호기
Fig. 4 Block diagram of VDPCM (a) encoder (b) decoder

선 스펙트럼 계수는 프레임사이에 곧, 시간축상으로 상관관계를 가지고 있을뿐 아니라 같은 프레임내에서 곧, 주파수축상에서도 상관관계를 가지고 있다(그림 5)⁽¹²⁾. 그러므로 선 스펙트럼 계수는 이러한 상관관계를 이용하여 과거의 선 스펙트럼 계수값들에 의해 현재의 계수값을 예측할 수 있다는 장점을 가진다. 이러한 장점을 살린 것이 그림 4에서와 같은 VDPCM구조를 가지고 프레임간에, 그리고 프레임내에서의 상관관계에 의해 과거의 계수들로 부터 예측한 값과 실제값사이의 예측 오차를 양자화하는 방법이다.

이 논문에서는 그림 5에서 보는 바와 같이 n번째 프레임의 i번째 선 스펙트럼 계수 $f_i(n)$ 을 예측하기 위하여 n번째 프레임의 i-1번째 선 스펙트럼 계수 $f_{i-1}(n)$, n-1번째 프레임의 i번째 선 스펙트럼 계수 $f_i(n-1)$, n-1번째 프레임의 i+1번째 선 스펙트럼 계수 $f_{i+1}(n-1)$ 을 사용하였다. 그리고 이 때 예측을 위하여 사용된 예측계수를 각각 a_i, b_i, c_i 라고 하였다. 위와 같은 관계를 식으로 표현하면 다음과 같다.

$$\tilde{f}_i(n) = a_i \tilde{f}_{i-1}(n) + b_i \tilde{f}_i(n-1) + c_i \tilde{f}_{i+1}(n-1) \quad (9)$$

$$\hat{f}_i(n) = \tilde{f}_i(n) + \hat{e}_i(n), \quad i = 1, 2, \dots, p \quad (10)$$

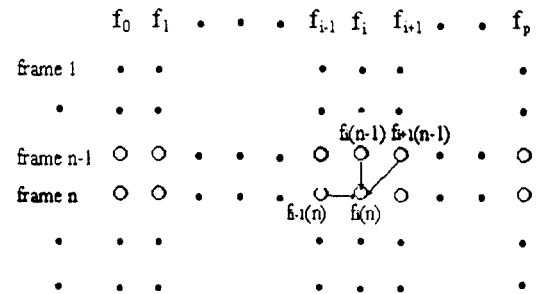


그림 5. 연속적인 프레임안에서 선 스펙트럼 계수들의 관계
Fig. 5 LSP parameters in consecutive frames

여기서 $\tilde{f}_i(n)$ 은 $f_i(n)$ 을 예측한 값이고, $\hat{f}_i(n)$ 는 $f_i(n)$ 을 양자화한 값이며, $\hat{e}_i(n)$ 은 예측오차를 격자벡터양자화한 값이다. 또한 a_i, b_i, c_i 는 예측계수로 식 (11)에 의해 표현되는 예측오차의 자승을 최소화하는 것(식

(12))에 의해 얻을 수 있다. 여기서 N_f 는 전체 프레임의 수이다.

$$E_i = \sum_{n=1}^{N_f} [r(n)]^2 = \sum_{n=1}^{N_f} [f_i(n) - \tilde{f}_i(n)]^2 \quad (11)$$

$$\frac{\delta E_i}{\delta a_i} = 0, \quad \frac{\delta E_i}{\delta b_i} = 0, \quad \frac{\delta E_i}{\delta c_i} = 0 \quad (12)$$

식 (11)과 식 (12)에 의해 얻은 예측계수 a_i, b_i, c_i 는 다음과 같이 표현되며

$$a_i = \frac{A}{B} \quad (13-1)$$

$$b_i = \frac{B \cdot \mu_{i,i-1}(0) - A \cdot \sigma_{i-1}^2(0) - C \cdot \mu_{i+1,i-1}(1)}{B \cdot \mu_{i,i-1}(1)} \quad (13-2)$$

$$c_i = \frac{C}{B} \quad (13-3)$$

식 (13)에서의 A, B, C는 다음과 같다.

$$A = [\mu_{i,i-1}(0) \sigma_i^2(1) - \mu_{i,i}(1) \mu_{i,i-1}(1)] [\sigma_{i+1}^2(1) \sigma_i^2(1) - \mu_{i,i+1}^2(0)] - [\mu_{i,i+1}(0) \sigma_i^2(1) - \mu_{i,i}(1) \mu_{i,i+1}(0)] [\mu_{i+1,i-1}(1) \sigma_i^2(1) - \mu_{i,i+1}(0) \mu_{i,i-1}(1)]$$

$$B = [\sigma_{i-1}^2(0) \sigma_i^2(1) - \mu_{i,i-1}^2(1)] [\sigma_{i+1}^2(1) \sigma_i^2(1) - \mu_{i,i+1}^2(0)] - [\mu_{i+1,i-1}(1) \sigma_i^2(1) - \mu_{i,i-1}(1) \mu_{i,i+1}(0)]^2$$

$$C = [\mu_{i,i+1}(1) \sigma_i^2(1) - \mu_{i,i}(1) \mu_{i,i+1}(0)] [\sigma_{i-1}^2(0) \sigma_i^2(1) - \mu_{i,i-1}^2(1)] - [\mu_{i,i-1}(0) \sigma_i^2(1) - \mu_{i,i}(1) \mu_{i,i-1}(1)] [\mu_{i+1,i-1}(1) \sigma_i^2(1) - \mu_{i,i+1}(0) \mu_{i,i-1}(1)]$$

여기서 $\sigma_i^2(m) = \sum_{n=1}^{N_f} f_i^2(n-m)$ 이고, $\mu_{i,j}(m) = \sum_{n=1}^{N_f} f_i(n) f_j(n-m)$ 이다.

위의 식 (9)와 (10)을 각각 벡터표현식으로 나타내면 다음과 같고

$$\tilde{f}(n) = \Delta \cdot \hat{f}(n-1) \quad (14)$$

$$\hat{f}(n) = \tilde{f}(n) + \hat{e}(n) \quad (15)$$

Δ 는 아래와 같이 표현된다.

$$\Delta = \begin{bmatrix} b_1 & c_1 & 0 \\ a_2 b_1 & a_2 c_1 + b_2 & c_2 \\ a_3 a_2 b_1 & a_3(a_2 c_1 + b_2) & a_3 c_2 + b_3 \\ \vdots & \vdots & \vdots \\ a_k \dots a_2 b_1 & a_k \dots a_3(a_2 c_1 + b_2) & a_k \dots a_4(a_3 c_2 + b_3) \\ \vdots & \vdots & \vdots \\ \dots & 0 & \dots \\ \dots & 0 & \dots \\ \dots & 0 & \dots \\ \vdots & \vdots & \vdots \\ \dots & a_k a_{k-1} + b_k & \dots \\ \vdots & \vdots & \vdots \end{bmatrix}$$

표 3은 모의실험에서 실제로 사용된 예측계수 a_i, b_i, c_i 의 값들이다. 이들은 한글 문장에 대한 최적 계수이다.

표 3. 모의실험에 사용된 a_i, b_i, c_i 의 예측계수
Table 3. Prediction coefficients for a_i, b_i, c_i

i	a_i	b_i	c_i
1	0.000000	0.927084	0.038652
2	0.601995	0.442978	0.087698
3	0.474166	0.609943	0.092592
4	0.315116	0.670496	0.091010
5	0.154858	0.750128	0.116463
6	0.328419	0.569296	0.126125
7	0.192260	0.651660	0.167712
8	0.308244	0.655115	0.062993
9	0.263998	0.619592	0.138620
10	0.253394	0.639372	0.110982

V. 모의실험 및 결과

V-1. 모의실험 환경

모의실험에 사용된 선형예측 계수는 40 ms의 헤밍 창을 사용한 후, 자기 상관방법을 사용하여 계산하였으며, 예측차수는 10이고, 창은 20 ms씩 이동시켰다.

각 양자화기의 학습벡터로는 공통적으로 한명의 남성과 두명의 여성으로 이루어진 10270개의 프레임을 갖는 한글을 사용하였으며, 실험용 음성으로는 FM 라디오에서 녹음한 음성으로서 세명의 여성과 한명의 남성으로 이루어진 10848개의 프레임을 갖는 한글을 사용하였다. 또한 양자화된 선 스펙트럼 계수는 순차성을 이용하여 양자화된 것이 아니므로 순차성을 어길수도 있는데 이런 경우에는 양자화기의 성능을 저하시키지 않도록 크기순으로 배제하였다. 두가지 종류의 양자화기에 4가지 종류의 10차원 격자벡터양자화기를 모두 적용하여 각각의 격자벡터양자화기의 성능을 비교하였다. 그리고 이 논문에서 설계한 양자화기의 성능과 비교하기위해서 국내의 CDMA기 음성부호기에 사용된 AQBW(Backward Adaptive Quantizer)를 사용하였다⁽¹³⁾. 성능 평가방법으로는 평균 주파수왜곡을 사용하였다.

2단 벡터양자화기의 경우 다음과 같은 모의실험이 이루어졌다. 학습벡터를 사용하여 벡터양자화기의 부호화를 설계하였다. 프레임당 부호화율은 벡터양자화기에 9 bit를 할당하고, 4가지 격자벡터양자화기의 부호화율을 변화시켜서 전체적으로 프레임당 부호화율을 26 bit에서 39 bit까지 변화시켰다. 격자벡터양자화기의 입력인 벡터양자화 오차는 학습벡터에 의한 벡터양자화 오차의 평균과 분산을 사용하여 정규화하였다.

VDPCM 격자벡터양자화는 먼저 학습벡터를 사용하여 예측계수를 계산한 후, 예측 오차를 4가지 종류의 격자벡터양자화기를 사용하여 23 bit에서 32 bit까지 증가시키면서 격자벡터양자화하였다. 이때, 격자벡터양자화기의 입력인 예측오차를 평균 0, 분산 1로 정규화하기 위해서 필요한 평균과 분산은 학습벡터를 양자화기의 입력으로 사용하였을 경우의 예측 오차의 평균과 분산을 사용하였다.

V-2. 모의실험 결과

다음의 그림들은 모의실험 결과로서, 그림 6은 2단 벡터양자화를 사용하였을 경우의 4가지 종류의 격자벡터양자화기와 AQBW의 성능을 비교한 것이고, 그림 7은 VDPCM에 적용한 4가지 종류의 격자벡터양자화기와 AQBW의 성능을 비교한 것이다. 그림 6에서 4가지 격자벡터양자화기중 구형의 의사균일 격자

벡터양자화기가 가장 나은 성능을 나타내고, 어느 정도 부호화율이 증가하면 피라미드형의 의사균일 격자벡터양자화기의 성능이 급격히 좋아짐을 볼 수 있으며, 나머지 격자벡터양자화기는 성능의 차이가 크지않음을 알 수 있다. 그리고 격자벡터양자화기의 성능이 전체적으로 AQBW보다 좋지않은 것을 알 수 있다. 이러한 성능의 차이는 벡터양자화기의 경우 그 성능이 학습벡터에 의해 영향을 받는 반면, 모의실험에 사용된 학습벡터가 다양하지 못했기 때문인것으로 해석된다. 따라서 보다 다양한 화자의 다양한 문장을 사용한다면 나은 성능을 보일 수 있을것으로 기대된다. 그림 7에서 나타내는 VDPCM 구조를 가지는 양자화기는 전체적으로 AQBW보다 현저히 나은 성능을 나타낸다. VDPCM 구조의 양자화기의 경우 스케일러 양자화기와 비교해 볼 때 복잡도가 약간 높으나 벡터양자화기에 비해서는 월등히 낮은 복잡도로 인하여 실시간 구현이 용이하다는 장점을 가진다. 4가지 격자벡터양자화기 중에서 구형의 의사균일 격자벡터양자화기가 가장 나은 성능을 나타내고, 2단 벡터양자화기의 경우와 같이 어느 정도 부호화율이 증가하면 피라미드형의 의사균일 격자벡터양자화기의 성능이 급격히 좋아지는 것을 볼 수 있다. 2단 격자벡터양자화기와 VDPCM구조의 양자화기에 있어서 이러한 현상이 발생하는 것은 각각의 격자벡터양자화기의 입력으로 들어가는 신호의 분포가 어느 특정모델로 확연히 모델링되지않기 때문인 것으로 해석된다.

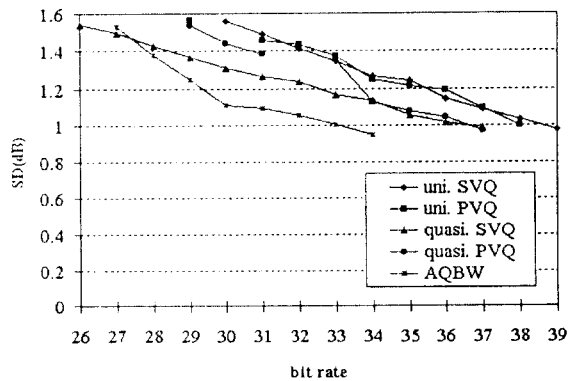


그림 6. 2단 벡터양자화기에 사용된 격자벡터양자화기들의 성능 비교
Fig. 6 Performance comparison of LVQs in two-stage VQs

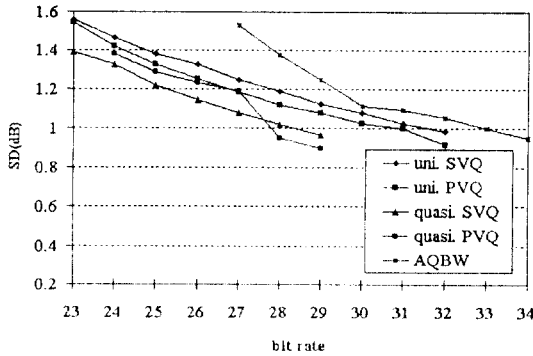


그림 7. VDPCM에 사용된 격자양자화기들의 성능 비교
Fig. 7 Performance comparison of LVQs in VDPCM

표 4. 28 bit에서의 양자화된 선 스펙트럼 계수의 평균 주파수 왜곡(VDPCM)

Table 4. Average spectral distortion for VDPCM at 28 bit rate

	SD (dB)	Outliers (%)	
		2~4 dB	4 dB >
uni. SVQ*	1.19	7.69	3.19
uni. PVQ**	1.118	3.8	0.9
quasi. SVQ***	1.02	5.96	1.549
quasi. PVQ****	0.954	4.34	0.728

- *: 구형 균일 격자벡터양자화기
- ** : 피라미드형 균일 격자벡터양자화기
- ***: 구형 의사균일 격자벡터양자화기
- ****: 피라미드형 의사균일 격자벡터양자화기

표 5. 29 bit에서의 양자화된 스펙트럼 계수의 평균 주파수 왜곡(VDPCM)

Table 5. Average spectral distortion for VDPCM at 29 bit rate

	SD (dB)	Outliers (%)	
		2~4 dB	4 dB >
uni. SVQ	1.126	7.69	3.19
uni. PVQ	1.082	3	0.553
quasi. SVQ	0.968	4.92	1.88
quasi. PVQ	0.9	4.17	0.98

표 6. 37 bit에서의 양자화된 스펙트럼 계수의 평균 주파수 왜곡(2단 벡터양자화기)

Table 6. Average spectral distortion for two-stage VQ at 37 bit rate

	SD (dB)	Outliers (%)	
		2~4 dB	4 dB >
uni. SVQ	1.089	11.107	0.654
uni. PVQ	1.094	11.84	0.829
quasi. SVQ	0.985	5.365	0.28
quasi. PVQ	0.97	5.06	0.47

표 7. 34 bit에서의 양자화된 선 스펙트럼 계수의 평균 주파수 왜곡(AQBW)

Table 7. Average spectral distortion for AQBW at 34 bit rate

	SD (dB)	Outliers (%)	
		2~4 dB	4 dB >
AQBW	0.95	3.973	0.2

표 4에서 표 7까지는 각 양자화기가 평균 주파수 왜곡 1 dB를 나타낼 때의 outlier를 나타낸 것이다. 전체적으로 outlier가 높음을 알 수 있는데, 이는 모의실험에 사용된 테스트용 음성의 음질이 다소 좋지 않았기 때문인 것으로 생각된다.

VI. 결 론

이 논문에서는 저 전송속도 음성부호화 방식에서 많이 사용되는 선형예측 계수를 선스펙트럼 계수로 변환하여, 선 스펙트럼 계수를 양자화하는 방법에 있어서, 두 가지 종류의 양자화방법을 사용하였다.

첫번째 방법인 벡터양자화기와 격자벡터양자화기를 결합한 2단 벡터양자화기의 경우, 구형 균일 격자 벡터양자화기, 피라미드형 균일 격자벡터양자화기, 구형 의사균일 격자벡터양자화기, 피라미드형 의사 균일 격자벡터양자화기 중 구형의 의사균일 격자 벡터양자화기의 성능이 36 bit이하에서는 다른 세가지 종류의 격자벡터양자화기에 비해 전반적으로 좋았고, 부호화율이 증가하면서 피라미드형의 의사균일 격자벡터양자화기의 성능이 급격히 증가하여 37 bit에서는 0.97의 평균 주파수 왜곡을 보여 가장 좋은 성능을 나타내었다. 한편, 벡터양자화기는 그 성능이 학습벡터에 의해 영향을 받는 반면, 모의실험에 사용된

학습벡터가 다양하지 못했기 때문에 부호책 설계를 위하여 사용되는 음성을 보다 다양하게 사용한다면 충분히 좋은 성능을 얻을 것으로 생각된다.

선 스펙트럼 계수의 두번째 양자화 방법인 VDDPCM 구조를 가지는 양자화기는 복잡도에 있어서는 벡터 양자화기보다 상당히 낮고, 스케일러 양자화기 보다 다소 높은 복잡도를 가지며, 성능면에 있어서 벡터양자화기를 사용했을 경우와 차이가 나지 않음을 보였다. 따라서 실시간으로 구현할 경우 기존의 양자화 방법에 비해 부호화율을 프레임당 10 bit 정도 낮출수 있으리라 생각된다. 또한 2단 벡터양자화기의 경우에서와 같이 구형의 의사균일 격자벡터양자화기가 27 bit에서 평균 주파수 왜곡이 1.08 dB로 전반적으로 가장 나은 성능을 나타내고, 부호화율이 증가할수록 피라미드형의 의사균일 격자벡터양자화기의 성능이 급격히 향상되어 28 bit에서 0.954 dB의 평균 주파수왜곡을 나타내었다. 이처럼 두가지 벡터양자화기의 성능이 어느 특정 격자벡터양자화기에 대해서만 두드러지게 나은 성능을 보이지 않는 것은 각 격자벡터양자화기의 입력의 통계적 모델이 가우스 또는 라플라스 함수 형태로 정확히 모델링 되지 않기 때문이다.

이상의 모의실험을 통해 볼때 2단 벡터양자화기의 경우, 부호책 생성을 위한 학습벡터를 보다 다양한 음성과 문장으로 사용한다면 구형의 의사균일 격자벡터양자화기를 적용한 2단 벡터양자화기에서 상당히 좋은 성능을 얻을 수 있으리라 기대된다. 그리고 VDDPCM 구조의 양자화기는 기존에 이동통신 분야에서 사용되고 있는 선형예측 계수의 양자화 방법보다 복잡도면에서 약간 높으나 부호화율면에서 매우 낮은 부호화율로 인하여 실시간 구현시 많은 이득을 볼 수 있으리라 예측된다.

참 고 문 헌

1. Sugamura and N. Farvardin, "Quantizer Design in LSP Speech Analysis-Synthesis," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, pp. 398-401, Apr. 1988.
2. F. K. Soong and B. H. Juang, "Optimal Quantization of LSP parameters," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, New York, pp. 394-397, Apr. 1988.
3. T. Mariya and M. Honda, "Speech coder using phase equalization and vector quantization," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Tokyo, Japan, pp. 1701-1704, Apr. 1986.
4. Y. Shoham, "Cascaded likelihood vector coding of the LSP information," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, Glasgow, Scotland, pp. 160-163, May 1989.
5. K. K. Paliwal and B. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," *IEEE Trans. on Speech and Audio Processing*, vol. 1, pp. 3-14, Jan. 1993.
6. J. Pan and T. R. Fischer, "Vector Quantization-Lattice Vector Quantization of Speech LPC Coefficients," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, Adelaide, South Australia, pp. 1-513~516, Apr. 1994.
7. 이황수, "저전송속도 음성부호화를 위한 음성처리 기술," *한국통신학회지*, 제9권, pp. 678-698, 1992년 9월.
8. Y. Linde, A. Buzo, and R. M. Gray, "An algorithm for Vector Quantizer Design," *IEEE Trans. on Comm.*, Com-28, NO. 1, 1980.
9. R. Laroia, N. Phamdo and N. Farvardin, "Robust and Efficient Quantization of Apeech LSP Parameters using Structured Vector Quantizers," *IEEE ICASSP 91*, Toronto, Canada, pp. 641-643, 1991.
10. D. G. Jeong and J. D. Gibson, "Uniform and Piecewise Uniform Lattice Vector Quantization for Memoryless Gaussian and Laplacian Source," *IEEE Trans. Inform. Theory*, vol. 39, pp. 786-804, May 1993.
11. Conway, J. H. and Sloane, N. J. A., *Sphere packings, Lattices and Group*, New York, NY:Spring Verlag. 1988.
12. C. C. Kuo, F. R. Jean, and H. C. Wang, "Low bit-rate quantization of LSP parameters using two-dimensional differential coding," *IEEE Proc. ICASSP*, San Francisco, U.S.A, pp. 1-97~1-100, 1992.
13. 이인성, 우홍채, 임명섭, "효율적인 벡터-스칼라

LSP 양자화 방법," 1994 신호처리합동학술대회, 인하대학교, pp. 202-205, 1994.

- 14. 강정원, 김성수, 정재호, 우홍제, 이인성, 정대권, "음성신호의 Line Spectrum Pair(LSP) 주파수의 2단 벡터양자화," 1994 신호처리합동학술대회, 인하대학교, pp. 88-291, 1994.

강 정 원(Jung-Won Kang)

정회원

1970년 3월 15일생

1993년 2월: 한국항공대학교 항공전자공학과(학사)
 1995년 2월: 한국항공대학교 대학원 항공전자공학과(석사)
 1996년 10월~현재: 미국 조지아 전기 및 컴퓨터공학과 박사과정

※주관심분야: 음성신호압축, 이동통신



정 재 호(Jae-Ho Chung) 정회원

1982년: 美 University of Maryland, College Park Campus(공학사)

1984년: 美 University of Maryland, College Park Campus(공학석사)

1990년: 美 Georgia Institute of Technology(공학박사)

1984년~1985년: 美 국방성 산하 해군연구소, 신호처리실, Electronic Engineer

1991년~1992년: 美 AT&T Bell 연구소, 음성 신호처리 연구실, 연구원(Member of Technical Staff)

1992년~현재: 인하대학교, 전자공학과 (협)부교수

1995년~현재: 한국전자통신연구소, 자연어처리연구실 초빙 연구원



정 대 권(Dae-Gwon Jeong) 정회원

1956년 11월 13일생

1979년 2월: 한국항공대학교 항공전자공학과(학사)

1987년 5월: Texas A&M대 Electrical Eng(석사)

1990년 12월: Texas A&M대 Electrical Eng(박사)

1979년 3월~1984년 5월: 국방과학연구소(ADD)연구원

1990년 10월~1991년 8월: 한국전자통신연구소 선임연구원

1991년 9월~현재: 한국항공대학교 항공전자공학과 부교수

1993년 5월~현재: 한국통신학회 학회지 편집위원

1996년 4월~현재: ITU-R TG8/1과 ITU-T SG15과의 Rapporteur

※주관심분야: 영상 및 음성신호 압축, 이동통신시스템 코덱