

예측 VQ-Pyramid VQ를 이용한 광대역 음성용 LSF 양자화기 설계

A LSF Quantizer for the Wideband Speech Using the Predictive VQ-Pyramid VQ

이 강 은, 강 상 원, 이 인 성
(Kang-Eun Lee, Sang-Won Kang, In-Sung Lee)

한양대학교 전자전기제어계측공학과
(접수일자: 2004년 1월 9일; 채택일자: 2004년 5월 17일)

본 논문에서는 벡터 양자화기와 피라미드 벡터 양자화기를 직렬로 결합하여 16차 벡터 소스에 대한 vector quantizer-pyramid vector quantizer (VQ-PVQ)를 개발하였으며, 예측 구조와 세이프티-넷 (safety-net) 개념을 결합시켜 광대역 음성 부호화기용 LPC 계수 양자화기를 설계하였다. 본 양자화기의 성능은 AMR-WB (ITU-T G.722.2)의 LPC 양자화기 성능과 비교하였는데, 스펙트럼 왜곡 및 메모리 요구량에서 상당한 이득을 얻었다.

핵심용어: 벡터 양자화, 광대역 음성 부호화기, 선형 예측 부호화 계수

투고분야: 음성처리 분야 (2,2)

This paper proposes the vector quantizer-pyramid vector quantizer(VQ-PVQ) structure. Also both predictive structure and safety-net concept are combined into the VQ-PVQ to quantize the LPC parameter of wideband speech codec. The performance is compared to the LPC vector quantizer used in the AMR-WB(ITU-T G.722.2), demonstrating reduction in both spectral distortion and encoding memory.

Keywords: Vector quantization, Wideband speech codec, Linear prediction coding coefficient

ASK subject classification: Speech signal processing (2,2)

I. 서론

일반적으로 음성 부호화 시스템은 선형 예측을 기본으로 구성되며, 선형 예측 부호화 (LPC: linear prediction coding) 계수를 유한 공간 안에서 최소한의 왜곡을 갖도록 하는 연구가 이미 오랜 기간 진행되어왔다. 많은 연구의 결과로 스칼라 양자화 또는 벡터 양자화를 이용한 여러 가지 효과적인 양자화 방법이 소개되어 왔고, 그 중 벡터 양자화는 큰 양자화 효과를 제공하나 그 복잡도와 메모리 크기가 구현에 많은 제약을 준다. 일반적으로 선형 예측 계수는 양자화 특성이 좋고 수학적으로 LPC와 동가인 line spectral frequency (LSF)[1]로

변환되어 양자화 된다. 대표적 벡터 양자화기인 Linde, Buzo, Gray (LBG) 알고리즘[2]을 이용하여 설계된 비구조적 벡터 양자화기의 복잡도는 부호화율과 벡터 차수의 곱에 지수적으로 증가한다. 이런 문제점을 해결하기 위하여 트리 구조 (tree-structured) 벡터 양자화 [3], 분할 (split) 벡터 양자화[4], 다 단계 (multi-stage) 벡터 양자화[5]등의 제한된 구조를 이용한 벡터 양자화기가 소개 되었다. 본 논문에서 소개하고자 하는 벡터 양자화-피라미드 벡터 양자화 (VQ-PVQ) 기법은 일종의 다 단계 벡터 양자화 방식으로 비구조적 벡터 양자화기와 구조적 벡터 양자화기인 피라미드 벡터 양자화 (PVQ: pyramid VQ)[6]를 직렬로 결합한 형태의 양자화기이다. PVQ는 고차원 피라미드 (hyper pyramid) 표면에 존재하는 격자점을 코드워드(code word)로 하는 일종의 격자 구조 양자화기로서, 단순한 부호화 복호화 알고리즘을 갖기 때문에 적은 복잡도를 요구한다.

책임저자: 강 상 원 (swkang@hanyang.ac.kr)
경기도 안산시 사1동 1271번지
한양대학교 전자컴퓨터공학부
(전화: 031-400-5215; 팩스: 031-418-8476)

PVQ는 고차원의 메모리를 가지지 않는 라프라시안 (Laplacian) 소스의 양자화에 좋은 성능을 제공하나, LSF의 예측 에러값의 분포는 정확한 라프라시안 분포를 이루지 않기 때문에 PVQ 단독으로 LSF 양자화를 수행할 때는 좋은 성능을 갖지 못할 것이다. 따라서 본 논문에서는 PVQ를 수행하기에 앞서 비구조적 벡터 양자화를 먼저 수행 함으로써 PVQ의 입력 벡터가 보다 라프라시안 구조에 근접하도록 하였다. 또한 LSF 벡터의 프레임간 상관도 (inter-frame correlation)를 이용하고자 auto-regressive(AR) 예측 구조를 이용하였고 프레임간 상관도가 떨어지는 외좌점 (outlier)에 대하여 보다 정확한 양자화를 위하여 세이프티-넷 (safety-net) 구조 [7]를 VQ-PVQ와 결합하여 LSF 양자화기를 구성하였다.

본 논문은 제 2 절에서 VQ-PVQ에 대하여 설명을 하고, 제 3 절에서는 예측기와 세이프티-넷이 결합된 VQ-PVQ에 대하여 설명을 한다. 제 4 절에서는 예측 VQ-PVQ를 이용한 광대역 음성 부호화기의 LSF 양자화기의 실험적 결과를 언급하고, 마지막으로 제 5 절에서 본 연구에 대한 결론을 논한다.

II. VQ-PVQ

PVQ[6]는 '96년 Fischer에 의해 제안된 양자화기로서 피라미드의 표면상에 존재하는 큐빅 격자점들을 코드워드로 사용하는 일종의 격자 양자화 방식이다. 따라서 PVQ는 코드북을 저장할 필요가 없으므로 메모리가 요구되지 않으며, 부호화율이 증가할수록 복잡도가 선형적으로 증가하는 특성을 가지고 있으므로, 적은 계산량으로 입력 벡터 전체를 양자화 할 수 있다. 특히 메모리를 가지지 않는 소스에 대한 PVQ의 성능은 입력 벡

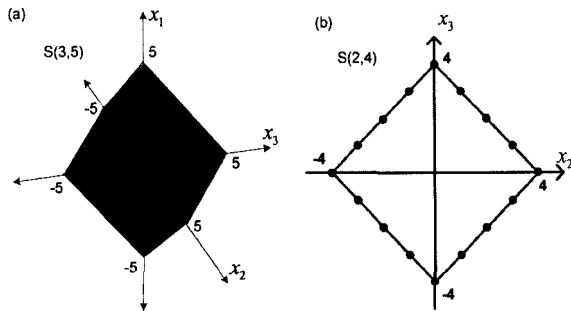


그림 1. 3-D 피라미드 S(3,5)와 2-D 피라미드 S(2,4)
Fig. 1. 3-D pyramid S(3,5) and 2-D pyramid S(2,4)

터 차수가 클 경우, 엔트로피 제한 스칼라 양자화와 비슷한 성능을 나타낸다[6]. PVQ의 코드워드는 정규화된 피라미드 위에 놓인 격자점의 스케일된 형태이다. 차수가 L인 입력 벡터 X가 라프라시안 분포를 가질 때 최적 코드워드는 식 (1)과 같이 표현되는 단일 피라미드 상에 나타낼 수 있다.

$$S(L,K) = \left\{ \mathbf{X} : \sum_{i=1}^L |X_i| = K \right\} \quad (1)$$

기하학적으로 S(L,K)는 L차원 피라미드의 표면을 나타낸 것이다. 예를 들어 입력 신호 벡터의 차수가 3이고 K가 5이면 그림 1의 (a)와 같은 3-차원 피라미드를 얻을 수 있고, x1축을 기준으로 x1=1인 곳에서 횡단 한다면 그림 1의 (b)와 같은 벡터의 차수가 2이고 K가 4인 2-차원 피라미드를 얻을 수 있다.

단일 피라미드 S(L,K)상에 있는 코드워드의 수를 N(L,K)라 한다면, K는 식 (2)를 만족하는 가장 큰 정수로 선택된다.

$$N(L,K) \leq 2^{RL} \quad (2)$$

여기서, R은 벡터 차수 당 할당된 비트이며, N(L,K)는 식 (3)을 이용하여 구할 수 있다.

$$N(L,K) = N(L-1,K) + N(L-1,K-1) + N(L,K-1) \quad (3)$$

코드워드를 정규화된 피라미드로 나타낼 필요성이 있으므로 정규화된 피라미드를 S(L,L/λ)로 표기한다. 여기서 λ는 L/E[||X||]를 나타낸다. 여기서 ||X||_v는 $\left(\sum_{i=1}^L |X_i|^v \right)^{1/v}$, v ≥ 1을 나타낸다. 양자화 입력 벡터에 대한 PVQ 부호화 알고리즘은 다음과 같다.

단계 1: 각 입력 벡터, X에 대해서, 프로젝션 공식 식 (4)에 의하여 피라미드 S(L,L/λ)의 표면상으로 프로젝션 하여 $\hat{\mathbf{X}}$ 을 얻는다.

$$\hat{\mathbf{X}} = \mathbf{X} - \left[\left(\mathbf{X}, \frac{\mathbf{S}}{\|\mathbf{S}\|_2} \right) - \frac{L\|\mathbf{S}\|_2}{\lambda\|\mathbf{S}\|_1} \right] \frac{\mathbf{S}}{\|\mathbf{S}\|_2} \quad (4)$$

여기서, S 는 식(5)와 같고, (\cdot, \cdot) 은 내적을 말한다.

$$S_i = \begin{cases} 1, & \text{if } X_i \geq 0 \\ -1, & \text{if } X_i < 0 \end{cases} \quad (5)$$

단계 2: 피라미드 $S(L, K)$ 표면으로 위치시키기 위하여 \hat{X} 을 $K\lambda/L$ 로 스케일링 한다.

단계 3: 피라미드 $S(L, K)$ 표면의 $N(L, K)$ 개의 정수 코드워드 중 가장 가까운 코드워드, \tilde{X} 를 찾는다. 찾는 과정은 다음과 같다.

(1) \tilde{X} 의 각 요소를 가장 가까운 정수로 올림 한다. 이 때 올림된 벡터를 \tilde{Y} 이라고 하자.

(2) $\|\tilde{Y}\|_1$ 이 K 와 같다면 $S(L, K)$ 표면의 가장 가까운 코드워드가 찾아진 것이다. $\|\tilde{Y}\|_1 < K$ 이면 $K - \|\tilde{Y}\|_1$ 만큼 \tilde{Y} 의 요소들 중 큰 오차를 가지는 것을 1 증가시키고, $\|\tilde{Y}\|_1 > K$ 이면 $\|\tilde{Y}\|_1 - K$ 만큼 \tilde{Y} 의 요소들 중 0이 아니면서 큰 오차를 가지는 것을 1 감소시킨다.

단계 4: \tilde{Y} 를 다시 $L/K\lambda$ 로 스케일링하여 양자화된 벡터 Y 를 얻는다.

PVQ는 입력 벡터의 차수가 충분히 클 때 충분히 우수한 성능을 나타낸다. 즉 입력 벡터의 차수가 약 20이상일 때, $\|X\|_1$ 이 일정한 값을 갖게 된다.

그러나 벡터의 차수가 20이하일 때 $\|X\|_1$ 의 분산값이 커져서 일정한 값을 갖지 않는다. 이러한 경우 단일 피라미드를 사용하여 양자화 할 때 많은 오차가 발생한다. 상기 단점을 보완하기 위하여 PCPVQ(productive coded PVQ)[6]가 제안되었다.

PCPVQ는 양자화기 입력 벡터를 정규화된 여러 개의 피라미드를 사용하는 개념으로, 정규화된 피라미드의 크기는 스칼라 양자화 된 형태, $\hat{f} = Q(\|X\|_1)$ 로 나타내어진다. 여기서 $Q(\cdot)$ 는 스칼라 양자화를 말하며, \hat{Y} 이 PVQ의 출력 벡터라면 PCPVQ의 출력 벡터, Y 는 식(6)과 같이 나타낼 수 있다.

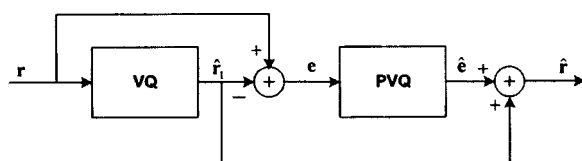


그림 2. VQ-PVQ의 구조
Fig. 2. Structure of VQ-PVQ

$$Y = \hat{f}\hat{Y} \quad (6)$$

이렇게 함으로써 스칼라 양자화기의 양자화 레벨 수만큼의 피라미드를 사용하는 효과를 얻을 수 있다. PVQ의 평균 벡터 차수당 할당 비트를 R_p , 스칼라 양자화기에 할당된 비트를 R_r 이라 하면, 전체 차수당 할당 비트 R 은 식(7)과 같이 나타낼 수 있다.

$$R_p L + R_r = RL \quad (7)$$

VQ-PVQ는 두 단계의 벡터 양자화기로서 첫 번째 단계에서는 비구조적 양자화를 사용하고 두 번째 단계에서는 PVQ를 사용하였다. Rate-distortion 이론에 따르면 메모리를 가지지 않는 소스에 대하여 부호화율이 증가할수록 부호화 잡음은 가우시안 (Gaussian) 분포를 갖는다. 따라서 두 번째 양자화기는 구면 격자 양자화기를 사용함이 적당하나 구면 격자 양자화기는 연산량 측면에서 그리 효율적이지 못하다. 따라서 복잡도 측면에서 볼 때 PVQ는 매우 적은 부호화 복잡도를 요구하기 때문에 전체적인 양자화 출력은 공동 최적 방식으로 탐색될 수 있다. 첫 번째 단계 양자화기의 코드북이 nVQ 의 코드 워드를 갖고, 두 번째 단계 양자화기가 $nPVQ$ 의 코드 워드를 갖는다면 전체적인 부호화율은 벡터 차수 당 $\lceil \log_2 n_v n_{pvq} \rceil / L$ 의 비트가 된다.

여기서 $\lceil A \rceil$ 는 A 보다 작지 않은 가장 큰 정수를 나타낸다. VQ-PVQ의 부호화 구조는 그림 2와 같다. 최종적으로 양자화된 출력 벡터는 식(8)과 같다.

$$\hat{f} = \hat{e} + \hat{f}_i \quad (8)$$

비구조적 벡터 양자화기와 PVQ간의 공동 최적 양자화를 위하여, 첫 번째 단계 양자화기에서 mean squared error (MSE)를 최소화 하는 여러 개의 후보 벡터를 선택한다. 그런 후 두 번째 단계 양자화기 PVQ에서는 각 후보 벡터의 양자화 에러를 각각 양자화 한다. 두 양자화기에서 뽑은 코드북 짝 $p(i, j)$ 는 식 (9)와 같이 표현된다. 여기서 i 는 첫 번째 단계 벡터 양자화기에서 찾은 코드북의 인덱스를 j 는 두 번째 단계 양자화기 PVQ에서 찾은 코드북의 인덱스를 나타낸다.

$$p(i, j) = \left\langle r : \min_{n_{1V}}^{-1} d(r, U_{n_{1V}}) = i, \right. \\ \left. e : \min_{n_{1V}}^{-1} d(e, P_{n_{1V}}) = j \right\} \quad (9)$$

여기서 $U_{n_{1V}}$ 는 코드북 개수가 nVQ인 첫 번째 단계 벡터 양자화기의 코드북을 나타내고, $P_{n_{1V}}$ 는 코드북 개수가 nPVQ인 두 번째 단계 양자화기 PVQ의 코드북을 나타낸다. 또한 $d(A, B)$ 는 벡터 A와 코드북 B 간의 MSE (mean squared error)값을 나타내는 벡터라 할 때 $\min_{n_{1V}}^{-1} c = l$ 는 n차 벡터 c중 그 크기가 최소인 요소 값의 인덱스 l을 나타낸다. 최종적으로 입력 벡터와의 에러값이 가장 적은 두 양자화기의 코드북 짝이 선택된다. PVQ의 부호화 복잡도가 매우 적기 때문에 첫 번째 양자화기에서 다수의 후보 벡터를 선택할 수 있다. 후보 벡터의 개수가 많아질수록 더욱 정확한 공동 최적의 탐색 과정이 얻어질 수 있으며 개수가 어느 정도를 넘었을 때는 성능 향상이 포화 된다.

III. 예측 VQ-PVQ

LSF 계수 양자화기의 설계는 인지적 음성의 품질, 요구 비트, 그리고 구현상 복잡도 등을 상호 고려하여야 한다. 본 논문에서 설계한 LSF 계수 양자화기는 주파수

영역 및 시간 영역에서 가지는 LSF 계수의 높은 상관성을 이용하였다. 대상 신호의 양자화는 최소의 비트로 최대한 정확히 대상 신호를 표현하도록 부호화 하는 것이 목적이다. 앞에서 언급하였듯이 LSF 계수는 프레임간 높은 상관성을 가지며, LSF 계수 양자화기가 효율적인 성능을 가지기 위하여 언급된 상관성을 충분히 이용하여야 한다. 그림 3은 광대역 음성 신호에 대한 LSF 계수 자취를 음성의 자취와 함께 나타내었으며, LSF가 프레임간 높은 상관성을 가짐을 보여 주고 있다. 이러한 성질을 프레임간 상관도 (inter-frame correlation)라 한다. LSF 계수 벡터는 음성 프레임의 주파수 특성을 나타내며, 일반적으로 10ms 내지 20ms의 음성 프레임은 음성의 단구간 주파수 특성이 아주 천천히 변화하므로 연속 프레임간에 큰 상관성이 존재 한다. 따라서 LSF 계수의 이러한 높은 상관성을 이용함으로써 양자화 이득을 높일 수 있다.

본 논문에서는 LSF의 프레임간 상관도를 이용하기 위하여 1차 AR 예측기를 사용하였으며, 그 구조를 그림 4에 나타내었다. $f'(n)$ 은 L차원의 평균값이 제거된 n 번째 프레임의 LSF 벡터이고, $f'_k(n)$ 는 $f'(n)$ 의 k번째 요소이다.

n번째 프레임의 예측 값이 제거된 LSF 벡터 $r(n)$ 은 식 (10)과 같이 나타내 질 수 있다.

$$r_k(n) = f'_k(n) - \alpha_k f'_k(n-1) \quad (10)$$

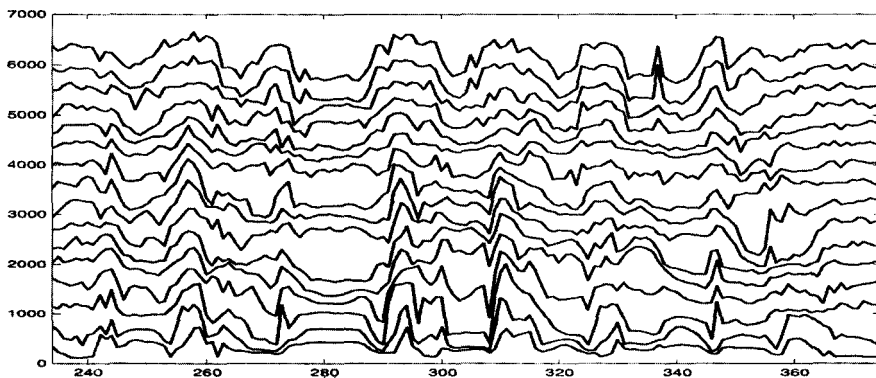
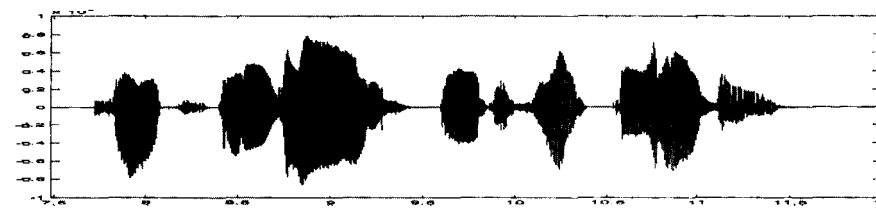


그림 3. 프레임에 따른 광대역 음성과 광대역 코덱의 LSF 변화
Fig. 3. Evolution with frame of wideband speech and wideband line spectral frequencies.



여기서 α_k 는 k번째 요소에 해당하는 AR 예측 계수 값이며, $\hat{f}_k(n-1)$ 은 n-1번째 프레임의 LSF벡터의 k번째 요소가 양자화된 값을 나타낸다. LSF 계수는 단구간 음성의 주파수 특성을 나타내며 입력 음성의 주파수적 특성이 급격히 변하는 프레임의 LSF 계수는 그림 3에서와 같이 급격히 변화함을 확인할 수 있다. LSF 계수의 프레임간 높은 상관도를 이용하는 프레임간 예측기를 사용하는 양자화기의 경우 이와 같이 급격히 변화하는 프레임에 대해서는 적절한 예측이 불가능하여 성능이 떨어질 수 밖에 없으며 외좌점 발생의 원인이 된다.

따라서 예측이 절절히 되지 않는 음성 프레임의 신호 특성에 대하여 최적화된 양자화기를 병렬로 구성하여 성능을 향상 시키도록 설계된 구조를 세이프티-넷 구조[7]라 한다. 세이프티-넷 양자화 구조는 두 개의 구조에 대한 양자화가 독립적으로 수행되어 각각 가장 좋은 코드워드를 결정한 후 둘을 비교하여 더 좋은 코드워드를 선택한다. 이와 같이 메모리 기반 양자화기와 비 메모리 양자화기를 함께 사용하여 얻을 수 있는 효과는 다음과 같다. 첫 번째로 예측 구조에 적합하지 않은, 즉 과거 프레임의 LSF 계수와 상관도가 떨어지는 외좌점을 효율적으로 부호화 할 수 있고, 두 번째로 메모리 기반 양자화기를 상관도가 큰 프레임에 초점을 두어 설계 할 수 있도록 하여 성능 향상을 극대화 할 수 있으며, 마지막으로 비 메모리 모듈에 의해 잡음 채널에서 에러 전달현상을 줄일 수 있다. 다음 그림 5는 AR 필터 예측기를 사용하는 메모리 기반 양자화기와 대상 신호에 대한 예측 없이 직접 양자화하는 비 메모리 양자화기로 구성된 세이프티-넷 구조 양자화기 전체 구조를 나타내었다. 그림 5에 나타낸 양자화기의 동작은 다음과 같이 이루어진다. 우선 각각의 구조로 입력된 LSF 계수는 벡터의 DC 성분이 제거된다. DC 성분이 제거된 LSF 계수 벡터는 메모리 모듈에서는 1차 AR 예측 필터를 사용하여 프레임 간 예측이 수행되고 예측 에러 값이 VQ-PVQ의 양자화기를 통해 양자화가 수행

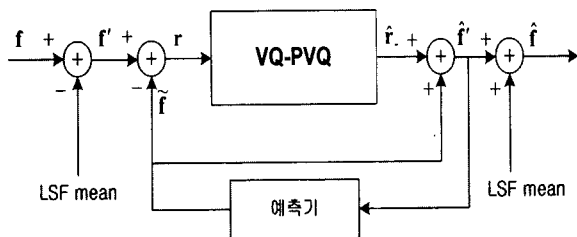


그림 4. 1차 AR 예측기를 사용한 VQ-PVQ의 구조
Fig. 4. Predictive VQ-PVQ with first order AR predictor

되며, 비메모리 모듈에서는 예측이 수행되지 않고 VQ-PVQ의 양자화가 수행된다. 다음은 세이프티-넷 구조를 이루는 두 양자화기의 최적화 설계 방법에 대하여 알아보겠다. 세이프티-넷 양자화기를 구성하기 위해 독립적으로 설계된 각 양자화기를 단순히 조합하면 입력 신호에 대하여 두 개의 양자화 구조는 상호 적절히 작용할 수 없다. 왜냐하면 각각의 양자화기가 예측 가능한 입력 신호 및 예측 불가능 입력 신호의 구분 없이 설계 되었기 때문이다. 본 논문에서 제시한 예측 VQ-PVQ 양자화기의 최적 코드북 설계 과정은 다음과 같다.

단계 1: 메모리 구조를 위하여 예측 에러 벡터를 추출하고, 비메모리 구조를 위하여 평균값만을 제거한 LSF 벡터를 추출한다. 예측 에러 벡터를 추출할 때 양자화는 이루어지지 않는 구조를 통하여 추출한다. 추출된 LSF 벡터를 사용하여 두 구조의 첫 번째 단계 양자화기의 코드북을 설계 한다.

단계 2: 설계된 코드북을 이용하여 두 구조에서 첫 번째 단계의 양자화기를 구성한다. 두 번째 양자화기 PVQ를 설계하기 위하여 첫 번째 단계의 양자화 에러 벡터를 추출한다. 추출된 LSF 벡터를 이용하여 PVQ를 설계한다. 이 때 두 번째 양자화는 수행되지 않는다.

단계 3: 각각 개루프 설계를 통해 설계된 코드북을 이용하여 양자화기를 구성하고 양자화를 수행하면서 두 구조에서 첫 번째 양자화기의 코드북을 설계하기 위한 LSF 벡터와 두 번째 양자화기 PVQ를 위한 벡터를 추출한다. 이 때 MSE에 의해 선택 되어진 구조에서의 값만을 추출한다. 추출된 LSF 벡터를 사용하여 각각의 코드북을 설계한다.

이러한 과정을 코드북의 갱신 변화율이 어느 정도 한 계 이하로 떨어질 때 까지 되풀이 한다. 언급한 반복적 과정을 거쳐 메모리 구조와 비메모리 구조에 적합한 코드북이 설계된다.

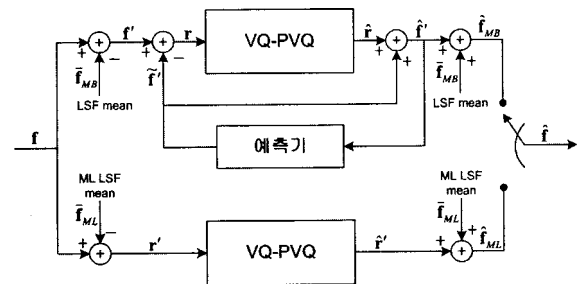


그림 5. 세이프티-넷 구조를 사용한 VQ-PVQ의 구조
Fig. 5. Predictive VQ-PVQ with safety-net

표 1. 예측 VQ-PVQ의 비트할당
Table 1. Bit allocation of the predictive VQ-PVQ.

		비트 할당	
모드 정보	1		
첫 번째 단계 VQ	7	6	
두 번째 단계 PVQ	38	39	
총	46		

표 2. 여러 가지 비트할당에 따른 예측 VQ-PVQ의 SD 성능
Table 2. SD performance of the predictive VQ-PVQ at different bit rate.

비트 할당	평균 SD [dB]	3 dB ~ 5 dB [%]	> 5 dB [%]
46	0.737	0.016	0
45	0.790	0.010	0
43	0.871	0.018	0
41	1.015	0.036	0
39	1.220	0.085	0
36	1.307	0.327	0

IV. 실험 및 결과

제 4 절에서는 본 논문에서 개발된 광대역 음성 부호화기 용 LSF계수 양자화기의 성능을 3GPP에 의해 광대역 음성 부호화기 표준으로 선택된 AMR-WB 음성 부호화기[8]의 LPC 계수 양자화기와 비교 분석 하였다. LPC 계수 양자화기의 성능 평가를 위해 스펙트럼 왜곡 (SD)값을 사용하였다. i번째 프레임의 SD 값 Di는 식 (11)과 같이 정의 된다.

$$D_i^2 = \frac{1}{F_s} \int_0^{F_s} [10 \log_{10}(P_i(f)) - 10 \log_{10}(\hat{P}_i(f))]^2 df \quad (11)$$

여기서 Fs는 샘플링 주파수 이며, $P_i(f)$ 및 $\hat{P}_i(f)$ 는 LPC 파워 스펙트럼 값이다. 설계된 광대역 LPC 양자화기에서 투명한(transparent) 성능이란 평균 SD가 1dB 이하이며, 3~5dB 외좌점 비율이 3% 이내이고, 또한 5dB 이상의 외좌점 비율은 0%로서 사람의 귀로 LPC 계수의 양자화 에러 영향을 거의 감지할 수 없는 수준이다[9]. 본 논문에서 개발한 양자화기의 성능을 AMR-WB 음성 부호화기의 LPC 계수 양자화기와 균등한 조건에서 비교하기 위하여, LPC 계수 분석의 전처리 과정을 AMR-WB 음성 부호화기와 똑 같이 적용하였다. 비트 할당은 AMR-WB와 동등하게 46비트를 할

표 3. AMR-WB의 LPC양자화기와 예측 VQ-PVQ의 SD 비교(46비트 할당)

Table 3. SD comparison of the AMR-WB S-MVQ and the predictive VQ-PVQ(at 46 bits).

	AMR-WB S-MVQ	예측 VQ-PVQ
평균 SD [dB]	0.842	0.737
3 dB ~ 5 dB [%]	0.013	0.016
> 5 dB [%]	0	0

표 4. AMR-WB의 LPC양자화기와 예측 VQ-PVQ의 WMOPS를 이용한 연산량과 코드북 메모리 비교(46 비트 할당)

Table 4. Comparison of computational complexity and ROM requirement of the AMR-WB LPC quantizer and predictive VQ-PVQ (at 46 bits).

	AMR-WB S-MVQ	예측 VQ-PVQ
WMOPS	2.25	2.12
ROM (words)	5280	3343

표 5. WMOPS측정 시 사용된 연산자의 가중치

Table 5. Weighting value to measure the WMOPS.

연산자	가중치
+ or -	1
×	1
MAC	1
논리 연산	2
절대값	1
Branch	2
+	18

당하였다. 또한 VQ와 PVQ에 할당될 최적의 비트를 찾기 위하여 예측 구조와 비예측 구조 각각에 대하여 다양한 비트를 할당하여 최적의 비트 조합을 찾았다. 찾아진 최적의 비트 조합은 표 1과 같다.

본 실험에서는 LPC 계수 양자화기 설계를 위하여 38500 프레임 (12분)의 16kHz 샘플링 음성을 사용하였으며, 성능 테스트를 위하여 38500 프레임의 음성 샘플을 사용하였다. 훈련용 및 테스트용 음성 샘플은 한국어 남자, 여자, 영어 남자, 여자의 음성 순으로 반복 구성하였다.

16차 LSF 계수에 대한 예측 VQ-PVQ의 성능은 표 2와 같으며 프레임 당 41비트에서 투명한 품질이 제공됨을 알 수 있다. 광대역 AMR-WB의 LPC 계수 양자화기와 예측 VQ-PVQ의 성능을 비교하여 표 3에 나타내었으며, 본 논문에서 개발한 예측 VQ-PVQ의 성능이 평

균 SD 측면에서 0.105 dB 더 우수함을 알 수 있다.

표 4는 광대역 AMR-WB의 LPC 계수 양자화기와 예측 VQ-PVQ의 계산량을 비교한 표이다. 연산량은 WMOPS를 사용하여 측정하였으며 WMOPS 계산 시 사용한 가중치는 표 5와 같다.

V 결론

본 논문에서는 광대역 음성 부호화기를 위한 효율적인 LSF계수 양자화 방식을 예측 VQ-PVQ를 이용하여 설계하였다. 이를 위해 LSF 계수의 프레임간 상관도를 이용하는 예측 구조 설계와 예측 에러 벡터를 위한 효율적인 양자화 방식을 연구 하였다. 기본적인 VQ-PVQ의 구조에 AR 예측 구조와 비예측 두 가지 구조를 병렬로 사용하는 세어프티-넷 구조를 사용함으로써 예측 구조에 적합치 않은 외좌점의 수를 줄일 수 있으며, AR 예측 구조에서 발생할 수 있는 에러 전달 현상을 방지할 수 있다.

본 논문에서 개발한 양자화기는 프레임 당 46 비트에서 3GPP 광대역 표준 음성 부호화기 AMR-WB의 LPC 계수 양자화기에 비하여 같은 연산량하에서 평균 SD가 0.105 dB 향상되었다. 또한 코드북을 저장하기 위한 메모리는 36.6%감소 하였다. VQ-PVQ는 약 41비트에서 광대역 음성 부호화기가 원하는 투명한 조건을 만족한다. 하지만 PVQ가 전송 오류에 대하여 민감한 특성을 가지고 있어 채널 에러하에 기존의 S-MVQ보다 다소 성능저하가 있다. 이부분에 대해서는 현재 연구가 진행중에 있다.

참고 문헌

1. F. Itakura, "Line spectral representation of linear predictive coefficients of speech signals," J. Acoust. Soc. Amer., 57, Supplement no. 1, S35, 1975.
2. Y. Linde, A. Buzo, R. M. Gray, "An algorithm for vector quantizer design," IEEE Trans., COM-28, 84-95, 1980.
3. L. Breiman, J. H. Friedman, R. A. Olshen, and C. J. Stone. Classification and Regression Trees. Wadworth, Belmont, California, 1984.
4. 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.
5. B. H. Juang and A. H. Gray, Jr., "Multiple stage vector

quantization for speech coding," International Conference on Acoustic, Speech, and Signal Processing., 1, 597-600, Paris, April, 1982.

6. Thomas R. Fischer, "A Pyramid Vector Quantizer," IEEE Trans. Information Theory, IT-32, 4, 568-583, July 1986.
7. T. Eriksson, J. Linden, J. Skoglund, "Exploiting interframe correlation in spectral quantization: a study of different memory VQ schemes," Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on, 2, 765-768, 1996.
8. B. Bessette, R. Salami, R. Lefebvre, M. Jelinek, J. Rotola-Pukkila, J. Vainio, H. Mikkola, K. Jarvinen, "The adaptive multirate wideband speech codec (AMR-WB)," IEEE Trans. Speech Audio Processing, 10(8), Nov. 2002.
9. M. Ferhaoui, S. Gerven, "LSP quantization in wideband speech coders," 1999 IEEE Workshop on speech coding processing, 25-27, 1999.

저자 약력

• 이 강 은 (Kangeun Lee)



1995년~2002년: 한양대학교 전자컴퓨터공학부 (공학사)
 2002년~2004년: 한양대학교 전자전자제어계측공학과 (공학 석사)
 2004년~현재: 삼성종합기술원 HCI Lab. 연구원

• 강 상 원 (Sangwon Kang)



1976년~1980년: 한양대학교 전자공학과(공학사)
 1980년~1982년: 서울대학교 전자공학과(공학석사)
 1985년~1990년: Texas A&M 대학교 전기공학과 (공학박사)
 1982년~1994년2월: 한국전통통신연구원 신호처리 연구실
 1994년 3월~현재: 한양대학교 전자컴퓨터공학부, 교수
 *주관심 분야: 음성부호화, 신호처리, 이동통신

• 이 인 성 (Insung Lee)

1979년~1983년: 연세대학교 전자공학과 (공학사)
 1983년~1985년: 연세대학교 전자공학과 (공학석사)
 1987년~1992년: Texas A&M 대학교 전기공학과 (공학박사)
 1986년~1987년7월: 한국통신연구개발단 전임 연구원
 1997년~1999년8월: 한국전통통신연구원 신호처리 연구실
 1995년~현재: 충북대 전자컴퓨터공학부, 부교수
 *주관심 분야: 음성부호화, 신호처리, 이동통신