

Principal component analysis를 이용한 LSP 계수의 압축기법

안해용, 이철희
연세대학교 전자공학과

Compression of LSP Coefficients Using Principal Component Analysis

Haeyong Ahn, Chulhee Lee
Department of Electronic Engineering, Yonsei University
E-mail: chulhee@yonsei.ac.kr

요 약

Line spectrum pair(LSP) 계수는 양자화 오류에 강하고, 선형 보간에 효율적이며, 필터의 안정성 판정이 용이하므로 LPC를 대신하여 음성 부호화에 널리 사용되고 있다. 일반적으로 LSP 계수간에는 일정한 상관관계가 나타나고, 이 특성을 이용하면 LSP 계수의 부호량을 줄일 수 있는 가능성이 있다. 본 논문에서는 LSP 계수를 압축하기 위해 principal component analysis(PCA)를 사용한 방법을 제안한다. 제안된 방법에서는 LSP 계수를 Karhunen-Loeve(KL) 변환해 에너지가 집중되는 고유치(eigenvalue)와 고유벡터(eigenvector)를 찾고, 그 값을 양자화 한다. 성능 평가를 위해 2.4kbps MELP(mixed excitation linear prediction)와 8kbps QCELP(qualcumn code excited linear prediction) 음성 부호화기를 사용해 결과 값을 비교했고, 압축률이 증가하는 것을 확인했다.

1. 서 론

선형 예측 모델은 음성 신호를 분석하고 합성하는데 사용되는 기본 모델이다. 선형 예측 모델에서 음성의 스펙트럼 포락 정보를 나타내는 선형 예측 계수(LPC)는 그 값의 범위가 일정치 않고, 특정 차수의 양자화에 어려움이 있다. 또한, 양자화 된 LPC로 결정된 전극 필터(all-pole filter)의 안정성(stability)을 검사할 수 있는 직접적인 방법이 없기 때문에 음성 부호화기에 사용될 파라미터로는 적합하지 못하다 [1]. 이와 같은 결점을 보완하기 위해 LPC 대신 PARCOR 계수를 사용하는 해석 방식이 개발되었지만, 이 PARCOR 시스템에서도 계수의 보간을 수행할 때 스펙트럼 왜곡이 크다는 단점이

있었다. 이것은 PARCOR 계수의 특성이 선형 보간에 맞지 않기 때문이다.

이러한 문제를 해결하기 위해 제안된 것이 LSP(line spectrum pair) 계수이다 [2]. LSP 계수는 양자화 오류에 강하고, 선형 보간에 효율적이며, 필터의 안정성 판정이 용이한 특징을 갖기 때문에 LPC를 대신하여 음성 부호화에 널리 사용되고 있다. 일반적으로 LSP 계수 사이에는 강한 상관관계가 나타나는데 이를 이용해서 LSP 계수의 부호량을 줄일 수 있다 [3].

본 논문에서는 LSP 계수의 할당되는 비트 수를 줄이기 위해 PCA를 사용해 에너지가 집중되는 계수들을 제거하고 정보량을 줄여서 양자화 하는 압축 기법을 제안한다.

2. Line spectrum pair

그림 1은 p 차의 PARCOR 합성 필터이다. 이 필터에서 Z 와 X 사이에 K_{p+1} 을 삽입할 때, X - Y 간의 전달 특성 $H(z)$ 는 다음의 식 (1)과 같다.

$$H(z) = \frac{1}{(A_p(z) - K_{p+1}B_p(z))} \quad (1)$$

$A_p(z)$ 는 X 에서 Y 까지의 전달함수이고, $B_p(z)$ 는 Y 에서 Z 까지의 전달함수라고 할때 각각 다음과 같은 값을 갖는다.

$$\begin{cases} A_p(z) = 1 - \sum_{k=1}^p a_k z^{-k} \\ B_p(z) = z^{-(p+1)} A_p(\frac{1}{z}) \end{cases} \quad (2)$$

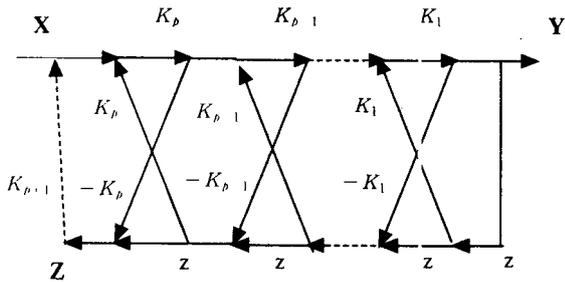


그림 1. PARCOR 합성 필터

$K_{p+1}=1$ 일 때의 전달 함수를 $P(z)$ 라 하고, $K_{p+1}=-1$ 일 때의 전달 함수를 $Q(z)$ 라 하면 식 (1), (2)로부터 다음과 같은 식을 얻을 수 있다.

$$\begin{cases} P(z) = A_p(z) - B_p(z) = A_p(z) - z^{-(p+1)}A_p(\frac{1}{z}) \\ Q(z) = A_p(z) + B_p(z) = A_p(z) + z^{-(p+1)}A_p(\frac{1}{z}) \end{cases} \quad (3)$$

식 (3)에서 $P(1)$ 과 $Q(-1)$ 의 경우의 $P(z)$ 와 $Q(z)$ 의 값은 0이 된다. 또한, $H(z)$ 는 무손실계로 가정했기 때문에 $P(z)$, $Q(z)$ 의 근은 z 평면의 단위 원주 상에 있다. 따라서 $P(z)$, $Q(z)$ 는 p 가 우수 일 때 식 (4)와 같이 표현할 수 있고, p 가 기수 일 때 식 (5)와 같이 표현할 수 있다.

$$\begin{cases} P(z) = (1 - z^{-1}) \prod_{i=2}^p (1 - 2 \cos \omega_i z^{-1} + z^{-2}) \\ Q(z) = (1 + z^{-1}) \prod_{i=2}^p (1 - 2 \cos \theta_i z^{-1} + z^{-2}) \end{cases} \quad (4)$$

$$\begin{cases} P(z) = (1 - z^{-2}) \prod_{i=2}^p (1 - 2 \cos \omega_i z^{-1} + z^{-2}) \\ Q(z) = \prod_{i=2}^p (1 - 2 \cos \theta_i z^{-1} + z^{-2}) \end{cases} \quad (5)$$

이때, ω_i 는 성대 개방($K_{p+1}=1$)에 대한 공진 주파수, θ_i 는 성대 폐쇄($K_{p+1}=-1$) 조건에 대한 공진 주파수이고, 양자를 포함한 것을 LSP 계수라 한다. LSP 계수 사이에는 다음과 같은 순서 관계가 있다.

$$0 = \omega_0 < \theta_1 < \omega_1 < \dots < \theta_{p/2} < \omega_{p/2} < \theta_{p/2+1} = \pi$$

3. Principal Component Analysis

여러 개의 변수들에 대하여 얻어진 다변량 자료를 분석대

상으로 하는 주성분 분석 기법은 다차원적인 변수들을 축소, 요약하는 차원의 단순화와 함께 일반적으로 서로 상관되어 있는 반응 변수들간의 복잡한 구조를 분석하는데 그 목적을 두고 있다. 이를 위하여 입력 데이터를 상관 관계가 없는 새로운 좌표계로 선형 변환시킨다. 즉, 변환 좌표계에서 최대 분산치에 해당하는 성분이 제 1 주성분이 되며 이에 해당하는 기저 벡터가 특징 벡터로 추출된다. 그림 2는 2차원 공간에서 주성분 분석의 예를 보여준다.

데이터 집합의 주성분을 찾는 일반적인 방법은 공분산 행렬(covariance matrix)의 고유벡터를 계산하는 것이다. 고유 벡터 값들에 대한 데이터의 투영이 주성분인데 고유 벡터에 대응하는 고유 값은 주성분을 표현하는 각각의 정보양을 표시한다. 따라서 큰 고유값에 대응하는 주성분은 데이터가 많이 분포하는 방향을 표시한다. 주성분을 계산하는 과정은 다음과 같다 [3].

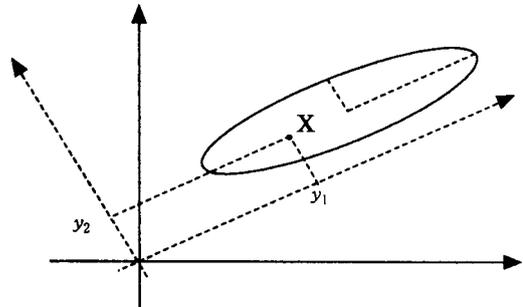


그림 2. 2차원 공간에서 주성분 분석의 예.

랜덤변수 x 의 벡터는 공분산 행렬(covariance matrix)인 Σ 를 갖는다고 가정하자. Σ 의 (i, j) 번째 성분은 $i \neq j$ 일 때 x 의 i 번째와 j 번째 성분 사이의 공분산을 나타내고, $i = j$ 일 때는 x 의 j 번째 성분의 분산을 나타낸다. $k=1, 2, \dots, p$ 에 대하여 a_k 가 Σ 행렬에서 k 번째 큰 고유치(eigenvalue)인 λ_k 에 일치하는 고유벡터(eigenvector)일 때, k 번째 주성분은 $z_k = a_k^T x$ 이다. 만약 a_k 가 단위길이($a_k^T a_k = 1$)를 가진다면 z_k 의 분산은 λ_k 이다.

주성분들을 유도하기 위해서 $a_1^T x$ 를 생각하자. a_1 은 $var[a_1^T x] = a_1^T \Sigma a_1$ 를 최대로 한다. 계산의 편리성을 위하여 $a_1^T a_1 = 1$ 로 정규화하여 사용한다. $a_1^T \Sigma a_1$ 의 최대값을 찾기 위해 Lagrange multiplier λ 를 사용한다.

$$a_1^T \Sigma a_1 - \lambda (a_1^T a_1 - 1) \quad (6)$$

식 (6)을 a_1 에 대하여 미분하면 다음과 같다.

$$\Sigma \alpha_1 - \lambda \alpha_1 = (\Sigma - \lambda I_p) \alpha_1 = 0 \quad (7)$$

여기서 I_p 는 $(P \times P)$ 단위 행렬이다. 따라서 λ 는 Σ 의 고유치이고, α_1 은 그것에 해당하는 고유벡터이다. 식 (7)을 풀어서 나오는 p 개의 고유벡터 중 어느 것이 식 (6)을 최대로 하는 α_1 인지 결정하기 위해 식 (7)로부터 $\Sigma \alpha_1 = \lambda \alpha_1$ 인 조건을 $\alpha_1^T \Sigma \alpha_1$ 에 대입하면 다음과 같다.

$$\alpha_1^T \Sigma \alpha_1 = \alpha_1^T \lambda \alpha_1 = \lambda \alpha_1^T \alpha_1 = \lambda$$

즉, $\alpha_1^T \Sigma \alpha_1$ 가 최대값을 가질때의 α_1 은 λ 가 최대 값을 가질 때 Σ 의 고유 벡터이다. 이 때 가장 큰 고유치는 $\text{var}(\alpha_1^T x) = \alpha_1^T \Sigma \alpha_1 = \lambda_1$ 이다. 위와 같은 방식으로 p 번째 주성분을 구하면 $\alpha_1, \alpha_2, \dots, \alpha_p$ 의 벡터 계수는 $\lambda_1, \lambda_2, \dots, \lambda_p$ 에 해당하는 고유 벡터가 되고 $\lambda_1, \lambda_2, \dots, \lambda_p$ 는 첫 번째, 두 번째, ..., p 번째로 가장 큰 고유치가 된다. 여기서 x 의 k 번째 원소 z_k 는 다음과 같다.

$$z = A^T x \quad (8)$$

이 때, A 는 k 번째 열이 α_k 인 직교 행렬이다. 따라서, 주성분들은 x 를 선형 변환으로 구할 수 있다. 위의 식 (8)을 PCA라고 한다. 그리고, λ 를 Σ 의 k 번째 고유치인 λ_k 를 k 번째 대각 성분으로 하는 대각 행렬이라고 하면 다음과 같은 관계가 성립한다.

$$\Sigma A = A \Lambda$$

4. PCA를 이용한 LSP 계수의 압축

공분산 행렬의 고유값을 내림차순으로 정렬했을 때, 가장 큰 값을 갖는 고유값에 대응하는 고유 벡터는 가장 큰 분산을 갖는다. 따라서 작은 고유값들에 대응하는 고유 벡터들은 입력 데이터에 대해 가장 적은 정보를 가지고 있으므로 제거할 수 있는 가능성이 있다. PCA를 이용해 LSP 계수를 압축하기 위해서는 변환된 고유값들을 조사하여 고유 벡터 중 고유값이 큰 k 개의 벡터를 선택해야 한다. 이 때, 선택의 기준이 되는 것은 다음과 같은 누적비이다.

$$\phi = \frac{\sum_{i=1}^k \lambda_i}{\sum_{i=1}^p \lambda_i}$$

그림 3은 고유값들의 누적비를 그래프로 표현한 것이다. 그림에서 총 고유값의 수는 50이다. 누적비가 99%보다 작은 차수를 선택한다고 가정하면 k 는 23이 된다. 따라서 50차의 데이터를 절반 이상 줄일 수 있다.

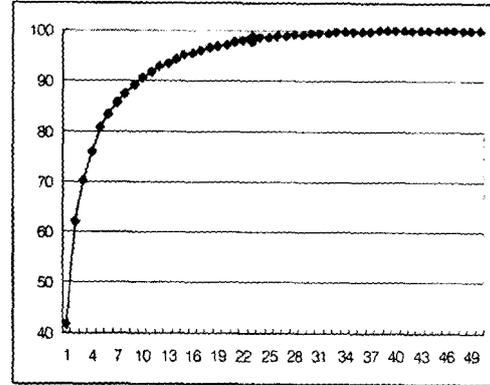


그림 3. 고유값들의 누적비.

5. 실험 결과

실험에는 3명의 화자(남자 2, 여자 1)가 긴 시간동안 녹음한 음성 파일을 사용했다. 음성 신호의 샘플링 주파수는 11 kHz이다. LSP 계수의 차수는 10차이고, 5 프레임을 묶어 50차의 LSP에 대해 PCA를 적용하였다. 음성압축 알고리즘은 2.4 kbps MELP와 8 kbps QCELP를 사용했다. MELP는 한 프레임당 LSP 계수로 25 bit가 쓰이고, QCELP는 40bit가 사용된다.

구현된 알고리즘의 성능 평가는 평균 스펙트럼 왜곡 (spectral distortion measure)을 사용했다. 평균 스펙트럼 왜곡의 계산식은 다음과 같다 [4].

$$SD = \sqrt{\frac{1}{\pi} \int_0^{\pi} (10 \log_{10} S(\omega) - 10 \log_{10} \hat{S}(\omega))^2 d\omega} \quad (9)$$

$S(\omega)$ 는 양자화 전의 LSP로부터 얻은 스펙트럼 포락이고, $\hat{S}(\omega)$ 는 양자화된 계수로부터 얻은 스펙트럼 포락이다. SD 값은 매 프레임마다 계산되며 최종적으로 평균치를 취한다. LSP 양자화의 성능 기준은 다음과 같다.

- i) 평균 SD가 1dB 이하의 값을 갖는다.
- ii) SD 값이 2dB ~ 4dB인 프레임 수의 비율이 2% 미만이다.
- iii) SD 값이 4dB를 넘는 프레임이 0%에 가까워야 한다.

이 조건을 모두 만족하면 스펙트럼 포락의 투명한 양자화

표 1. 성능 비교

음성 부호화기 \ 양자화 방법	기존의 방법			PCA를 사용한 방법		
	평균 SD (dB)	2<SD<4 (%)	4<SD (%)	평균 SD (dB)	2<SD<4 (%)	4<SD (%)
2.4 kbps MELP(k=25)	1.08	2.48	0.02	1.09	2.58	0.02
8 kbps QCELP(k=46)	0.95	1.86	0.01	1.38	3.36	0.03

(transparent quantization)라 하며 실험적으로 입증된 사실에 따르면 부명한 양자화에 의한 정보 손실은 무시될 수 있다.

표 1은 MELP와 QCELP에 제안된 압축 기법을 적용했을 때 스펙트럼 왜곡 정도를 구한 것이다. MELP의 경우 50개의 고유 벡터 중 25개의 고유 벡터를 선택했을 때 기존의 음성과 비슷한 수준의 음질을 얻을 수 있었다. 하지만 QCELP의 경우에는 46개의 고유 벡터를 선택했을 때도 스펙트럼 왜곡이 심한 것을 볼 수 있다. 실제로 음성을 들어보면 47 개의 고유 벡터를 선택했을 때부터 합성된 음성에 왜곡이 심해졌다. PCA를 사용할 경우 선형 변환된 LSP 계수뿐만 아니라 차수가 제한된 고유 벡터 값도 같이 전송해야 하므로 QCELP에서 보여지는 결과에서는 정보의 압축을 기대하기 어려운 것으로 관찰되었다.

표 2. 한 프레임 당 비트 수

	기존 방법	PCA 적용
MELP	25 bit	20 bit
QCELP	40 bit	37 bit

표 2는 PCA를 사용했을 때, 한 프레임 당 LSP 비트수를 보여준다. 단, 고유 벡터의 비트 수는 계산에 포함시키지 않았다. 이는 프레임의 크기가 일정 크기 이상이 되면 상쇄되는 값이기 때문이다. MELP의 경우 계수의 차수가 절반으로 감소된 것에 비해 비트 수의 감소량은 크지 않다. 이것은 변환된 LSP 계수에 스칼라 양자화를 적용했기 때문이다. 기존의 MELP는 MSVQ(Multi-stage vector quantization)을 사용하기 때문에 25bit만으로 LSP 계수를 표현할 수 있다. 전송해야 할 고유 벡터의 크기를 고려하면 25개의 고유 벡터를 선택했을 때 프레임의 크기가 200개 이상이면 부호량이 줄어들기 시작한다. 따라서 PCA를 사용해 LSP 계수를 압축하기 위해서는 장시간 음성이 필요하다.

6. 결론

본 논문에서는 PCA를 이용해 LSP 계수를 압축하는 방법에 대해 분석했다. QCELP와 MELP 두 개의 음성 부호화기를 통해 실험한 결과 MELP에서 보다 나은 성능을 보여주었다. 하지만 PCA가 요구하는 계산량과 지연 시간을 고려하면 이 방식을 실시간 전송에 적용하기는 어려울 것으로 보인다. 특히, 고유 벡터를 같이 전송해야 한다는 점과 양자화 방식의 한계 때문에 PCA를 사용해 LSP를 압축하는 기법은 장시간의 음성에 대해서만 제한적으로 적용할 수 있음을 볼 수 있다. MELP에서 25개의 고유 벡터를 선택했을 때, 프레임 수가 200 개 이상이면 프레임 당 5비트 씩 데이터 양이 감소한다는 것을 확인했다.

참 고 문 헌

1. 김승주, 오영환, "제한된 시간적 분해법에 기반한 선스펙트럼 주피수의 효과적인 양자화," 한국음향학회지, 1998 vol.17, no. 4, pp. 45-53
2. Itakura, "Line spectrum representation of linear predictive coefficients of speech signals," J. Acoustic Soc. America, vol.57, Apr. 1975.
3. Hai Le Vu, Laszlo Lois, "Optimal transformation of LSP parameters using neural network," Proc. ICASSP, 1997, pp. 1339-1342
4. I. T. Jolliffe, Principal component analysis, Springer-Verlag, 1986
5. Kuldip K. Paliwal, Bishnu S. Atal, "Efficient vector quantization of LPC parameters at 24 bits/frame," IEEE Trans. on Speech and Audio Processing, vol.1, no.1, pp. 3-14, Jan. 1993