

SAW 필터를 이용한 KLT 기반 Entropy-Constrained Quantizer 성능 향상

임동석, 김무영*

세종대학교, 정보통신과

lds1988lds@naver.com, mooyoung@sejong.ac.kr* (교신저자)Perceptual Quality Improvement of KLT based Entropy-Constrained
Quantizer using a SAW FilterDong-Seok Lim, Moo Young Kim*
Sejong University

요 약

KLT-AECQ 는 지각적인 성능 향상을 위하여 formant weighting 필터를 사용한다. Code Excited Linear Prediction(CELP) 코더는 사람의 음성신호를 압축하는 대표적인 방식이다. CELP 의 Rate-Distortion 성능을 향상 시키기 위해서 Karhunen-Loeve-Transform (KLT) 기반의 Classified Vector Quantization (KLT-CVQ) 방식이 제안되었으며, 이는 KLT 기반의 Adaptive Entropy-Constrained Quantization (KLT-AECQ) 방식으로 확장되었다. 기존의 KLT-AECQ 에서는 지각적인 성능 향상을 위하여 formant weighting 필터를 사용한다. 본 논문에서는 이 필터 대신에 Spectral Amplitude Warping (SAW) 필터를 적용함으로써, KLT-AECQ 코더의 지각적인 성능을 향상하였다.

1. 서론

음성을 압축하고 전송하는 방법으로 Code Excited Linear Prediction(CELP) 코더가 많은 곳에서 사용되고 있다. CELP 코더는 음성이 입력으로 들어 왔을 때 Linear Prediction Coding (LPC) 계수를 구하고, LPC 계수를 사용하는 합성필터를 이용하여 음성 신호를 합성한다. 하지만, 기존의 CELP 코더는 음성을 양자화 할 때, vector quantization (VQ) 방식의 장점을 충분히 이용하지 못하는 단점이 있다. 기존 CELP 코더의 단점을 보완하기 위하여 Karhunen-Loeve-Transform (KLT) 기반의 Classified Vector Quantization (KLT-CVQ) 방식이 제안되었다. 제안된 KLT-CVQ 는 KLT 도메인에서 양자화 과정이 이루어지기 때문에, vector quantization (VQ) 방식의 장점을 충분히 이용할 수 있는 장점이 있으며, 이를 통해서 기존의 CELP 방식보다 뛰어난 rate-distortion (R-D) 성능을 제공한다. 제안된 KLT-CVQ 를 Entropy-Constrained Quantization (ECQ) 방식으로 확장하여 KLT 기반의 Adaptive ECQ(KLT-AECQ)가 제안되었다 [1]. KLT-AECQ 는 KLT-CVQ 에서 사용하는 양자화 방법인 Resolution-Constrained Quantization(RCQ) 방식이 아닌 ECQ 방식을 사용하기 때문에 더 뛰어난 R-D 성능을 기대할 수 있다.

음성을 압축하는 소스코딩의 목적은 가능한 적은 비트를 이용하여 압축하고, 압축한 음성을 복원하였을 때 복원된 음성과 원래 음성 사이의 왜곡을 최소화 하는 것이다. 적은 비트를 이용하여 압축을 하기 위해서는 사람이 인지하지 못하는 불필요한 정보인 irrelevancy 를 없애는 과정이 필요하다. 즉, 사람이 느끼지 못하는 정도의 왜곡은 허용하는 범위에서 비트를 할당하는 양자화를 진행하는 방법이 있다. 대표적인 방법으로는 오디오 코딩의 masking 을 이용하는 방법과 CELP 코더의 formant weighting 필터가 있다. Masking 이나 formant weighting 필터는 사람의 귀에 들리지

않는 주파수 대역의 양자화 잡음에 대해서는 비트를 적게 사용하여 양자화를 진행한다. 반대로, 사람의 귀에 잘 들리는 대역에서는 비트를 더 사용하여 양자화를 진행한다. 따라서 비트를 청각 특성에 맞게 효율적으로 분배하여 비트율을 줄이고 지각적인 성능을 향상 시킬 수 있다. [2]에서는 지각적인 성능을 향상시키기 위한 새로운 방법으로 SAW 필터가 제안되었다. 본 논문에서는 SAW 필터를 제안된 KLT-AECQ 코더에 적용함으로써 지각적인 성능의 향상을 얻을 수 있었다.

2. 본론

CELP 코더의 단점을 보완하고 기존의 KLT-CVQ 를 ECQ 방식으로 확장한 KLT-AECQ 는 기존 CELP 코더에 비해 향상된 R-D 성능을 제공한다. KLT-AECQ 는 0.625ms 의 low delay 를 가지는 음성 코딩 방식이며, 따라서 프레임 간 신호의 연관성이 높게 된다. 이러한 높은 연관성으로 인하여 LPC 계수를 입력신호가 아닌 복원 신호에서 backward adaptation 방식을 사용하여 구할 수 있으며, 따라서 추가적인 비트 전송 없이 인코더와 디코더 단에서 같은 LPC 계수를 공유할 수 있다.

입력신호 x^k 에 대한 공분산 행렬을 고유치 분해를 이용하여 고유값 행렬 D 와 고유 벡터 U 로 분해하면 다음과 같다.

$$E[x^k x^{kT}] = U^T D U \quad (1)$$

입력신호 x^k 에 대해서 formant weighting 이 적용된 zero state response 신호를 s^k 라고 하면, 고유벡터 행렬 U 를 곱하여 다음과 같이 KLT 도메인의 신호를 얻을 수

있다.

$$y^k = Us^k \quad (2)$$

제안된 KLT-AECQ 에서는 y^k 를 입력 신호로 고려하여 arithmetic 코딩 방법을 사용한 ECQ 방식으로 코딩 하게 된다. 이 때, 인코딩과 디코딩 단계에서 입력 신호에 대한 probability density function (PDF)를 별도의 비트 전송 없이 adaptive 하게 추정할 수 있으므로, arithmetic 코딩 방법 또한 추정된 PDF 를 사용하여 adaptive 하게 디자인 할 수 있다. 이렇게 인코딩된 입력신호를 디코딩 부분으로 전송하여 복원한 후, inverse KLT 과정을 통하여 원래의 음성신호 도메인으로 변환하여 복원된 음성신호를 구할 수 있다.

본 논문에서는 KLT-AECQ 코더의 지각적인 성능을 높이기 위해서 SAW 필터를 적용하였다. KLT-AECQ 코더의 입력신호 $x^k = \{x(0), x(1), \dots, x(k-1)\}$ 를 FFT 한 신호를 $X^k = \{X(0), X(1), \dots, X(k-1)\}$ 라고 하면, 그림 1 의 SAW filter 부분은 다음과 같은 식으로 표현할 수 있다.

$$X_w(k) = X(k) \frac{|X(k)|^\alpha}{|X(k)|} \quad (3)$$

즉, $X(k)$ 를 주파수 도메인 상에서 변형하는 과정을 통하여 $X_w(k)$ 를 구하게 된다. 수식 (3)에서 α 값은 상수로써, α 값에 의해 주파수 도메인에서 magnitude 가 변형되는 정도를 결정할 수 있다. 주파수 도메인에서 변형이 이루어진 스펙트럼 $X_w^k = \{X_w(0), X_w(1), \dots, X_w(k-1)\}$ 는 inverse FFT 를 거쳐 시간 도메인 신호인 $x_w^k = \{x_w(0), x_w(1), \dots, x_w(k-1)\}$ 를 구할 수 있다. 이렇게 얻어진 x_w^k 를 KLT-AECQ 의 입력신호로 이용하면 양자화된 신호인 \hat{x}_w^k 를 얻게 된다. \hat{x}_w^k 를 FFT 해서 얻은 신호를 $\hat{X}_w^k = \{\hat{X}_w(0), \hat{X}_w(1), \dots, \hat{X}_w(k-1)\}$ 라고 하면, (3)의 역과정인 de-emphasis 를 아래와 같이 수행하게 된다.

$$\hat{X}(k) = \hat{X}_w(k) \frac{|\hat{X}_w(k)|^{(1/\alpha)}}{|\hat{X}_w(k)|} \quad (4)$$

그리고, 마지막으로 $\hat{X}^k = \{\hat{X}(0), \hat{X}(1), \dots, \hat{X}(k-1)\}$ 를 inverse FFT 를 거쳐서 최종적인 출력신호 $\hat{x}^k = \{\hat{x}(0), \hat{x}(1), \dots, \hat{x}(k-1)\}$ 를 얻을 수 있다.

3. 실험 및 결과

실험 데이터로는 TIMIT database 에서 여성 화자 40 명과 남성 화자 70 명에 대해서 각각 한 문장씩 총 110 문장을 추출해서 8kHz 로 다운 샘플링 하여 구성 하였다. 지각적인 성능을 테스트 하기 위해서 ITU-T 에서 표준화된 Perceptual Evaluation of Speech Quality (PESQ) measure 를 사용하였다.

지각적인 성능 향상을 확인하기 위하여 KLT-

AECQ 코더에 perceptual weighting filter 와 SAW filter 를 적용한 경우와 아무 필터도 사용하지 않은 경우에 대해서 비교 실험을 실시 하였다. SAW 필터의 상수 α 는 실험을 통해서 0.6 으로 고정하여 사용하였다.

표 1 의 실험 결과를 보면 알 수 있듯이 필터 사용에 따라 지각적 성능이 변하게 되는데 기존 KLT-AECQ 에서 사용하는 formant weighting 필터를 적용하였을 때 보다 SAW 필터를 적용한 경우에 약 0.1 정도의 PESQ score 성능이 증가한 것을 알 수 있다. 또한, SAW 필터를 적용한 경우 아무 필터도 적용하지 않은 경우와 비교하면 0.2 정도의 PESQ 향상을 보였다.

3. 결론

기존의 KLT-AECQ 에서 사용한 formant weighting 필터 는 지각적인 성능을 향상 시키는 것에 한계가 있었다. 따라서 본 논문에서는 SAW 필터를 적용하여 기존의 KLT-AECQ 의 지각적 성능을 향상시키는 방법을 제안하였다. KLT-AECQ 의 경우는 프레임의 길이를 매우 짧게 사용하므로 지각적인 측면에서 큰 향상을 기대 할 수 없었다. 따라서, KLT-AECQ 에서 프레임의 길이를 길게 사용하게 되면 SAW 필터를 적용하였을 때 더 높은 지각적 성능의 향상을 기대할 수 있을 것이다.

Acknowledgement

이 논문은 2012 년도 정부(지식경제부)의 재원으로 산업원천기반구축사업의 지원을 받아 수행된 연구임(No. 10037244).

참고문헌

- [1] Y. Lee and M. Y. Kim, "KLT-based adaptive entropy-constrained quantization with universal arithmetic coding," IEEE Trans. Consumer Electron., vol. 56, no.4, pp.2601-2605, 2010.
- [2] R. Lefebvre, C. Laflamme, "Spectral amplitude warping (SAW) for noise spectrum shaping in audio coding," ICASSP. Acoustics, Speech, and Process., vol.1, pp. 335-338, 1997.

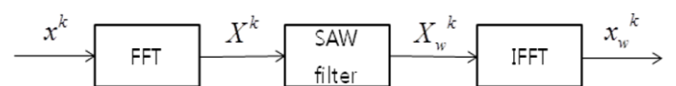


그림 1. SAW filter 블록 다이어그램

KLT-AECQ	PESQ score
without filter	4.0763
with formant weighting filter	4.1857
with SAW filter	4.2811

표 2. 필터 사용에 따른 PESQ 성능