

ZINC 함수 여기신호를 이용한 분석-합성 구조의 초 저속 음성 부호화기

*서상원, *김영준, *김종학, *김영주, *이인성
 *충북대학교 전자공학과
 e-mail : rokaf960@hanmail.net

Very Low Bit Rate Speech Coder of Analysis by Synthesis Structure Using ZINC Function Excitation

*Sang-Won Seo, *Young-Jun Kim, *Jong-Hak Kim, *Young-Ju kim *In-Sung Lee
 *Dept. of Radio Engineering, ChungBuk National University

Abstract

This paper presents very low bit rate speech coder, ZFE-CELP(ZINC Function Excitation-Code Excited Linear Prediction). The ZFE-CELP speech codec is based on a ZINC function and CELP modeling of the excitation signal respectively according to the frame characteristic such as a voiced speech and an unvoiced speech. And this paper suggest strategies to improve the speech quality of the very low bit rate speech coder.

I. 서론

초저속 음성 부호화 방법은 디지털 셀룰러 폰의 출현 및 군사적 목적을 위한 보안통신 시스템과 VOIP의 이용을 위하여 그 중요성이 증가하고 있다. CELP는 4~6KBPS의 중 전송률 부호화기에서 우수한 음질을 갖지만 초 저속 코더에서는 한계를 보여주고 있다.[1]

이에 본 논문에서는 저 전송률 및 초 저 전송률 음성부호화기를 위해 ZINC 함수와 CELP 모델링을 이용한 방법에 대해 서술한다. 또한 음질 향상을 위한 방법으로 분석-합성구조와 샘플 오버 샘플링 합성기법 및 위상제한에 대해 기술한다.

II. ZFE-CELP음성 부호화기

2.1 ZFE-CELP 부호화기 및 복호화기의 구조

제안된 초 저속 부호화 방법은 4khz의 대역폭을 갖는 음성 신호를 1.2kbps의 비트율을 갖는 음성 데이터로 변환하는 부호화 방법을 제공한다. 초 저속 부호화기는 입력 신호의 LPC를 분석 한 후 유/무성음 판별한 후 유성음일 경우는 잔여 신호로부터 1차, 2차 검색을 통해 피치를 얻은 후 표준 파형을 추출하여, Zinc 함수 여기 신호에 대한 파라미터를 얻는 과정을 거친다. 그리고, 무성음일 경우 피치 분석이 생략된 stochastic 코드북을 이용해 이득 값과 모양 파라미터를 추출 하는 CELP구조를 이용한다. 이러한 과정을 그림 1에 나타내었다.

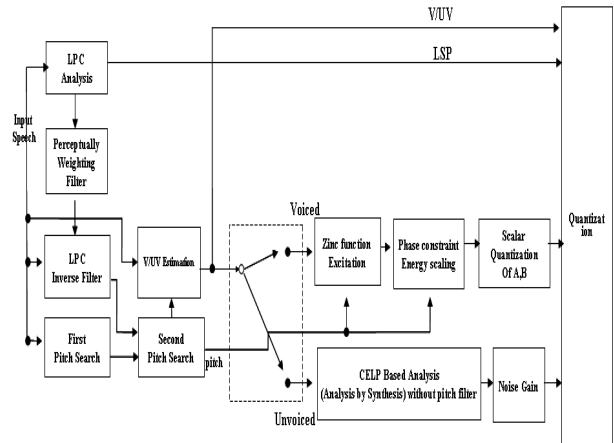


그림 1 ZINC 모델과 CELP 모델을 이용한 음성 부호화기의 인코더 블록다이어그램

2.2 분석-합성 구조를 이용한 표준 파형의 추출

일반적으로 표준파형은 이전 프레임의 표준파형과의 상관관계가 가장 큰 것을 표준파형으로 채택한다.[3] 하지만 한 프레임에서 표준파형이 잘못 채택되었을 경우 그 이후의 표준파형에도 영향을 미쳐 결과적으로 전체적인 음질의 저하를 초래할 수 있다. 본 논문에서 제안하는 ZINC 파라미터 분석-합성구조는 LPC 잔여 신호를 타겟으로 하여 전 구간에서 표준파형을 추출한 후, Sub-Optimal 한 ZINC 함수를 이용하여 합성된 신호와 타겟 신호와의 에러를 최소화 하는 파형을 표준파형으로 채택한다.

하지만 이렇게 전 구간에서 표준파형을 찾아내어 에러가 최소가 되는 표준파형을 채택하는 방식은 상당히 많은 계산량이 요구된다. 따라서 계산량을 줄이기 위해 상관관계 방법과 분석 합성 구조를 결합한 Fine prototype search method를 제안한다. 이 방법은 우선, 전단의 표준파형과 현재단의 표준파형의 상관관계를 계산하여 후보 표준파형을 정한다. 이렇게 결정된 후보 표준파형을 제한된 수만큼 분석-합성 구조를 이용하여 최적의 표준파형을 찾아낸다.

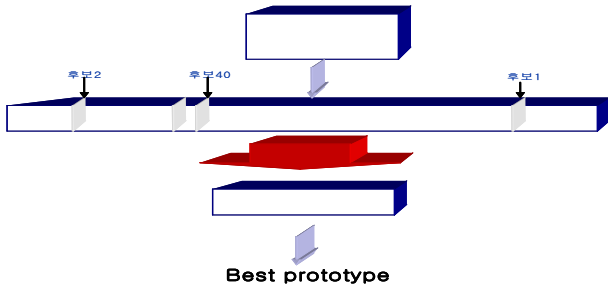


그림 2 Fine prototype search method의 구조

III. 음질 향상을 위한 전략

3.1 샘플 by 샘플 오버 샘플링

1.2kbps 초 저 전송률 보코더는 비트를 줄이기 위해 한 프레임의 길이를 40ms로 한다. 하지만 이럴 경우 프레임 간 하모닉의 불연속이 발생하게 되는데 이러한 불연속성을 해결하기 위해 선형위상 보간 합성법을 사용한다. 선형위상인 $\psi^k(l, w_0^k, n)$ 는 복호기에서 합성되는 정보이며, 다음과 같은 식으로 선형 위상을 합성할 수 있다.

$$\psi^k(l, w_0, n) = \psi^{k-1}(l, w_0^{k-1}, n) + \frac{l(w_0^{k-1} + w_0^k)}{2} n \quad (3.1.1)$$

식 (3.1.1)에서 볼 수 있듯이, 선형위상은 이전 프레임과 현 프레임의 시간에 따른 피치 각주파수를 선형 보간 하여 얻어진다. 복호기에서의 과형 합성은 부호화기로부터 전송된 ZINC 함수의 파라미터를 사용하여 수행 된다. 그런 다음 식 (3.1.1)에서 제시한 선형 위상합성방법을 사용하여 각 하모닉 크기들에 해당하는 위상정보를 만들어낸 후, 128-point IFFT(Inverse Fourier Transform)를 통해 기준 과형을 만들어 낸다. 이렇게 만들어진 기준 과형은 피치정보를 포함하지 않은 상태이기 때문에 순환형태로 재구성한 다음, 피치 주기로부터 얻은 오버 샘플링 비율로 피치변화를 고려하여 보간, 샘플링하여 최종 여기신호를 얻어낸다.

$$ov = \frac{256}{2T_p} = \frac{256/4}{T_p/2} = \frac{64}{l} \quad (3.1.2)$$

$$p_{ov}[n] = \sum_{i=0}^n \left(\frac{N-i}{N} ov^{k-1} + \frac{i}{N} ov^k \right) \quad (3.1.3)$$

3.2 위상제한

위에서 언급한 Fine method search method는 과거의 상관관계를 이용한 방법 보다 현재 프레임의 정보를 더 우선시하기 때문에 현재 값과 과거 값 사이에 프레임 간 불연속이 발생할 수 있다. 즉, 과거 값과 현재 값의 위상 값이 변하기 때문에 사이에 채워지는 값이 0에 가까운 되는 경우 위상반전이 일어난다. 이러한 위상 값에 의한 영향을 없애기 위해 과거 프레임 N-1과 현재 프레임 N의 위상이 반전될 경우 식 (3.2.1) 와 식(3.2.2) 와 같이 과거 프레임과 현재 프레임의 LPC 잔여 신호에 대한 RMS값의 비로서 과거 프레임의 A와 B를 스케일링 해 준다. 그리고 그 값을 현재의 표준과형에 대한 파라미터로 사용함으로써 심한 위상의 변화를 없애도록 하였다.

$$A_1(N) = \alpha A_1(N-1) \quad (3.2.1)$$

$$B_1(N) = \alpha B_1(N-1) \quad (3.2.2)$$

IV. 성능 평가 및 결론

4.1 성능평가

제안된 초저속 음성 부호화방법의 성능을 평가하기 위해 1.2kbps의 음성 부호화기로 구현하여 2.4kbps의 MELP 부호화와 비교해보았다. 성능평가 방법으로는 2001년 ITU-P.862로 승인된 PESQ(Perceptual Evaluation of Speech Quality)를 이용하여 MOS를 측정하는 방법과 미국 국방부(DoD)에서 2.4 kbps 음성 부호화기를 개발할 당시 사용했던 화자인식 시험을 시행하였다[3]. Table 1 에서는 MOS TEST 결과를 나타내었는데 한조의 1.9kbps ZFE-PWI coder의 MOS 2.602에 비해 거의 비슷한 MOS 2.595로 측정되었다. 그리고 Table 2 에서는 화자인식 시험 결과를 나타내었는데 제안된 ZFE-CELP 코더는 85%의 화자인식률을 보였는데 2.4kbps 코더에 비해 낮은 비트율에도 불구하고 높은 인식률을 나타내었다.

Table I. MOS TEST 결과

| 음성 부호화기 | MOS 값 |
|------------------------------|-------|
| 2.4KBPS MELP coder | 3.305 |
| 1.9KBPS Hanzo의 ZFE-PWI coder | 2.602 |
| 제안된 1.2KBPS coder | 2.595 |

Table II. 화자인식 시험 결과

| System | Processed-Processed | | | Processed-Processed (DIFFERENTs) | | |
|--------------------|---------------------|-----|-----|----------------------------------|------|------|
| | Avg. | 남자 | 여자 | Avg. | 남자 | 여자 |
| Unprocessed coder | 96% | 95% | 96% | 4.58 | 4.66 | 4.50 |
| 2.4KBPS MELP coder | 89% | 93% | 85% | 4.15 | 4.41 | 3.89 |
| 제안된 1.2KBPS coder | 85% | 87% | 82% | 3.94 | 4.25 | 3.62 |

4.2 결론

본 논문은 1.2kbps 초저전송률 음성 코더의 개발을 위해 Zinc function을 사용하여 과형 보간을 함으로써 유성음을 부호화하고, 무성음 부호화를 위해 CELP구조를 사용하는 방법을 제안하였다. 또한 표준과형의 정확한 추출을 위해 분석-합성 구조를 도입하였고, 계산량을 줄이기 위한 방법으로 Fine prototype method를 제안하였다. 그리고 하모닉의 연결성을 보장하기 위해 샘플 by 샘플 오버 샘플링 기법을 이용하였고, 위상반전 문제를 강제적으로 위상 제한함으로써 해결하였다. 결과적으로 Hanzo의 1.9kbps ZFE-PWI 합성 음성과 비교해 볼 때 비트율이 낮아졌음에도 불구하고 MOS 0.1정도의 차이로 유사한 성능을 얻을 수 있었다.

" 이 논문은 2006년도 교육인적자원부 지방연구중심대학 육성사업의 지원에 의하여 연구되었음."

참고문헌

- [1] A. M. Kondoz, "Code-Excited Linear Predictive Coding", Digital Speech, Chapter 6, pp. 174-212, 1994.
- [2] F. C. A. Brooks, Lajos Hanjo, "A 1.9kbps ZINC Function Excited, Waveform Interpolated Speech Codec", 1998 IEEE Vehicular Technology Conference, Ottawa, Canada, May 18-21, 1998
- [3] Schmidt-Nielsen, A. Brock, D.P. "Speaker recognizability testing for voice coders", Acoustics, Speech, and Signal Processing, 1996. ICASSP-96. Conference Proceedings., 1996 IEEE International Conference on Volume 2, 7-10 May 1996