한국음향학회지 제25권 제6호 pp. 282-290 (2006)

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

A Very Low-Bit-Rate Analysis-by-Synthesis Speech Coder Using Zinc Function Excitation

서 상 원*, 김 종 학*, 이 창 환*, 정 규 혁*, 이 인 성*
(Sang Won Seo*, Jong Hak Kim*, Chang Hwan Lee*, Gyu-Hyeok Jeong*, In Sung Lee*)

*충북대학교 전파공학과

(접수일자: 2006년 8월 23일; 채택일자: 2006년 9월 4일)

본 논문에서는 1.2 kbps의 전송률을 가지는 초 저속 음성 부호화기를 위한 방법과 구조를 제안한다. ZFE-CELP (Zinc Function Excitation-Code Excited Linear Prediction) 음성 부호화기는 선형예측 분석 후, 추출된 잔여 신호가 유성음일 경우 Zinc Function을 이용하여 부호화하고, 무성음일 경우에는 CELP구조를 이용하여 부호화한다. 또한 Super-frame (40ms)의 영향으로 발생하는 하모닉의 불연속 문제를 해결하기 위해 오버 샘플링을 이용한 선형 위상 합성 기법을 이용하고 Zinc 함수의 정확한 표준파형을 추출하기 위하여 분석-합성 구조를 제안한다. 제안된 초 저속 음성 부호화기의 성능을 2.4 kbps의 MELP (Multi Pulse Linear Prediction) 부호화기 및 1.9kbps의 ZFE-PWI (Zinc Function Excitation-Prototype Waveform Interpolation) 음성 부호화기와 비교하였다. 제안된 부호화 방법은 1.9 kbps ZFE-PWI 부호화기와 유사한 성능을 보이는 것을 확인하였다.

핵심용어: 초 저속 음성 부호화기, Zinc Function, CELP, 표준 파형, 분석-합성 구조

투고분야: 음성처리 분야 (2.2)

This paper presents the algorithm and the structure for very low bit rate speech coder with 1.2 kbps. ZFE-CELP (Zinc Function Excitation-Code Excited Linear Prediction). After LPC analysis, the ZFE-CELP speech coder 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. Also we designed the linear phase synthesis method using over-sampling to solve the discontinuation of harmonic in Super-frame (40 ms) and the analysis by synthesis structure to extract optimum prototype. The performance of proposed speech coder compared with 2.4kbps MELP (Multi Pulse Linear Prediction) speech coder and 1.9 kbps ZFE-PWI (Zinc Function Excitation-Prototype Waveform Interpolation) speech coder. In spite of the lower bit rate, the performance of the proposed 1.2 kbps ZFE-CELP speech coder is similar to 1.9 kbps ZFE-PWI speech coder.

Key words: Very low bit rate speech coder, Zinc function, CELP, Prototype waveform, Analysis by synthesis structure

ASK subject classification: Speech Signal Processing (2.2)

I. 서 론

초 저속 음성 부호화 방법은 디지털 셀룰러 폰의 출현 및 군사적 목적을 위한 보안 통신 시스템과 VoIP의 이용을 위하여 그 중요성이 증가하고 있다. 특히 군사적 목적을 위한 보안 통신시스템과 위성 통신에서 채널상태에 따른 가변 비트율의 저속 음성 부호화기의 연구가 활발히 진행되고 있다 [1]. 4~16 kbps의 전송률의 부호화기에서는 우수한 음질을 갖는 CELP (Code Excited Linear Prediction) [2] 구조가 QCELP, CS-ACELP, VSELP등의 다양한 방법으로 GSM, CDMA등의 이동통

책임저자: 이 인 성 (inslee@cbucc.chungbuk.ac.kr)
충북 청주시 흥덕구 개신동 충북대학교 전파공학과
(전화: 043-261-3223; 팩스: 043-261-4647)

신시스템에서 사용되고 있다. 하지만 CELP 코더는 국제 표준으로 성능이 뛰어 나지만 단독 모델로는 저 전송률 및 초 저 전송률에서는 고품질의 음질을 얻을 수 없는 한계를 보여주고 있다. 그래서 저 전송률 및 초 저 전송률에서도 좋은 음질을 낼 수 있는 음성부호화기의 개발을 위해 4 kbps이하의 음성 부호화기에 대한 연구가 활발히 진행되었고 LPCe를 개선하기 위한 목적으로 개발된 MELP (Mixed Excited Linear Prediction) [3] 부호화기가 2.4 kbps DoD코더로 표준화 되었다. 뿐만 아니라, STC (Sinusoidal Transform Coding), MBE (Multiband Excitation)와 같은 하모닉 코더와 표준 파형을 이용한 PWI (Prototype Waveform Interpolation)와 같은 후보군들도 계속된 개발로 매우 우수한 음질을 내고 있다 [4][5][6]. 하지만 낮아진 비트 전송률에 의한 스펙트럴 왜곡이나 프레임간 불연속성이 문제로 제기되고 있다.

이에 본 논문에서는 초 저 전송률 음성 부호화기를 위해 유성음 구간에서는 인지적 음질과 비트율에서 우수한 성능을 나타내는 Zinc 함수를 이용하여 부호화하고, 무성음 구간은 CELP를 이용한 효율적인 부호화방법을 서술한다. 그리고 종래의 Zinc 함수는 이전 프레임과의 상관도만을 이용하여 현재 프레임의 표준파형을 추출하는데 [7], 만약 과거의 표준파형이 전이구간에서 잘못 선택되었을 경우 프레임이 증가함에 따라 음질의 저하가 가중 되는 문제가 발생한다. 이에 본 논문에서는 상관도와 분석-합성 구조를 결합하여 표준파형을 추출하는 방법을 제안 한다. 또한 비트 전송률이 낮아짐에 따라 생기는 주파수 왜곡 문제를 오버 샘플링을 이용한 선형 위상 합성 기법으로 해결하고, 프레임 사이에서 발생하는 위상 반전 문제를 강제적으로 제한 해줌으로써 프레임간의 불연속 문제를 해결한다.

본 논문의 2장에서는 ZFE-CELP 초 저 전송률 음성 부호화기의 전체적인 구조에 대해 설명하고, 3장에서는 Zinc함수를 이용한 유성음 모델링 방법을 4장에서는 CELP 구조를 이용하여 무성음을 부호화 하는 방법에 대해 설명한다. 그리고 5장에서는 음성 파라미터의 양자화 방법 및 비트할당에 대해 서술한다. 6장에서는 성능평가를 7장에서는 결론을 맺는다.

II. ZFE-CELP 초 저속 음성 부호화기의 구조

2.1. 인코더의 구조

ZFE-CELP 음성 부호화기는 4 kHz의 대역폭을 갖는 음성 신호를 1.2 kbps의 비트율을 갖는 음성 데이터로 변환하는 부호화 방법을 제공한다. ZFE-CELP의 프레임 구간은 40 ms이며 두 개의 부 프레임으로 나뉜다. 부호화 과정은 우선 입력 신호의 LPC를 분석한 후, 유/무성음을 판별한다. 모드 결정은 20 ms 프레임마다 결정된다. 이러한 결정방법은 합성된 스펙트럼과 원본 스펙트럼의 유사성 및 신호 파워 값, LPC 잔여신호의 파워로 정규화된 최대 자기 상관 값, 영 교차율 (Zero Crossing Rate) 값을 사용한다. 정규화 된 자기상관 값은 지연 값이 커질수록 작은 값을 가지는데, 주기성이 클수록 각 피크치의 감소율이 낮은 특성을 나타낸다. 이러한 정규화 된 자기상관 값의 첫 번째와 두 번째 피크의 비율이 유/무성음을 판별하는데 이용된다. 유/무성음을 판별하여 유성음일 경우는 잔여 신호로부터 1차, 2차 검색을 통해 피치를 얻고 표준 파형을 추출하여 Zinc 함수 여기신호에 대한 파라미터를 얻는 과정을 거친다. 그리고 무성음일 경우 피치 분석이 생략된 Stochastic 코드북을 이용해 이득 값과 모양 파라미터를 추출하는 CELP구조를 이용한다 [8]. 이러한 과정을 그림 1에 나타내었다.

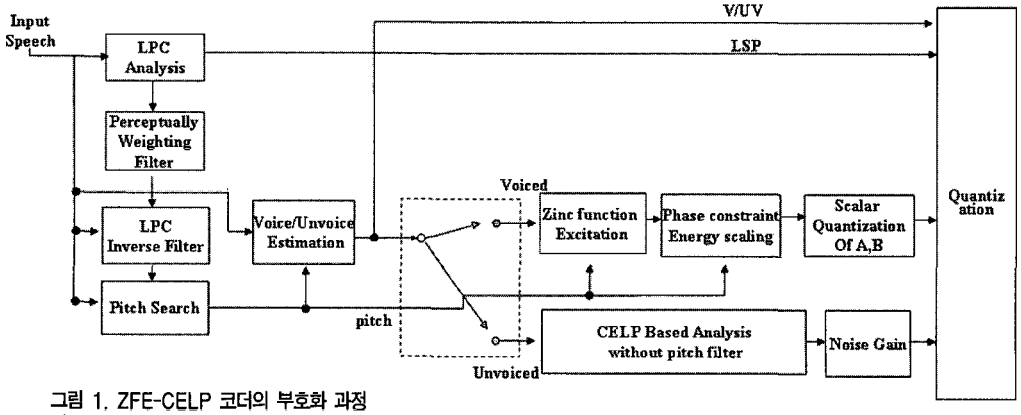

그림 1. ZFE-CELP 코더의 부호화 과정
Fig. 1. Block diagram of ZFE-CELP encoder.

2.2. 디코더의 구조

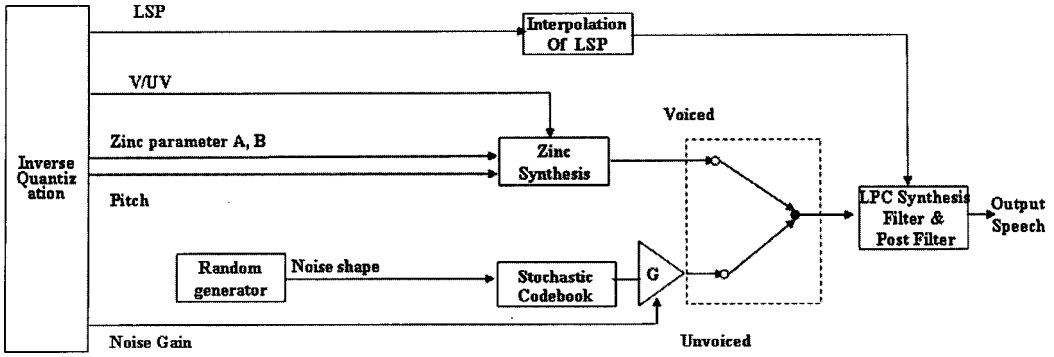

그림 2. ZFE-CELP 코더의 복호화 과정
Fig. 2. Block diagram of ZFE-CELP decoder.

그림 2는 ZFE-CELP 음성 부호화기의 복호화 과정을 나타내고 있다. 우선 LPC 계수는 2.5 ms 마다 과거와 현재의 LSP를 보간 함으로써 얻는다. 그리고 구해진 LPC 합성 필터에 유/무성음의 여기 신호를 통과시킨 후 후단 여파기를 통과시킴으로써 최종 합성 음성을 얻는 과정을 거친다. 이러한 과정을 그림 2에서 도시하고 있다.

III. Zinc 함수를 이용한 유성음 모델링

3.1. Zinc 함수 파라미터 추출

아산 시간의 경우 8 kHz의 샘플링 주파수를 가지며, cutoff 주파수로 4 kHz를 가지는 Zinc 함수 [9]는 식 (1)과 같이 $\sin c$함수와 $\cos c$함수로 나타낼 수 있다.

$$z_k(n) = A_k \sin c(n-\lambda_k) + B_k \cos c(n-\lambda_k) = \begin{cases} A_k & , n-\lambda_k = 0 \\ \frac{2B_k}{n\pi} & , odd \\ 0 & , even \end{cases} \quad (1)$$

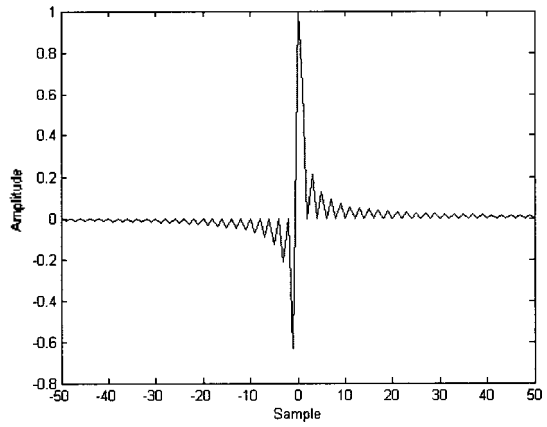

그림 3. 표준 파형의 여기 신호를 위한 Zinc 기저 함수
Fig. 3. Typical shape of a Zinc basis function for excitation signal of prototype.

여기에서 A_k, B_k는 Zinc 함수의 진폭 값을 나타내며, λ_k는 위치를 나타낸다. 그림 3은 $A_k = B_k = 1$이고, λ_k가 0인 경우를 도시한 것이다.

3.2. 표준 파형의 추출

음성은 연속된 유성음 구간에서 준-정적 (quasi-stationary)한 특성을 가지지만, 음성 시작점이나 끝점의 경우 순간적인 증가나 감소가 일어나는 경우가 많다. 따라서 무성음에서 유성음으로 변환하는 구간과 이 후의 구간에 대해 표준 파형을 추출하는 과정을 다르게 한다. 우선 무성음에서 유성음으로 변하거나 반대의 경우에는 프레임의 중심을 기준으로 피치 주기에 해당하는 구간에서 최대 값을 검색하여 준다. 그리고 이 값을 기준으로 영 교차 값을 시작점으로 설정하여 피치 주기 만큼의 길이를 가지는 표준파형을 추출한다 [9]. 반면에 이 후의 유성음 구간은 자기 상관함수에 의해전단의 표준 파형과 가장 유사한 파형을 추출한다.

3.3. 분석-합성 구조를 이용한 표준 파형의 추출

인간의 귀는 유성음 구간에서 민감도가 증가하게 된다. 따라서 유성음의 모델링은 음질에 많은 영향을 끼치게 된다. 일반적으로 Zinc 함수를 이용하여 유성음을 모델링 할 경우 표준 파형의 선택을 이전단의 표준파형과의 상관도에 의존하지만 표준파형이 잘못 선택되어질 경우, 그 영향이 계속해서 다음 프레임에 미칠 수 있어 결과적으로 전체적인 음질저하의 문제를 야기할 수 있다. 따라서 이러한 문제를 해결하기 위해 분석-합성 구조를 이용하여 표준파형을 추출하는 구조를 제안 한다.

본 논문에서 제안하는 Zinc 파라미터 분석-합성구조는 프레임마다 최적의 표준파형을 추출함으로써 음질의

향상을 가져올 수 있다. 기본적인 구조는 LPC 잔여 신호를 타겟으로 하여 전 구간에서 표준파형을 추출한 후 Sub-Optimal Zinc 함수를 이용해 합성하여 에러가 최소가 되는 표준파형을 추출하는 것이다. 하지만 이렇게 전 구간에서 표준파형을 찾아내어 에러가 최소가 되는 표준파형을 채택하는 방식은 많은 계산량이 요구된다. 따라서 계산량과 음질 두 가지 측면을 모두 고려한 방법을 제안 한다. 우선은 프레임 크기에서 피치주기를 뺀 수만큼의 표준파형을 만들어낸 후, 이전단과의 상관관계를 계산한다. 그리고 상관관계가 큰 순으로 후보 표준파형을 N개 선정하여 분석-합성 구조를 통해 에러가 제일 작은 표준파형을 최종 표준파형으로 채택한다. 여기서 N개는 계산량을 고려하여 선정해야 하는 개수이며 본 논문에서의 실험을 위해 N개를 40으로 선택하였다. 이수치는 225 MHz의 동작 속도를 가지는 TI사의 DSK6713 보드에서 실시간 구현이 가능한 수치이다. 이러한 과정을 그림 4에 나타내었다.

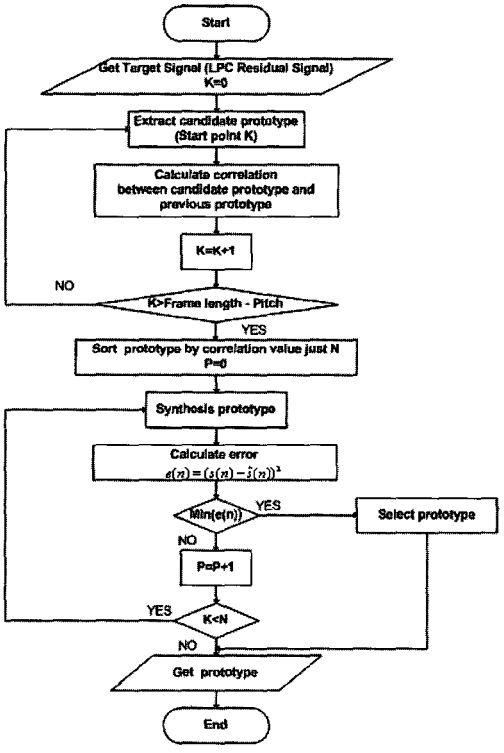

그림 4. 표준 파형 추출 순서도
Fig. 4. Flow chart to extract prototype.

3.4. 표준 파형간의 보간

초기 유성음 구간에 대해서는 보간을 하지 않는다. 초기 유성음 구간은 무성음에서 유성음으로 변하는 구간을 의미하며, 이러한 경우에는 현재 프레임에서 추출된 Zinc 파라미터를 이용하여 삼각 윈도우에 의한 합성을 하여 준다. 즉 프레임 경계를 기준으로 하여 피치 주기 만큼의 환형 버퍼 (circular buffer)를 이용하여 반복된 표준파형을 삽입하여 주게 된다. 여기에서 프레임의 경계를 기준으로 표준파형을 환형버퍼에 삽입하기 때문에 Zinc 함수의 위치정보를 1로 처리해야 한다. 이것은 위치 정보를 1로 처리하지 않을 경우, 그림 2와 같은 Zinc 함수를 표현할 수 없기 때문이다. 이후의 프레임은 과거와 현재의 표준 파형을 각각 보간 하여 준 후 중첩-합산을 통해 합성이 된다.

3.5. 오버 샘플링을 이용한 선형 위상 합성 기법

본 논문에서 제안하는 1.2 kbps 초 저 전송률 보코더는 비트를 줄이기 위해 한 프레임의 길이를 40 ms로 한다. 하지만 이 경우 프레임 간 하모닉의 불연속이 발생하게 되는데 이러한 불연속성을 해결하기 위해 선형위상 보간 합성법을 사용한다.

선형위상인 $\phi^k(l, w_0^k, n)$는 복호기에서 합성되는 정보이며, 식 (2)와 같이 선형 위상을 합성할 수 있다 [10][11].

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

여기에서 ϕ, w_0는 선형위상과 피치 각주파수를 의미하고 k, l, n은 각각 프레임 번호, 하모닉의 개수, 샘플 번호를 의미한다. 식 (2)에서 볼 수 있듯이, 선형위상은 이전 프레임과 현 프레임의 시간에 따른 피치 각주파수를 선형 보간 하여 얻어진다. 인간의 청각 시스템은 위상 연속성이 보존되는 동안 선형 위상에 비 감각적이며, 부정확한 또는 완전히 관이한 이산 위상을 허용할 수 있다 [11]. 이러한 지각적 특성은 저 전송률 코딩에 있어 하모닉 모델의 연속성에 대한 중요한 조건이 된다. 따라서 합성 위상은 측정된 위상을 대체할 수 있게 된다.

복호기에서의 파형 합성은 부호화기로부터 전송된 Zinc 함수의 파라미터와 스펙트럴 파라미터, 그리고 피치 파라미터 값을 사용하여 수행 된다. 우선, 기준 파형을 합성하기 위해, 스펙트럴 파라미터에서 역 양자화과정을 통해 하모닉 크기들을 추출한다. 그런 다음 식 (2)에서 제시한 선형 위상 합성방법을 사용하여 각 하모닉 크기들에 해당하는 위상정보를 만들어낸 후, 128-point IFFT (Inverse Fourier Transform)를 통해 기준 파형을 만들어 낸다. 이렇게 만들어진 기준 파형은 피치정보를 포함하지 않은 상태이기 때문에 순환형태로 재구성한 다음, 피치 주기로부터 얻은 오버 샘플링 비율로 피치변

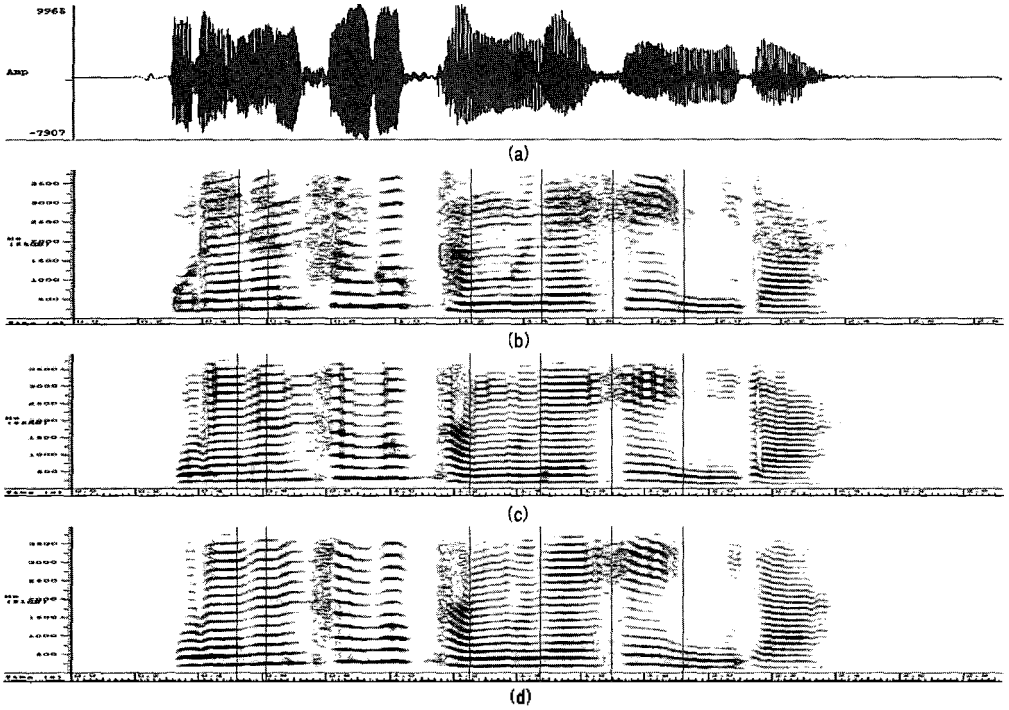

그림 5. 스펙트로그램 결과 ; (a) 시간 영역에서의 원본 음성 (b) 원본 음성 신호의 스펙트로그램 (c) 1.2kbps 음성 부호화기의 스펙트로그램 (d) 오버 샘플링을 이용한 선형 위상 합성 기법이 적용된 1.2kbps 음성 부호화기의 스펙트로그램

Fig. 5. Spectrogram results ; (a) Original speech signal in time domain (b) Spectrogram of original signal (c) Spectrogram of ZFE-CELP synthesized speech (d) Spectrogram of ZFE-CELP synthesized speechLinear phase synthesize method using over-sampling.

화를 고려하여 보간, 샘플링 하여 최종 여기신호를 얻어 낸다.

$$r_{ov} = \frac{256}{2T_p} = \frac{256/4}{T_p/2} = \frac{64}{l} \tag{3}$$

$$p_{ov}[n] = \sum_{i=0}^{n}\left(\frac{N-i}{N} r_{ov}^{k-1} + \frac{i}{N} r_{ov}^{k}\right) \tag{4}$$

식 (3)과 (4)는, 각각 오버 샘플률 (r_{ov})과 샘플링위치 ($p_{ov}[n]$)를 나타낸다. 여기서 N은 프레임 길이, T_p는 피치주기, l은 하모닉 개수, k는 프레임 번호를 나타낸다. 식 (3)에서 구해진 오버 샘플링 비율을 이용하여 식 (5)에서 샘플링 된 여기신호는 표준파형의 축소/확장을 의미한다. 또한, 순환 파형을 만들어내는 과정 중 프레임간의 부드러운 연속성을 보장하기 위해 현 순환 파형의 시작 지점을 바꾸는 방법이 필요하게 된다. 이를 위해 *offset*값을 정의하며, 이를 이용하여 각 프레임 순환 파형의 끝 지점을 한 개의 기준 파형 구간길이인 128로 맞춤으로써 다음 프레임의 첫 지점과 자연스럽게 이어진다 [12].

$$w'^{k-1}(l) = w^{k-1}(\mathrm{mod}(l, 128)) \tag{5}$$

$$w'^{k}(l) = w^{k}(\mathrm{mod}(offset + l, 128)) \tag{6}$$

$$offset = 128 - \mathrm{mod}(L, 128) \tag{7}$$

여기서, L은 N개의 샘플을 복원시키기 위해 오버 샘플링되는 데이터 개수이며, $\mathrm{mod}(x, y)$는 x를 y로 나눈 나머지 값을 돌려준다. 또한 $w'^{k}(l)$은 k 번째 순환 파형을, $w^{k}(l)$은 k 번째 기준 파형을 나타낸다. 식 (7)의 *offset*값은 이전 프레임의 마지막 순환파형의 잘린 위치를 의미한다. 따라서 현제 프레임의 순환파형은 이전 프레임의 마지막 순환파형이 잘리지 않도록 *offset*값을 시작점으로 하여야 한다. 실제로 구현된 IFFT 합성은 피치의 빠른 변화를 고려하여 그 비율이 고정 값 이상으로 크면 과거 및 현재 순환 파형에 대한 각각의 피치로 샘플링을 한 다음 부분선형 보간 합성하는 방식을 취한다. 반면, 피치의 변화가 작다면 과거 현재 순환 파형을 미리 선형 보간 한 다음 과거와 현재 피치의 중간 값으로 샘플링 하여 최종 여기 합성신호를 얻는다. 그림 5는 오버 샘플링을 이용한 선형 위상 합성기법에 대한 실험결과이다.

그림 5에서 (c)와 (d)의 표시된 부분의 스펙트로그램을 살펴보면 선형 위상 합성 기법이 사용되지 않은 음성 부호화기의 출력 스펙트로그램에서는 하모닉이 계단처럼 끊어지는 것을 볼 수 있지만 오버 샘플링을 이용한 선형 위상 합성 기법을 이용한 음성 부호화기의 스펙트로그램에서는 하모닉이 부드럽게 연결되어 원본의 스펙

트럼 왜곡을 해결하는 것을 볼 수 있다.

3.6. 위상 제한

Zinc 함수를 이용한 여기신호의 모델링에서는 pulse를 사용하여 여기신호를 추출하기 때문에 별도의 위상 값을 할당하지 않는다. 다만, A_1, B_1의 부호에 의해 과거 프레임과 현재 프레임의 위상 성분을 결정하게 된다. 과거프레임에서 추출된 A_1, B_1이 현재 프레임에서 반대로 될 경우 여기 신호가 급작스런 위상 반전이 자주 나타나는 것을 확인할 수 있다. 이것은 인접한 두 표준파형 간에 보간 과정에서 위상의 변화로 인해, 표준파형 사이에 생성된 파형이 0에 가까운 값으로 바뀌면서 연결이 되기 때문이다. 따라서 이러한 위상 값에 의한 영향을 제거하기 위해 과거 프레임 $N-1$과 현재 프레임 N의 위상이 반전될 경우 식 (8)과 식 (9)와 같이 과거 프레임과 현재 프레임의 LPC 잔여 신호에 대한 RMS (Root Mean Square)값의 비로서 과거 프레임의 A와 B를 스케일링 해 준다 [7]. 그리고 그 값을 현재의 표준파형에 대한 파라미터로 사용함으로써 급격한 위상 변화가 제거되도록 하였다.

$$A_1(N) = \alpha A_1(N-1) \tag{8}$$

$$B_1(N) = \alpha B_1(N-1) \tag{9}$$

여기서, α는 N번째 프레임의 LPC 잔여 신호의 RMS를 $N-1$번째 프레임의 LPC잔여 신호의 RMS로 나눈 값이다.

IV. CELP 기반 무성음 모델링

우선 무성음 구간의 잡음과 같은 여기신호를 부호화하기 위해 Stochastic 코드북을 사용한 분석-합성 방법이

표 1. 1.2 kbps 음성 부호화기의 비트 할당
Table 1. Bit allocation of 1.2 kbps speech coder.

전송률	1.2 kbps				
LSPs	1차 상태 : 6 bits				
	2차 상태 :10 bits				
Mode	2 bits				
Frame	Unvoiced (1st Subframe)			Voiced (2nd Subframe)	
Excitation 관련 인덱스 값	10ms	이득 : 5 bits			피치 : 7 bits
	10ms	이득 : 5 bits		20ms	진폭 A : 4 bits
	Reserved : 5bit				진폭 B : 4 bits
	Voiced (1st Subframe)			Unvoiced (2nd Subframe)	
	20ms	피치 : 7 bits		10ms	이득 : 5 bit
		진폭 A : 4 bits		10ms	이득 : 5 bit
		진폭 B : 4 bits		Reserved : 5bit	
	Voiced (1st Subframe)			Voiced (2nd Subframe)	
	20ms	피치 : 7 bits			피치 : 7 bits
		진폭 A : 4 bits		20ms	진폭 A : 4 bits
		진폭 B : 4 bits			진폭 B : 4 bits
	Unvoiced (1st Subframe)			Unvoiced (2nd Subframe)	
	10ms	이득 : 5 bits		10ms	이득 : 5 bits
	10ms	이득 : 5bits		10ms	이득 : 5 bits
	Reserved : 1bit			Reserved : 1 bit	
	3차 상태 : 8 bits				
Total	48 bits				

사용되며, 식 (10)과 같이 왜곡측정치가 최소가 되는 이득과 모양벡터를 찾아낸다.

$$e = \sum_{n=0}^{N-1} (ref(n) - G \cdot syn(n))^2 \tag{10}$$

그림 6은 이러한 방법을 기본으로 하는 Stochastic Pulse 여기 코더의 구조를 나타내고 있다. 여기서, $H(z)$은 LPC 합성 필터이며, $ref(n)$ 은 지각 가중치 된 LPC 합성 필터를 사용한 입력신호의 ZSR (Zero State Response), $syn(n)$은 모양벡터 코드북 값에 의한 여기입력 신호로부터 유도된 ZSR이다. 여기에서 모양벡터의 코드북 값은

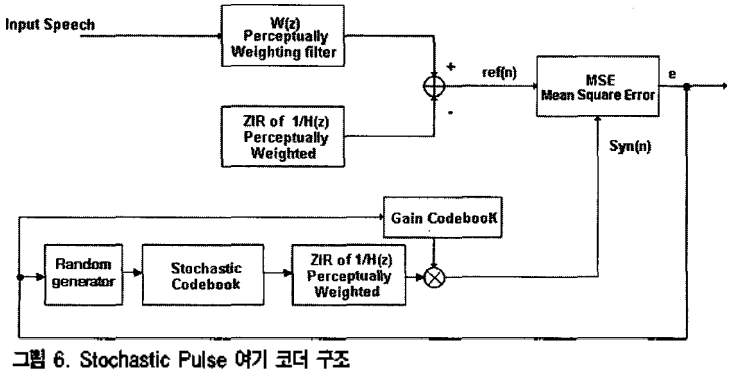

그림 6. Stochastic Pulse 여기 코더 구조
Fig. 6. Structure of Stochastic Pulse excitation coder.

랜덤 넘버 발생기에 의해 결정된다. 이것은 모양 벡터에 할당된 비트를 스펙트럴 정보에 추가로 할당함으로써 적은 비트를 효율적으로 사용하고, 성능을 개선하기 위함이다. G는 크기벡터 코드북에 의한 크기 값이며, 식 (10)의 에러를 최소화하는 값을 찾는다 [2].

V. 비트 할당 및 양자화 방법

양자화 방법은 크게 LSP 양자화 과정과 여기신호 양자화 과정으로 구성된다. 우선 LSP 양자화 과정은 기본적으로 2단계 분할 벡터 양자화 (2-stage Split Vector Quantization)를 적용하여 16비트를 할당하였고, 한 프레임 (40 ms) 동안의 유성음과 무성음 및 혼합구간을 표현하기 위하여 각각의 모드에 2비트를 할당하였다. 따라서 1.2 kbps에 할당된 총 48비트 중 30비트가 여기신호 표현을 위해 할당되며 모드에 따라 비트할당이 바뀌게 된다. 이 중 유성음의 경우 Zinc 파라미터는 진폭값 A, B 두 개와 위치 파라미터 λ이지만, 디코더 단에서 인접한 표준파형의 보간 과정을 통해 위치 파라미터 값은 생략이 가능하다. 따라서 양자화는 진폭값 A, B만 수행한다. A와 B는 벡터 양자화를 이용하여 각각 4비트로 양자화 한다. 또한 유성음의 여기신호 합성 시 필요한 피치정보는 7비트가 할당된다. 무성음의 경우는 효율적인 비트 할당을 위해 2단계 부 프레임을 두어 이득에 5비트를 할당하고 형태는 디코더에서 랜덤으로 생성된다. 따라서 형태에 쓰였던 비트는 LSP의 잔차를 표현하기 위해 할당된다. 즉 두 개의 부 프레임이 무성음일 경우에 LSP는 3단계 분할 양자화기를 사용하게 되며 8비트가 할당된다. 그리고 각각의 프레임에 남겨진 비트는 FEC (Forward error correction)에 할당할 예정이다. 이것을 표 1에 나타내었다

VI. 성능 평가

실험에 사용한 음성은 KIST 음성 DB[13]중 잡음이 없는 한국어 음성 40개 (남자 20개, 여자 20개)를 사용하였고, MOS 값을 통해 2.4 kbps MELP 음성 부호화기와 Hanzo의 1.9 kbps ZFE-PWI 음성 부호화기와 비교하였다. MOS 값은 2001년 ITU-P .862로 승인된 PESQ (Perceptual Evaluation of Speech Quality)를 이용하였다. 표 1은 MOS 실험 결과를 나타내고 있다. 2.4 kbps MELP와 1.9 kbps ZFE-PWI 의 코더의 낮아진 비트수에도 불구하고 좋은 성능을 나타내는 것을 볼 수 있다. 특히 1.9 kbps 음성 부호화기와 유사한 성능을 나타내는 것을 확인할 수 있다.

표 2. MOS 실험 결과
Table. 2. Results of MOS TEST.

음성 부호화기	MOS 값
2.4kbps MELP coder	3.065
1.9kbps Hanzo의 ZFE-PWI coder	2.602
1.2kbps ZFE-CELP coder	2.595

현재 저속 음성 부호화기의 사용자들은 좋은 음성의 질과 명료함뿐만 아니라 화자의 인식을 요구한다. 화자인식은 미국 국방부 (DoD) [14]에서 2.4 kbps 음성 부호화기를 개발할 당시 선택 제한조건 중 하나였다. 따라서 화자 인식 실험 방식은 미국 국방부에서 개발한 화자인식 실험을 기본으로 한다. 기본적으로 화자인식 실험은 두 개로 나누어질 수 있는데 첫 번째는 각각의 코더가 화자의 동일성을 어느정도 유지하고 있는가를 측정하는 것이고, 두 번째는 화자들을 구별하는데 필요한 정보를 얼마나 잘 유지하고 있는가를 측정하는 것이다. 이 방법들을 SAME-DIFFERENT method [14]라 한다. 실험에 참가할 인원은 남자 5명과 여자 5명으로 화자인식에 필요한 sentences pairs를 제공한다. 또한 실험 data를 제공할 사람들은 테스트를 할 실험자와는 알지 못하는 사람으로 구성한다. 이는 화자의 버릇이나 습관을 알고 있으면 순수하게 코더에 의한 화자인식능력을 측정할 수 없기 때문이다. 화자인식 실험은 8명의 훈련되지 않은 청취자들이 실시하였다. 이 결과를 표 3에 나타내었다. 첫 번째 실험은 각각 음성 부호화기의 출력 쌍이 같은지 다른지를 평가하는 실험이고 두 번째 실험은 다름의 정도를 5점 (아주 다름)에서 1점 (아주 같음)까지 점수를 표시하는 실험이다. 실험 결과에서 1.2 kbps ZFE-CELP 코더는 85%의 화자인식률을 보였는데, 2.4 kbps 코더에 비해 절반의 비트율에도 불구하고 거의 유사한 인식률을 나타내었다.

표 4는 표준 파형 추출 방법에 따른 위상 제한 발생 횟수를 나타내고 있다. 위상 제한은 전 프레임과 현재프레임의 위상 반전으로 인하여 강제적으로 Zinc 파라미터를 제한시키는 방법이다. 따라서 위상 제한 횟수의 증가는 위상 반전의 발생 빈도가 높다는 것을 나타내고, 이는 표준파형의 추출이 정확하지 않다는 것을 의미한다.

표 3. 화자 인식 실험 결과.
Table 3. Results of speech recognition TEST.

음성 부호화기	Processed-Processed			Processed-Processed (DIFFERENTs)		
System	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
1.2kbps ZFE-CELP coder	85%	87%	82%	3.94	4.25	3.62

표 4. 표준파형 추출 방법에 따른 위상 제한 발생 횟수
Table. 4. The number of phase restrict as Extracting method.

위상 제한	위상 제한 (횟수)
상관도만을 이용한 표준 파형 추출	71
상관도와 분석-합성 구조를 이용한 표준파형 추출	57

표 4에서 종래의 상관도만을 이용한 표준파형 추출에 비해 분석-합성 구조가 결합된 구조에서 위상제한의 횟수가 크게 줄었음을 나타내고 있다. 이것은 제안된 구조의 표준파형 추출 방법이 좋은 성능을 나타내고 있음을 알 수 있다.

VII. 결 론

본 논문은 1.2 kbps 초저전송률 음성 코더의 개발을 위해 Zinc function을 사용하여 파형 보간을 함으로써 유성음을 부호화하고, 무성음 부호화를 위해 CELP구조를 사용하는 방법을 제안하였다. 실험 결과 Hanzo의 1.9 kbps ZFE-PWI 합성 음성과 비교해 볼 때 비트율이 낮아졌음에도 불구하고 MOS 0.01정도의 차이로 유사한 성능을 얻을 수 있었다. 제안된 부호화 방법은 LPC 분석 구간을 40 ms 단위로 수행함으로써 비트율을 효과적으로 낮추었고, 그에 따라 발생한 신호의 왜곡은 오버 샘플링을 이용한 선형 위상 합성법으로 해결했다. 또한 음질에 중요한 영향을 미치는 유성음의 모델링 시, 최적의 표준 파형을 추출하기 위하여 분석-합성 구조를 이용하였고, 이로 인한 음질 향상을 확인 할 수 있었다. 제안된 코더는 1.2 kbps의 적은 비트로 유성음 구간을 합성할 수 있기 때문에 1.2 kbps의 초 저 전송률 코더의 개발을 위한 적절한 모델링 기법으로 활용할 수 있다. 하지만 2.4 kbps의 코더나 1.9 kbps 코더와 비교해서는 할당된 비트가 적고, Mixed 음성에 대한 합성방법이 추가되지 않아 고주파 성분이 잘 표현되지 않기 때문에 이를 위한 별도의 방법이 필요함을 알 수 있었다. 이러한 점을 개선하고 파라미터에 대한 보간 방법을 잘 활용하면 1.9 kbps 코더보다 더 나은 성능을 얻을 수 있을 것이라 생각된다.

감사의 글

본 연구는 2006년도 충북대학교 학술연구 지원사업의 연구비 지원으로 수행되었습니다.

참 고 문 헌

1. R. C. de Lamare, A. Alcaim, "Strategies to improve the performance of very low bit rate speech coders and application to a variable rate 1.2 kb/s codec", Proc. IEE, 152, 74-86, Feb. 2005.
2. M. Schroeder, B. Atal, "Code excited linear prediction: Hight quality speech at low bit rates", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 937-940, 1985.
3. A. McCree, K. Truong, E. George, T. Barnwell, and V. Viswanathan, "A 2.4kbit/s coder candidate for the new U.S. federal standard." Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 200-203, 1996.
4. D. W. Griffin and J. S. Lim, "Multiband excitation vocoder", IEEE Trans. Acoustics, Speech and Signal Processing, 36 1223-1235, 1988.
5. R. J. McAulay and T. F. Quatieri, "The application of subband coding to improve quality and robustness of the sinusoidal transform coder", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2 439-442. 1993.
6. W. B. Kleijn and J. Haagen, "A speech coder based on decomposition of characteristic waveforms", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 508-511, 1995.
7. F. C. A. Brooks, Lajos Hanzo, "A 1.9kbps Zinc Function Excited, Waveform Interpolated Speech Codec", IEEE Vehicular Technology Conference, 2 18-21, 1998.
8. 김종학, 이인성 "Low Rate Speech Coding Using the Harmonic Coding Combined with CELP Coding", 한국음향학회지, 20 (7), 37-46, 2000.
9. D. J. Hiotakakos and C. S. Xydeas, "Low bit rate coding using an interpolated Zinc excitation model", in proceedings of the ICCS 94 865-869, 1994.
10. P. Hedelin, "Phase compensation in all-pole speech analysis", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 339-342, 1998.
11. E. Shlomot, Vladimir Cuperman and A. Gersho, "Combined Harmonic and Waveform Coding of speech at low Bit Rate", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 585-588, 1998.

12. Masayuki, Nishiguchi and J. Matsumotor, "Harmonic and Noise Coding of LPC Residuals with Classified Vector Quantization," Proc. International Conference on Acoustics, Speech and Signal Processing, 484-487, 1995.
13. 언어 자원 은행, http://bola.or.kr
14. Schmidt-Nielsen, A. Brock, D.P. "Speaker recognizability testing for voice coders", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, 2 7-10 1996.

저자 약력

•서 상 원 (Sang Won Seo)

2005년 2월: 충북대학교 전자공학과 (공학사)
2005년 3월~현재: 충북대학교 전파공학과 (석사과정)
※주관심분야: 음성/오디오 부호화, 디지털 신호처리

•김 종 학 (Jong Hak Kim)

1998년 2월: 충북대학교 전자공학과 (공학사)
2000년 2월: 충북대학교 전파공학과 (공학석사)
2000년 3월~현재: 충북대학교 전파공학과 (박사과정)
※주관심분야: 음성/오디오 부호화, 영상압축, 적응필터

•이 창 환 (Chang Hwan Lee)

2003년 2월: 충북대학교 전자공학과 (공학사)
2005년 2월: 충북대학교 전파공학과 (공학석사)
2005년 3월~현재: LG 전자 연구소 근무
(이동통신 기술 연구원)
※주관심분야: 음성/오디오 부호화, 채널 코딩

•정 규 혁 (Gyu-Hyeok Jeong)

2004년 2월: 충북대학교 전기전자공학과 (공학사)
2006년 2월: 충북대학교 전파공학과 (공학석사)
2006년 3월~현재: 충북대학교 전파공학과 (박사과정)
※주관심분야: 음성/오디오 부호화, 디지털신호처리

•이 인 성 (In Sung Lee)

1983년 2월: 연세대학교 전자공학과 (공학사)
1985년 2월: 연세대학교 전자공학과 (공학석사)
1992년 2월: Texas A&M University 전기공학과 (공학박사)
11993년 2월~1995년 9월: 한국전자 통신연구원 이동통신 기술연구단 선임연구원
1995년 10월~현재: 충북대학교 전기전자공학부 정교수
※주관심분야: 음성/영상 신호 압축, 이동통신, 적응필터