

구조화된 다중 펄스열을 이용한 낮은 전송률 음성 부호화기

正會員 姜泓求*, 徐廷泰*, 車日煥*, 尹大熙*

A Low Bit-Rate Speech Coder using the Structured Multiple Pulses

Hong Goo Kang*, Jeong Tae Seo*, Il Whan Cha*, Dae Hee Youn* Regular Members

要 約

낮은 전송률 음성 부호화 방법의 대표적 예인 CELP(Code Excited Linear Prediction) 부호화기는 8kbps 내외의 전송률에서는 우수한 성능을 보이지만, 4.8kbps 이하의 전송률에서는 급격히 성능이 저하된다. 이는 전송률을 낮추기 위해 분석 프레임의 길이를 길게함에 따라 잔차 신호에 피치 주기를 갖는 펄스 성분이 존재하게 되어 백색 가우시안 코드북만으로는 이를 정확히 모델링하기 어렵기 때문에 나타나는 현상이다.

본 논문에서는 이러한 문제를 극복하기 위해 몇 가지 제한 조건에 의해 구성된 구조화된 다중 펄스로 장구간 예측 후의 여기 신호를 구성하고, 고속 알고리즘을 이용한 직접 탐색을 통해 연산량을 줄일 수 있는 방법을 제안한다. 부프레임에 사용할 펄스 개수는 실험을 통해 두 개로 제한하였고, 각 펄스의 양자화 레벨은 실시간 처리에 적합하도록 $\pm 2^*$ 으로 구성하였다. 4.8kbps 전송률을 갖는 DoD-CELP와 VSELP(Vector Excited Linear Prediction)와의 성능 평가 결과 4.2kbps에서도 연산량 증가 없이 기존 방법과 유사한 성능을 나타내었다.

ABSTRACT

CELP(Code Excited Linear Prediction) which is the promised method of low bit-rate speech coding shows good performance at the rate of 8kbit/s. However, the quality degrades considerably at bit rates below 4.8kbit/s, mainly because of the inappropriate degree of periodicity in the voiced region

To overcome this problem, a new residual signal which represents the signal only a few structured pulses is proposed. It can also be applied fast full search algorithm. The number of pulses are fixed with two, and the levels of each pulses are quantized with $\pm 2^*$, which is proper to the real-time system.

As a result of simulation, the proposed method has as good or better quality than 4.8 kbps VSELP and DoD-CELP at the rate of 4.2kbit/s without increasing the computational amount.

*연세대학교 전자공학과
Dept. of Electronics Eng., Yonsei Univ.
論文番號 : 95088-0227
接受日字 : 1995年 2月 27日

I. 서론

무선 및 네트워크 통신 시스템의 적용 분야가 다양해짐에 따라 음성 부호화 알고리즘의 중요성은 커지고 있다. 4kbps 내외의 전송률을 목표로 하는 half-rate 시스템에서도 기존 방법과 비슷한 성능을 유지하기 위한 연구를 통해 음성 부호화 연구는 급격한 발전을 이루고 있으며, 이를 바탕으로 각국에서는 half-rate 표준안을 결정하기 위한 연구가 활발히 진행 중이다[1].

전송률을 낮추기 위한 가장 간단한 방법은 분석하고자 하는 음성 구간의 길이를 길게 하는 것이다. 같은 비트를 사용하더라도 구간의 길이가 길어지면 그만큼 전송률도 낮아지기 때문이다. 그러나, 이런 환경에서는 장구간 예측 성능이 저하되어 유성음 구간의 잔차 신호(residual)에는 여러 개의 펄스형 성분이 나타나게 되고, 이를 기존의 잡음 성분으로 구성된 코드북만으로 구성할 경우에는 만족할 만한 성능을 유지하기 어려워진다. 이를 해결하기 위한 연구가 다양한 각도로 진행되고 있다 [2][3][4].

미국 방성에서 표준안으로 제안한 DoD-CELP(Department of Defence CELP)[2]에서는 장구간 예측시에 보간(interpolation)을 통해 지연시간을 실수 간격으로 구함으로써 이를 해결하려고 하였으며, MB-CELP(Multi-Band CELP)[3]에서는 장구간 예측 후의 신호를 다시 한번 여러 대역으로 나누어 피치의 영향을 감소시키려 시도하였다. PSI-CELP(Pitch Synchronous Interpolation CELP)[4]에서는 백색 가우시안 잡음으로 구성된 통계 코드북 자체를 피치에 의존하도록 적응적으로 구성하여 문제를 해결하려 하였다. 그러나, 이러한 방법들은 코드

북을 구조화시키기 어려우므로 계산량이 많고, 복잡한 방법에 비해 성능 향상 정도도 미흡한 실정이다.

본 논문에서는 코드북을 몇 개의 펄스 조합으로 구성하고, 미리 정해놓은 제한 조건에 따라 최적의 펄스 위치와 간격을 구함으로써 이러한 두가지 문제를 해결하려 하였다. 실험 결과 2 개의 펄스로도 만족할 만한 결과를 얻을 수 있었으며, 제안된 고속 탐색 알고리즘을 사용하여 계산량도 급격히 감소시킬 수 있었다.

본 논문의 구성은 다음과 같다. 2 장에서는 분석-합성(analysis-by-synthesis) 방법 부호화기의 기본 구조에 대해 살펴보고, 3 장에서는 본 논문에서 제안한 부호화기를 서술하였다. 4 장에서는 제안된 부호화기의 성능 평가를 위한 실험 및 결과에 대해 알아보고, 5장에서 결론을 맺는다.

II. 분석-합성 부호화기

디지털 셀룰라와 같이 낮은 전송률을 필요로 하는 응용 분야에는 선형 예측(linear prediction) 방법을 사용하여 합성 필터 계수를 얻은 후, 원 신호와 재생 신호 간의 차이가 최소가 되도록 하는 여기 신호를 선택하여 부호화하는 분석-합성 방법이 효과적이다[8]. 그림 1은 이에 대한 블럭도이다.

과거의 여기 신호로 구성된 적응 코드북(adaptive codebook)은 음성 신호의 피치 주기마다 존재하는 상관성을 모델링하기 위한 것이고, 통계 코드북(stochastic codebook)은 피치 성분을 제거한 후의 잔차 신호를 모델링 하기 위한 것이다. 이때 잔차 신호를 모델링하기 위해 일정한 간격으로 구성된 펄스열로 코드를 구성하는 방법을 RPE (Regular Pulse Excitation)[5], 비정규

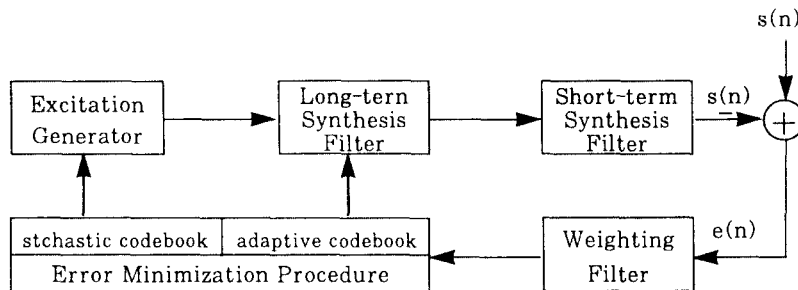


그림 1. 분석-합성 부호화기의 블럭도

Fig. 1. block diagram of a analysis-by-synthesis coder

적인 펄스의 위치와 크기 정보로 표현하는 방법을 MPE(Multi Pulse Excitation)(6). 그리고 백색 가우시안 잡음으로 구성된 벡터 코드북을 이용하는 방법을 CELP(Code Excited Linear Prediction)(7)라고 한다.

각 방법에서 최적의 여기 신호를 얻기 위해서는 코드북에서 선택된 코드를 장구간 및 단구간 예측기의 합성 필터에 통과시켜 재생된 음성 신호를 얻고, 이 신호와 원 신호간의 오차를 청각 특성을 반영한 가중 필터(weighting filter)에 통과시켜 평균 자승 오차를 최소로 하는 값을 찾아야 한다(8).

그 중 CELP는 8kbps 이하의 낮은 전송률에서도 우수한 음질을 가지며, 구조화된 코드북을 통해 연산량 및 복잡도를 줄일 수 있어 실시간 시스템으로 응용되고 있으며(12), 최근에는 전송률을 4kbps 이하로 낮추려는 연구가 이루어지고 있다(10)(11). 지금까지 발표된 결과에 의하면 4 kbps 이하의 낮은 전송률에서는 여기신호가 성능에 중요한 영향을 미친다고 알려져 있다(1)(15). 프레임의 길이가 길어질 경우에는 장구간 예측 성능이 저하되므로 유성음 구간의 잔차 신호에는 펄스형 성분들이 다수 존재하게 되고, 이를 백색 가우시안 잡음으로 모델링 할 경우에는 재생된 신호에 왜곡을 초래하게 된다(1)(15). 왜곡 발생의 주요 원인은 피치 주위의 신호 왜곡에 기인한다. 피치를 정확히 모델링하지 못하면 재생된 음에 명료도가 떨어지므로 왜곡 정도가 심하다고 판단하며, 원 신호와 재생 신호사이의 피치 간격에 차이가 생겼을 경우에도 비슷한 결과를 낳기 때문

이다. 이러한 문제는 전체적으로 음질이 저하되는 낮은 전송률에서는 주관적 성능 저하를 일으키는 중요한 요인으로 작용하는 요인이다. 코드북의 개수를 많이 하면 해결 가능하겠지만, 연산량 및 기억 장치 면에서 비효율적 이므로 실시간 시스템 구현이 어렵다는 단점이 있다.

이러한 문제를 극복하고, 합성음의 명료성을 높이기 위한 한 방법은 MPE(6)와 같이 몇 개의 펄스 조합을 바탕으로 직접 탐색 방법을 이용하는 것이다. 이 방법은 코드북을 사용하지 않으므로 코드북을 저장할 메모리가 필요없으며, 혼린 과정에서 포함되지 못한 데이터에서 발생할 수 있는 왜곡 문제를 고려하지 않아도 되기 때문이다. 그러나, 여기 신호를 MPE 형태로 구성하려면 최적 변수 정보를 찾기 위해 복잡한 연산이 필요하며(9), 구해진 각 펄스 위치와 크기를 모두 전송해야 하므로 half-rate 전송률을 유지하기는 어렵다. 다중 펄스의 장점을 이용하면서 전송률을 낮추기 위해 본 논문에서는 펄스 개수와 이득 정보를 미리 고정시킨 후, 오차가 최소가 되는 펄스의 위치만을 찾아 전송하는 방법을 제안한다. 펄스 수는 부프레임의 길이에 따라 조정하며, 이득 정보는 실시간 처리에 편한 수로 양자화된 값을 사용하므로 전송률 및 연산량은 펄스 수와 펄스 위치 탐색 간격에 따라 결정된다. 이 방법을 구조화된 다중 펄스(structured multi-pulse) 방법이라 명칭하며 자세한 내용은 다음 장에 설명한다.

III. 구조화된 다중 펄스 부호화기

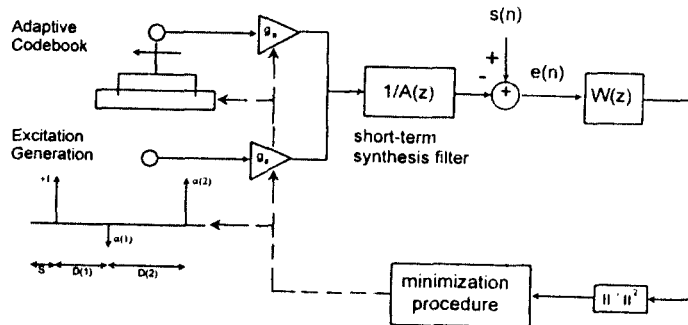


그림 2. 제안한 부호화기의 블록도
Fig. 2. block diagram of the proposed coder

제안한 음성 부호화기의 블록도는 그림 2와 같다. 기본 구조 자체는 기존의 분석 합성 방법과 동일하며, 장 구간 예측 후의 여기 신호를 구성하는 방법에만 차이가 있음을 알 수 있다.

3.1 다중 펄스 발생 방법

전송률과 여기 신호에 사용하는 펄스 개수와와 상관 관계에 따라 프레임의 길이는 20 msec에서 30 msec 내외로 결정한다. 구조화된 다중 펄스열을 생성하기 위해 펄스 개수 및 시작 펄스의 위치, 펄스 탐색 영역 등에 주어진 제한 조건은 다음과 같다.

제한조건-1 : 펄스 갯수는 M개로 고정한다.

제한조건-2 : 첫번째 펄스의 크기는 "1"이며, 부프레임의 어느 위치에나 존재할 수 있다.

제한조건-3 : 두번째 이후의 펄스는 $\pm 2^n$ 값으로 양자화하며, 이전 펄스로부터의 일정 간격에 대해 완전 탐색(full search)을 통해 최적 위치와 양자화 값을 결정한다.

제한 조건-3 은 필터링 과정 및 최적 코드 탐색을 위한 오차 연산을 쉬프트(shift)와 덧셈 및 뺄셈으로만 표현할 수 있도록 하기 위해 정해진 값으로서 실시간 시스템에 쉽게 적용할 수 있다.

3.2 최적 펄스열을 찾기 위한 알고리즘

펄스 개수 및 형태는 전송률 및 계산량에 영향을 미치며 성능과도 밀접한 관계가 있는 변수이다. 한 예로서 세 개의 펄스로 구성된 코드 변수는 그림 3과 같이 이루어진다.

그림 3에서 시작 펄스의 위치를 S, 부프레임 내의 펄스 수를 K, 각 펄스의 크기를 $\alpha(k)$, 그리고 펄스 사이

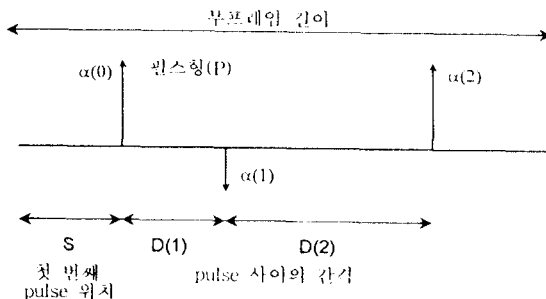


그림 3. 세 개의 펄스로 구성된 구조화된 다중 펄스
Fig. 3. structured multi-pulse with three pulses

의 간격을 $D(k)$ 라고 한다면, 구성된 코드는 다음과 같이 표현할 수 있다.

$$r(n) = \sum_{k=1}^K \alpha(k) \delta[n - S - D(k)] \tag{1}$$

여기서, $\delta(\cdot)$ 는 dirac-delta 함수를 나타낸다. 또한, 제한 조건에 의해 첫번째 펄스의 크기 $\alpha(0)$ 는 1.0으로 주어지며, 첫번째 펄스에서는 이전 펄스로부터의 간격이 존재하지 않으므로 $D(0)$ 는 0으로 주어진다.

합성 필터의 임펄스 응답을 $h(n)$ 이라고 하면, 필터링 후의 신호 $q(n)$ 은 다음과 같이 표현된다.

$$q(n) = \begin{cases} 0 & , 0 \leq n < S \\ \sum_{l=0}^{m-1} \alpha(l) h(n - S - \sum_{t=0}^l D(t)) & , S + \sum_{k=1}^m D(k) \leq n < S + \sum_{k=1}^{m+1} D(k) \\ \sum_{l=0}^{K-1} \alpha(l) h(n - S - \sum_{t=0}^l D(t)) & , S + \sum_{k=1}^{K-1} D(k) \leq n \leq N-1 \end{cases} \tag{2}$$

가중 필터를 통과한 신호를 $p(n)$ 이라고 하고, 이 신호와 $q(n)$ 에 이득 β 를 곱하여 구성된 합성 신호 사이의 오차 에너지 E는

$$E = \sum_{n=0}^{N-1} [p(n) - \beta q(n)]^2 \tag{3}$$

으로 주어진다. 오차를 최소로 하기 위한 최적의 β 값은

$$\beta_{opt} = \frac{\sum_{n=0}^{N-1} [p(n)q(n)]}{\sum_{n=0}^{N-1} [q^2(n)]} \tag{4}$$

으로 주어지며[6], β_{opt} 을 대입하면 최소 오차에너지는

$$E_{min} = \sum_{n=0}^{N-1} [p^2(n)] - \frac{[\sum_{n=0}^{N-1} p(n)q(n)]^2}{\sum_{n=0}^{N-1} [q^2(n)]} \tag{5}$$

이다.

E_{min} 의 앞 항은 상수값이므로 오차를 최소로 하기 위해서는 오른쪽 항

$$\frac{[\sum_{n=0}^{N-1} p(n)q(n)]^2}{\sum_{n=0}^{N-1} [q^2(n)]} \tag{6}$$

을 최대로 하는 $q(n)$ 을 찾고, 이때의 펄스 구성 변수를

찾으면 최적 여기 신호를 얻을 수 있다.

전송률을 낮추고, 고속 알고리즘을 적용하기 위해서는 부프레임에 사용할 최대 펄스 개수와 형태를 제한해야 한다. 일반 사람의 피치 간격을 고려할 때 부프레임의 길이가 64이면 피치 성분은 세 개를 넘지 않으며, 두 개일 경우에도 그림 4에 주어진 합성음 파형에서 볼 수 있는 바와 같이 세 개일 경우와 별 차이가 없음을 알 수 있다. 물론 간단한 파형 비교를 통해 펄스 개수를 선택하는 것은 다소 무리가 있지만, 펄스 수가 세 개 이상으로 증가할 경우, 본 논문에서와 같이 완전 탐색을 수행하려면 경우의 수가 너무 많아지므로 급격히 연산량이 증가되고, 펄스 정보를 표현하기 위해 할당되는 비트도 증가한다는 문제도 발생한다[15]. 따라서 본 논문에서는 펄스 개수를 2개로 제한하였으며, 이 경우에는 다음

절에 설명한 고속 탐색 알고리즘을 이용하여 연산량을 줄일 수 있다.

3.3 두개의 펄스로 구성된 코드

제한 조건에 의해 두 개의 펄스로만 코드를 구성할 경우에는 두 번째 펄스의 탐색 영역과 크기 정보만이 필요하므로 잔차 신호는 (7)식과 같이 더욱 간단히 표현할 수 있다. 펄스 사이의 간격을 D , 두 번째 펄스의 크기를 a 라고 하면, 구성된 코드는 다음과 같이 표현할 수 있다.

$$r(n) = \delta(n-S) + a\delta(n-S-D) \tag{7}$$

(2)식으로부터 필터링된 후의 신호 $q(n)$ 은 다음과 같이 표현할 수 있다.

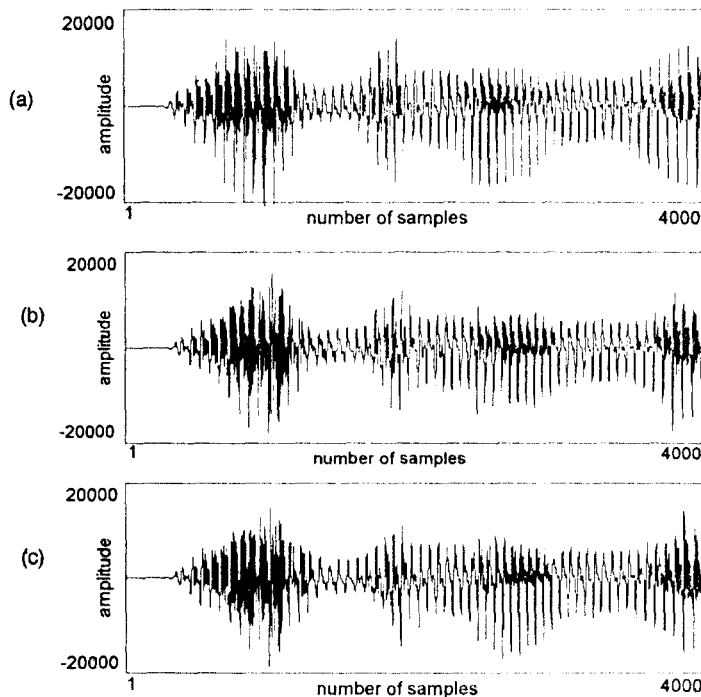


그림 4. 펄스 개수 변화에 따른 합성 파형
(a) 원음 (b) 펄스 개수 = 2 (c) 펄스 개수 = 3

Fig. 4. synthesized speech with changing the pulse number
(a) original speech (b) pulse number = 2 (c) pulse number = 3

$$q(n) = \begin{cases} 0, & 0 \leq n < S \\ h(n-S), & S \leq n < S+D \\ ah(n-S-D) + h(n-S), & S+D \leq n \leq N-1 \end{cases} \quad (8)$$

펄스가 두 개일 경우 필터링 과정은 간단하게 표현될 수 있다. 그러나, (6)식에 주어진 바와 같이 오차에너지를 최소화 하는 최적의 펄스 조합을 찾기 위해서는 부프레임 전체에 대한 제곱 계산과 상호 상관 함수 계산이 필요하므로 연산량은 아직도 상당함을 알 수 있다. 이러한 문제는 고속 반복 알고리즘을 사용하여 해결할 수 있으며, 그 방법은 다음과 같다.

3.4 펄스열을 찾기 위한 고속 알고리즘

펄스의 시작 위치를 S라 하고, 두 펄스 사이 간격을 D라고 하면, (6)식의 분모항은 다음과 같이 주어진다.

$$\begin{aligned} P_{S,D}(n) &= \sum_{n=0}^{N-1} q^2(n) = 0 + \sum_{n=S}^{S+D-1} h^2(n-S) + \\ &\sum_{n=S+D}^{N-1} (ah(n-S-D) + h(n-S))^2 \\ &= \sum_{n=S}^{S+D-1} h^2(n-S) + \sum_{n=S+D}^{N-1} [a^2 h^2(n-S-D) + 2ah(n-S-D)h(n-S) + h^2(n-S)] \\ &= \sum_{n=S}^{N-1} h^2(n-S) + a^2 \sum_{n=S+D}^{N-1} h^2(n-S-D) + \\ &2a \sum_{n=S+D}^{N-1} h(n-S-D)h(n-S) \end{aligned} \quad (9)$$

펄스 시작 위치가 한 샘플 증가하여 S+1이 되었다면, 위의 식은 다음과 같이 표현될 수 있다.

$$\begin{aligned} P_{S+1,D}(n) &= \sum_{n=S+1}^{N-1} h^2(n-(S+1)) \\ &+ a^2 \sum_{n=S+D+1}^{N-1} h^2(n-S-1-D) \\ &+ 2a \sum_{n=S+D+1}^{N-1} h(n-S-1-D)h(n-S-1) \\ &= [\sum_{n=S}^{N-1} h^2(n-S) - h^2(N-1-S)] \\ &+ a^2 [\sum_{n=S+D}^{N-1} h^2(n-S-D) - h^2(N-1-S-D)] + 2a [\sum_{n=S+D}^{N-1} h(n-S-D) \\ &h(n-S) - h(N-1-S-D)h(N-1-S)] \\ &= P_{S,D}(n) - [h^2(n-1-S) + a^2 h^2(N-1-S-D) + 2ah(N-1-S-D)h(N-1-S)] \end{aligned}$$

$$= P_{S,D}(n) - [ah(N-1-S-D) + h(N-1-S)]^2 \quad (10)$$

따라서 분모항의 계산은 S가 변함에 따라 덧셈 2번, 곱셈 2번 만으로 재추정이 가능하다.

(6)식의 분자 항에 주어진 상호 상관 값도 다음과 같은 과정을 통해 연산량을 줄일 수 있다.

$$\begin{aligned} \sum_{n=0}^{N-1} p(n)q(n) &= \sum_{n=S}^{N-1} p(n)q(n) \\ &= \sum_{n=S}^{S+D-1} p(n)h(n-S) + \sum_{n=S+D}^{N-1} p(n)[\\ &ah(n-S-D) + h(n-S)] \\ &= \sum_{n=S}^{N-1} p(n)h(n-S) + a \sum_{n=S+D}^{N-1} p(n)h(n-S-D) \\ &= \sum_{n=0}^{N-1-S} p(n+S)h(n) + a \sum_{n=0}^{N-1-S-D} p(n+S+D)h(n) \end{aligned} \quad (11)$$

위의 식에서 첫 번째 항과 두 번째 항에 주어진 p(n)과 h(n)의 상호 상관값을 (12)식과 같이 미리 구해 놓으면, D가 변하더라도 추가 연산없이 메모리 이동 연산을 이용하여 반복 사용할 수 있다.

$$R(S) = \sum_{n=0}^{N-1-S} p(n+S)h(n) \quad (12)$$

S=0, 1, 2, ..., N-1

(12)식에서 R(S)는 S가 변함에 따라 이전 값을 이용하는 반복 연산을 통해 구할 수 있으며, 부프레임에 대해 한번만 구해 놓으면 된다.

$$\begin{aligned} \sum_{n=0}^{N-1} p(n)q(n) &= R(S) + aR(S+D) \quad (13) \\ S &= 0, 1, 2, \dots, N-1 \end{aligned}$$

따라서, (6)식의 분자항에 주어진 식도 D가 변함에 따라 단지 메모리로부터의 이동 연산 한번과 곱셈 한번 만으로 구성할 수 있으므로 연산량은 매우 감소됨을 알 수 있다. 또한, 앞에서 주어진 임펄스 응답 h(n)을 다음과 같이 축소해서 사용한다면 연산량을 더욱 줄일 수 있다[12].

$$h(n) = \begin{cases} h(n), & 0 \leq n \leq N_T \\ 0, & N_T \leq n \end{cases} \quad (14)$$

보통 N_T 는 부프레임 길이의 절반을 사용한다.

제안 방법의 고속 탐색 알고리즘에 소요되는 연산량은 다음 장에 주어진 연산량 비교절에서 자세히 언급하기로 한다.

IV. 실험 및 결과 고찰

4.1 펄스 탐색 영역 및 양자화

부프레임에 사용되는 펄스 수와 마찬가지로 두번째 펄스의 위치를 찾기 위한 탐색 영역 및 양자화 값도 성능 및 전송률에 영향을 미치는 변수이다. 두번째 펄스의 탐색 영역은 첫번째 펄스부터의 간격으로 정한 후, 실험을 통해 구간 신호대 잡음비를 구하고, 이로부터 최적값을 정한다. 또한, 두번째 펄스의 크기는 다음과 같이 양자화된 표를 이용하여 고정한 후, 같은 실험을 통해 최적 비트 할당 정보를 얻는다.

탐색 영역 변화에 따른 결과는 그림 5와 같으며, 두번째 펄스에 할당된 양자화 레벨에 따른 결과는 그림 6과 같다.

그림 5와 그림 6의 실험 결과에 의하면 두 번째 펄스의 탐색 영역이 증가할 때마다 구간 신호 대 잡음비

레벨 변화에 대해서는 큰 증가가 없으므로 같은 전송률 (SegSNR)은 계속 증가하지만, 두 번째 펄스의 양자화에서는 펄스 형태보다는 탐색 영역이 더 중요한 변수임을 알 수 있다. 또한, 남성 화자의 경우 탐색 영역이 24 를 넘어서면 성능 향상이 별로 없지만, 여성 화자일 경우에는 계속 향상되므로 전체적인 향상을 위해 최소한 5비트는 할당해야만 한다. 이러한 결과가 나온 이유는 실험에 사용된 여성 음성의 피치 간격이 30~40 정도의 값을 유지하기 때문에 한 프레임에 두 개 이상의 펄스 성분이 존재할 경우가 발생하기 때문으로 생각된다. 두 번째 펄스의 탐색 영역을 전 영역으로 확장하면 쉽게 해결 가능하겠지만, 연산량 및 전송률이 증가되는 문제가 발생하므로 본 연구에서는 펄스 탐색 영역을 5부터 36 까지로 제한하는 방법을 통해 성능에 미치는 영향을 최소화하였다.

또한, 그림 6으로부터 두 번째 펄스의 양자화 레벨이 4 이상으로 증가해도 성능 향상 정도가 적으므로 2 비트를 할당하여 전송률 및 연산량의 이득을 취한다.

그림 5와 6의 결과로부터 구조화된 다중 펄스 발생기에 할당된 비트는 표 2와 같이 정리할 수 있다.

표 1. 두번째 펄스 양자화 레벨

비트수	양자화 레벨
1	+1, -1
2	+1, -1, +0.5
3	+1, -1, +0.5, -0.5 +0.25, -0.25, +0.125, -0.125

표 2. 구조화된 다중 펄스 발생기에 할당된 비트

	할당 비트수	양자화표
펄스 시작 위치(S)	6	1-64
두 번째 펄스 탐색 영역(D)	5	5-36
두 번째 펄스 양자화 레벨(P)	2	1.0, -1.0, 0.5, -0.5

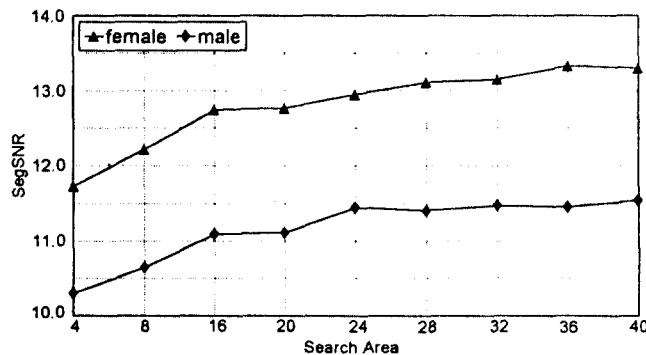


그림 5. 탐색 영역 변화에 따른 구간 신호대 잡음비
Fig. 5. segmental SNR with changing the search area

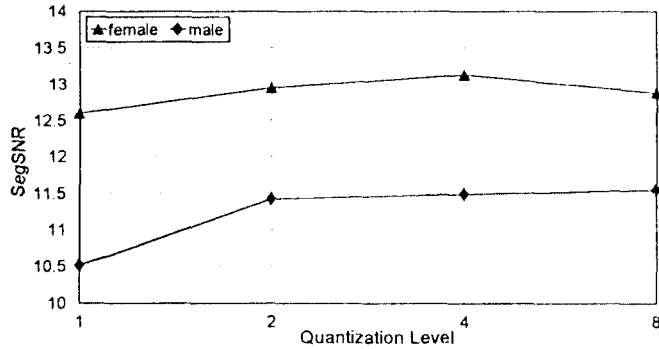


그림 6. 두번째 펄스 양자화 레벨 변화에 따른 구간 신호대 잡음비
Fig. 6. segmental SNR with changing the quantization level of the second pulse

표 3. 구조화된 다중 펄스 방법에 필요한 연산량

기본 연산량	(9)식	4
	(10)식	4
	(13)식	2
부프레임(8 msec)에 필요한 총 연산량		기본 연산량 * S(1-64) * D(5-36) * P(1-4) * 기타 = 10 * 64 * 32 * 3 + 6000 = 67440
MIPS		0.06744 * 1000 / 8 = 8.4

표 4. 각 부호화 방법의 여기 신호 생성에 필요한 연산량

부호화 방법	여기 신호 생성에 필요한 연산량(MIPS)
DoD-CELP	8.8
VSELP 4.8kbps	9.7
구조화된 다중 펄스	8.4

4.2 연산량 비교

부호화기의 연산량을 측정하기 위해서는 부호화 및 복호화 전반 과정에 대한 단계별 고찰이 필요하지만, 구현할 시스템에 내장된 연산 장치 분해능, 레지스터(register) 개수, 누산기(accumulator) 특성에 따라 각기 다른 방법으로 최적화 과정이 진행되므로 모든 경우를 포함하는 결과를 얻을 수는 없다. 따라서, 본 논문에서는 범용 디지털 신호 처리기의 구조를 근간으로 장구간 예측 후의 여기 신호를 생성하는 방법에 따른 연산량 차이만을 언급한다. 최근에 개발된 디지털 신호처리 프로세서의 연산 장치는 디지털 신호처리의 기본이 되는 MAC(Multiply and Accumulate)를 기본으로 설계되어 으며, 모든 연산은 한 연산 싸이클(instruction cycle)에 수행함을 원칙으로 한다[12]. 본 논문에서도 이와 같은 특징을 이용하여 초당 연산 회수를 계산하고, 이를 MIPS(Millions Instruction Per Second) 단위로 표현하여 복잡도를 평가한다.

제안한 구조화된 다중 펄스 방법의 고속 탐색 알고리즘을 기반으로 하여 구한 연산량은 표 3과 같다. 기본 연산량에 주어진 값은 (9)식에 급수로 표현된 부분이 모두 미리 연산되어 메모리에 저장되어 있다고 가정하여 얻은 결과이다. 급수 연산에 필요한 값은 부프레임에 필요한 총 연산량에 주어진 '기타' 부분에 포함되어 있다. S는 첫번째 펄스 위치, D는 두번째 펄스의 탐색 영역을 나타내며, P는 두번째 펄스에 할당된 양자화 레벨을 의미한다. 표 3에 표시한 값들은 임펄스 응답 $h(n)$ 의 길이를 줄이지 않고 계산한 값이므로 (14)식을 이용한다면 연산량은 더욱 줄어든다.

표 3에 표시된 값들은 C 언어로 구현된 고속 알고리즘을 어셈블리어로 변환하는 과정에서 필요한 연산량만으로 구성된 값이므로 실시간 시스템 구현시에 절대적인 수치는 아니지만, 이 방법을 비교 기준 부호화기에도 동일하게 적용한다면 부호화기 간의 상대적 평가는 가능할 것이다. 표 4는 이를 이용하여 구한 결과이

다.

표 4로부터 구조화된 다중 펄스 방법에 필요한 연산량은 DoD-CELP, VSELP 보다는 약 0.5~1.5 MIPS 정도 적음을 알 수 있다. 또한, DoD-CELP의 장구간 예측에서 실수 간격 탐색에 필요한 연산량과 VSELP의 이득 양자화 시 필요한 벡터 양자화에 소요되는 연산량을 고려한다면 전체 시스템의 복잡도는 더욱 낮을 것으로 판단된다.

4.3 성능 평가

제한한 부호화기의 성능 평가를 위해 모의 실험을 수행하였다. 실험에 사용된 데이터는 조용한 환경에서 남녀 화자 각 2명이 발음한 2 개의 문장을 8KHz 샘플링하여 사용하였으며, 전체 길이는 약 30 초 정도이다. 앞 실험 결과에 의해 표 5와 같이 각 변수에 비트를 할당할 결과 전송률은 4.2kbps로 나타났다.

선형 예측 계수는 LSP(Line Spectrum Pair)로 변환한 후, 벡터-스칼라 분리 양자화 방법(15)을 사용하였다. 장구간 예측 래그(lag) 탐색 범위는 20부터 146으로 제한하였으며, 구조화된 다중 펄스에는 4.1절에서 실험을 통해 구한 결과를 이용하여 할당하였다.

성능 비교를 위해 4.8kbps 전송률을 갖는 DoD-CELP(2)와 VSELP(13)를 사용하였으며, 평가 방법은 객관적 평가 방법으로 구간 신호대 잡음비

(SegSNR)와 신호대 잡음비(SNR)를 그리고, 주관적 평가 방법으로는 A-B test[14]와 MOS(Mean Opinion Score) [14]를 사용하였다. 주관적 평가는 연구원 10명을 대상으로 실시하였으며, 각 평가 방법에 따른 결과는 다음과 같다.

주관적 평가 결과로 부터 제안 방법은 기존 방법에 비해 비슷하거나 우수한 성능을 나타내었다. 이는 펄스 위치를 정확히 모델링함에 따라 피치 성분을 비교적 정확히 모델링할 수 있었고, 이에 따라 위상 왜곡이 줄어들었기 때문으로 판단된다. A-B Test 결과를 통해서 알 수 있는 바와 같이 제안된 방법은 명료성이 뛰어나므로 그만큼 선택 비율이 높았다. 그러나, 무성음 특성이 강하게 나타나는 신호나 여성 데이터의 경우에는 다소 왜곡 정도가 심하게 나타났으며, 이러한 영향으로 기존 방법 보다 낮거나 비슷한 평가를 얻는 경우도 발생하였다. 그 이유는 프레임의 특성에 따라 두 개의 펄스만으로는 여기 신호 구성이 어려운 부분도 상당히 존재하기 때문이라고 생각된다. 결국, 낮은 전송률 음성 부호화 방법에서는 명료성이 중요한 역할을 차지하지만, 무성음 부분에서 발생하는 거친 성분을 어느 정도 해결해야만 더 좋은 평가를 받을 수 있으며, 제안한 방법이 적합하지 않은 음성 구간도 존재한다는 것을 알 수 있다. 그러나, 제안한 방법이 갖는 계산량 및 메모리 효율은 실시간 처리에 용이하므로 이러한 장점을 이용하고, 단점을 극복

표 5. 제안 방법의 비트 할당

parameter	bit/8ms subframe	bit/32ms frame
LPC coef. (LSP)		30
LTP lag	7	28
LTP gain	3	12
structured multi-pulse codebook(S+D+P)	13	52
codebook gain	3	12
Total		134(4.2 kbps)

표 7. 주관적 성능 평가(A-B Test)

VSELP	제안 방법	DoD-CELP	제안 방법
47.5%	52.5%	44.0%	56.0%

표 6. 객관적 성능 평가

	SegSNR	SNR
VSELP	13.60	10.35
DoD-CELP	9.40	8.30
제안 방법	12.65	9.95

표 8. 주관적 성능 평가(MOS Test)

VSELP	DoD-CELP	제안 방법
3.25	3.13	3.18

할 수 있는 방안이 연구되어야 할 것이다.

V. 결 론

본 논문에서는 제한된 크기와 위치를 갖는 다중 펄스를 사용하는 직접 탐색 방법을 사용하여 장구간 예측 후의 여기 신호를 적응적으로 모델링함으로써 4.2kbps에서 우수한 성능을 보이는 음성 부호화기를 제안하였다. 제안된 부호화기는 고속 알고리즘을 통해 계산량이 감축되고, 필요 메모리 또한 적으므로 실시간 시스템 구현이 쉽다는 장점이 있다. 또한, 4.8kbps DoD-CELP, VSELP와의 주, 객관적 성능 평가도 비슷하거나 우수한 성능을 유지하였다.

그러나, 유성음의 특성이 변화하는 부분이나 무성음 구간에서의 왜곡으로 인해 다소 성능 저하가 나타나는 부분도 있었다. 이를 해결하기 위한 가장 좋은 방법은 구간 특성에 따라 다른 부호화 방법을 사용하는 다중 모드(multi-mode) 방법일 것이다. 제안된 방법의 장점을 유지하면서 다중 모드로의 적용이 가능한 연구가 진행 중에 있다.

참고문헌

1. H.Su and P. Mermelstein, "Delayed Decision Coding of Pitch and Innovation Signals in CELP of Speech" *Speech and Audio Coding for Wireless and Network Applications*, Kluwer Academic Publisher, 1993.
2. J. P. Campbell, T. A. Tremain, V. C. Welch, "DoD 4.8Kbps Standard (Proposed Federal Standard 1016)," *Advances in Speech Coding*, Kluwer Academic Publisher, 1991.
3. G. Yang, H. Leich and R. Boite, "Multiband Code-Excited Linear Prediction(MBCELP) for Speech Coding," *Signal Processing*, Vol. 31, pp.215-227, 1993.
4. S. Miki et al, "A Pitch Synchronous Innovation CELP(PSI-CELP) Coder for 2-4 kbit/s," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.113-116., Adelaide, May, 1994.
5. P. Kroon, E. F. Deprettere, R. J. Sluyter, "Regular Pulse Excitation : A Novel Approach to Effective And Efficient Multipulse Coding of Speech," *IEEE Trans. Acoust. Speech, Signal Processing*, vol. ASSP-34, pp.1054-1063, Oct., 1986.
6. B. S. Atal, J. R. Remde, "A New Model of LPC Excitation for Producing Natural Sounding Speech at Low Bit Rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.614-617, Paris, May, 1982.
7. M. R. Schroeder, B. S. Atal, "Code-Excited Linear Prediction(CELP) High-Quality Speech at Very Low Bit Rates," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 25.1.1~25.1.4, Florida, Mar., 1985.
8. P. Kroon, E. F. Deprettere, "A Class of Analysis-by-Synthesis Predictive Coders for High Quality Speech Coding at Rates between 4.8 And 16 Kbit/s," *IEEE Journal on Selected Areas in Communications*, Vol. 6, no. 2, pp.353-363, Feb., 1988.
9. S. Singhal, B.S. Atal, "Amplitude Optimization and Pitch Prediction in Multipulse Coders," *IEEE Trans. Acoust. Speech, Signal Processing*, vol.ASSP-37, no. 3, pp.317-327, Mar., 1989.
10. K. Ozawa, M. serizawa, T. Miyano and T. Nomura, "M-LCELP Speech Coding at 4kbps," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.269-272, Adelaide, May, 1994.
11. J. Martin, B. Wachter, "A CODEC Candidate for the GSM Half-Rate Speech Channel" *Proc. Int. Conf. Acoust., Speech, Signal Processing*, 1994.
12. 윤대희, 낮은 비트율 음성부호화기(VSELP)에 대한 연구, 연세대학교 전자정보 통신 연구소, 1994.
13. I. A. Gerson, M. A. Jasuik, "Vector-Sum Excited Linear Prediction(VSELP) Speech Coding at 8Kbps," *Proc. Int. Conf. Acoust., Speech, Signal Processing*, pp.641-644,

Albuquerque, Apr., 1990.

14. J. R. Deller, Jr., J. G. Proakis, J. H. L. Hansen, *Discrete-Time Processing of Speech Signals*, Macmillan Publishing Company, 1993.

15. 강 홍구, "시간축 청각 특성을 이용한 고정 모드 음성 부호화기", 연세대학교 전자공학과 박사학위 논문, 1995. 6.



姜 泓 求(Hong Goo Kang) 정회원

1989년 2월 : 연세대 전자공학과 졸업
1991년 2월 : 동 대학원 전자공학과 졸업(공학석사)
1991년 3월~현재 : 동 대학원 박사과정

※주관심 분야 : 음성 부호화, 오디오 부호화, 실시간 신호처리 시스템



徐 廷 泰(Jeong Tae Seo) 정회원

1985년 2월 : 연세대 전자공학과 졸업
1987년 2월 : 동 대학원 전자공학과 졸업(공학석사)
1990년 3월~현재 : 동 대학원 박사과정

※주관심 분야 : 음성 부호화, 영상 압축, 적응 디지털 신호처리



車 日 煥(Il Whan Cha) 정회원

1959년 2월 : 연세대 전기공학과 졸업(공학석사)
1983년 8월 : 동 대학원 전자공학과 졸업(공학박사)

1969년~1970년 : 영국 University of Southampton Institute of Sound and Vibration 연구원

1986년 9월~1987년 7월 : University of Iowa 교환교수

1988년 8월~현재 : 연세대학교 전자공학과 교수

※주관심 분야 : 건축 음향, 소음진동, 초음파, 음성 부호화, 오디오 부호화

尹 大 熙(Dae Hee Youn)

정회원

한국통신학회 논문지 Vol.20, No. 2, Feb., 1995. 참조.