

# 발성 속도에 따른 가변전송률 CELP 부호화기 설계에 관한 연구

나 덕 수, 배 명 진  
승실대학교 정보통신공학과  
전화 : (02) 824-0906

## A Study on a Design of the Variable Bit-Rate Vocoder by Measuring of the Speaking Rate

Duck Su Na, Myung Jin Bae  
Dept. of Information and Telecommunication Engineering Soongsil University  
E-mail : mjbae@saint.soongsil.ac.kr

### Abstract

CELP 부호화기는 선형 예측 합성에 의한 분석 부호화의 원칙에 기본을 두고 있다. 그리고 음성 신호의 스펙트럼을 LPC 분석을 통해 부호화하는데 고정 윈도우를 사용하여 부호화 한다. 그러나 음성신호는 화자의 발성속도에 따라 파형의 변화가 시간적으로 빠르게 변화하기도 하고, 반대로 유사한 파형이 일정시간 유지되기도 한다. 따라서 윈도우의 크기를 발성속도에 맞추어 분석한다면 보다 효율적인 부호화를 할 수 있다.

본 논문에서는 발성속도에 따라 전송률을 달리 적용하는 방법을 제안한다. 발성속도의 측정은 스펙트럼 변화도를 이용하여 측정하였고, 발성속도가 빠를 때는 프레임 크기를 줄여 시간적으로 빠르게 변화하는 신호에 적응적으로 분석하고 대신 파라미터 표현에 비트를 줄인다. 반대로 발성속도가 느릴 때는 프레임 크기를 키우고 파라미터 표현에 비트를 더 할당한다. 제안한 방법을 실험하기 위해 G.723.1 5.3kbps ACELP 부호화기를 이용하였다. 음질의 열하 없이 평균 16.34% 전송률 감소효과를 얻을 수 있었다.

### 1. 서론

표준 음성 부호화기는 1972년 ITU-T 권고안 G.711

로 채택된 64kbps PCM(Pulse Code Modulation) 방식으로부터 출발하여 32kbps의 ADPCM(adaptive DPCM), 16kbps의 LD-CELP(Low-Delay CELP) 방식으로 표준화되었다. 현재 ITU-T에서는 PCS, IMT-2000등에서 사용할 수 있는 8kbps 음성 부호화기에 대한 표준화 작업으로 1996년에 CS-CELP (conjugated structure algebraic CELP)를 G.729로, 그리고 인터넷폰 및 화상통신용 음성 부호화기로 ACELP/MP-MLQ(algebraic CELP/Multipulse Maximum Likelihood Quantization)의 5.3/6.3kbps dual rate를 G.723.1 권고안으로 선정하였다[1]. 그리고 현재 국내에서는 유선망을 통한 화상회의를 목적으로 표준화된 G.723.1 음성부호화기 사용하여 인터넷폰이나 화상회의에 응용하기 위해 많은 연구가 이루어지고 있다. 전화는 64kbps이지만 인터넷은 6-16kbps로 상황이 나쁘면 음성이 단절되거나 잡음이 섞이므로 안정적인 통화를 보장할 수 없어 전송률을 낮추는 기술과 음질을 향상시키는 기술이 매우 중요하다.

본 논문에서는 발성속도에 따라 전송률을 달리 적용하여 보다 효율적인 부호화 방법을 제안한다. 즉, 발성속도가 빠를 때는 프레임 크기를 줄여 시간적으로 빠르게 변화하는 성실에 초점을 맞추고, 반대로 발성속도가 느릴 때는 프레임 크기를 키우고 각 프레임의 파라미터 표현에 비트를 더 할당한다.

2. CELP 부호화기의 원리

2-1. 개요

CELP(Code Excited Linear Prediction) 부호화기는 코드북내에 저장된 입력 여기 신호열을 두 개의 시변 선형 회귀(Time-varying Linear Recursive) 필터를 통과시킴으로써 얻은 신호 중 주어진 충실도 판정을 최적화 시키는 것을 선택하도록 구성되어 있다[2]. 이러한 CELP 부호화기는 입력으로 얻어진 음성신호를 분석하여 필요한 파라미터를 추출하고 이를 이용, 음성신호를 합성하여 입력음성 신호와 비교하는 소위 합성에 의한 분석(Analysis By Synthesis) 방법을 사용함으로써 음질이 매우 우수하다. 그림 2-1은 CELP 부호화기의 기본 구조를 나타낸 것이다[2].

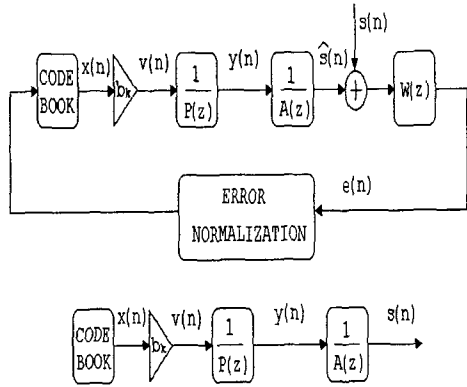


그림 2-1. CELP 부호화기의 기본 구조

그림 2-1에서 s(n)은 입력된 음성 신호이며 x(n)은 코드북내에 저장되어 있는 입력 여기 신호열을 나타낸다. 이 입력 여기 신호열 x(n)을  $b_k$ 를 이용하여 적절하게 크기를 조절한 후, 두 시변 회귀필터  $1/P(z)$ ,  $1/A(z)$ 를 통과시키면 합성 음성신호  $\hat{s}(n)$ 을 얻는다. 두 음성신호 s(n)과  $\hat{s}(n)$ 과의 차이를 오차 가중 필터 W(z)에 통과시키면 오차 신호인 e(n)을 얻게 되는데, 이 e(n)을 평균 제곱오차를 이용하여 비교함으로써 가장 적은 오차 신호를 나타내는 x(n)을 얻게 된다. 이때, P(z), A(z)의 계수는 입력된 음성 신호를 이용하여 구하고, 코드북 내의 각 원소들은 백색 가우시안 불규칙 수열들로부터 구성한다.

2-2. 스펙트럼 필터

포먼트 예측기 또는 단기 예측기(Short-term Predictor)라고도 불리는 스펙트럼 필터는 보통 10차의 차수를 갖는 LPC 계수를 얻기 위하여 자기 상관방법을 사용하게 된다.

다음은 본 논문에 사용된 G.723.1 ACELP의 LPC 분석에 대한 것이다.

G.723.1 ACELP의 LPC 분석은 60샘플의 4개의 부프레임에서 10차의 선형예측분석이 수행된다. 180표본의 해밍 윈도우(Hanning Window)가 각 부프레임의 중앙에 위치하게 되며 11개의 자기 상관계수가 윈도우 처리된 신호로부터 계산되어진다. 선형예측계수(LPC)는 Levinson-Durbin recursion을 사용하여 계산되어지며, 모든 입력 프레임의 각 부프레임마다 하나씩 계산되어 네 집합의 LPC 계수가 계산되어진다. LPC 합성 필터는 다음과 같이 정의된다[3].

$$A_i(z) = \frac{1}{1 - \sum_{j=1}^{10} a_{ij}z^{-j}}, \quad 0 \leq i \leq 3 \quad (2.1)$$

여기서  $i$ 는 0과 3사이에서 정의되는 부프레임 인덱스이며  $j$ 는 차수를 나타낸다.

3. 제안한 알고리즘

3-1. 발생속도의 측정

발생속도는 1초의 음성신호 중 포함된 음소의 수로 정의할 수 있다. 본 논문에서는 음소를 직접 측정하지 않고 단지 스펙트럼의 변화를 추정하여 음소의 변화를 측정하였다. 그리고 스펙트럼 변화가 3프레임 60msec 동안 발생하지 않으면 발생속도가 느린 부분으로 간주하고, 3프레임동안 스펙트럼의 변화가 발생하면 빠른 발생 부분으로 간주한다.

식 3.1은 스펙트럼 변화를 추정하기 위해 사용한 파라미터를 나타낸 것이다.

$$D_n = \frac{R_n(0)}{R_n(1)}, \quad R_n(j) = \sum_{m=0}^{N-1-j} s_n(m)s_n(m+j) \quad (3.1)$$

$D_n$ 은 스펙트럼 기울기를 나타내는 파라미터이고,  $R_n(j)$ 는 자기상관계수이다.

그림 3-1에서 3-4까지는 60msec동안의 스펙트럼 기울기의 변화도를 나타낸 것이다. 프레임 크기를 30msec로 하여 15msec 겹쳐 분석하였다. 그림 3-1, 3-2, 3-4의 경우에는 변화가 일어나는 부분이고, 3-3은 변화가 일어나지 않는 구간이다. 그림 (b)는 각 구간의 스펙트럼 기울기를 나타내고 있다. 세 개의 그래프가 차이가 없으면 발생속도가 느린 부분으로 간주하고, 차이가 크면 발생속도가 빠른구간으로 간주한다.

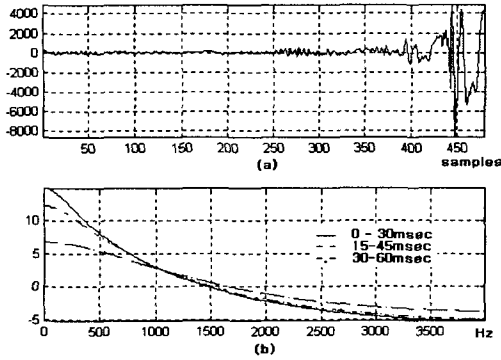


그림 3-1. 스펙트럼 변화의 추정  
(a) 음성파형 (b) 스펙트럼 기울기

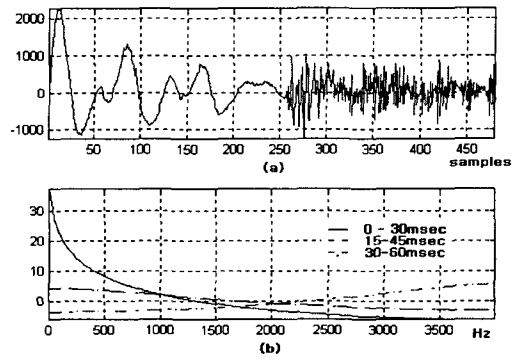


그림 3-4. 스펙트럼 변화의 추정  
(a) 음성파형 (b) 스펙트럼 기울기

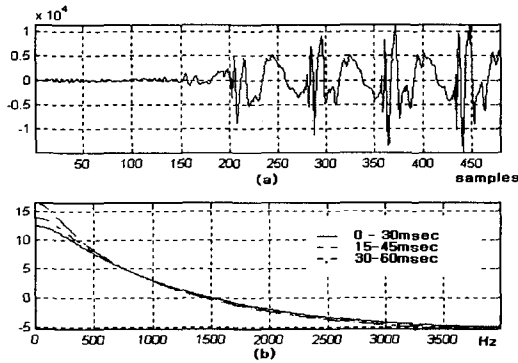


그림 3-2. 스펙트럼 변화의 추정  
(a) 음성파형 (b) 스펙트럼 기울기

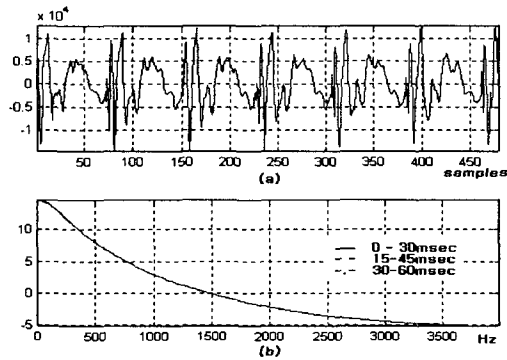


그림 3-3. 스펙트럼 변화의 추정  
(a) 음성파형 (b) 스펙트럼 기울기

그림 3-1에서 3-4까지는 60msec동안의 스펙트럼 기울기의 변화도를 나타낸 것이다. 프레임 크기를 30msec로 하여 15msec 겹쳐 분석하였다. 그림 3-1, 3-2, 3-4의 경우에는 변화가 일어나는 부분이고, 3-3은 변화가 일어나지 않는 구간이다. 그림 (b)는 각 구간의 스펙트럼 기울기를 나타내고 있다. 세 개의 그래프가 차이가 없으면 발성속도가 느린 부분으로 간주하고, 차이가 크면 발성속도가 빠른구간으로 간주한다.

#### 4. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC 586(300MHz)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다. 처리결과와 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 각 5번씩 발성하여 시료로 사용하였다. 음성 시료는 SNR이 30dB인 환경 하에서 녹음하였다. 음성 시료는 다음과 같다.

- 발성1: /인수내 꼬마는 천재소년을 좋아한다/
- 발성2: /예수님께서 천지창조의 교훈을 말씀하셨다/
- 발성3: /창공을 헤쳐 나가는 인간의 도전은 끝이 없다/
- 발성4: /숭실대학교 정보통신과 음성통신 연구팀이다/

제안한 알고리즘의 시뮬레이션은 G.723.1 ACELP 부호화기에 C-언어로 구현하여 수행하였다. 제안한 알고리즘의 성능 비교는 G.723.1 Annex A를 통과한 음성과 제안한 알고리즘을 통과한 음성의 MOS 측정하였다. 부호화기는 240샘플 프레임마다 처리한다(8kHz 샘플링에서 30ms). 각 프레임은 DC 성분을 제거하기 위해 하

이 패스 필터를 통과하고, 60샘플의 4개의 부프레임으로 나누어지고 모든 두 부프레임에서 발생속도를 측정한다. 윈도우 크기는 240, 480의 2종류를 사용하였다. 부호화기의 블록도는 그림 4-1.에 나타내었다.

각의 경우에 따라 전송률을 달리하여 부호화 하였다. 결과 전송률은 감소하고, 음질에서는 큰 차이를 보이지 않았다.

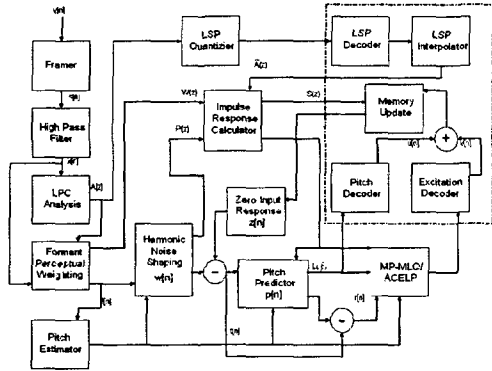


그림 4-1. G.723.1 음성 부호화기 블록도[3]

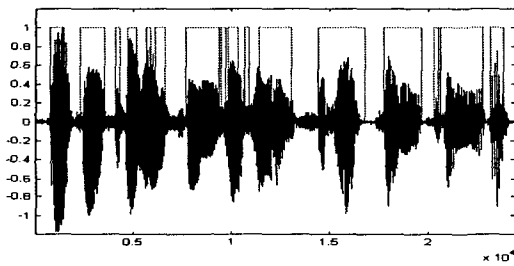


그림 4-2. 느린 발생속도를 가진 위치 측정 결과

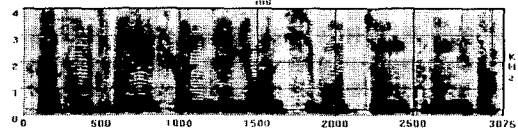
그림 4-1은 발생속도가 느린 부분으로 간주된 부분을 나타낸 그림이다. 발생속도가 느린 부분은 480샘플(60msec)크기의 윈도우를 적용하여 분석하고 부호화하였다. 그리고, 발생속도가 빠르게 측정된 부분은 5.3kbps로 부호화 하였고, 발생속도가 느린 부분은 6.3kbps로 부호화 하였다. 전송률을 가변적으로 적용한 결과 20%의 전송률이 감소하였고, 음질의 열하는 발생하지 않았다.

### 5. 결론

CELP 부호화기는 선형 예측을 통한 합성에 의한 분석을 기본원리로 한 부호화기이다. 그러나 고정 윈도우를 사용하여 분석함으로써 시간적으로 발생속도가 변하는 음성신호를 적응적으로 분석하지 못하는 단점이 있다. 따라서 본 논문에서는 스펙트럼 변화를 측정하여 발생속도를 빠르고, 느린 부분으로 분류한 후 각



(a) 음성파형의 스펙트로그램



(b) G.723.1 보코더를 통과한 신호의 스펙트로그램



(c) 제안한 보코더를 통과한 신호의 스펙트로그램

그림 4-3. 스펙트로그램의 비교

표 4-1. 전송률 비교(bit/sec)

	발성1	발성2	발성3	발성4
G.723.1	5306	5324	5322	5326
제안한 알고리즘	4495	4332	4585	4391
전송률감소 (bit/sec)	807	992	737	935
감소율(%)	15.21	18.64	13.86	17.65

### 참고 문헌

- [1] 이미숙, "IMT-2000을 위한 음성부호화연구", 가일자망연구소 정보통신연구 제13권 제1 호, 1999.3
- [2] A.M. Kondoz, "Digital Speech", John Wiley & Sons, 1994.
- [3] ITU-T Recommendation G.723.1, March, 1996.
- [4] 정찬중, 나덕수, 신동성, 배명진, "스펙트럼 누설에너지를 이용한 음성신호의 창함수 적용에 관한 연구", 한국통신학회, 하계종합학술대회 발표회 논문집(상), PP. 487-490, 1999.7
- [5] W. B. Kleijn et. al, "Speech Coding and Synthesis", Elsevier Science B.V., 1995.
- [6] "음성 신호 처리 기술", 한국과학기술원, 삼성첨단기술센터, 제 3권 음성부호화, PP.171