

CELP 부호화기의 코드북 탐색 시간 개선

Reduced Search for a CELP Adaptive Codebook

이지웅* · 나 훈* · 정대권*

Ji-Woong Lee*, Hoon Na*, and Dae-Gwon Jeong*

요 약

본 논문에서는 웨이블릿 변환계수를 이용하여 CELP 부호기의 코드북 탐색 시간을 개선하는 새로운 방법을 제안하고 있다. CELP 음성 부호화기에서 적응 코드북 탐색시 개루프(open loop) 및 폐루프(closed loop) 탐색을 이용하여 피치를 추출하는 과정은 많은 연산량을 필요로 한다. 특히 개루프 탐색에서 전구간에 걸쳐 자기상관함수를 이용한 피치 추출은 과도한 탐색시간에 비효율적인 것으로 지적되어 왔다. 본 논문에서는 다양한 스케일리 계수에 대한 다이애덕 웨이블릿 변환계수의 최대값의 위치가 변하지 않는 주파수에 피치가 존재하는 특성을 이용하여 새로운 적응코드북 탐색 방법을 제안하고 있다. 제안한 알고리즘은 기존의 적응 코드북 탐색 시간을 25% 감소시켰고, 합성된 음성신호의 음질의 차이는 거의 느낄 수 없었다.

Abstract

This paper proposes a reduction scheme for codebook search time in the adaptive codebook using wavelet transformed coefficients. In a CELP coder, pitch estimation with a combined open loop and closed loop search in adaptive codebook needs a lengthy search. More precisely, the pitch search using autocorrelation function over all possible ranges has been shown inefficient compared to the consuming time. In this paper, we propose a new adaptive codebook search algorithm which ensures the same position for the pitch with maximum wavelet coefficient over various scaling factors in Dyadic wavelet transform. A new adaptive codebook search algorithm reduces 25% conventional search time with almost the same quality of speech.

I. 서 론

정보통신 문화가 발달함에 따라 디지털 이동통신이나, 멀티미디어, 음성우편 시스템 등 음성을 이용한 여러 가지 새로운 산업들이 급속히 성장하고 있다. 이 중에서도 특히 디지털 이동통신분야에 대한 연구가 활발한데, 이는 디지털 시스템에서는 부호화기를 사용하여 음성신호를 압축하기 때문에 아날로그 시스템에 비해 채널 증대를 가져올 수 있기 때문

이다. 이에 따라 음성 부호화 알고리즘에 대한 연구가 매우 활발히 진행되고 있다. 특히 많은 연구가 되고 있는 부호화기는 선형예측을 기본으로 하는 음성 분석-합성(analysis-by-synthesis) 방식이다.

음성 부호화란 음성 신호를 디지털 부호를 변환시키는 것을 말하여 음성의 전송 또는 저장 등에 주로 사용된다. 음성 부호화의 방법은 크게 파형 부호화(waveform coding), 보코딩(vocoding: voice coding) 그리고 혼성 부호화(hybrid coding)로 분류할 수 있다.

* 한국항공대학교 항공전자공학과 (Dept. of Avionics Eng., Hankuk Aviation Univ.)

· 논문번호 : 2000-1-7

· 접수일자 : 2000년 6월 15일

디지털 통신 시스템에 사용될 음성 부호화기는 낮은 전송률에서도 아날로그 통신에서 제공하던 정도로 양호한 품질의 서비스가 가능해야 하므로 앞에서 세가지 부호화 방식 중에서 주로 혼성 부호화 방식에서 연구가 활발히 진행되고 있다. 최근에는 혼성 부호화기 중에서 4~13 kbps의 CELP 형 부호화기가 널리 사용되고 많은 연구가 진행되고 있는데 ITU-T 에서 1995년에 8 kbps 음성부호화기로 표준화된 ACELP(algebraic CELP)기반의 CS-ACELP G.729와, 멀티미디어 기술에서 응용할 수 있는 MPEG 4 오디오에서 6~24 kbps 중간 비트율의 CELP 부호화기, ETSI(european telecommunication standard institute)에서 1996년 ACELP 기반의 GSM EFR 음성부호화기, 그리고 ACELP 기반의 부호화기로 채널의 상태에 따라 다양한 전송률을 제공하는 IMT 2000 음성 부호화기인 AMR (Adaptive Mult-Rate)부호화기 등이 있다[4],[5].

이런 CELP 부호화기는 우수한 품질과 낮은 전송률을 가진 장점이 있으나 적응 코드북과 고정 코드북 탐색시에 막대한 연산량이 소모되는 단점이 있다. 특히 적응 코드북 탐색 과정에서 후보 피치를 찾기 위해 전구간에 걸친 개루프 탐색과정은 비효율적이므로 효율적이며 연산량을 줄일 수 있는 연구가 필요하다. 따라서 기존의 CELP 음성 부호화기 보다 연산량을 줄이며 품질의 저하를 일으키는 않는 새로운 알고리즘 연구가 필요하다.

이러한 관점에서 본 논문에서는 적응 코드북 탐색시의 연산량을 줄이기 위해 웨이블릿 변환을 이용한 적응 코드북 탐색 방법을 제안하였다. 실제로 웨이블릿 변환을 이용한 적응 코드북 탐색방법을 GSM EFR 음성 부호화기에 적용하였을 때 프레임 당 약 25%의 적응 코드북 탐색 시간을 감소시켰다.

II. CELP 부호화기

CELP(codebook excited linear prediction) 부호화기는 합성에 의한 분석법을 사용하는 음성 부호기들 중에서 가장 널리 쓰이고 있다. CELP 부호기는 1982년에 Atal에 의해서 처음 제안된 후에 최근까지도 변형된 알고리즘이 계속 제안되고 있다.

CELP 부호기에서는 여기 신호를 하나의 벡터로 보고, 원음과의 에러를 최소화할 수 있는 여기 신호 벡터를 미리 준비된 코드북에서 선택하는 방법을 사용한다. 이러한 방법은 스펙트럼 포락 정보와 피치 정보를 원음 신호에서 제거하고 남은 신호가 랜덤 잡음 형태로 나타나는 것을 이용한 것이다. 따라서 서로 다른 랜덤 잡음들을 코드 벡터로 가진 코드북을 미리 준비하여 이 중에서 하나를 선택하여 피치 정보와 스펙트럼 포락선 정보를 부가함으로써 합성음을 생성해 보고, 원음과의 에러를 최소화하는 파라미터들을 선택한다. 초기의 CELP 부호기는 여기 신호 분포를 가우스(Gauss) 분포로 가정하여 가우스 코드북을 이용하였다. 그러나 코드북 탐색을 위한 과정에서 많은 처리시간이 요구되므로 탐색 시간을 줄이기 위해 겹친 가우스 코드북(overlapped Gaussian codebook)을 이용하여 탐색 시간을 줄이려 하였다. 그러나 실제 음성 신호는 가우스 분포가 아니므로 품질 향상을 기대하기 어렵다[2]. 이런 문제를 해결하고자, 다수의 0과 몇 개의 -1과 1로 구성된 대수 코드북(algebraic codebook)이 제안되었다. 대수 코드북은 통계학적 분포를 이용한 가우스 코드북에 비해 그 탐색시간이 대폭 줄었으며, 원음에 더 가깝게 음성을 합성할 수 있다.

CELP 부호화기는 크게 스펙트럼 포락 정보를 얻기 위한 선형 예측(linear prediction, LP) 분석 과정과 피치 주기에 해당하는 피치 파라미터를 예측하기 위한 LTP 탐색과정, 그리고 랜덤 잡음 성분을 모델링하기 위한 고정 코드북(fixed codebook) 탐색 과정으로 나누어진다. 피치 파라미터를 탐색하는 과정은 개루프(open loop) 탐색 과정과 폐루프(closed loop) 탐색 과정이 있다. 그림 1은 CELP 부호화의 블록 다이어그램이다.

만약 과거의 복호화된 여기 신호를 이용하여 현재 프레임 내 신호와의 유사성을 찾다면 이 과정은 적응 코드북(adaptive codebook) 탐색 과정이다.

2-1 선형 예측 분석

음성 생성 모델에서 식 (1)과 같이 여기 신호 스펙트럼 $G(z)$ 가 스펙트럼 포락(spectrum envelo-

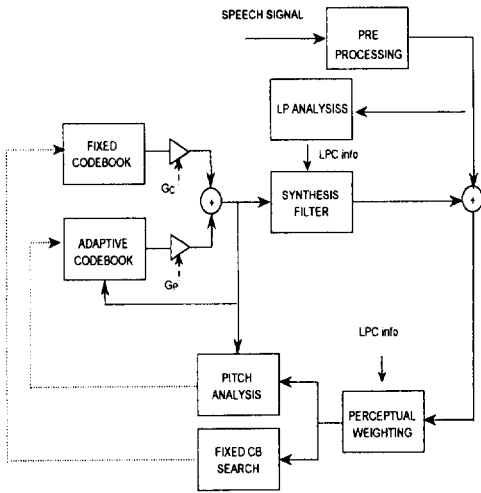


그림 1. 일반적인 CELP 부호화기 블록도
Fig. 1. Block diagram of the general CELP encoder.

pe)을 나타내는 조음 필터 $H(z)$ 를 통과함으로써 합성음 스펙트럼 $S(z)$ 가 생성된다.

$$S(z) = H(z)G(z) \tag{1}$$

식 (1)에서 $H(z)$ 는 일반적으로 식 (2)와 같이 all-pole 모델로 표현되며 선형 예측 방법에 의해서 쉽게 파라미터를 구할 수 있다.

$$H(z) = \frac{1}{1 - \sum_{i=1}^p a_i z^{-i}} \tag{2}$$

이 필터의 계수 a_i 를 구하기 위한 방법에는 자기상관법(autocorrelation method)와 공분산법(covariance method)이 있는데, CELP 부호화기에서는 선형 예측 합성 필터의 안정성을 보장해 주는 자기상관법을 이용한다[2]. 자기상관법에 의해 얻어진 선형 예측 필터의 계수들은 양자화 과정을 거치게 되는데, 양자화 오차를 최소화하기 위해 LSF(Line Spectral Frequency)로 변환된 후 양자화된다[2].

여기 신호는 $H(z)^{-1}$ 을 전달 함수로 하는 역 필터를 사용하여 구해낸다.

$$H(z)^{-1} = 1 = \sum_{i=1}^p a_i z^{-i} \tag{3}$$

2-2 적응 코드북 탐색

피치 주기에 해당하는 파라미터를 예측하기 위한 적응 코드북 탐색 과정은 개루프 탐색 과정과 폐루프 탐색 과정으로 이루어진다. 먼저 개루프 탐색 과정은 프레임 단위로 이루어지고 개루프 탐색 과정에서 찾은 후보 피치의 주위에서 폐루프 과정을 통해 정확한 적응 코드북의 피치 인덱스를 찾기 위한 피치 분석의 단순화가 목적이다.

개루프 탐색 과정은 지각적으로 가중된 음성 신호 $s_w(n)$ 에 대해 다음 식 (4)와 같이 자기상관함수를 이용한다.

$$O_k = \sum_{n=0}^L s_w(n)s_w(n-k) \tag{4}$$

여기에서 $k=18, 19, \dots, 143$ 은 피치 지연을 나타내고, L 은 프레임 크기를 나타낸다. 각 피치 지연 k 에 대해 자기상관함수 결과 O_k 를 다음과 같이 세 부분으로 나눈 다음, 각 부분에서 하나씩 최대값을 찾는다.

$$\begin{aligned} i=3 &: 18, \dots, 35 \\ i=2 &: 36, \dots, 71 \\ i=1 &: 72, \dots, 143 \end{aligned} \tag{5}$$

마지막으로 남은 세개의 최대값 $O_i, i=1, 2, 3$ 에 대해서 $\sqrt{\sum_n s_w^2(n-t_i)}$ 으로 나누어 정규화(normalize)하고 (M_i, t_i) 로 나타낸다. 여기에서 M_i 는 정규화된 최대값, t_i 는 각 3개의 범위에서 O_k 가 최대값을 갖는 피치 지연 k 를 나타낸다. 3개의 피치 지연 중에서 최종 후보 피치 T_{op} 은 피치 지연 값이 적은 곳에 가중치를 주어 가장 큰 피치 M_i 를 갖는 t_i 를 T_{op} 로 선택한다.

개루프 탐색 과정에서 찾은 최종 후보 피치 T_{op} 근처에서 부프레임 단위로 피치를 찾는다. 이때 정확한 피치를 찾기 위해서 부분 피치 지연을 사용하고 원 음성과 합성된 음성사이의 평균 자승 오차를

최소화(minimum mean square error)하는 정수 피치 지연을 찾고, 정수 피치 지연에 대해 더 세밀하게 부분 피치 지연을 찾는다.

피치 지연 D 에 대한 입력 신호 $\hat{s}(n)$ 와 합성된 음성 신호 $\tilde{s}(n)$ 사이의 오차 $E(D)$ 는 다음과 같다.

$$E(D) = \sum_{n=0}^{L-1} (\hat{s}(n) - \tilde{s}(n))^2 \quad (6)$$

합성된 음성 신호 $\tilde{s}(n)$ 은 지각적으로 가중된 선형 예측 필터 응답 $h_w(n)$ 과 D 만큼 지연된 과거 여기 신호와의 컨벌루션으로 다음과 같이 표시된다.

$$\tilde{s}(n) = \beta \sum_{k=0}^n \hat{\gamma}(n-k-D) h_w(k) \quad (7)$$

페루프 탐색과정은 $E(D)$ 를 최소로 하는 피치 지연 D 와 이득 β_i 를 찾는 것으로 한다. 그림 2는 페루프 탐색 과정을 보여 주고 있다.

2-3 고정 코드북 탐색

초기의 CELP 모델은 랜덤한 잡음성분을 모델링 하기 위해서 가우스 코드북을 사용하였다. 그러나 실제 여기 신호는 가우스 분포가 아니며 존재할 수 있는 모든 신호를 가우스 코드북으로 나타낼 수 없다[2]. 또한 전송률이 낮아질수록 많은 정보가 손실 되므로 코드북의 역할은 중요하게 되었다. 따라서 음질을 높이기 위해 새로운 코드북의 필요성이 고조

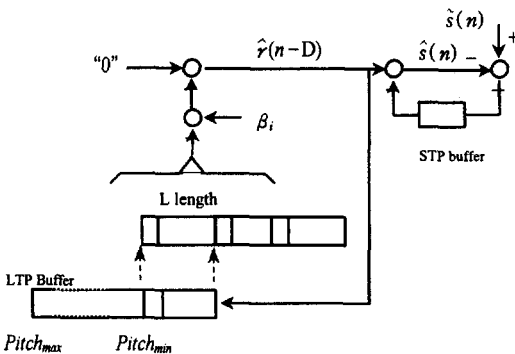


그림 2. 페루프 탐색 과정

Fig. 2. Block diagram of the closed loop search analysis.

표 1. GSM EFR 부호화기에서 사용하는 고정 코드북의 구조

Table 1. Structure of fixed codebook was used at GSM EFR encoder.

Track	Pulse	position
1	i_0, i_5	0, 5, 10, 15, 20, 25, 30, 35
2	i_1, i_6	1, 6, 11, 16, 21, 26, 31, 36
3	i_2, i_7	2, 7, 12, 17, 22, 27, 32, 37
4	i_3, i_8	3, 8, 13, 18, 23, 28, 33, 38
5	i_4, i_9	4, 9, 14, 19, 24, 29, 34, 39

되었다. 최근의 CELP형 부호화기는 ISSP(Interleaved Single Pulse Permutation)을 사용하는 대수 코드북(algebraic codebook)을 기본으로 하고 있다. 이 대수 코드북은 4개의 0이 아닌 펄스를 포함한다. 각 펄스는 1과 1을 진폭으로 가진다. GSM EFR 부호화기에서 사용하는 고정 코드북의 구조는 표 1과 같다[5].

코드북 탐색은 부프레임 단위로 행하며 원 신호와 합성 신호 사이의 평균자승 오차를 최소화하는 코드북 인자와 코드북 이득을 찾는다. 페루프 피치 탐색에서 사용된 타겟 신호를 $x(n)$ 이라 하면, 적응 코드북 타겟 신호 $\hat{x}(n)$ 은 다음과 같다.

$$\hat{x}(n) = x(n) - g_k y(n) \quad (8)$$

여기에서 $y(n)$ 은 합성 필터를 거친 적응 코드북 벡터이고, g_k 는 적응 코드북 이득이다. k 번째 대수 코드북 벡터를 c_k 라 하면 대수 코드북 탐색은 식 (9)를 최대화하는 코드 벡터를 찾는 과정이다.

$$\frac{c_k^2}{E_k} = \frac{\left(\sum_{n=0}^{39} d(n) c_k(n) \right)^2}{c_k^t \Phi c_k^t} \quad (9)$$

여기 $d(n)$ 은 역방향 필터링된 신호이며 $\Phi = H^t H$ 는 임펄스 응답으로 이루어진 lower triangular Toeplitz 행렬이다. $d(n)$ 은 타겟 신호 $\hat{x}(n)$ 와 임펄

스 응답 $h(n)$ 으로부터 다음과 같이 얻어진다.

$$d(n) = \sum_{i=n}^{39} \hat{x}(i)h(i-n), \quad n=0, \dots, 39 \quad (10)$$

III. 웨이블릿 변환

3-1 개 요

시간영역에서 정의된 신호를 시간영역이 아닌 다른 변환된 영역, 즉 주파수 영역에서 분석하고자 할 때 사용되는 대표적인 변환기법으로는 푸리에 변환(Fourier transform)을 예로 들 수 있다. 그러나, 푸리에 변환에서는 변환에 사용되는 기저함수인 복소 기저함수가 시간영역의 무한구간에서 정의되기 때문에 다음과 같은 제약을 갖고 있다. 먼저, 신호가 가지고 있는 시간영역에서의 국부적인 정보가 변환된 주파수 영역에서는 넓게 퍼져서 나타나고, 또 기저함수가 모든 시간에 대해 정의되므로 실제 계산상의 어려움이 있다.

이런 문제점을 해결하기 위해 윈도우 기법을 도입한 단시간 푸리에 변환(Short Time Fourier Transform, STFT)이 정의되었다[1]. STFT은 특정한 시간구간과 주파수대역 안에 에너지가 집중되어 있는 신호의 분석에 유리하다. 그러나 모든 주파수 대역에서 동일한 길이를 갖는 윈도우를 사용하기 때문에 시간-주파수 평면에서 제공되는 분석의 해상도

는 균일하다. 그림 3(a)에서 보는 바와 같이, STFT에 의해 신호를 주파수 영역에서 분석할 때, 임의의 시간구간에 대해 전 주파수 대역에서 주파수축상의 해상도는 일정하다.

이와는 달리, 주파수 변환 파라미터와 시간천이 파라미터의 함수로 주어지는 웨이블릿 변환(wavelet transform)에서는 시간-주파수 평면상에서 신호의 다해상도 분석(multiresolutional analysis)이 가능하다. 그림 3(b)에서 보는 바와 같이, 웨이블릿 변환에서는 고주파영역에서 시간축상의 해상도가 높고 주파수축상의 해상도가 낮은 반면에, 저주파영역에서는 주파수축상의 해상도가 높고 시간축상의 해상도가 낮은 특성을 지닌다. 실제의 음성에서 짧은 시간동안에 급격한 변화를 하는 부분은 고주파성분에 해당하고, 긴 시간에 걸쳐 일어나는 완만한 변화는 저주파성분에 해당하므로, 웨이블릿 변환이 갖는 다해상도 특성은 음성신호의 분석에 적절함을 알 수 있다.

3-2 다이애딕 웨이블릿 변환

웨이블릿(wavelet)은 식 (11)과 같이 모 웨이블릿(mother wavelet)를 $\Psi(t)$ 를 확장(dilation)하고 이동(translation)하여 얻어지는 함수들의 집합을 말한다.

$$\Psi_{a,b}(t) = a^{-1/2} \Psi\left(\frac{t-b}{a}\right) \quad (11)$$

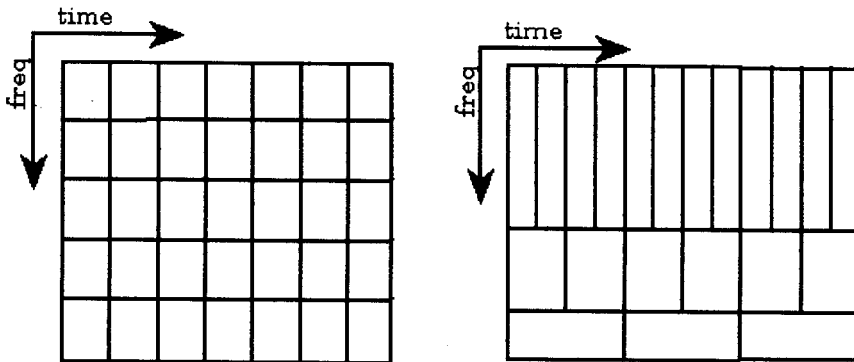


그림 3. (a) 푸리에 변환과 (b) 웨이블릿 변환의 시간-주파수 해상도

Fig. 3. (a) Fourier transform (b) Temporal Frequency Resolution of Wavelet transform.

여기서 a 를 스케일(scale), b 를 이동(translation) 인자라고 한다.

함수 $f(t)$ 의 연속 웨이블릿 변환(CWT, Continuous Wavelet Transform)은 식 (12)와 같이 정의된다.

$$CWT_f(b, a) = a^{-1/2} \int_{-\infty}^{+\infty} \frac{|\Psi(w)|^2}{|w|} dw < \infty \quad (12)$$

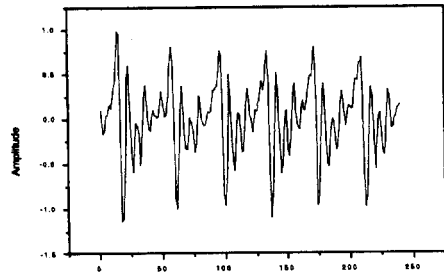
매개변수 a 는 웨이블릿 기저의 크기를 조정하고, b 는 시간축 상에서의 변이를 나타내며, b 를 변화시킴으로써 웨이블릿 기저를 원하는 곳에 놓을 수 있다. CWT는 계산적으로 매우 복잡하고 많은 잉여 정보를 가지고 있다. 그러나 스케일 인자 a 나 이동 인자 b 와 스케일 인자 a 둘 다를 이산화하면 계산의 복잡도와 잉여 정보를 감소시킬 수 있다[8],[9].

다이애딕 웨이블릿 변환(DyWT)은 스케일 인자 a 가 다이애딕 열 2^j , $j=1, 2, 3, \dots$ 을 따라서 이산화 된 것으로 다음과 같이 정의된다.

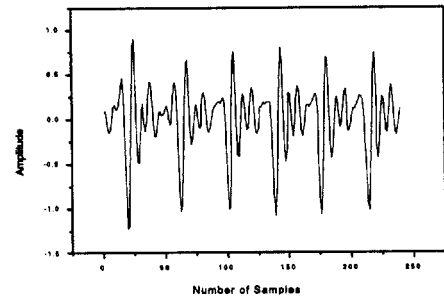
$$DyWT_f(b, 2^j) = 1/2^j \int_{-\infty}^{+\infty} f(t) \Psi\left(\frac{t-b}{2^j}\right) dt \quad (13)$$

DyWT는 음성 신호 분석에 유용한 다음과 같은 특성을 갖는다. 첫째로, 선형성과 시불변성으로서 음성 신호는 주로 움직임 감쇄하는 사인 곡선의 선형 결합으로 모델링되므로 음성 신호 분석에 적합하다. 둘째로, 웨이블릿 변환을 이용하여 신호의 급격하고 느린 변화를 감지할 수 있다. 즉, 신호나 신호의 도함수가 불연속성을 가지면 $|DyWT(b, 2^j)|$ 는 불연속 점에서 국부 신호의 최대값을 갖는다. 따라서 성문의 폐쇄(glottal closure)와 같은 음성 신호의 급격한 변화를 감지할 수 있다[6],[7].

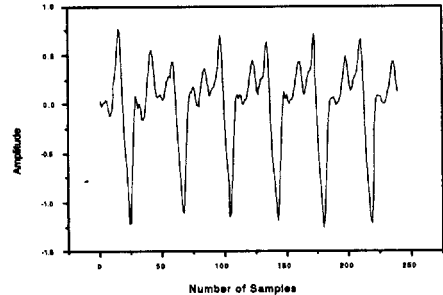
유성음의 기본 주파수(피치)는 저주파 성분(40~500 Hz)을 가지며, 무성음은 고주파 성분을 가지므로 스케일 변수의 수를 제한하여 연산량을 줄일 수 있다. 그림 4는 스케일 $j=3, 4, 5$ 음성 신호들의 DyWT을 보여 주고 있다. 각 스케일에서 DyWT의 국부 최대값(local maxima)들을 나타내고 있으며,



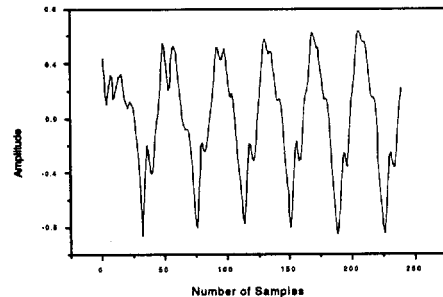
(a) 원 음성 데이터
(a) Original voice data



(b) 스케일 $j=3$ 에서 DyWT
(b) DyWT at Scale $j=3$



(c) 스케일 $j=4$ 에서 DyWT
(c) DyWT at Scale $j=4$



(d) 스케일 $j=3$ 에서 DyWT
(d) DyWT at Scale $j=3$

그림 4. 음성 신호의 다이애딕 웨이블릿 변환
Fig. 4. DyWT of voice data.

스케일 파라미터가 증가할수록 고주파에 해당하는 성분은 제거됨을 알 수 있다. 스케일 $j=3,4,5$ 에서 DyWT국부 최대값들의 위치를 통해 피치를 찾을 수 있다.

IV. 웨이블릿 변환을 이용한 적응 코드북 탐색

본 논문에서는 개루프 탐색 과정에서 전 구간에 걸쳐 자기상관함수를 이용하여 최종 후보 피치를 찾는 대신 웨이블릿 변환을 이용하여 찾은 7개의 후보들에 대해 자기상관함수 적용하는 방법을 통해 적응 코드북 탐색 과정의 연산량을 줄이는 방법을 제안한다.

4-1 다이애딕 웨이블릿 변환을 이용한 피치 분석

Shuba[7]는 성문의 폐쇄는 DyWT의 연속적인 스케일 $j=3,4,5$ 에서 국부 최대값이 일치하는 지점에서 발생한다고 가정하였다. 그러나 실질적으로 국부 최대값이 작은 구간에서 연속적으로 발생하는 경우에 정확한 피치를 찾아내는 데 어려움이 있고, 스케일 $j=3,4,5$ 에 대한 DyWT의 연산량도 무시할 수 없다.

본 논문은 적은 연산량으로 정확한 피치를 찾기 위해 웨이블릿 변환을 통해 얻어진 후보들에 대해 자기상관함수를 이용하는 방법을 사용한다. 연속적인 스케일 $j=3,4,5$ 에 대해서 DyWT를 취하는 대신, 스케일 $j=5$ 에서만 DyWT를 취한 다음, 국부 최대값들의 위치들의 조합을 통해 7개의 후보를 선

택한다. 7개의 후보 피치 지연들에 대해서 정확한 최종 후보 피치를 찾기 위해 자기상관함수를 적용하였다. 또한 DyWT에서 연산량을 줄이기 위해, 식 (14)로 정의되는 Haar 웨이블릿(Haar Wavelet)을 사용한다.

또한 웨이블릿 변환을 가하기 전 효율적인 국부 최대값을 찾아내기 위해 저역통과필터(Low Pass Filter: LPF) 특성을 가지고 있는 Haar 스케일링 함수를 이용한다. Haar 스케일링 함수는 식 (15)로 정의된다. Haar 웨이블릿과 스케일링 함수의 특성 상 입력신호에 대해 컨벌루션을 취할 때 과거의 컨벌루션 값을 이용할 수 있으므로 추가되는 연산은 새로운 입력신호와 제외되는 신호에 대한 연산만이 필요로 하므로 연산량을 감소시킬 수 있다.

$$\Psi_H(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ -1, & 1/2 \leq t < 1 \\ 0, & \text{otherwise} \end{cases} \quad (14)$$

$$\Phi_H(t) = \begin{cases} 1, & 0 \leq t < 1/2 \\ 0, & \text{otherwise} \end{cases} \quad (15)$$

4-2 다이애딕 변환을 이용한 적응 코드북 탐색 알고리즘

그림 7은 제안한 알고리즘 블록도를 나타낸다. 알고리즘은 다음과 같이 수행된다.

- Step 1 : 지각적으로 가중된 입력 음성 신호 $s_w(n)$ 가 들어오면 Haar 스케일링 함수를 이용하여 LPF를 취한다.
- Step 2 : 스케일 $j=5$ 에서 DyWT를 이용하여 5개의 국부 최대값들을 찾는다.

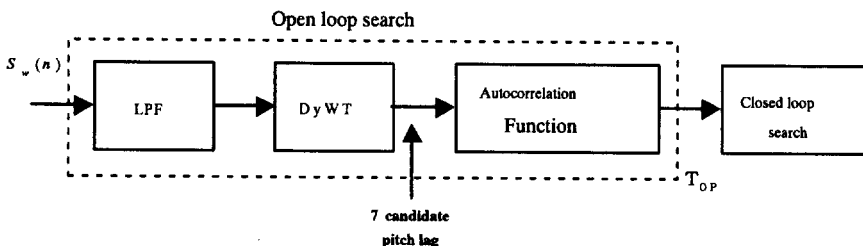
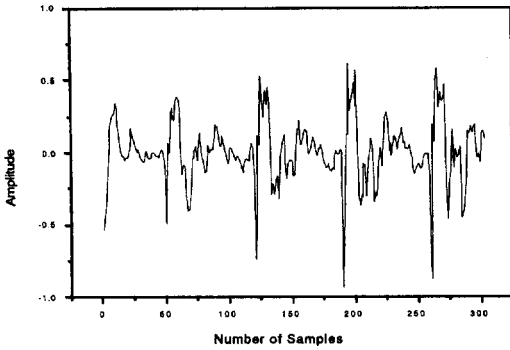
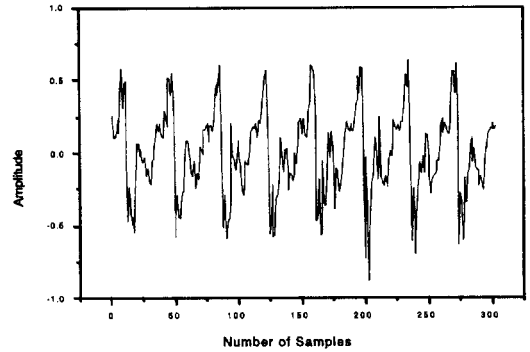


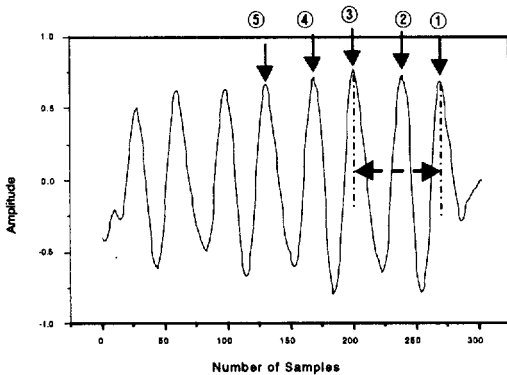
그림 5. 제안하는 알고리즘 블록도
Fig. 5. Proposed algorithm block diagram.



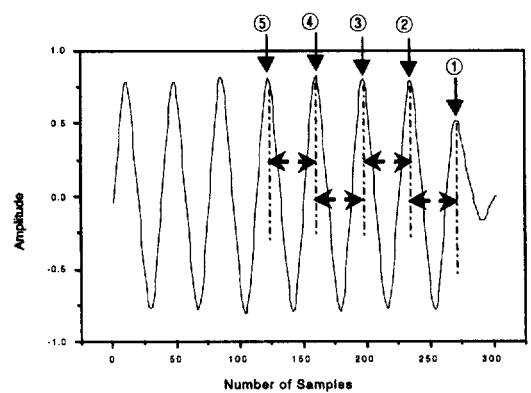
(a) 8 kHz로 샘플링된 남성 음성 '아'의 샘플 일부분
(a) A part of man voice, [a:] at sampling rate 8 kHz



(a) 8 kHz로 샘플링된 여성 음성 '에'의 샘플 일부분
(a) A part of woman voice, [e:] at sampling rate 8 kHz



(b) (a)신호에 대해 제안한 알고리즘을 적용한 신호
(b) Signal using proposed algorithm for signal (a)



(b) (a)신호에 대해 제안한 알고리즘을 적용한 신호
(b) Signal using proposed algorithm for signal (a)

그림 6. 제안한 알고리즘 적용 예 (1)

Fig. 6. Example of proposed algorithm (1).

그림 7. 제안한 알고리즘 적용 예 (2)

Fig. 7. Example of proposed algorithm (2)

Step 3 : 5개의 국부 최대값들의 간격들을 조합하여 다음과 같이 7개의 후보 피치 지연 $L_i, i=1, 2, \dots, 7$ 를 설정한다.

$$\begin{aligned} L_1 &: |T_{M1} - T_{M2}|, & L_2 &: |T_{M2} - T_{M3}| \\ L_3 &: |T_{M3} - T_{M4}|, & L_4 &: |T_{M4} - T_{M5}| \\ L_5 &: |T_{M1} - T_{M3}|, & L_6 &: |T_{M3} - T_{M5}| \\ L_7 &: |T_{M1} - T_{M5}| \end{aligned}$$

여기에서 $T_{Mi}, i=1, 2, \dots, 5$ 는 국부 최대값들의 위치를 나타낸다.

Step 4: 7개의 후보 피치 지연에 대해서만 입력 신호 $s_w(n)$ 에 대해서 자기상관함수 식 (4.3)을 이용하여 최종 후보 피치 T_{op} 를 찾는다. 이때 피치의 더블링을 피하기 위해서 후보 L_1, L_2, L_3, L_4 에 가중치를 준다.

$$O_k = \sum_{n=0}^L s_w(n)s_w(n-k) \quad (16)$$

Step 5: 최종 후보 피치 T_{op} 근처에서 페루프 탐

색 과정을 수행하게 된다. 그림 6과 7은 본 논문에서 제안한 방법을 이용하는 예를 보여준다.

그림 6과 7에서 실선의 화살표는 5개의 국부 최대값들을 나타낸다. 이 5개의 국부 최대값들은 현재 프레임에서 가까운 부근에서부터 선택하였다. 점선으로 표시된 화살표는 실제의 피치 지연을 나타낸다. 실제의 피치 지연은 5개의 국부 최대값 위치값의 조합 중의 하나로 만들어 낼 수 있다.

V. 실험 결과

5-1 실험 환경

제안한 부호화의 성능과 음질 평가를 위해서 GSM EFR 음성 부호화기에 적용하여 제안한 알고리즘의 성능을 평가하였다. GSM EFR 부호화기는 12.2 kbps 음성 부호화기로서 1996년에 GSM 이동통신 시스템(mobile communication system)의 부호화기로 표준화되었고, 최소한 32 kbps ADPCM과 동등한 음질을 제공한다. ACELP(Algebraic Code Excited Linear Prediction) 알고리즘을 기반으로 하고 있으며, 한 프레임의 크기는 20 ms 이며 4개의 부프레임(subframe)으로 구성된다.

평가에 사용된 음성데이터는 표 2와 같이 8kHz로 샘플링한 음성 데이터를 사용하였고 펜티엄 II 333MHz PC를 사용하여 성능 평가를 수행하였다.

5-2 결 과

적응 코드북 탐색시간을 비교하기 위해 시험용 음성 신호들의 프레임당 평균 탐색시간을 구하였다.

표 2. 평가에 사용된 음성 데이터

Table 2. Voice data were used at estimation.

	문 장	프레임 길이	총 프레임수
kf1(한국 여성)	미는 피부 한꺼풀 차이입니다.	20 ms	124
kf2(한국 여성)	지나친 흡연은 건강을 해칩니다.	20 ms	160
km1(한국 남성)	이번 겨울은 예년과 달리 포근합니다.	20 ms	130
km2(한국 남성)	개인통신시대가 조만간에 개막될 것입니다.	20 ms	158

제안한 알고리즘을 적용한 적응 코드북 탐색 시간과 개선되지 않은 EFR 부호화기의 적응 코드북 탐색 시간 비교는 표 3과 같다. 표 3의 결과에서처럼 개선이 이루어지지 않은 EFR 음성 부호화기(Original EFR)의 프레임당 적응 코드북 탐색 시간은 약 36 ms이지만 제 4장에서 제안한 알고리즘을 적용한 개선된 EFR(Modified EFR)의 적응 코드북 탐색 시간은 약 27 ms로 탐색시간이 약 25% 정도 단축되었다.

표 3. 제안한 알고리즘의 적응 코드북 탐색 시간 비교

Table 3. Search time comparison of adaptive codebook using proposed algorithm.

	Original EFR [ms/frame]	Modified EFR [ms/frame]	감소율 [%]
kf1(한국 여성)	37.26	28.16	24.4
kf2(한국 여성)	36.41	27.77	23.7
km1(한국 남성)	36.78	27.42	25.4
km2(한국 남성)	36.97	27.14	26.5

표 4. 개선된 EFR 부호화기 SegSNR 비교

Table 4. SegSNR comparison of improved EFR coder.

	Original EFR [dB]	Modified EFR [dB]	비교
kf1(한국 여성)	6.4930	6.5734	0.0804
kf2(한국 여성)	4.8596	4.8658	0.0062
km1(한국 남성)	5.9647	5.8546	-0.1101
km2(한국 남성)	5.9316	5.8595	-0.0702

음질의 측정은 객관적 척도인 SegSNR(Segmental Signal-to-Noise Ratio)와 청취에 의한 주관적 테스트를 병행하였다. 표 4는 GSM EFR 부호화기와 개선된 GSM EFR 부호화기의 SegSNR을 나타낸다. 표 4에서와 같이 한국 여성 1의 경우 약 0.08 dB 정도 높게 나오고, 한국 여성 2의 경우 거의 같고, 남성 음성의 경우 약 0.07~0.1 dB 정도 낮게 나왔다. 이는 남성의 경우 유/무성음이 혼재하는 프레임에 의한 영향으로 전체적인 음질면에서는 영향을 미치지 못했다. 전체적으로 음질의 저하는 느낄 수 없었다.

VI. 결 론

본 논문은 CELP 형 부호화기의 적응 코드북 탐색 과정에서 탐색 시간을 개선하는 방법에 대한 연구로서, 웨이블릿 변환을 이용한 적응형 코드북 탐색 방법의 알고리즘을 제안하고, 이를 GSM EFR 음성부호화기에 적용하였다. 기존이 CELP형 부호화기는 적응 코드북 탐색과 고정 코드북 탐색 과정에 대부분의 처리 시간이 소요된다. 특히 적응 코드북 탐색에서 비효율적이며 많은 연산량을 필요로 하는 개루프 탐색 과정에 대해 개선된 방법을 제안하였고, 기존의 방법과 프레임당 적응 코드북 탐색 시간에 대해 비교하고 기존의 부호화기와 음질을 비교 평가하였다.

기존의 CELP 형 음성 부호화기는 여기신호의 주기적인 성분을 구성하는 유성음을 모델링하기 위해 적응 코드북을 사용한다. 적응 코드북 탐색과정은 개루프 탐색과 페루프 탐색으로 이루어진다. 개루프 탐색 과정에서 모든 구간에 걸쳐 자기상관함수를 이용하여 후보 피치 지연을 찾아야 하므로, 비효율적이며 페루프 탐색 과정과 비슷한 연산량을 필요로 한다.

기존의 이런 문제점을 해결하기 위해서 웨이블릿 변환을 이용하여 찾은 후보 피치들에 대해서 자기상관함수를 이용하는 방법을 제안하였다. 성문의 폐쇄(glottal closure)와 같은 음성 신호의 급격한 변화는 웨이블릿 변환시 국부 최대값을 나타내므로 이런 국부 최대값들의 위치값들을 이용하여 7개의 후보 피

치 지연을 선정하여, 7개의 후보 피치 지연에 대해서만 자기상관함수를 이용하여 최종 후보 피치 후보를 찾았다. 웨이블릿 변환에서 연산량을 감소하기 위해서 Haar 웨이블릿을 사용하였으며, 7개의 후보 피치 지연에 대해서만 자기상관함수를 이용한 결과 프레임당 적응코드북 탐색 시간이 약 25% 감소하다. 그리고 객관적 척도인 SegSNR 결과는 여자 음성의 경우는 거의 비슷했으며 남자 음성의 경우 약 0.08 dB 정도 낮게 나왔으나 주관적인 음질의 차이는 거의 느낄 수 없었다.

앞으로 진행되어야 할 연구로는 개루프 탐색뿐 아니라 많은 연산량을 차지하는 페루프 탐색 과정에서 복잡한 분석-합성 방법을 사용하지 않고 좀더 효율적이고 정확한 피치정보를 이용하여 연산량을 감소시킬 수 있는 알고리즘 연구가 필요하다.

참 고 문 헌

- [1] L. R. Rabiner, *Digital Speech Processing of Speech Signal*, 2nd, Prentice Hall, 1996.
- [2] A. M. Kondoz, *Digital Speech Coding for Low Bit rate Communication Systems*, John Wiley & Sons, 1996.
- [3] 이황수, "이동통신을 위한 음성 부호화 방식", 음성통신 및 신호처리워크샵, pp. 3-10, 1998.
- [4] IU-T Recommendation G.729, Coding of speech at 8 kb/s using conjugate-structure algebraic code-excited linear prediction (CS-ACELP), June, 1995.
- [5] "Digital Cellular Telecommunication system: Enhanced Full Rate(EFR) speech Transcoding (GSM 06.60)", version 5.1.2.
- [6] Shubha & G. F. Boudreaux-Bartels, "A comparison of a wavelet functions for pitch detection of speech signals", *Proc. of ICA-SSP*, pp. 449~452, May, 1991.
- [7] Shubha & G. F. Boudreaux-Bartels, "Application of the wavelet transform for pitch detection of speech signals", *IEEE Trans. on Information Theory.*, vol. 38, no. 2,

March, 1992.

[8] R. K. Young, *Wavelet theory and its applications*, Kluwer Academic Publishers, 1994.

[9] Raghuvver M. Rao and Ajit S. Bopardikar, *Introduction to the Theory and Applications of Wavelet Transforms*, Addison Wesley,

1998.

[10] Obaidat M. S, Brodzik A and Sadoun B, "A performance evaluation study of four wavelet algorithms for the pitch period estimation of speech signals", *Information Sciences*, V.112 N.1-4, 213-221, 1998. 12. 01.

이 지 응(李智雄)



1998년 : 한국항공대학교 항공전자공학과(공학사)
 2000년 : 한국항공대학교 항공전자공학과 대학원(공학석사)
 관심분야 : 음성 신호처리, 멀티미디어 신호처리

정 대 권(丁大權)



1979년 : 한국항공대학교 항공전자공학과(공학사)
 1987년 : Texas A&M대 대학원(공학석사)
 1990년 : Texas A&M대 대학원(공학박사)
 1990년~1991년 : 한국전자통신연구원 선임연구원
 1991년~현재 : 한국항공대학교 전자·정보통신·컴퓨터공학부 부교수
 관심분야 : 정지영상 및 동영상 압축 기법, 저속음성 부호기법, IMT-2000 부호기, JPEG2000 부호기, 멀티미디어 신호처리

나 훈(羅訓)



1996년 : 한국항공대학교 항공전자공학과(공학사)
 1998년 : 한국항공대학교 항공전자공학과 대학원(공학석사)
 1999~현재 : 한국항공대학교 항공전자공학과 대학원 박사과정
 관심분야 : 정지영상 및 동영상 압축 기법, 멀티미디어 신호처리, 음성부호기, VoIP