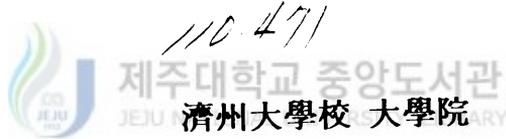


碩士學位論文

효율적인 음성 부호화기 설계를 위한
개선된 LP필터에 관한 연구



通信工學科

金 香 辰

2000年 12月

효율적인 음성 부호화기 설계를 위한 개선된 LP필터에 관한 연구

指導教授 林 載 允

金 香 辰

이 論文을 工學 碩士學位 論文으로 提出함



제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY
2000年 12月

金香辰의 工學 碩士學位 論文을 認准함

審査委員長 李 鎔 鶴 印

委 員 梁 斗 榮 印

委 員 林 載 允 印

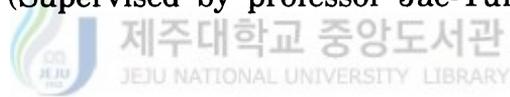
濟州大學校 大學院

2000年 12月

A study on the revised LP filter for the effective voice CODEC design

Hyang-Jin Kim

(Supervised by professor Jae-Yun Lim)



A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF ENGINEERING

DEPARTMENT OF TELECOMMUNICATION ENGINEERING
GRADUATE SCHOOL
CHEJU NATIONAL UNIVERSITY

2000. 12.

< 목 차 >

Abstract	1
I. 서 론	2
II. 음성 생성 과정 및 모델링	4
1. 음성의 선형 예측 부호화	5
2. 자기 상관(autocorrelation) 방법	7
III. 개선된 LP 필터 설계	11
1. 음성 부호화기(CODEC)	11
2. LP 보간의 대역폭 확장	13
3. LSF(Line Spectrum Frequency)	16
1) LPC에서 LSF로의 변환	16
2) LSF에서 LPC로의 변환	22
IV. 실험 및 고찰	25
V. 결 론	37
참고 문헌	38

Abstract

The LP filter is the important element of the speech CODEC. This filter is what human's vocal tract is modeled, and is constructed by coefficients as LPC.

This paper studies about the method which is progressing the CODEC performance.

First, this paper treats of the linear prediction interpolation that is the method of the bandwidth expanding, progressing the synthesised voice quality at the CODEC.

The one following, this paper proposes the LSF or LSP that is the method of using to transfer the speech parameters after processed the speech to LPC, which is digital coding transferring efficiently. The LPC parameters which represent short-term speech spectra, are not very efficient for quantisation because of large dynamic range and filter instability problems. But The LSF parameters provide very good performance in terms of both quantisation and filter stability preservation, and the LSF are currently one of the most efficient choices of transmission parameters for the LPC. The new revised transform algorithms between LSF and LPC coefficients are proposed.

I. 서 론

디지털 기술이 발달하고 인터넷의 급속한 성장과 더불어 고기능 디지털 시스템이 저렴하게 생산됨에 따라 최소의 채널 용량, 최대의 음질, 최소의 비용으로 음성을 전송하는 효율적인 음성의 디지털 코딩 및 전송기법에 많은 관심이 집중되고 있다.

디지털 부호화는 아날로그 신호를 디지털 신호로 부호화 하는 과정이다. 디지털 신호의 음질은 저장 매체에 독립적이고, 데이터가 단순한 이진수의 나열이므로 음질의 저하 없이 무한히 복사가 가능하며, 오류 정정 기술을 사용하면 아날로그 신호에 비해 전송상의 잡음 영향이 적고, 간섭이나 잡음에 의한 데이터의 손실이 적다는 크나큰 이점이 있다. 그러나, 디지털 신호는 아날로그 신호에 비해 표현정도가 클수록 데이터 량이 증가하므로 효율적인 디지털 신호의 처리를 위해서는 데이터 압축이 필수적이다.(Woodard, 1994)

이런 데이터 압축 중에 아날로그 음성 신호를 그대로 코딩하는 넓은 의미로의 음성 압축인 파형 부호화가 있다. 이 파형 부호화는 분석이 용이하고 합성 시 음질도 뛰어나다는 장점이 있지만, 음성 신호에 대한 가정이 없고 음성 신호의 특성을 사용하지 않아 많은 압축을 할 수 없어 비트율이 크기 때문에, 이를 음성 전송 시 사용하기에는 비효율적이다. 이와는 달리 좁은 의미로의 음성 압축으로 보이스 코더인 보코더는 음성 신호가 가지고 있는 특정 파라미터를 추출하여 디지털 신호로 부호화 한다. 이렇듯 파라미터를 추출하여 부호화 하기에 많은 압축이 가능해서 비트율이 작고,(Degroat, 1993) 전송된 파라미터 값을 수신 시스템에서 합성하기 때문에 파형 부호화에 비해 전송 시간과 전송 데이터 량을 크게 줄일 수 있다. 따라서 음성 신호를 분석 시, 전송 시스템에는 보코더가 적합하다.

보코더는 음성의 특성을 추출하여 진행하는 안으로, 물리적으로 사람의 성도에 해당하는 모델링 필터의 파라미터를 측정하고, 성도를 흐르는 공기인 여기신호 파라미터를 측정해 넣으로써 이루어진다. 여기에서 전송되는 파라미터 중 여기 신호

의 양자화 방법에 따라 보코더가 구분된다.

보코더중 선형 예측 부호화(Linear Predictive Coding) 방식은 음성 신호 특성을 전 극점 모델링 필터를 구성하는 예측 계수와 이 모델링 필터에 가해지는 여기 신호를 이득 상수, 그리고 피치 주기로 분석하고 합성 시 이 값들을 이용한다.(Atal, 1982) 효과적으로 압축을 할 수 있으나 예측 계수 값의 동적 영역이 크기 때문에 이를 양자화할 경우에는 많은 비트 율을 필요로 하게된다. 이 동적 영역의 제한을 위한 하나의 방법으로 LSF(Line Spectrum Frequency)기법이 사용된다. LSF는 선형 예측 필터 시스템 함수에서 비 선형적인 가공에 의해 양자화하여 전송하는 파라미터의 동적 영역을 작은 영역으로 제한하여 효율적으로 전송하는 기법이다. 주파수 영역으로 바꾼 LSF의 동적 영역은 $0 \leq \omega \leq \pi$ 로 한정되어 있다. 전송하여야 하는 값의 동적 영역이 작은 범위에서 한정적이라는 것은 그 만큼 작은 영역에서 양자화가 이루어져 적은 채널 용량으로 음성 파라미터 전송이 이루어짐을 의미한다. LSF는 음성의 선형 예측 코딩에 사용되는 분석 및 합성에서 대안의 파라미터이다.(Chung-Hsien Wu, 1984)(G. S. Kang, 1985)

분석을 위한 시스템 다항식은 단위 원 위에서 서로 엇갈려있는 근을 갖는 두 개의 짝수 차수의 대칭 다항식으로 변환되고, 단위 원 위의 이들 각 다항식의 응답은 체비세프 다항식의 확장으로 표현된다. LSF는 실 함수의 실근을 찾는 알고리즘에 의해 구해진 서로 엇갈려있는 근을 이용하여 찾을 수 있다. 방법은 매우 엄밀하며, 삼각 테이블의 저장 또는 삼각 함수의 계산을 피하는 형태로 사용할 수 있다. 예측 계수로의 LSF 복원은 체비세프의 확장인 근 인자의 표현을 통하여 유도된 효율적인 알고리즘을 사용한다.(Fox, 1968)(Kabal, 1986)

그리고 전송 음성의 음질 향상을 위하여, 예측 계수에 $0 < \gamma < 1$ 인 가중치 인자 γ 사용으로 z -평면의 원점에서 방사 적인 극점 이동을 도모하여 대역폭 확장을 시행함으로써, LP(Linear Prediction) 필터에서 신호 에너지를 더욱 많이 분석해 내는 시스템을 구성한다.(Kabal, 1988)

본 논문에서는 대역폭 확장 시스템 구성 및 직접 형태의 예측 계수와 LSF와의 변환 기법에 대한 개선된 방법을 제안한다.

II. 음성 생성 과정 및 모델링

음성이란 사람의 음성 기관에서 나타나는 발음으로 의미의 전달을 내포하고 있는 것을 말한다. 음성은 성대의 진동으로 발생하는 맥동적 공기의 흐름과 구강 안을 좁혀서 발생하는 잡음성의 공기 흐름이 성도라고 하는 음향 관을 전파하여 음향 관의 형상에 따라 결정되는 공명 특성에 의한 변화를 받아 공간으로 방사하는 것이다. 이 중에서 성대의 진동을 수반하는 것을 유성음, 수반하지 않은 것을 무성음이라 한다.(Douglas, 1990)

유성음 중 단음을 주파수 분석하였을 때 공진으로 인하여 특정한 주파수 영역에 에너지 산이 생기는 것을 볼 수 있는데 이를 포만트(formant)라 하고, 이 영역의 중심 주파수를 포만트 주파수라 한다. 또, 무성음과 달리 유성음의 파형은 거의 주기적인 성질을 갖고 있는데 이 주기를 피치(pitch)라고 한다.

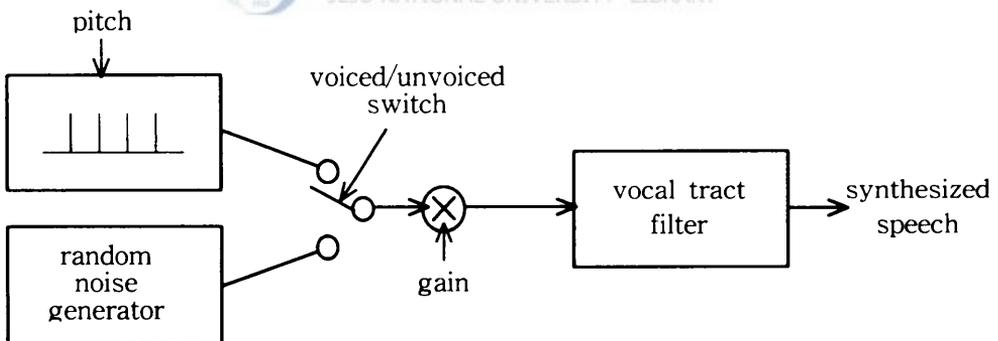


Fig. 1. speech generator model

앞에서 설명한 음성 생성 과정은 일반적으로 Fig. 1과 같이 모델링된다. 성도는 시변 선형 필터로 모델링되고 성도를 지나는 공기 흐름은 성도 필터(vocal tract filter)의 여기 신호(excitation signal)로 모델링된다. 일반적으로 음성 부호화기에서는 시변 선형 필터로 선형 예측 모델을 가장 많이 사용하는데, 선형 예측 필터 계수들은 선형 예측 분석을 통하여 음성 신호로부터 직접 구해지고 성도의 모

양이 시간에 따라 변화하므로 필터 계수들도 주기적으로 갱신되어야 한다. 그리고, 이런 선형 예측 모델이 여러 가지 여기 신호 모델과 조합되어 서로 다른 형태의 음성 부호화를 이루게 된다. Fig. 1에 나타나 있는 여기 신호 모델은 보코더에서 사용되는 것으로, 유성음인 경우는 피치를 주기로 하는 임펄스 열을 여기신호로 사용하고 무성음인 경우는 불규칙 잡음을 여기 신호로 사용한다.

1. 음성의 선형 예측 부호화

선형 예측은 어떤 시스템의 과거 몇 개의 입력과 출력을 가지고 출력을 예측하는 것을 말한다.

$$\hat{y}(n) = \sum_{j=0}^L b(j)x(n-j) - \sum_{i=1}^L a(i)y(n-i) \quad (1)$$

여기에서, $x(n)$ 은 입력, $y(n)$ 은 출력, $\hat{y}(n)$ 은 예측치를 의미하며, $a(i)$, $b(j)$ 는 입력과 출력의 가중치를 의미한다. 즉, 과거의 입력과 출력에 가중치를 주어, 예측을 하는 방법을 이용한다. 이는 어느 정도의 과거가 현재의 출력에 영향을 주는가를 알려주는 인자이다.

다음으로 과거의 입력은 출력에 아무 영향을 주지 않는 시스템을 고려해볼 수 있는데, 보통 선형 예측기가 이에 속한다.

$$\hat{y}(n) = - \sum_{i=1}^L a(i)y(n-i) \quad (2)$$

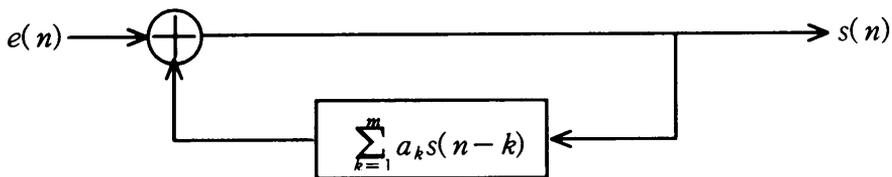


Fig. 2. LP filter

Fig. 2에 사람의 성도를 모델링하는데 가장 많이 쓰이는 LP 필터를 나타내었다. 이 그림에서 예측치와 예측 에러의 결합으로 원 신호인 음성 $s(n)$ 이 생성됨을 알 수 있다. 즉, 음성 $s(n)$ 은 m 개의 과거 값들의 선형 결합으로 예측되고 이 때의 예측 오차 $e(n)$ 은

$$e(n) = s(n) - \sum_{k=1}^m a_k s(n-k) \quad (3)$$

이고, 총 평균 제곱 오차는

$$E = \sum_n [e(n)]^2 = \sum_n \left[s(n) - \sum_{k=1}^m a_k s(n-k) \right]^2 \quad (4)$$

이다. 여기서 m 은 LP필터의 차수를 나타내고, a_k 는 LP 필터의 예측계수 LPC(Linear Prediction Coefficients)로서 E 를 최소화하도록 결정된다. E 를 최소화하는 최적의 LPC를 구하는 방법에는 자기 상관 방법과 공분산 방법과 Lattice 방법이 있다. (Douglas, 1990) 자기 상관 방법은 예측 계수를 얻는 과정에서 생기는 반사 계수(reflection coefficients)가 필터의 안정성을 보장할 수 있지만, 입력 신호에 원도잉을 취해야 하므로 신호의 해상도가 떨어지게 된다는 단점이 있다. 또, 공분산 방법은 입력 신호에 원도잉을 취하지 않아도 되지만 필터의 안정성을 보장할 수 없다. Lattice 방법은 입력 신호에 원도잉을 취하지 않고 필터의 안정성을 보장해 주지만 필터 계수를 구하기 위해서는 순방향 오차와 역 방향 오차를 계산해야 하므로 많은 계산량을 필요로 한다. 음성 부호화기인 코덱(CODEC)에서의 많은 표준안들은 계산량이 많지 않으면서 필터의 안정성을 보장하는 자기 상관 방법을 채택하고 있다.

2. 자기 상관(autocorrelation) 방법

예측기를 실현하기 위해 식 (4)의 E 를 최소화하는 최적의 예측 계수를 계산해야 한다. 이 최적의 예측계수를 찾기 위해 직교 원리(Orthogonality principle)를 이용한다.(Parson, 1987) 이 직교 원리는 $s(n-1), s(n-2), \dots, s(n-m)$ 샘플에서 직교 에러를 구하는 계수들을 설명한다. 식 (4)의 E 를 최소화하는 최적의 예측 계수를 계산하기 위해 직교 원리를 이용하여 다음 식을 얻을 수 있다.

$$\sum_{i=0}^m a(i)r_{i-j} = 0 \quad j = 1, 2, \dots, m \quad (5.a)$$

$$E = \sum_{i=0}^m a(i)r_i \quad (5.b)$$

$$r_i = \sum_{n=-\infty}^{\infty} s(n)s(n-i), \quad \text{windowed } r_i = w(n) \cdot \sum_{n=-\infty}^{\infty} s(n)s(n-i) \quad (5.c)$$

여기에서 본 논문에서 다루는 음성은 8 kHz로 샘플링하고, 한 프레임을 240샘플로 하기에, 다음과 같이 정의된 30ms 해밍 윈도우를 사용한다.

$$w(n) = \begin{cases} 0.54 - 0.46 \cos\left(2\pi \frac{n}{N-1}\right) & ; 0 \leq n \leq N-1 \\ 0 & ; \text{otherwise} \end{cases} \quad (6)$$

식 (5)의 해 예측 계수 $a(i)$ 를 구하기 위해 Durbin 알고리즘이라 하는 다음과 같은 단계를 거친다.

먼저 식 (5)을 다음 행렬식으로 전개한다.

$$\begin{bmatrix} r_1 & r_0 & r_1 & r_2 & \cdots & r_{m-1} \\ r_2 & r_1 & r_0 & r_1 & \cdots & r_{m-2} \\ r_3 & r_2 & r_1 & r_0 & \cdots & r_{m-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_m & r_{m-1} & r_{m-2} & r_{m-3} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ a(2) \\ \cdots \\ a(m) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \quad (7)$$

$$\begin{bmatrix} r_0 & r_1 & r_2 & \cdots & r_{m-1} \\ r_1 & r_0 & r_1 & \cdots & r_{m-2} \\ r_2 & r_1 & r_0 & \cdots & r_{m-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ r_{m-1} & r_{m-2} & r_{m-3} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ a(3) \\ \cdots \\ a(m) \end{bmatrix} = - \begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ \cdots \\ r_m \end{bmatrix} \quad (8)$$

다음으로 식 (8)에 보조 방정식으로 예측 에러 공식을 삽입한다.

prediction - error formula : $\sum_{i=0}^m a(i)r(i) = E$

$$\begin{bmatrix} r_0 & r_1 & r_2 & r_3 & \cdots & r_m \\ r_1 & r_0 & r_1 & r_2 & \cdots & r_{m-1} \\ r_2 & r_1 & r_0 & r_1 & \cdots & r_{m-2} \\ r_3 & r_2 & r_1 & r_0 & \cdots & r_{m-3} \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ r_m & r_{m-1} & r_{m-2} & r_{m-3} & \cdots & r_0 \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ a(2) \\ a(3) \\ \cdots \\ a(m) \end{bmatrix} = \begin{bmatrix} E \\ 0 \\ 0 \\ 0 \\ \cdots \\ 0 \end{bmatrix} \quad (9)$$

결국 식 (9)의 해 $a(i)$ 를 구하는 것으로 전개된다. 이 식 (9)을 풀기 위해 다음과 같이 순차적으로 진행한다.

i) $i = 0$ 일 때

$$[r_0][a_0] = [E_0] \quad (10)$$

ii) $i = 1$ 일 때

$$\begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} E_1 \\ 0 \end{bmatrix} \quad (11.a)$$

$$\begin{bmatrix} r_0 & r_1 \\ r_1 & r_0 \end{bmatrix} \left\{ \begin{bmatrix} a_0 \\ 0 \end{bmatrix} + k_1 \begin{bmatrix} 0 \\ a_0 \end{bmatrix} \right\} = \begin{bmatrix} E_0 \\ q \end{bmatrix} + k_1 \begin{bmatrix} q \\ E_0 \end{bmatrix} \quad (11.b)$$

$$r_1 a_0 + k_1 r_0 a_0 = q + k_1 E_0 \quad (11.c)$$

$$\therefore k_1 = -\frac{q}{E_0} = -\frac{1}{E_0} r_1 a_0, \quad E_1 = E_0(1 - k_1^2) \quad (11.d)$$

여기에서 q 는 임의의 변수이다.

iii) $i = 2$ 일 때



$$\begin{bmatrix} r_0 & r_1 & r_2 \\ r_1 & r_0 & r_1 \\ r_2 & r_1 & r_0 \end{bmatrix} \left\{ \begin{bmatrix} a_0 \\ a_1 \\ 0 \end{bmatrix} + k_2 \begin{bmatrix} 0 \\ a_1 \\ a_0 \end{bmatrix} \right\} = \begin{bmatrix} E_1 \\ 0 \\ p \end{bmatrix} + k_2 \begin{bmatrix} q \\ 0 \\ E_1 \end{bmatrix} = \begin{bmatrix} E_2 \\ 0 \\ 0 \end{bmatrix} \quad (12.a)$$

$$r_2 a_0 + r_1(a_1 + k_2 a_1) + k_2 r_0 a_0 = q + k_2 E_1 \quad (12.b)$$

$$r_2 a_0 + r_1 a_1 + k_2(r_1 a_1 + r_0 a_0) = q + k_2 E_1 \quad (12.c)$$

$$\therefore k_2 = -\frac{q}{E_1} = -\frac{1}{E_1} (r_2 a_0 + r_1 a_1), \quad E_2 = E_1(1 - k_2^2) \quad (12.d)$$

iv) $i = 3$ 일 때

$$\begin{bmatrix} r_0 & r_1 & r_2 & r_3 \\ r_1 & r_0 & r_1 & r_2 \\ r_2 & r_1 & r_0 & r_1 \\ r_3 & r_2 & r_1 & r_0 \end{bmatrix} \left\{ \begin{bmatrix} a_0 \\ a_1 \\ a_2 \\ 0 \end{bmatrix} + k_3 \begin{bmatrix} 0 \\ a_2 \\ a_1 \\ a_0 \end{bmatrix} \right\} = \begin{bmatrix} E_2 \\ 0 \\ 0 \\ q \end{bmatrix} + k_3 \begin{bmatrix} q \\ 0 \\ 0 \\ E_2 \end{bmatrix} = \begin{bmatrix} E_3 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad (13.a)$$

$$r_3 a_0 + r_2(a_1 + k_3 a_2) + r_1(a_2 + k_3 a_1) + k_3 r_0 a_0 = q + k_3 E_2 \quad (13.b)$$

$$r_3 a_0 + r_2 a_1 + r_1 a_2 + k_3(r_2 a_2 + r_1 a_1 + r_0 a_0) = q + k_3 E_2 \quad (13.c)$$

$$\therefore k_3 = -\frac{q}{E_2} = -\frac{1}{E_2} \sum_{i=0}^2 a_i r_{3-i}, \quad E_3 = E_2(1 - k_3^2) \quad (13.d)$$

모든 단계를 거침으로 다음과 같은 일반식을 얻을 수 있다.

$$\begin{bmatrix} a_n(0) \\ a_n(1) \\ a_n(2) \\ \dots \\ a_n(n-1) \\ a_n(n) \end{bmatrix} = \begin{bmatrix} a_{n-1}(0) \\ a_{n-1}(1) \\ a_{n-1}(2) \\ \dots \\ a_{n-1}(n-1) \\ 0 \end{bmatrix} + k_n \begin{bmatrix} 0 \\ a_{n-1}(n-1) \\ a_{n-1}(n-2) \\ \dots \\ a_{n-1}(1) \\ a_{n-1}(0) \end{bmatrix} \quad (14.a)$$

$$a_n(i) = a_{n-1}(i) + k_n a_{n-1}(n-i) \quad (14.b)$$

$$k_n = \frac{-1}{E_{n-1}} \sum_{i=0}^{n-1} a_{n-1}(i) r_{n-i} \quad (14.c)$$

$$E_n = E_{n-1}(1 - k_n^2) \quad (14.d)$$

여기서 a_n 은 이 알고리즘을 통해 구하고자 하는 예측 계수이며, k_n 은 LP 필터의 반사 계수로서 그 크기가 1보다 작아 필터의 안정성을 보장해 준다. LP 필터의 계수가 많아지므로 해서 식 (14.d)의 예측 에러 E_n 이 작아진다. 즉, 예측기의 성능이 더욱 좋아짐을 의미한다. 그러나 음성 압축 시 대부분이 $n=10$ 으로 하는 10차 필터가 널리 쓰이고 있으며, 본 논문에서도 이를 사용한다.

III. 개선된 LP 필터 설계

1. 음성 부호화기(CODEC)

분석·합성 구조로 음성 부호화기 블록도를 Fig. 3에 나타내었다. 여기 신호 $u(n)$ 은 스케일링된 적응 코드북 신호(유성음 구간동안 Long-term 예측이 더해진)와 크기가 방대한 고정 코드북의 스케일링된 신호와의 합성으로 주어진다. 이 여기 신호는 성도효과를 모델링한 합성 필터를 운용하기 위해 사용된다. 디코더에서 여기 신호는 음성 신호 $\hat{s}(n)$ 을 생성하기 위해 합성 필터를 거쳐 통과된다. 전형적으로 필터의 파라미터를 먼저 측정하고, 코드북 색인들 a 와 k 계다가 이득 G_1 과 G_2 을 찾는다.

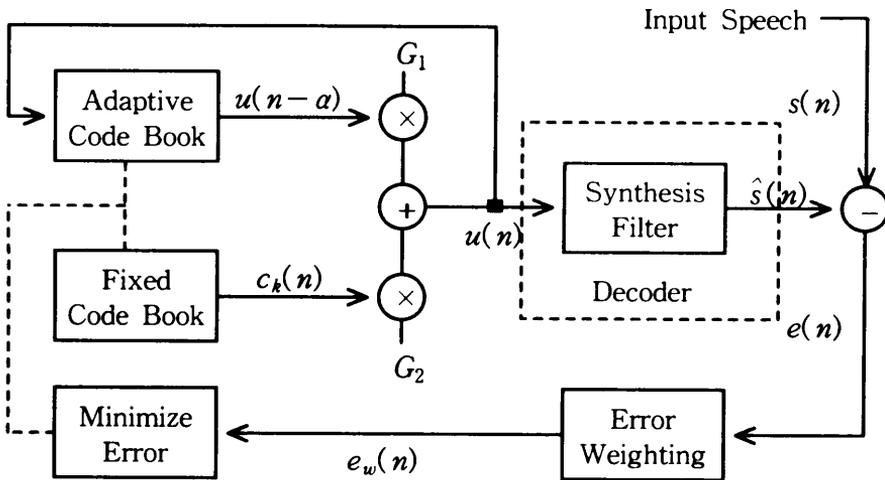


Fig. 3. CODEC encoder block

이런 구조의 음성 부호화기는 CELP(Code Excited Linear Prediction) 방법으로 서 많은 연구가 이루어지고 있다.(Mahesan, 1990) 이 CELP에서 코드북 파라미

터들은 다시 합성된 것과 원 음성 신호와의 가중 에러를 최소화하도록 선택되어진다. 효율적으로 가능한 각 코드북 엔트리는 인지할 수 있게 가중된 민감도에서 입력 신호에 가장 가까운 출력이 주어지는가를 테스트하기 위해 합성 필터를 거치게 된다. 이 방대한 폐 루프 구조는 가능한 한 원 음성에 가깝게 신호를 다시 생성해 내는 수행을 하기 위해 사용된다.

그러나 이 부호화기에서 엄격한 폐 루프접근의 두 가지 예외가 있다. 첫 번째로 예측 잉여 에러의 에너지를 최소화하는 Short-term 선형 예측 에러 필터의 단순한 역인 합성 필터의 결정에서 있다. 이는 여기 신호 $u(n)$ 이 합성 필터의 형태를 고려하고 운용될지라도, 합성 필터 파라미터가 결정되어질 때 여기 신호의 형태가 고려되지 않는다. 이것은 결함이고, 합성 필터가 모델링하기에 더 낮게 적응 코드북을 들 수 있는 Long-term 예측들을 고려 시도할 수 있을지도 모른다는 예를 의미한다.

엄격한 폐 루프접근에서의 두 번째 예외는 코드 파라미터들의 결정에서 있다. 오히려 적응 및 고정 코드북 파라미터들이 가중된 에러 신호에서 총괄적인 최소화를 수행하기 위해 함께 결정되기보다는, 먼저 고정 코드북 신호가 '영'이라 추정된 상태에서 적응 코드북 지연과 이득이 결정된다. 그런 후에 적응 코드북 신호가 주어지고 고정 코드북 파라미터들을 찾는다. 이 부분에서의 최소화 접근은 CELP 코덱의 복잡성을 적당한 레벨로 줄이기 위해 채택된다. 그러나 음성의 재생성에서 어떤 저하가 생긴다는 것이 명백하다.

이렇게 이런 부호화기 설계에서의 파라미터 계산은 Short-term 선형 예측, Long-term 적응 코드북 탐색과 고정 코드북 탐색이라는 세 가지 기초 함수에 의하여 수행된다. 세 가지 기초 함수의 수행으로 계산된 LPC, 적응 코드북 지연과 이득, 고정 코드북의 인덱스와 이득의 파라미터를 양자화해서 전송하거나 저장해 두어, 수신 측이나 복원 시에 이들을 이용하여 음성을 합성해낸다.

2. LP 보간의 대역폭 확장

Fig. 3에서의 합성 필터(Synthesis Filter)는 다음과 같은 시스템 함수로 표기되는 LP 필터로 구성된다.

$$H(z) = \frac{1}{A(z)} = \frac{1}{\sum_{k=0}^m a(k)z^{-k}} \quad (15)$$

이런 LP 필터에서 LP 분석 계수인 $a(k)$, 대신 $a(k)\gamma^k$ 로 대치를 통하여 음성 분석 시 대역폭 확장을 시행한다. 이 시행은 $0 < \gamma < 1$ 인 가중치 인자 γ 에 의해 z -평면의 원점에서 방사 적으로 극점을 이동시킴으로 이루어진다. 여기에서 $a(0) = 1.0$ 이다. 이 확장된 계수들은 LP 필터 시스템 함수 $1/A(z)$ 로 재 정의된다. 15Hz확장을 위해 다음 식 (16)에 의하여 $\gamma = 0.994127$ 을 이용한다.(Tohkura, 1987)

$$\text{bandwidth expand} = -f_s/\pi \ln \gamma [\text{Hz}] \quad (16)$$

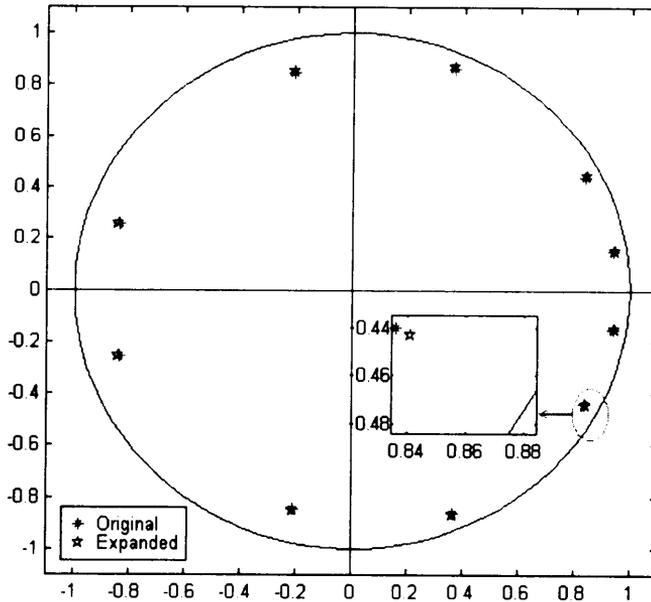


Fig. 4. The original roots and the expanded roots

예측 계수들은 프레임 당 하나 뿐만 아니라 각 서브 프레임마다 필요하기 때문에 네 서브 프레임의 각각에 중간 설정 형태로 선형적으로 보간 된다.

이런 대역폭 확장은 다음 Fig. 4에서 확인할 수 있듯이 시스템 함수의 극점들을 방사 적으로 원점에서 단위 원 쪽으로 이동시킨다. 그리하여 LP 필터가 신호의 에너지를 더욱 많이 포함해내는 분석을 할 수 있도록 보간 된다.

즉, LP 필터의 예측이 더욱 많은 신호의 에너지를 포함하고 있다는 의미인데 이는 다음 Fig. 5에서 확인할 수 있다.

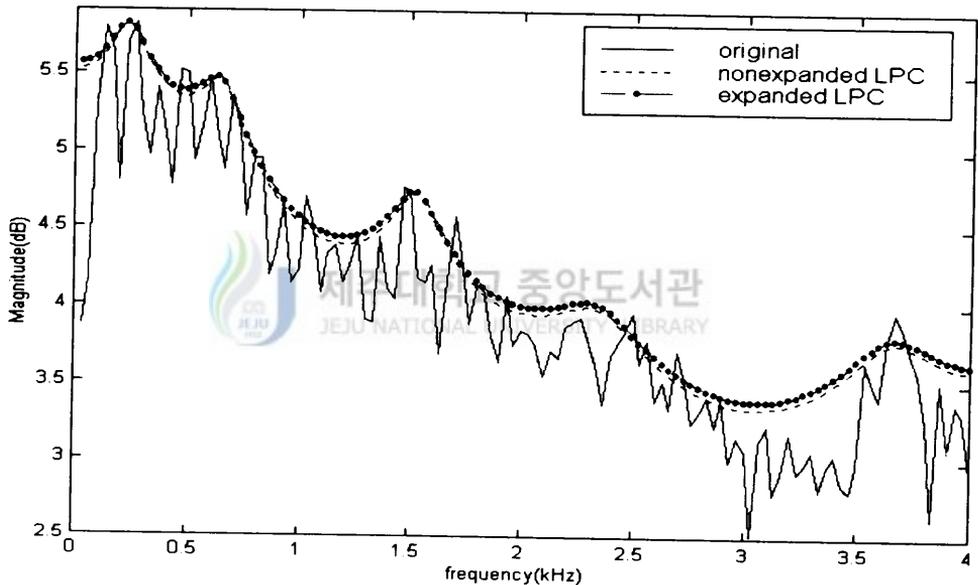


Fig. 5. Bandwidth Expanding

Fig. 4, 5와 같이 대역폭을 확장함으로써 LP 필터가 신호의 에너지를 더욱 많이 포함하는 형태가 되어 여기 신호로 양자화되는 잉여 성분이 더욱 작아진다. 그리하여 Fig. 6과 Fig. 7에 나타내었듯이 효율적인 적응 코드북 탐색을 유도하여 고정 코드북 탐색에서의 타겟 잉여 성분이 더욱 작아져 고정 코드북 탐색에서의 효율을 보장한다. 합성 필터에 가하는 여기 신호를 코드북의 인덱스와 이득으로 파라미터화하는 분석 필터의 잉여성분에 더욱 더 접근시킬 수 있다. 즉, 더욱 양질의 음성을 보장할 수 있음을 의미한다.

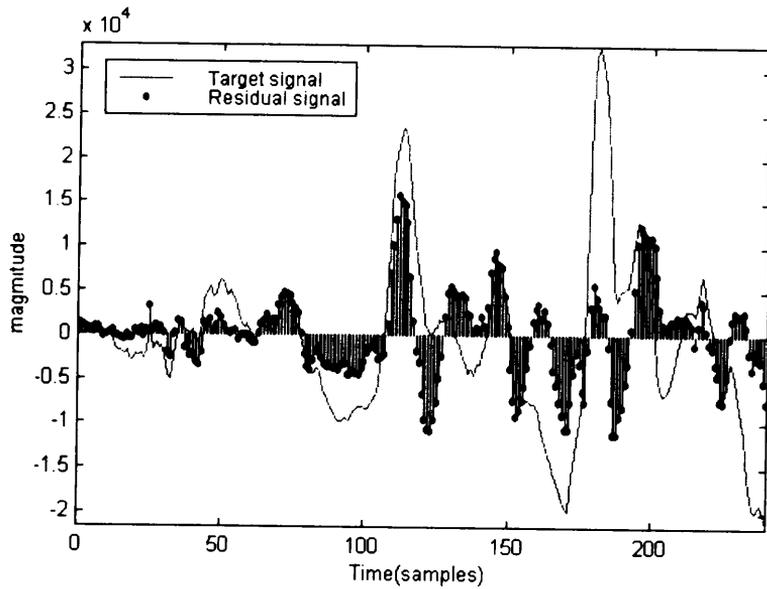


Fig. 6. Codebook search by nonexpanded

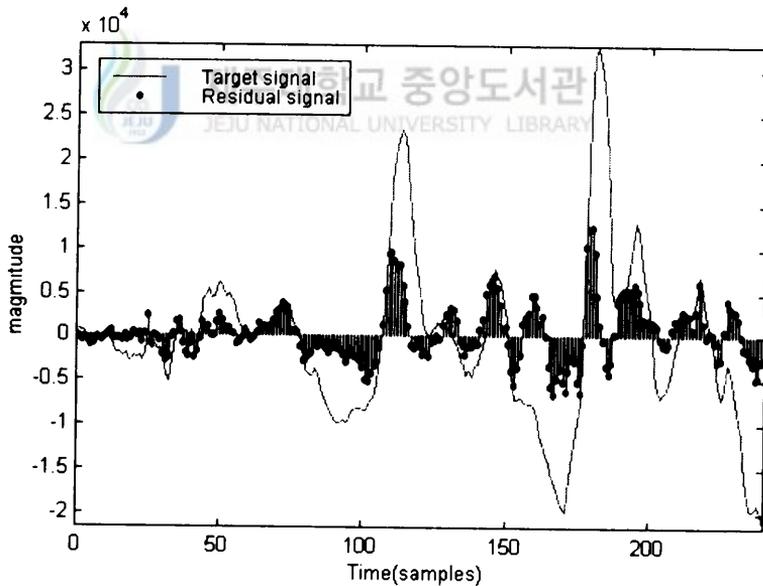


Fig. 7. Codebook search by expanded

게다가 음성을 향상시키기 위한 이 대역폭 확장은 다음절에서 설명될 LSF 양자화와 양자화된 LSF들에서 직접적으로 예측 계수들로 변환하는 고속 탐색 방법을 효율적으로 할 수 있게 해준다.

3. LSF(Line Spectrum Frequency)

LPC 값의 동적 영역이 크기 때문에 저장이나 전송을 위하여 LPC를 양자화할 경우에는 많은 비트율이 필요하게된다. 음성 신호를 분석할 때 분석 데이터가 클 경우 전송 시간 소요도 클 뿐 아니라 수신단의 데이터 저장 매체도 대용량이 되어야 하기 때문에 분석 데이터가 적을수록 전송면에서 유리하다. 따라서 전송시스템에서는 LPC 동적 영역의 제한을 위한 하나의 방법으로 LSF 기법이 사용된다. LSF는 LPC 필터 시스템 함수에서 비 선형적인 가공에 의해 양자화하여, 전송하는 파라미터의 동적 영역을 작은 영역으로 제한하여 효율적으로 전송하는 기법이다. 주파수 영역으로 바꾼 LSF의 동적 영역은 $0 \leq \omega \leq \pi$ 로 한정되어 있다. 전송하여야 하는 값의 동적 영역이 작은 범위에서 한정적이라는 것은 그 만큼 작은 영역에서 양자화가 이루어진다. 그리고, LPC는 그 동적 영역이 크기 때문에 이를 양자화할 경우 데이터의 왜곡이 우려되나, 반면에 LSF는 그 동적이 정해져 있으므로 양자화로 인한 데이터의 왜곡이 심각하지 않다. 이렇게 함으로써 적은 채널 용량으로 음성 파라미터 전송이 가능케 된다. LSF는 음성의 선형 예측 코딩에 사용되는 분석 및 합성에서 대안의 파라미터이다.

1) LPC에서 LSF로의 변환

우선 최소 위상(Minimum Phase) LPC 다항식에서 시작한다.

$$A(z) = \sum_{k=0}^m a(k)z^{-k} \quad (17)$$

여기에서 m 은 LPC 필터의 차수이며, $a(k)$ 는 필터의 계수이다. LPC 필터는 음성음 중 단음을 주파수 분석했을 때 공진으로 인하여 생기는 특정한 주파수 영역에서의 주파수 산인 포먼트(Formant)를 모델링하여 음의 형태를 만들어 내는 역할을 한다. 즉, 사람의 성도를 모델링한 것이다. 그리고, 선형 예측 코딩 다항식은 자기 상관, 공분산, Lattice 등의 방법으로 구해질 수 있다.

다음으로 대칭적(Symmetric)인 다항식과 비대칭적(Antisymmetric)인 다항식으로 구성한다.

$$F_1(z) = A(z) + z^{-(m+1)}A(z^{-1}), \quad F_2(z) = A(z) - z^{-(m+1)}A(z^{-1}) \quad (18)$$

이는 비 선형적인 가공으로 전송 파라미터의 동적 영역을 한정하기 위한 방법으로 가공에 의하여 주파수 영역으로 바뀐 이 보조 다항식의 근은 다음과 같은 특징이 있다.

첫째, 근의 위치가 모두 단위 원상에 존재한다.

둘째, 근은 $z = \pm 1$ 을 제외한 모든 값이 공액 복소 쌍으로 존재한다.

위 특징 중 최소 위상 LPC다항식이라면 근의 위치가 단위 원 상에 존재한다는 것이 명백하다.

여기에서 $F_1(z)$ 는 $z = -1$ 인 근과 $F_2(z)$ 는 $z = 1$ 인 근을 포함하고 있다.

그 다음으로, $z = \pm 1$ 에서의 근을 제거한다. 이는 다항식 (18)의 근이 이 위치를 제외하고는 모두 공액 복소 쌍으로 존재하여 단위 원의 상위 부분만 고려하여도 된다는 이점이 있기 때문이다.

$$G_1(z) = \frac{F_1(z)}{1+z^{-1}} \quad \text{and} \quad G_2(z) = \frac{F_2(z)}{1-z^{-1}} \quad m \text{ even} \quad (19)$$

다항식 $G_1(z), G_2(z)$ 의 근은 LPC의 대안 파라미터 LSF(Line Spectrum Frequency) 또는 LSP(Line Spectrum Pairs)라 불린다.

Fig. 8에서 { 1.00000 -1.20224 0.00223 -0.03548 0.08875 0.19893 0.04769 0.04036 -0.11231 0.04571 -0.01723} LPC 계수로 모델링되었을 때, $A(z)$ 근의 위치를 도시하였다

Fig. 8에서 근의 위치가 모두 단위 원 안에 존재하여 최소 위상임을 알 수 있다.

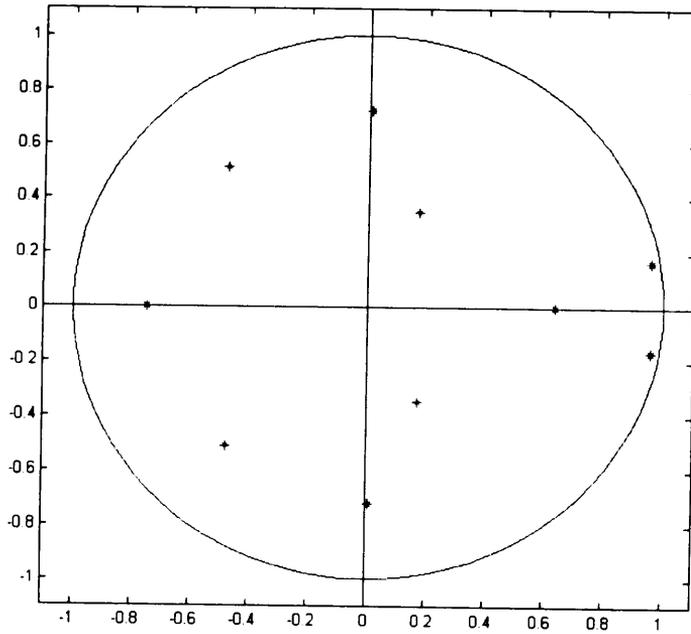


Fig. 8. The root's position of the $A(z)$ polynomials

그리고, Fig. 9에서 $F_1(z), F_2(z)$ 근의 위치를 도시하였다.

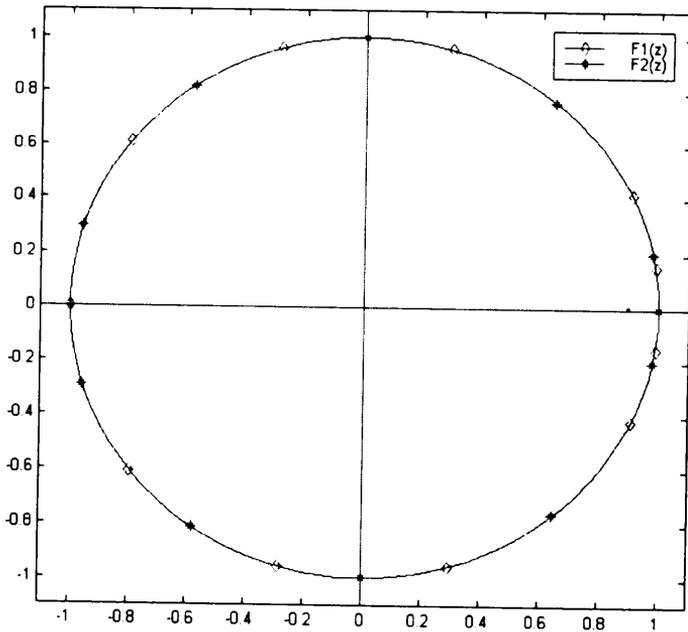


Fig. 9. The root's position of the $F_1(z), F_2(z)$ polynomials

이 그림을 살펴보면, $F_1(z), F_2(z)$ 의 근이 단위 원상에 있으며, 서로 엇갈려 존재한다는 특징을 확인할 수 있다. $F_1(z), F_2(z)$ 에서 $z = \pm 1$ 에서의 근을 제거한 $G_1(z), G_2(z)$ 모두 결과는 짝수 차수의 대칭 다항식이다. 그러므로, 근은 복소 공액 쌍으로 존재하여, 위 반원에 있는 근만 구하여도 해석이 가능하게 된다. 이는 더욱 효율적인 양자화를 가능케 해준다.

식 (18)과 식 (19)를 이용하여 풀어쓰면,

$$\begin{aligned} G_1(z) &= \prod_{i=2,4,\dots,m} (1 - 2z^{-1} \cos \omega_i + z^{-2}), \\ G_2(z) &= \prod_{i=1,\dots,m-1} (1 - 2z^{-1} \cos \omega_i + z^{-2}) \end{aligned} \quad (20)$$

여기에서 $G_1(z), G_2(z)$ 의 근을 구하기 위하여, 근의 위치가 단위 원이라는 점과 체비셰프(Chebyshev) 다항식을 이용하여 x 의 다항식으로 고친다.(Kabal,1986)

$$z = e^{j\omega} \Rightarrow z^n + z^{-n} = 2 \cos n\omega \quad (21)$$

여기에서 $p = m/2$ 이고, 식 (21)을 식 (20)에 적용하면, 다음 식을 얻을 수 있다.

$$G_1(e^{j\omega}) = e^{-j\omega p} G_1'(\omega), \quad G_2(e^{j\omega}) = e^{-j\omega p} G_2'(\omega) \quad (22)$$

$$G_1'(\omega) = 2 \cos p\omega + 2g_1(1) \cos (p-1)\omega + \dots + 2g_1(p-1) \cos \omega + g_1(p) \quad (23.a)$$

$$G_2'(\omega) = 2 \cos p\omega + 2g_2(1) \cos (p-1)\omega + \dots + 2g_2(p-1) \cos \omega + g_2(p) \quad (23.b)$$

이때 체비셰프 다항식은 다음과 같다.

$$x = \cos \omega, \quad T_m(x) = \cos(m \arccos(x)) \quad (24.a)$$

$$T_m(x) = 2xT_{m-1}(x) - T_{m-2}(x) \quad (24.b)$$

식 (24)의 체비셰프 다항식을 식 (23)에 적용하여, 다음 식을 얻는다.

$$G_1'(x) = 2T_p(x) + 2g_1(1)T_{p-1}(x) + \dots + 2g_1(p-1)T_1(x) + g_1(p) \quad (25.a)$$

$$G_2'(x) = 2T_p(x) + 2g_2(1)T_{p-1}(x) + \dots + 2g_2(p-1)T_1(x) + g_2(p) \quad (25.b)$$

식 (25)를 다음과 같이 표현하여, 이를 계속 확장한다.

$$Y(x) = c_0 T_0 + c_1 T_1 + \dots + c_{N-1} T_{N-1} + c_N T_N \quad (26)$$

$$Y(x) = c_0 T_0 + \dots + (c_{N-2} - c_N) T_{N-2} + c_{N-1} T_{N-1} + 2c_N x T_{N-1} \quad (27)$$

$$\begin{aligned} Y(x) &= c_0 T_0 + \dots + (c_{N-3} - c_{N-1}) T_{N-3} + (c_{N-2} - c_N) T_{N-2} \\ &\quad + 2c_{N-1} x T_{N-2} + 2c_N x T_{N-1} \\ &= c_0 T_0 + \dots + (c_{N-4} - 2c_{N-2} + 2c_N) T_{N-4} + (c_{N-3} - c_{N-1}) T_{N-3} \\ &\quad + 2(c_{N-2} - c_N) x T_{N-3} + 2c_{N-1} x T_{N-2} + 2c_N x T_{N-1} \end{aligned} \quad (28)$$

이렇게 함으로써, 식 (29)와 같은 일반식을 얻는다.

$$Y(x) = x Y'(x) + c_0' \quad (29)$$

$$Y(x) = x(x Y''(x) + c_1') + c_0' \quad (30)$$

$$Y(x) = \sum_{k=0}^N c_k' x^k \quad (31)$$

식 (31)식을 식 (25)에 적용하여 $G_1'(x)$, $G_2'(x)$ 를 구한다. 마지막으로, $G_1'(x)$, $G_2'(x)$ 에서 근 x_i 를 계산하여, LSF를 구한다. 이때, LSF $\omega_i = \arccos x_i$ 이다.

Fig. 10에서 $G_1'(x)$, $G_2'(x)$ 근의 위치를 도시하였는데, 살펴보면 근 x_i 의 위치가 $[-1, 1]$ 사이의 값임을 확인 할 수 있고, $G_1'(x)$, $G_2'(x)$ 근이 서로 엇갈려 존재함을 알 수 있다.

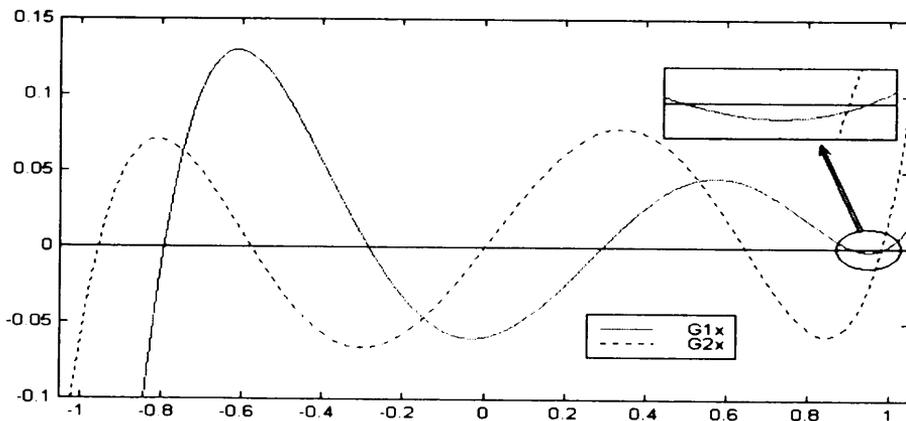


Fig. 10. The root's position of the $G_1'(x)$, $G_2'(x)$ polynomials

이는 앞에서 언급한 $G_1(z), G_2(z)$ 모두 결과가 짝수 차수의 대칭 다항식으로 근인 LSF가 복소 공액 쌍으로 존재한다는 것을 의미한다.

다음으로 본 논문에서 제시한 LPC에서 LSF의 변환 알고리즘에 대한 흐름도를 나타내면 Fig. 11과 같다.

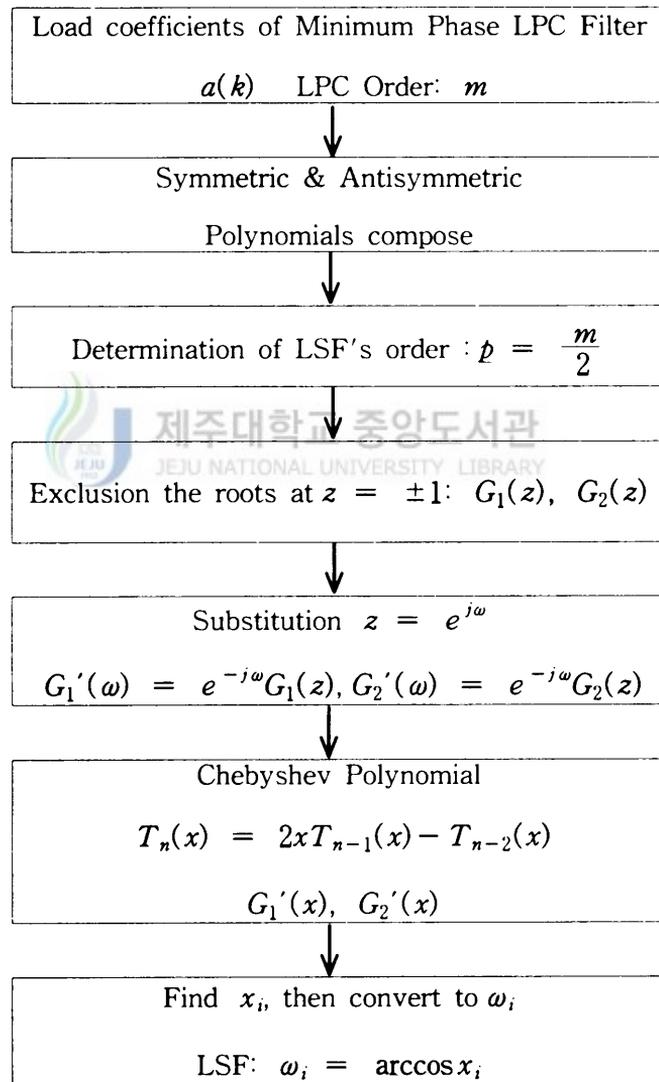


Fig. 11. The Algorithm to transform from LPC to LSF

2) LSF에서 LPC로의 변환

먼저 전송된 LSF ω_i 를 다음과 같이 바꾼다.

$$x_i = \cos \omega_i \quad (32)$$

$$G_1'(x) = \prod_{k=1}^p 2(x - x_{2k}), \quad G_2'(x) = \prod_{k=0}^{p-1} 2(x - x_{2k+1}) \quad (33)$$

식 (33)은 근 x_i 의 인수분해 형태로 표현된 다항식 $G_1'(x), G_2'(x)$ 이다. 이를 확장함으로써 다항식의 계수를 구한다. 이 수행을 하기 위해 근 x_i 를 하나씩 대입 하면서, 식 (34)를 이용하여 체비셰프 다항식의 형태로 바꾼다.

$$Y_{N+1}(x) = 2(x - x_r)Y_N(x) = \sum_{k=-1}^{N+1} [c_{k-1} - 2x_r c_k + c_{k+1}] T_k(x) \quad (34)$$

$T_n(x) = \cos n\omega$ 를 이용하여, cosine의 항목으로 바꾸고, $G_1'(\omega), G_2'(\omega)$ 다항식을 구한다.

$$G_1(\omega) = G_1'(\omega) \times e^{-j\omega p}, \quad G_2(\omega) = G_2'(\omega) \times e^{-j\omega p} \quad (35)$$

그리고, 체비셰프의 정의 식 (32)를 이용하여, $G_1(z), G_2(z)$ 의 계수를 구한다.

$$2 \cos k\omega = z^k + z^{-k} \quad (36)$$

다음으로, $G_1(z), G_2(z)$ 에 $z = \pm 1$ 에서의 근을 포함시켜, $F_1(z), F_2(z)$ 의 계수를 구한다. 그런 후에 다음의 식으로 LPC 다항식 $A(z)$ 를 구한다.

$$A(z) = \frac{1}{2} \{F_1(z) + F_2(z)\} \quad (37)$$

이 알고리즘은 체비셰프 방정식의 특성을 이용하여 Soong and Juang의 알고리즘을 개선한 Kabal and Ramachandran의 알고리즘이다.

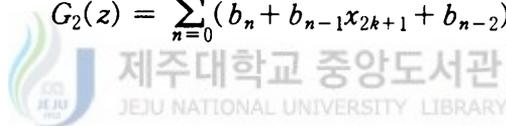
차수가 짝수인 경우에는 체비셰프 방정식의 특성을 이용한 Kabal and Ramachandran의 알고리즘보다 더욱 곱셈 연산과 덧셈 연산을 줄일 수 있도록 개선한 알고리즘을 다음과 같이 제안한다.

x_i 의 인수분해 형태로 표현된 다항식 $G_1'(x), G_2'(x)$ 을 구하지 않고, 식 (20)에서 바로 $G_1(z), G_2(z)$ 를 구한다. 식 (20)을 다시 써보면 다음과 같다.

$$\begin{aligned} G_1(z) &= \prod_{i=2,4,\dots,m} (1 - z^{-1} \cos \omega_i + z^{-2}), \\ G_2(z) &= \prod_{i=1,\dots,m-1} (1 - z^{-1} \cos \omega_i + z^{-2}) \end{aligned} \quad (38)$$

식 (38)을 다음 식으로 풀어 쓸 수 있다.

$$\begin{aligned} G_1(z) &= \sum_{n=1}^p (b_n + b_{n-1}x_{2k} + b_{n-2})z^{-n}, \\ G_2(z) &= \sum_{n=0}^{k-1} (b_n + b_{n-1}x_{2k+1} + b_{n-2})z^{-n} \end{aligned} \quad (39)$$



식 (39)에 x_i 를 직접 대입하여 다항식을 확장하게 되면, $g_1(n) = g_1(m-n)$ 임을 알 수 있다. $g_2(n)$ 도 마찬가지로이다. 그러므로, 계수를 구할 때 LPC 차수 m 까지 구하지 않고, $m/2$ 인 p 까지만 구하고, 이를 앞의 식에 적용시켜 모든 계수를 구할 수 있다.

식 (39)에 x_i 를 직접 대입하여 계수를 구함으로써, Kabal 알고리즘에서 x 의 다항식으로 고친 후에 체비셰프의 다항식으로 변환하는 과정을 한 단계로 줄여, Kabal의 알고리즘에서 곱셈 및 덧셈의 연산 숫자를 개선한다.

개선된 LSF에서 LPC로의 변환 알고리즘을 나타내면 Fig. 12와 같다.

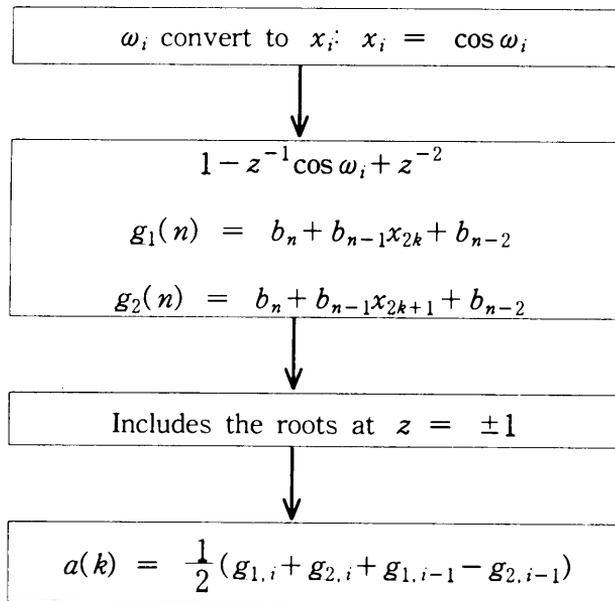


Fig. 12. The Algorithm to transform from LSF to LPC

IV. 실험 및 고찰

CELP 코딩은 샘플링한 입력 신호를 하나의 단위로서 처리되는 샘플의 블록으로 만든 프레임지향의 기술이다. 일반적으로 8kHz로 샘플링한 PCM을 이용하며 한 프레임은 240개 샘플로 이루어지며, 60개의 샘플을 단위로 한 한 프레임 당 4개의 서브 프레임을 두어 음성의 처리를 하는 코딩 방법중 하나이다. CELP 코딩은 analysis-to-synthesis 탐색, 인지적으로 가중된 VQ(Vector Quantization)와 선형 예측에 기초를 두고 있다.

10차 LP 필터는 음성 신호의 Short-term 포먼트 구조를 모델링하며, Long-term LP 또는 피치는 적응 코드북 VQ로 모델링하고, Short-term LP와 피치 VQ로부터 잉여는 고정 코드북으로 양자화하는 벡터이다. 이런 파라미터를 이용하여 음성을 분석 코딩한다.

Short-term 포먼트 구조를 모델링한 LP 필터 계수의 가공으로 필터 극점을 확장하여 탐색의 효율을 도모하고 양질의 음성을 합성하도록 하였는데, 이 때 대역폭의 확장을 위하여 LP 필터 계수인 a_i , 대신 $a_i\gamma^i$ 로 대체에서 $\gamma = 0.994127$ 을 사용하였다.

다음 Fig. 13은 모의 실험에 사용된 “통신공학과입니다”의 남성 음성이다.

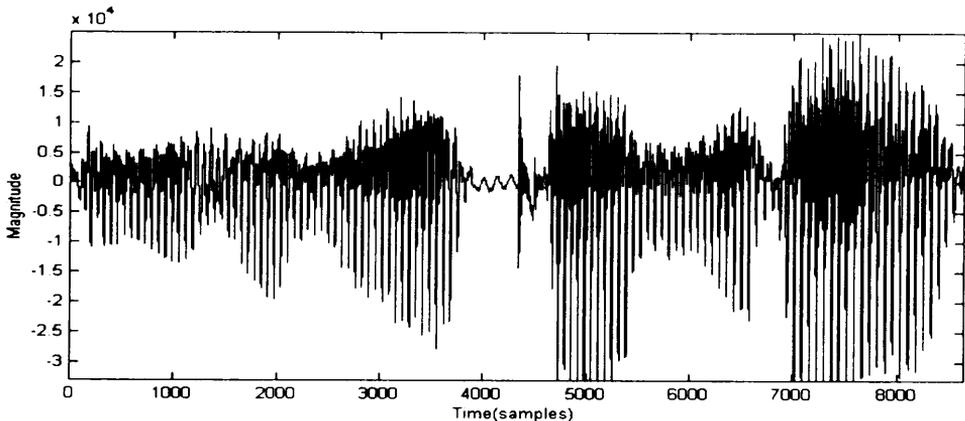


Fig. 13. The original speech signal

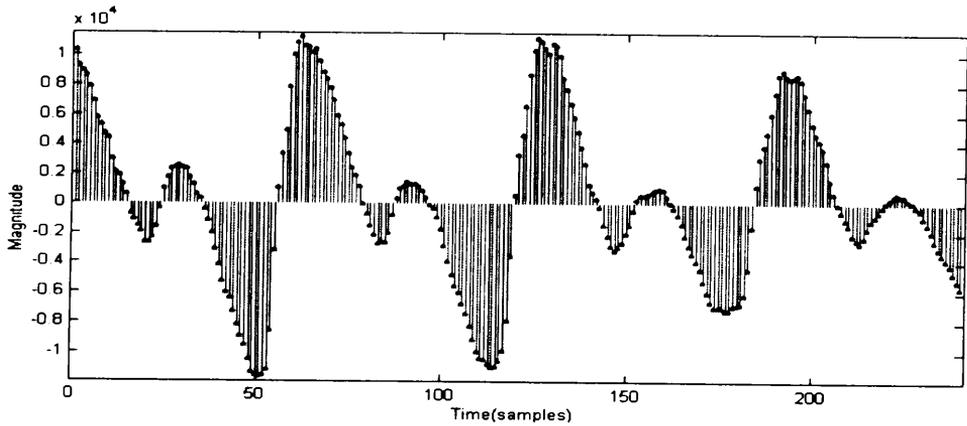


Fig. 14. The 9'th frame of the original speech signal

Fig. 14는 실례에 쓰인 아홉 번째 프레임의 신호로, Fig. 15에 주파수 분석으로 LPC와의 관계를 도시하였다.

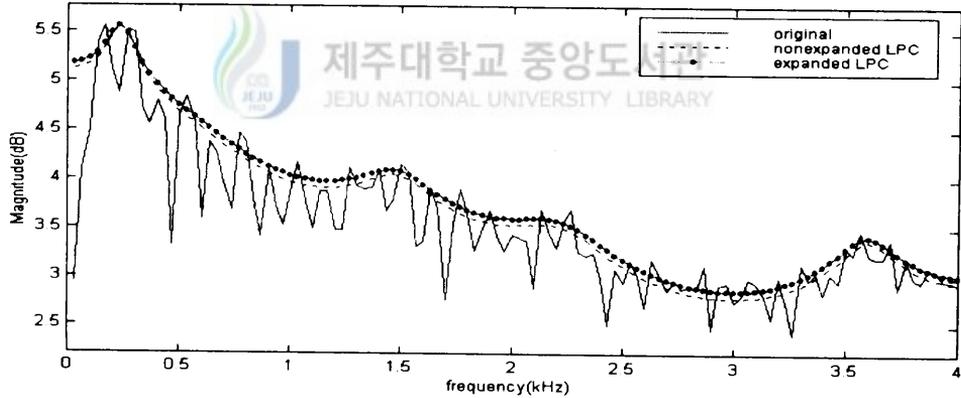


Fig. 15. The frequency analysis by LPC

Fig. 15에서 확인할 수 있듯이 LP보간에 의한 대역폭 확장의 모델링이 조금 더 큰 dB형태이다.

Table 1에 실례로 “통신공학과입니다”의 음성을 샘플링한 후 아홉 번째 프레임의 선형 예측 필터의 계수 및 확장한 필터 계수에 의한 극점의 위치를 나타내었다.

Table 1. The pole position of the 9'th frame

LPC	1.00000	-2.12793	1.53486	-0.15420	-0.89320	1.04058
	-0.29519	-0.49133	0.98792	-0.86528	0.29892	
극점의 위치	$-0.8655 \pm 0.3126i$	$-0.1264 \pm 0.8554i$	$0.3749 \pm 0.8105i$			
	$0.7265 \pm 0.3245i$	$0.9546 \pm 0.1548i$				
확장한 LPC	1.00000	-2.11543	1.51689	-0.15150	-0.87240	1.01038
	-0.28494	-0.47148	0.94244	-0.82061	0.28182	
극점의 위치	$-0.8707 \pm 0.3145i$	$-0.1272 \pm 0.8604i$	$0.3771 \pm 0.8153i$			
	$0.7308 \pm 0.3264i$	$0.9602 \pm 0.1557i$				

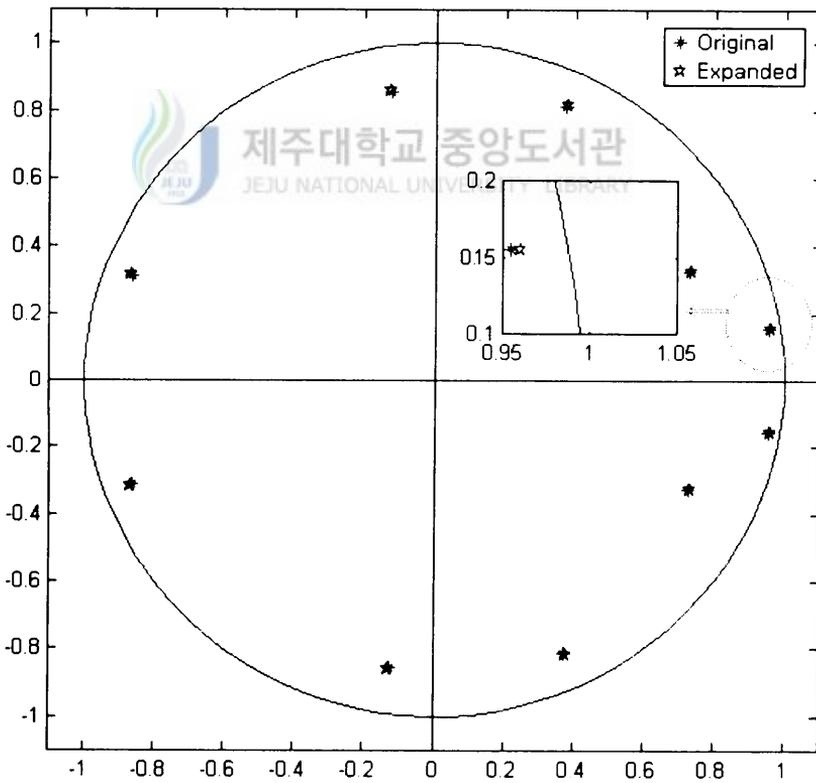


Fig. 16. The position of Table 1's pole

표와 그림을 통하여 알 수 있듯이 필터 계수 a_i , 대신 $a_i \gamma^i$ 로 대치함으로 필터의 극점을 방사 적으로 원점에서 먼 쪽으로 단위 원 가까이 이동 시켰음을 확인할 수 있다. 이런 대역폭확장을 시행함으로써 다음과 같은 효과를 얻을 수 있다.

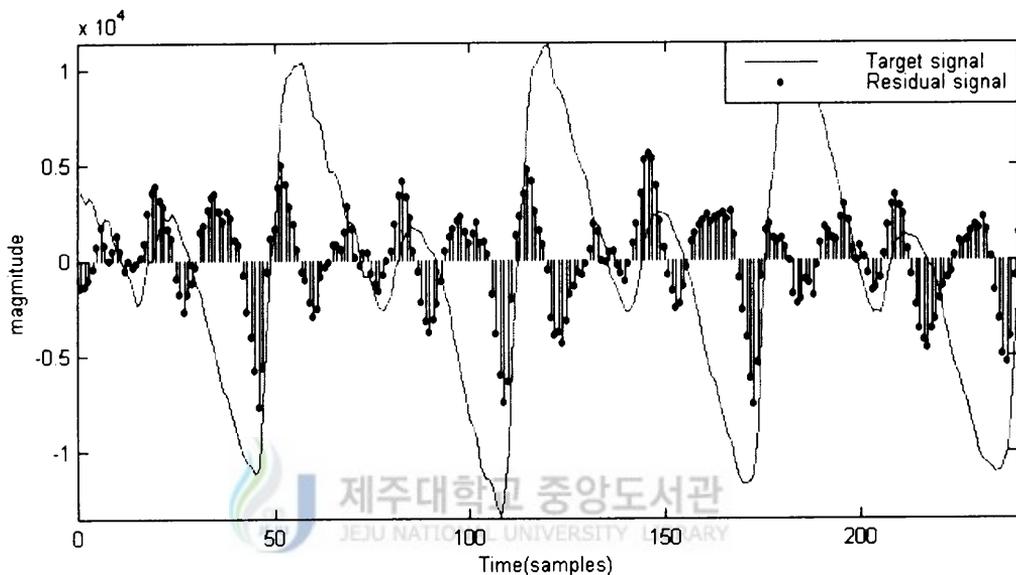


Fig. 17. The residual signal after the adaptive search by nonexpanded

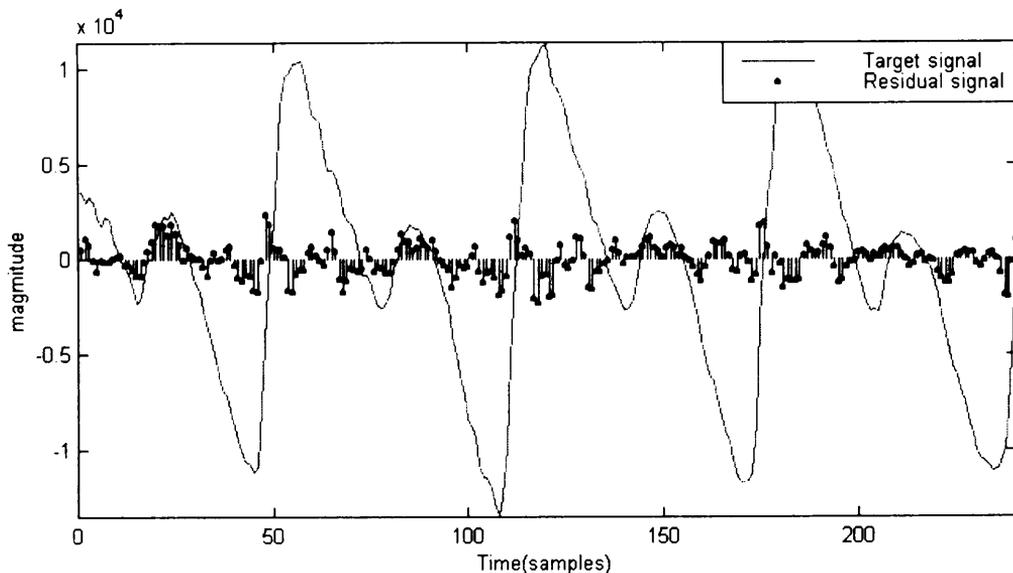


Fig. 18. The residual signal after the adaptive search by expanded

Fig. 17에는 LP 계수의 가공 없이 그대로 파라미터로 사용하여 적응 코드북으로 코딩한 나머지 성분으로 다음 단계인 고정 코드북 탐색의 타겟 벡터들을 도시하였으며, Fig. 16에서는 LP 계수를 a_i , 대신 $a_i\gamma^i$ 로 대치한 파라미터를 이용하여 그 잉여 성분을 도시하였다. Fig. 17과 18을 비교하여 보면 Fig. 18에서의 잉여 성분 크기가 작음을 알 수 있다. 이는 다음 단계인 고정 코드북 탐색에서의 효율을 보장할 수 있으며, 더욱 잉여 벡터와 일치하는 코드북 벡터를 찾아 낼 수 있다. 즉, 양질의 음성을 보장함을 의미한다.

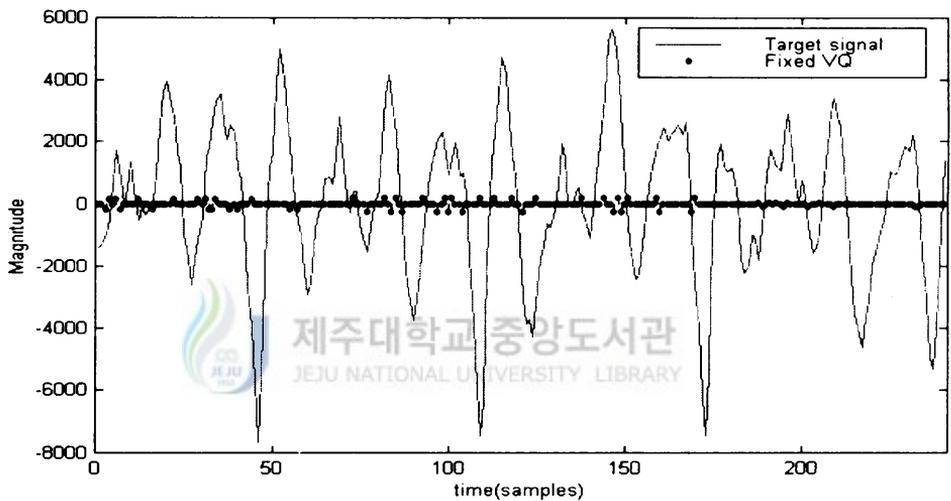


Fig. 19. The stochastic codebook VQ by nonexpanded LP

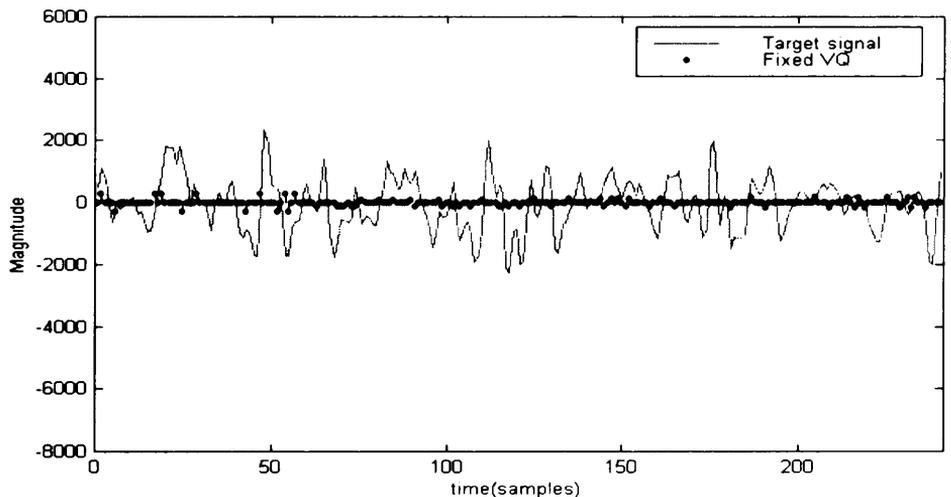


Fig. 20. The stochastic codebook VQ by expanded LP

Fig. 19와 Fig. 20에서 적응 코드북으로 분석 후의 잉여 성분을 코딩하는 과정인 고정 코드북 탐색으로 비 확장 때와 확장 시의 그래프를 도시하였다.

지금까지의 분석은 유성음 구간이다. 다음으로 무성음 구간에서 유성음 구간으로 전환할 때의 구간을 분석한다.

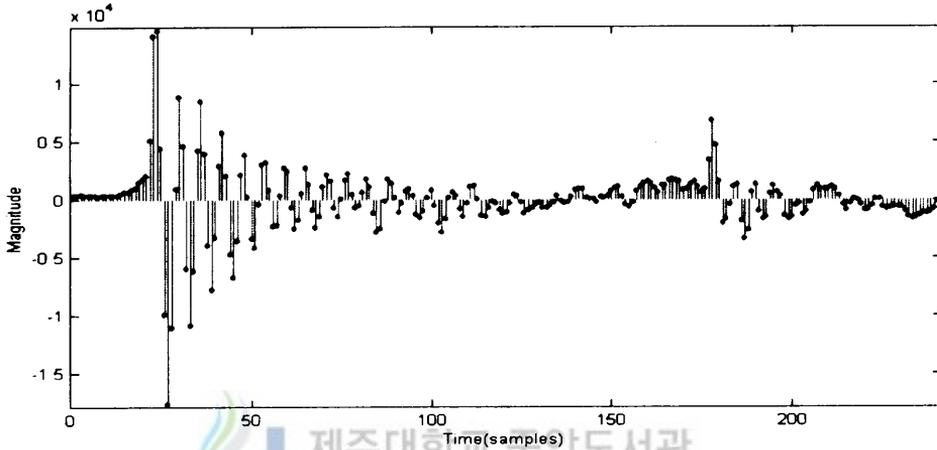


Fig. 21. The voiced and unvoiced transferred frame of the original speech

Fig. 21은 무성음에서 유성음으로 전환하는 구간으로 “통신공학과입니다”의 한 프레임이다. 앞에서와 마찬가지로 다음 Fig. 22에 주파수 분석으로 LPC와의 관계를 도시하였다. Fig. 22에서도 Fig. 15와 같이 대역폭 확장 기법을 이용한 LPC 분석이 높은 형태를 확인할 수 있다.

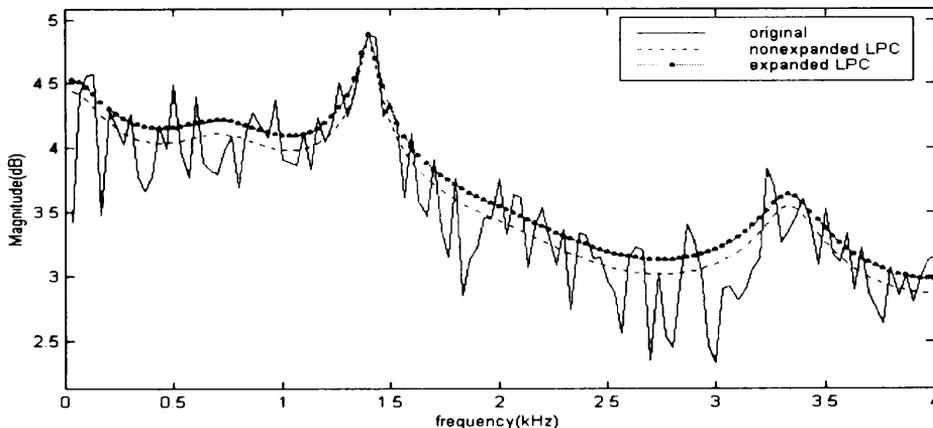


Fig. 22. The frequency analysis by LPC

이 구간에서 분석된 LPC와 근의 위치를 확인해 보면 다음과 같다.

Table 2. The pole position of the voiced and unvoiced transferred subframe

LPC	1.00000	-1.45762	0.91193	0.05577	-0.23912	-0.45715
	0.78404	-0.65210	0.38734	-0.12481	-0.06670	
극점의 위치	-0.2452	-0.7904 ± 0.4802i	-0.0091 ± 0.7104i			
	0.9207	0.7214 ± 0.4354i	0.4691 ± 0.8626i			
확장한 LPC	1.00000	-1.44906	0.90125	0.05480	-0.23355	-0.44389
	0.75681	-0.62576	0.36951	-0.11836	-0.06289	
극점의 위치	-0.2466	-0.7951 ± 0.4830i	-0.0091 ± 0.7146i			
	0.9262	0.7257 ± 0.4380i	0.4719 ± 0.8676i			

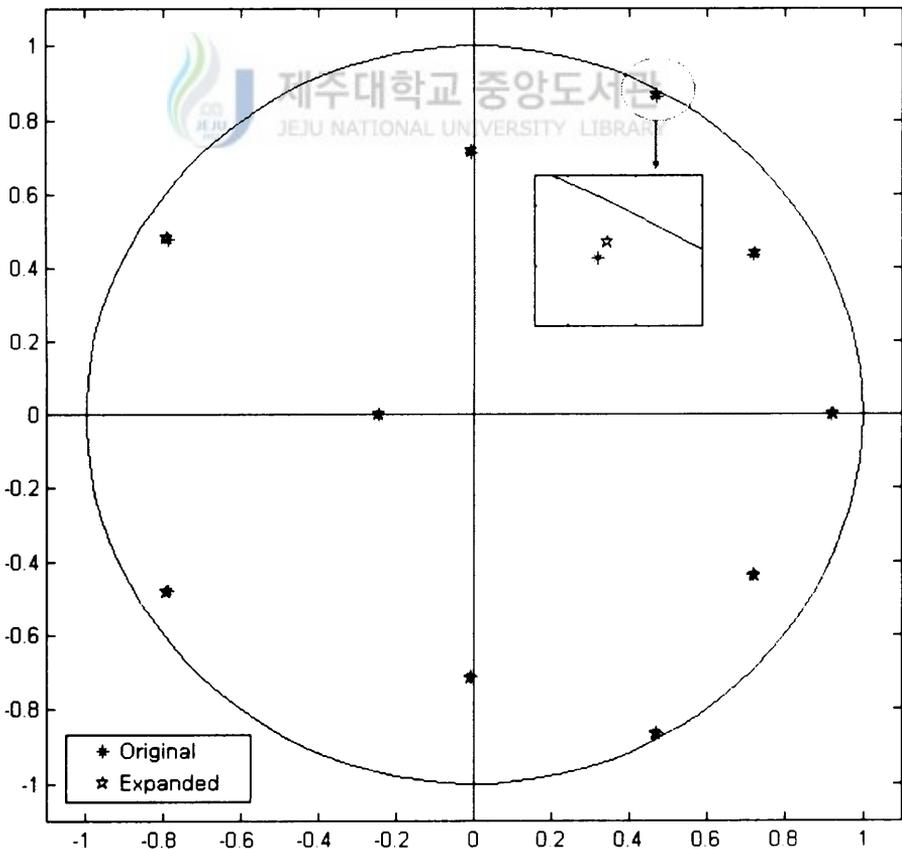


Fig. 23. The position of table 2's pole

위 Table. 2와 Fig. 23에서 확인할 수 있듯이 확장된 LPC의 근의 위치는 원점에서 단위 원 쪽으로 방사 적으로 이동하였다. 이런 과정을 거쳐 앞에서 언급하였듯이 다음과 같은 효과를 얻을 수 있다.

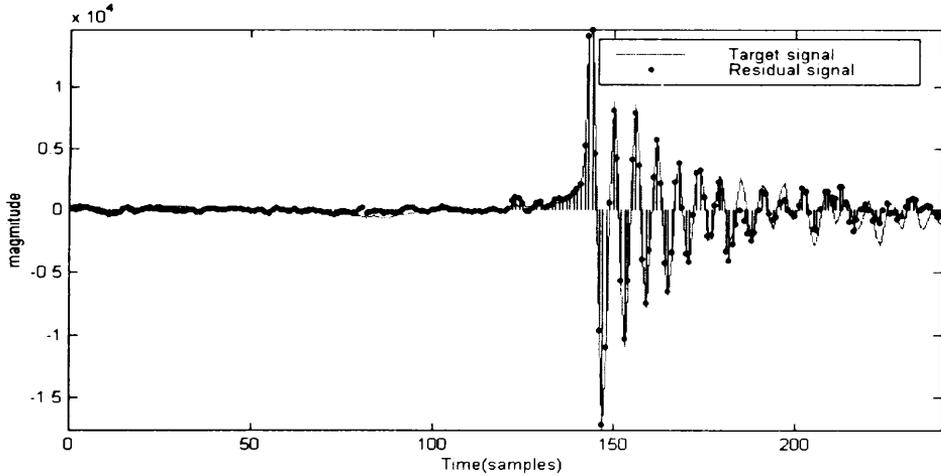


Fig. 24. The residual signal after the adaptive search by nonexpanded

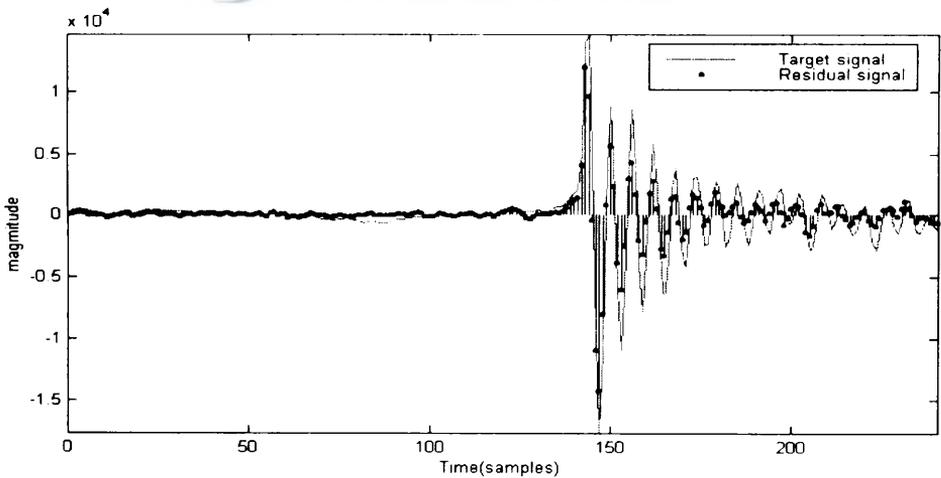


Fig. 25. The residual signal after the adaptive search by expanded

Fig. 24는 확장을 시행하지 않은 분석이며, Fig. 25는 확장을 시행하여 분석한 결과이다. 이 그림에서 나타나듯이 확장을 시행하여 분석한 것이 더욱 효율적인 적용 코드북 탐색을 초래함을 알 수 있다.

다음 Fig. 26과 Fig. 27은 다음 단계인 고정 코드북 탐색 결과이다.

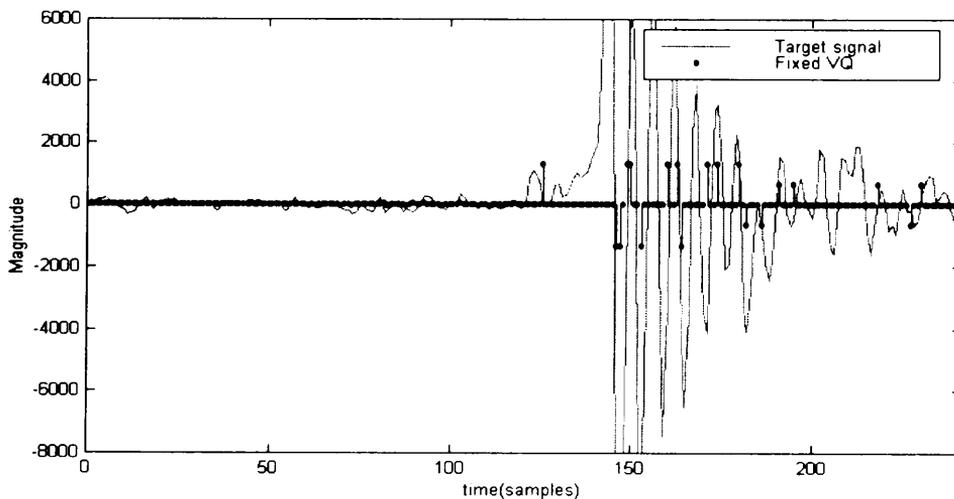


Fig. 26. The stochastic codebook VQ by nonexpanded LP

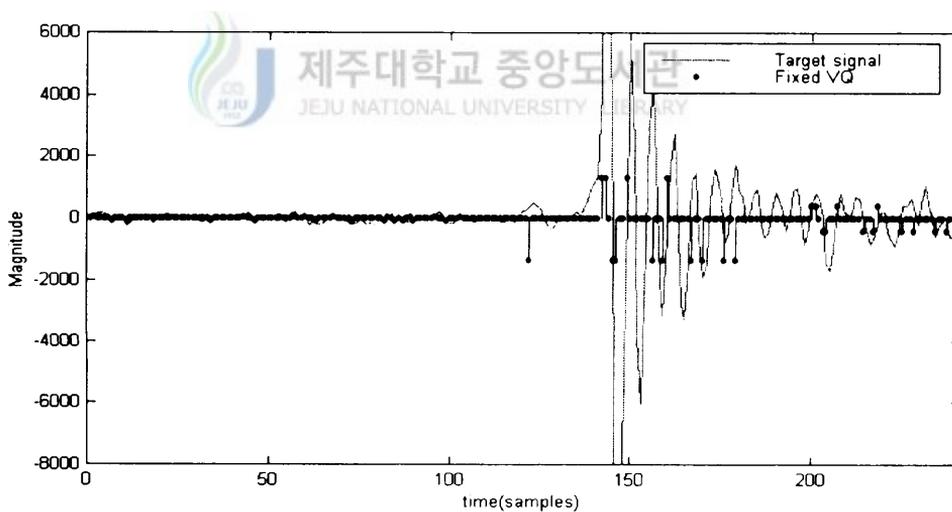


Fig. 27. The stochastic codebook VQ by expanded LP

이러한 과정을 모두 거쳐서 다음 아래와 같은 합성음을 만들어 낸다.

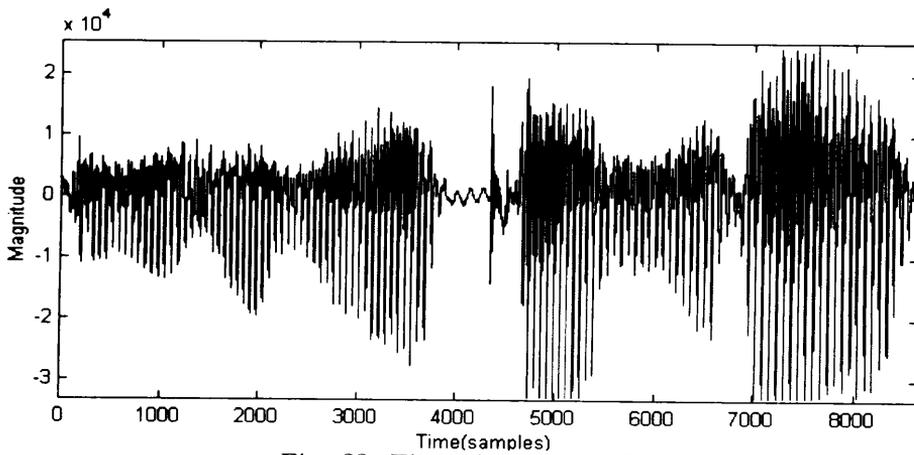


Fig. 28. The original speech

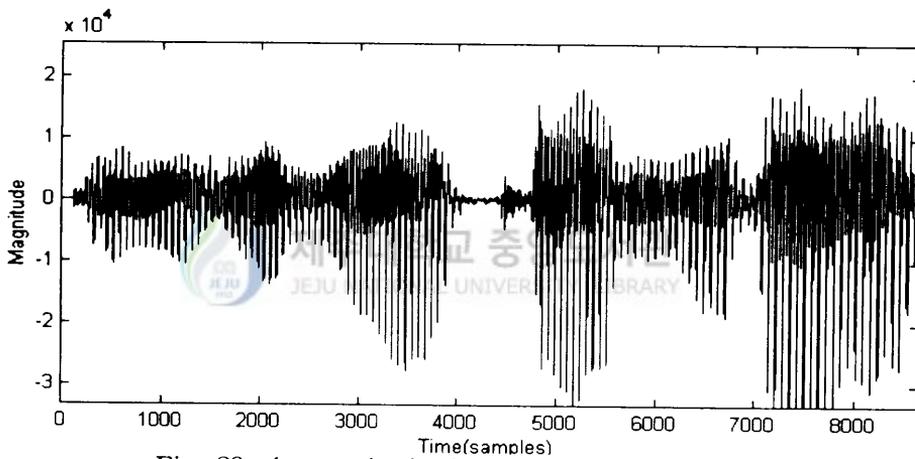


Fig. 29. the synthesised speech by nonexpanded

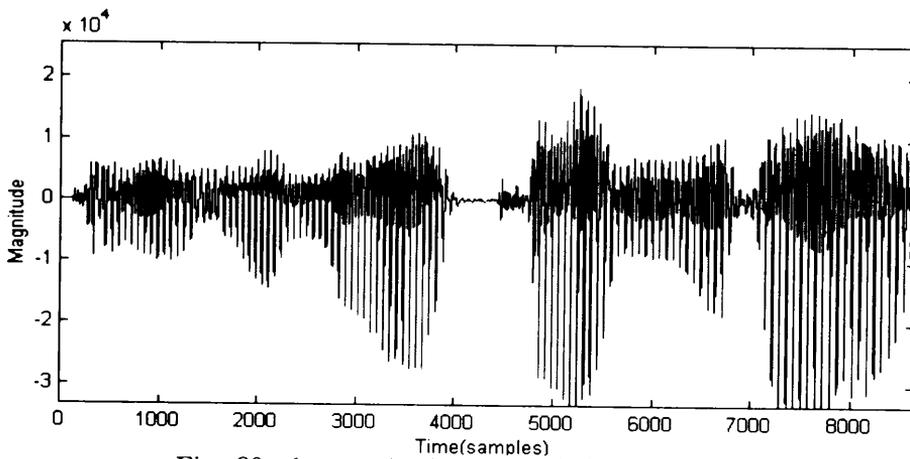


Fig. 30. the synthesised speech by expanded

Fig. 29와 Fig. 30에서 합성음의 비교 검토로서 음성 부호화기의 성능 평가를 위한 객관적인 성능 척도인 SEGWSNR(SEGmentally Weighted SNR)을 사용한다.

$$SEGWSNR = 10 \log \frac{\sum \|s\|^2}{\sum \|s - \hat{s}\|^2} \quad (40)$$

SEGWSNR를 구하여 보면, 확장을 이용하지 않았을 때의 SEGWSNR은 3.3322dB이고, 확장을 했을 때의 SEGWSNR은 4.4280dB이다. 즉 확장을 했을 때가 1.0958dB정도의 향상을 보였다. 그리고 주관적으로도 합성음을 들어보면 확장을 이용했을 때의 음성이 더 나음을 알 수 있었다.

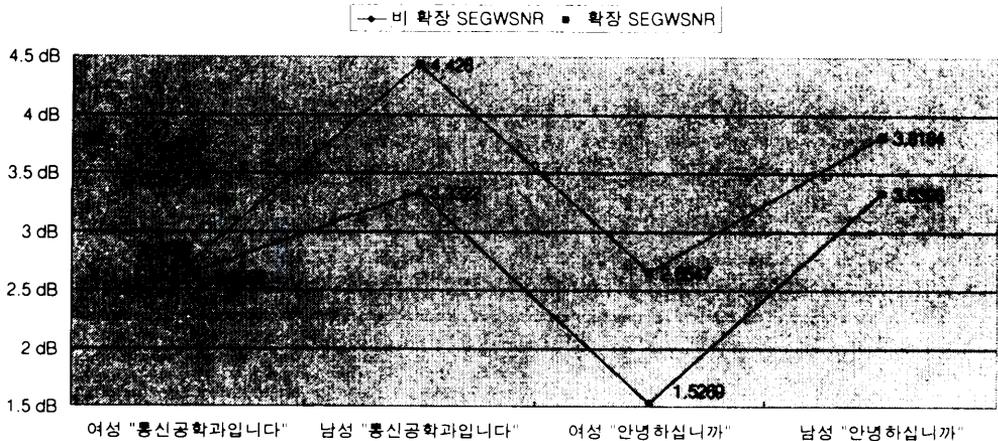


Fig. 31. The SEGWSNR graph of some sample voices

일반적인 음성들에 대한 음성 부호화기의 성능 평가를 위하여 남성·여성의 음성으로 “통신공학과입니다”와 “안녕하십니까”를 모의 실험에 사용하였다.

Fig. 31은 이 음성들에 대한 음성 부호화기 성능 평가인 SEGWSNR값들이다. 이 그래프에서도 살펴볼 수 있듯이 일반적으로 확장시의 합성음이 더욱 나은 SEGWSNR 값을 보이고 있다. 즉 부호화기의 성능이 일반적으로 향상하고 있음을 알 수 있다.

그에 반하여 이 부호화기의 단점으로 CELP 특성 상 생기는 지연 15msec가 있는데, 대역폭 확장으로 인하여 총 16msec의 지연이 생긴다.

다음으로 음질 향상이 아닌 다른 면에서의 효율적인 LP 필터 설계 제안인 개선된 LPC·LSF 상호 변환 알고리즘 수행결과이다.

LSF를 LPC로 변환시의 알고리즘을 식 (39)에 직접 대입으로 전체 연산의 과정이 줄어들고, 연산에서 요구되어지는 곱셈 및 덧셈의 연산이 현저히 줄어들었다.

본 논문에서는 홀수 차수의 LPC 필터는 고려치 않고, 많이 이용되고 있는 짝수 차수에 대한 개선 알고리즘이다.

이를 Fig. 32에서 이전에 제안된 KABAL의 알고리즘(Kabal, 1986)과 비교해 보았다.

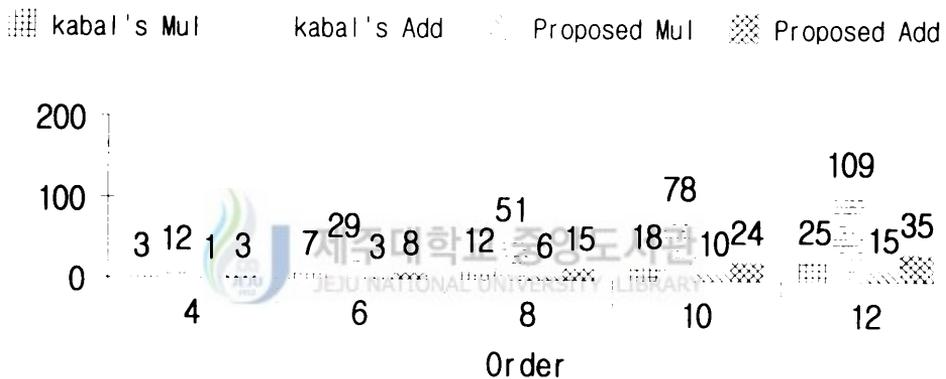


Fig. 32. G_1 's coefficients operations counts for the kabal algorithm and the revised algorithm

앞 그림을 살펴보면 가장 많이 이용되고 있는 10차를 예를 들면 곱셈 연산이 18개에서 10개로 줄고, 덧셈 연산이 78개에서 24개로 줄었음을 확인할 수 있다. 즉 곱셈 연산이 약 40%, 덧셈 연산이 약 70%의 감소를 확인할 수 있다.

V. 결 론

본 논문에서는 효율적인 음성 부호화기의 설계를 위하여 LP필터의 연구로서 LP보간에 의한 대역폭 확장과 개선된 LPC·LSF와의 상호 변환 알고리즘을 다루었다.

먼저 LP보간에 의한 대역폭 확장으로 효율적인 코드북 탐색을 야기하여 양질의 음성을 합성해 낼 수 있었다. 모의 실험을 통하여 이런 갱신으로 SEGWSNR 이득 최대화가 가능하다는 것을 찾았다. 이러한 LP보간의 대역폭 확장 기술을 사용하므로 코드북 탐색에 효율성을 높여 결과적으로 합성음에 SEGWSNR 1.0958dB 향상을 보였다. 그러나 대역폭 확장으로 인하여 인코더 지연을 1msec더 증가시킴을 알 수 있었다.

다음으로 LPC·LSF 상호변환의 효율적인 알고리즘의 개선으로 부호화기의 성능을 향상시켰다. 기존의 방법에 비하여 곱셈과 덧셈의 연산과정을 줄였으며, 10차 LP인 경우 곱셈은 40%, 덧셈은 3/1정도로 줄일 수 있어 본 알고리즘의 유용성으로 보였다.

이런 LP필터의 연구를 통하여 부호화기의 성능을 향상시킬 수 있었으며, 추후 인코더의 지연 증가를 줄일 수 있는 LP필터의 연구가 필요함을 알 수 있었다. 그리고 대역폭 확장 기술을 사용함으로 인코딩하기에 더욱 적합한 적응 코드북을 두어 더욱 더 부호화기의 성능을 향상시킬 수 있을 것이다.

참고 문헌

- A. Nejat Ince, Digital speech processing speech coding, synthesis and recognition, Kluwer Academic Publishers.
- Bishu S. Atal and Joel R. Remde, 1982, A New Modal of LPC Excitation for Producing Natural-Sounding Speech at Low Bit Rates, *Proc. ICASSP*, pp. 614-617
- Chung-Hsien Wu and Jau-Hung Chen, March 1984, A novel two-level method for the computation of the lsp frequencies using a decimation-in-degree algorithm, *IEEE Trans, on Speech and Audio Processing*, Vol.5, No.2 pp.106~115.
- Douglas O'Shaughnessy, 1990, Speech Communication Human and Machine.
- G. S. Kang and L. J. Fransen, April 1985, Application of line spectrum pairs to low bit rate speech encoders, in *Proc. Int. Conf. Acoust. , Speech, Signal Processing, Tempa, FL*. pp. 7.3.1~7.3.4.
- Gene H. Golub and Charles F. Van Loan, 1980, An Analysis of the Total Least Squares Problem, *SIAM Journal of Numerical Analysis*, vol. 17, no. 6, pp. 883-890
- J. D. Markel, Linear prediction of Speech, Speech Communication Research laboratory, Inc..

- J. P. Woodard and L. Hanzo, Improvements to the analysis-by-synthesis loop in CELP CODECS.

 - Jason P. Woodard, June 1994, Digital Speech Coding. Mini-Thesis, Department of Electronics and Computer Science, University of Shouthampton.

 - John Bellamy, Digital Telephony, A Wiley-Interscience Publication.

 - L. Fox and I. B. Parker, 1968, Chebyshev Polynomials in Numerical Analysis, London, England Oxford University Press. pp.5.1~5.3.

 - M. Fratti, G.A. Miani and G. Riccardi, 1992, On The Effectiveness of Parameter Reoptimization in Multipulse Based Coders, *Proc. ICASSP, vol. 1*, pp. 73-76
- 
제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY
- Mahesan Niranjana, 1990, CELP Coding with Adaptive output-error Model Identification, *Proc. ICASSP*, pp. 225-228

 - Mammone, Robert J. ,Digital Signal Estimation, A Wiley-Interscience Publication.

 - MD. Anisur Rahham and Kai-Bor Yu, 1987, Total Least Squares Approach for Frequency Estimation Using Linear Prediction, *IEEE Transactions on Acoustics, Speech and Signal Processing*, pp. 1440-1454

 - P. Kabal, J. L. Moncet and C. C. Chu, 1988, Synthesis Filter Optimization and Coding: Applications to CELP, *Proc. ICASSP, vol. 1*, pp. 147-150

· Peter Kabal and Ravi Prakash Ramachandran, December 1986, The Computation of Line Spectral Frequencies Using Chebyshev Polynomials, *IEEE Trans, on Acoustics, Speech, and Signal Processing, ASSP. 34, No. 6*, pp.1419~1426.

· Ronald D. Degroat and Eric M. Dowling, 1993, The Data Least Squares Problem and Channel Equalization, *IEEE Transactions on Signal Processing*, pp. 407-411.

· Sharad Sigal and Bishnu S. Atal, 1983, Optimizing LPC Filter Parameters for Multi-Pulse Excitation, *Proc. ICASSP*, pp. 781-784

· Thomas W. Parsons, Voice and speech processing, McGraw-Hill Book Company.



제주대학교 중앙도서관
JEJU NATIONAL UNIVERSITY LIBRARY

· Yoh'Ichi Tohkura, Fumitada Itakura and Shin'Inchiro Hashimoto, 1987, Spectral Smoothing Technique in PARCOR Speech Analysis-Synthesis, *IEEE Trans. on Acoustics, Speech and Signal Processing*, pp. 587-596.

감 사 의 글

학부 생활서부터 제 마음의 길을 일깨워 주시고, 본 논문이 완성이 되어 결실을 맺을 수 있도록 세심한 배려와 관심을 아끼지 않으신 임재윤 교수님께 진심으로 감사를 드립니다. 그리고, 논문을 잘 마무리할 수 있도록 조언과 격려를 해 주신 이용학 교수님, 문건 교수님, 김홍수 교수님, 양두영 교수님, 강진식 교수님께도 감사 드립니다.

학부, 대학원 생활을 함에 있어 매사에 힘이 되어주신 박경배 선배님과 나를 무척이나 아껴주시는 함영희 선배님, 그리고 또 나와 같이 연구실에서 부대끼며 서로 도움을 주고받던 형준이에게 고마움을 전하며, 이제 연구실에서 열심히 하고 있는 재필 형과 영애에게도 고마움을 전합니다.

아버지, 어머니 당신의 아들이 이렇게 성장할 수 있도록 헌신적인 사랑으로 보살펴 주심에 어떻게 보답해야 할지 또 보답하기는 어렵겠지만 정말로 감사 드립니다. 그리고 언제나 따뜻한 마음으로 나에게 힘을 주는 영진과 이 기쁨을 함께 나누고 싶습니다.

마지막으로 다음의 글을 마음에 깊이 새기며, 저를 아시는 모든 분과 기쁨을 함께 하고 싶습니다.

오늘 하루라는 오솔길을

열심히 가야 인생이라는 넓은 길을 만날 수 있다.

최선이라는 오솔길을

열심히 가야 성공이라는 넓은 길을 만날 수 있다.

고통이라는 오솔길을

열심히 가야 성취라는 넓은 길을 만날 수 있다.

그래서 오솔길을 열심히 가야한다.

넓은 길을 만나려거든 오솔길을 열심히 가야 한다.