

## 상관관계 특성을 이용한 CELP 코더의 피치검색시간 단축법의 비교

( On a Performance Comparison of Pitch Search Algorithms with the Correlation Properties for the CELP Vocoder )

°김 대식, 권 현일, 주 한서, 배 경진  
충실대학교 정보통신공학부

°D.S. Kim, H.I. Kwon, H.S. Joo, M.J. Bae  
Dept. of Telecommunication Engineering Soongsil University

### ABSTRACT

Code Excited Linear Prediction(CELP) speech coders exhibit good performance at data rates as low as 4800bps. But the major drawback to CELP type coders is their large computational requirements. Therefore, in this paper a comparative performance study of three pitch searching algorithms for the CELP vocoder was conducted. For each of the algorithms, a standard pitch searching algorithm was used by the full pitch searching algorithm that was implemented in the QCELP vocoder. The algorithms used in this study is to reduce the pitch searching time 1) using the skip table(TABLE), 2) using the symmetrical property of the autocorrelation(SYMMPT), and 3) using the pre-processing autocorrelation(PREPC), 4) using the positive auto-correlation(POS-AUTO), 5) using the preliminary pitch (PRELP). Performance scores are presented for each of the five pitch searching algorithms based on computation speed and on pitch prediction error.

### 1. 서 론

음성신호를 메모리에 저장하거나 전송하기 위한 음성 부호화법에는 크게 파형부호화법, 신호원 부호화법, 혼성부호화법 등이 있다. 혼성부호화법에서 포먼트성모는 선형 예측부호화법으로 보통 부호화하고, 그 나머지 잔여신호들을 어떻게 부호화하느냐에 따라 RELP법, VELP법, MPLP법, CELP법 등이 제안되어서 있다. 혼성부호화법들 중에서도 최근 이동통신용 부호화법으로 채택되고 있는 것은 CELP (code excited linear prediction)법이다.

CELP부호화법에서는 피치필터를 적용하여 음성신호의 피치주기 성분을 부호화하고 있다. 이러한 피치필터에 주로 적용되고 있는 피치검색법은 피치치인에 따른 상관관계법이다. 상관관계 피치검색법은 피치가 존재하는 모든 피치치인에 대해 원래음성과 합성된 음성의 상관관계를 검색하고, 최상의 상관관계를 갖는 피치치인의 피치치인과 이득을 결정하게 된다. 그러나 이러한 피치 검색과정은 모든 피치구간에 대해 반복적으로 검색해야 하기 때문에 그 처리시간이 많이 소요된다. 따라서 피치 검색시간을 줄이는 연구가 최근 몇년 동안 지속적으로 연구되어져 왔다.

피치 검색시간을 절감하는 방법들은 음절단위의 상관관계 특성을 이용하는 방법과 표본단위의 상관관계 특성을 이용하는 방법으로 나누어 연구되어지고 있다. 음절단위의 특성을 이용하는 방법은 음성신호가 20ms정도까지 정상상태를 유지한다는 성질을 이용하는 것으로서, 자기 여기된 방법[6], 델타 피치검색 방법[9] 등이 제안되어져 있다. 표본단위의 상관관계 특성을 이용하여 피치 검색시간을 단축하는 방법은 음성표본이 갖는 인접 표본간의 자기 상관관계 특성을 이

용하는 것으로서, 상관관계 이력표(검색구간표)를 사용하는 방법[12], 상관관계 특성의 대칭성을 이용하는 방법[13], 전치리용 상관관계식을 사용하는 방법[14] 등이 있다.

본 논문에서는 이들 피치검색법들의 원리를 간단히 살펴본 다음에 피치검색시에 얻어지는 시간단축효과 이때 야기되는 피치 예측이득의 손실정도를 파악하고자 한다.

### 2. CELP 코더의 원리

그림 2-1은 CELP 코더에 대한 블록도를 나타내었다. CELP 코더에서 포먼트 합성필터는 10개의 LPC 전압형 구조가 보통 적용된다. LPC계수는 양자화시에 왜곡이 크기때문에 LSP계수로 변환되어 부호화되고, 복호화시에 다시 LPC계수로 변환시킨다. LPC계수는 한 프레임 구간(20ms)마다 부호화되고, 인터플레이션을 통해 5ms의 각 부분레이마다 별도로 공급해 준다. 여기서의 과라미터는 또한 5ms 부분레이마다 새롭게 바꾸어 준다.

CELP부호/복호기는 2개의 여기원은 사용된다. 첫 번째 여기원은 장기(피치) 예측상태 또는 적응코더북이다. 두 번째 여기원은 여기코더북에서 취한다. 지전송율의 경우에는 코더북의 크기는 128개 정도이다. 이들 2개의 여기원은 이득형이 곱해진 다음에 합해진다. 이것은 결합된 여기치역수가 된다. 각 부분레이의 여기중력인 다음 부분레이에서 취할 수 있도록 적응코더북의 장기 피치상태를 바꾸는데 사용된다.

CELP코더의 부호화법에서는 포먼트성모가 잘리지고 남은 잔여신호에 대해 벡터 양자화기법을 적용하여 부호화하기 때문에 잔여신호성분 전송에 적용되는 데이터는 코더북의 인덱스기 된다. 따라서 전송율은 4.8kpbs이하로 낮출 수 있으며, 이러한 과라미터를 음성성분용 부가코드와 함께 전송하면 전송잡음에도 강인한 부호화법이 된다. 또한 이 부호화법은 합성에 의한 분석법을 적용하여 피치의 유절이 유지되도록 반복적으로 분석하기 때문에 주어진 지 전송율에서도 음성은 우수하다.

CELP 코더는 어떤 음성을 합성해서 비교해야 하므로 복잡한 구조를 갖는다. 특히 부호화시에 계산량이 많이 요구하는데, 코더북

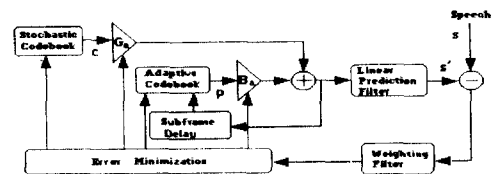


그림 2-1. CELP 음성 부호화기  
Fig. 2-1 A CELP speech coder

에서 입력 여기신호를 찾는 과정과 피치필터  $P(z)$ 의 계수를 구하는 과정에서 대부분의 계산시간을 소요하게 된다.

피치검색 과정은 음성신호의 장기 상관관계에 해당하는 피치 주기정보를 얻는 과정이다. 이것은 8kHz로 표본화된 음성의 경우 피치 분석은 5ms 마다 한번씩 수행된다. 스펙트럼 분석이 개화로 구조로 수행되어짐에 비해 피치분석은 폐회로 구조에 의해 수행되어야 한다. 즉, 피치 지연조건을 최적으로 만족하는 값을 반복적인 비교를 통해 결정하게 된다. 피치이득은 이때 얻어진 피치시연을 이용하여 상대적인 이득값을 갖도록 양자화하게 된다. 이러한 피치 검색과정은 코덱검색과 함께 CELP 부호화기의 계산량에 큰 영향을 미치는 중요한 부분이다.

확장 적응코덱을 사용한 피치 검색구조는 그림 2-2와 같이 LPC 합성필터의 입력 여기신호로 사용되는  $y(n)$ 을 이용하여 피치필터의 적응코덱을 변화시키는 구조가 된다[8]. 적응코덱 내에는  $y(n)$ 으로부터 얻은 각 지연값  $L$ 에 대한  $y(n-L)$ 값을 저장하고 나서 이들 다음 프레임에서 피치시연을 구할 때 사용하게 된다. 각 코덱워드를 대해서 자신화  $d(n)$ 을 오차로 구한 후 이  $d(n)$ 을 최소화 하는 코덱워드를 선택하게 된다.

피치 검색과정은 원래음성과 합성된 음성간의 피치 지연조건을 최적으로 만족하는 값을 반복적인 비교를 통해 결정하는 과정이다. 즉, 이것은 원래의 음성신호에 대해 시간지연을 점차 변화하면서 자기 상관관계치를 감출하고, 이들중 최대의 상관관계치를 나타내는 시간지연을 피치주기로 검출하는 과정이 된다. 이렇게 계산된 상관관계치는 시간지연에 따라 그림 2-3(b)처럼 나타난다. 여기서 시간지연이 음성파형의 주기상의 정수배에 해당될 때마다 상관관계가 최대치를 이루게 된다.

### 3. QCELP의 피치검색 알고리즘

1/8음의 패킷으로 부호화되고 있는 프레임은 제외하고는, 모든 코덱워드의 프레임은 동일한 길이의 피치 부프레임으로 재차 분할된다. 전송율 1패킷에 대해서는 피치 부프레임이 4개가 있고, 전송율 1/2 패킷에 대해서는 피치 부프레임이 두개, 그리고 전송율 1/4패킷에 대해서는 피치 부프레임이 하나가 있다. 전송율 1/8패킷에 대해서는 피치 부프레임이 없다. 피치 합성필터는 다음과 같이 나타낼 수 있다:

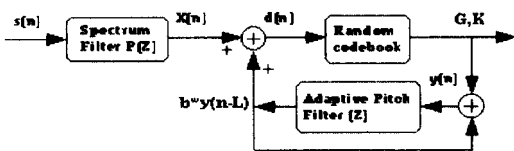


그림 2-2. 확장 적응 코덱을 이용한 피치 분석기

Fig.2-2 A pitch search by using the extended adaptive codebook

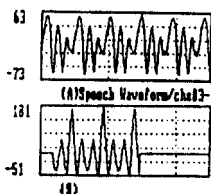


그림 2-3. 시간지연에 따른 상관관계값

(a)음성파형 (b) 자기 상관관계 파형

Fig.2-3 Autocorrelation for time delay

$$1 / P(z) = 1 / 1 - bz^{-L} \quad (3-1)$$

피치지연  $L$ 은 7비트로 표현되어 17에서 143까지의 범위를 갖는다.  $L=16$ 은  $b=0$ 인 경우에 대해 사용한다. 피치이득  $b$ 는 3비트로 표현되어 0에서 2까지의 범위를 갖는다. 각 피치 부프레임에 대해 음성 부호화기는 피치지연  $L$ 과 피치이득  $b$ 를 결정하고 부호화하게 된다.

피치 파라미터를 선택하는데 적용되는 방법은 합성에 의한 분석법이며, 부호화는 입력 음성과 합성된 음성간의 가중된 오차를 최소화하는 파라미터를 선택하는 과정이다. 합성된 음성은 LPC필터에 의해 여파된 피치 합성필터의 출력이다. 피치합성은 입력으로 모두 영값 요소로 구성된 코덱북 벡터를 사용한다. 피치지연  $L$ 은 집합{17, 18, ..., 143}에서 선택되고, 피치이득  $b$ 는 집합 {0, 0.25, 0.5, ..., 2.0}(0과 2.0사이를 0.25간격으로 선형적으로 양자화됨)에서 선택된다. 가중화 이득  $G$ 는 다음 형식을 갖는다:

$$W(z) = A(z) / A(z; \xi) \quad (3-2)$$

여기서  $A(z)$ 는 포먼트 예측오차필터이고,  $\xi$ 는 인치 가중화된 파라미터로 0.8이다. 가중화 필터에 적용된 LTP계수는 원 피치 부프레임에 대한 것과 같다. 음성부호화기에 적용된 합성필터는 다음과 같다:

$$H(z) = (1 / A(z)) W(z) = 1 / A(z; \xi) \quad (3-3)$$

여기서 인치 가중화 필터가 뒤따르는 부호기 음성 합성필터이고, 이를 가중화된 합성필터(weighted synthesis filter)라 한다.

### 4. 검색구간표에 의한 피치검색법

CELP코덱에서 피치검색은 잔여신호에서 합성된 음성신호가 원래음성과 가장 유사하게 나타나는 자기 상관관계가 최상인 피치지연값과 이 때의 피치이득을 구하게 된다[1-3]. 따라서 피치지연값 중에서 자기 상관관계가 낮은 구간을 피치검색구간에서 제외하는 방법을 적용하여 피치 검색시간을 줄이고자 한다. 먼저 상관관계가 낮은 구간을 파악하기 위해 포먼트의 영향을 제거하면서 파형의 주기성만 강조하고, 자기 상관관계가 높은 주된 봉우리 구간을 검출하여 피치 검색시 자기 상관관계가 낮은 구간은 검색구간표를 작성하여 건너뛰는 방법을 적용한다.

음성신호의 피치는 음성파형의 반복되는 봉우리에서 봉우리가 지나는 곳에서 골까지로 정의된다. 파형의 봉우리 위주로 피치를 검출하는 경우에는 두드러진 봉우리가 존재하는 시간지연에 대해서만 자기 상관관계가 높게 존재한다. 반면, 파형의 골에 의해 피치를 검출하는 경우에는 두드러진 골이 존재하는 시간지연에 대해서만 자기상관관계가 높게 존재한다.

음성파형의 주요 봉우리에 대한 구간을 검출하기 위해 음성신호의 피치가 2.5ms(20표본) 이내에는 존재하지 않는다는 특성을 이용한다. 따라서 봉우리의 조사는 2.375ms(19표본) 마다 수행한다. 이렇게 하여 피치주기를 이루는 봉우리를 검출하면 파형의 봉우리가 de offset의 영향을 받는 경우에도 19표본 내에서의 최대신폭을 이루는 봉우리를 반드시 선택하기 때문에 피치봉우리가 검출될 수 있다. 또한 한 프레임단위로 피치를 검출하면 이 프레임내에서 잔폭의 빠른 변화에 의해 피치주기를 이루는 봉우리를 잘 검출하지 못하는 경우가 있으나, 여기서는 2.375ms마다 봉우리를 이루는 선택을 고려하기 때문에 이러한 문제도 해결할 수 있게 된다.

검색시간을 단축하기 위해 파형 봉우리들의 진폭 25% 이상인 자기 상관관계 계산에 사용한다. 이렇게 하면 피치검색시 포먼트의 영향을 제거할 수 있게 된다. 한 프레임에 대해 상위 25% 이상의 진폭 봉우리를 검출하기 위해 부프레임 단위로 그림 4-1과 같이 처리

## 상관관계 특성을 이용한 CELP코더의 피치검색시간 단축법의 비교

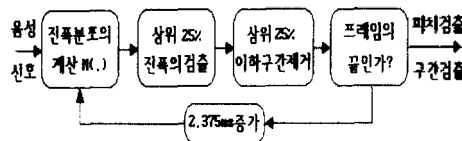


그림 4-1. 부프레임 마다 상위 25%의 봉우리를 검출하는 과정  
Fig. 4-1 A detecting peaks with upper 25% magnitude each subframe

한다. 상위 25%의 진폭을 구하기 위해 먼저 다음의 진폭분포도를 구한다:

$$M(s(n)) = M(s(n)) + 1, \quad n = 0, 1, \dots, N-1 \quad (4-1)$$

여기서  $M(n)$ 은 진폭분포 벡터이고,  $s(n)$ 은 음성신호이며 그리고  $N$ 은 한 프레임의 길이로 19 표본수이다. 진폭배열에 대해 변수를 최대치에서부터 점차 낮추면서 벡터값의 누적이  $N \times 25\%$  개수가 될 때까지 변수를 카운트-다운한다. 이 때의 변수는 25% 이상의 진폭값을 나타낸다. 25% 이상의 진폭만 남기고 나머지는 영으로 대체시킨다. 상기의 과정을 2.375ms 구간마다 반복한다.

피치 검색구간을 검출하려면 검출된 상위 25%의 파형 진폭봉우리 구간(그림 4-2(a))에 대해 피치봉우리의 시점이 되는 중심봉우리를 찾아야 한다. 중심봉우리를 찾는 과정을 실제의 파형에 대해 그림 4-2에 나타내었다. 한 프레임에서 첫 두 부프레임(4.75ms) 구간 내에서 최대진폭의 주된 봉우리 중심구간을 찾는다. 여기서부터 다음 봉우리 구간이 나올 때마다 피치 검색구간으로 선택한다(그림 4-2(b)). 그리고 실제 피치의 시점이 첫 두개의 부프레임 내에 포함되지 않는 경우를 대비하여, 끝 두개의 부프레임(4.75ms) 구간 내에서 최대진폭의 주된 중심 봉우리구간을 찾은 다음에, 여기서부터 거꾸로 진행하면서 다음 봉우리구간이 나올 때마다 피치 검색구간으로 선택한다(그림 4-2(c)). 따라서 피치주기가 존재할 피치 검색구간은 찾아진 각 중심 봉우리의 중심에서부터 앞 또는 뒤로 시간지연시킨 때 나타나는 봉우리구간(그림 4-2(d))이다. 중심 봉우리에서부터 실제의 피치는 2.5ms가 경과된 시간지연 이후에 존재하게 된다(그림 4-2(e)).

## 5. 상관관계에 의한 피치검색시간의 단축

피치검색시 음성신호  $s(n)$ 에 대해 시간지연에 따른 상관관계치  $E(L)$ 를 다음과 같이 계산한다:

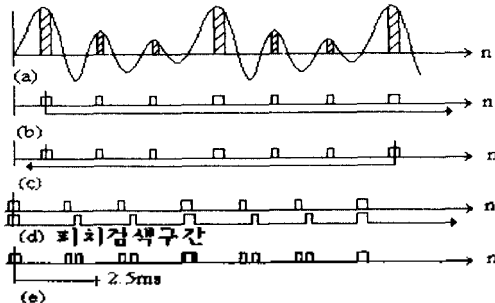


그림 4-2. 피치검색 구간의 설정과정

(a)음성을 파형과 25%의 봉우리, (b)파에서 위로 중심봉우리 결정, (c)우에서 좌로 중심봉우리 결정, (d)중심봉우리에서부터 두 영역의 결합, (e) 실제 피치 검색영역 검출

Fig. 4-2 A Detection process of the pitch search durations

$$E(L) = \frac{\sum_{n=0}^{L-1} [s(n)s(n-L)]}{\sum_{n=0}^{L-1} [s(n)s(n)]} \quad (5-1)$$

여기서  $L$ 은 피치지연을 나타낸다. 이때 상관관계치는 피치 주기마다 100%에 근접한 값이 얻어지며, 유사한 정도가 어느 정도인지는 피치검색 구간 내에 있는 파형의 주기성과 파형 진폭의 변화에 따라 다르다. 그러나 피치의 주기성이 강한 음성일 경우에는 보통 70% 이상이 얻어지며, 상관관계 값은 시간지연에 대해 그 변화가 작아지게 나타난다.

피치검색법에서 가장 바람직한 피치지연을 구하기 위해서는 식 (5-1)과 같은 상관관계식을 가능한 모든 피치지연에 대해 반복적으로 수행해야 한다. 이것은 매 피치지연  $L$ 에 대해  $L$ 의 범위는 20에서 147 표본까지 곱셈과 덧셈을  $L$ 번씩 수행해야하는 많은 계산량이 된다. 이 때문에 CELP 코더를 가장 최근의 DSP칩으로 구현하여도 5 MIPS 이상이 소요되며 코더 구현시간의 반을 차지하게 된다. 따라서 피치 검색에차는 변동되지 않는 상황에서 피치 검색시간만 단축시킬 수 있어야 한다.

식 (5-1)을 통해 음성파형의 상관관계를 계산해 보면 그 파형은 다음과 같은 특징이 있다:

1. 진폭변화: 음성파형에서 인접 표본간의 상관관계가 높기 때문에 이러한 상관관계를 나타내는 상관관계 파형 봉우리도 작아지게 변화된다.
2. 봉우리 폭: 음성음의 파형에서 제 1 포먼트의 에너지가 가장 크기 때문에 한 피치주기내의 파형변화는 제 1 포먼트의 구조로 감쇄진동은 하게 된다. 이에 대한 자기 상관관계 파형도 제 1 포먼트의 주기를 그대로 유지하는 파형이 구성된다. 따라서 상관관계 파형은 피치주기 사이에서 일정된 폭을 이루게 된다.
3. 음의 봉우리: 음성파형은 양과 음의 봉우리가 교대로 반복되어 한 피치주기를 이루고 있다. 양의 파형 봉우리를 기준으로 하여 상관관계를 구하면 양의 파형 봉우리마다 양의 값이 얻어지며, 음의 파형을 기준으로 하면 음의 파형 봉우리마다 양의 값이 얻어지게 된다. 따라서 상관관계 파형도 시간지연에 따라 음과 양의 봉우리가 반복적으로 나타나는 구조를 이루게 된다.

피치 검색과정에서 최대의 상관값이 얻어지는 피치지연은 피치 주기로 고려하기 때문에 피치검색에 적용되는 상관관계값은 양의 봉우리를 이루는 경우만을 필요로 한다. 또한 상관관계의 특성상 양의 봉우리와 음의 봉우리가 반복되어 나타난다. 이러한 사실은 피치지연에 따라 양의 봉우리가 나타나면 그 봉우리가 끝나는 부분에서부터 양의 봉우리폭 만큼은 음의 봉우리가 존재하는 구간이기 때문에 이 구간은 상관관계값을 계산할 필요가 없어진다. 본 방법에서는 이러한 단점을 이용하여 불필요한 피치 검색시간을 줄이게 된다.

## 6. 전처리에 의한 피치검색 시간의 단축

피치검색시에는 잔여신호  $s(n)$ 에 대해 시간지연에 따른 상관관계치  $E(L)$ 를 다음과 같이 계산한다:

$$E(L) = \frac{\sum_{n=0}^{L-1} [s(n)s(n-L)]}{\sum_{n=0}^{L-1} [s(n-L)s(n-L)]} \quad (6-1)$$

여기서  $M$ 은 부프레임의 길이를 나타내고,  $L$ 은 시간지연을 나타

낸다. 이렇게 하여 시간지연에 따라 계산된 상관관계치는 그림 2-3(b)처럼 나타난다. 이때 상관관계치는 피치 주기마다 100%에 근접한 값이 얻어지며, 유사한 정도가 어느 정도인지는 피치 검색 구간 내에 있는 파형의 주기성과 파형 진폭의 변화에 따라 다르다. 그리고 시간 지연이 음성파형의 주기성의 경우 배에 해당될 때마다 상관관계가 최대치를 이루게 된다.

파형의 봉우리 위주로 피치를 검출하는 경우에는 두드러진 봉우리가 존재하는 시간지연에 대해서만 자기 상관관계가 높게 존재한다. 반면, 파형의 골에 의해 피치를 검출하는 경우에는 두드러진 골이 존재하는 시간지연에 대해서만 자기 상관관계가 높게 존재한다. 파형의 봉우리와 골을 사전에 검출할 수 있다면, 이 때의 상관관계는 다음과 같이 계산할 수 있다:

$$R(L) = \sum_{k=1}^L |s(n-L)(s(n)+s(n-2L)) + \sum_{k=1}^L s(k-L)(s(k) + s(k-2L))|$$

$$L = 20, 21, \dots, 147. \quad (6-2)$$

여기서  $s(n)$ 은 잔여신호의 파형봉우리를 나타내고,  $s(k)$ 는 잔여신호의 파형골을 나타내고,  $n=0$ 은 봉우리의 정점을, 그리고  $k=0$ 은 골의 계곡점을 나타낸다. 그리고 상관관계값을 봉우리(또는 골)의 정점  $n=0$ 을 기준으로  $n+1$ 에서  $n-1$ 까지 고려한 것은 임펄스성 압축에 의해 상관관계값이 큰 영향을 받지 않게 하기 위함이다. 두드러진 파형 봉우리를 기준으로 피치주기에 해당하는 봉우리를 찾는 방법은, 식 (6-1)의 상관관계값이 봉우리 정점마다 최상의 상관관계 봉우리를 이루는 원리를 적용하면 된다.

잔여파형에 대해 식 (6-2)의 상관관계를 계산하면, 파형 봉우리가 존재할 때마다 상관관계값이 양의 봉우리를 이루게 된다. 따라서 양의 상관관계 봉우리가 존재하는 정점의 구간은 예비피치로 고려하여 이들의 조합 ( $L_1, L_2, \dots, L_N$ )을 만들게 된다. 검출된 예비피치들의 조합에 대해 식 (6-1)의 상관관계식에 대입하여 최대의  $E(L)$ 를 이루는  $L_N$ 을 피치핀더의 피치값으로 설정하고, 피치핀더의 계수는 다음식으로 결정한다:

$$b_1 = E_{xy} / E_{yy} = \frac{\sum_{n=0}^{L-1} (s(n)s(n-L))}{\sum_{n=0}^{L-1} (s(n-L)s(n-L))} \quad (6-3)$$

이상의 과정으로 예비피치를 검출하는 데는 한 표본의 피치지연당 6번의 곱셈, 10번의 덧셈, 한번의 비교가 추가되지만, 식 (6-1)을 계산해야하는 예비피치들의 갯수가 적기 때문에 피치검색 시간이 많이 줄어든다.

검출될 수 있는 예비피치들의 갯수는 피치주기 사이에 나타나는 제 1 포먼트의 주파수에 관계된다. 제 1 포먼트의 주파수는 250Hz에서 750Hz 사이에 존재하기 때문에, 피치검색 구간에서 파형 봉우리가 가장 많이 나타나는 경우에는 750Hz/(8000/147)=13.78개 정도이다. 순차 피치검색법의 경우에는 식 (6-1)을 128번 수행해야 하지만, 이 방법은 간단한 선처리 연산만 추가함으로써 식 (6-1)의 연산은 14번 이하로 줄어들게 된다. 또한 14개 이상의 예비피치가 찾아지게 되면 이 경우에는 현재의 프레임이 무성음, 혼합음, 배경잡음 등으로 고려될 수 있기 때문에, 피치검색은 음성음원 경우에 의미 있으므로 예비피치의 갯수를 14개까지로 제한할 수 있다.

## 7. Positive ACF에 의한 피치검색 시간 단축

음성신호의 피치지연은 약 20ms 정도까지 정상상태를 거의 유지한다. 이러한 피치 지연의 성질을 이용한 것이 엘타 피치검색 구조이다[9]. 첫 번째(또는 세 번째) 부프레임에서는 20에서 147사이의 모든 피치 지연값 모두에 대해 피치검색을 수행한다. 두 번째(또는

네 번째) 부프레임에서는 첫 번째(또는 세 번째) 프레임의 피치지연 값을 중심으로 32표본 범위 안에 있는 피치 지연값들에 대해서만 피치검색을 실시한다. 이렇게 하면 이전 부프레임과의 상관관계들을 이용하므로 계산량과 부호화율이 감소(128번→32번)되고, 이전 부프레임의 피치주기가 연결되어 합성 음질이 우수하게 된다.

피치검색시 음성신호  $s(n)$ 에 대해 시간지연에 따른 상관관계치  $E(L)$ 를 다음과 같이 계산한다:

$$E(L) = \sum_{n=0}^{N_n-1} (s(n)s(n-L)) / \sum_{n=0}^{N_n-1} (s(n)s(n)) \quad (7-1)$$

여기서  $N_n$ 는 부프레임의 길이로 40표본을 나타내고,  $L$ 은 피치지연을 나타내며 피치 검색구간이다. 상관관계치  $E(L)$ 은 피치 주기마다 1에 근접한 값이 얻어지며, 유사한 정도가 어느 정도인지는 피치검색 구간 내에 있는 파형의 주기성과 파형 진폭의 변화에 따라 다르다. 그러나 피치의 주기성이 강한 음성음원 경우에는 이 값의 크기가 보통 0.9 ~ 1.1 정도가 얻어지며, 또한 시간지연에 대해 느리게 변화한다.

피치검색법에서 가장 바람직한 피치지연을 구하기 위해서는 식 (7-1)과 같은 상관관계식을 가능한 모든 피치지연에 대해 반복적으로 수행해야 한다. 이것은 매 피치지연  $L$ 에 대해  $L$ 의 범위는 20에서 147표본까지 곱셈과 덧셈을  $2N_n$ 번씩 수행해야 하는 많은 계산량이 된다. 이 때문에 CELP 코더를 가장 최근의 DSP칩으로 구현하여 50 SMIPS 이상이 소요되며 코더내 구현시간의 1/2 이상을 차지하게 된다. 따라서 피치 검색오차가 최소로 유지되는 상황에서 피치 검색 시간만 단축시킬 수 있는 알고리즘이 필요하게 된다.

피치 검색과정에서 최대의 상관값이 얻어지는 피치지연을 피치 주기로 고려하기 때문에 피치검색에 적용되는 상관관계값은 양의 봉우리를 이루는 경우만을 필요로 한다. 또한 상관관계의 특성상 양의 봉우리와 음의 봉우리가 반복되어 나타난다. 이러한 사실은 피치지연에 따라 양의 봉우리가 나타나면 그 봉우리가 끝나면 부분에서부터 양의 봉우리도 만들어진 음의 봉우리가 존재하는 구간에서 때문에 이 구간은 상관관계값을 계산할 필요가 없어진다. 즉, 지금까지 양의 상관관계값이 얻어지고 있고 이 구간이  $L_N$ 까운으로 추정되고 있을 때, 지금  $E(L)$ 이 영 이하로 얻어지게 되면, 다음에 검색할 피치지연 구간은 온 다음과 같다:

$$L = L_N + d \text{ samples} \quad (7-2)$$

여기서  $d$ 는 스킵-배율이며, 양의 봉우리쪽에 비해 건너 뛴 음의 봉우리 쪽을 설정하는 값이다. 이 값이 1에서부터 어느 정도 증가함에 따라 피치 검색구간은 감소되나, 표 7-1에서처럼  $d$ 가 1.5 이상이 되면 피치핀더의 예측오차가 급속히 감소되어 바람직하지 못하다. 따라서 이 값은 1.2정도가 적당하다.

표 7-1. 스킵-인자에 따른 검색시간율 및 피치핀더의 예측오차

Table 7-1 Searching time rate and the prediction gain of pitch filter according to the skip factors.

		Skip Factors					
		1.0	1.2	1.3	1.4	1.5	1.6
피치이득(dB)		11.63	11.60	11.28	10.61	9.75	8.05
검색시간(%)		77.2	70.9	68.9	67.5	66.7	63.5

상기와 같이 상관관계의 대칭성을 이용할 경우에 음성구간에서는 40% 이상의 계산시간 절감을 얻었다. 그렇지만 무성음 또는 묵음구간에서는 상관관계값이 양의 또는 음의 값으로 치우쳐 나타나기 때문에 상관관계의 대칭성을 이용할 수 없다. 즉, 이러한 경우에

## 상관관계 특성을 이용한 CELP보코더의 피치검색시간 단축법의 비교

는 전체 검색을 수행하는 경우에 비해 시간절감을 얻을 수 없게 된다. 따라서 이는 주기성이 나타나지 않는 경우이기 때문에, 피치 검색구간을 전체구간으로 잡지 않고 임의적으로 감소시킬 수 있다.

또한 유성을 구간에서도 지연에 따른 상관관계값이 영래백을 기준으로 배열이 된다면, 양의 상관관계로 고려되는 지연의 갯수  $L_c$ 는 전체 검색구간(=20에서 148까지)의 1/2을 넘을 필요가 없다. 여기에 양의 상관봉우리 폭보다 음의 봉우리폭은  $d=1.2$ 의 배율로 건너뛰기 때문에 이러한 사항을 고려하여 양의 봉우리로 고려되는 지연의 총 갯수는 다음과 같다:

$$L_c = (148-20)/(1+d) = 58 \text{ samples} \quad (7-3)$$

이렇게 하여 피치 검색구간을 상관관계가 양의 값으로 고려되는 구간으로 제한하면 피치 검색시간은 전체검색에 비해 다음과 같이 감소된다:

$$T_d = 58/128 * 1.05 * 100 = 47.6\% \quad (7-4)$$

여기서 시간계산에 105%를 곱한 것은 양의 상관관계를 카운트 할 때에 추가로 소요될 수 있는 시간을 포함시킨 것이다

## 8. 예비피치에 의한 피치검색시간의 단축

CELP 보코더에서 피치 검색과정은 잔이신호로 합성된 음성 신호가 원래 음성과 가장 유사하게 나타나는 피치지연값과 이때의 피치 이득을 구하게 되는데(1-3), 이때는 시간지연에 따른 상관관계가 최상인 경우를 찾으면 된다. 상관관계가 최상인 시간 지연을 찾기 위해서는 피치가 존재하는 영역을 차례차례 조사해 보아야 한다. 이러한 순차 피치검색법은 많은 시간이 소요되기 때문에, 전처리 단계에 의해 상관관계가 높게 나타나는 구간을 미리 파악하고, 이들 구간에 대해서만 본격적인 피치검색법을 수행하게 되면 피치 검색시간을 경감시킬 수 있다.

음성 신호의 피치는 음성 파형의 반복되는 봉우리에서 봉우리까지 또는 골에서 골까지로 정의된다. 파형의 봉우리 위치로 피치를 검출하는 경우에는 두드러진 봉우리가 존재하는 시간지연에 대해서만 자기 상관관계가 높게 존재한다. 또한 피치 주기는 2.5ms 이내에서 찾아지자 않기 때문에 피치검색에 적용되는 과정에 대해 다음과 같은 네치메이션을 수행하여 예비 피치를 구한다.

먼저 한 프레임은 19표본단위로 나누어 구간번호  $i$ 를 붙인다. 이때  $i$ 번째 19표본에 대해 최대 피치를 계산하여  $p(i)$ 에, 위치는  $p(i,0)$ 에 저장한다. 또한 최소의 밸리를 측정하여  $v(i)$  높이와 위치를  $v(i,0)$  및  $v(i,0)$ 에 각각 저장한다. 이렇게 피크와 밸리를 찾게 되면 음성신호의 제 3표본단의 위치변화에 따른 영향 때문에 예비피치가 몇 표본정도 오차를 가질 수 있다. 따라서 음성 신호에 대해 다음과 같은 해당필터를 수행한 후에 상기의 네치메이션을 수행하면 고차의 포먼트에 의한 영향을 제거할 수 있다.

$$s'(n-2) = \{s(n) + 2s(n-1) + 3s(n-2) + 2s(n-3) + s(n-4)\} / 9 \quad (8-1)$$

여기서 해당필터의 차단주파수는 2.67kHz이다. 검출된 피크와 밸리를 예비피치로 사용하기 위해서는 처음 찾아진 두드러진 피크(밸리)가준으로 그 다음의 피크(밸리)의 차이가 다음과 같은 구간이내에 있는 경우에만 식(3-1)의 본격적인 자기 상관관계를 수행해야 한다:

$$T_p(2i) = p\%(i, 0) - T_{hp} \quad (8-2) \\ T_v(2i+1) = v\%(i, 0) - T_{hv}, \quad i=1,2,\dots,12$$

여기서  $T_{hp}$ 는 두드러진 첫 피크의 위치를,  $T_{hv}$ 는 첫 밸리의 위치를 나타낸다. 검출된 예비피치들의 조합에 대해  $E(L) = E_{xy} / E_{yy}$  상관관계식에 대입하여 최대의  $E(T_p(i))$ 를 이루는  $T_p(i)$ 를 피치핀더의 피치값  $L$ 로 설정하고, 이 때 피치핀더의 계수는 다음식으로 결정한다:

$$b_L = E_{xy} / E_{yy} = \frac{\sum_{n=L}^{n-L} [s(n)s(n-L)]}{\sum_{n=L}^{n-L} [s(n-L)s(n-L)]} \quad (8-3)$$

이렇게 네치메이션을 수행하면 피크 및 밸리의 갯수가 19표본당 하나씩 찾아지게 되며, 피크와 밸리구간을 번들로 고려하여 예비 피치 구간을 각각 찾으면 검색시간은 순차 피치검색을 적용했을 때에 비해 다음과 같이 단축된다:

$$T_R = 2/19 * 105 = 11\% \quad (8-4)$$

여기서 계산시간에 5%를 더 고려한 것은 예비피치를 구하기 위해 네치메이션을 수행하는 시간이다.

## 9. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC/486DX2(50)이며 여기에 음성신호를 입출력하기 위한 AD/DA변환기를 인터페이스하여 8kHz의 표본율로 표본당 12비트의 데이터를 입출력하였다. 처리길이의 성능을 평가하기 위해 다음과 대표적인 문장은 20세의 여성, 22세의 남성, 28세의 남성 화자가 각 5원씩 작성하여 서로로 사용하였다.

- 문장 1: "원수대 모래는 전체수원을 좋아한다."
- 문장 2: "해수면에서 진자장치의 교훈을 받았습니다."
- 문장 3: "순정대 정보통신공학과 음성처리 연구실이다."
- 문장 4: "공인이검사우유전판지"

피치검색의 성능을 비교하기 위해 컴퓨터 시뮬레이션 과정은 크게 두 가지로 나뉘어 수행하였다. 우선 피치검색 구간(20~147표본까지)에 대해 피치지연  $L$ 은 1씩 증가하여 검색하는 순차 피치검색법으로, 일반적인 CELP 보코더의 피치 검색과 상을 C-언어로 구현하여 수행하였다. 여기서  $1/A(2N)$  보코더 여파기의 전달함수이고,  $A(2)/A(2n)$  위치가중화 여파기 응답이며, ZIR은 이전 상태의 영입력 여파기응답을 나타낸다. 그리고  $y_L(n)$ 은 피치지연  $L$ 에 대해 합성된 음성파형이다.  $E_{xy}$  원래음성과 합성된 음성의 교차 상관관계치이고,  $E_{yy}$ 는 합성된 음성의 자기 상관관계치이다. 피치검색에 대해 순차 처리한  $b-E_{xy}/E_{yy}$ 의 한 결과파형을 그림 9-6(b)에 나타내었다. 이 경우에 피치주기가 두드러진 상관관계가 얻어짐을 알 수 있다.

두 번째 과정은 피치검색의 시간단축을 위해 음성 표본들간의 상관관계 특성을 이용하여 우리가 제안했던 다음과 같은 방법들을 구현하여 그 특성을 조사하였다:

- 1) 검색구간표를 활용하는 방법(TABLE).
- 2) 상관관계의 대칭특성을 이용한 방법(SYMMET).
- 3) 전지리용 상관관계식을 이용한 방법(PREPC).
- 4) Positive Autocorrelation을 이용하는 방법(POS-AUTO).
- 5) 예비피치를 이용하는 방법(PRELD).

이상의 피치 검색과정은 처리단위로 그림 9-1, 2, 3, 4, 5에 각각 나타내었다. 각 단계에서 검출으로 나타낸 결과대의 처리과정은

피치 검색과정을 단순하기 위해 부가한 것이다. 점선 박스 내에서 처리의 결과인  $K_5$  값은 피치검색을 수행하지 않고 건너뛰는 구간을 나타낸다. 음성신호의 한 프레임구간에 대해 피치검색을 건너뛰는 구간을 상관관계값이 영이 되도록 표시하여, 각 방법으로 그림 9-6(c)~(g)에 나타내었다. 이 결과 과정에서 양의 상관관계 봉우리 특히 절두부분은 순차 검색방법의 경우와 모두 일치하고 있음을 알 수 있다.

순차 피치검색법에 대해 단순된 처리과정들의 피치검색 시간차이를 비교하기 위해, 상기 발생들에 대해 1초 단위의 평균 검색시간을 구하여 표 9-1에 제시하였다. 기존의 단순 피치검색법은 평균 12.2초가 소요되었고, 제안된 방법에 의해서: 평균 5.1초(TABLE), 7.4초(SYMMT), 2.9초(PREPC), 4.8초(POS-AUTO), 1.02초(PRELI)가 소요되어 약 58%, 약 40%, 약 77%, 약 59%, 약 87%의 시간절감이 각각 이루어졌다. 여기서 시간 측정치는 컴퓨터의 자중에 따라 다르기 때문에 상대적인 시간 단축율만을 평가에 고려하였다.

순차 피치검색법에 비해 처리시간은 단축되었지만 이때 나타나는 피치편타의 예측이득의 손실은 조사가기 위해 상기의 발생에 대한 프레임별 제 1번의 예측이득의 평균을 구하였다. 이러한 결과로 표 9-1에 함께 나타내었다. 각 방법은 평균 2.5%(TABLE), 2.1%(SYMMT), 6.1%(PREPC), 3.5%(POS-AUTO), 6.4%(PRELI) 정도의 예측이득이 저하되었다. 그렇지만 이는 값은 피치편타의 계수를 3비트(0과 2사이)로 0.25 레벨(단위)로 양자화할 때 발생하는 피치 예측이득의 저하가 평균 8.7%인 경우를 고려한다면 크게 문제되지 않으며, 코덱 검색단계에서 이러한 오차를 흡수하면 문제가 해결될 수 있다.

표 9-1. 시간단축율과 예측오차의 저하율

Table 9-1. Reducing the pitch search time and lowering the prediction of pitch

	시간단축율 (%)	예측이득의 저하율 (%)
TABLE	58	2.5
SYMMT	40	2.2
PREPC	77	6.1
POS-AUTO	58	3.5
PRELI	87	6.4

## 10. 결 론

CELP 부호화기: 음성신호를 합성하여 입력 음성신호와 비교하는 분석에 의한 합성법을 사용함으로써 음성이 매우 우수하다. 그렇지만 방대한 계산량 때문에 기존의 DSP로도 실시간 구현에 어려움이 있다. CELP 코덱에서 피치검색 시간은 부호화시간의 약 1/2을 차지하고 있다. 이 때문에 피치검색 시간을 단축하는 것이 중요하다고 생각되어지고 있다. 그렇지만 이러한 피치검색을 단순화시키려는 모든 피치 예측이득의 저하를 동반하게 된다. 따라서 본 논문에서는 우리가 이미 제안했던 피치 검색시간 단축법 다섯 가지에 대한 비교평가를 수행하였다.

비교평가에 적용한 방법은 불필요한 피치검색구간을 건너뛰기 위해 구간표준치 설정하는 방법, 상관관계 대칭성을 이용하는 방법, 선지리용 상관관계식을 사용하는 방법, 예비피치값 사용하는 방법이 적용되었다. 이들의 피치검색 시간단축율은 각각 58%, 40%, 77%, 59%, 87%를 차지한 반면, 피치 예측오차율은 2.5%, 2.1%, 6.1%, 3.5%, 6.4%를 각각 순서대로 차지하였다. 그렇지만 이는 값은 피치편타의 계수를 3비트(0과 2 사이)로 0.25 레벨(단위)로 양자화할 때 발생하는 피치 예측이득의 저하가 평균 8.7%인 경우에 비해 높지않기 때문에 피치검색을 단순화시키더라도 손색없이 적용될 수 있다.

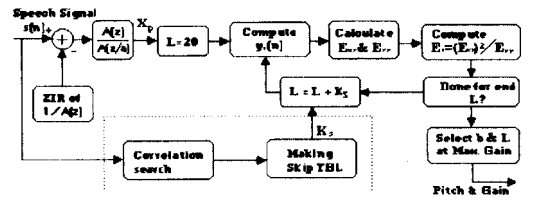


그림 9-1. TABLE법 피치 검색과정

Fig. 9-1 The pitch search algorithm by TABLE method.

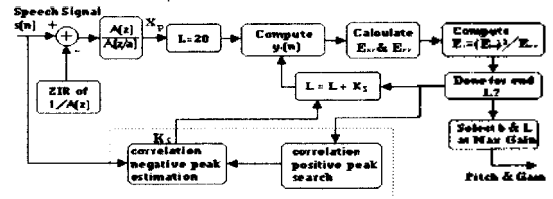


그림 9-2. SYMMT법 피치 검색과정

Fig. 9-2 The pitch search algorithm by SYMMT method.

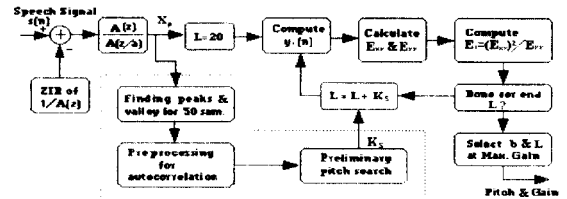


그림 9-3. PREPC법 피치 검색과정

Fig. 9-3 The pitch search algorithm by PREPC method.

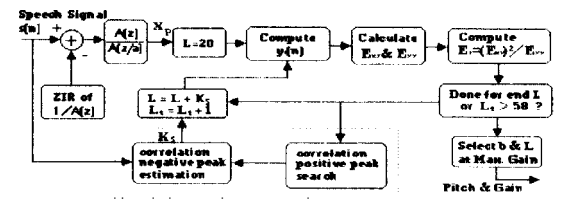


그림 9-4. POS-AUTO법 피치 검색 과정

Fig. 9-4 The pitch search algorithm by POS-AUTO method

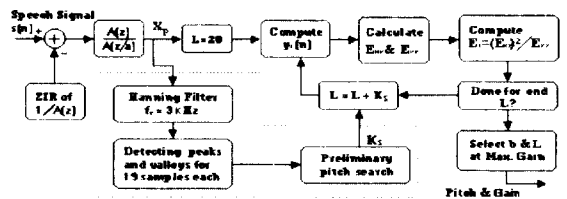


그림 9-5. PRELI법 피치 검색 과정

Fig. 9-5 The pitch search algorithm by PRELI method.

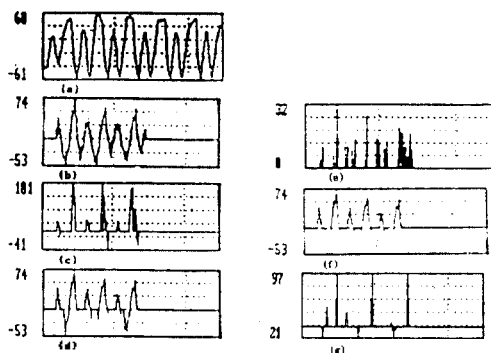


그림 9-6 발성1에 대해 처리한 결과

- (a)음성신호파형, (b)순차 피치검색 결과파형,  
(c)TABLE법 검색 결과파형, (d)SYMMT법 검색 결과파형,  
(e)PREPC법 검색 결과파형, (f) POS-AUTO법 검색 결과파형  
(g) PRELI법 검색 결과파형.

Fig.9-6 A result for utterance 1.

## 참고 문헌

- [1] J.D. Markel and A.H. Gray, Jr., Linear Prediction of Speech, Springer Verlag, New York, 1976.
- [2] L.R. Rabiner and R.W. Schafer, Digital Processing of Speech Signal, Prentice-Hall, 1978.
- [3] A.N. Ince, Digital Speech Processing( speech coding, synthesis, and recognition), Kluwer Academic Publishers, 1992.
- [4] M.R. Schroeder and B.S. Atal, "Code-Excited Linear Prediction (CELP): High-Quality at Low Bit Rates", Proc. Int.Conf. on Acoust., Speech and Signal Processing, pp.25.1.1-25.1.4, 1985.
- [5] G. Davidson and A. Gersho, "Complexity Reduction Methods for Vector Excitation Coding", Proc.Int.Conf. on Acoustics, Speech and Signal Processing, 1986.
- [6] R.C. Rose and T.P. Barnwell III, "Quality Compression of Low Complexity 4800 bps Self Excited and Code Excited Vocoders", Proc.Int.Conf. on Acoustics, Speech and Signal Processing, 1987.
- [7] A.Le Glayader, D. Massaloux, and J.P. Petit, "Robust and Fast Code-Excited Linear Predictive Coding of Speech Signals", Proc.Int.Conf. on Acoustics, Speech and Signal Processing, 1989.
- [8] J. Menez, C. Galand, M. Rosso, and F. Bottau, "Adaptive Code Excited Linear Predictive Codet(ACELPC)", Proc.Int.Conf. on Acoustics, Speech and Signal Processing, 1989.
- [9] J.P. Campbell, Jr., V.C. Welch, and T.E. Tremain, "An Expandable Error-Protected 4800 bps CELP Codet(U.S.Federal Standard 4800bps Voice Codet)", Proc.Int.Conf. on Acoustics, Speech and Signal Processing, 1989.
- [10] W.B. Kleijn et al., "Fast Methods for the CELP Speech Coding Algorithm", IEEE Trans.Acoustics, Speech and Signal Processing, Vol.38, No.8, pp.1330-1341, Aug.1990.
- [11] R.C. Rose and T.P. Barnwell, "Design and Performance of an Analysis-by-Synthesis Class of Predictive Speech Coders", IEEE Trans.Acoustics, Speech and Signal Processing, Vol. 38, No.9, pp.1489-1503, Sep.1990.
- [12] 배명진, "CELP 코더에서 피치검색법의 개선", ETRI-CAD seminar, 1993년 5월 18일.
- [13] 임재원, 배성근, 배명진, 인수진, "상관관계 특성을 이용한 CELP 코더의 피치 검색시간 단축에 관한 연구", ETRI-CAD seminar, 1993년 7월 27일.
- [14] 이 해근, 김 홍, 배 명진, "신자리에 의한 CELP 코더의 피치 검색시간 단축에 관한 연구", ETRI-CAD seminar, 1993년 8월 18일.
- [15] 장 진남, 배 성근, 배 명진, 김 형래, 한 기환, 변 경관, 유 하영, "Positive ACF를 이용한 CELP 코더의 피치검색시간 단축에 관한 연구",Proc. of KITE Summer Conference '94, pp.291-294, VOL.17, NO.1, 1994.
- [16] 김 대식, 김 홍, 배 명진, 한 진인, 김 종래, 유 하영, "예비피치에 의한 CELP 코더의 피치 검색 시간의 단축에 관한 연구", Proc. of KITE Summer Conference '94, pp.285-298, VOL.17, NO.1, 1994.