

논문 2012-49-10-11

# 보코더에서 서브프레임 수의 변화를 이용한 피치검색 성능 개선에 관한 연구

( A Study on Improving Pitch Search by Varying the number of  
Subframes for Vocoder )

백 금 란\*, 배 명 진\*\*

( Geum-Ran Baek and Myung-Jin Bae )

## 요 약

보코더에서의 피치 검색 과정은 매우 중요하다. 일반적으로 먼저 신호의 주기성을 강조한 후 피치를 검색하는 방법을 주로 사용한다. 프레임 내에서 두 개의 펄스 간격을 변화시켜 가며 신호와의 상관관계를 구하는 것으로 상관관계 값이 가장 높을 때가 주기가 가장 두드러지는 반복 구간이므로 이때의 펄스 간격을 피치 주기로 찾는다. 그러나 프레임 내에서의 신호의 진폭이 일정하지 않고 갑자기 변하는 구간이 있는 경우 정확하지 않은 피치가 얻어진다. 이를 해결하기 위해 프레임을 여러 개의 서브프레임으로 다시 나누어 피치를 검색하는 방법을 사용하고 있는데 이 방법은 피치를 정확하게 검색할 수 있지만 계산량이 많아진다. 본 논문에서는 프레임의 에너지 비율로 프레임 내의 진폭 변화율을 예측하여 서브프레임 수를 가변시키는 방법을 제안한다. 이 방법들을 적용하면 합성 음질에는 영향을 주지 않는 상태에서 피치검색 시간을 단축할 수 있고, 피치 검색의 정확도를 높일 수 있어 전반적인 피치 검색에 관한 성능 개선이 된다.

## Abstract

The pitch searching is a very important process in a vocoder. Generally, the method of pitch searching method is used by highlighting the periodicity, where a correlation is identified with the signal by changing the interval of two pulses. When the correlation value is highest, the pitch can be found by the pulse interval because it is the repetition interval with most striking period. There are many methods to solve this problem and search the pitch by dividing a frame into many subframes, but there is too much calculation to solve. A method in this paper is suggested to vary the number of subframes by predicting the amplitude change rate in a frame. If this method is applied, the general pitch searching performance will be improved because the accuracy may be enhanced without affecting the sound quality in the synthesized signal after parameter transmission; and the pitch searching time may be reduced.

**Keywords :** vocoder pitch search, correlation, subframe, energy change ratio

## I. 서 론

음성신호를 메모리에 저장하거나 전송하기 위한 음성 부호화법에는 크게 과형부호화법, 신호원부호화법,

혼성부호화법 등 세 가지로 나눌 수 있다. 과형부호화법은 음성의 성분 분리 없이 과형 자체의 잉여 성분만을 제거한 후 부호화하여 전송하고 다시 합성하는 방식으로 고음질과 화자의 개성이 유지되는 반면 전송과형을 유지하기 위한 데이터양이 많기 때문에 전송률이 높으며 대용량의 메모리가 필요하게 된다<sup>[1]</sup>. 이에 비해 신호원 부호화법은 음성의 발성모델에 근거하여 음성 신호의 여기 성분과 여파기 성분을 분석하여 각각을 독

\* 정회원, \*\* 정회원-교신저자, 송실대학교 정보통신학과

(Dept. of Information Telecommunication Engineering, Soongsil University)

접수일자:2012년6월14일, 수정완료일:2012년9월21일

립적으로 분리시켜 부호화하는 방법을 사용하기 때문에 전송 대역폭이 작고 메모리 사용량이 적다. 그러나 분석 및 합성 시에 오차가 누적되어 명료성이 떨어진다는 단점을 갖고 있다<sup>[1]</sup>.

혼성 부호화화법은 고음질의 화자의 개성을 유지시켜 주는 과형 부호화법과 메모리 효율성이 높은 신호원 부호화법을 결합시킨 방법이다. 즉 포먼트 정보는 선형 예측 부호화법으로 부호화하고 나머지 잔여 신호를 어떻게 부호화하느냐에 따라 RELP, VELP, MPLPC, CELP법 등이 제안되어져 있다<sup>[1~2]</sup>. 음성이 선형예측을 통해 얻어 질 수 있다는 가정으로 개발된 CELP(Code Excitation Linear Prediction)계열의 음성부호기들은 음성을 부호화하는 데 있어서 선형예측 계수와 여기신호로 이루어진다. 일반적으로 선형 예측 계수는 LSP를 이용하여 부호화되며 여기신호는 몇 개의 코드북을 이용하여 부호화한다. CELP를 기반으로 발전한 부호화 기법으로는 ACELP, CS-CELP 등이 있다. 또한 CELP 계열의 압축 방식은 저 전송률에서 좋은 통화 품질을 제공해 준다<sup>[1~2, 7]</sup>.

앞에서 살펴 본 CELP보코더의 과정 중 피치를 검색하는 과정은 코드북 검색과 함께 매우 중요한 과정이다. CELP부호화법에서는 피치필터를 적용하여 음성신호의 피치 주기 성분을 부호화하고 있다. 이러한 피치 필터에 주로 적용되고 있는 피치검색법은 피치지연에 따른 상관관계법이다. 상관관계 피치검색법은 프레임 단위로 피치가 존재하는 모든 피치지연에 대해 원래음성과 합성된 음성의 상관관계를 검색하고, 최상의 상관관계를 갖는 피치필터의 피치지연과 이득을 결정하게 된다. 또한 좀 더 정확한 계수 값들을 얻기 위해 서브프레임 단위로 수행하는데 이는 많은 피치 검색 시간을 필요로 한다<sup>[2]</sup>.

본 논문에서는 음성 신호의 진폭변화에 따라 각 프레임마다 서브프레임 수를 달리하여 피치를 구하는 방법을 제안한다. 이 방법으로 피치 검색의 전체적인 성능을 그대로 유지하면서 피치 검색 시간을 단축시켜 주어 전송하고자하는 음성신호의 파라미터들을 코딩하는 데도 용이하게 해준다. 먼저 II장에서 CELP계열의 보코더에서의 피치 검색 알고리즘을 살펴보고 본 논문에서 제안하는 알고리즘을 III장에서 설명한다. IV장과 V장에서는 실험 및 결과를 살펴보고 결론을 맺는다.

## II. 기존의 피치 검색 방법

### 1. CELP계열의 보코더에서의 피치 검색

보코더 부호화시에 수행되는 피치 검색과정은 음성 신호의 자기 상관관계에 해당하는 피치 주기 정보를 얻는 과정이다. 스펙트럼분석이 개회로 구조로 수행되어지는 것에 비해 피치분석은 폐회로 구조에 의해 수행되어져야한다. 즉, 피치 지연조건을 최적으로 만족하는 값을 반복적인 비교를 통해 결정하게 된다. 피치 이득은 이때 얻어진 피치지연을 이용하여 상대적인 이득 값을 갖도록 양자화하게 된다. 이러한 피치 검색과정은 코드북 검색과 함께 CELP 부호화기의 계산량에 크게 영향을 미치는 중요한 부분이다<sup>[2]</sup>.

그림 1은 전형적인 피치검색 과정 중 한 과정만을 보여 주고 있다. 피치검색은 합성에 의한 분석을 기초로 하고 피치 예측 필터를 위한 입력신호와 합성된 신호사이의 가중화된 에러를 최소화 하는 피치지연 값인  $L$ 과 피치이득 값인  $b$ 를 선택하는 과정이다. 그림 1에서 ZIR(Zero Input Response)은 영(zero) 입력 응답이고 피치검색을 더욱 효과적으로 하기위해 가중여파기,  $W(z)$ 를 사용하는데 식 (1)과 같다.

$$W(z) = \frac{A(z)}{A(z/\alpha)} \tag{1}$$

여기서  $A(z)$ 는 포먼트 예측 오차 필터이고  $\alpha$ 는 인자 가중화된 파라미터이다. 피치 합성 필터는 다음 식 (2)와 같이 나타낼 수 있다.

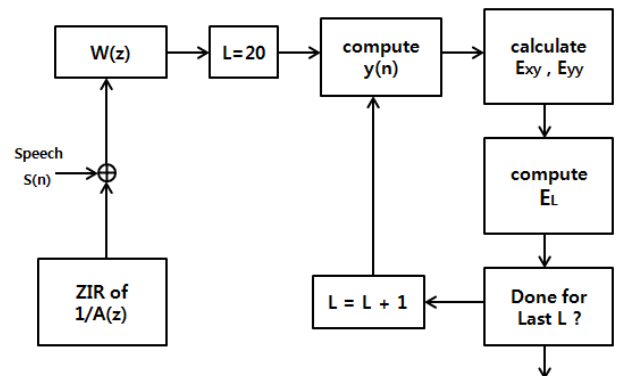


그림 1. 피치 검색을 위한 구현 과정의 예  
 Fig. 1. An example of implementation flow for pitch search.

$$\frac{1}{P(z)} = \frac{1}{1 - bz^{-L}} \quad (2)$$

가장 최적의 피치지연  $L$ 과 피치 이득  $b$ 를 구하는 과정은 식 (3)의 MSE(Mean Square Error)를 최소화하는  $L$ 과  $b$ 을 구하는 것과 같다.

$$\begin{aligned} MSE &= \frac{1}{L_p} \sum_{n=0}^{L_p-1} [x(n) - by_L(n)]^2 \\ &= \frac{1}{L_p} \sum_{n=0}^{L_p-1} [x(n) - by_L(n-L)]^2 \end{aligned} \quad (3)$$

여기서  $L_p$ 는 피치분석 프레임 길이이고  $x(n)$ 과  $y(n)$ 은 각각 입력신호와 인지 가중화된 합성 신호이다. 이것은 식 (4)의 최대값을 구하는 것과 같다.

$$E_L = \frac{(E_{xy})^2}{E_{yy}} \quad (4)$$

여기서  $E_{xy} = \sum_{n=0}^{L_p-1} x(n)y_L(n)$   
 $E_{yy} = \sum_{n=0}^{L_p-1} y_L(n)y_L(n)$  이다.

주어진  $L$  값에 대해 최적의  $b$ 는 식 (5)와 같이 구한다. 이러한 과정은 각각의 모든  $L$ 값에 대해 반복적으로 수행된다<sup>[3]</sup>.

$$b_L = \frac{E_{xy}}{E_{yy}} \quad (5)$$

QCELP의 피치검색 알고리즘을 자세히 살펴보면 1/8 율의 패킷으로 부호화되는 프레임을 제외하고는 모든 보코더의 프레임은 동일한 길이의 피치 부프레임으로 재차 분할된다. 즉 전송율 1 패킷에 대해서는 피치 프레임이 4개, 전송율 1/2 패킷에 대해서는 피치 프레임이 2개, 전송율 1/4 패킷에 대해서는 피치 프레임이 1개가 있고 전송율 1/8 패킷에 대해서는 피치 프레임이 없다. 피치 합성필터는 식 (2)와 같고 피치지연은 7비트로 표현되어 17에서 143까지의 범위를 갖는다.  $L = 16$ 은  $b = 0$ 인 경우에 대해 사용한다. 피치 이득  $b$ 는 3비트로 0에서 2.0까지의 범위를 갖는다. 각 부프레임에 대해 음성 부호화기는 피치지연  $L$ 과 피치이득  $b$ 를 결정하고 부호화하게 된다.

피치 파라미터를 선택하는데 적용되는 방법은 합성에 의한 분석법이며 부호화는 입력음성과 합성된 음성간의 가중된 오차를 최소화하는 파라미터를 선택하는 과정이다. 합성된 음성은 LPC 필터에 의해 여파된 피치 합성필터의 출력이다. 피치합성은 입력으로 모두 영값 요소로 구성된 코드북 벡터를 사용한다. 피치지연  $L$ 은 집합{ 17, 18, ....., 143 }에서 선택되고 피치이득  $b$ 는 집합 { 0, 0.25, 0.5, ..., 2.0 } (0과 2.0 사이를 0.25 간격으로 선형적으로 양자화)에서 선택한다. 가중화 여파기는 식 (1)과 같고  $\alpha$ 는 인지 가중화된 파라미터로 0.8이다. 가중화 필터에 적용된 LPC계수는 현 피치 부프레임에 대한 것과 같다. 음성부호화기에 적용된 합성필터는 식 (6)과 같다.

$$H(z) = \frac{1}{A(z)} \cdot W(z) = \frac{1}{A(z/\alpha)} \quad (6)$$

여기서 인지가중화 필터가 뒤따르는 복호기 음성 합성필터이고 이를 가중화된 합성필터라 한다<sup>[2]</sup>.

## 2. 듀얼펄스에 의한 피치 검색

일반적으로 피치를 검색하기 전에 그림 2와 같이 신호의 주기성을 먼저 강조한 다음 그 신호에서 두드러진 반복 구간을 찾는 방법을 사용한다.

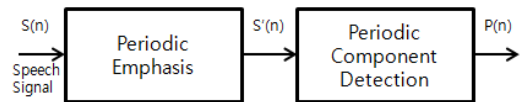


그림 2. 피치 검색 과정  
Fig. 2. Pitch Detection.

자기상관법에 속하는 듀얼펄스(Dual Pulse) 기법은 두 개의 임펄스 신호 쌍과 원래신호를 상관관계를 구하여 가장 두드러진 반복구간을 찾는다. 그러므로 신호의 레벨은 중요하지 않고 간격이 중요하다. 즉 한 프레임 내에서 식 (7)과 같은 전달특성을 갖는 두 개의 펄스와 원래 음성신호와의 가장 높은 상관관계 값을 구하여 주된 주기를 찾는 방법이다<sup>[4~5]</sup>.

$$H(z) = 1 - bz^{-L} \quad (7)$$

여기서  $b$ 는 필터의 이득이며  $L$ 은 두 펄스간의 간격이다.

$$\Phi_i = \sum_{n=0}^{N-1} s(n) * b_i s(n - L_i) \quad (8)$$

(단,  $N$ 은 프레임사이즈)

식 (8)을 이용해서 프레임 내에서 샘플을 증가시켜가며 원래 신호와 두 펄스 간격만큼 지연된 신호 사이의 상관관계(Correlation)를 계산한다. 여기서  $b_i, L_i$ 는 각각  $i$  번째 프레임의 이득과 펄스 간격이다. 이 식에서 얻어진 파형에서의 두드러진 피크와 피크 간격 중에 첫 번째로 찾아지는 가장 큰 값을 피치 주기라 한다. 그러나 이 경우에 찾아진  $L_i$ 값이 반드시 피치라는 보장이 없다. 이유는 이 간격이 피치 주기가 아닌 반주기, 배주기 및 세배주기일 수 있다.

반주기나 배주기가 피치 주기로 찾아지는 것을 막기 위해 그림 3에서 나타냈듯이 처음에  $b$ 을 1로 놓고  $L_i$ 값을 증가 시켜가며 원래 신호와  $H(z)$ 의 상관관계를 구한다. 이때 유사도가 높은 값들 사이의 비율을  $b$ 값으로 한다. 두 번째 단계로 이렇게 얻은  $b$ 값을 다시  $H(z)$ 에 대입하여  $L_i$ 값을 증가 시켜가며 원래 신호와  $H(z)$ 의 상관관계를 구한다. 이때 상관관계 값이 가장 높을 때의  $L_i$  값을 피치 주기로 한다. 이렇게 하면 반주기, 배주기 및 세배주기가 피치 주기로 찾아지는 것을 막는다. 그러나 신호의 진폭 변화가 크게 없는 프레임에서는 정확한 계수를 얻을 수 있지만 갑작스런 신호의 변화가 있는 구간

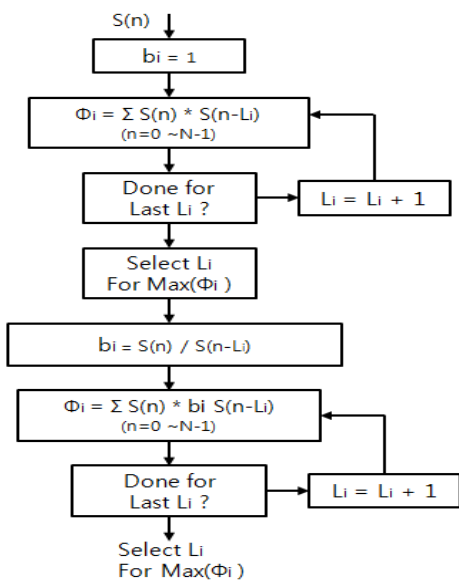


그림 3. Dual Pulse에 의한 피치 검색 과정  
Fig. 3. Pitch search by using Dual Pulse.

에서는 신뢰할 수 없는 피치를 얻는다. 이러한 문제점을 해결하기 위해 분석구간을 더욱 짧게 할당하는 방법을 적용한다. 즉, 프레임을 다시 서브프레임으로 세세하게 나누어 분석 하는데 이 방법은 피치를 검색하는데 필요한 시간을 증가시킨다.

### III. 제안한 알고리즘

보코더에서 정확한 피치 주기와 이득을 검색하기 위해 프레임을 두 개에서 네 개의 서브프레임으로 좀 더 세세하게 나누어 처리하는 이유는 음성 신호 중에 급격한 변화 구간에서의 피치 검색의 정확도를 증가시키기 위함이다. 그림 4에서처럼 진폭의 변화가 일정한 음성 신호 구간에서는 피치 변화 또한 거의 없다는 것을 실험을 통해 알 수 있었다. 즉, 파형의 변화가 갑자기 발생하는 전이구간에서는 피치 주기 또한 변화한다는 것을 알 수 있다. 그림 4는 진폭 변화가 거의 없는 음성 파형의 피치를 점으로 나타내었고 기본주파수도 보여 주고 있다. 그림에서 알 수 있듯이 신호의 진폭 변화가 거의 없는 구간에서는 피치 변화도 거의 없이 얻어짐을 알 수 있다.

본 논문에서는 프레임내의 에너지 변화율을 이용하여 진폭 변화의 상태를 추정된 후에 서브프레임의 간격을 결정하는 방법을 제안한다. 먼저 피치를 검색하기 전에 프레임 앞부분 일정 간격 동안의 평균에너지인  $E_{1av}$ 과 뒷부분 일정 간격 동안의 평균에너지인  $E_{2av}$ 의 평균에너지를 각각 구한다. 식 (9)에서처럼 프레임 앞뒤에서 구한 에너지의 값을 이용하여 에너지비율인  $i$  번째 프레임의  $a_i$  값을 정한다.

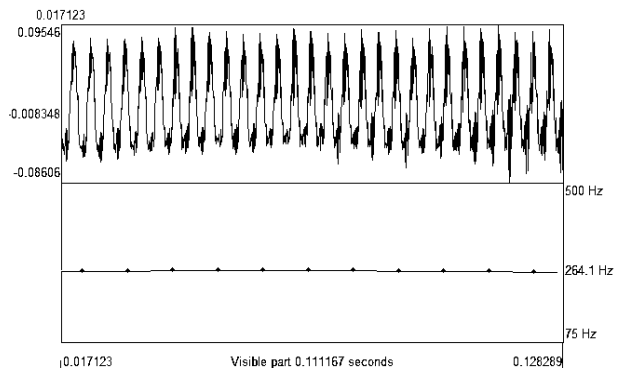


그림 4. 전이구간이 없는 음성과 피치  
Fig. 4. The stationary speech signal and pitch.

$$a_i = \sqrt{\frac{E_{1av}}{E_{2av}}} \quad (9)$$

이때  $a_i$  값이 1에 가까우면 신호의 진폭 변화가 없는 경우이므로 서브프레임 수를 반으로 줄이거나 프레임 단위로 분석을 하는 결정 논리를 추가한다. 그림 5에서 본 논문의 알고리즘 과정을 그림으로 설명하였다. 피치를 검색하기 전에 먼저 프레임의 에너지 비를 구한 후 서브프레임의 수를 결정하는 과정을 보여주고 있다.

이 방법은 피치의 정확도를 높이기 위해 서브프레임 단위로 피치 검색하는 과정을 수정하여 계산량을 줄이는데 그 목적이 있다. 즉 프레임의 에너지비율을 미리 계산하여 음성신호의 상태를 미리 추정하여 서브프레임수를 가변시켜 가며 피치주기를 구한다. 그 결과 피치검색 시간을 줄일 수 있고 이를 이용하여 다른 파라미터들을 추출하는데 시간을 더 할당할 수 있어 전체적인 보코더의 성능을 개선할 수 있으리라 기대된다.

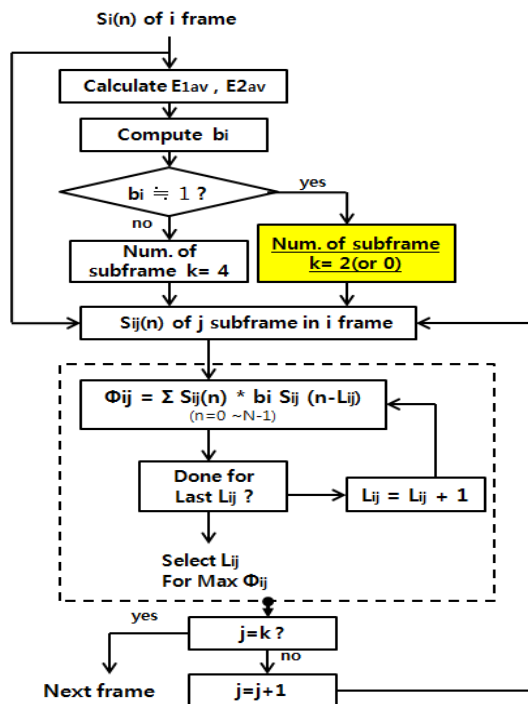


그림 5. 제안한 알고리즘의 블록도  
Fig. 5 Block diagram of proposed algorithm.

#### IV. 실험 및 결과

본 논문에서 제안한 방법을 실험하기 위해 시물레이

션 장비로는 워크스테이션 Z400을 사용하였고 음성 시료는 8 kHz로 샘플링 시켰으며 16 bit A/D 컨버터로 디지털 신호로 변환된 20대 남/녀의 음성을 이용하였다. 음성 시료로는 '아, 예, 이, 오, 우'를 약 0.5초 간격으로 발음하게 한 것과 약 3초에서 10초 사이에 발음한 문장을 5개 사용하였다. 프레임의 간격은 160 샘플로 정하였고 서브프레임 길이는 40샘플로 하였다. 식 (9)의 값 즉,  $i$  번째 프레임의 이득인  $b_i$ 를 구하기 위해 프레임의 앞과 뒤 파형의 에너지 비를 구할 때는 각각의 프레임 시작과 끝부분으로 부터 일정 간격인 10 msec 만큼의 에너지를 구하여 계산하였다. 또한 제안한 알고리즘의 성능을 평가하기 위해 피치 검색 과정을 Visual C와 MatLab 프로그램으로 구현하여 기존의 QCELP 보코더에서의 피치 검색 방법과 본 논문에서 제안한 방법을 비교하였다. 그림 6은 QCELP 보코더에서 얻어진 합성음과 제안한 방법으로 얻어진 합성음의 파형을 나타낸 것이다.

처리시간 측정은 수행시간 측정함수를 사용하였고 각각의 시료에 대해 QCELP 알고리즘과 제안한 알고리즘에 대한 처리시간을 측정하였다. 또한 각 시료에 7번씩 수행한 다음 최대치와 최소치를 뺀 나머지 5개 값에 대해 평균을 측정하였다. 실험 후 표 1에서 나타난 것처럼 QCELP에 비해 제안한 방법은 26.84% 만큼의 처리 시간이 감소되었음을 알 수 있었다. 여기서 음성 시료 1과 시료 2는 평균값에 비해 낮게 나타나고 시료 3,

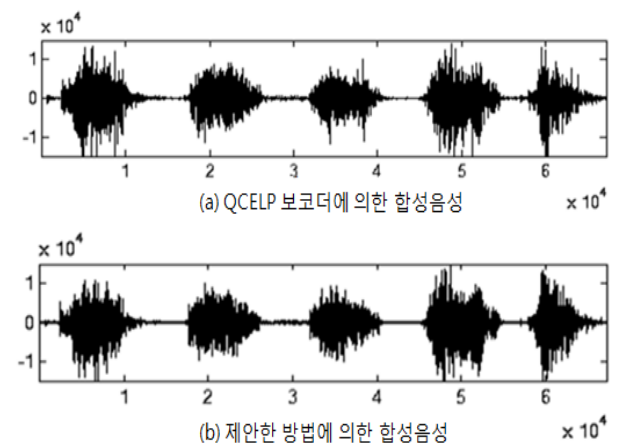


그림 6. QCELP 보코더에서의 합성음성과 제안한 방법으로 얻어진 합성음 파형의 비교

Fig. 6. Comparison between synthesis signal by QCELP and synthesis signal by proposed method.

표 1. 처리시간 결과 비교  
Table 1. Result of the process time.

	QCELP (단위sec)	제안한 알고리즘 (단위sec)	감소율 (단위 %)
시료 1	45.049	33.610	25.39
시료 2	49.378	36.668	25.74
시료 3	2.130	1.557	26.90
시료 4	6.960	4.975	28.52
시료 5	4.050	2.930	27.65
평균			26.84

시료4 그리고 시료 5는 높은 이유는 음성 시료 1과 시료 2에 비해 시료 3, 시료 4 그리고 5의 피치의 변화가 적기 때문이다.

### V. 결 론

음성신호처리 분야에서 피치를 검색하는 과정은 중요하다. 피치 검색의 정확도를 높이기 위해 프레임들 다시 나누어 서브프레임 단위로 피치를 검색하는 방법을 사용한다. 그러나 이 방법은 피치를 정확하게 검색할 수 있지만 계산량이 많아진다는 단점이 있다. 본 논문에서는 미리 프레임 내 에너지 비를 구하여 이를 이용하여 전체 음성의 진폭 변화를 추정하여 그 변화에 따라 서브프레임 수를 조절하는 방법을 제안한다. 이 알고리즘들을 적용하면 각 프레임별 피치검색 시간을 26.84% 단축할 수 있다. 즉 이 방법으로 합성음의 음질은 그대로 유지하면서 피치 검색 시간을 단축시켜 주고 전송하고자 하는 음성신호의 파라미터들을 코딩하는 데도 용이하게 해주어 전반적인 보코더 성능 개선에도 기여할 수 있다.

### 참 고 문 헌

[1] M. J. Bae, Digital Speech Coding, p17-29, p30-48, pp49-68 Dongyoung, 1996.  
 [2] M. J. Bae, "On a Performance Comparison of Pitch Search Algorithms by using a Correlation Properties for the CELP Vocoder", In ASK, the proceeding Vol. 11, pp. 165-171, 2004.  
 [3] Wang-Rae Jo, Seong-Young Choi, MyungJin

Bae, "A Study on Pitch Search Time Reduction of G.723.1 Vocoder by Improved Hybrid Domain Cross-correlation" Trans. KIEE. Vol.59, No.12, DEC, 2010.

[4] JooHun Lee, MyungJin Bae, SouGil Ann, "A Fast Pitch Searching Algorithm Using Correlation Characteristics in CELP Vocoder", The Journal of ASK, Vol. 13. No. 2E, 1994.  
 [5] JaeOk Bae, KiChun Han, JinJong Cha and MyungJin Bae, "On a Fast Searching of the Random Codebook using the Searching Technique of the Regular Pulse Excitation," International Conference on Speech Processing, Proceedings of ICSP'97, pp.281-286, August 26-28, 1997.  
 [6] J. S. Han, Speech Processing, Osung Media, pp.5- pp7, pp103. 2003년.  
 [7] L. R. Rabiner, R. W. Schafer, Digital Processing of Speech Signal, pp.131-149.  
 [8] A. M. Kondo, Digital Speech coding for low bit rate communications systems, John Wiley & Sons, new York, 2004.

### 저 자 소 개



백 금 란(정회원)  
 1994년 2월 숭실대 전자공학과 (공학사)  
 1999년 2월 KAIST 전지 및 전자공학과 (공학석사)  
 2012년 2월 숭실대 정보통신공학과(박사수료)  
 2010년 3월~2011년 12월 숭실대전산원 교수  
 2012년 1월~현재 숭실 호스피탈리티 교수  
 <주관심분야 : 정보통신, 신호및 신호처리, 음성 신호처리>



배 명 진(정회원)-교신저자  
 1986년~1992년 호서대학교 전자공학과 조교수  
 1992년~현재 숭실대학교 정보통신공학과 정교수  
 <주관심분야 : 음성신호처리, 신호 및 신호처리 >