

# SAMDF를 이용한 IMBE VOCODER의 피치 검색 시간 단축에 관한 연구

홍성훈, 김종광, 배명진

송실대학교 정보통신공학과

## On a Reduction of Pitch Search Time for IMBE Vocoder by Using the Spectral AMDF

SeongHoon HONG, JongKwang KIM, MyungJin BAE

Dept. of Telecomm. Engr., Soongsil Univ. Seoul 156-743, Korea

sublime@assp.soongsil.ac.kr, mjbae@saint.soongsil.ac.kr

### ABSTRACT

IMBE(Improved Multi-Band Excitation) vocoders exhibit good performance at low data rates. The major drawback to IMBE coders is their large computational requirements. In this paper, thus, we propose a new pitch search method that preserves the quality of the IMBE vocoder with reduced complexity.

The basic idea is to reduce computation complexity of the pitch searching by using the SAMDF. Applying the proposed method to the IMBE vocoder, we can get approximately 52.02% searching time reduction in the pitch search. There is no difference in voice quality between conventional IMBE and proposed IMBE.

### 1. 서론

음성신호를 메모리에 저장하거나 전송하기 위한 음성 부호화법에는 크게 파형부호화법, 신호원 부호화법, 혼성부호화법이 있다[1]. 이들 중에서 혼성부호화법은 신호원 부호화법의 메모리 효율성과 고음질의 파형부호화법을 결합시킨 것이다. 혼성 부호화법에서 포먼트정보는 선형 예측 부호화법으로 부호화하게 되고, 그 나머지 잔여신호를 어떻게 부호화 하느냐에 따라 RELP법, VELP법, MPLPC법, CELP법 등이 제안되어져 있다. 혼성 부호화법들 중에서 주파수 영역에서 음성을 다루

고 있는 IMBE (Improved Multi-Band Excitation) 법 또한 최근 이동 및 휴대통신용 부호화법으로 채택되어 사용되고 있다[2]. 그림 1-1은 IMBE형 보코더의 한 블록도를 나타낸다.

IMBE 음성부호화기는 MBE 음성 모델에 기반한 음성 부호화기로 주파수 영역에서 피치 예측을 하여 기존의 CELP 계열 음성 부호화기에 비해 상당히 자연스러운 음성의 합성이 가능하다. 기존의 음성 부호화기는 각각의 프레임을 유성음이나 무성음으로 판별하여 유/무성음이 섞여 있는 혼합영역의 특성을 살려주지 못한다. 그러나 IMBE 음성부호화기는 신호를 주파수 대역으로 변환한 후 대역분할 하여 보다 고음질의 음성합성이 가능하므로 자연스러운 음성합성을 할 수 있다. 그러나 이를 위해서는 보다 정교한 피치검색이 요구되어 복잡한 피치 탐색 과정을 필요로 하므로 음성 처리시 많은 시간이 소요된다[2]. 즉, 시간영역에서 초기피치검색과정에서 음성의 피치가 천천히 변하는 성질을 이용하여 자연성을 줄 수 있으며, 이를 위하여 현재 프레임의 앞, 뒤에서 피치추정과정을 수행하여 초기피치를 찾게 된다[4]. 이렇게 찾아진 초기피치에 대하여 주파수영역에서 피치개선(pitch refinement)과정을 수행하기 때문에 피치 검색시 많은 시간이 소요된다. 처리시간이 많이 소요되면 저가형 정수처리 DSP칩으로 실시간 처리가 어려워져서 비용이 높아진다. 또한 보코더의 처리과정이 복잡하면 이에 비례적으로 전력소모가 증가하기 때문에 휴대전화기 등에서 강전지 사용시간이 감소하게 된다[2][3].

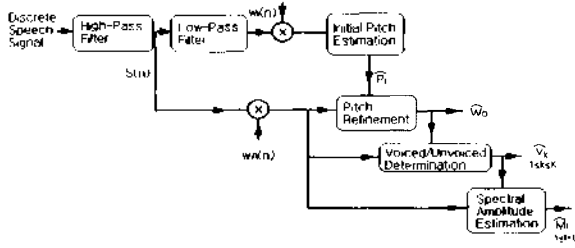


그림 1-1 IMBE 음성 부호화기의 블록도

본 논문에서는 IMBE보코더의 처리시간에서 50% 정도를 차지하는 피치검색과정에 대해 음질의 일화를 최소화하면서 피치 검색시간을 줄일 수 있는 피치검색법을 새로이 제안하고자 한다.

## 2. IMBE 보코더의 피치검색법

IMBE는 피치검색시 피치 파라미터값의 평균 자승 오차가 최소가 되도록 하는 값을 구한다. 피치검색은 우선 시간 영역에서 초기 피치를 구한 후 초기피치를 근거로 하여 정밀한 피치검색을 수행한다. 초기피치검색과정은 첫째로, 이전의 프레임과 피치의 연속성을 유지하는 후방추정,  $\hat{P}_B$ , 둘째로 미래의 음성 프레임과 피치의 연속성을 유지하는 전방추정,  $\hat{P}_F$ 가 그것이다. 후방 피치 추정은 후방피치추정 알고리즘에 의해 계산되고, 전방 피치 추정은 전방피치추정 알고리즘에 의해 계산된다. 피치 추정의 목적은 현재의 음성 프레임  $S_w(n)$ 과 관계되는 피치  $P_0$ 를 검출하는 것이다.  $P_0$ 는 (2-1)식에 의해 기본 주파수  $\omega_0$ 와 연관된다. 여기서  $\omega_0$ 는 라디안 값이다.

$$P_0 = 2\pi/\omega_0 \quad (2-1)$$

즉, 피치추정 알고리즘은 현재 프레임의 피치가 결정되었을 때 이전 프레임과 앞 프레임의 피치까지 고려하게 된다. 두 개의 앞선 음성 프레임과 관련되어있는 피치들은  $P_1$ 과  $P_2$ 로 나타낸다. 비슷하게 두 개의 이전 음성 프레임의 피치는  $P_{-1}$ 과  $P_{-2}$ 로 나타낸다. 초기의 피치 추정은  $\{21, 21.5, \dots, 121.5, 122\}$  집합의 한 값으로 제한된다. 초기피치 값은 4분의 1 샘플 정확성을 갖는 기본주파수  $\hat{\omega}_0$ 를 측정할 후에 원하는 피치를 얻을 수 있다.

두 부분으로 이루어진 순서는 피치검출의 정확성을 향상시키기 위한 것이다. 피치검출 알고리즘에서 한가지 중요한 것은 초기의 피치검출 알고리즘이 피치개선 알고리즘과는 다른 윈도우를 사용한다는 것이다. 초기의 피치검출을 위해서 사용되는 윈도우  $\omega_1(n)$ 은 301샘

플의 길이를 갖는다. 피치개선을 위해 사용되는 윈도우  $\omega_R(n)$ 은 221샘플의 길이를 갖는다. 여기서 두 윈도우의 중심점은 일치해야만 한다.  $\omega_R(n)$ 이 사용될 때는 61샘플 오버랩되고,  $\omega_1(n)$ 이 사용될 때는 141샘플이 오버랩된다. 피치 추정은  $E(P)$ 의 결과값을 비교하여 수행되고, 21부터 122범위 내에서 가장 알맞은 후보값을 골라  $\hat{P}_1$ 로 명명한다. 이 과정은 그림 2-1에 나타내었다.  $E(P)$ 함수는 다음과 같이 정의된다.[4]

$$E(P) = \frac{\sum_{i=1}^{150} s_{LPF}^2(i) \omega_1^2(i) - P \cdot \sum_{n=P}^{150-P} r(n \cdot P)}{\left[ \sum_{i=1}^{150} s_{LPF}^2(i) \omega_1^2(i) \right] \left[ 1 - P \cdot \sum_{n=P}^{150-P} \omega_1^2(i) \right]} \quad (2-2)$$

여기서  $\omega_1(n)$ 은 초기 피치 검색 윈도우이고,  $r(n)$ 과  $s_{LPF}(n)$ 식은 다음과 같다.

$$r(t) = \sum_{i=1}^{150} s_{LPF}(i) \omega_1^2(i) s_{LPF}(i+t) \omega_1^2(i+t)$$

$$s_{LPF}(n) = \sum_{i=1}^{150} s(n-i) h_{LPF}(i)$$

초기피치값  $\hat{P}_1$ 는  $E(\hat{P})$ 값을 가장 작게 하는 값으로 선택한다. 이때 측정값이 갑자기 변화할지도 모른다. 피치값의 갑작스런 변화는 합성음의 질을 떨어뜨리게 되는 결과를 초래할 수 있다. 피치가 천천히 변한다는 성질 때문에 이웃하는 프레임으로부터 얻어지는 피치측정값은 현재 프레임의 피치추정에 도움을 줄 수 있다[4].

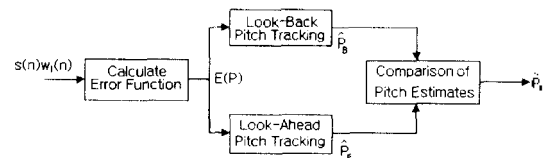


그림 2-1. 초기피치검색 블록도

### 2.1 Look-Back Pitch Tracking

$E(P)$ 의 값을 최소화시키고 식 2-3을 만족시키는  $P$  값을  $\hat{P}_B$ 로 정의한다[4].

$$0.8 \hat{P}_{-1} \leq P \leq 1.2 \hat{P}_{-1} \quad (2-3)$$

### 2.2 Look-Ahead Pitch Tracking

전방 피치추정은 미래의 음성 프레임 사이에서 피치의 연속성을 보존하기 위한 시도로 이루어진다[4].

$CE_F(P_0)$ 의 값을 최소로 하는 집합내의  $P_0$ 값을  $\hat{P}_0$  값으로 정의한다.

전방누적오차와 후방누적오차가 계산되어진 후 각각의 대표값을 비교하여 결정논리에 의하여  $\hat{P}_F$  또는

$\hat{P}_B$ 중의 하나의 값이 초기 피치 측정  $\hat{P}_1$ 로 선택된다[4].

### 2.3 Pitch Refinement

피치 개선 알고리즘은 정확도를 2분의 1 샘플에서 4분의 1 샘플까지 피치 측정의 결과를 향상시켜준다. 열개의 후보 피치들은 초기 피치 측정으로부터 만들어낸다. 이것들은 각각 수식적으로 다음과 같다.

$$\hat{P}_1 - 9/8, \hat{P}_1 - 7/8, \dots, \hat{P}_1 + 7/8, \hat{P}_1 + 9/8 \quad (2-4)$$

후보피치들은 기본주파수로 바뀌어진다. 식 2-5에 정의된 에리함수  $E_R(\omega_0)$ 는 후보 기본 주파수  $\omega_0$ 를 대입하여 계산된다.  $E_R(\omega_0)$ 를 최소로 하는 값으로부터 얻어진 후보 기본 주파수는 수정된 기본주파수  $\hat{\omega}_0$ 로 선택되어진다[4].

$$E_R(\omega_0) = \sum_{m=0}^{N-1} |S_w(m) - S_w(m, \omega_0)|^2 \quad (2-5)$$

### 3. 제안한 방법

기존의 IMBE에서 초기 피치검색은 시간영역에서 전방피치추정과 후방피치추정 알고리즘을 이용하여 21부터 122 사이에서 초기피치값을 선택한다. 이 방법은 피치의 갑작스런 변화를 방지하기 위한 것이다. 이렇게 찾아진 초기치에 대하여 다시 주파수영역에서 피치 개선 과정을 수행하기 때문에 피치 검색시 많은 시간이 소요된다. 이러한 IMBE의 복잡한 피치검색방법은 음질의 향상을 가져오지만 피치검색시간이 많이 소요되는 단점을 갖고 있다. 따라서 본 논문에서는 SAMDF (Spectrum Area Magnitude Difference Function) 를 적용하여 피치 검색시간을 단축하는 방법을 제안하였다 [6].

그림 3-1 에 본 논문에서 제안한 블럭도를 나타내었다. 시간영역에서 처리하는 초기피치 검색법을 생략하고 바로 주파수 영역으로 변환한 다음 본 논문에서 제안한 방법인 SAMDF를 이용하여 기본 주파수를 찾는 다음 피치 개선 과정을 수행하였다. 그러나 이렇게 할 경우 피치의 갑작스런 변화를 방지할 수 없으므로 홀수 프레임에 대해서는 기본주파수가 존재하는 건 범위 내에서 검색하고 짝수 프레임에서는 이전 프레임의 기본주파수의  $\pm 20\%$ 내의 범위 내에서만 검색을 하였다. 만일 짝수 프레임의 기본주파수가 이전 프레임의

기본주파수의  $\pm 20\%$  범위를 벗어날 경우에는 이전 프레임의 기본주파수로 대체하였다.

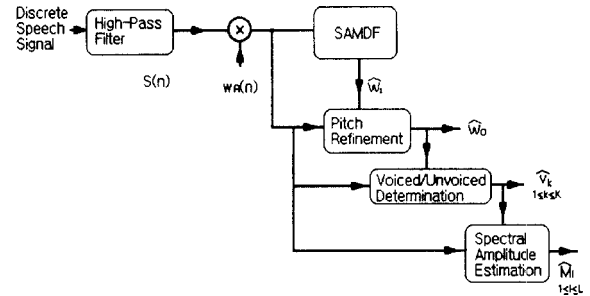


그림 3-1. 본 논문에서 제안한 처리 블럭도

SAMDF식은 다음(식 3-1)과 같다. 이 SAMDF법은 유성음의 경우 스펙트럼상에서 제 1 포먼트의 에너지가 다른 포먼트들보다 약 10dB정도 크다는 점에 착안하여 최대 에너지 위치인  $F_m$ 까지만 처리한다.

$$SAMDF(d) = \sum_{i=0}^N |S_w(i) - S_w(i-d)|, \quad (3-1)$$

$$(d = 1, 2, 3, \dots, F_m)$$

여기서  $N$ 은 한 프레임의 길이이고,  $S_w(\cdot)$ 는 음성신호의 진폭스펙트럼이다.

그림 3-2 에 본 논문에서 제안한 블럭도를 나타내었다

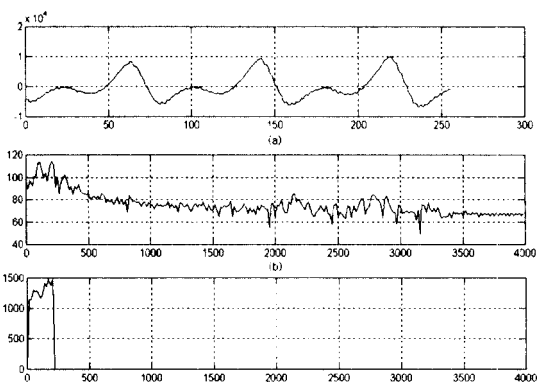


그림 3-2. 유성음에 대한 제안한 방법의 결과에서  
(a) 원 음성 파형  
(b) 원 음성에 대한 진폭스펙트럼  
(c) SAMDF를 통과한 진폭스펙트럼

그림 3-3에서는 원 음성과 합성음성에 대해서 나타내었다.

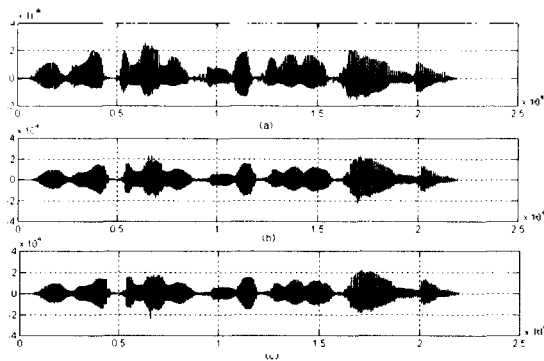


그림 3-3. 원 음성과 합성음성의 비교  
 (a) 원 음성 파형  
 (b) 기존 IMBE의 합성음성 파형  
 (c) 제안한 IMBE의 합성음성 파형

#### 4. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC/586(200) 시스템이며 여기에 음성신호를 입출력하기 위한 상용화된 16비트 AD/DA변환기를 인터페이스하여 8kHz의 표본율로 데이터를 입력하였다. 각 시료에 대해 한 프레임의 길이를 160표본으로 하여 처리하였다. 처리결과와 성능을 측정하기 위해 다음의 대표적인 문장을 연령층이 다양한 남녀 5명의 화자가 5번씩 발성하여 시료로 사용하였다.

- 발성1: /인수네 꼬마는 천재소년을 좋아한다./  
 발성2: /예수님께서 천지창조의 교훈을 말씀하셨다./  
 발성3: /충실대 정보통신공학과 음성통신연구팀이다./  
 발성4: /감사합니다./

피치 검색은 IMBE 보코더의 피치 검색 과정을 C-언어로 구현하여 수행하였다. 성능비교를 위해서 기존의 IMBE 보코더 피치검색과 제안한 방법을 시뮬레이션하여 그 결과를 비교하였다. 기존의 IMBE vocoder는 초기 피치검색과 피치 개선 과정을 모두 수행하였고, 본 논문에서 제안한 알고리즘은 초기피치 검색과정은 수행하지 않고 바로 주파수 영역에서 SAMDF법을 이용하여 초기기본주파수를 구하여 피치 개선과정을 수행하였다. 표 4-1.에서는 기존의 IMBE의 처리시간과 제안한 IMBE의 처리시간을 비교하여 나타내었다. 표 4-2.에서는 MOS 테스트 결과를 나타내었다.

표 4-1. 기존의 방법과 제안한 방법의 처리시간 결과 비교 (sec)

	발성1	발성2	발성3	발성4	평균
기존 IMBE	7.470	8.680	8.900	3.630	7.17
제안한 IMBE	3.854	4.458	4.683	1.924	3.73

표 4-2. 기존의 방법과 제안한 방법의 MOS 결과

	발성1	발성2	발성3	발성4	평균
기존 IMBE	3.6	3.8	3.6	3.7	3.675
제안한 IMBE	3.6	3.75	3.65	3.7	3.67

#### 5. 결론

IMBE 보코더에서 피치 검색시간은 저가형 정수처리 DSP를 사용할 때 총 부호화 처리시간의 약 50%정도를 차지한다. 따라서 본 논문에서는 SAMDF법을 이용하여 IMBE 보코더의 피치 검색시간을 개선하는 검색법 을 새로이 제안하였다.

실제 음성에 대해 제안한 피치검색을 사용하여 IMBE 부호화를 수행하였을 때, 기존의 IMBE 피치검색법에 비해 처리시간이 평균 52.02%로 감소하였다. MOS 테스트결과 음질의 열하는 거의 없었다.

#### 6. 참고문헌

- [1] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech Signal", Prentice Hall, 1978.
- [2] A. M. Kondoz, "Digital Speech", John Wiley & Sons, 1994.
- [3] J.C.Hardwick and J.S.Lim, "A 4800kbps Multi-band Excitation Speech Coder", Proc. of IEEE int. Conf. on Acoustics, Speech and Signal Proc., pp. 374-377, New York, April 1988
- [4] "APCO project 25 Vocoder Description", Digital Voice Systems, Inc., 1993
- [5] 나훈외 6명, "IMBE 음성부호기의 효율적인 피치 탐색을 이용한 실시간 구현에 관한 연구", KITE Autumn Conference '95, Vol.19, No.2 pp. 1621-1624
- [6] 홍성훈, 이주현, 배명진, 김종재, 유하영, "SAMDF에 의한 CELP형보코더의 피치검색법", 1995년도 한국음향학회 학술발표대회 논문집, Vol.9, No.1(s), pp.171-174, Nov.11, 1995.