1999년도 한국음향학회 학술발표대회 논문집 제18권 제1(s)호

IMBE Vocoder 실시간 처리를 위한 피치 검색 시간 개선에 관한 연구

장경아, 김정진, 민소연, 배명진
숭실대학교 정보통신공학과

On a Improvement of Pitch Search Time for Real Time Implementation in IMBE Vocoder

KyungA Jang, JeongJin KIM, MyungJin Bae
Dept. of Telecomm. Engr., Soongsil Univ. Seoul 156-743, Korea
kajang@assp.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 skip unnecessary range of the pitch searching by using the quantization error. Applying the proposed method to the IMBE vocoder, we can get approximately 45.88% processing time reduction and there is no difference in voice quality between conventional IMBE and proposed IMBE.

1. 서론

음성신호를 메모리에 저장하거나 전송하기 위한 음성부호화법에는 크게 파형부호화법, 신호원 부호화법, 혼성부호화법 등이 있다[1]. 이들 중에서 혼성부호화법은 신호원 부호화법의 메모리 효율성과 고음질의 파형부호화법을 결합시킨 것이며, 포만트정보는 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 다루고 있는 IMBE (Improved Multi-Band Excitation)법 또한 최근 이동 및 휴대통신용 부호화법으로 채택되어 사용되고 있다[2].

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 한 프레임 내에서 여러 대역을 유/무성음 판정을 하여 자연스러운 음성합성을 할 수 있기 때문에 기존의 음성부호화기의 문제점을 해결할 수 있다. 따라서 IMBE형 보코더는 자연스러운 음성을 합성하기 위하여 적합한 피치를 찾아내는 과정이 필요하다[3]. 이를 위한 피치 탐색 과정은 음성 처리시 많은 시간이 소요된다. 시간영역에서 초기피치트래킹과정에서 음성피치값이 천천히 변하는 성질을 이용하여 자연성을 줄 수 있으며, 이를 위해 초기 피치는 현재 프레임의 앞, 뒤에서 피치트래킹과정을 수행하여 얻을 수 있다[4]. 이 초기피치값에 대하여 주파수영역에서 피치리파인먼트 과정을 수행하기 때문에 피치 검색시 많은 시간이 소요되고, 이는 저가형 정수처리 DSP칩으로 실시간 처리가 어려워져서 비용이 높아진다. 보코더의 처리과정이 복잡하면 비례적으로 전력소모가 증가하여 휴대전화기 등에서 건전지의 사용시간이 감소하게 된다[2][3].

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

2. IMBE 보코더의 피치검색법

IMBE는 피치검색시에 피치 파라미터값의 평균 자승오차가 최소가 되도록 하는 값을 구하기 때문에 합성에 의한 분석법으로 볼 수 있다. 피치검색은 우선 시

간 영역에서 초기 피치를 구한 후 초기피치를 근거로 하여 정밀한 피치검색을 수행한다. 초기피치검색과정은 첫째로, 이전의 프레임과 피치의 연속성을 유지하는 백워드측정, $\hat{P}_B$, 둘째로 미래의 음성 프레임과 피치의 연속성을 유지하는 포워드측정, $\hat{P}_A$,가 그것이다. 백워드 피치 측정은 룩백피치트래킹 알고리즘에 의해 계산되고, 포워드 피치 측정은 룩어헤드피치트래킹 알고리즘에 의해 계산된다. 피치 측정의 목적은 "현재의" 음성 프레임 $S_w(n)$과 관계되는 피치 P_0는 식 2-1에 의해 기본 주파수 ω_0와 연관되는데, ω_0는 라디안 값이다.

$$P_0 = 2\pi/\omega_0 \qquad (2.1)$$

즉, 피치트래킹 알고리즘은 현재 프레임의 피치가 결정되었을 때 이전프레임과 앞프레임의 피치까지 고려하게 된다. 두 개의 앞선 음성 프레임과 관련되어있는 피치들은 P_1과 P_2로 나타낸다. 비슷하게 두 개의 이전 음성 프레임의 피치는 P_{-1}과 P_{-2}로 나타낸다. 초기의 피치 측정은 {21,21.5,.....121.5,122} 집합의 한 값으로 제한된다. 초기피치 값은 4분의 1 샘플 정확성을 갖는 기본주파수 $\hat{\omega}_0$를 측정한 후에 원하는 피치를 얻을 수 있다.

두 부분으로 이루어진 순서는 피치검출의 정확성을 향상시키기 위한 것이다. 피치검출 알고리즘에서 한가지 중요한 것은 초기의 피치검출 알고리즘이 피치리파인먼트 알고리즘과는 다른 윈도우를 사용한다는 것이다. 초기의 피치검출을 위해서 사용되는 윈도우 $\omega_1(n)$은 301샘플의 길이를 갖는다. 피치개선을 위해 사용되는 윈도우 $\omega_R(n)$은 221샘플의 길이를 갖는다. 두 윈도우의 중심점은 일치해야만 한다. $\omega_R(n)$이 사용될 때는 61샘플이 오버랩되고, $\omega_1(n)$가 사용될때는 141샘플이 오버랩된다. 피치 추정은 E(P)의 결과값을 비교하여 수행되고, 21부터 122범위 내에서 가장 알맞은 후보값을 골라 $\hat{P}_I$로 명명한다. 이 순서는 그림 2-1에 나타내었다. E(P)함수는 다음과 같이 정의된다.

$$E(P) = \frac{\sum_{j=-150}^{150} s^2_{LPF}(j)\,\omega^2_1(j) - P\cdot\sum_{n=\cdots}^{\cdots} r(n\cdot P)}{\left[\sum_{j=-150}^{150} s^2_{LPF}(j)\,\omega^2_1(j)\right]\left[1 - P\cdot\sum_{n=-150}^{150}\omega^4_1(j)\right]} \qquad (2.2)$$

여기서 $\omega_1(n)$은 초기 피치검색시 윈도우이고

$$r(t) = \sum_{j=-150}^{150} s_{LPF}(j)\,\omega^2_1(j)\,s_{LPF}(j+t)\,\omega^2_1(j+t) \qquad (2.3)$$

$$s_{LPF}(n) = \sum_{j=-10}^{10} s(n-j)h_{LPF}(j),\quad h_{LPF}(n):\text{FIR필터} \qquad (2.4)$$

초기피치 검출값 $\hat{P}_I$는 $E(\hat{P})$값을 가장 작게하는 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 피치측정에 도움을 줄 수 있다[4].

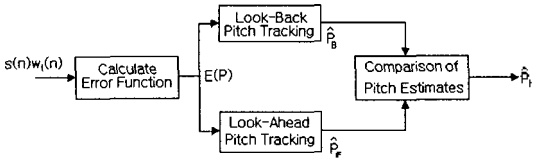

그림 2-1. 초기피치검색 블록도

2.1 Look-Back Pitch Tracking

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

$$.8\,\hat{P}_{-1} \le P \le 1.2\,\hat{P}_{-1} \qquad (2.5)$$

2.2 Look-Ahead Pitch Tracking

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

$CE_F(P_0)$의 값을 최소로 하는 집합내의 P_0값을 $\hat{P}_0$ 값으로 정의한다.

포워드누적에러와 백워드누적에러가 계산되어진 후 각각의 대표값을 비교하여 결정논리에 의하여 $\hat{P}_F$ 또는 $\hat{P}_B$중의 하나의 값이 초기 피치 측정 $\hat{P}_I$로 선택된다[4].

2.3 Pitch Refinement

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

$$\hat{P}_I - 9/8,\ \hat{P}_I - 7/8,\ \ldots,\ \hat{P}_I + 7/8,\ \hat{P}_I + 9/8 \qquad (2.6)$$

후보피치들은 기본주파수로 바뀌어진다. 식 2-5에 정의된 에러함수 $E_R(\omega_0)$는 후보 기본 주파수 ω_0를 대입하여 계산된다. $E_R(\omega_0)$를 최소로 하는 값으로부터

얻어진 후보 기본 주파수는 수정된 기본주파수 $\hat{\omega}_0$로 선택되어진다[4][6].

$$E_R(\omega_0) = \sum_{m=50}^{\left(\left\lfloor \frac{.9254\pi}{\omega_0} - .5 \right\rfloor \frac{256}{2\pi}\omega_0\right)} |S_w(m) - S_w(m,\omega_0)|^2 \qquad (2.7)$$

3. 제안한 방법

기존의 IMBE에서 초기 피치검색은 시간영역에서 룩백피치 트래킹과 룩어헤드피치트레킹 알고리즘을 이용하여 21부터 122 사이에서 초기피치값을 선택한다. 이렇게 찾아진 초기치에 대하여 다시 주파수영역에서 피치 리파인먼트과정을 수행하기 때문에 피치 검색시 많은 시간이 소요된다. 이러한 IMBE의 복잡한 피치검색 방법은 음질의 향상을 가져오지만 피치검색시간이 많이 소요되는 단점을 갖고 있다. 본 논문에서는 완전한 피치검색을 수행하기 전에 상관관계가 높은 구간을 양자화 오차의 추림을 통해 구한 다음에 양의 상관관계를 갖는 구간에 대해서만 피치검색을 수행하는 방법을 제안하고자 한다. M비트로 선형 양자화된 음성신호 s(n)은 다음과 같이 나타낼 수 있다:

$$\begin{aligned} s(n) &= \sum_{i=0}^{M-1} a_i 2^i \\ &= \sum_{i=0}^{N-1} a_i 2^i + \sum_{i=N}^{M-1} a_i 2^i \\ &= Q_L + Q_H \end{aligned} \qquad (3.1)$$

여기서 Q_L은 음성신호를 (M-N)비트로 부호화할 때 발생하는 양자화 오차이다. 유성음 파형의 경우에 낮은 쪽 포만트는 높은 쪽의 포만트에 비해 에너지가 아주 높다. 따라서 그림 3-1(b)와 같이 에너지가 우세한 기본주파수와 제 1,2 포만트성분들은 Q_L의 최대진폭을 유지하게 된다. 한편 에너지가 낮은 고차의 포만트들은 Q_L의 진폭범위내에서 파형의 빠른 변화를 이루게 된다. 양자화 오차 Q_L의 또다른 특징은 진폭변화의 범위가 2^N-1 이내로 제한되어 정규화된 진폭특성을 얻게 된다는 점이다. 이것은 시간영역에서 파형진폭의 변동에 따른 피치주기 검색에 미치는 영향을 감소시킬 수 있게 된다. 양자화 오차 Q_L을 사용하여 저역특성이 강한 제 1,2 포만트 위주의 정규화된 파형을 추출하여 그림 3-1(c)에 나타내었다. 이 정규화된 파형을 사용하여 예비피치를 구하려면 먼저 주기성 강조를 수행해야 한다. 주기성 강조법에는 자기상관관계법, AMDF법, 확률분포도법 등이 제안되어져 있으나[2][5], 본 논문에서는 다음과 같이 자기 상관관계법을 적용하였다:

$$R(L) = \sum_{n=0}^{fr-1} s(n)s(n-L) \qquad (3.2)$$

여기서 fr은 프레임 길이, L은 시간지연값, s(n)은 저대역신호를 각각 나타낸다.

검출된 주기신호가 두가지 레벨만을 갖기 때문에 다음과 같이 파형의 부호 파악만으로 자기 상관함수값을 계산할 수 있게 된다:

$$\begin{aligned} s(n)s(n-L) &= C, \quad \text{if } s(n) = s(n-L) \\ &= -C, \quad \text{if } s(n) \neq s(n-L) \end{aligned} \qquad (3.3)$$

여기서 C는 양자화 오차의 최대값에 대한 제곱을 나타낸다. 이처럼 양자화 오차에 대한 상관관계를 구하기 위해 곱셈대신에 덧셈을 수행하여도 DSP(디지탈 신호처리칩)로는 계산량이 줄어들지 않는다. 따라서 본 논문에서는 상관관계파형에서 음의 봉우리 부분을 스킵하여 피치검색시간을 단축하는 방법을 제안하였다.

유성음의 경우 실제 피치 지연은 항상 상관관계함수의 파형의 양의 봉우리에 위치한다[1]. 때문에 피치검색시 양의 봉우리에 대해서만 수행하게 되면 불필요한 피치 검색시간을 줄일 수 있게된다. 이것은 다음에 나타나는 음성신호의 상관관계 함수의 특성들에 의해 가능하다: (1)유성음에서 파형은 천천히 변화하기 때문에, 단구간에서 음성신호는 매우 높은 상관관계를 갖는다. (2)양의 봉우리 구간과 음의 봉우리 구간은 번갈아 나타난다. (3)유성음의 제1포만트의 영향이 지배적이기 때문에 각 봉우리 구간의 폭은 거의 변화하지 않는다. 이들 특성을 이용하여 음의 봉우리 구간은 전의 양의 봉우리 구간의 폭을 계산함으로써 쉽게 제외할 수 있다. 즉, 양의 봉우리 구간의 폭을 계산하고 음의 봉우리가 시작하면 양의 봉우리구간만큼을 제외하고 상관관계함수를 수행함으로써 계산량을 감소시킬 수 있다[6-7]. 이렇게 하여 상관관계 파형의 봉우리가 양의 값을 나타내는 구간에 대해서 완전한 피치검색을 수행하게 되면, 음양의 봉우리가 교대로 나타날 확률이 50%이기 때문에 검색시간을 약 50% 정도로 절약할 수 있게 된다. 더 필요하다면 양의 봉우리 정점을 기준으로 전후 각각 3표본 정도에 대해서만 수행하게 되면, 양의 봉우리가 평균 6.4개 정도 나타나기 때문에 완전검색에 비해 30% 정도의 피치 검색시간만 필요로 하게 된다[8-10].

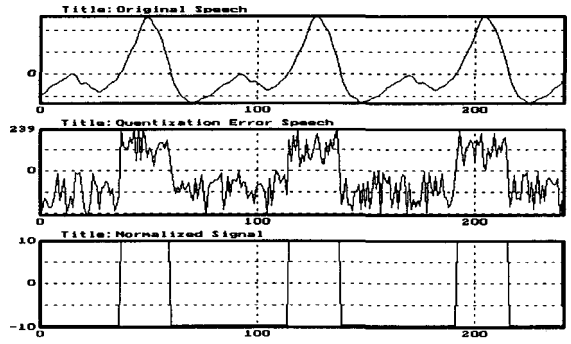

그림 3-1. 유성음에 대한 양자화 오차 파형
(a) 원 음성 파형
(b) 양자화 오차 파형
(c) 규준화된 파형

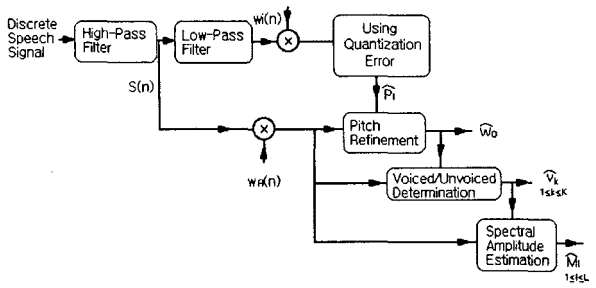

그림 3-2. 본논문에서 제안한 처리 블록도

4. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC/586(200) 시스템이며 여기에 음성신호를 입출력하기 위한 상용화된 16비트 AD/DA변환기를 인터페이스하여 8㎑의 표본율로 데이터를 입력하였다. 각 시료에 대해 한 프레임의 길이를 160표본으로 하여 처리하였다. 처리결과의 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: /숭실대 정보통신공학과 음성통신연구팀이다./
발성4: /감사합니다./

피치 검색은 IMBE 보코더의 피치 검색 과정을 C-언어로 구현하여 수행하였다. 성능비교를 위해서 기존의 IMBE 보코더 피치검색과 제안한 방법을 이용하여 피치검색을 수행하였다. 본 논문에서는 피치 검색의 정확성을 기하기 위해 양의 상관관계 파형에 대해서만 완전한 피치검색을 수행하였다. 표 4-1에 기존의 방법과 본 논문에서 제안한 방법의 결과를 나타내었다. 표 4-2에서는 MOS 테스트 결과를 나타내었다.

표 4-1 검색시간 결과[sec]

	발성1	발성2	발성3	발성4	평균
기존 IMBE	7.470	8.680	8.900	3.630	7.17
제안한 IMBE	4.210	4.560	4.760	1.980	3.88

표 4-2 기존의 방법과 제안한 방법의 MOS 결과

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

5. 결론

IMBE 보코더에서 피치 적응코드북의 검색시간은 저가형 정수처리 DSP를 사용할 때 총 부호화 처리시간의 약 50%정도를 차지한다. 따라서 우리는 IMBE 보코더의 피치 검색시간을 개선하는 검색법을 새로이 제안하였다. 먼저, 초기피치검색시 규준화된 양자화 오차 신호를 구하여 상관관계 함수를 적용하여 상관관계 파형 중에서 양의 봉우리부분에 대해서만 피치검색을 수행하였다. 이렇게 해서 찾아진 피치를 초기피치로 정하고 이 초기피치값을 기존의 피치 리파인먼트과정에 인수로 넘겨주어 음성을 처리하게 하였다. 즉, 간단하게 약식 상관관계 함수를 통해 얻어진 상관관계 값중에서 양의 봉우리를 이루는 구간을 예비피치 구간으로 선택하였다. 그런 다음에 예비피치 구간에 대해서만 피치검색을 수행하여 피치 검색시간을 절약하는 새로운 방법을 제안하였다.

실제 음성에 대해 제안한 피치검색을 사용하여 IMBE 부호화를 수행하였을 때, 기존의 IMBE 피치검색법에 비해 처리시간이 평균 45.88%의 처리시간이 감소하였다. MOS 테스트결과 음질의 열하는 거의 없었다.

6. 참고문헌

[1] L. R. Rabiner and R. W. Schafer, "Digital Proce-ssing 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 Siganl Proc., pp. 374-377, New York, April 1988

[4] "APCO project 25 Vocoder Description", Digital Voice Systems, Inc., 1993

[5] E.F. Deprettere and P. Kroon, "Regular Excitation Reduction for Effective and Efficient LP-Coding of Speech", IEEE, Proc. Int. Conf. on Acoustics, Speech and Signal Processing, pp.965-968, 1985.

[6] J.H.LEE, M.J.BAE, and H.Y.YOO, "A New Fast Pitch Search Algorithm Using the Abbreviated Correlation Function in CELP Vocoder", IEEE Comm. Soc., Proceeding of MILCOM'96 pp.653 - 657, Oct. 21-24, 1996.

[7] D. KIM, M. BAE, J. KIM, K. BYUN, K. HAN, H. YOO, "On a Reduction of Pitch Searching Time by Preliminary Pitch in the CELP Vocoder," J. Acoust. Soc. Korea, Vol.13, No.2E, pp.51-57, July 1994.