

## 균일양자화기의 잔여신호를 이용한 음성신호의 피치검출

### On a Pitch Extraction of Speech Signal using Residual Signal of the Uniform Quantizer

배 명 진\*, 한 기 천\*\*, 차 진 중\*\*  
(MyungJin Bae\*, KiCheon Han\*\*, JinJong Cha\*\*)

※본 연구는 한국전자통신연구소의 1997년도 수탁과제 연구지원비에 의해 이루어졌습니다.

#### 요 약

음성신호처리 분야에서 정확한 피치검출은 중요하고 필요하다. 지금까지 제안된 피치검출 알고리즘들은 음성신호의 다양성으로 인해 피치를 정확히 검출하기가 어렵다. 본 논문에서는 PCM과 같은 균일 양자화기의 잔여신호에 대해 음성신호의 기본주기를 검출하는 새로운 피치검출법을 제안하였다. 제안한 방법은 무잡음 음성에 대해 평균 0.25%의 조오율이 그리고 0dB의 SNR에 대해서는 평균 3.39%의 조오율이 나타나는 정확성을 보였다. 또한 음소의 천이영역이나 배경잡음 하에서도 피치검출의 정확도가 개선된 피치검출의 결과를 얻었다.

#### ABSTRACT

In speech signal processing, it is necessary and important to detect exactly the pitch. The algorithms of pitch extraction which have been proposed until now are difficult to detect exactly pitches over wide range speech signals. In this paper, thus, we proposed a new pitch detection algorithm that finds the fundamental period of speech signal in the residual signal quantized by the uniform quantizer as PCM. The proposed method shows little gross error of average 0.25% for clean speech and average 3.39% for SNR of 0dB. It also achieves results of the pitch contours, improving the accuracy of pitch detection in transient phonemes and noise environments.

#### I. 서 론

음성인식, 합성 및 분석과 같은 음성신호처리 분야에 있어서 기본주파수 즉, 피치를 정확히 검출하는 것은 중요하고 어려운 일이다. 만일 음성신호의 기본주파수를 정확히 검출할 수 있다면 음성인식에 있어서 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성합성 시에 자연성과 개성을 쉽게 변경하거나 유지할 수 있고, 음성부호화시에는 압축율을 높여 고품질의 저전송율이 가능하다. 또한 분석시에는 피치에 동기시켜 분석하면 성문의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있다.

이러한 피치검출의 중요성 때문에 지금까지 피치검출

에 대한 방법들이 다양하게 제안되었으며, 시간영역법, 주파수영역법, 시간-주파수영역법으로 구분할 수 있다. 시간영역 검출법은 음성파형의 주기성을 강조한 후에 결정논리에 의해 피치를 검출하는 방법으로 병렬처리법, AMDF법, ACM법 등이 있다. 이러한 방법은 보통 시간영역에서 수행되므로 영역의 변환이 불필요하고, 합, 차, 비교논리 등 간단한 연산만 필요하게 된다. 그러나, 음소가 천이 구간에 걸쳐 있는 경우에는 프레임 내의 레벨변화가 심하고 피치주기가 변동하기 때문에 피치검출이 어렵다. 특히 잡음이 섞인 음성의 경우에는 피치검출을 위한 결정논리가 복잡해져서 검출 오류가 증가되는 단점이 있다[1]-[3].

주파수영역의 피치검출법은 음성 스펙트럼의 고조파 간격을 측정하여 유성음의 기본주파수를 검출하는 방법으로 고조파분석법[4], Lifter법, Comb-filtering법 등이 제안되어 있다[1][2]. 일반적으로 스펙트럼은 한 프레임(20-40ms) 단위로 구해지므로, 이 구간에서 음소의 천이나 변동이 일어나거나 배경잡음이 발생하여도 평균화되므로 그 영향을 적게 받는다. 그러나 처리 과정상 주파수영역으

\* 숭실대학교 정보통신공학과  
\*\* 전자통신연구소 VLSI 구조 연구실  
접수일자: 1996년 10월 20일

로의 변환과정이 필요함으로 계산이 복잡하며, 기본주파수의 정밀성을 높이기 위해 FFT의 포인트 수를 늘리면 그만큼 처리시간이 길어지고 변화특성에 둔해지게 된다.

시간-주파수 혼성영역법은 시간영역법의 계산시간 감소와 피치의 정밀성, 그리고 주파수영역법의 해상도를 떠나 음소변화에 대해서도 피치를 정확히 구할 수 있는 장점을 취한 것이다. 이러한 방법으로는 Cepstrum법, 스펙트럼비교법 등이 있고, 이 방법은 시간과 주파수영역을 왕복할 때 오차가 가중되어 나타나므로 피치추출의 영향을 받을 수 있고, 또한 시간과 주파수영역을 동시에 적용하기 때문에 계산과정이 복잡하다는 단점이 있다[6].

따라서 본 논문에서는 상기에 열거한 문제점들 중 시간영역 피치검출기의 복잡성을 해결하고 검출의 정확도를 높일 수 있는 시간영역 피치검출법을 새로이 제안하고자 한다. 제안한 방법은 선형 PCM 데이터의 양자화 오차가 갖는 특성을 이용하여 음성신호의 피치를 검출하게 된다. 따라서 제 2 절에서는 유성음 신호의 특성을 간단히 소개하였고, 제 3 절에서는 피치검출시의 문제점을 파악하였으며 제 4 절에서는 양자화된 오차신호의 특성에 대해 다루면서 음성신호의 피치주기를 추출하는 방법에 대해 제안하게 된다. 그리고 5 절에서는 실험 및 결과에 대해 설명한 후에 6장에서 결론을 짓게 된다.

### II. 유성음 신호의 분석

음성신호는 음성 여기원에 따라 유성음, 무성음, 혼합음으로 구분할 수 있다. 무성음의 경우에는 백색 가우시안 불규칙사륵스가 그 여기원이므로 주기성은 나타나지 않지만, 주로 3kHz 근방에서 첫 번째의 공진봉우리를 갖기 때문에 유성음에 비해 평균 영교차율이 크다. 유성음은 폐에서 올라온 공기가 성문을 통하여 배출될 때 진동되고, 성도에서의 공명으로 인하여 그림 2-1(a)처럼 에너지가 크고 준-주기적인 형태의 신호가 된다. 이를 주파수영역에서 살펴보면 그림 2-1(b)와 같이 성도의 공명봉우리에 유성신호의 기본주파수  $F_0$ 가 세세하게 나타나고 있다. 성도 공명 봉우리의 주파수들을 포먼트라고 하고 가장 낮은 주파수의 봉우리를 제 1 포먼트( $F_1$ )라 한다.

일반적인 유성음 구간에서  $F_1$ 의 에너지봉우리는 다른 포먼트들보다 10dB이상 높기 때문에 이를 시간영역의 파형으로 표현하면  $F_1$ 의 영향이 주로 나타난다. 한 피치구간에서 Zero Crossing Interval(ZCI)의 역수는  $2F_1$ 의 주파수와 거의 같게 된다. 그리고 포먼트들은 대역폭을 갖게 되므로 시간영역 파형의 한 피치구간에서는 감쇄진동을 하게 된다. 따라서 한 피치구간에서 처음 양의 봉우리가 다른 봉우리를 보다 두드러지게 나타나게 되고, 이 봉우리가 한 피치구간내에서 성문의 영향이 크게 나타나는 성문봉우리로 고려할 수 있다[7].

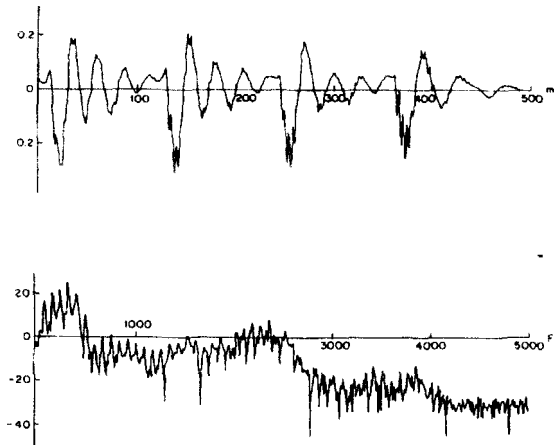


그림 2-1. 유성음의 파형과 spectrum:(a) 유성음에 대한 파형, (b) 유성음에 대한 spectrum.

### III. 피치검출의 문제점

음성신호의 피치는 음성 파형의 반복되는 봉우리에서 봉우리까지 또는 골에서 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 시간 지연에 대해서만 자기 상관관계가 높게 존재한다.

음성 파형에 대해 피치주기를 검출하려고 하면 성도 포먼트에 따른 영향을 받게 된다. 성도포먼트들은 문장을 구성하는 음소에 따라 변화하게 되고, 음소는 10ms 정도의 범위 내에서는 안정상태를 이룬다. 또한 피치검출시에 크게 영향을 주는 포먼트들은 기본주파수에 근접한 제 1 포먼트이고, 이 포먼트의 에너지가 파형을 지배하기 때문에 이 성분을 적용적으로 제거 또는 억압시킬 필요가 있다.

유성음의 포먼트성분을 적용적으로 추정하여 억압하는 고전적인 방법으로는 간략화된 역-역파기추정(SIFT)법이 있다. 이 방법은 선형예측분석에 의해 포먼트성분을 추정하는데 여성이나 어린이 화자인 경우에는 기본주파수가 제 1 포먼트와 근접되어 있기 때문에 포먼트 성분과 함께 기본주파수성분도 감쇠된다는 문제점이 존재한다[7].

시간영역에서 음성파형의 제 1 포먼트성분을 간단히 억압시키는 방법에는 저역통과여파기(LPF)에 음성신호를 통과시키는 방법이 있다. 그렇지만, 이 경우에는 LPF의 차단주파수를 제 1 포먼트주파수와 기본주파수의 중간 값으로 적용하는 것이 바람직하다. 그러나, 이를 사전에 계산하는 것은 아주 어렵다[8].

또한 시간영역에서 음성신호의 진폭은 음소에 따라 크게 변화한다. 특히 초성으로 유성자음이 발생되는 경우에는 이 유성자음의 피치를 구하기 전에 진폭을 정규화

시켜야 한다. 한 프레임에서 음성신호의 진폭을 정규화시키는 시키는 간단한 방법으로는 프레임을 두부분으로 분할하고 각 부분의 평균적인 진폭비로 파형을 정규화시키는 방법이 있다. 그렇지만 이러한 방법은 사전에 파형의 기울기를 정확히 추정해야하는 문제가 존재한다[1-2].

따라서 본 논문에서는 시간영역파형에서 피치를 검출할 때 나타나는 포먼트의 영향과 천이진폭의 영향을 동시에 제거할 수 있는 새로운 피치검출법을 제안하였다.

**IV. 양자화 오차를 이용한 피치검출법**

M비트로 선형 양자화된 음성신호 s(n)은 다음과 같다:

$$\begin{aligned}
 s(n) &= \sum_{i=0}^{M-1} a_i 2^i \\
 &= \sum_{i=0}^{N-1} a_i 2^i + \sum_{i=N}^{M-1} a_i 2^i \\
 &= Q_L + Q_H
 \end{aligned}
 \tag{4.1}$$

여기서 Q<sub>L</sub>은 음성신호를 (M-N)비트로 부호화할 때 발생하는 양자화 오차이다.

음성신호에서 양자화 오차 Q<sub>L</sub>에는 신호의 에너지가 우세한 성분과 미약한 성분이 함께 나타난다. 유성음 파형의 경우에 낮은 쪽 포먼트는 높은 쪽의 포먼트에 비해 에너지가 아주 높다. 따라서 그림 4-1(b)와 같이 에너지가 우세한 기본주파수와 포먼트성분들은 Q<sub>L</sub>의 최대진폭을 유지하게 된다. 한편 에너지가 낮은 고차의 포먼트들은 Q<sub>L</sub>의 진폭범위내에서 파형의 빠른 변화를 이루게 된다.

양자화 오차 Q<sub>L</sub>의 또다른 특징은 진폭변화의 범위가 2<sup>N</sup> 이내로 제한되어 정규화된 진폭특성을 얻게된다는 점이다. 이것은 시간영역에서 음소천이에 따른 파형진폭의 변동에 의해 입게되는 피치검출의 부정확성을 개선할 수 있게 된다.

양자화 오차 Q<sub>L</sub>을 사용하여 에너지가 우세한 제 1 포먼트 위주의 정규화된 파형을 추출하여 그림 4-1(c)에 나타내었다. 이 정규화된 파형을 사용하여 피치를 구하려면 먼저 주기성 강조를 수행해야 한다. 주기성강조법에는 자기상관계법, AMDF법, 확률분포도법 등이 제안되어져 있으나[1-2], 본 논문에서는 다음과 같은 자기상관계법을 적용하였다:

$$R(L) = \sum_{n=0}^{fr-L} e(n)e(n-L)
 \tag{4.2}$$

여기서 e(n)은 양자화기의 잔여신호이고, fr은 프레임 길이, L은 시간지연 값을 나타낸다.

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

$$\begin{aligned}
 e(n)e(n-L) &= C, \quad \text{if } e(n) = e(n-L) \\
 &= -C, \quad \text{if } e(n) \neq e(n-L)
 \end{aligned}
 \tag{4.3}$$

여기서 C는 양자화기 잔여신호의 최대 값의 곱을 나타낸다. 자기상관함수를 계산하여 그림 4-1(d)에 나타내었다.

피치주기를 검출하기 위한 결정논리로 자기상관값에 대해 피크피킹법을 적용하였다. 먼저 시간지연에 따른 상관값이 최대를 이루는 시간지연을 택하여 예비피치로 결정하였다. 예비피치보다 시간지연이 작은 구간에 대해 재검색을 수행한다. 재검색은 예비피치의 상관관계값에 비해 90%를 넘는 상관관계봉우리를 찾는 것이다. 예비피치의 상관관계 봉우리에서 90% 이상을 이루는 면적보다 재검색에서 예비피치 상관관계값의 90% 이상을 이루는 면적이 더 크다면 재검색에서 구한 시간지연이 피치로 결정되고, 같거나 또는 더 작은 경우에는 예비피치를 피치주기로 결정하게 된다. 이것은 한 프레임내에서 피치주기의 변동이 심한 경우에 실제 피치의 상관관계값이 시간지연에 대해 봉우리를 날카롭게 갖지 못하고 일정 폭을 유지하는 경우가 발생할 수 있기 때문이다. 또한 이러한 재검색 작업을 예비피치의 약수에 대해서만 적용하는 이유는 한 프레임내에서 피치주기가 변동하는 경우에는 배주기나 삼배주기에서 좀더 무딘 상관관계 봉우리를 구성하기 때문이다.

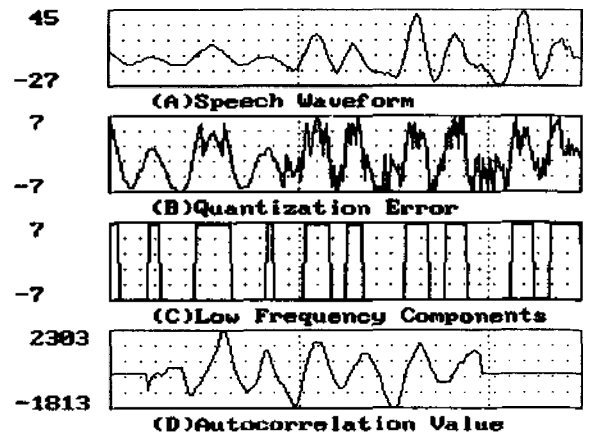


그림 4-1. 피치 검출과정의 예시:  
 (a)음성신호 /메-/의 천이파형, (b)양자화된 오차신호,  
 (c)검출된 기본주기파형, (d)상관계함수의 파형.

**V. 실험 및 결과**

이상의 과정을 컴퓨터 시뮬레이션하기 위하여 IBM PC/586-133에 마이크가 부착된 16-비트 A/D변환기를 인터페이스시키고, 아래의 문장들을 남녀 각 3명에게 발성시키면서 8kHz의 표본화 주파수로 표본화하여 저장한 다음에 시뮬레이션의 시료로 사용하였다:

- 발성1) "인수네 꼬마는 천재소년을 좋아한다."
- 발성2) "예수님께서 천지창조의 교훈을 말씀하셨다."
- 발성3) "승실대 정보통신과 음성통신 연구팀이다."
- 발성4) "창공을 헤쳐나가는 인간의 도전은 끝이없다."
- 발성5) "공일이합사오육칠팔구."

위의 각 음성시료에 대해 한 프레임의 길이를 256샘플로 하여 128샘플 단위로 겹치게하여 피치검출을 수행하였다. 본 논문에서 제안한 피치 검출과정을 블록도로 나타내면 그림 5-1과 같다.

주어진 프레임내의 음성파형에 대해 음성표본마다 상위 6바트로 양자화 하고, 이때의 양자화 오차를 구한다. 이 양자화 오차에 대해 ± 최대값을 유지하는 성분만 추출하고 나서, 자기상관함수식을 통해 주기성강조를 수행하였다. 주기성이 강조된 자기상관파형에 대해 문턱값 이상의 피크피킹을 수행하여 예비피치주기를 검출한 다음에 재검색을 통해 피치주기를 검출하였다.

상기의 음성시료에 대해 가우시안 백색잡음을 신호의 에너지에 비해 적으로 가미하면서 피치검출에 대한 조오율(gross error)을 측정하여 표 5-1에 결과로 제시하였다. 표를 살펴보면 조오율은 무잡음 음성신호에 대해 평균 0.25% 정도로 비교적 우수한 결과가 얻어졌으며, 0dB SNR 환경 하에서도 평균 3.39%의 오류만 발생하였다.

결과적으로 제안한 방법은 시간영역에서 직접 피치를 검출함으로써 파형의 위상특성을 유지할 수 있으며, 처리과정에서 정규화가 수행되어 음소천이에 따른 영향이 억압된다. 또한 양자화 오차의 최대 범위내를 변동하는 고차의 포먼트성분을 제거하기 때문에 고차 포먼트의 영향이나 배경잡음에도 강인한 특성을 나타내었다.

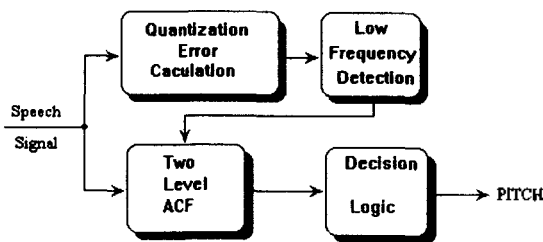


그림 5-1. 피치 검출에 대해 제안한 처리 블록도

표 5-1. 각 음성문장에 대한 gross 오류율

발성	분석 프레임 수	Gross Error Rates (%)			
		clean	SNR 6dB	SNR 3dB	SNR 0dB
1	192	0.12	1.02	1.25	3.21
2	192	0.27	1.19	1.27	3.23
3	192	0.72	1.27	1.34	3.91
4	64	0.25	1.20	1.30	3.19
평균		0.34	1.17	1.29	3.39

## VI. 결 론

음성신호처리분야에서 피치를 정확히 검출하면 음성 인식시에 화자에 따른 영향을 줄일 수 있기 때문에 인식의 정확도를 높일 수 있고, 음성합성시에 자연성과 개성을 유지하거나 쉽게 변경할 수 있다. 또한 분석시 피치에 동기시켜 분석하면 성분의 영향이 제거된 정확한 성도 파라미터를 얻을 수 있게 된다. 따라서 본 논문에서는 균일 양자화기의 잔여신호에 나타나는 파형적인 특성을 사용하여 음성신호의 피치를 검출하는 새로운 방법을 제안하였다.

제안한 방법은 시간영역에서 직접 피치를 검출하므로 파형의 위상특성을 유지할 수 있으며, 처리과정에서 정규화가 수행되어 음소천이에 따른 영향이 제거된다. 그리고 양자화 오차의 최대 범위내를 변동하는 고차의 포먼트성분을 제거하기 때문에 고차 포먼트의 영향이나 배경잡음에도 강인한 특성을 나타내었다. 또한 피치검출을 위해 필요한 전처리용 저역통과여파기를 필요로하지 않으며, 처리하는 데이터의 범위도 양자화된 잔여신호로 제한되어 계산처리용 메모리버퍼의 규모가 작아진다. 따라서 피치검출을 위한 시스템의 복잡성을 감소시킬 수 있었다.

## 참 고 문 헌

1. L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech signals*, Englewood Cliffs, Prentice-Hall, New Jersey, 1978.
2. P. E. Paparnichalis, *Practical Speech Processing* Prentice-Hall, Inc, Englewood Cliffs, New Jersey, 1987.
3. M. Bae, and S. Ann "The High Speed Pitch Extraction of Speech Signals Using the Area Comparison Method," KITE, Korea, Vol. 22, No. 2, pp. 13-17, March 1985.
4. S. Seneff, "Real Time Harmonic Pitch Detection," IEEE Trans. Acoust. Speech, and Signal Processing, Vol. ASSP-26, pp. 358-365, Aug. 1978.
5. M. Bae, and S. Ann, "Fundamental Frequency Estimation of Noise Corrupted Speech Signals Using the Spectrum Comparison," J., Acoust., Soc., Korea, Vol. 8, No. 3, June 1989.
6. A. M. Noll, "Cepstrum Pitch Determination," J. Acous. Soc. Am., Vol. 41, pp. 293-309, Feb. 1967.
7. J. D. Markel, "The SIFT Algorithm for Fundamental Frequency Estimation," IEEE Trans. Audio Electroacous., Vol. AU-20, pp. 367-377, Decem. 1972.
8. H. HONG, G. R. BAEK, M. J. BAE, H. S. JANG, "Pitch Detection using Variable Bandwidth LPP", J., Acoust., Soc., Korea, Vol. 13, No. 5, pp. 77-82, October 1994.
9. Hans Werner Strube, "Determination of the instant of glottal closure from the speech wave," J., Acoust., Soc., Am, Vol. 5, No. 5, pp. 1625-1629, November 1974.

10. M. Bae, I. Chung, and S. Ann. "The Extraction of Nasal Sound Using G-peak in Continued Speech," KIEE, Korea, Vol. 24, No. 2, pp. 274-279, March 1987.
11. J. H. LEE, M. J. BAE and C. H. KWON, "A Simple and Fast Pitch Search Algorithm using a Modified Skipping Technique in CELP Vocoder," J. Acoust. Soc. Korea, Vol. 14, No. 2E, pp. 33-38, July 1995.

▲배 명 진(Myung Jin Bae)

현재: 숭실대학교 정보통신공학과 교수  
(재14권 1E호 참조)

▲한 가 천(Ki Cheon Han) 1964년 1월 12일생

1986년 2월: 광운대학교 전자계산학과 졸업(학사)  
1988년 2월: 광운대학교 대학원 전자계산학과 졸업(석사)  
1988년 2월~현재: 한국전자통신연구원 통신회로연구실 선임연구원

※주관심분야: DSP 설계, 디지털 통신용 시스템 프로그래밍

▲차 진 중(Jin Jong Cha) 1956년 9월 17일생

1980년 2월: 한양대학교 전자공학과 졸업(공학사)  
1982년 2월: 한국과학기술원 전기 및 전자공학과 졸업(석사)  
1982년 3월~1985년 4월: 한국전자기술연구소 선임연구원  
1985년 3월~현재: 한국전자통신연구원 책임연구원 VLSI 구조연구실장

※주관심분야: 디지털 통신용 VLSI 설계, 영상통신용 VLSI 설계, DSP 설계