ISSN : 1226-7244 (Print)
ISSN : 2288-243X (Online)
논문번호 23-04-15
j.inst.Korean.electr.electron.eng.Vol.27,No.4,472~476,December 2023
https : //doi.org/10.7471/ikeee.2023.27.4.472
105

변곡점 검출에 기반한 음성의 기본 주파수 추정

Fundamental Frequency Estimation of Voiced Speech Signals Based on the Inflection Point Detection

임 병 관*

Byeonggwan Iem†

Abstract

Fundamental frequency/pitch period are major characteristics of speech signals. They are used in many speech applications like speech coding, speech recognition, speaker identification, and so on. In this paper, some of inflection points are used to estimate the pitch which is the inverse of the fundamental frequency. The inflection points are defined as points where local maxima, local minima or the slope changes occur. The speech signal is preprocessed to remove unnecessary inflection points due to the high frequency components using a low pass filter. Only the inflection points from local maxima are used to get the pitch period. While the existing pitch estimation methods process speech signals in blockwise, the proposed method detects the inflection points in sample and produces the pitch period/fundamental frequency estimates along the time. Computer simulation shows the usefulness of the proposed method as a fundamental frequency estimator.

요 약

피치 혹은 기본 주파수는 음성 신호의 주요 특성 인자이며 음성 부호화, 음성인식, 화자인식 등의 다양한 음성 관련 응용에 활용된다. 본 논문에서는 기본 주파수의 역수인 음성의 피치 주기를 추정하기 위해서 음성 신호의 변곡점을 이용한다. 변곡점은 국소적인 최대값, 최소값 혹은 신호의 기울기가 변하는 지점으로 정의된다. 음성 신호는 저역통과 필터로 먼저 전처리되어 고주파 성분이 제거된다. 이를 통해 불필요한 변곡점들이 제거되며, 피치 주기 추정에 유용한 국소적인 최대값만을 변곡점 검출법을 이용하여 추출한다. 얻어진 변곡점 간의 시간 간격을 측정하여 피치 주기를 추정하며, 그 역수로 기본 주파수 추정치를 얻는다. 기존의 피치 추정 방법은 음성이 국소적으로 시불변이라는 가정하에 음성을 블록 단위로 처리하여 블록당 피치 주기를 구하지만, 제안된 방법은 음성을 샘플 단위로 처리하여 변곡점을 검출하며, 그 결과 피치 주기를 시간 경과에 따라 얻게 되어 음성의 시변성이 반영된 기본 주파수 추정치를 얻는다. 컴퓨터 모의실험으로 기본 주파수 추정기로서 제안된 방법의 유용성을 볼 수 있다.

Key words : fundamental frequency, pitch, voiced speech, inflection point detection, autocorrelation.

Ⅰ. 서론

기본 주파수는 음성이 갖는 특성 가운데 가장 대표적인 인자로 음성코딩, 음성분석, 음성인식/화자인식, 음성합성 등 음성과 관련된 다양한 응용에서 필수적으로 검출해야 하는 요소이다. 음성 발생모델에서, 허파에서 나오는 공기 흐름을 통제하는 물리적인 부분인 성대(vocal

* Dept. of Electronic Eng., Gangneung-Wonju Nat. Univ.
† Corresponding author
E-mail : ibg@gwnu.ac.kr, Tel : +82-33-640-2426
Manuscript received Nov. 24, 2023; revised Dec. 4, 2023; accepted Dec. 7, 2023.

fold)의 개폐 주기에 직접 연관된 기본 주파수는 음성 혹은 화자의 고유한 특성이 되어 음성의 다양한 응용에서 활용된다[1, 2]. 이러한 기본 주파수는 응용에 따라 다양한 방법으로 검출될 수 있다. 예를 들어, 단순 음성분석의 경우에는 푸리에 분석으로 주파수 영역에서 얻을 수 있으며, 켑스트럼을 이용하여 얻는 방법, 음성코딩 등의 경우는 블럭 단위로 처리하며 다양한 short-time 기법으로 얻을 수 있다[1, 2, 3, 4]. 음성이 기본적으로 시변 신호이기 때문에 기본 주파수가 시간과 함께 변화할 것이라는 고찰과 달리, 기존의 방법은 음성이 국소적으로 시불변이라고 가정하여 블록별로 일정한 기본 주파수를 갖는 것으로 본다. 본 연구에서는 변곡점 검출이라는 간편한 방법으로 유성음의 기본 주파수를 추정하고자 한다. 변곡점은 신호의 추세가 바뀌는 지점을 지칭하며 대표적으로는 국소적인 최대값/최소값 등이 해당된다[5, 6, 7]. 이러한 변곡점 검출을 이용한 기본 주파수 추정은 기존의 방법과 달리 음성의 시변성(nonstationarity)을 반영할 수 있으며, 계산량을 줄여 실시간 처리시에 도움이 될 것으로 예상된다. 본 논문은 다음과 같이 구성된다. 먼저 음성의 기본 주파수와 관련된 음성발생모델을 설명한다. 그리고 음성의 기본 주파수를 검출하는 대표적인 방법인 short-time 상관함수(correlation function)를 소개한다. 그리고 변곡점 검출법을 소개하며 변곡점 검출법을 활용한 기본주파수 추정 방법을 제안한다. 컴퓨터 시뮬레이션으로 제안한 방법의 유용성을 보이고 결론을 맺는다.

II. 본론

1. 음성발생모델과 기본 주파수

유성음과 무성음으로 분류되는 음성 신호는 기본적으로 날숨에서 발생되므로, 허파에서 발생되는 공기의 흐름에 영향을 받는다. 특히, 유성음의 경우, 음성 발생 시스템에 주기적인 흐름을 보이는 공기가 입력되는 것으로 모델링 된다. 아래 그림 1은 인체의 음성 발생과 관련된 기관을 보이는 해부도와 이를 모델링한 그림이다[4]. 허파는 음성신호 발생을 위한 입력신호의 발생기 역할을 하고 후두(larynx)는 스위치의 역할을 한다. 후두 이후 입술까지가 선형시변시스템(linear time-varying system)으로 음성을 발생시킨다. 즉, 유성음의 경우 허파에서 발생된 날숨을 후두가 단속적으로 개폐하며 주기적인 공기 흐름을 만들고, 이를 입력으로 받은 구강과, 비강, 혀, 입술 등의 상호작용으로 소리가 만들어진다[1, 2]. 여기에서 후두의 주기적인 개폐에 의하여 유성음의 주기성이 나타나며 이를 유성음의 기본 주파수라 정의한다. 성별, 나이 등의 요인에 따라 개인적인 차이가 있지만, 통상 성인남성의 경우 200Hz 이내, 성인 여성의 경우 300Hz 이내에서 분포를 보인다.

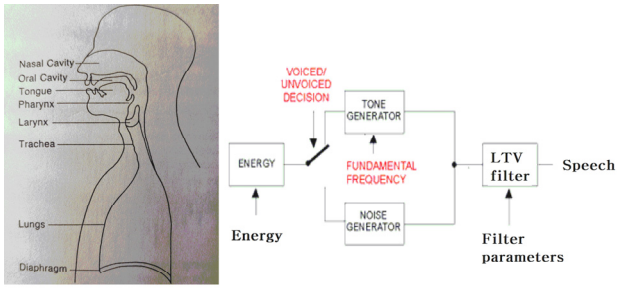

Fig. 1. (a) Anatomy of speech organ [4] and (b) Speech production model.
그림 1. (a) 해부도 [4]와 (b) 음성발생모델

2. 대표적인 기본 주파수 추정법

시간 영역에서 기본 주파수를 구하는 방법으로 자기상관함수를 이용하는 방법이 있으며, 주파수 영역에서는 푸리에 스펙트럼으로 하모닉 피크를 매칭하는 방법이 있다[1, 2]. 주기신호의 자기상관함수 또한 동일한 주기성을 갖는다는 고찰에 기반하여 자기상관함수를 이용한 기본 주파수 추정법이 널리 사용된다. 자기상관함수와 유사한 방식으로 AMDF(Average Magnitude Difference Function)를 이용하는 방법이 있으며 신호간 차의 평균치라는 면에서 구현시 오버플로를 방지한다는 장점이 있다[1, 2]. 자기상관함수 추정방법은 음성 신호 $s(n)$ 에 대하여 아래와 같은 구조의 상관함수 추정치를 적용하여 얻는다.

$$R(k) = \sum_{n=0}^{N-1} s(n)s(n+k) \quad (1)$$

3. 변곡점 검출 알고리즘 [5]

변곡점은 신호의 기울기 혹은 추이가 바뀌는 지점으로 정의된다. 대표적인 예로는 국부적인 최고점 또는 최저점을 들 수 있다. 즉, 신호가 증가하다가 감소하는 지점 혹은 감소하다가 증가하는 지점을 변곡점이라 할 수 있다. 아래 그림 2는 변곡점의 여러 유형을 보여준다. 그림 2의 점 A, B는 국소적인 최고점과 최저점을 나타내며, 점 C는 신호의 증가 혹은 감소 국면에서 기울기가 변화하는 단순 변곡점을 보여준다. 이러한 변곡점은 다음과

같이 검출할 수 있다.

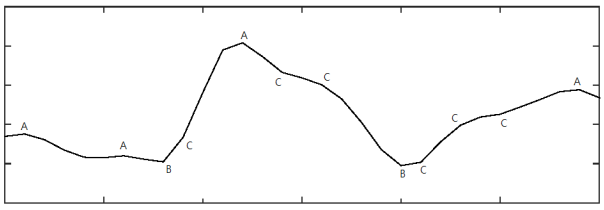

Fig. 2. A signal with various inflection points [5].
그림 2. 다양한 변곡점을 갖는 신호[5]

세 개의 연속하는 신호값 x_1, x_2, x_3에 대하여 연속하는 두 신호 값의 차를 $d_{ij}=x_i-x_j$라 할 때, 만약 $d_{21} \cdot d_{32} < 0$ 이면 x_2는 국부적인 최고점 A 혹은 최저점 B에 해당한다. 신호의 증가 혹은 감소의 추이가 바뀌는 기울기가 변화하는 변곡점 C는 아래 식에 의하여 검출될 수 있다[5, 6].

$$ID = \frac{|d_{21} - d_{32}|}{|d_{21}| + |d_{32}|} \geq threshold \qquad (2)$$

threshold값은 0과 1 사이의 값으로 설정된다. 그림 3은 변곡점 검출 알고리즘을 보인다.

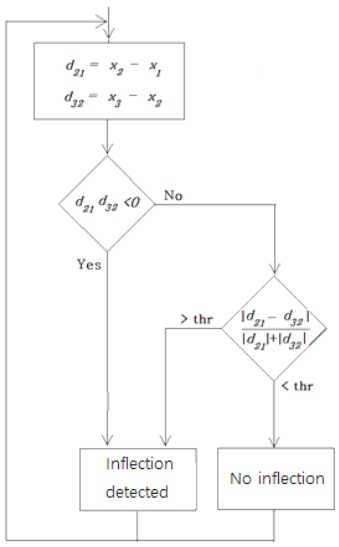

Fig. 3. Inflection detection algorithm [5].
그림 3. 변곡점 검출 알고리즘[5]

4. 제안된 기본 주파수 추정 방법

음성의 기본 주파수는 음성의 피치(pitch)와 역수 관계를 갖는다. 피치는 주기적인 신호의 시간 간격을 말한다. 따라서 음성의 기본 주파수를 추정하는 것은 피치를 추정하는 것과 동일하다. 이러한 피치는 음성의 변곡점 가운데 국소적인 최대값 사이의 시간 간격을 측정하여 얻을 수 있다.기본 주파수는 대체로 300Hz 이내의 저주파 영역에 존재하는 반면 음성은 주요 주파수 성분이 약 4000Hz까지 분포하므로 고주파에 따른 국소적인 최대값이 피치 값 사이에도 많이 존재하여 실제 피치 값 추정을 어렵게 한다. 따라서, 전처리로 저역통과필터로 음성을 처리하여 고주파 성분을 제거한 후에 피치를 추정한다. 아래 그림 4는 정상 음성의 유성음과 이를 저역통과필터로 전처리한 후의 모습을 확대한 결과이다. 그림 4. (a)의 경우 유성음과 무성음이 시간 구간별로 발생하는 모습을 보이며, 유성음의 경우에도 고주파 성분에 따른 다수의 국소적인 최대값을 볼 수 있다. 그림 4. (b)는 (a)의 음성을 저역통과필터로 전처리한 결과 신호를 보인다. 고주파 특성을 보이는 무성음 구간은 완전히 제거되고 유성음의 경우에도 고주파 성분이 제거된 모습을 보이며, 국소적인 최대값 간격에서 피치를 정확히 추정할 수 있음을 알 수 있다. 따라서 제안된 기본 주파수 추정법은 아래 그림 5와 같다. 음성은 먼저 저역통과 필터로 전처리된 후, 변곡점 검출기로 입력된다. 검출된 변곡점 IP_n 가운데 국소적인 최대값(양수값을 갖는 변곡점) 사이의 시간 간격을 측정하여 피치 정보를 얻어낸다. 즉, 연속하는 국소적인 최대값의 시간이 각각 $t(IP_n)$, $t(IP_{n-1})$ 일 때, 피치 P는 아래와 같이 주어진다.

$$P = t(IP_n) - t(IP_{n-1}) \qquad (3)$$

그리고 피치 P의 역수를 취하여 기본 주파수를 구한다. 계산량의 경우 식 (1)의 자기상관함수를 이용하는 방법은 N개의 샘플로 이뤄진 음성신호 블록에 대하여 N^2 만큼의 곱셈이 필요하나, 제안된 방법은 그림 3에서 보이듯 샘플당 4개의 덧셈과 3개의 곱셈만으로 결과를 얻

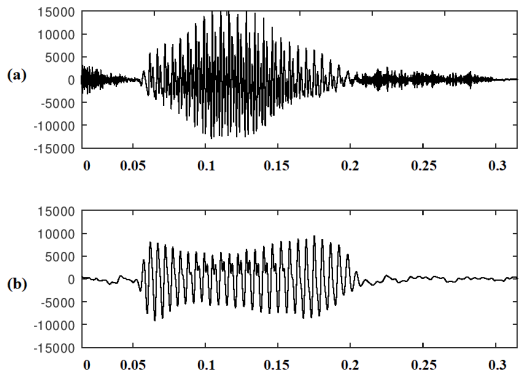

Fig. 4. (a) A speech signal and (b) preprocessed speech signal by a low pass filter.
그림 4. (a) 정상적인 음성신호와 (b) 저역통과필터로 전처리된 음성신호

어서 구현시 계산량을 절감할 수 있다. 아울러 기존의 방법은 음성이 국소적으로 시불변이라는 가정하에 기본주파수를 추정하기 때문에 일정 구간의 음성 블럭 내에서는 기본 주파수가 변하지 않고 일정한 값으로 추정된다. 반면 본 연구에서 제안된 방법은 식 (3)과 같이 음성의 이웃하는 변곡점 사이의 시간 간격으로 피치를 추정하여 기본 주파수를 구하기 때문에, 음성이 국소적으로 시불변이라는 가정이 불필요하며 음성의 시변성을 적절하게 반영한 결과를 얻는다.

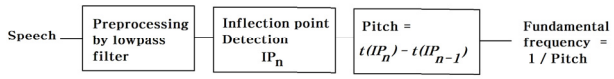

Fig. 5. Block diagram of the proposed fundamental frequency estimator.

그림 5. 제안된 기본 주파수 추정기의 구조도

5. 컴퓨터 모의실험

컴퓨터 모의실험으로 제안된 방법의 유용성을 확인한다. 모의실험은 유성음만으로 이뤄진 문장을 대상으로 한다. 동일한 문장을 성인 여성과 성인 남성이 발성한 신호를 대상으로 실험하여 남녀의 기본 주파수 차이 또한 비교한다. 그림 6과 그림 7은 여성의 음성을 처리한 결과이다. 그림 6 (a)와 (b)는 각각 정상신호와 저역통과필터로 전처리된 신호를 보인다. 그림 7 (a)는 그림 6 (b)의 전처리된 신호에서 변곡점을 검출한 결과를 보인다. 그림 5의 기본 주파수 추정기는 그림 7 (a)의 변곡점 가운데 국소적인 최대값을 취하여 유성음의 피치 즉 기본 주파수를 추정한다. 그림 7 (b)는 제안된 방법으로 추정된 유성음의 기본 주파수를 보인다. 함께 표시된 적색 별표는 기존의 자기상관함수를 이용한 방법의 결과이다. 두 방법 모두 시험 구간에서 대체로 200Hz 부근에서 기본 주파수의 추정치를 보인다. 기존의 자기상관함수를 이용한 방법은 매 20msec의 블록에 대한 결과를 보여주는 반면, 제안된 방법은 모든 시간 구간에서 변화되는 기본 주파수 추정치를 보여준다.

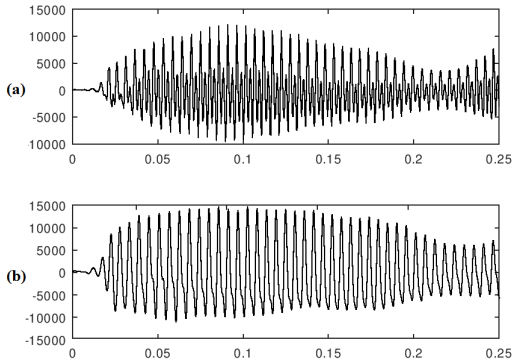

Fig. 6. Female speech signals (a) normal signal, (b) preprocessed signal.

그림 6. 여성의 음성신호 (a) 정상신호, (b) 전처리된 신호

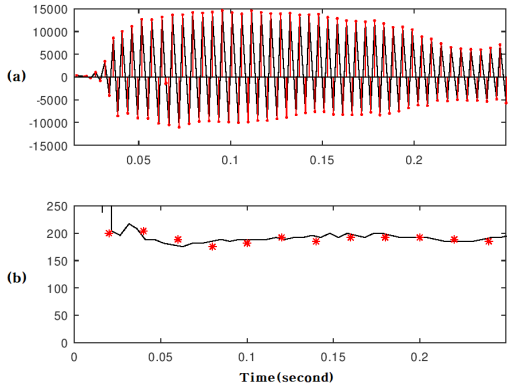

Fig. 7. Female speech case (a) inflection point detection results, (b) proposed fundamental frequency estimation result with autocorrelation method results in red star.

그림 7. 여성의 음성신호 처리결과 (a) 변곡점 검출결과, (b) 제안된 기본 주파수 추정 결과와 자기상관함수 방법 결과 (적색별표)

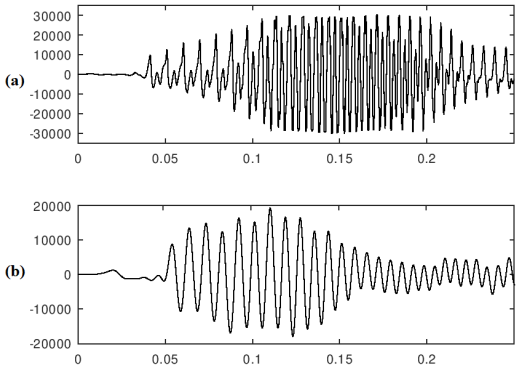

Fig. 8. Male speech signals (a) normal signal, (b) preprocessed signal.

그림 8. 남성의 음성신호 (a) 정상신호, (b) 전처리된 신호

그림 8 (a)와 (b)는 각각 남성의 음성신호와 전처리된 신호를 보인다. 그림 9 (a)는 전처리된 신호에서 변곡점을 검출한 결과를 보인다. 그림 9 (b)는 대체로 100~180Hz 사이에서 시간과 함께 기본 주파수가 변화하는 양상을 볼 수 있다. 여기에서도 적색 별표는 기존의 자기상관함수를 이용한 기본 주파수 추정치 결과를 보이며 매 20 msec 블록에 대한 결과를 보인다. 그림 7 (b)와 그림 9 (b)의 기본 주파수 추정치를 비교하면 여성과 남성의 기본 주파수의 차이를 볼 수 있다. 대체로 남성의 기본 주파수 추정치가 여성의 추정치보다 낮은 값을 보인다.

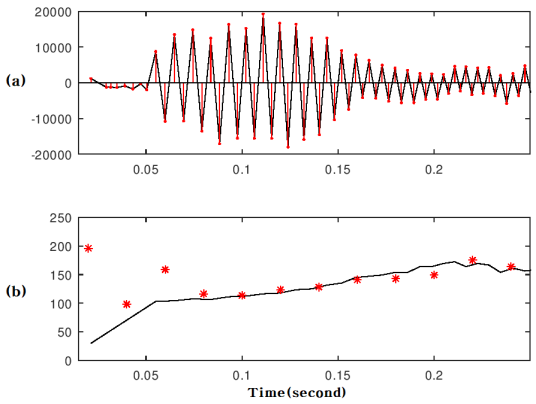

Fig. 9. Male speech case (a) inflection point detection results, (b) proposed fundamental frequency estimation result with autocorrelation method results in red star.

그림 9. 남성의 음성신호 처리결과 (a) 변곡점 검출결과, (b) 제안된 기본 주파수 추정 결과와 자기상관함수 방법 결과 (적색별표)

Ⅲ. 결론

변곡점 추출 방법을 이용하여 음성의 피치 혹은 기본 주파수를 추정하는 방법을 제안하였다. 기본 주파수는 해부학적인 특성으로 음성인식, 음성코딩, 화자인식 등에 활용된다. 기존의 방법은 음성을 블록 단위로 처리하여 음성의 시변성을 적절하게 반영하지 못하는 반면, 제안된 방법은 시간에 따른 기본 주파수의 변화를 적절하게 보여준다. 컴퓨터 시뮬레이션을 통해 음성의 기본 주파수를 적절하게 추정함을 확인하여 제안된 기본 주파수 추정법의 유용성을 보였다.

References

[1] L. R. Rabiner and R. W. Schafer, *Digital Processing of Speech Signals*, Englewood Cliffs, NJ, Prentice-Hall, 1978.

[2] T. F. Quatieri, *Discrete-Time Speech Signal Processing: Principles and Practice*, Upper Saddle River, NJ, Prentice-Hall, 2002.

[3] A. M. Kondoz, *Digital Speech: Coding for Low Bit Rate Communication Systems*, West Sussex, England, Wiley, 1994.

[4] D. O'Shaughnessy, *Speech Communication: Human and Machine*, MA, Addison-Wesley, 1987.

[5] B. Iem, "Instantaneous frequency estimation of AM-FM signals using the inflection point detection," *Journal of Inst. Korean Electrical and Electronics Engineers*, vol.24, no.4, pp.1081-1085, 2020. DOI: 10.7471/ikeee.2020.24.4.1081

[6] B. Iem, "Power disturbance detection using the inflection point estimation," *Journal of Inst. Korean Electrical and Electronics Engineers*, vol.25, no.4, pp.710-715, 2021. DOI: 10.7471/ikeee.2021.25.4.710

[7] B. Iem, "A Nonuniform Sampling Technique based on Inflection Point Detection and its Application to Speech Coding," *Journal of Acoustical Society of America*, vol.136, no.2, pp.903-909, 2014. DOI: 10.1121/1.4884882

BIOGRAPHY

Byeong-Gwan Iem (Member)

1988 : BS degree in Electronic Engineering, Yonsei University.
1990 : MS degree in Electronic Engineering, Yonsei University.
1998 : PhD degree in Electrical Engineering, University of Rhode Island.
1999~2001 : Senior Research Engineer, Samsung Electronics.
2002~Present : Professor, Gangneung-Wonju National University