

Fundamental Frequency Estimation based on Time-Frequency Analysis

林 炳 寬[†]
(Byeong-Gwan Iem)

Abstract - A simple robust fundamental frequency estimator on the time-frequency domain is proposed. Combined with the appropriately designed low-pass filter, the instantaneous frequency estimator based on the Teager-Kaiser energy function can detect the fundamental frequency of speech signal. The Teager-Kaiser function can be obtained through real computation and show the change of frequency as time goes. And when a speech block with N samples is processed with a lowpass filter with length of L , it requires $O(N \cdot (L+5))$ operations, compared to $O(N \cdot 2\log_2 N + L)$ operations in the recently introduced wavelet and conventional instantaneous frequency method. The computer simulation confirms the usefulness of the proposed fundamental frequency estimation method.

Key Words : Instantaneous Frequency, Fundamental Frequency, Pitch, 순간주파수, 기본주파수

1. 서 론

음성의 기본주파수 혹은 피치의 검출은 음성코딩, 음성인식, 화자인식 등을 위한 오래된 과제이다. 따라서 많은 연구가 있어왔고, 이미 상용화된 신뢰성 있는 다양한 기법들이 있다[1]. 이러한 기법들로는 자기상관함수(Autocorrelation function)을 이용한 기법, Average Magnitude Difference Function (AMDF)을 이용한 방법, Zero Crossing Rate(ZCR)를 이용하는 방법, 캡스트럼을 이용한 기법 등이 있다[1]. 이러한 방법들은 음성의 특성이 국지적으로 일정 시간동안 변하지 않는다는 가정하에 적용한다.

시간의 경과에 따라 특성이 변하는 시변신호의 분석 방법으로 시간주파수 분석법(Time-Frequency Analysis)은 1990년대에 주목을 받았고, 많은 연구성과가 보고되었다[2]. 대표적인 시간주파수 분석법으로는 Short-Time 푸리에 변환, Wavelet 변환, 스펙트로그램, 스케로그래프, 위그너 함수, 순간주파수 등이 있다 [2]. 음성신호와 같이 일반적으로 그 특성이 시변인 신호의 분석에 시간주파수 분석분야의 연구성과를 적용하는 시도가 보고되고 있다[3-6]. 대표적인 연구결과는 S. Kadambe와 Boudreaux-Bartels의 wavelet 변환을 이용한 음성분석[3], Koh 등의 wavelet 변환과 순간주파수를 이용한 기본주파수의 검출[4], L. Janer의 modulated Gaussian wavelet을 갖는 스케로그래프로 음성신호를 Bark scale로 17개의 밴드로 나눈 피치검출법[5], STFT을 이용한 피치검출법[6] 등이 있다. 본 논문에서는 복소 연산과 비선

형 연산이 필요한 고전적인 순간주파수 추정법[4]에 비하여 계산량이 적게 소요되는 Teager-Kaiser 함수에 기반한 순간주파수 검출기를 이용한 음성의 기본주파수 검출법을 제안하고자 한다. 고전적인 순간주파수 추정법에서는 실신호를 먼저 아날리틱 신호화하여 위상성분을 얻는다. 순간주파수는 이렇게 얻어진 위상성분을 미분하여 얻어진다 [4]. 반면에 Teager-Kaiser 함수에 기반한 순간주파수 검출기는 이웃하는 샘플간의 곱과 뺄셈만으로 의미있는 통계치를 얻으며, 이러한 통계치를 이용하여 순간주파수를 얻는다. 본 논문에서는 저역통과 필터를 사용하여 음성신호에 포함된 고주파수 성분을 제거하고, 이 신호에 Teager-Kaiser함수에 기반한 순간주파수 추정기를 적용하였다. 본 논문의 구조는 다음과 같다. 2장에서는 본 논문에서 제안하는 기본주파수 검출기의 주요한 구성요소인 Teager-Kaiser 함수를 소개한다. Teager-Kaiser함수는 순간주파수와 신호의 진폭이 결합된 형태를 보이므로, Teager-Kaiser 함수에서 순간주파수를 추출하는 기법을 소개한다. 아울러, 얻어진 순간주파수에서 음성의 기본주파수를 결정하는 방법을 소개한다. 3장에서는 본 논문에서 제안하는 기본주파수 검출기의 계산량을 분석한다. 4장에서는 컴퓨터 모의시험으로 본 논문에서 제안한 기본주파수 검출법의 성능을 검증한다

2. 저역통과필터와 Teager-Kaiser함수를 이용한 기본주파수 추정기

그림 1.은 본 논문에서 제안하는 간단한 기본주파수검출기의 구조를 보인다. 먼저 음성신호는 저역통과필터로 처리되어, 고주파 성분이 제거된 신호만을 추출한다. 이는 전 처리 단계로서 기본주파수 성분이 통상 저주파수 대에 존재한다는 관찰에 따른 것이다. 이렇게 전처리를 거친 음성신호를

[†] 교신저자, 正會員 : 江陵大 工大 電子工學科 助教授 · 工博

E-mail : ibg@kangnung.ac.kr

接受日字 : 2005年 10月 6日

最終完了 : 2005年 12月 14日

Teager-Kaiser 함수로 처리하여 기본주파수 정보를 검출한다. 아래에서는 본 논문에서 제안하는 기본주파수 검출기의 주요 부분을 간략히 살펴본다.



그림 1 기본주파수 검출기의 구조

Fig. 1 Structure of the fundamental frequency estimator

2.1 저역통과필터

순간 주파수 추정기의 입력으로 적정한 신호를 얻기 위하여 본 논문에서는 다음과 같은 저역통과필터로 음성신호를 전처리하였다. 통상 남성음성인 경우 250 Hz이하에 여성의 경우 500Hz이하에 기본 주파수 성분이 있음을 감안하여 저역통과 필터의 통과대역을 결정한다. 본 논문에서는 간단한 필터설계기법인 window함수 기법을 이용하여 저역통과 필터를 설계하였다. 남성 음성을 처리하는 경우 샘플링주파수는 10kHz, 통과대역은 250Hz, window함수로는 Hanning window를 사용하였다 [7]. 아래 그림 2.은 얻어진 필터의 임펄스 응답과 주파수 특성이며, 그림 3.은 남성이 발성한 /b/소리를 저역통과 필터로 처리하여 고주파 성분은 제거된 결과를 보인다.

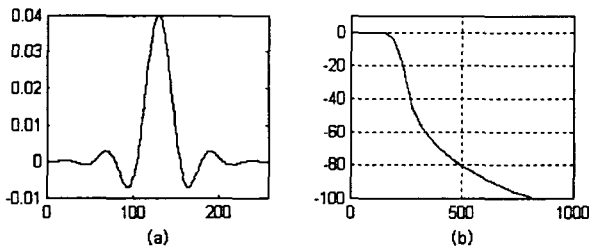


그림 2 저역통과 필터의 (a)임펄스응답과 (b)주파수 특성

Fig. 2 (a) Impulse response and (b) Frequency response of the lowpass filter

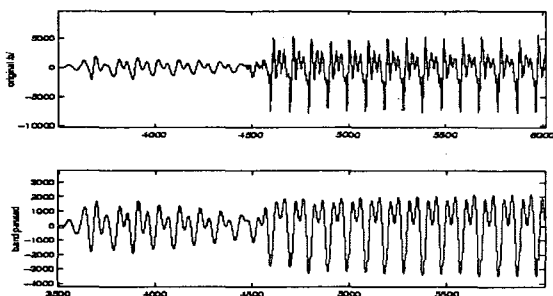


그림 3 음성/b/의 원신호와 저역통과 필터된 신호

Fig. 3 Raw signal and lowpass filtered signal of a male speech sound /b/

2.2 Teager-Kaiser 함수를 이용한 순간 주파수 검출

이산신호 $x[n]$ 에 대하여 Teager-Kaiser 함수는 아래와 같이 정의된다 [8,9]

$$\Psi(x[n]) = x^2[n] - x[n-1] \cdot x[n+1]. \quad (1)$$

이산 AM-FM 신호 $x[n]$ 이 아래와 같이 정의될 때,

$$x[n] = a[n] \cos(\phi[n]) = a[n] \cos(\omega_c n + \omega_m \int q[m] dm + \theta)$$

Teager-Kaiser 함수를 이용하여 순간 주파수 성분을 유도할 수 있다. 일반적으로 순간주파수 $\omega[n]$ 는 아래와 같이 정의된다.

$$\omega[n] = \frac{d}{dn} \phi[n] = \omega_c + \omega_m q[n]$$

Maragos 등은 [8,9]에서 진폭 $a[n]$ 이 대역 제한되어 있고, FM 인자 ω_m 이 반송파 ω_c 에 비하여 작을 때, $x[n]$ 의 Teager-Kaiser함수와 이산미분으로서 symmetric difference 값 $y[n] = (x[n+1] - x[n-1])/2$ 의 Teager-Kaiser 함수는 각각 아래와 같이 근사됨을 보였다.

$$\Psi(x[n]) \approx a^2[n] \sin^2(\omega[n])$$

$$\Psi(y[n]) \approx a^2[n] \sin^4(\omega[n])$$

위 식을 정리하면 순간 주파수는 아래 식과 같이 근사적으로 주어진다.

$$\omega[n] \approx \sin^{-1} \left(\sqrt{\frac{\Psi(x[n+1]) - \Psi(x[n-1])}{4\Psi(x[n])}} \right) \quad (2)$$

Maragos 등은 [9]에서 순간 주파수 $\omega[n]$ 를 음성신호의 formant성분을 추정하는데 사용하였다. 본 논문에서는 기본 주파수 추정을 위하여 위식의 $\omega[n]$ 를 더욱 간략하게 근사화한다. 음성신호의 표본 주파수를 고려할 때, 기본 주파수는 통상 500Hz 미만에 존재하여 비교적 낮은 주파수라 할 수 있다. 위 식 (2)에서 $\sqrt{\frac{\Psi(x[n+1]) - \Psi(x[n-1])}{4\Psi(x[n])}}$ 의 값이 비교적 작으면 순간 주파수 $f[n]$ 는 다음과 같이 근사화를 할 수 있다.

$$f[n] \approx \frac{1}{2\pi} \sin^{-1} \left(\sqrt{\frac{\Psi(x[n+1]) - \Psi(x[n-1])}{4\Psi(x[n])}} \right) \approx \frac{1}{2\pi} \sqrt{\frac{\Psi(x[n+1]) - \Psi(x[n-1])}{4\Psi(x[n])}}$$

2.3 Decision rule

2.2절에서 구한 순간주파수 $f[n]$ 는 무성음과 유성음 구간에서 다른 특성을 보인다. 기본적으로 시간의 경과에 따른 주파수의 변화를 보이는 $f[n]$ 은 그자체로서 고주파 신호로 볼 수 있지만, 유성음 구간과 비교해서 무성음 구간에서 더욱 큰 폭의 변화 범위를 갖는다. 따라서 본논문에서는 기본 주파수를 구하는 중간과정으로 이동분산값을 구하여, 유성음과 무성음을 구분하는 처리과정을 포함시킨다. 이동분산값 $\sigma^2[n]$ 은 다음과 같이 주어진다.

$$\sigma^2[n] = \frac{1}{M} \sum_{m=0}^{M-1} (f[n-m] - \overline{f[n]})^2$$

여기에서 $\overline{f[n]}$ 은 순간주파수의 이동평균치이며, M은 이동평균을 구하는 구간의 길이이다. 본 논문에서는 global variance의 85%를 기준으로 유성음 구간과 무성음 구간을 구분하였다. 아래 그림 4은 남성의 /b/ 소리를 본 논문에서

제안한 주파수 추정법으로 처리하여 얻어진 순간주파수의 이동분산값을 구한 결과이다. Global variance의 85% 값 (1.25×10^7)도 표시되어 있다.

얻어진 순간주파수가 의미있는 값인지를 판단하는 decision규칙으로는 Koh 등이 제안한 방법을 준용하였다 [4]. Koh 등은 음성의 특성을 고려하여, 얻어진 순간주파수가 기본주파수로서 합당한지를 다음과 같은 기준에 따라 판단하였다.

1. 순간주파수 500Hz보다 크다.
2. 순간 주파수가 50 Hz보다도 작다.
3. 이웃하는 순간주파수 간에 차이가 1.4Hz보다도 크다.
4. 1.4Hz보다도 적은 변화를 보이는 주파수가 20 msec 이상 유지되지 않는다.

위의 조건중 하나라도 만족되면, 그 순간주파수는 기본주파수가 아닌 것으로 판단한다. 즉 그 음성구간은 무성음이거나 묵음 구간인 것으로 판단한다.

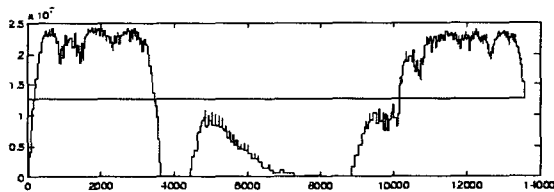


그림 4 /b/ 소리의 순간주파수의 이동분산값 $\sigma_f^2 [n]$
 Fig. 4 Moving variance value of the instantaneous frequency of /b/ sound

3. 기본주파수 검출기의 계산량 분석

이장에서는 본 논문에서 제안하는 기본주파수 검출기와 Koh등이 제안한 기본주파수 검출기 [4]의 계산량을 비교한다. Koh 등이 사용한 순간주파수 추정기는 아래와 같이 정의된다.

$$f[n] = \frac{f_s}{4\pi} \left\{ \tan^{-1} \left(\frac{\text{Im} [z[n+1]]}{\text{Re} [z[n+1]]} \right) - \tan^{-1} \left(\frac{\text{Im} [z[n-1]]}{\text{Re} [z[n-1]]} \right) \right\}_{\text{mod } 2\pi}$$

여기에서 $z[n]$ 은 음성신호 $x[n]$ 의 analytic 신호이다. 아래 표에서는 기본주파수를 검출해야할 음성데이터 블록의 길이가 N 으로 정해져 있을 때, 제시된 방법을 수행하는데 요구되는 계산량을 비교한다. 순간주파수의 저역통과 필터의 길이는 L 을 가정한다.

표 1 기본주파수 검출방법의 계산량 비교

Table 1 Comparison of the computational cost of the fundamental frequency estimator

구분	Wavelet+ 순간주파수	Teager-Kaiser 추정기를 이용한 방법
Analytic 신호 검출	$2N \log_2 N$	
순간주파수	$2N$ (+2N 비선형연산)	$5N$ (+N 비선형연산)
저역통과 필터	$LN+N$ (+N 비선형연산)	LN
합	$N(2 \log_2 N + L + 3)$ (+3N 비선형연산)	$N(L+5)$ (+N 비선형연산)

표 1은 기본주파수 검출법에 소요되는 계산량을 비교분석한 결과이다. 기존의 wavelet변환과 고전적인 순간주파수를 이용한 방법[4]이 대략 $O(N \cdot 2 \log_2 N + L)$ 의 계산량을 요구하는데 비하여 본 논문에서 제안하는 방법은 $O(N \cdot (L + 5))$ 의 계산량을 필요로 한다. 특히, wavelet변환과 순간주파수를 이용하는 방법은 arctan나 exponential 함수와 같은 많은 비선형 연산이 소요된다. 따라서, 같은 시간주파수 분석법의 범주에 속하지만, 본 논문에서 제안된 방법이 wavelet변환과 순간주파수를 이용한 방법에 비교해서 경제적이다.

4. 컴퓨터 모의실험

이장에서는 실제 컴퓨터 모의실험을 통하여 본 논문에서 제안하는 방법의 기본주파수 검출기로서의 효용을 확인한다. 시험에 사용하는 음성데이터는 10kHz로 샘플링된 신호로 남성이 발성한 /b/소리이다. 처리된 결과는 아래 그림 5와 같다. 그림에서 위에는 원 음성신호를 보여주고 있고, 두 번째 신호는 본 논문에서 제안한 방법으로 처리한 결과이다.

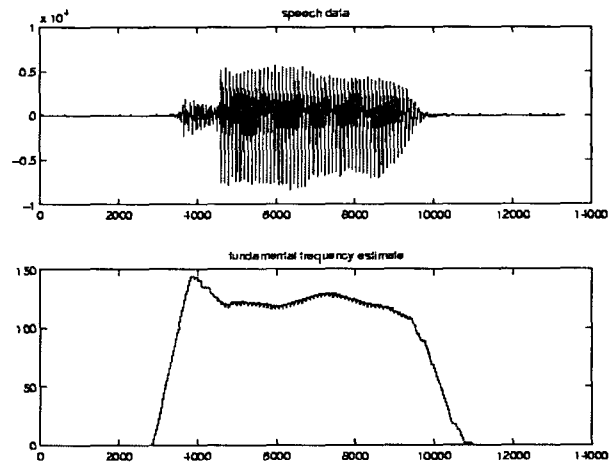


그림 5 남성의 /b/ 소리의 처리 결과
 Fig. 5 Processing result of a male /b/ sound

위 그림의 처리결과를 보면, 원 음성은 유성음 성분이 강함을 예측할 수 있다. 아래의 기본주파수값을 보면 /b/ 소리중에 기본주파수가 소리의 진행과 함께 완만하게 변하는 모습을 볼 수 있다. 따라서 주기인 피치 또한 완만히 변할 것을 예측할 수 있으며, 현재의 음성구간이 유성음 구간임을 알 수 있다. 따라서, Teager-Kaiser 함수에 기반한 기본주파수 추정법을 다른 피치 검출기법들과 마찬가지로 유/무성음 판정기로도 사용될 수 있음을 알 수 있다. 완전한 한 문장에 대하여 처리한 결과는 아래 그림 6과 같다. 음성은 영어 문장으로 "Should we chase those cowboys?"라는 문장이다. 남성이 발성한 문장이며, 10kHz로 샘플링되었다.

20ms에서 40ms에 이르는 프레임 단위로 처리하는 기존의 기본주파수 추정법은 프레임별로 특정한 기본주파수를 검출하는데 비하여, 순간주파수 추정기를 이용한 기본주파수 검출기는 매순간마다 변화하는 기본주파수의 추이를 관찰할

수 있다. 본 논문에서 제안한 기본주파수 추정기의 이러한 특징은 음성인식 또는 화자인식과 같은 분야에서 활용할 수 있을 것이다.

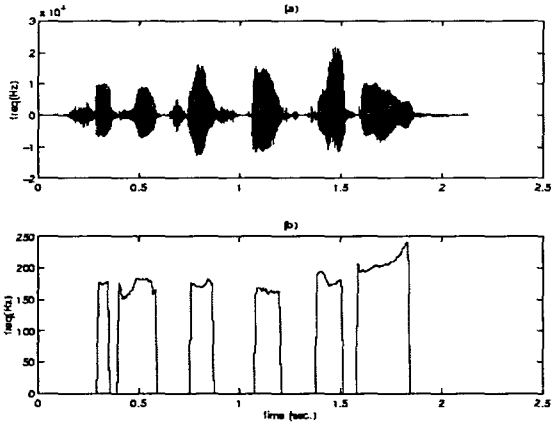


그림 6 완전한 한 문장에 대하여 처리한 결과
Fig. 6 Processing result of a complete sentence

5. 결 론

본 논문에서는 비교적 간단한 구조를 갖는 시간 주파수 영역에서의 음성의 기본주파수 검출방법을 제안하였다. Teager-Kaiser함수는 시간의 경과에 따른 신호의 주파수 변화를 보여주는 함수로서 실수값연산을 하는 장점이 있다. 저역통과필터 및 적절한 decision 규칙과 결합될 때, 음성의 기본주파수 검출에 사용될 수 있음을 보였다. 컴퓨터 모의 실험으로 제안한 방법이 기본주파수 검출법으로 타당함을 확인하였다. 계산량에 있어서도 기존의 wavelet변환과 고전적인 순간주파수를 이용한 방법[4]이 대략 $O(N \cdot 2 \log_2 N + L)$ 의 계산량을 요구하는데 비교하여 본 논문에서 제안하는 방법은 $O(N \cdot (L + 5))$ 의 계산량을 필요로 하여 경제적이다.

감사의 글

이 논문은 2003년도 강릉대학교 학술연구조성비 지원에 의하여 수행되었음

참 고 문 헌

[1] T. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice, Prentice-Hall, 2002.
 [2] F. Hlawatsch and G.F. Boudreaux-Bartels, "Linear and Quadratic Time-Frequency representations," IEEE Signal Processing Magazine, 1990.
 [3] S. Kadambe, G.F. Boudreaux-Bartels, "Application of the Wavelet transform for pitch detection of speech signals", IEEE Trans. On Information theory, vol. 38,

pp. 917-924, March, 1992.

[4] L. Qiu, H. Yang, S. Koh, "Fundamental frequency determination based on instantaneous frequency estimation", Signal Processing, vol. 44, pp.233-241, 1995.
 [5] L. Janer, "New pitch detection algorithm based on wavelet transform", IEEE TF/TS symposium, pp. 165-168, Pittsburgh, PA, Oct. 1998.
 [6] D. Liu and C. Lin, "Fundamental frequency estimation based on the joint time-frequency analysis of harmonic spectral structure", IEEE Trans. On Speech and Audio Processing, vol. 9, pp. 609-621, September, 2001.
 [7] P. Lynn and W. Fuerst, Introductory Digital Signal Processing with Computer Applications, Wiley, 2000.
 [8] P. Maragos, J. Kaiser, T. Quatieri, "On amplitude and frequency demodulation using energy operator", IEEE Trans. On Signal Processing, vol. 41, pp. 1532-1550, April, 1993.
 [9] P. Maragos, J. Kaiser, T. Quatieri, "Energy separation in signal modulations with application to speech analysis", IEEE Trans. On Signal Processing, vol. 41, pp. 3024-3051, October, 1993.

저 자 소 개



임 병 관 (林炳寬)

1965년 4월 3일생.
 1988년 연세대 전자공학과 졸업.
 1998년 Univ. of Rhode Island 전기공학과 졸업(공박)
 2002~현재 강릉대 전자공학과 조교수
 Tel : 033-640-2426
 Fax : 033-646-0740
 E-mail : ibg@kangnung.ac.kr