

대칭구조를 갖는 일반적인 고차의 미분 에너지함수를 기반한 순간주파수를 이용한 음성의 기본주파수 추정

논 문
60-12-31

Estimation of Fundamental Frequency Using an Instantaneous Frequency Based on the Symmetric Higher Order Differential Energy Operator

임 병 관*
(Byeong Gwan Iem)

Abstract - The fundamental frequency of the voiced speech is estimated using the instantaneous frequency based on the symmetric higher order differential energy operator. The instantaneous frequency based on the symmetric higher order energy operator shows better frequency estimation result since it is aligned to the time instance of the signal. The speech is pre-processed by a lowpass filter to remove higher frequency components. Then, it is processed by the instantaneous frequency to obtain the fundamental frequency estimates. The symmetric higher order energy operator has been used as an indicator to determine the voiced/unvoiced speech. The fundamental frequency estimates are further processed by a moving average filter to obtain the monotonically changed estimates. The obtained fundamental frequency estimates have been compared with the spectrogram of the speech to confirm its accuracy.

Key Words : Fundamental frequency, Instantaneous frequency, Pitch estimation

1. 서 론

음성발생모델에 따르면 유성음은 주기적으로 반복되는 단위 충격파를 입력신호로 갖는 선형시변시스템(Linear Time Varying system)의 출력으로 볼 수 있다[1]. 이때 단위 충격파신호의 주기를 피치라고 부르며, 피치의 역수가 유성음의 기본주파수가 된다. 이러한 피치 또는 기본주파수 추정은 음성인식, 음성코딩 등 다양한 응용분야에서 중요하다. 따라서 많은 연구가 진행되어 왔으며, 대표적인 결과로 자기상관함수(Autocorrelation function), AMDF(Average Magnitude Difference Function), ZCR(Zero Crossing Rate), 캡스트럼 방법 등이 있다[1]. 이러한 기본주파수 추정 방법들은 음성이 국소적으로 시불변이라는 가정하에 음성을 10~40 msec의 블록단위로 취하여 기본주파수를 추정한다. 하지만 음성은 본질상 시간에 따라 변하는 특성을 보인다[1]. 시간에 따라 변하는 특성을 반영한 기본주파수를 추정한다면 음성인식, 화자인식과 같은 분야에서 신뢰도를 향상시킬 수 있을 것이다.

시간과 함께 특성이 변하는 신호의 분석도구로서 시간주파수 분석법이 있으며, 음성과 같은 시변신호의 분석에 적합하다 [1,2]. 따라서 시간주파수 분석법을 활용하면 음성이 국부적으로는 시불변이라는 가정이 불필요하다. 시간주파수 분석법으로는 신호의 크기 또는 에너지를 표현하는 시간주파수 표현법(Time-frequency representations), 매 시간의 주파수를 보이는 순간주파수 등 다양한 방법들이 연구, 소개되

어 왔다 [2,3]. 시간주파수 표현법은 시간과 주파수의 함수로서 신호의 크기 또는 에너지를 표현하는 함수이며, 대표적인 예로는 스펙트로그램, 스켈로그램, 위그너함수 등이 있다 [2]. 이러한 시간주파수 표현법은 시간의 변화에 따른 음성의 에너지 변화를 주파수 별로 보여주므로 음성의 피치분석, 포먼트 분석 등에서 유용하다. 하지만, 시간과 주파수라는 2차 평면상에서의 값으로 주어지므로 계산량이 많은 단점이 있으며, 위그너함수와 같은 2차 함수의 경우는 자체적으로 간섭신호가 존재하여 신호해석을 어렵게 하는 문제가 있다 [2]. 이와 달리 순간주파수는 시간의 함수로서 주파수를 표현하며 매 순간 신호의 주파수 값을 출력한다. 따라서 신호의 크기나 에너지 보다는 주파수의 값이 요구되는 응용에 적합하다. 다만, 한 순간에 하나의 주파수 값을 출력하므로 다수의 주파수 성분이 있는 신호의 분석에는 적합하지 않다. 그러나 적절하게 전 처리를 한 음성의 기본주파수 추정과 같은 응용에 적용할 수 있다 [4].

순간주파수를 구하는 방법으로는 신호의 위상을 미분하는 고전적인 방법 [3], Teager-Kaiser 에너지 함수를 이용하여 정의하는 방법 [5], 고차의 미분에너지를 이용하는 방법 [6], 대칭구조를 갖는 고차의 미분에너지 함수를 이용하는 방법 [7] 등이 있다. 신호의 위상을 미분하는 방법의 경우, 신호가 실신호인 경우 위상을 얻기 위하여 힐베르트 변환을 필요로 한다. Teager-Kaiser 에너지 함수를 이용한 방법은 힐베르트 변환을 필요로 하지 않으며, 덧셈과 곱셈 그리고 역 삼각 함수를 이용하여 순간주파수를 구할 수 있다. 고차의 미분 에너지 함수를 이용하는 방법은 Teager-Kaiser방법에 비교해서 계산량이 적고 추정 성능이 우수하지만 [6], 신호와 시간 동기가 되지 않는 단점이 있다. 대칭구조를 갖는 고차의 미분에너지 함수에 기초한 순간주파수는 신호와 시간 동기가 되며, 따라서 여타의 방법에 비교해서 우수한 주파수 추

* 정 회 원 : 강릉원주대 전자공학과 부교수

E-mail : ibg@gwnu.ac.kr

접수일자 : 2011년 6월 3일

최종완료 : 2011년 10월 25일

정성능을 보인다 [7]. 본 논문에서는 대칭구조를 갖는 고차의 미분에너지 함수에 기초한 순간주파수를 이용하여 음성의 기본주파수를 추정한다.

순간주파수를 음성에 적용하려면 음성신호가 분석 시점에 하나의 주요 주파수 성분을 가져야 하므로 적절하게 설계된 저역통과필터로 음성을 처리한다. 전 처리된 음성신호는 순간주파수 추정기의 입력으로 사용되며, 얻어진 순간주파수는 음성의 에너지 함수와 비교하여 무성음과 목음 구간의 순간주파수는 무시된다. 얻어진 순간주파수는 매 표본순간에 급격하게 변화하는 값을 가질 수 있으므로, 이동평균 필터로 처리하여 시간에 따라 단조변화(monotonically changing)하는 함수로서 음성의 기본주파수를 추정한다.

본 논문의 구조는 다음과 같다. 2장에서는 순간주파수의 고전적인 정의와 대칭구조를 갖는 고차의 에너지 함수를 이용한 순간주파수를 소개한다. 아울러 사용되는 순간주파수 추정기의 추정성능을 소개한다. 3장에서는 제안된 음성의 기본주파수 추정기의 구조를 소개한다. 아울러 3장에서는 유성음, 무성음 및 목음 구간을 분류하기 위한 에너지함수의 특성을 고찰한다. 4장에서는 컴퓨터 모의시험으로 순간주파수를 이용한 제안된 기본주파수 추정기의 예제를 소개하고 5장에서 결론을 정리한다.

2. 순간주파수와 에너지 함수

순간주파수는 한 순간에 하나의 주파수, 즉 시간의 함수로서 주파수를 보인다. 고전적인 정의는 아래와 같다.

2.1 순간주파수의 고전적인 정의

순간주파수는 신호의 위상을 시간에 대하여 미분을 취하는 것으로 정의되었다 [2,3]. 복소 정현파 신호의 위상과 주파수 사이의 관계에서 이러한 고전적인 정의를 확인할 수 있다. $x(t) = a(t)e^{j\phi(t)} = a(t)e^{j2\pi f_0 t}$ 라면 위상 $\phi(t) = 2\pi f_0 t$ 이며, 위상의 미분 $f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} = f_0$ 로 직관적인 주파수 f_0 와 같다. 만약 신호가 실수 값을 갖는다면, 신호는 먼저 힐베르트 변환을 이용하여 아날리틱신호로 만든 후, 위상신호를 미분하여 순간주파수를 구한다. 즉 실신호 $x(t)$ 에 대하여 아날리틱신호 $x_a(t)$ 는 다음과 같다. $x_a(t) = x(t) + jHx(t) = a(t)e^{j\phi(t)}$. 여기에서 $x_a(t)$ 는 $x(t)$ 의 아날리틱 신호로 복소신호, $Hx(t)$ 는 $x(t)$ 의 힐베르트 변환, $a(t)$ 는 $x_a(t)$ 의 진폭신호, $\phi(t)$ 는 $x_a(t)$ 의 위상신호이다. 순간주파수 $f(t)$ 는 위상의 미분이므로 $f(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt}$ 로 정의된다 [2]. 이산신호에 대하여는 순간주파수는 다음과 같이 차동방정식을 이용하여 구할 수 있다. $f(n) = \frac{1}{4\pi} \cdot (\phi(n+1) - \phi(n-1))$ 여기에서 $\phi(n)$ 은 이산시간 영역에서의 위상이다. 이산신호의 경우는 위상의 차동방정식으로 미분을 근사화하므로 다양한 정의가 가능하지만, 일반적으로 분석시간 n 에 대하여 대칭인 위의 정의가 사용된다. 이와 같은 고전적인 정의는 힐베르트 변환을 필요로 하기 때문에 상대적으로 많은 계산

량을 요구한다.

2.2 대칭구조를 갖는 고차의 미분에너지함수를 이용한 순간주파수 정의

k 차의 미분에너지함수 $\Gamma_k\{x(t)\}$ 는 다음과 같이 정의된다 [6].

$$\Gamma_k\{x(t)\} = x'(t)x^{(k-1)}(t) - x(t)x^{(k)}(t). \quad (1)$$

$k=2$ 일 때, $\Gamma_k\{x(t)\}$ 는 Teager-Kaiser 에너지 함수 $\Gamma_2\{x(t)\} = (x'(t))^2 - x(t)x''(t)$ 가 된다. 그리고 $k=3, k=4$ 일 때, 미분에너지함수는 각각 다음과 같다.

$$\Gamma_3\{x(t)\} = x'(t)x^{(2)}(t) - x(t)x^{(3)}(t)$$

$$\Gamma_4\{x(t)\} = x'(t)x^{(3)}(t) - x(t)x^{(4)}(t).$$

이러한 고차의 미분에너지함수를 이용하여 Maragos와 Potamianos는 신호 $x(t) = A \cdot \cos(2\pi ft + \theta)$ 일 때, 주파수 성분 $f(t)$ 를 구하는 알고리즘을 아래와 같이 유도하였다 [6].

$$f(t) = \frac{1}{2\pi} \sqrt{\frac{-\Gamma_4\{x(t)\}}{\Gamma_2\{x(t)\}}} \quad (2)$$

이산신호에 대하여는 고차의 미분에너지함수를 다음과 같이 정의된다 [6].

$$\Gamma_k\{x(n)\} = x(n)x(n+k-2) - x(n-1)x(n+k-1) \quad (3)$$

$k=1$ 일 때, (3)의 미분에너지함수는 $\Gamma_1\{x(n)\} = 0$ 이다. $k=2$ 일 때 미분에너지함수는 Teager-Kaiser 에너지함수 $\Gamma_2\{x(n)\} = x^2(n) - x(n+1)x(n-1)$ 와 같다. $k=3, k=4$ 일 때, 미분에너지함수는 각각 다음과 같다.

$$\Gamma_3\{x(n)\} = x(n)x(n+1) - x(n-1)x(n+2)z \quad (4)$$

$$\Gamma_4\{x(n)\} = x(n)x(n+2) - x(n-1)x(n+3) \quad (5)$$

이와 같은 고차의 미분에너지함수를 이용하여 이산신호에 대하여 순간주파수를 아래와 같이 정의할 수 있다 [6].

$$f(n) = \frac{1}{2\pi} \cos^{-1} \left(\frac{\Gamma_3\{x(n)\}}{2 \cdot \Gamma_2\{x(n)\}} \right) \quad (6)$$

위의 이산신호에 대한 순간주파수는 임의의 차수 k 에 대하여 다음과 같이 일반화 될 수 있다.

$$f(n) = \frac{1}{2\pi} \frac{1}{(k-1)} \cos^{-1} \left(\frac{\Gamma_{2k-1}\{x(n)\}}{2 \cdot \Gamma_k\{x(n)\}} \right) \quad (7)$$

이와 같은 고차의 미분에너지함수는 분석 시간 n 에 대하여

사용되는 표본 시간이 (n-1), (n), (n+k-2), (n+k-1), (n+2k-3), (n+2k-2)로 비대칭적인 구조를 갖는다. 따라서 고차의 에너지함수를 이용한 순간주파수 추정기 (6), (7)은 시간동기가 이루어지지 않아 주파수 추정 성능이 저하되는 현상이 발생한다. 따라서 아래와 같은 대칭구조를 갖는 미분에너지함수를 정의 한다 [7].

$$\Xi_k \{x(n)\} = \begin{cases} \frac{\Gamma_k \{x(n)\} + \Gamma_k \{x(n-k+2)\}}{2} k odd \\ \Gamma_k \left\{ x \left(n - \frac{k}{2} + 1 \right) \right\}, & k even \end{cases} \quad (8)$$

이와 같은 대칭구조를 갖는 미분에너지함수를 이용한 순간주파수 추정기는 다음과 같이 정의된다[7].

$$f(n) = \frac{1}{2\pi} \frac{1}{(k-1)} \cos^{-1} \left(\frac{\Xi_{2k-1} \{x(n)\}}{2 \cdot \Xi_k \{x(n)\}} \right) \quad (9)$$

순간주파수 추정기의 주파수 추정성능은 다음과 같이 얻어진 다. 시험신호로는 다음과 같은 AM-FM신호가 사용되었다.

$$x(n) = \left(1 + \kappa \cos \left(\frac{\pi n}{100} \right) \right) \cdot \cos \left(\frac{\pi}{5} n + 20\lambda \sin \left(\frac{\pi}{100} n \right) \right), 0 \leq n \leq 399,$$

$(\kappa, \lambda) \in \{(0.05i, 0.05j) | i, j = 1, 2, 3, \dots, 10\}$. 즉, κ 는 AM 변조부분을 조절하며, λ 는 FM 변조부분을 조절한다. κ, λ 는 i, j 값에 따라 5%에서 50%까지 5%간격으로 증가한다. 이 신호의 순간주파수는 $f(n) = f_s \cdot \left(\frac{1}{10} + \frac{1}{10} \lambda \cdot \cos \left(\frac{\pi}{100} n \right) \right)$ 로 주어진다. 여기에서 f_s 는 표본화 주파수이다. 순간주파수의 추정오차는 다음과 같이 계산되었다.

$$\begin{aligned} \text{추정오차 (\%)} &= \\ &= \frac{1}{40000} \sum_{n=0}^{399} \sum_{i=1}^{10} \sum_{j=1}^{10} (f_{ij}(n) - \hat{f}_{ij}(n)) / f_{ij}(n). \end{aligned}$$

여기에서 $f_{ij}(n)$ 은 특정 (i, j)에서의 순간주파수의 참값이고 $\hat{f}_{ij}(n)$ 은 순간주파수 추정결과이다. 아래 표 1은 순간 주파수

표 1 순간주파수 추정기에 따른 주파수 추정성능
Table 1 Estimation performance of various instantaneous frequency estimators

순간주파수 추정방법	잡음이 없을때	SNR=30dB
고차의 순간주파수, k=2	2.92	39.14
고차의 순간주파수, k=3	0.78	4.86
고차의 순간주파수, k=4	1.78	3.21
대칭구조순간주파수, k=2	0.14	38.08
대칭구조순간주파수, k=3	0.27	3.78
대칭구조순간주파수, k=4	0.70	2.59

추정기에 따른 추정성능을 보인다. 일부 결과는 [7]에서 인용하였으며 대칭구조의 순간주파수 추정기의 k=2에서의 결과는 본 논문에서 추가되었다. 표의 결과에서 알 수 있듯이 전반적으로 대칭구조의 순간주파수추정기가 고차의 순간주파수보다 우월한 주파수 추정성능을 보인다. 표 1의 추정성능을 고려해서 본 논문의 음성의 기본주파수 추정에는 차수 k=2의 대칭구조의 순간주파수 추정기를 사용한다.

3. 기본주파수 추정기의 구조

3.1 기본주파수 추정기의 구조

음성의 기본주파수는 여성의 경우 대략 500 Hz이하, 남성의 경우 300 Hz 이하의 주파수 영역에 존재하므로 이러한 기본주파수를 순간주파수 추정기로 구하려면 음성신호의 대역을 제한하여야 한다. 따라서 음성을 저역통과필터로 처리한다. 처리된 음성은 기본주파수 성분을 주요한 성분으로 갖는 신호로 볼 수 있다. 대역 제한된 음성신호를 2장에서 소개된 대칭구조를 갖는 고차의 미분에너지 함수를 이용한 순간주파수 추정기로 처리하여 기본주파수의 추정 값을 구한다. 얻어진 순간주파수는 음성의 구간에 따라 상이한 특성을 보인다. 목음 혹은 무성음 구간에서는 급격히 변화하며 일정한 주파수 패턴을 보이지 않으며, 유성음 구간에서는 500 Hz 이내의 범위에서 변화하는 형태를 보인다. 따라서 기본주파수가 존재하는 유성음과 존재하지 않는 무성음 혹은 목음 구간을 구별할 필요가 있다. Koh 등은 유성음과 무성음 구간을 Heuristic한 방법으로 결정하는 규칙을 제안하였다[4]. 즉 순간주파수가 일정시간동안 일정편차 이내에서 유지되는 경우에만 기본주파수로 가정하였다. 이와 같은 결정규칙은 화자에 따른 변화에 적절하게 대응치 못하는 문제점이 있다. 본 논문에서는 순간주파수를 구하는 과정에서 사용되는 고차의 미분에너지 함수 값을 관찰하여 유성음과 무성음을 구별하는 인자로 사용한다. 이렇게 얻어진 유성음 구간에서의 순간주파수 값을 이동평균 필터로 처리하여 기본주파수의 추정값을 구한다. 그림 1은 이와 같은 기본주파수 추정기의 구조를 보인다.

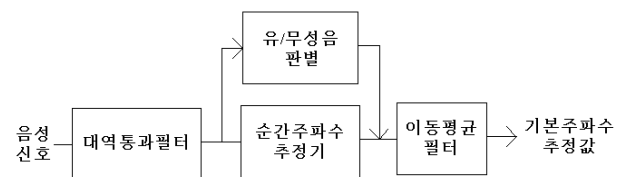


그림 1 제안된 기본주파수 추정기의 구조
Fig. 1 Structure of the proposed fundamental frequency estimator

3.2 유·무성음 판별기로서의 에너지함수

앞 절에서 지적된 것처럼 제안된 기본주파수 추정기 가운데 순간주파수는 음성신호의 구간별로 상이한 특성을 보인다. 즉 목음 및 무성음 구간에서는 난잡음(random noise)과 같은 특성을 보이며 유성음 구간에서는 비교적 단조롭게 변화하는 파형을 보인다. 아래 그림 2는 저역통과된 음성신호의 순간주파수를 보인다. 음성신호는 남성의 목소리가 사

용되었다. 대역통과필터로는 차단주파수 250 Hz의 저역통과 필터가 사용되었고, 순간주파수로 식 (9)의 대칭구조를 갖는 순간주파수가 사용되었다. 그림 2(b)에서 보이듯이 묵음 혹은 무성음 구간에서는 순간주파수가 복잡하게 변하는 임의의 패턴을 보인다. 기본주파수는 유성음 구간에서만 의미가 있으므로 유성음과 무성음 구간을 구분하여 유성음 구간만을 취하는 일종의 코스트함수가 필요하다. 별도의 함수를 사용하는 대신 순간주파수 추정에 사용된 미분에너지함수 (8)을 사용하면 계산량을 절감할 수 있다. 즉, 식(9)에서 $k=2$ 인 순간주파수를 본 논문에서 사용하므로 2차와 3차의 대칭구조를 갖는 고차의 미분에너지 함수를 얻으며, 이 가운데 하나를 유무성음 판별을 위한 코스트함수로 사용할 수 있다. 3차의 에너지함수가 2차의 에너지함수보다 유성음구간과 무성음 구간의 구별에 더 적합하므로 3차의 에너지 함수를 유무성음 구간을 구별하기 위하여 사용한다. 아래 그림 3은 대역 제한된 음성신호를 3차의 미분에너지함수로 처리한 결과이며, 미분에너지함수가 유성음구간과 무성음구간, 묵음구간에서 진폭에 있어서 확인한 차이를 보임을 알 수 있다. 음성 구간에 따른 이러한 미분에너지함수의 차이는 다음과 같이 해석될 수 있다.

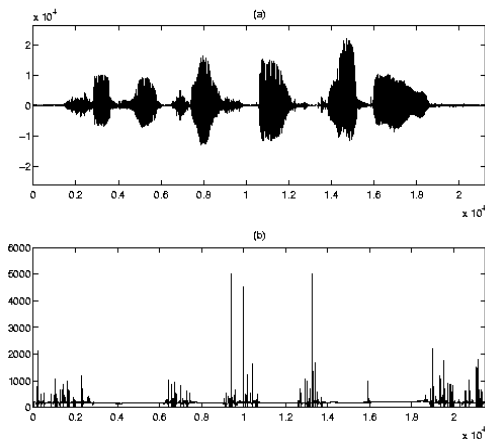


그림 2 저역필터로 처리된 음성의 순간주파수 (a) 원음성신호 (b) 저역 통과 필터로 처리된 신호의 순간주파수
Fig. 2 Instantaneous frequency of a band pass filtered signal

$k=3$ 으로 홀수일 때, 식(8)의 대칭구조의 고차미분에너지 함수는 $\Xi_3\{x(n)\} = [\Gamma_3\{x(n)\} + \Gamma_3\{x(n-1)\}]/2$ 로 식(3)의 고차미분에너지함수의 산술평균이며, 식(3)의 고차미분 에너지 함수 $\Gamma_3\{x(n)\} = x(n)x(n+1) - x(n-1)x(n+2)$ 는 순간상관함수 $\phi_{xx}(n, k) = x(n)x(n+k)$ 의 차 $\Gamma_3x(n) = \phi_{xx}(n, 1) - \phi_{xx}(n-1, 3)$ 로 볼 수 있다. 따라서 식(3)의 고차 미분에너지함수의 구성성분인 순간상관함수의 유무성음 구간에서의 특성을 살펴본다. 묵음이나 무성음 구간과 같이 신호가 랜덤 신호에 가까운 특성을 보여서 상관관계가 작은 경우, 이러한 순간상관함수는 시간차 k 값에 따른 변화가 거의 없다. 반면에 상관관계가 큰 유성음 구간의 경우, 시간차 k 에 따라 상관관계가 큰 변화를 보인다. 아래 그림 4는 순간상관함수를 시간차 k 에 대하여 그린 것이다. 그림 4(a)는

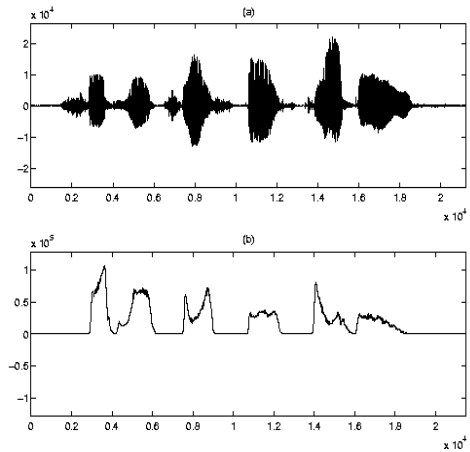


그림 3 저역필터로 처리된 음성신호의 고차 미분에너지 함수 값 (a) 원음성신호, (b) 고차 미분에너지 함수값
Fig. 3 Higher order differential energy function value of a lowpass filtered speech signal (a) speech signal (b) higher order energy function

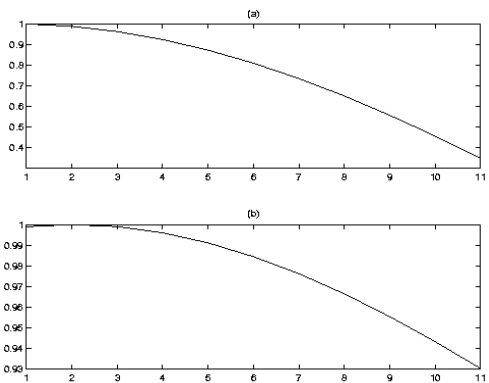


그림 4 저역필터로 처리된 음성신호의 고차 미분에너지 함수 값 (a) 원음성신호, (b) 고차 미분에너지 함수값
Fig. 4 Higher order differential energy function value of a lowpass filtered speech signal (a) speech signal (b) higher order energy function

유성음 구간의 순간상관함수이며, (b)는 무성음 구간의 순간상관함수이다. 두 경우 모두 순간상관함수는 최대값 1로 정규화된 결과이다. 유성음 구간에서 순간상관함수가 시간차 k 와 함께 급격하게 감소함을 알 수 있다. 따라서, 순간상관관계의 차로 정의되는 식(3)의 고차미분에너지함수 그리고 그 평균인 (8)의 대칭구조의 고차미분에너지함수는 유성음 구간에서 큰 값을 보이고 무성음 구간에서는 작은 값을 보이며 그림 3은 이러한 특성을 보여준다.

4. 컴퓨터 모의 실험

이절에서는 본 논문에서 제안된 기본주파수 추정기의 유용성을 컴퓨터 모의실험을 통해 확인한다. 추정된 기본주파수는 대표적인 시간주파수표현법인 스펙트로그램 결과와 비교하여 정확성을 검증하였다. 즉, 저역 통과 필터로 전처리

된 음성신호의 스펙트로그램을 구하고, 이를 기본주파수 추정치와 비교하여 제안된 기본주파수 추정기가 음성의 기본주파수를 적절하게 추정하는지 검증하였다. 순간주파수 추정기는 식(8)의 대칭구조를 갖는 고차의 미분에너지 함수를 이용한 식(9)의 정의가 사용된다. 순간주파수 추정기에서 차수는 2로 설정되었다. 유·무성음을 판별하기 위한 고차의 미분에너지 함수의 차수는 3으로 설정되었으며, 이는 식(9)의 순간주파수 추정기에서 분자에 해당한다. 유성음, 무성음을 판별하기 위한 고차미분에너지함수 값의 임계값은 800으로 설정되었다. 최종 기본주파수 추정치를 얻기 위하여 길이가 200인 이동평균필터가 사용되었다. 사용된 음성은 무성음과 유성음이 혼재되어 있는 문장과 유성음만으로 이루어진 문장을 사용하였다. 음성의 표본화주파수는 10kHz이다. 첫 번째 문장은 영문으로 “Should we chase those cowboys ?” 이며 두 번째 문장은 “We are away a year ago”라는 문장이다 [8].

그림 5는 첫 번째 음성의 파형, 저역 통과된 음성의 파형 그리고 추정된 기본주파수, 스펙트로그램과의 비교를 보인다. 대략 200 Hz 내외에서 시간과 함께 변하는 특성을 보인다. 음성의 뒷부분에서는 의문문의 특성상 주파수가 상승하면서 끝맺는 특성을 보이고 있다. 그림 5(d)는 얻어진 기본주파수를 대역 통과된 신호의 스펙트로그램에 중첩해서 그린 그림이다. 스펙트로그램은 시간주파수 평면 위에 같은 에너지를 갖는 위치를 연결한 등위선(contour)으로 표현했다. 등위선은 최대값의 3%에서 99%까지를 3%간격으로 취했다. 즉 폐곡선의 안쪽이 에너지가 높은 시간주파수 성분을 나타낸다. 스펙트로그램을 통해 얻어진 기본주파수가 원음성의 기본주파수와 근사함을 볼 수 있으며 따라서 제안된 방법이 적절하게 기본주파수를 추정함을 알 수 있다.

그림 6은 두 번째 문장 “We were away a year ago”의 기본주파수 추정결과를 보여준다. 그림 6(d)에서 보이듯 제안된 기본주파수 추정기가 적절하게 신호의 기본주파수를 추정함을 알 수 있다. 그림 5와 6의 결과에서 알 수 있듯이 순간주파수 추정기를 음성의 기본주파수 추정에 사용함으로써, 기존의 방법에서 필요한 국부적으로 시불변이라는 가정이 더 이상 필요치 않으며, 짧은 시간동안 변하는 음성의 특성을 적절하게 분석할 수 있다. 따라서 음성분석, 인식, 부호화 등의 응용에서 우수한 결과를 얻을 것으로 예측된다.

5. 결 론

본 논문에서는 대칭구조를 갖는 고차의 미분 에너지함수에 기반한 순간주파수 추정기를 이용하여 음성의 기본주파수를 추정했다. 제안된 기본주파수 추정기는 전처리기로 저역통과필터, 순간주파수 추정기, 유·무성음 판별기로서의 고차에너지함수, 그리고 이동평균필터로 구성된다. 저역통과필터는 음성의 기본주파수 성분을 주로 갖는 신호를 얻기 위하여 전 처리기로서 사용된다. 제안된 기본주파수 추정기는 음성의 기본주파수를 시간주파수 분석법의 일종인 순간주파수를 이용하여 추정하기 때문에, 기존의 기본주파수 추정기에서 필요로 했던 국부적인 시불변이라는 가정이 불필요하다. 순간주파수 추정기의 한부분인 고차의 미분에너지함수가 유성음/무성음/목음을 구분하는 역할을 수행해서 계

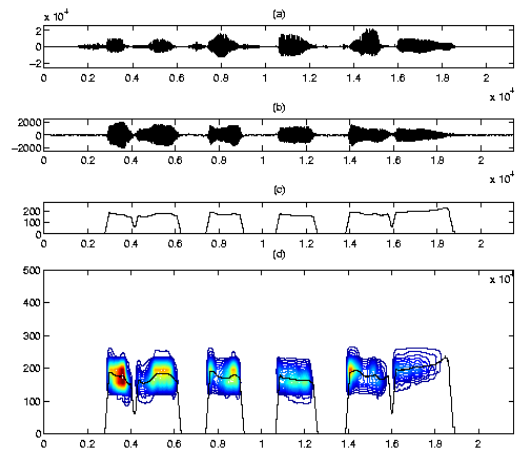


그림 5 기본주파수추정결과 (a) 원음성신호, (b) 저역통과신호 (c) 기본주파수 추정값 (d) 스펙트로그램과 기본주파수 추정값

Fig. 5 Fundamental frequency estimation result (a) speech signal (b) preprocessed speech signal (c) fundamental frequency estimates (d) fundamental frequency superimposed onto the Spectrogram

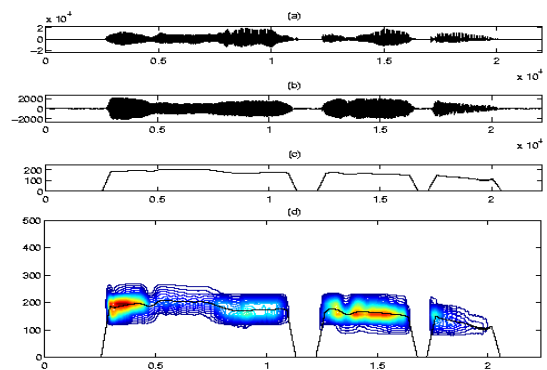


그림 6 기본주파수추정결과 (a) 원음성신호, (b) 저역통과신호 (c) 기본주파수 추정값 (d) 스펙트로그램과 기본주파수 추정값

Fig. 6 Fundamental frequency estimation result (a) speech signal (b) preprocessed speech signal (c) fundamental frequency estimates (d) fundamental frequency superimposed onto the Spectrogram

산량을 절감했다. 순간주파수는 이동평균필터를 이용하여 단순변화하는 기본주파수의 추정치를 얻는다. 컴퓨터 모의 실험을 통해 얻어진 음성의 기본주파수가 음성의 스펙트로그램에서 보이는 기본주파수와 일치함을 보였다. 이와 같이 얻어진 기본주파수는 음성의 시변특성을 반영한 것으로 좀 더 세밀한 음성인식, 화자인식 등에 활용될 것으로 기대된다.

감사의 글

이 논문은 2010학년도 강릉대학교 장기해외파견 연구지원에 의하여 수행 되었음.

참 고 문 헌

- [1] L.R. Rabiner and R. W. Schafer, Digital Processing of Speech Signals, Englewood Cliffs, NJ: Prentice Hall, 1978.
- [2] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," IEEE Signal Processing Magazine, vol. 19, pp. 21-67, Apr. 1992.
- [3] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal-Part 2: algorithms and applications," Proc. of the IEEE, vol. 80, pp. 540-568, Apr. 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] P. Maragos, J. F. Kaiser, and T. F. Quatieri, "On amplitude and frequency demodulation using energy operators," IEEE Trans. Signal Processing, vol. 41, pp. 1532-1550, Apr. 1993.
- [6] P. Maragos, and A. Potamianos, "Higher order differential energy operators," IEEE Signal Processing Letters, vol. 2, pp. 152-154, Aug. 1995.
- [7] B. Iem, "An Instantaneous frequency estimators based on the symmetric higher order differential energy operator," IEICE Trans. Fundamentals, vol.E93-A, no. 1, pp. 227-232, Jan. 2010.
- [8] D. G. Childers, Speech Processing and Synthesis Toolboxes, Wiley, September. 1999.

저 자 소 개



임 병 관 (林炳寬)

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