

윈도우의 영향이 제거된 에너지 파라미터에 관한 연구

조태수, 신동성, 배명진
숭실대학교 정보통신공학과
Tel. 820-0016 / Fax. 820-0018

A Study of Energy Parameter without Windowing Influence in Speech Signal

Taesu Cho, Dongsung Shin, Myungjin Bae
Dept. of Telecommunication, Soongsil University
E-mail : jungbotongsin@weppy.com

Abstract

The preprocessing is very important course in speech signal processing. It influence the compression-rate in speech coding and the recognition-rate in speech recognition etc.

In this paper, we propose that minimizing window-influence method with pitch period and start points. The proposed method is available for voiced detection and word labeling.

1. 서론

음성신호처리에 있어서 전처리 과정은 매우 중요한 역할을 담당하고 있다. 어떻게 전처리를 하느냐에 따라서 그 결과에 많은 영향을 미칠 수 있다. 특히 프레임의 에너지 검출은 음성구간 검출에서 유성음 구간 검출에 중요한 역할을 담당한다[1][2][3].

또한 음성신호의 에너지 정보는 어떠한 발성에 대한 음운학적 특성을 나타낼 수도 있다. 예를 들면 마찰음은 유성음보다 에너지가 작다는 특성을 갖고 있다. 따라서 에너지 정보는 워드를 구별하는데 유용하게 사용될 수 있다[4]. 하지만 기존의 방법에서는 윈도우의 크기를 고정시켜 사용함으로써 정확한 에너지 값을 구하지 못하였다.

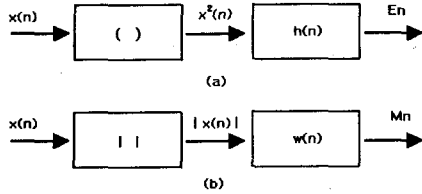
본 논문에서는 피치주기를 먼저 구하고 피치주기에 따라서 에너지 값을 추출하는 방법을 제안하였다.

피치 주기를 구하는 방법은 시간영역에서 구하는 방법과 주파수 영역에서 구하는 방법, 그리고 시간-주파수 영역에서 구하는 방법이 있다. 시간영역 피치검출법은 시간영역에서 직접 처리하기 때문에 다른 영역으로의 변환과정이 불필요하며 합, 차, 비교논리 등에 의해서만 처리가 가능해진다. 피치의 범위는 보통 2.5에서 25msec로 알려져 있고 음성을 8kHz로 표본화하여도 그 범위가 20-200 표본 사이에 나타나기 때문에 시간영역에서의 검출은 분해능이 높다는 장점이 있다. 따라서, 본 논문에서는 각 프레임의 에너지를 구하는데 있어서 피치의 시작점과 주기를 이용해서 윈도우의 크기를 결정함으로써 윈도우의 영향을 최소화하는 에너지 계산방법을 제안하였다.

2. 단구간 에너지 검출

에너지와 평균진폭의 측정방법들은 음성신호를 더 작은 음서단위 즉 프레임으로 나누어 처리한다. 유성음과 무성음 사이에서 진폭은 크게 변화하고, 조음기관의 다른 방법에 의한 음소들 사이의 진폭은 크게 변화하고, 조음기관의 다른 방법에 의한 음소들 사이의 진폭은 작게 변화한다는 특징을 가지고 있으며, 이러한 특징을 이용하여 자동인식시스템에서는 에너지 $Q(n)$ 를 특정하여 음성의 끝점을 검출한다[3].

그림 2-1은 단시간 에너지와 평균 진폭의 블록도를 나타내었다.



(a) 단시간 에너지의 블록도
(b) 단시간 평균진폭의 블록도

그림 2-1. 단시간 에너지와 평균진폭 블록도
신호의 이산시간 에너지는 식(2.1)로 정의되고,

$$E = \sum_{m=-\infty}^{\infty} x^2(m) \quad \text{식(2.1)}$$

단시간 에너지는 식(2.2)와 같이 정의된다.

$$E_n = \sum_{m=n-N+1}^n x^2(m) \quad \text{식(2.2)}$$

즉 표본 n에서 단시간 에너지는 n-N+1에서 n까지 N개의 표본(sample)의 제곱의 합으로 정의된다.

그림 2-2는 단시간 에너지의 처리방법을 나타내고 있다.

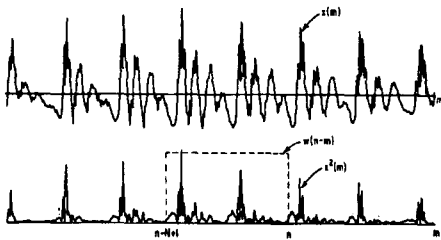


그림 2-2. 단구간 에너지검출

식(2.2)에 대해서 창함수를 사용하여 나타내면, 식(2.3)과 같이 정의할 수 있고,

$$E_n = \sum_{m=-\infty}^{\infty} [x(m)w(n-m)]^2 \quad \text{식(2.3)}$$

임펄스응답으로 표현하면 식(2.4)와 같이 정의할 수 있다.

$$E_n = \sum_{m=-\infty}^{\infty} x^2(m) \cdot h(n-m) \quad \text{식(2.4)}$$

$$h(n) = w^2(n)$$

임펄스 응답 또는 창함수의 선택에 따라 단시간 에너지 표현에 영향을 준다[2].

3. 시간영역에서 피치 검출법

시간영역에서의 피치검출법들은 일반적으로 음성 파형의 주기성을 강조하고 나서 결정논리에 의해 주기성을 검출하게 된다. 주기성 강조는 성도의 공명현상에 나타나는 포먼트들의 영향을 제거시키고 여기원의 피치만을 강조하는데 그 목적이 있다. 따라서 주기성 강조가 복잡하면 할수록 결정논리가 간단해진다[5].

그림 3-1은 전형적인 시간영역 피치검출법에 대한 블록도이다.

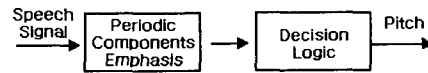


그림 3-1. 시간영역 피치검출법

피치검출의 결정논리법에는 강조된 음성파형의 기본 주기를 실험적인 문턱값이나 무게치를 적용하여 결정하는 것이 보통이다. 그 외에도 위상이나 진폭에 따라 몇 가지의 기준패턴을 만들어 입력된 음성의 파형에서 추출한 패턴과의 유사도를 측정하여 주기성을 결정하기도 한다. 결정논리가 실험적인 문턱값에 의존하면 검출오차가 확률적으로 커지기 때문에 경우의 수에 따라서 결정 문턱값을 신중히 고찰해야 하는 복잡성이 따른다. 따라서 최근에는 Fuzzy나 신경망의 결정논리를 적용한 사례가 계속해서 연구 중에 있다. 먼저 시간영역 피치검출법으로는 Parallel Processing, Autocorrelation, AMDF, Data Reduction Harmonic Matching법 등이 있다[6].

3-1. 중앙 클리핑 자기상관관계에 의한 피치검출법

자기상관함수를 통과하기 전에 식(2.1)과 같이 중앙클리핑을 수행한다.

$$y(n) = x(n) - C_L \quad \text{if } s(n) \geq C_L$$

$$y(n) = x(n) + C_L \quad \text{if } s(n) \leq -C_L \quad \text{식(3.1)}$$

$$y(n) = 0 \quad \text{otherwise}$$

클리핑 레벨은 일반적으로 식(3.2)와 같이 한다.

$$C_L = 0.3A_{\max} \quad \text{식(3.2)}$$

여기서 A_{\max} 는 음성 세그먼트의 절대값의 최대값이다. 중앙클리핑을 수행하는 이유는 성도정보를 제거하여 정확하게 피치를 검출하기 위해서 사용된다[7].

그림3-2은 중앙클리핑 자기상관관계에 의한 피치검출법의 블록도이다[2].

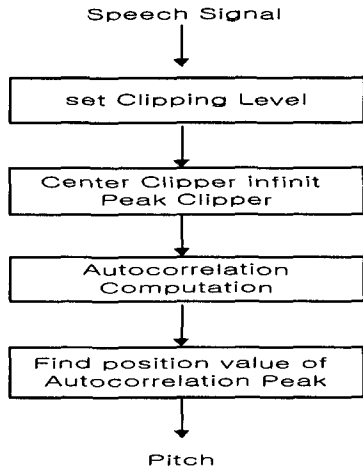


그림 3-2 중앙클리핑 자기상관관계법

3-2. AMDF를 이용한 피치 검출법

AMDF방법은 자기상관함수법에 수행하는 곱 연산을 절대값과 차분으로 대신하기 때문에 상대적으로 빠르다는 장점을 가지고 있다. 이러한 이유로 실시간에 많이 적용한다. 자기상관관계법에서는 피치주기 배수에 최대값을 이루지만, AMDF법에서는 피치주기 배수에 최소값을 갖는다. AMDF는 식(3.3)로 정의된다[1][2].

$$AMDF(k) = \sum_{m=-\infty}^{\infty} |x(m) - x(m-k)| \quad \text{식(3.3)}$$

그림 3-3은 AMDF 방법을 도식화하여 나타낸 것이다.

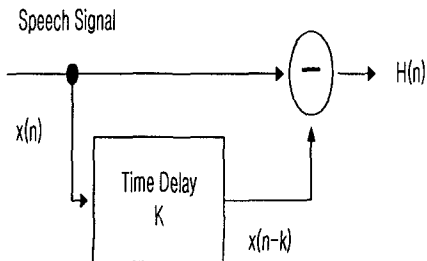


그림 3-3. AMDF의 블록도

4. 제안한 알고리즘

기존의 방법은 윈도우 사이즈를 고정 시켜 에너지를 구함으로써 정확한 프레임 에너지를 구할 수 없었다. 따라서 본 논문에서는 피치주기를 먼저 구하고 그 주기에 따라 에너지를 구하는 방법을 제안하였다.

제안한 알고리즘의 식은 식(4.1)과 같다.

$$E_n = \frac{\sum_{i=0}^{Pitch\ period} E(i)}{Pitch\ period} \quad \text{식(4.1)}$$

제안한 알고리즘의 블록도는 그림 4-1에서 보여 주고 있다.

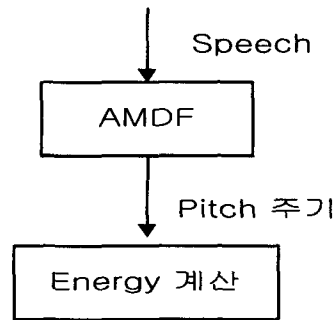


그림 4-1. 제안한 알고리즘

5. 실험 및 결과

컴퓨터 시뮬레이션에 이용한 장비는 IBM-PC 586(500MHz)에 상용화된 AD/DA 컨버터를 인터페이스한 시스템이다.

그림 5-1은 실제 실험한 결과 파형을 나타내고 있다. 실제 음성과 가변윈도우의 해상도를 맞추어 그린 에너지 그래프이다.

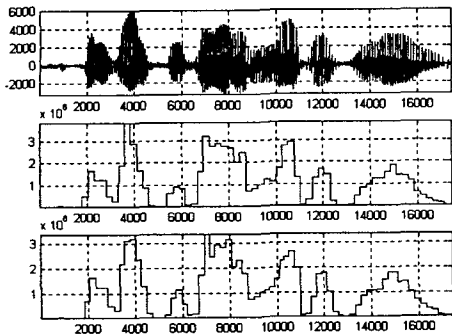


그림 5-1. 실험한 결과 파형

실험에 사용한 발성음은 다음과 같다.

발성 : /깊은 산속에 엄마 돼지와 아기돼지 삼 형제가 살고 있었어요/

6. 결론

음성신호처리에서 전처리 과정은 매우 중요하다. 전처리를 어떻게 하느냐에 따라서 코딩에 있어서는 압축률이 달라질 수 있고, 음성인식에 있어서는 인식률에 큰 영향을 미치게 된다.

본 논문에서는 각 프레임의 에너지를 구하는데 있어서 피치의 시작점과 주기를 이용해서 윈도우의 크기를 결정함으로써 윈도우의 영향을 최소화하는 에너지 계산방법을 제안하였다.

실제 음성파형에 기존의 고정 윈도우 방법과 윈도우의 영향을 제거한 방법을 실험한 결과 윈도우의 영향을 제거한 방법이 묵음과 유성음을 확실하게 구별할 수 있게 해주어, word labeling을 더 효율적으로 할 수 있었다.

7. 참고 문헌

- [1] L. R. Rabiner and R. W. Schafer, "Digital Processing of Speech signals", Englewood Cliffs, Prentice-Hall, New Jersey, 1978.
- [2] D.O'Shaughnessy "Speech Communications", IEEE PRESS, 2000
- [3] 이상효, 배명진, "디지털 음성분석", 1998
- [4] L. R. Rabiner & Biing-Hwang Juang,

"Fundamentals Of Speech Recognition", Prentice-Hall, AT&T, U.S.A, 1993

[5] M.J. BAE, K.C. HAN and J.J. CHA, "On a Pitch Extraction of Speech Signal using Residual Signal of the Uniform Quantizer," J., Acoust., Society, Korea, Vol.16, No.2, pp.36-40, February 1997.

[6] W.R. JO, M.J. BAE and D.S. KIM, "On a Pitch Alteration Technique in the V/UV Spectrum for High Quality Speech Synthesis Technique," J., Acoust., Society, Korea, Vol.15, No.6, pp.99-103, December 1996.

[7] 함명규, 최성영, 박종철, 배명진, "자기상관함 수에서 위상 성분의 보존에 의한 피치시점 검출에 관한 연구", 대한전자공학회, 제 13회 신 호처리합동학술대회 논문집, Vol.13. No.1, pp.799-802,