EDGE ENHANCEMENT METHOD를 이용한 음성구간 추출

°임 준 석, 성 굉 모
서울대학교 공과대학 전자공학과

Speech Interval Extraction by Edge Enhancement Method

Jun Suk Lim, Koang Mo Sung
Dept. of Electronics Eng., Seoul Nat'l Univ.

요 약

본 실험에서는 nonlinear product average 법을 사용하여 에너지만을 parameter로 써서 음성구간을 순차적으로 검출할수 있음을 보였다.

Abstarct

In this experiment, we have the result that the speech interval can be extracted with a energy parameter by noise suppression through nonlinear product average in edge enhancement method.

I. 서론

음성 구간 검출의 문제는 DSI(digital speech interpolation) 이나 TASI(time assigned speech interpleation), voice mail 시스템 분야뿐아니라 음성인식 분야에서도 관심을 갖고 있는 문제이다.

DSI나 TASI를 위한 초기의 음성 구간 검출 방법에서는 신호 level이나 포락선 검출을 이용하여 일정한 임계치를 초과 하면 음성이라 판별하고 임계치 이하이면 무음(silence,noise,pause)이라 결정하였다.

또한 음성인식에서의 음성 검출기로서는 Rabiner 와 Sambur가 영교차율과 에너지를 이용한 음성구간 검출법이 많은 사람들에게 알려져 있다[1]. 그외에도 Atal과 Rabiner는 LPC방법을 이용하여 영교차율,short time energy,인접 음성표본치의 상관관계, LPC 분석의 1차 예측계수,예측오차의 에너지 등 5가지를 parameter로 하는 음성 검출 방법을 제안하였다[2].

이런 방법들은 상당히 정확한 음성구간 검출을 할 수 있으나,그 수행과정이 복잡하다는 단점이 있다.

Rabiner와 Sambur의 방법에서는 data의 양 끝점을 미리 알아야 하는 불편이 있으나, 본 실험에서는 음성의 에너지만을 고려해서 data를 순차적으로 처리하여 음성구간의 검출을 시도하였다.

II. 이론적 배경

본 음성 검출 실험에서의 기본적인 개념은 보통의 무음 구간(silence or noise interval)에서는 에너지의 시간에 따른 변동이 심하지 않은 것을 이용해서 음성 구간을 좁다

두드러지게 하여 무음구간과 음성구간의 차이를 크게 만드는 데 있다.

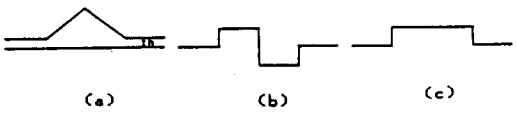


그림1. edge enhancement 기본 개념

예로 그림1 (a)와 같은 모양의 변화를 갖는 에너지 신호를 생각하자. 여기에 미분 계남을 넣은 operater를 적용하면 그림1 (b) 와 같이 될것이고 여기에 절대치를 취하면 그림1(c) 와 같이 되어 높이 h 이상 부분을 쉽게 알아낼 수 있다.

이와 같은 일을 구현하기 위하여 다음과 같이 화상 신호처리에서 edge enhancement의 한 방법으로 사용하는 nonlinear product average 법을 이용했다[3].

$$Dm(j)=(f(j+m-1)+f(j+m-2)+\ldots+f(j) -f(j-1)-f(j-2)..-f(j-m))/m \quad -(1)$$

$$Pm(j)=D1(j)*D2(j)*D3(j)*\ldots*Dm(j) \quad -(2)$$

여기서 m=2**n n:integer

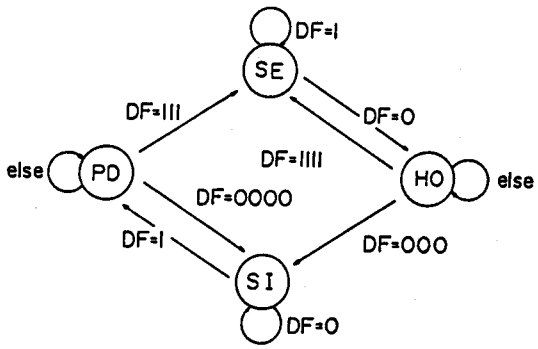
위 product를 각 frame 마다 실시했을때 order n을 높힐수록 더 많은 잡음 억압 효과를 얻을 수 있고 따라서 잡음에 대해 강해질 수 있다. 반면에 낮은 order의 n을 사용할 경우 잡음 억압 효과가 떨어지서 잡음에 대해 민감해진다.

위 식으로써 음성 구간 추출을 할 때는, 우선 처음 5 frame이 무음 이라 가정하고 그 무음 구간에서 얻어진 다음 식과 같은 임계치와 (1)식을 각 frame 마다 실시한 결과를 비교해서

$$ETH=0.5*0.2*(\sum_{n=1}^{5}E(n)) \quad -(3)$$

E(n):energy of n-th frame

임계치보다 낮아서 무음이라 판정되면 flag DF=0으로 설정하고, 음성이면 flag DF=1로 설정한다. 그 다음에는 다음과 같은 state-diagram에 따라 음성 구간을 추출하게 된다.



여기서 SE:speech existence
HO:hangover
PD:primary detection
SI:Silence

그림2. state diagram [4].

또 무음 구간 동안에 click음이나 발음끝에 생기는 펄스형태의 잡음이 있는 경우에 오류를 방지하려면 음성으로 추출된 구간의 시간간격 또한 생각해 주어야 하는데, 그 기준이되는 시간간격은 160msec 로 잡았다. 다시 말해서 추출된 음성의 시간이 160msec이하 일 때 그 구간은 음성 구간으로 판정하지 않는다. 이로써 한 음성 data에서 음성 아닌 것이 음성으로 추출되는 오류를 방지하였다.

III. 실험결과

다음에 보인 실험결과는 두 명의 성인 남자에 의해서 발음된 숫자음 "일","삼","칠","구"에 대해서 본 음성구간 추출법을 적용한 결과 이다. 이 실험에서 한 frame의 길이는 8msec로 하였다.

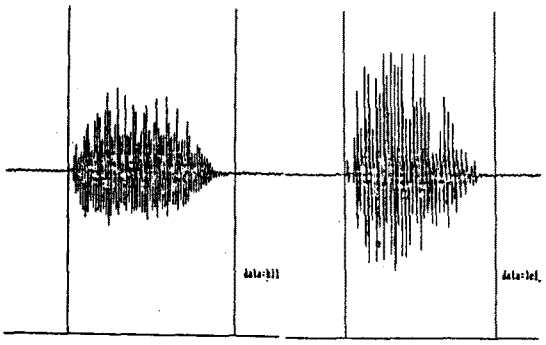


그림3. 두명의 화자에의해 발음된 "일"의 시간축 파형과 검출된 음성 구간

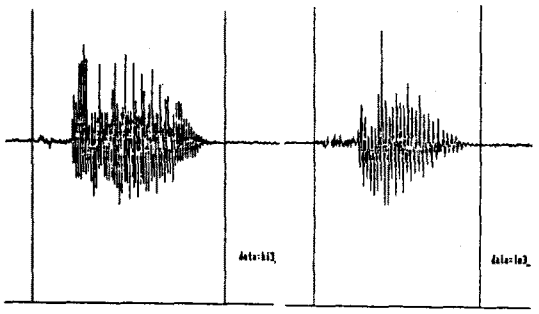


그림4. 두명의 화자에의해 발음된 "삼"의 시간축 파형과 검출된 음성 구간

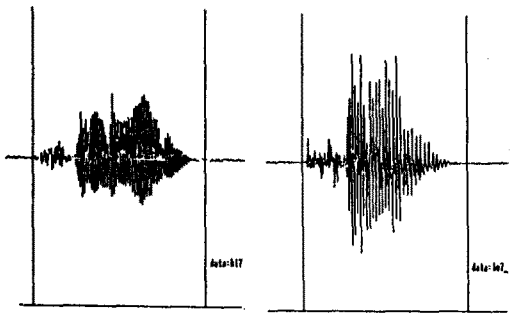


그림5. 두명의 화자에의해 발음된 "칠"의 시간축 파형과 검출된 음성 구간

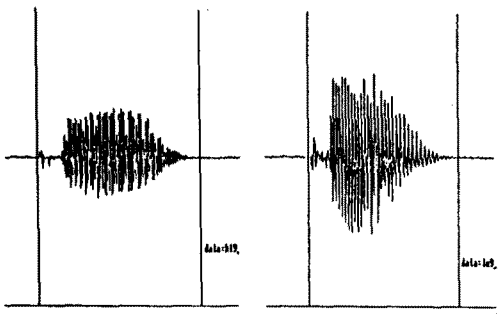


그림6. 두명의 화자에의해 발음된 "구"의 시간축 파형과 검출된 음성 구간

IV. 결론

본 실험에서는 화상 신호처리 기법중 edge detection에서 사용하는 방법을 이용하여 무음과 음성구간의 경계를 에너지만을 parameter로 써서 순차적인 방법으로 쉽게 검출할 수 있음을 보였다.

V. 참고 문헌

[1] L.R.Rabiner and M.R.Sambur,"An algorithm for determining the endpoints of isolated utterances," B.S.T.J, pp 297-315,Feb 1975

[2] B.S.Atal and L.R.Rabiner," A pattern recognition approach to voice-unvoice-silence classification with applications to speech recognition," IEEE Trans on ASSP,Vol.ASSP-24, No.3,June 1976

[3] W.K.Pratt,Digital Image Processing,John Wiely & Sons, New York, 1978

[4] 강철희 et al.,"한국어 음성인식 합성 시스템 개발에 관한 연구," ETRI 최종보고서,July 1986