끝점 검출 알고리즘에 관한 연구

○양 진 우*, 이 형 준**, 김 순 협*
* 광운대학 전자계산기 공학과, **삼성 연구소

A Study on the Endpoint Detection Algorithm

○ Jin Woo Yang*, Hyoung Jun Lee**, Soon Hyob Kim*
* Dept. of Computer Eng. Kwang Woon University, ** Sam Sung Research Lab.

(Abstract)

This paper is a study on the Endpoint Detection for Korean Speech Recognition.
In speech signal process, analysis parameter was classification from Zero Crossing Rate(Z.C.R), Log Energy(L.E), Energy in the predictive error (E_p) and fundamental Korean Speech digits,/영/-/구/ are selected as data for the Recognition of Speech.
The main goal of this paper is to develop techniques and system for Speech input to machine.
In order to detect the Endpoint, this paper makes choice of Log Energy(L.E) from various parameters analysis, and the Log Energy is very effective parameter in classifying speech and nonspeech segments.
The error rate of 1.43% result from the analysis.

1. 서론

음성 인식은 인간과 기계의 대화를 위해서 화자의 어휘수와 상관없이 모든 음성을 인식할 수 있는 시스템의 개발에 중점을 두고 있다.

음성 신호에 대한 정확한 끝점 검출은 주위 환경의 잡음으로 부터 정확한 음성 신호만을 추출해 내는 일이다.

끝점이 정확하게 검출되면 음성 신호의 데이터 처리량을 최소로 줄일수 있고 정확한 처리를 할수있다.

본 연구에서는 한국어 단독 숫자음의 중성 특징이 유성자음과 모음이라는 점을 고려해서 에너지 패턴의 기준값만 정확하게 잡으면 끝점은 비교적 간단하면서 정확하게 검출할 수 있다.

2. 한국어 음성 및 시간 영역에서의 음성 신호 분석

(1) 한국어 음성의 발성상 특징

1) 단독 숫자음의 초성은 무성자음 ㅅ(s), ㅊ(ㅎ), ㅍ(p), ㄱ(k) 이다.

2) 단독 숫자음의 종성은 유음 ㄹ(l) 과 비음 ㅁ(m), ㅇ(ŋ) 과 같은 유성자음으로 끝난다.

3) 단독 숫자음 육 (juk) 의 종성 ㄱ(k) 은 묵음 (stop consonant)로 거의 무음 (silence)에 가깝다.

4) 한국어 단독 숫자음은 모두 단음절어이다.

(2) 시간 영역에서의 음성 신호 분석

1) 영교차율 : 분석구간 프레임내에서 파형이 영점축과 교차하는 회수이고 과거 몇년동안 음성분할, 분석, 인식에 많이 사용되었다 또한 화자의 성량에 독립적이고 외관상으로는 주파수 평면의 파라 메터보다 화자에 덜 의존적이므로 디지틀 신호처리 과정에 매우 적합하다.

일반식 표현

$$Z = \sum_{m=1}^{N} \frac{|1 - sgn(X_{(m+1)}) sgn(X_{(m)})|}{2} \quad \text{-----(1-1)}$$

여기서 $sgn[X_{(m)}] = \begin{cases} 1, & X_{(m)} \geq 0 \\ -1, & X_{(m)} < 0 \end{cases}$

단, X(m) : 단 구간내에서 샘플의 크기

본 연구소에서는 샘플수 (Sample point) N =256 으로 정했다.

2) 대수 에너지

일반적으로 음성신호에서 대수 에너지는 무성음 부분보다 유성음 부분이 크다.

에너지에 대한 일반식 표현

$$E = \sum_{m=-\infty}^{\infty} X^2(m) \quad \text{-----(2-1)}$$

(2-1)식을 짧은 시변계 단구간으로 제한시키면

$$E_m = \sum_{m=m-N+1}^{m} X^2_{(m)} \qquad \text{-----(2-2)}$$

이 식은 신호가 자승되어 들어오므로 큰 신호에선 매우 민감하고 무성음 부분으로 부터 유성음 부분을 구분하는데 사용된다.

특히 좋은 음질의 신호에 대해서는 무음 (silence)과 무성음 (unvoice)을 구분하는데 매우 유용하다.

그러므로 대수를 취해준 것이 대수 에너지이며 일반적인 표현식은

$$LE_{(m)} = 10\log\left[\sum_{m=m-N+1}^{m} X^2_{(m)}\right] \text{-----(2-3)}$$

3. 음성 신호의 끝점 검출 알고리즘

끝점 결정;

음성신호의 단구간 에너지에 대한 기본 개념을 바탕으로하여 전 구간중에 각 프레임의 대수 에너지를 측정하고 다음 식 들로 부터 결정된다.

K2 = 0.73 X MAXLE ---------- (3 - 1)

K3 = 0.705 X MAXLE ----------(3 - 2)

T2 = 5 (Frame) ----------(3 - 3)

여기서, MAXLE : 최대에너지 (Maximum Log Energy)

T2 : 프레임 수

(3-1)식은 전 구간중 최대에너지의 73(%)값이고

(3-2)식은 전 구간중 최대에너지의 70.5(%)값이며

(3-3)식은 최적 프레임 길이 이다.

그러므로 끝점 프레임 구간은

본 알고리즘에 의해 추정된 끝점 검출길이의 $\pm$ 1 (Frame)으로 설정하여 실험 결과 결정된 최종 프레임은 (3-4)식과 같다.

$$EF = FR \pm 1 \text{----------(3 - 4)}$$

여기서 EF : 결정된 최종 프레임

FR : 알고리즘에 의한 추정된 끝점 프레임

끝점 검출 파형 ; 그림 1. 참 조

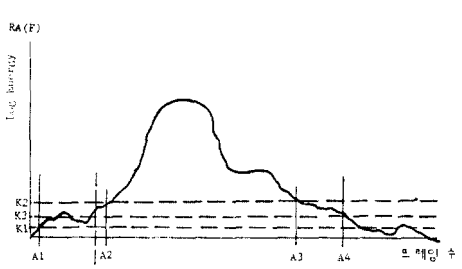


그림 1. 대수 에너지의 기준 파형.

Fig.1 Threshold waveform of Log Energy.

4. 음성 인식 실험 및 고찰

실험단계는 그림 2.와같다.

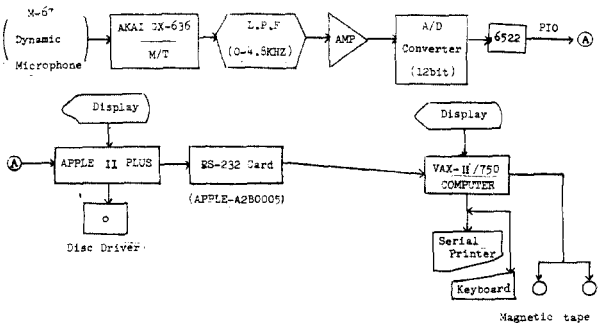
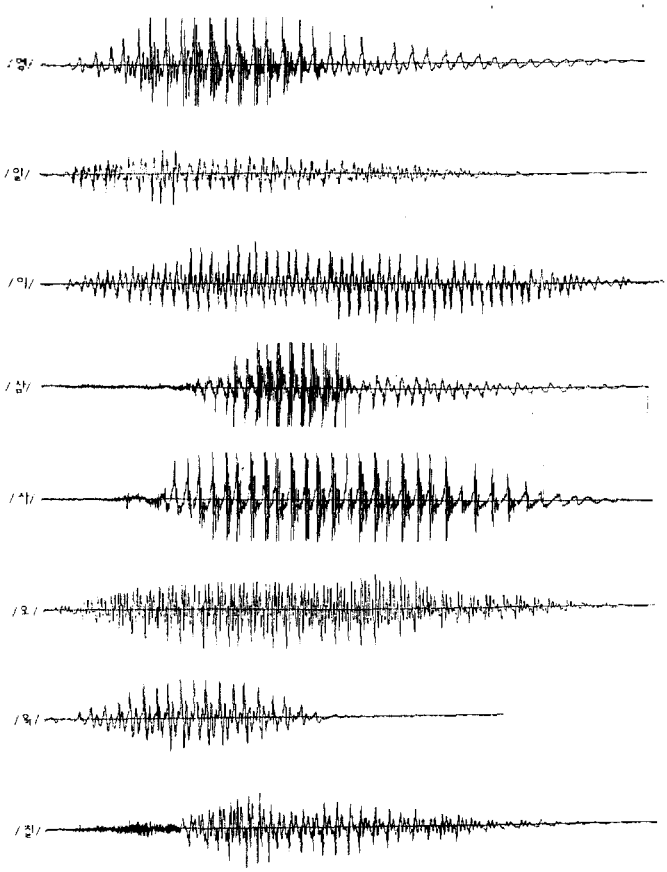


그림 2. 한국어 음성분석 및 검출을 위한 블록도.

Fig.2. A block diagram for Korean Speech analysis and detection.

- 화자수 : 성인 남성 3명
- 음 성 : /영/-/구/의 한국어 숫자음
- 어휘수 : 8개 숫자음 (140x24= 3360 프레임)

표준 끝점 검출실험 그림 3.과 같다.



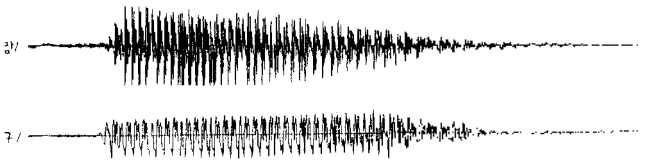


그림 3.표준 끝점 검출에 사용한 음성 파형.

Fig.3. Speech waveform using reference Endpoint Detection.

끝점 검출 분석을 위한 흐름도는 그림 4.와 같다.

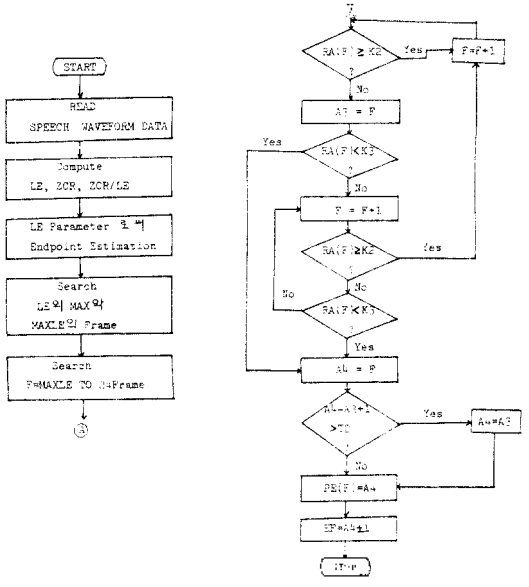


그림 4.끝점 검출 분석을 위한 전체 흐름도.

Fig.4. Flow Chart for Endpoint Detection analysis.

5. 결과 및 고찰

여러가지 파라메타로서 분석한 결과 2.C.3.은 유성음과 무성음을 구분하는데 좋은 파라메타로 추정되었고 1.C.는 음성부분과 비음성부분을 추정하는데 최적의 파라메타로 분석 결과 입증 되었다. 본 알고리즘에 의하여 분석한 끝점 검출 파형은 그림 5.와 같다.

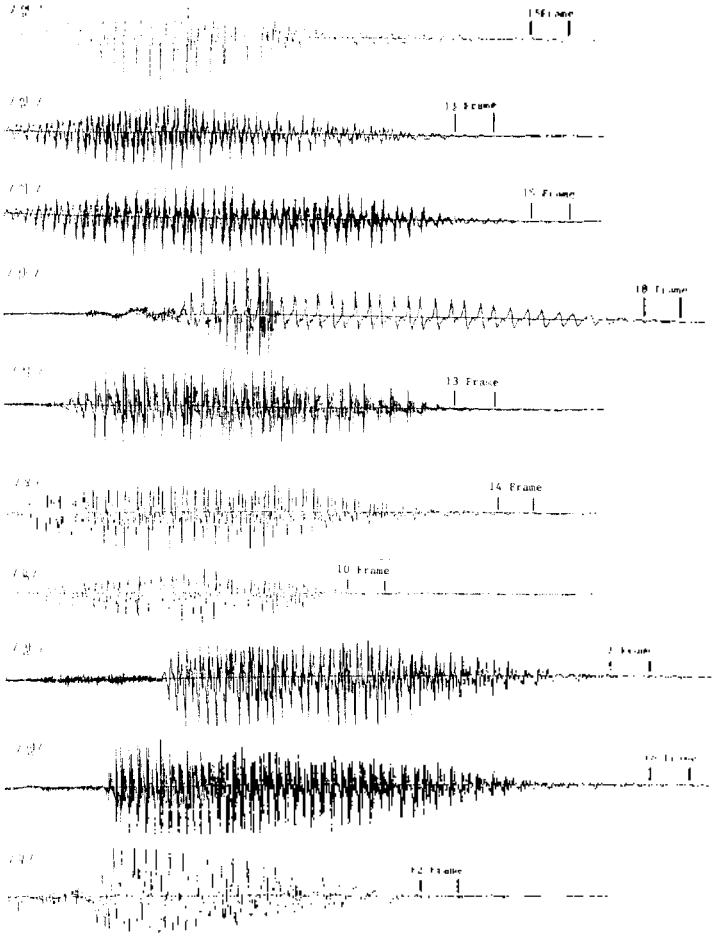


그림 5. 실험 결과의 끝점검출 파형.

Fig. 5. Endpoint Detection waveform of experiment result.

6.결 론

1)음성 신호에대한 파라메타는 영교차율, 대수 에너지, 에측 오차 에너지를 사용 했다.

2) 표준 끝점 검출은 음성 신호 부류에 대해 숫자 식별한 음성 데이타들로 부터 처리 되었다.

3) 끝점 검출은 각 파라메타에 대한 프레임 길이의 차이로써 분석하여 결정 했다.

4) 그러므로 분석 결과 가장 좋은 대수 에너지로써 선정하여 검출한 결과 1.4 % 이하의 오차율을 얻었다.

(참 고 문 헌)

1. L.R.Rabiner and M.R.Sambur," An Algorithm for Determining the Endpoints of Isolated Utterance," Bell System Technical Journal,vol.54, no.2,pp.297-315,Fib.1975.

2. T.B. Martin, "Practical Applications of Voice Input to Machines," Proc. of the IEEE, vol. 64, no. 4, pp. 487-502, April 1976.

3. T.B. Martin, Applications of Limited Vocabulary Recognition system," in Speech Recognition : Invited Papers Presented at the 1974 IEEE Symposium. New York: Academic Press,1975,pp. 55-72.

4. J.D.Markel and A.H.Gray, Linear Prediction of Speech, New York: Springer Verlag, 1976.

5. L.F. Lamel, L.R.Rabiner, A.E.Rosenberg, and J.G.Wilpon, "An improved Endpoint detector for islated word recognition," IEEE Trans. Acoust.,Speech,Signal Processing, vol ASSP-29, pp. 777-785, August, 1981.

6. 김 순협 "한국어 음성의 분석과 자동인식에 관한 연구" 박사학위 논문, 연세대학원, 1982. 12.