# 음성인식을 위한 자동차 소음환경에서의 끝점 검출

# Endpoint Detection in the Car Noise Environment for Speech Recognition

서 동 권*, 신 원 호*, 양 태 영*, 김 원 구**, 윤 대 희*
(Dong-Kwon Seo*, Won-Ho Shin*, Tae-Young Yang*, Weon-Goo Kim**, Dae-Hee Youn*)

※이 연구는 95년도 한국과학재단 연구비지원에 의한 결과임(과제번호:951-0915-059-1)

요 약

소음이 존재하지 않는 환경에서는 에너지 파라메터만으로도 정확한 끝점 검출을 수행할 수 있으나 신호대 잡음비가 0dB에 가까운 자동차 환경에서는 끝점 검출이 거의 불가능하다. 본 논문에서는 자동차 소음 환경에서 음성 구간 검출을 위하여 단구간 영교차율과 2~4 kHz의 주파수 영역 에너지를 사용한 끝점 검출 방법을 제안하였다. 제안된 방법과 기존의 방법의 성능을 DTW를 이용한 단독음 인식 시스템에 적용하여 인식률로 비교하였으며 제안된 음성 구간 검출 방법을 적용한 경우가 보다 좋은 인식률을 나타내었다.

ABSTRACT

If no noise is present, energy parameter does well in speech detection. However, in the car noise environmnt where the signal to noise ratio is nearly 0dB, the speech detection is almost impossible. In this paper, a method that utilize a short-term zerocrossing rate and 2-4kHz frequency band energy is proposed for the speech detection in the car noise environment.

The comparion of the performance of proposed method and that of the conventional ones is conducted using the DTW based isolated word recognizer. Experimental results show that proposed endpoint detection method is superior to the conventional ones.

## Ⅰ. 서 론

끝점 검출은 입력 신호로부터 묵음과 음성을 구분하는 과정이다. 끝점 검출은 특히 단독음 인식 시스템에서 필수적인 요소로서 정확한 끝점 검출은 인식 시스템의 성능을 크게 좌우한다. 예를 들어 DTW[1]를 이용한 단독음 인식 시스템의 경우 검출된 음성 구간만의 특징 벡터를 비교하기 때문에 인식 시스템의 성능을 향상 시키기 위해서는 정확한 끝점 검출이 요구된다. 또한 검출된 음성 구간이 불필요한 묵음을 포함하고 있다면 단어 인식에 소요되는 시간이 증가하게 된다.

일반적으로 끝점 검출에 사용되는 파라메터로는 단구간 에너지와 영교차율이 있다[2]. 에너지는 음성 구간과 묵음 구간을 구분하는데 이용되며, 영교차율은 음성의 모음과 자음 구간을 구분하는데 이용된다. 그러나 이러한 끝점 검출 과정은 자동차 소음이 존재하는 상황에서는 매우 어려워진다. 예로서, 소음이 존재하지 않는 환경에서는 에너지 파라메터만으로도 끝점 검출을 비교적 정확히 수행할 수 있으나 SNR이 0dB에 가까운 자동차 소음 환경에서는 끝점 검출이 거의 불가능하다[3][4].

본 논문에서는 자동차 소음 환경에서 음성을 검출하기 위한 방법으로 음성 신호의 단구간 영교차율과 2~4 kHz 주파수 영역의 에너지를 사용하는, 자동차 소음의 영향을 적게 받는 끝점 검출 방법을 제안하였다.

제안된 끝점 구간 검출 방법의 성능을 기존의 에너지 및 영교차율을 이용한 방법[2], Teager 에너지를 이용한 방법[5], 그리고 TF 파라메터를 이용한 방법[6]과 비교하였다. 성능 평가는 각 끝점 검출 방법을 DTW를 이용한 단독음 인식 시스템에 적용하여 시스템의 인식률을 비교하였다.

Ⅱ장에서는 제안된 끝점 검출 알고리듬에 대하여 설명하였고 Ⅲ장에서는 제안된 음성 구간 검출 방법과 기존의 음성 구간 검출 방법을 DTW를 이용한 단독음 인식 시스템

*연세대학교 전자공학과
**군산대학교 전기공학과

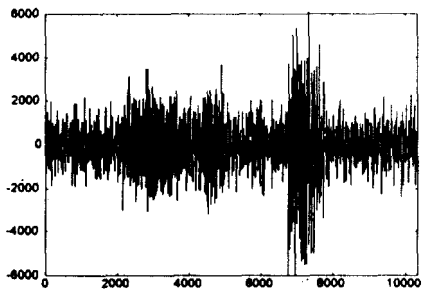
접수일자 : 1997년 11월 18일

에 적용하여 성능 비교한 결과를 나타내었다. Ⅳ에서는 본 논문의 결론을 맺었다.

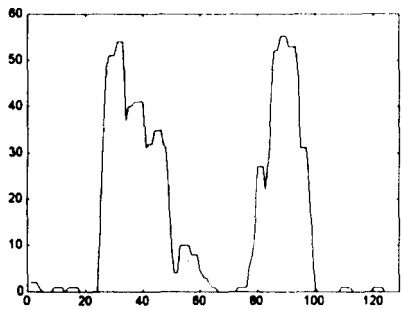
## Ⅱ. 끝점 검출 알고리듬

일반적으로 단구간 에너지와 영교차율을 사용하는 끝점 검출 방법에서는 먼저 단구간 에너지를 사용하여 음성의 안정적인 구간을 찾은 후에 영교차율을 이용하여 검출된 안정적인 음성 구간 양단의 자음 부분을 교정하여 최종적인 음성구간을 결정한다. 그러나 자동차 환경과 같이 SNR이 0dB에 가까운 소음 환경에서는 일반적인 단구간 에너지로 안정적인 음성 구간을 검출하는 것이 어려우며, 또한 영교차율을 이용하여 음성 구간의 처음과 끝 부분에서 자음과 모음을 구분하는 것은 거의 불가능하다.

그림 1(a)(b)는 자동차 소음이 SNR 0dB로 첨가된 음성 신호의 시간축 파형과 신호의 단구간 영교차율이다. 그림1(b)에서 알 수 있듯이 오히려 소음 환경에서는 단구간 영교차율로 첫 모음의 시작과 마지막 모음의 끝을 안정적으로 검출할 수 있다. 따라서 본 논문에서는 단구간 영교차율을 이용하여 음성에서 안정된 모음 구간을 검출하는 방법을 제안하였다.




(a)




(b)

그림 1. 자동차 소음이 SNR 0dB로 첨가된 음성의 끝점 검출
(a) 신호의 시간축 파형 (b) 신호의 단구간 영교차율

Fig. 1. speech detection in noisy speech signal to which car noise is added for 0dB SNR.
(a) speech signal (b) short-term zerocorssing rate

다음으로 자동차 소음 환경에서 음성의 처음 부분의 마찰음이나 파열음 그리고 음성의 마지막 부분의 유음 등을 검출하기 위하여 2~4 kHz 주파수 영역의 에너지를 사용하여 교정하는 방법을 제안하였다. 2~4 kHz의 주파수 영역의 에너지는 대부분의 자동차 소음과 모음의 영향을 받지 않기 때문에 음성의 처음과 끝 부분의 자음 구간을 비교적 정확히 검출할 수 있다.

따라서 먼저 음성 신호의 단구간 영교차율을 이용하여 음성의 안정한 구간을 찾은 후 2~4 kHz 주파수 영역의 에너지를 이용하여 음성의 최종 끝점 검출을 교정하여 소음환경에서 비교적 정확하게 끝점을 검출할 수 있다.

그림 2에 제안된 끝점 검출 방법을 도시하였다. 먼저 잡음 구간으로 여겨지는 처음의 5 프레임으로 부터 끝점 검출에 사용되는 문턱값을 결정한다. 프레임 영교차율의 문턱값은 영교차율의 평균과 분산을 이용하여 구하여지고 2~4 kHz 영역의 프레임 에너지의 문턱값은 에너지의 평균을 사용하여 구해진다.

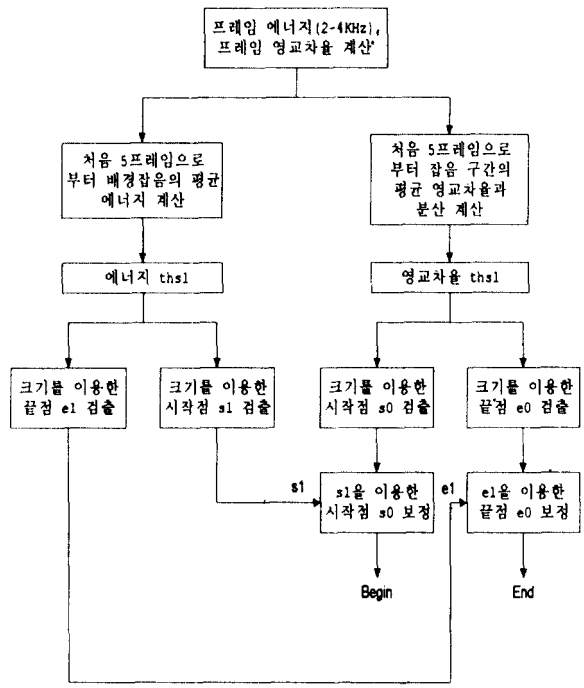



그림 2. 끝점 검출 과정
Fig. 2. endpoint detection procedure

다음으로 입력 신호의 각 프레임 영교차율을 구하여 영교차율이 연속적으로 6 프레임 이상 문턱값을 넘으면 처음으로 문턱값을 넘는 시점을 안정적인 음성 구간의 시작점으로 결정하고, 프레임 영교차율이 연속적으로 10 프레임 이상 문턱값보다 작으면 처음으로 문턱값보다 작아진 시점을 음성 구간의 끝점으로 정한다.

이와 동일한 과정을 적용하여 2~4 kHz 영역의 프레임 에너지 파라메터로 음성의 시작점과 끝점을 구하고 영교

차율로 구한 음성 구간과 비교하여 최종적인 음성 구간을 결정하게 된다.

## Ⅲ. 실험 및 고찰

제안된 음성 검출 방법의 성능을 평가하기 위하여 DTW를 이용한 단독음 음성 인식 시스템을 사용하여 인식률을 비교하였다.

### 1. 데이터 베이스 및 인식 시스템 구성

음성 인식 실험에 사용된 데이터 베이스는 30개의 단독음으로 구성되었다. 기준 패턴으로는 20대의 남성 화자 3명과 20대의 여성 화자 2명이 조용한 사무실 환경에서 각 단어를 두 번씩 발음한 음성을 사용하였으며 시험 패턴으로는 기준 패턴의 음성과 동일한 화자가 2개월 후에 같은 환경에서 한 번씩 발음한 음성에 잡음을 첨가하여 사용하였다.

실험에 사용된 소음 신호는 시속 100 km/h로 주행중인 자동차에서 녹음한 소음을 이용하여, 소음이 포함되지 않은 음성 신호와 각각 20dB, 10dB, 0dB의 신호대 잡음비를 가진 신호를 만들어 인식률을 비교하였다.

실험에 사용된 음성 신호는 PC의 사운드카드로 8 kHz, 16비트로 표본화되었다. 음성 신호는 20msec의 크기를 갖는 해밍(hamming) 윈도우를 사용하여 10msec씩 이동하면서 분석되었으며 특징 파라메터로는 10차의 LPC 켑스트럼을 사용하였다.

기준 패턴은 각 단어당 5개씩 생성하였으며 시험 패턴과 기준 패턴의 비교는 normalize/warp 방법을 이용한 수정된 DTW 알고리듬[7]을 사용하였다.

### 2. 실험 결과 및 고찰

본 논문에서는 끝점 검출 방법의 성능을 비교하기 위하여 DTW를 이용한 고립 단어 인식 시스템에서 자동차 소음이 첨가된 음성을 대상으로 제안된 끝점 검출 방법의 성능을 기존의 에너지 및 영교차율을 이용한 방법, Teager 에너지를 이용한 방법, TF 파라메터를 이용한 방법, 그리고 손으로 검출한 끝점 정보를 사용한 경우와 비교하였다.

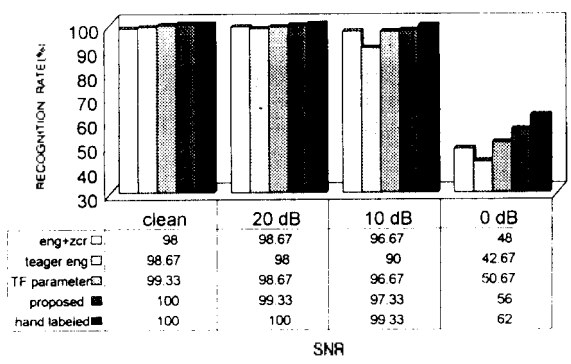



그림 3. 끝점 검출 방법과 SNR에 따른 인식률

Fig. 3. recognition rates for the endpoint detection methods and SNR

그림 3에 제안된 끝점 검출 방법과 기존의 방법들을 이용한 경우의 인식 시스템의 인식률을 SNR별로 나타내었다.

그림 3에서 알 수 있듯이 소음이 첨가되지 않은 경우나 SNR이 20dB, 10dB인 경우에는 끝점 검출 방법에 따른 인식률의 차이가 거의 없었으나 전체적으로 제안된 끝점 검출 방법이 약간 좋은 성능을 나타내었다. SNR 0dB의 경우에는 끝점 검출 방법에 따른 인식률의 차이가 거의 없었으나 전체적으로 제안된 끝점 검출 방법이 약간 좋은 성능을 나타내었다. SNR 0dB의 경우에는 제안된 방법을 사용한 경우의 인식률이 기존의 방법을 사용한 경우보다 많은 성능 향상을 나타내었으며 손으로 검출한 끝점 정보를 사용한 경우의 인식률에 접근하는 것을 알 수 있었다.

Teager 에너지를 사용하는 끝점 검출 방법의 경우에는 SNR이 10dB 이하인 경우에 다른 방법보다 큰 오차를 나타내었다.

## Ⅳ. 결 론

본 논문에서는 자동차 소음 환경에서 음성 구간 검출을 수행하기 위하여 단구간 영교차율과 2~4 kHz 영역의 에너지를 사용하는 방법을 제안하였다.

먼저 음성 신호의 단구간 영교차율을 이용하여 음성의 안정한 구간을 찾은 후 2~4 kHz 주파수 영역의 에너지를 이용하여 음성의 최종 끝점 검출을 교정하여 소음환경에서 보다 정확하게 끝점을 검출할 수 있었다.

제안된 방법과 기존의 방법의 성능을 DTW를 이용한 단독음 인식 시스템에 적용하여 인식률로 비교하였으며, 특히 SNR이 낮은 경우 제안된 음성 구간 검출 방법을 적용한 경우가 보다 많은 인식률 향상을 나타내었다.

## 참 고 문 헌

1. L. R. Rabiner, A. E. Rosenberg and S. E. Levinson, "Considerations in Dynamic Time Warping Alogrithms for Discrete Word Recognition," IEEE Trans. Acoust., Speech, Signal Processing, Vol. ASSP-26, No. 6, pp. 575-582, Dec. 1978.

2. L. R. Rabiner and M. R. Sambur, "An Algorithm for Determining the Endpoint of Isolated Utterance," *Bell Syst. Tech. J.*, Vol. 54, No. 2, pp. 297-315, Feb. 1975.

3. H. Kobatake and K. Tawa, "Speech/Nonspeech Discrimination for Speech Recognition System under Real Life Noise Enviroments," in *Proc. ICASSP*, pp. 356-368, May 1989.

4. J. Junqua, B. Reaves and B. Mak, "A Study of Endpoint Detection Algorithms in Adverse Conditions: Incidence on a

DTW and HMM Recognizer," in *Proc. EUROSPEECH*, Vol. 3, pp. 1371-1374, 1991.

5. G. S. Ying, C. D. Mitchell, L. H. Jamieson, "Endpoint detection of isolated utterances based on a modified teager energy measurement," in *Proc. ICASSP*, pp. 732-735, 1993.
6. Jean-Claude Junqua, Brian Mak, and Ben Reave, "A robust Algorithm for Word Boundary Detection in the Presence of Noise," *IEEE Trans. Acoust., Speech, and Signal Processing*, Vol. 2, No. 3, pp. 406-412, 1994.
7. C. Myers, L. R. Rabiner and A. E. Rosenberg, "Performance Tradeoffs in Dynamic Time Warping Algorithm for Isolated Word Recognition," *IEEE Trans. Acoust., Speech, Signal Processing*, Vol. ASSP-28, No. 6, pp. 623-635, Dec. 1980.

▲서 동 권(Dong-Kwon Seo) 1970년 10월 28일생
1995년 2월 : 연세대학교 전자공학과 졸업(공학사)
1997년 2월 : 연세대학교 본대학원 전자공학과 졸업(공학석사)
1997년 3월~현재 : 만도기계중앙연구소 연구원
※주관심분야 : 음성인식, 음성신호처리

▲신 원 호(Won-Ho Shin)
한국음향학회지 1996년 15권 2호 참조
현재 : 연세대학교 대학원 전자공학과 박사과정

▲양 태 영(Tae-Young Yang)
한국음향학회지 1996년 15권 2호 참조

▲김 원 구(Weon-Goo Kim)
한국음향학회지 1996년 15권 2호 참조
현재 : 군산대학교 전기공학과 조교수

▲윤 대 희(Dae-Hee Youn)
한국음향학회지 1996년 15권 2호 참조
현재 : 연세대학교 전자공학과 교수