잡음음성에서의 음성 활성화 구간 검출 방법

이광석 · 추연규 · 김현덕
진주산업대학교

Speech Active Interval Detection Method in Noisy Speech

Kwang-seok Lee · Yeon-gyu Choo · Hyun-deok Kim
Jinju National University
email : kslee@jinju.ac.kr

요 약

음성통신 및 음성인식에 있어서 잡음이 섞인 음성으로 부터 음성의 활성화 구간의 검출은 대단히 중요한 과정으로 알려져 있다. 따라서 본 연구에서는 잡음음성으로 부터 음성의 활성화 구간을 검출하기 위하여 스펙트럴 엔트로피와 복합으로 구성하는 특징 파라미터를 제안하고 에너지를 기반으로 음성 활성화 구간을 검출하는 방식과 성능 비교 실험을 행하였다. 실험결과, 노이즈 환경에서 다른 파라미터에 비하여 제안한 파라미터에 의한 음성 활성화 구간 검출의 성능이 우수함을 확인할 수 있었다.

ABSTRACT

It is important to detect speech active interval from Noisy Speech in speech communication and speech recognition. In this research, we propose characteristic parameter with combining spectral Entropy for detect speech active interval in Noisy Speech, and compare performance of speech active interval based on energy. The results shows that analysis using proposed characteristic parameter is higher performance than others in noisy environment.

키워드

Speech Processing, Speech Characteristic Parameter

Ⅰ. 서 론

음성통신 및 음성인식의 전처리 단계로 음성인식의 인식률에 결정적인 영향을 미치는 중요한 과정의 하나로 음성 활동구간 검출(VAD: Voice Activity Detection)이 이에 해당한다. 음성 활동구간 검출은 여러 잡음과 다양한 사운드를 분류하여 실제 음성구간만을 검출하는 것이다. 그러나 실제, 잡음과 주위환경에 독립적이고 안정된 음성 활동 구간 검출 알고리듬을 구현한다는 것은 어려운 작업이다.

일반적인 음성 활동 구간 검출은 에너지 기반 방식이다.[1],[3] 그러나, 에너지를 이용할 경우, 깨끗한 환경에서는 뛰어난 성능을 보이는 반면 실제 잡음이나 외부 사운드가 부가되는 환경에서는 적용하기 어렵다. 특히 음성의 시작점에서의 자음구간, 에너지가 적은 모음은 검출하기가 어렵다. 그리고 외부의 기침소리 혹은 음성구간의 시작과 끝에 있는 숨소리 잡음과 같은 부가적인 사운드 성 잡음 역시 인식처리 시에는 무시되어야 한다. 이를 해결하기 위하여, 검출할 음성 구간을 음성의 임계 지속시간보다 긴 경우는 임계값을 초과하는 단구간 평균 에너지로 보통 검출하고 음성구간의 시작점을 에너지 임계값에 의하여 검출된 위치보다 일정 정도 앞에 위치시켜 결정하고 있다. 또한, 더욱더 신뢰할만한 음성구간 검출을 위하여 영 교차율과 더불어 비교하여 검출하기도 한다.[1] 또 다른 음성 활동 구간 검출 방법으로는 스펙트럴 분석을 통한 접근법으로 대표적인 방법이 입력신호와 레퍼런스 잡음 스펙트럼간의 스펙트럴 차이를 이용하여 음성 활동 구간을 검출하는 방법이다.[2]

본 연구에서는 안정된 음성 구간검출을 위하여 정보이론의 주요 개념인 Entropy를 적용한 스펙트럴 Entropy를 사용하고자 한다. 이 방법은 J.L. Shen이 처음으로 음성에 적용하여 사용하였으며 Shen은 실험을 통하여 음성의 스펙트럴 Entropy가 비음성의 그것과 매우 다름을 보여주었다.[5] 따라서 여기서는 이를 보완하여 여러 형태의 잡음에 대하여 스펙트럴 Entropy를 재구성한 새로운 특징 파라미터를 제안한다. 그리고 제안한 재구성 특징파라미터와 에너지를 서로 비교 분석하여 음성 활동 구간 검출 파라미터로서의 적용 가능성을 확인하기 위한 연구이다.

Ⅱ. Entropy

1. Entropy

Entropy는 Shannon의 정보이론에 기반한 정

보량을 측정하는 척도로서 출력(x_i)로부터 유도된 정보는 그것의 확률 값에 의존한다. 만약, 확률 $P(x_i)$가 작다면 드물게 발생하므로 출력으로부터 많은 정보를 얻을 수 있다. 반대로, 확률이 크다면 쉽게 예상되므로 정보는 적다는 이론이다. 정보량은 식(1)과 같이 정의된다.

$$I(x_i) = \log\frac{1}{P(x_i)} \quad (1)$$

X를 유한 표본공간 $S=\{x_1, x_2, \cdots, x_i, \cdots\}$에서 x_i 값을 취하는 이산랜덤변수라고 가정하면 심볼 x_i는 랜덤변수 X의 확률분포에 따른다. 여기서, 랜덤변수 X의 Entropy $H(X)$를 다음과 같은 정보량의 평균값(기대값)으로 정의한다.

$$\begin{aligned} H(X) &= E\,[I(X)] \\ &= \sum_S P(x_i) I(x_i) \\ &= \sum_S P(x_i) \log\frac{1}{P(x_i)} \\ &= E\,[-\log P(X)] \end{aligned} \quad (2)$$

2. 스펙트럴 Entropy

스펙트럴 Entropy는 입력신호에 대하여 FFT하여 음성의 주파수영역의 파워 스펙트럼의 확률밀도를 구하고 Entropy를 계산하는 단계로 이루어진다. 스펙트럼에 대한 확률밀도는 식(3)과 같이 일종의 주파수 성분에 대한 정규화의 효과를 가지는 방식으로 추정된다.

$$p_i = \frac{s(f_i)}{\sum_{k=1}^{M} s(f_k)}, \quad i = 1, \cdots, M \quad (3)$$

여기서, $s(f_i)$는 주파수 성분 f_i에 대한 파워 스펙트럼이고, p_i는 대응하는 확률밀도함수이다. M은 FFT에서의 주파수 성분의 총 개수이며 위에서 설명한 Entropy를 계산한다. 그러나 위와 같은 처리는 잡음과 비 음성 부분이 Entropy가 강조되는 결과를 얻게 되므로 이 값의 역수를 취함으로서 음성부분이 부각되는 처리를 가한다. 그리고 마지막 단계로 계산된 Entropy를 재구성하게 되며 그 과정을 그림 1에 나타내었다.

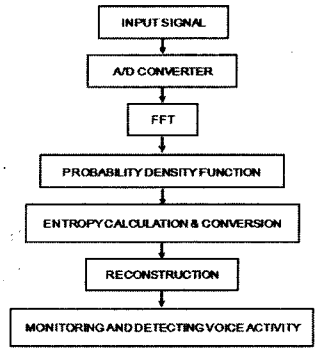


그림 1. 스펙트럴 Entropy 산출

3. 스펙트럴 Entropy의 재구성

스펙트럴 Entropy의 재구성은 음성구간 검출시에 임계값 설정에 여유를 주고 음성부분만을 강조하는 방식으로 이루어진다. 본 연구에서는 다음과 같은 다양한 방법으로 VAD용 특징 파라미터를 재구성하였다.

(1) 특징 파라미터 1 : Entropy
(2) 특징 파라미터 2 : Entropy×Log energy
(3) 특징 파라미터 3 : Entropy×(ZCR_max-ZCR)
(4) 특징 파라미터 4 : Entropy×Log energy×(ZCR_max-ZCR)
(5) 특징 파라미터 5 : Entropy×Speech entropy guassian distribution function
(6) 특징 파라미터 6 : Entropy×Speech entropy guassian distribution function×Log energy

여기서, ZCR_max는 영 교차율이 음성구간에 낮고 잡음구간에서 높으므로 음성구간을 강조하기 위하여 반전시키기 위하여 설정하는 값이다. 특징파라미터 5, 6에서 적용한 가우시안 분포함수는 미리 음성 샘플데이터로 얻은 Entropy의 가우시안 분포함수를 적용하여 음성부분이 강조되도록 한 것이다. 적용된 가우시안 분포함수는 그림2와 같다.

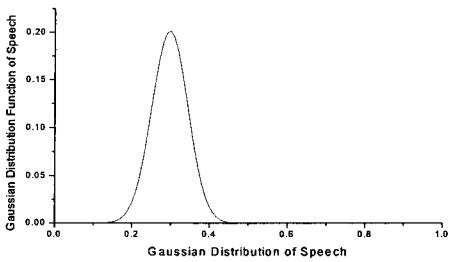


그림 2. 음성 Entropy의 가우시안 분포

III. 실험결과 및 고찰

1. Database

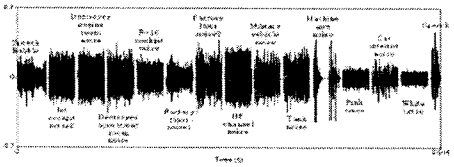
본 연구에서는 NOISEX92의 데이터베이스를 이용하였다.[9] 본 데이터는 표1과 같은 다양한 종류의 잡음들로 구성되어 있으며 19.98㎑-16bit으로 anti-aliasing 필터링 데이터를 16㎑-16bit로 재 샘플링 후 실험에 이용하였다.

표 1. NOISEX-92 D/B

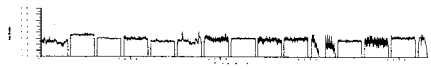
Noisex-92 Database			
1	Speech babble	8	HF channel noise
2	Jet cockpit noise2	9	Military vehicle noise
3	Destroyer engine room noise	10	Tank noise
4	Destroyer operation room noise	11	Machine gun noise
5	F-16 cockpit noise	12	Pink noise
6	Factory floor noise1	13	Car interior noise
7	Factory floor noise2	14	White noise

2. 결과 및 고찰

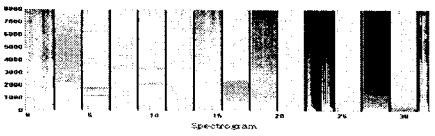
평가용 샘플데이터로 2절에서 언급한 재구성한 특징파라미터들과 에너지를 사용하여 실험한 여러 결과들을 그림3에 나타내었다.



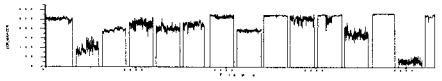
(a) Source Signal



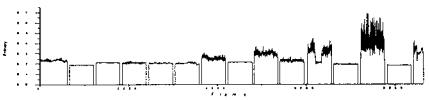
(b) Log Energy



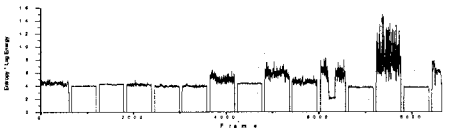
(c) Spectrogram



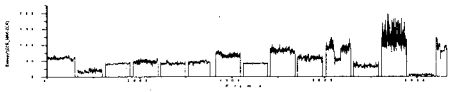
(d) ZCR_max-ZCR



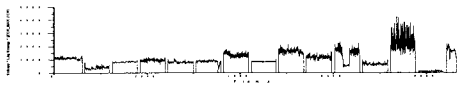
(e) Entropy



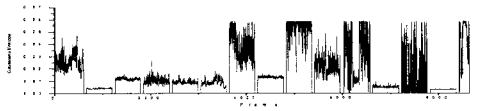
(f) Entropy×Log Energy



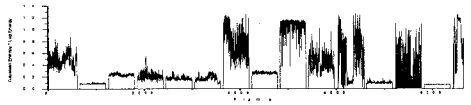
(g) Entropy×(ZCR_max-ZCR)



(h) Entropy×Log Energy×(ZCR_max-ZCR)



(i) Gaussian Entropy



(j) Gaussian Entropy×Log Energy

그림 3. 에너지(I)과 재구성 파라미터의 비교

그림 4는 5초간의 각 샘플 데이터에 대하여 재구성 특징파라미터들의 평균값을 도표로 나타낸 것이다. 모든 특징파라미터는 에너지의 경우와 비교하기 위하여 음성을 기준으로 설정하였다. 에너지의 경우 거의 구별을 위한 임계값 설정이 어려운 반면, 재구성 특징파라미터들에서는 어느 정도의 임계값 설정에 여유가 있는 것을 확인할 수 있다. 특히 가우시안 분포함수를 적용한 특징파라미터 5, 6이 가장 우수함을 알 수 있었으며 여기서, 신호 15는 음성이다.

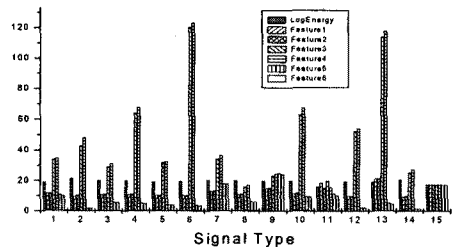


그림 4. 에너지(II)와 재구성 파라미터의 비교

IV. 결 론

최근 하드웨어 처리속도 향상의 도움으로 단순한 에너지를 기반으로 한 음성구간 검출에서 더 나아가 다양한 잡음과 다양한 오디오 사운드의 분류를 통한 보다 신뢰성 있고 견고한 음성구간 검출에 대한 다각적인 방식이 모색 연구되고 있다.

본 연구에서는 음성의 활동구간 검출을 위하여 잡음에 강인한 특징 파라미터로 스펙트럴 Entropy와 이를 재구성한 여러 가지 특징 파라미터를 이용하여 에너지에 의한 경우와 서로 비교 실험 분석하였다.

음성구간의 검출을 위한 실험한 결과, 에너지에 의한 경우보다 스펙트럴 Entropy를 재구성한 특징파라미터가 보다 효과적이며 특히, 음성에 대한 Entropy의 가우시안 분포함수를 적용하여 음성을 강조한 특징파라미터가 가장 우수한 성능을 보이는 것을 확인할 수 있었다.

참고문헌

[1] Sadaoki Furui: "Digital Speech Processing Synthesis, and Recognition", Maecel Dekker, Inc., pp. 248-249, 2007.

[2] Xuedong Huang, Alex Acero, Hsiao-Wuen Hon: "Spoken Language Processing", Prentice Hall, pp120-130, 2007.

[3] Nikos Doukas, Patrick Naylor and Tania Stathaki: "Voice Activity Detection Using Source Separation Techniques", Signal Processing Section, Proc. Eurospeech '06

[4] J.L. Shen, J.Hung, L.S.Lee : "Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environments", Preceeding of ICLP-05, 2005.

[5] J.Sohn and W.Sung : "A Voice Activity Detector Employing Soft Decision Based Noise Spectrum Adaptation", in Proc. IEEE International Conference on Acoustics, Speech, and Signal Processing, pp. 356-368, 2006.

[6] http://spib.rice.edu/spib/select_noise.html