차량 잡음 환경에서 엔트로피 기반의 음성 구간 검출
Voice Activity Detection Based on Entropy in Noisy Car Environment

노용완* · 이규범* · 이우석* · 홍광석*
Yong-Wan Roh · Kue-Bum Lee · Woo-Seok Lee and Kwang-Seok Hong

요 약

정확한 음성 구간 검출은 음성 인식 및 음성 코딩 그리고 음성 통신 시스템 등과 같은 음성 어플리케이션의 성능에 큰 영향을 미친다. 본 논문에서는 실제 운전하고 있는 상태에서 다양한 차량 노이즈 환경의 음성 구간 검출 방법을 제안한다. 기존의 음성 구간 검출은 시간 에너지, 주파수 에너지, 영 교차율, spectral entropy 등 다양한 방법을 사용하였으며 잡음 환경에서 급격하게 성능이 저하되는 단점이 있었다. 본 논문에서는 기존의 spectral entropy를 기반으로 하여 MFB(Mel-frequency Filter Banks) spectral entropy, 기울기 FFT(Fast Fourier Transform) spectral entropy, 기울기 MFB spectral entropy를 이용한 음성 구간 검출 방법을 제안한다. MFB는 멜 스케일과 FFT를 곱한 것으로 멜 스케일은 인간이 소리를 인지할 때 주파수에 대해 비선형적인 스케일이며 음성의 특징을 잘 반영한다. 제안한 MFB spectral entropy 방법은 다양한 차량 잡음 환경에서 음성 및 비음성 분별 능력을 향상 시킬 수 있으며 실험 결과 93.21%의 음성 구간 검출율을 나타내었다. 이는 기존의 spectral entropy 방법과 비교할 때 MFB를 이용한 음성 구간 검출 방법이 3.2%의 검출율이 향상 되었다.

Abstract

Accurate voice activity detection have a great impact on performance of speech applications including speech recognition, speech coding, and speech communication. In this paper, we propose methods for voice activity detection that can adapt to various car noise situations during driving. Existing voice activity detection used various method such as time energy, frequency energy, zero crossing rate, and spectral entropy that have a weak point of rapid decline performance in noisy environments. In this paper, the approach is based on existing spectral entropy for VAD that we propose voice activity detection method using MFB (Mel-frequency filter banks) spectral entropy, gradient FFT(Fast Fourier Transform) spectral entropy, and gradient MFB spectral entropy. FFT multiplied by Mel-scale is MFB and Mel-scale is non linear scale when human sound perception reflects characteristic of speech. Proposed MFB spectral entropy method clearly improve the ability to discriminate between speech and non-speech for various in noisy car environments that achieves 93.21% accuracy as a result of experiments. Compared to the spectral entropy method, the proposed voice activity detection gives an average improvement in the correct detection rate of more than 3.2%.

Keywords : Voice activity detection, spectral entropy, delta spectral entropy, MFB

I. 서 론

음성 인식 기술의 향상으로 음성 인식 제품의 실용화가 점차이루어지고 있지만 아직까지 만족할 만한 결과를 얻지 못하고 있으며 연속 음성 인식의 경우 주위의 잡음 환경에 의해 음성 구간 검출에 상당한 어려움이 있다[1][13]. 음성 검출 (VAD: Voice Activity Detection)은 음성 부호화, 음성 인식, 음성 향상, 핸즈프리 컨퍼런스, 에코 제거기와 같은 다양한 응용 예에서 음성과 비음성을 구별해 내는 작업을 한다[2]. 그 중에서도 자동차 환경에서 발생하는 잡음이나 마이크로폰 등과 같은 전송 채널에 의한 음성 신호 왜곡 현상으로 인한 인식 성능 저하의 문제는 아직까지

* 성균관대학교 정보통신공학부
논문 번호 : 2008-1-10 접수 일자 : 2008. 3. 12
심사 완료 : 2008. 4. 18
* 본 연구는 정보통신부 및 정보통신연구진흥원의 IT신성장동력핵심기술 개발사업과 지식경제부 및 정보통신연구진흥원의 대학 IT 연구센터 지원사업의 연구결과로 수행되었음.
[2007-S025-01, VDMS기술개발]
(IITA-2008-(C1090-0801-0046))

완전히 해결하지 못한 분야 중의 하나이다[14]-[16].

최근 많은 국가에서 무선 통신 기술을 이용하여 운전자, 자동차, 지명, 여러 가지 정보 네트워크 시스템과 연결시킬 뿐 아니라 자동차 운행 중에 필요한 명령어 등 다양한 정보를 제공하는 지능형 교통 시스템 (Intelligent Transport System)을 개발 하고 있다[3]. 특히 운행을 위한 교통 상황, 도로 구조, 상황에 따른 항법 가이드 등 다양한 정보를 운전자가 쉽게 조작하기 위해서는 HCI(Human Computer Interface)가 반드시 필요하며 잡음 환경에서 우수한 성능을 갖는 음성 인식 기술을 위하여 VAD는 필수적이다[17].

음성 검출을 위해 사용되는 특징들로는 신호의 에너지, 영교차율과 같은 시간 영역의 특징들과, 주파수 스펙트럼의 통계적인 특징에 기반을 둔 방법들로 나눠진다[4]-[5]. 시간 영역의 특징을 이용한 음성 구간 검출 방법들은 간단한 수학적 계산에 의해 빠르게 음성 구간을 검출할 수 있다는 장점은 있으나 음성 입력에 잡음이 첨가되면 성능이 저하되는 단점이 있다[6]-[8].

기존의 spectral entropy를 이용한 음성 검출 방법은 잡음 환경에서의 인식 성능이 시간 영역의 특징을 사용한 경우 보다 우수하였다[9]-[10]. 하지만 spectral entropy를 이용한 음성 검출 방법도 만족할 만한 결과를 얻지 못했으며, 실시간 응용에 적합하지 않은 단점이 있다.

본 논문에서는 기울기 spectral entropy와 MFB 기반의 spectral entropy를 사용하여 차량 잡음 환경에서의 음성 검출 방법을 제안한다. 차량 단말기에 적용할 것을 감안하여 모든 MFB(Mel-frequency Filter Banks)를 사용하는 것이 아니라 5개의 필터를 선별하여 사용하였다. 또한 기존의 spectral entropy의 VAD 성능 제안한 방법들의 VAD 성능을 비교 평가 하였다.

본 논문의 구성은 다음과 같다. 2장에서는 기존에 제안한 spectral entropy를 이용한 VAD 방법에 대해 살펴보고, 3장에서는 MFB 기반 spectral entropy 방법 및 제안한 기울기 spectral entropy를 이용한 음성 구간 검출 방법에 대해 기술한다. 4장에서는 제안한 방법과 기존 spectral entropy 방법을 이용한 음성 구간 검출 비교 실험을 한 후 5장에서 결론 및 향후 연구 방향에 대해 기술한다.

II. Spectral Entropy

Entropy는 Shannon의 정보 이론에서 데이터에 포함되어 있는 정보의 양을 나타낸다. Spectral entropy는 entropy 이론을 음성 신호 처리에 적용한 것으로 잡음 환경에서 신호의 에너지나 영교차율 등으로 음성검출의 문제점을 개선하기 위한 특징들 중의 하나로 사용되고 있다[9]. Spectral entropy를 이용한 음성 구간 검출 방법은 그림 1과 같다.

그림 1에서 음성은 먼저 프리엠퍼시스 필터를 사용하여 진폭 주파수의 특성을 평탄하게 유지하기 위해 사용한다. 프리엠퍼시스 필터를 통과한 음성 신호는 단구간 신호로 분할되며 이를 프레임이라 한다. 프레임 길이는 32ms(512 sample)로 하였으며 두 인접 프레임 사이의 오버랩은 16ms(256 sample)로 하였다. 그 후 각 프레임마다 해밍 윈도우를 사용하였다. 시간 영역의 음성 신호를 주파수 영역으로 변환 하기위해 FFT를 사용하였다. FFT는 $X(i,n)$으로 정의하였으며 식 (1)로 구한다.

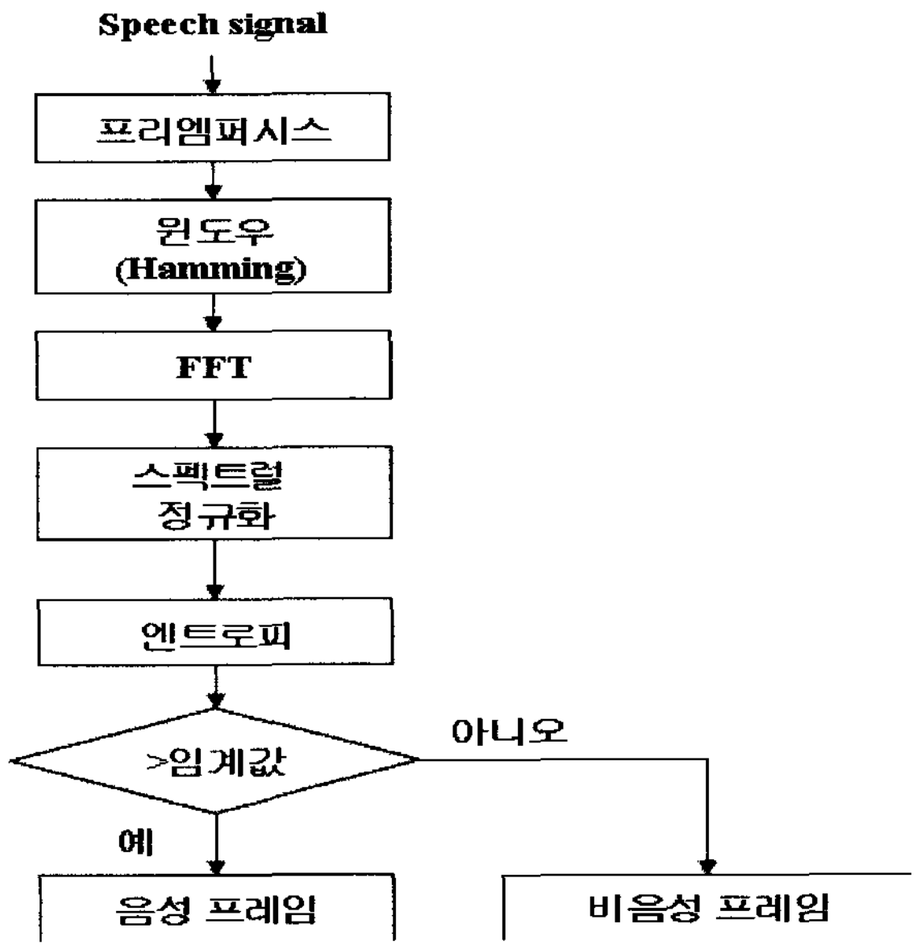


그림 1. 음성 구간 검출을 위한 기존 spectral entropy의 블록 다이어그램

Fig. 1. The block diagrams of existing spectral entropy method for VAD

$$X(i,n)=\sum_{m=1}^{M} x(m,n)e^{-j\frac{2\pi}{N}im} \tag{1}$$

여기서, $X(i,n)$은 n번째 프레임의 i번째 주파수 성분을 나타낸다. M은 FFT 포인트의 개수를 나타낸다. $x(m,n)$은 시간 영역의 음성신호 n번째 프레임의 m번째 샘플을 의미한다.

FFT를 취한 후 각 프레임의 파워스펙트럼 $S(i,n)$을 구하며 식 (2)에 나타내었다.

$$S(i,n)=|X(i,n)|^2 \tag{2}$$

스펙트럼의 확률밀도는 주파수 성분의 정규화 방법을 사용하여 얻을 수 있다. 파워 스펙트럼 정규화는 $P[S(i,n)]$로 정의되며 식 (3)에 나타내었다.

$$P[S(i,n)]=\frac{S(i,n)}{\sum_{m=1}^{M/2} S(m,n)} \tag{3}$$

마지막 단계로 음성과 비음성의 구별을 위한 entropy를 계산하는 것이며 기존의 entropy 방법 $H(n)$은 식 (4)에 나타내었다.

$$H(n) = -\sum_{i=1}^{M/2} P[S(i,n)]\log P[(S(i,n)] \quad (4)$$

Spectral entropy 방법은 다양한 노이즈 환경에서 에너지 기반의 음성 구간 검출 방법보다 효율적이다[9]-[10].

Ⅲ. 제안한 spectral entropy 기반의 VAD 방법

본 논문에서는 3가지의 spectral entropy 기반의 VAD 방법을 제안하였으며 MFB spectral entropy, 기울기 spectral entropy와 기울기 MFB spectral entropy 방법이다. MFB spectral entropy는 스펙트럼 정규화 및 entropy를 계산하기 전에 FFT 파워 스펙트럼에 멜-스케일의 삼각 필터를 곱한 것이다. 기울기 spectral entropy는 2개의 인접한 FFT 파워 스펙트럼 계수(주파수 성분)의 차를 사용하여 entropy를 구한 것이며 기울기 MFB spectral entropy는 2개의 인접한 필터 뱅크 에너지 차를 사용하여 entropy를 구한 것이다. 멜-주파수는 일반적으로 사용되는 선형적인 주파수 축이 아닌 인간의 청각 감도 특성을 반영한 비선형 주파수 축이다.

3.1 MFB spectral entropy

심리 음향 연구에 따르면 인간의 소리를 인지할 때 주파수에 대해 비선형적인 스케일을 가지며 이를 멜 스케일이라 한다[11]-[12]. 멜 스케일은 식 (5)에 정의 하였다.

$$f_{mel} = 2595\log_{10}(1 + \frac{f}{700}) \quad (5)$$

멜 스케일을 선형 주파수로 변환은 식 (6)에 나타내었다.

$$f = 700(10^{f_{mel}/2595} - 1) \quad (6)$$

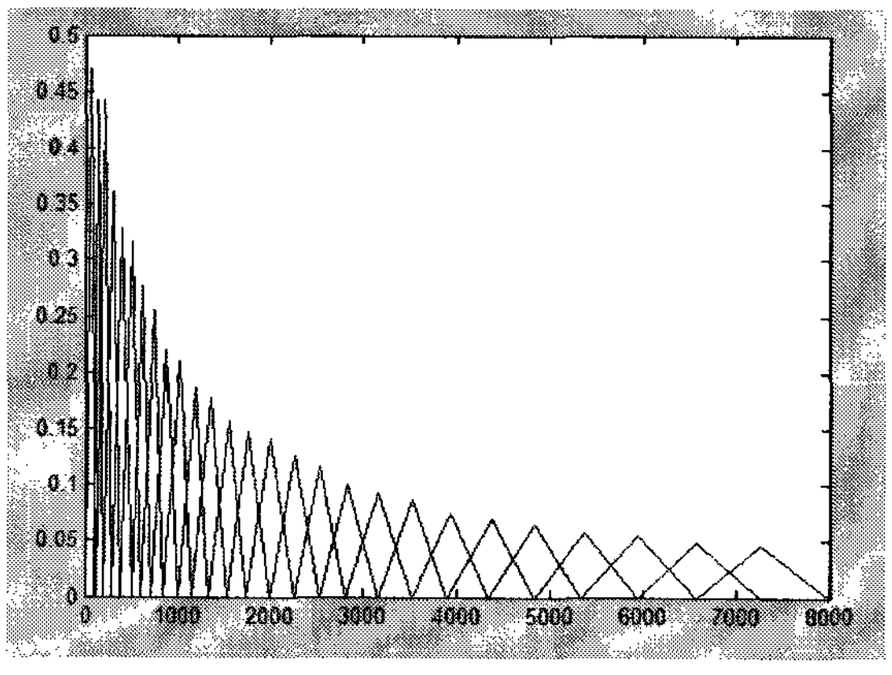


그림 2. 멜-주파수 필터 뱅크의 스케일
Fig. 2. Mel-frequency filter banks scale

멜-스케일은 인간의 저주파 대해 민감한 청각 특성을 반영한 것이다. MFB는 27개의 서브밴드 필터로 구성되어 있으며 FFT에 멜-스케일 필터를 곱한 것으로 그림 2에 나타내었다.

27개의 서브밴드로 된 것은 16kHz로 음성을 sampling 한 경우 FFT 변환하면 최대주파수 8000Hz이 되며, 0~8000Hz 주파수는 Mel-sclae 변환식인 식 (5)에 의해 0~2840.02 Mel 값으로 변환되며 28번째 필터의 중심 값이 8000Hz가 된다. 즉, 그림 2와 같이 삼각 필터의 중심주파수가 8000Hz에는 포함되지 않으며 mel-filter의 개수는 총 27개가 된다.

MFB spectral entropy는 주파수 정규화와 spectral entropy를 구하기 전에 멜-스케일 필터를 곱한 것으로 그림 3에 나타내었다. 그림 3에서와 같이 음성은 프리엠퍼시스 필터 $1-aZ^{-1}$을 사용하며 음성 입력 신호의 고주파 영역을 강조하기 위해 사용된다. 여기서 "a"의 범위는 0.9에서 1사이의 값을 가지며 기본 값으로 a는 0.97를 사용하였다.

프리엠퍼시스된 음성 신호는 단구간으로 분할되며 분할된 프레임마다 해밍윈도우 처리를 하게 된다. 해밍 윈도우 처리 방법은 입력 신호를 일정한 프레임으로 분할함으로써 생기는 불연속을 보완할 수 있는 방법이다. 윈도우 처리에 의하여 분할된 프레임에 존재하는 각각의 입력 신호를 주파수 영역의 신호로 변환하며 FFT를 이용하여 시간 영역의 입력 신호를 주파수 영역의 신호로 변환한다. 변환된 주파수 영역의 신호를 멜-스케일 필터와 곱하여 MFB 신호를 얻게 된다. MFB는 $M(b,n)$으로 정의 되며 식 (7)에 나타내었다.

$$M(b,n) = \frac{\sum_{i=L_b}^{U_b} V_b(i)S(i,n)}{\sum_{i=L_b}^{U_b} V_b(i)} \quad (7)$$

여기서 $M(b,n)$는 $S(i,n)$에 b번째 MFB의 i번째 멜-스케일을 곱하여 얻을 수 있다. L_b와 U_b는 i번째 멜 필터의 시작 주파수와 끝 주파수를 나타내며 멜-주파수 spectral 에너지를 $M(b,n)$로 정의 한다. $M(b,n)$는 n번째 프레임에 b번째 MFB 에너지이다. MFB spectral 에너지에 절대 값 및 로그 연산을 취한 후 정규화를 한다. 정규화는 $P[M(b,n)]$으로 정의되며 식 (8)에 나타내었다.

$$P[M(b,n)] = \frac{M(b,n)}{\sum_{m=1}^{B} M(m,n)} \quad (8)$$

여기서 B는 멜-필터의 개수를 나타내며 본 논문에서는 27개의 필터를 사용하였다. 제안된 MFB spectral entropy는 식 (9)에 나타내었다.

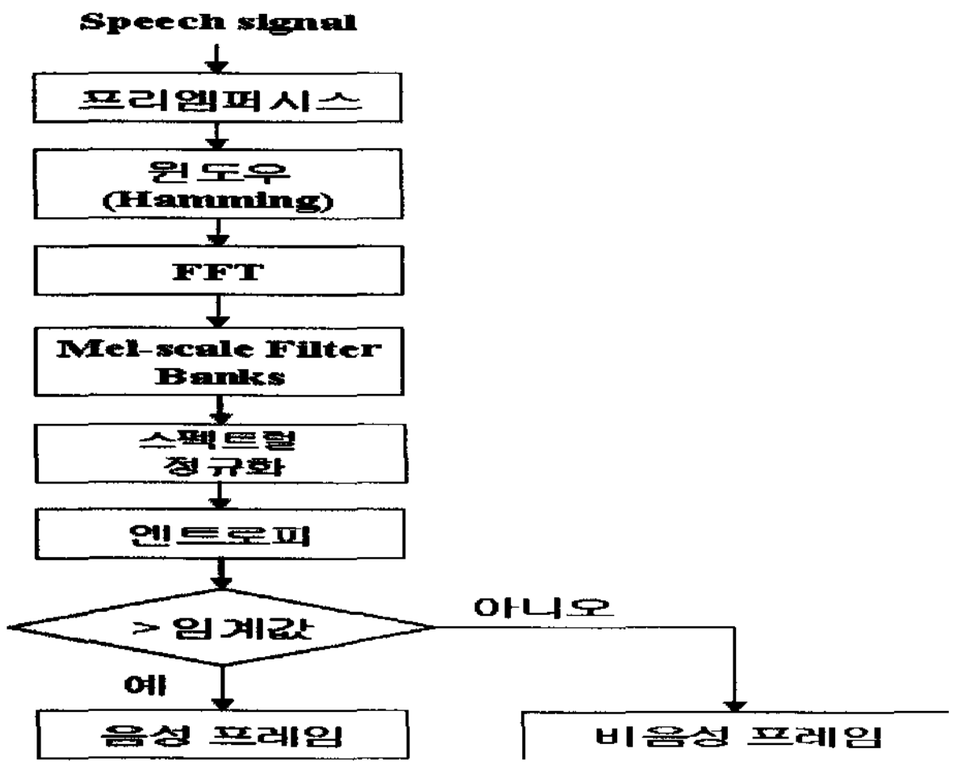


그림 3. 음성 구간 검출을 위해 제안된 MFB spectral entropy의 블록 다이어그램

Fig. 3. The block diagrams of proposed MFB spectral entropy method for VAD

$$H_{MFB}(n) = -\sum_{b=1}^{B} P[M(b,n)]\log P[M(b,n)] \qquad (9)$$

이전에 계산된 MFB spectral entropy 값을 사용하여 임계치를 설정하고 설정된 임계치를 사용하여 입력되는 각각의 프레임에 대해 음성 및 비음성을 구별하게 된다. 그리고 MFB spectral entropy의 값이 음성 및 비음성에 큰 차이가 나는 일부 필터만을 사용 가능하며 본 논문에서는 5개의 멜 필터를 선정하였다.

3.2 기울기 FFT spectral entropy

기울기 FFT spectral entropy는 인접한 스펙트럼 성분의 차로 entropy를 구하며 주파수의 변화를 반영한다. 기울기 FFT spectral entropy는 그림 4(a)에 나타내었다.

기울기 FFT spectral entropy는 정규화와 entropy를 구하기 전의 스펙트럼의 기울기이며 n번째 프레임의 $i+1$의 주파수 성분에서 n번째 프레임에서 i번째 주파수 성분을 차감 한 것이다. 기울기 FFT 스펙트럼은 $S^g(i,n)$으로 정의되며 식 (10)에 나타내었다.

$$S^g(i,n) = S(i+1,n) - S(i,n) \qquad (10)$$

기울기 FFT 스펙트럼에 로그와 절대 값을 씌우고 기울기 스펙트럼 정규화를 한다. $P[S^g(i,n)]$은 기울기 스펙트럼의 정규화이며 식 (11)에 나타내었다.

$$P[S^g(i,n)] = \frac{|S^g(i,n)|}{\sum_{m=1}^{M/2} |S^g(m,n)|} \qquad (11)$$

제안한 기울기 FFT spectral entropy는 $H^g(n)$으로 정의되며 식 (12)에 나타내었다.

$$H^g(n) = -\sum_{i=1}^{M/2} P[S^g(i,n)]\log P[S^{g(i,}n)] \qquad (12)$$

이전에 계산된 기울기 FFT spectral entropy 값을 사용하여 임계값을 구하며 설정된 임계값을 사용하여 입력되는 각각의 프레임에 대해 음성 및 비음성을 구별한다.

3.3 기울기 MFB spectral entropy

기울기 MFB spectral entropy는 인접한 MFB 사이의 차로 구하며 주파수의 변화에 대한 정보를 포함한다. 기울기 MFB spectral entropy는 그림 5(b)에 나타내었다.

기울기 MFB spectral 에너지는 $M^g(b,n)$으로 정의되며 n번째 프레임의 b+1번째 필터에서 n번째 프레임의 b번째 필터를 차감하였다. 기울기 MFB spectral 에너지를 로그 및 절대 값을 취한 후 기울기 MFB spectral 정규화를 한다.

기울기 MFB spectral 정규화는 $P[M^g(b,n)]$으로 나타내며 식 (13)에 나타내었다.

$$P[M^g(b,n)] = \frac{|M^g(b,n)|}{\sum_{m=1}^{B} |M^g(m,n)} \qquad (13)$$

제안한 기울기 MFB spectral entropy는 식 (14)에 나타내었다.

$$H^g_{MFB}(n) = -\sum_{b=1}^{B} P[M^g(b,n)]\log P[M^g(b,n)] \qquad (14)$$

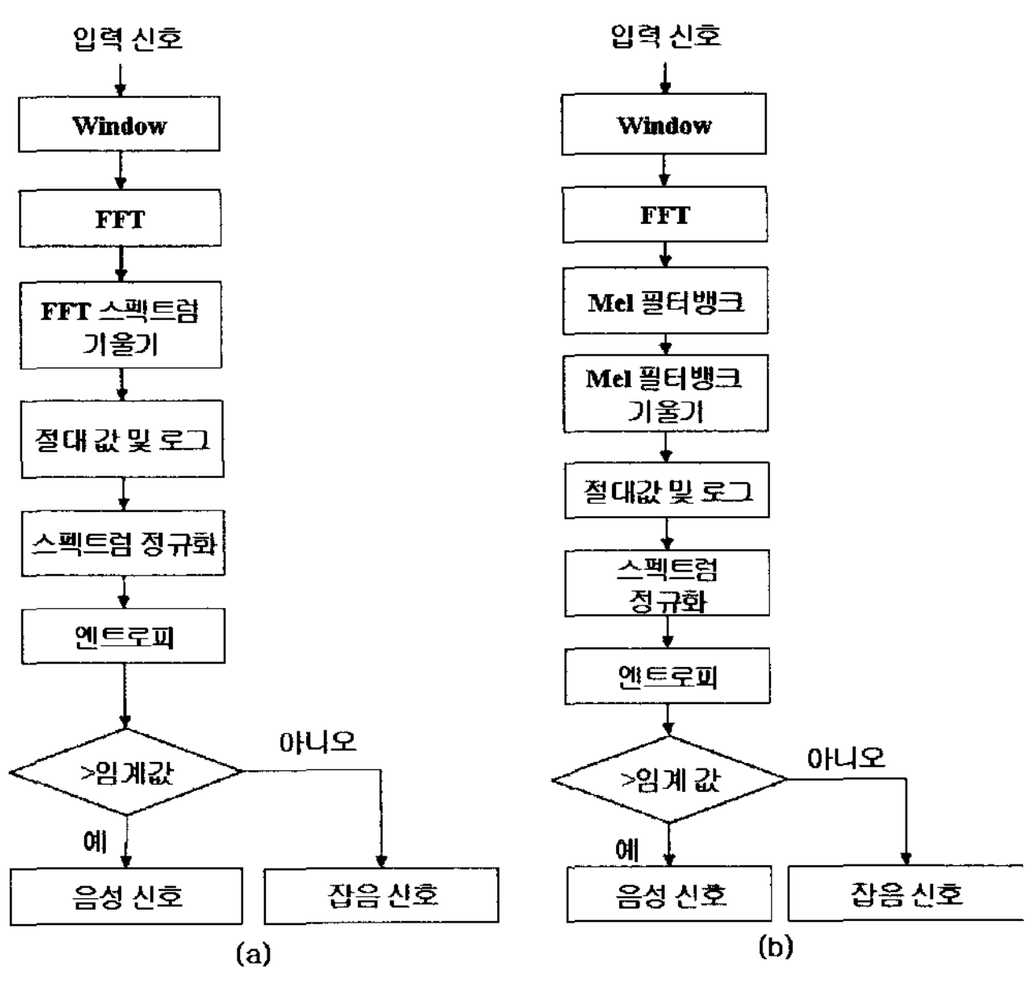


그림 54. 음성 구간 검출을 위해 제안된 방법에 대한 블록 다이어그램: 기울기 FFT spectral entropy 방법(a)과 기울기 MFB spectral entropy 방법(b)

Fig. 4. The block diagrams of the proposed VAD methods: the FFT gradient spectral entropy method (a) and the MFB gradient spectral entropy method (b).

Ⅳ. 실험 및 결과

4.1 실험 환경

실제 운전 환경과 동일한 상태에서 DB 구축을 하였으며 0km/h(정지상태) ~ 140km (20km 간격)으로 DB를 구축하였다 [13]. DB는 SM3 및 쏘렌토 차량에서 남성 및 여성 화자에 대해 A-J까지의 10가지 상태이며 10가지 상태는 표1에 나타내었다. 녹음 음성은 차량 내에서 자연스럽게 발성한 음성이며 그 예로 "어! 차가가네.", "80km로 이동 중입니다", "120km로 달리니까 무섭다" 등 DB의 목록은 정하지 않고 실제 차량에서 대화하는 듯이 음성 DB를 구축하였다. A-J 상태 DB는 각 상태마다 5분 이상에서 10분 이하로 발성을 하였으며 각 상태의 DB는 1분에 약 15-20 문장을 발성하였다.

음성 DB 수집은 Visual C++ 6.0을 사용하여 음성 녹취 프로그램을 구현 하였으며 16kHz 샘플링, 16비트 양자화를 한 PCM 파일이다. 한 프레임은 32ms, 50%의 오버랩 하였으며 해밍 윈도우를 사용하였다. FFT, MFB 기반의 entropy를 사용하여 실험을 하였으며 MFB는 총 27개를 사용하였다. PDA에서의 음성 구간 검출은 8kHz 샘플링, 16비트 양자화하여 실험을 하였으며 프로그램 구현을 위해 Embedded Visual C++ 4.0을 사용하였다. 한 프레임은 64ms, 50% 오버랩을 하였으며 제안한 방법 중 성능이 가장 우수한 MFB entropy를 사용하여 음성 구간 검출 방법을 포팅하였다.

표 1. 차량용 DB의 10가지 상태
Table 1. Ten state of Car DB

상태	차량의 상태
A	창문 close, 히터 off, 라디오 off, 시내도로이동
B	창문 open, 히터를 off, 라디오 off, 시내도로이동
C	창문 close, 히터 on, 라디오 off, 시내도로이동
D	창문 close, 히터 on, 라디오 on, 시내도로이동
E	창문 open, 히터를 off, 라디오 off, 시내도로이동
F	창문 close, 히터를 off, 라디오 off, 고속도로이동
G	창문 close, 히터 on, 라디오 off, 고속도로이동
H	창문 close, 히터를 off, 라디오 on, 고속도로이동
I	창문 close, 히터 on, 라디오 on, 고속도로이동
J	창문 open, 히터를 off, 라디오 off, 고속도로이동

4.2 실험 과정 및 결과

실제 차량 환경에서 녹취한 차량용 음성 DB를 음성 및 비음성으로 분할하였다. 실험은 기존의 spectral entropy 방법 및 제안한 3가지 방법, 총 4가지의 방법을 사용하여 각각에 대한 음성/비음성의 임계치를 결정하였다. 제안한 spectral entropy 기반의 방법은 3장에서 설명한 바와 같이 MFB spectral entropy, 기울기 FFT spectral entropy, 기울기 MFB spectral entropy로 총 3가지이다. 필터 선택은 계산량을 줄이기 위해 5개를 선택하였으며 상위 5개의 필터의 경우 제안한 entropy 기반의 방법들은 각 필터의 VAD 성능이 80% 이상이었다. 사용된 임계치는 표 2에서 나타내었다. MFB entropy의 경우 PAD의 포팅 하였을 때 8.5로 선정하였으며 실제 프로그램에서는 8.0, 8.5, 9.0과 같이 변화가 가능하도록 하였다.

표 2. 음성 검출을 위한 entropy의 임계값
Table 2. Threshold of entropy for VAD

파라메타	임계 값
spectral entropy	1.0
MFB spectral entropy(PC)	0.6
MFB spectral entropy(PDA)	8.5
기울기 FFT spectral entropy	1.1
기울기 MFB spectral entropy	0.9

또한 각 파라메타에 대한 모든 필터를 같은 값을 적용하였으며 임계값 이상의 경우 음성, 이하인 경우 비음성으로 판단하였으며 5개의 필터가 모두 음성인 경우에만 음성으로 판단하였다.

그림 5. 은 MFB entropy를 이용한 필터 선정의 예를 나타내었다. DB는 음성/잡음, 현재 속도, 프레임 수로 나타내었으며 기존에 마킹한 음성/잡음과 일치하는 프레임 수를 나타내었다. 필터는 차량마다의 특성, 차량의 속도 등에 따라 인식률이 다르며 그림 5의 경우 21번 필터에서 가장 높은 인식률을 90.38%을 나타내었다.

표 3. spectral entropy 기반의 방법을 사용한 음성 및 표비음성 프레임 판단 결과
Table 3. The result of speech and non-speech frame decision using spectral entropy-based method

파라메타	전체 프레임수	매칭된 프레임수	검출률
spectral entropy	129206	116324	90.03%
MFB spectral entropy	129206	120436	93.21%
기울기 FFT spectral entropy	129206	119257	92.30%
기울기 MFB spectral entropy	129206	120085	92.94%

표 3. 은 기존의 spectral entropy와 제안된 spectral entropy 기반의 방법들에 대한 음성/비음성 프레임을 판단한 결과이다. 실

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27
20Km 음성 (128프레임)	82	81	81	83	89	94	100	100	99	103	104	93	87	74	76	86	83	70	70	88	90	78	90	95	80	88	90
20Km 노이즈 (152프레임)	5	8	11	21	54	79	101	139	149	139	147	152	152	152	152	152	152	152	152	152	151	137	128	139	148	149	128
40Km 음성 (90프레임)	66	66	65	66	66	70	68	73	69	78	77	75	65	52	42	43	48	38	56	69	68	50	56	62	58	47	62
40Km 노이즈 (223프레임)	5	5	5	10	16	32	50	75	107	148	196	216	223	223	223	223	223	223	223	226	226	225	224	220	226	225	220
60Km 음성 (95프레임)	65	65	65	66	66	65	68	66	62	76	78	75	74	64	56	52	58	59	70	80	79	62	78	75	72	71	63
60Km 노이즈 (147프레임)	5	5	7	11	9	15	28	26	34	56	71	114	142	147	147	147	147	147	147	147	147	147	147	147	147	147	147
80Km 음성 (128프레임)	87	87	87	87	87	91	89	85	88	88	98	99	111	93	87	90	106	99	91	103	107	108	79	94	89	82	97
80Km 노이즈 (71프레임)	5	5	7	5	11	15	12	16	14	20	45	57	69	71	71	71	71	71	71	71	71	71	63	71	71	71	66
합 계 (1039프레임)	320	322	328	349	398	481	518	580	622	706	816	888	923	881	859	869	893	862	885	936	939	878	865	903	891	860	893
인식 율	30.80	30.99	31.57	33.59	38.31	44.37	49.86	55.82	59.87	67.95	78.54	85.47	88.84	84.79	82.68	83.64	85.95	82.96	85.1	90.09	90.38	84.50	83.25	86.91	85.76	82.77	85.95

그림 55. MFB entropy를 이용한 필터 선정의 예

Fig. 5. Example of selected filter using MFB entropy

험에서 사용된 전체 프레임수는 129206개를 사용하였으며 차량용 음성 DB를 분할할 때 음성/비음성 프레임으로 판단한 데이터와 비교하였다. 매칭된 프레임 수는 음성/비음성으로 마킹된 결과와 VAD 방법으로 판정된 음성/비음성의 결과가 매칭된 수를 나타낸다. 실험 결과 기존의 spectral entropy는 90.03%의 정확도를 가졌으며 제안한 모든 음성 구간 검출 방법들은 기존 spectral entropy 방법보다 성능이 우수하였다.

4.3 실제 차량 환경에서의 VAD

제안한 VAD 방법 중 성능이 가장 우수한 MFB spectral entropy를 이용한 음성 구간 검출 방법을 포팅하였으며 제안한 VAD 방법을 실제 차량 환경에서 실험 하였다.

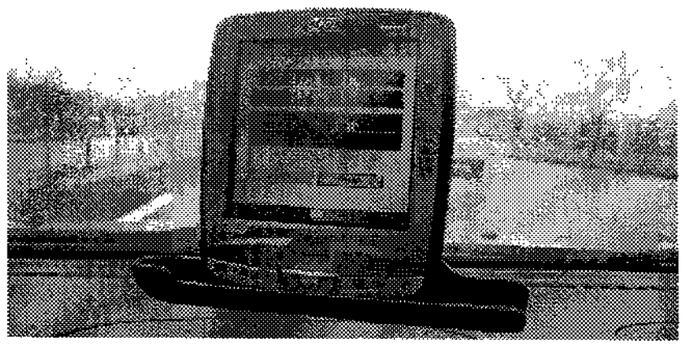
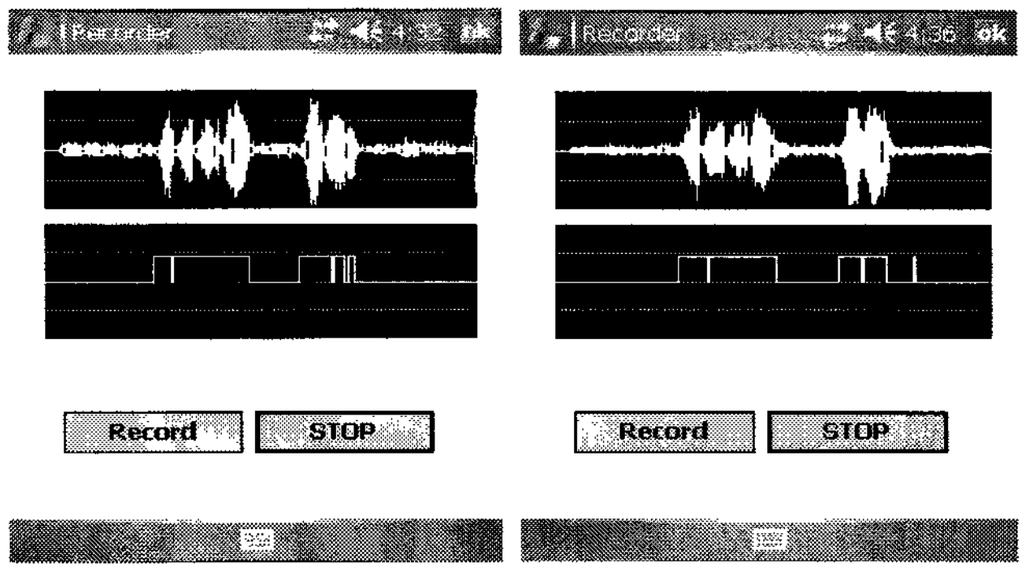


그림 6. 실제 차량 환경에서 제안한 MFB spectral entropy를 사용한 음성 구간 검출

Fig. 6. VAD for proposed MFB spectral entropy in real car environment

그림 6은 PDA에서 MFB spectral entropy를 사용하여 음성 구간 검출 하는 것을 나타내었다. 27개의 MFB에서 음성/비음성의 MFB entropy 값이 가장 큰 차이를 보이는 5개의 필터를 선정하여 사용하였다. 그림 7은 5개의 필터를 사용하여 음성 구간 검출 결과를 캡처 하였으며 (a)는 sixty kilometers driving (b)는 eighty kilometers driving로 발성을 하였다.



(a) (b)

그림 7. PDA에서 발성음 및 음성 구간 검출 결과 캡션; (a) "sixty kilometers driving" (b) "eighty kilometers driving

Fig. 7. Capture of utterance and VAD in the PDA; (a)"sixty kilometers driving" (b) "eighty kilometers driving"

선택된 5개의 필터를 사용할 경우 차량 노이즈에 대해 강인하며 계산량이 감소하기 때문에 차량용 디바이스에 실시간으로 음성/비음성 판별이 가능한 장점이 있다. 선택된 필터는 7, 12, 17, 20, 25이며 이것은 20km/h 간격의 속도로 필터를 선택하여 음성/비음성의 차이가 큰 5개의 필터를 사용한 것이다. 한 가지 예로 20km/h 속도로 운전 중에는 8, 12, 17, 20, 23번 필터가 음성/비음성의 차이가 가장 컸다.

Ⅵ. 결 론

본 논문에서는 실제 차량 잡음 환경에서의 음성 구간 검출이 가능한 spectral entropy 방법들을 제안하였다. 기존의 spectral entropy 방법을 기반으로 하여 5가지 VAD 방법을 제안하였으며 제안한 방법은 MFB spectral entropy, 델타 MFB spectral entropy, 기울기 FFT spectral entropy, 기울기 MFB spectral entropy이다. 실제 차량 환경에서 녹취한 DB를 사용하여 실험을 하였으며 기존 논문에서 제시된 spectral entropy와 제안한 방법들을 비교 실험 하였다. MFB spectral entropy 방법을 사용한 경우 가장 높은 인식률을 얻었으며 93.21% 였다. 기존 spectral entropy와 비교 했을 경우 약 3.2% 성능 향상이 있었다.

또한 노트북 기반의 음성 및 잡음 구간 검출 방법을 PDA에 포팅하여 실제 차량 환경에서 운전자가 발성한 음성에 대해 음성 및 비음성 구간을 검출 하였다. 실제 PDA 환경에서 제안한 MFB spectral entropy를 사용하여 음성 및 비음성 구간을 검출하였다. 차량의 속도의 변화에 따른 MFB 필터 선택이 가능하며 5개의 필터만을 사용하여 음성/비음성 프레임을 검출 하였다. 차후 잡음 환경에서 음성 구간 검출을 위해 다양한 방법을 결합하여 사용을 하면 더 좋은 성능을 얻을 것으로 사료 된다.

참 고 문 헌

[1] 김태석, 장종칠, “연속음성인식을 위한 음성구간과 피치 검출에 관한 연구”, 멀티미디어학회논문지, 제8권, 제1호, pp. 56-61, 2005

[2] Nemer E., Goubran R., Mahmoud S., "A Robust Voice Activity Detection Using Higher-Odor Statistics in the LPC Residual Domain", IEEE Transactions on Speech Audio Processing, Vol. 9, No. 3, pp. 217-231, 2001

[3] 정용주, 이승욱, “자동차 잡음환경 고립단어 음성인식에서의 VTS와 PMC의 성능비교”, 음성과학 제 10권 제 3호, pp. 251-261, 2003

[4] Rongqing H., Hansen J.H.L.; "Advances in Unsupervised Audio Classification and Segmentation for the Broadcast News and NGSW Corpora", IEEE transaction on Audio, Speech and Language Processing, Vol 14, No 3, pp.907-919, 2006

[5] Mauuary, L., Monne, J "Speech/non-speech Detection for Voice Response Systems," in Eurospeech'93, Berlin, Germany, pp. 1097-1100, 1993

[6] Yong-Wan Roh, Jong-Woo Choi, Dae-Sub Yoon, Hyun-Suk Kim, and Kwang-Soek Hong, "Delta FBLC based Speech/Non-Speech Frame Decision in Real Car Environment", The 4thConference on New Exploratory Technologies (NEXT 2007), pp. 244 - 247, 2007

[7] Almpanidis, G., Kotropoulos C.,"Voice Activity Detection Using Generalized Gamma Distribution", Proceeding of SETN 2006, LNAI 3955, pp.3-12, 2006

[8] F. Beritelli et al. "A Robust Voice Activity Detetor for Wireless Communications Using Soft Computing", IEEE Journal on selected areas in communication, Vol. 16, No. 9, pp. 1817-1829, 1998

[9] Shen J.-L., Hung J.-W. and Lee L.-S. "Robust Entropy-based Endpoint Detection for Speech Recognition in Noisy Environment", Proc. Int. Conf. on Spoken Lang Processing, Sydney ICSLP-98, 1998, CD-ROM

[10] Renevey, P., Drygajlo, A., "Entropy based voice activity detection in very noisy conditions" In: Proc. Eurospeech-2001, 2001, 9, pp. 1887 - 1890.

[11] Chakroborty S., Saha G., "Improved Text-Independent Speaker Identification using Fused MFCC & IMFCC Feature Sets based on Gaussian Filter", International Journal of Signal Processing Vol 5, Number 1, 2008

[12] Ben Gold and Nelson Morgan, "Speech and Audio Signal Processing", Part-IV, Chap.14 John Willy&Sons, pp. 189-203, 2002

[13] Javier Ranirez, Jose C. Segura, Carmen Benitez, Angel de la Torre, and AntonioJ. Rubio, "A New Kullback-Leibler VAD for Speech Recognition in Noise", IEEE Signal processing letters vol.11, no.2, 2004

[14] George Almpanides and Constantine Kotropoulos, "Voice Activity Detection Using Generalized Gamma Distribution" , SETN 2006, LNAI 3995, pp.3-12, 2006

[15] Ramirez.J, Yelamos.P, Puntonet.C.P, and Segura.J.C, "VSM-Enable Voice Activity Detection" , ISNN 2006, LNCS 3972, pp.676-681, 2006

[16] Gorriz.J.M, Ranirez.J, Puntonet.C.G, Lang.E.W and Stadlthanner.K, "Independent Component Analysis Applied to Voice Activity Detection", ICCS 2006, Part I, LNCS 3991, pp.234-241, 2006

[17] Martin A., Mauuary L.,"Robust speech/non-speech detection based on LDA-derived parameter and voicing parameter for speech recognition in noisy environments"Journal of Speech Communication pp. 191 - 206, 2006

[18] 노용완, 이우석, 홍광석외 “FFT와 MFB 기반의 VAD 성능 평가” 한국신호처리시스템학회 추계학술대회, 8권 2호, pp. 193-198, 2007

노 용 완(Yong-Wan Roh)

2001년남서울대학교 정보통신공학과(공학사)
2003년성균관대학교 정보통신공학부(공학석사)
2005년 ~ 현재 성균관대학교 정보통신공학부 박사과정

※주관심분야 : 오감 인식 및 오감 표현

이 규 범(Kue-Bum Lee)

2006년 서울보건대학 전산정보처리학과
2006년~현재 성균관대학교 정보통신공학부 (석사과정)

※주관심분야 : HCI, 멀티모달 인식 및 인터페이스

이 우 석(Woo-SeokLee)

2006년 대진대학교 전자공학과(학사)
2006년~현재 성균관대학교 정보통신공학부 (석사과정)

※주관심분야 : 음성 인식, 감정 인식

홍 광 석(Kwang-Seok Hong)

1985년 성균관대학교 전자공학과(학사)
1988년 성균관대학교 전자공학과(공학석사)
1992년 성균관대학교 전자공학과(공학박사)
1990년~1993년 서울보건전문대학 전산정보처리과 전임강사
1993년~1995년 제주대학교 정보공학과 전임강사
1995년~현재 성균관대학교 정보통신공학부 교수

※주관심분야 : 음성인식 및 합성, HCI