말소리와 음성과학 제5권 제1호 (2013. 3. 31)
ISSN 2005-8063, pp. 11~16

음성구간검출을 위한 비정상성 잡음에 강인한 특징 추출

Robust Feature Extraction for Voice Activity Detection in Nonstationary Noisy Environments

홍 정 표[1] · 박 상 준[2] · 정 상 배[3] · 한 민 수[4]
Hong, Jungpyo · Park, Sangjun · Jeong, Sangbae · Hahn, Minsoo

ABSTRACT

This paper proposes robust feature extraction for accurate voice activity detection (VAD). VAD is one of the principal modules for speech signal processing such as speech codec, speech enhancement, and speech recognition. Noisy environments contain nonstationary noises causing the accuracy of the VAD to drastically decline because the fluctuation of features in the noise intervals results in increased false alarm rates. In this paper, in order to improve the VAD performance, harmonic-weighted energy is proposed. This feature extraction method focuses on voiced speech intervals and weighted harmonic-to-noise ratios to determine the amount of the harmonicity to frame energy. For performance evaluation, the receiver operating characteristic curves and equal error rate are measured.

Keywords: voice activity detection, robust feature extraction, harmonic-to-noise ratio, harmonic-weighted energy

1. 서론

음성이 차세대 인터페이스로 각광받기 시작하면서 음성인식 기술에 대한 기대가 증대되고 있다. Tablet-PC, 스마트 폰 등 휴대용 전자 장치 뿐만 아니라 TV, 청소기 등의 가전제품, 자동차에 이르기 까지 음성을 인터페이스로 접목하려는 연구가 활발히 진행되고 있다.

정확한 음성인식 결과를 얻기 위해서는 음성구간을 정확히 찾아서 입력하는 것이 중요하다. 음성 구간 검출 (voice activity detection, VAD)이 음성 부호화기, 잡음제거, 음성인식기의 성능에 직접적인 영향을 미치기 때문에 정확한 음성구간 검출을 위한 연구가 지난 수십 년간 활발히 수행되었다[1-5]. 음성 입력이 조용한 환경에서는 충분히 정확하게 검출 되지만 실생활은 다양한 잡음에 노출되어있다. 이러한 편리성과 안전성을 갖춘 음성 인터페이스를 활용하기 위해서 가장 중요한 문제 중에 하나가 잡음의 개입이다.

잡음은 에어컨, PC 팬 등에서 유발하는 정상적인(stationary) 잡음과 그밖에 TV, 음악, 사람 목소리 등 시간에 따라 상태가 급격히 변하는 비정상성(nonstationary) 잡음으로 나뉜다. 정상성 잡음은 위너필터(Wiener filter) 와 칼만 필터(Kalman filter)에 의해 충분히 제거 할 수 있기 때문에 음성인식에 큰 문제가 되지 않는다[6-7]. 그러나 비정상성 잡음의 경우, 잡음의 종류가 다양하고, 잡음의 주파수 특성이 시간에 따라 급격히 변하기 때문에 잡음의 스펙트럼 파워 추정이 쉽지 않다. 이러한 비정상성 잡음을 제거하기 위해 멀티채널 기반의 빔포밍과 암묵신호 분리 방법이 널리 연구되어 왔으나, 음성입력이 단채널일 경우, 위의 두 방법은 사용할 수 없으며, SNR이 매우 낮은 음성입력의 경우 비정상성 잡음은 음성의 구간검출에 영향을 미칠 만큼 남아있는 경우가 많다[8].

기존의 음성인식 전처리 단은 정확한 음성구간 검출을 위해 잡음제거 단을 필수적으로 수반한다. 하지만 본 논문에서는 잡음 제거의 관점이 아닌 비정상성 잡음이 존재하는 상황에서 비정상성 잡음에 강인한 특징을 추출하는 것을 목표로 한다. 기존의 특징추출 방법 중, 가장 대표적인 방법인 에너지

1) 한국과학기술원, 전기 및 전자공학과, hansin@kaist.ac.kr
2) 한국과학기술원, 전기 및 전자공학과, psj@kaist.ac.kr
3) 경상대학교, 전자공학과(공학연구원), jeongsb@gnu.ac.kr, 교신저자
4) 한국과학기술원, 전기 및 전자공학과, mshahn@ee.kaist.ac.kr

접수일자: 2012년 11월 6일
수정일자: 2013년 2월 29일
게재결정: 2012년 3월 13일

와 영교차율 (zero crossing rate, ZCR)을 이용한 특징 추출 방법은 잡음의 개입에 매우 취약하다. 보다 잡음에 강인한 특징 추출을 위해 spectral entropy, mean delta function 등 다양한 연구가 수행 되었으나, 에너지와 영교차율 보다 크게 성능이 좋지 못하고, 알고리즘 복잡도를 감안하면 에너지와 영교차율 보다 뛰어난 특징 추출 방법이라 할 수 없다.

따라서 본 연구에서는 비정상성 잡음이 존재하는 SNR이 낮은 환경에서도 음성의 구간을 검출하기 위해 주기성 (harmonicity)에 주목하였다. 음성의 70% 이상이 유성음으로 구성되어 있기 때문에 유성음에 초점을 맞추어 유성음 구간을 정확히 검출하기 위한 특징 추출 방법으로 harmonicity를 파라미터화한 harmonic-to-noise ratio (HNR)을 프레임 에너지에 가중하여 harmonic-weighted 에너지를 추출한다. 제안한 특징 추출 방법을 receiver operating characteristic (ROC) 커브 와 equal error rate (EER)를 통해 비교하여 성능의 우수성을 증명하였다.

논문의 구성은 다음과 같다. 기존의 특징 추출 방법을 소개하고, 제안한 특징 추출 방법을 자세히 설명한 후에 실험 및 결과를 분석하고 최종 결론을 맺고자 한다.

2. 기존의 특징 추출 방법

2.1 프레임 에너지와 영교차율

지난 수십 년 동안, 음성구간 검출 (end-point detection, EPD) 에 대한 연구가 활발했다 [1-5]. 그중 가장 간단하면서도 효과적인 방법이 프레임 에너지와 영교차율을 이용한 EPD 방법이다. 이 방법은 음성의 시작과 끝에서 영교차율이 급격히 증가하는 현상, 유성음(voiced speech), 무성음(unvoiced speech), 묵음 (silence) 간의 에너지 차이가 크다는 점을 활용하여 효과적으로 음성의 시작점과 끝점을 검출하는 방법이다 [1].

2.2 Spectral entropy

최근 각광받고 있는 특징 추출 방법 중에 하나로, SNR이 낮은 환경에서도 스펙트럼의 크기는 음성구간이 잡음구간보다 조직적으로 나타난다는 가정을 기본으로 한다. 이런 스펙트럼의 "조직적인 정도 (measure of organization)"를 Shannon의 정보의 엔트로피를 이용하여 표현 하고자 한데서 비롯되었다[2]. Shannon의 정보의 엔트로피는 수식 (1)과 같이 표현된다.

$$H(s)=\sum_{i=1}^{N_s} P(s(i)) \cdot \log_2(P(s(i))) \tag{1}$$

N_s, $s(i)$, $P(\cdot)$은 각각 부호 (symbol)의 개수, 부호 i, 후험적 (a posteriori) 확률을 의미한다. 이것을 스펙트럼 영역에 적용한 spectral 엔트로피는 수식 (2)와 같다.

$$H(|Y(t_0)|^2) = -\sum_{k=1}^{N_{FFT}/2+1} \{P(|Y(t_0,k)|^2) \cdot \log_2(P(|Y(t_0,k)|^2))\} \tag{2}$$

$$P(|Y(t_0,k_0)|^2)=\frac{|Y(t_0,k_0)|^2}{\sum_{k=1}^{N_{FFT}/2+1}|Y(t_0,k_0)|^2} \tag{3}$$

수식 (3) 은 스펙트럼 영역에서의 확률을 의미하며, $|Y(t_0,k)|^2$는 t_0 프레임에서의 스펙트럼 에너지를 나타낸다.

2.3 Mean delta function

Mean delta function (MDF) 또한 spectral 엔트로피와 함께 많이 연구되는 특징추출 방법의 하나로써, 파워스펙트럼의 spectral autocorrelation의 delta 의 절대값을 평균 취한 값이다 [3]. MDF는 다음과 같이 정의된다.

$$M_d(t)=\frac{1}{\triangle L}\sum_{l=L_1}^{L_2}|\triangle R_p(t,l)| \tag{4}$$

$$R_p(t,l)=\sum_{k=0}^{N_{FFT}/2-1-l} S(t,k)S(t,k+l) \tag{5}$$

$$\triangle R_p(t,l)=\frac{\sum_{q=-Q}^{Q} qR_p(t,l+q)}{\sum_{q=-Q}^{Q} q^2} \tag{6}$$

수식 (4)는 t번째 프레임의 MDF 값이다. L_1 과L_2 는 수식(6)의 값이 안정적인 값을 갖는 경계값을 설정한다. 수식 (5)는 파워스펙트럼($S(k)=|X(k)|^2$)의 spectral autocorrelation function (SACF)을, 수식 (6)은 SACF의 delta인 delta SACF (DSACF)를 나타낸다. N_{FFT} 는 FFT 크기를 Q는 delta의 차수를 의미한다.

3. 제안한 특징 추출 방법

기존의 에너지와 영교차율을 이용한 특징 추출 방법은 잡음의 개입에 매우 취약하다. 영교차율은 잡음의 개입 시기에 급격히 늘어 날 수 있으며, 특히, 비정상성 잡음이 존재할 경우, 잡음의 크기가 음성의 크기만큼 또는 그 이상 클 경우에는 false alarm이 증가한다. 따라서, 비정상성 잡음이 존재하는 환경에서 에너지를 보완할 수 있는 효과적인 특징추출 방법으로 harmonic-weighted 에너지를 제안하였다. <그림 1>는 제안

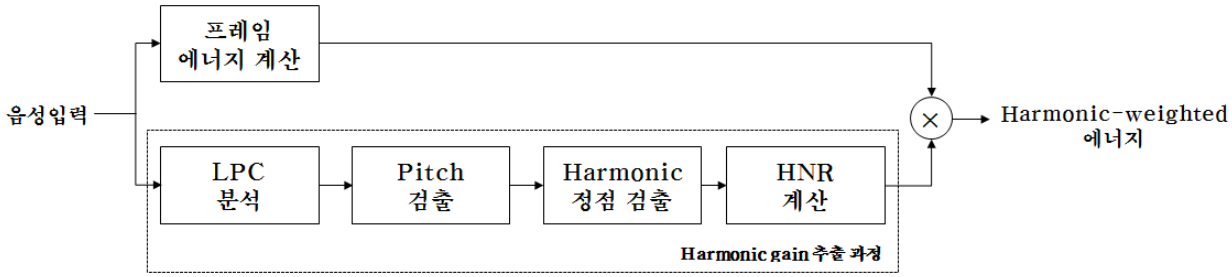

그림 1. 제안한 harmonic-weighted 에너지 추출 방법에 대한 블록도
Figure 1. Block diagram of the proposed feature extraction, harmonic-weighted energy

한 특징추출 방법에 대한 전체적인 블록도이다. 단구간 프레임 에너지와 주기성을 파라미터화한 harmonicity를 조합한 특징 추출 방법으로써, 비정상성 잡음이 존재하는 환경에 강인하다. 블록의 구체적인 내용은 다음과 같다.

3.1 프레임 에너지 계산

프레임 에너지는 음성 구간 검출에서 가장 간단하면서도 일반적인 특징 추출 방법이다. 프레임 에너지는 수식 (7) 과 같이 단구간 에너지 (short-time energy, STE)를 계산한다.

$$E(t)=\frac{1}{N_t}\sum_{n=0}^{N_t-1}x^2(n) \tag{7}$$

N_t은 한 프레임 당 샘플 개수를 나타낸다.

3.2 선형 예측 부호화 분석

선형 예측 부호화(Linear predictive coding, LPC) 분석 방법은 가장 널리 쓰이는 음성 분석 방법 중 하나로써, 인간의 발성 과정을 소스-필터 모델(source filter model) 이라는 이론적인 토대를 바탕으로 한 방법이다. 선형 예측 부호화 분석은 음성 샘플 간의 단구간 상관도(formants)를 모델링하고 필터링을 통해 엔벨로프(envelope)를 효과적으로 제거 할 수 있는 분석 방법이다[7]. 입력신호의 엔벨로프(envelope)를 제거하기 위해 Durbin 의 자기상관도 (autocorrelation)을 이용한 재귀적 방법 (recursive method)을 활용하였다[5].

3.3 피치 검출

피치 검출 알고리즘 (pitch detection algorithm, PDA)은 인간의 음성을 특징짓는 중요한 파라미터 중 하나이다. 시간영역, 주파수영역, 켑스트럼 영역에서 다양하게 구할 수 있으나 구현 복잡도 및 정확도를 감안했을 때 시간영역에서 입력 신호의 자기상관도(autocorrelation)를 이용한 방법을 일반적으로 사용한다[5]. 본 논문에서는 잡음의 개입에 의해 음성 입력신호를 이용한 피치 검출의 성능이 저하되기 때문에, 보다 정확한 피치 검출을 위해 음성의 엔벨로프(envelop)를 제거한 여기신호(LP residual)의 자기상관도를 이용하여 피치를 검출하였다. 또한, 인간의 피치는 표본화율(sampling rate)이 56~571 Hz의 범위를 가진다는 사실을 적용하였다[5].

$$F_0=\frac{f_s}{\tau} \tag{8}$$

$$F_{0,k}=\left\lfloor \frac{f_s}{\tau\bullet\left(\frac{f_s/2}{N_{FFT}/2+1}\right)}\right\rfloor=\left\lfloor\frac{N_{FFT}+2}{\tau}\right\rfloor \tag{9}$$

수식 (8)과 수식 (9)는 각각 Hz 단위와 입력신호를 Fast Fourier Transform (FFT)을 이용하여 주파수 분석 했을 때, 주파수 빈으로 수식 (8) 이 양자화 된 것이다. τ, f_s 는 각각 Pitch, 표본화 율을 나타낸다. 수식 (9)의 ( )안의 값은 FFT 해상도 (resolution)을 나타내며, $\lfloor\bullet\rfloor$ 연산자는 소수점이하 버림을 뜻한다.

3.4 Harmonic 정점 검출

검출된 Pitch를 이용하여 Harmonic 정점 (peak) 의 위치와 해당 harmonics의 amplitude를 구하고, 구해진 harmonic 정점 값들을 이용하여, 인접한 정점의 위치 평균값을 harmonic 정점 사이의 저점(valley)의 위치로 추정하였다. harmonic 정점과 저점은 각각 다음과 같이 표현된다.

$$f_{p,i}=i\bullet\left(\frac{f_s}{\tau}\right) \qquad i=1,\ldots,N_p \tag{10}$$

$$f_{v,j}=\frac{f_{p,j}+f_{p,j+1}}{2} \qquad j=1,\ldots,N_v \tag{11}$$

단, $N_v=N_p-1,\qquad N_p\bullet\left(\frac{f_s}{\tau}\right)<\frac{f_s}{2}$

N_p 와 N_v 는 각각 harmonic 정점과 저점의 개수를 의미한다. 수식 (10)과 (11)는 Hz단위의 값이기 때문에 해당 harmonic의 amplitude 값을 알기 위해서는 FFT 주파수 빈 단위로 양자화 해야 한다. 양자화된 harmonic 정점과 저점의 값은 다음과 같다.

$$k_{p,i} = \left\lfloor \frac{f_{p,i}}{\left(\frac{f_s/2}{N_{FFT}/2+1}\right)} \right\rfloor = \left\lfloor i \cdot \left(\frac{N_{FFT}+2}{\tau}\right) \right\rfloor \quad (12)$$

$$k_{v,j} = \left\lfloor \frac{f_{v,j}}{\left(\frac{f_s/2}{N_{FFT}/2+1}\right)} \right\rfloor = \left\lfloor \left(\frac{j \cdot (j+1)}{2}\right) \cdot \left(\frac{N_{FFT}+2}{\tau}\right) \right\rfloor \quad (13)$$

수식 (12) 과 수식(13)은 각각 harmonic 정점 및 저점의 주파수 빈 값으로 분모는 주파수 해상도로서 주파수 빈 하나당 대역폭을 의미한다. harmonic 정점 및 저점의 위치를 추정할 때, 주파수 빈의 정수배를 하지 않는 이유는 거듭되는 양자화 에러를 예방하여 보다 정확한 위치를 찾기 위함이다.

3.5 HNR 계산

주기성의 정도를 산술적으로 나타내기 위해서 harmonic-to-noise ratio (HNR)의 개념을 활용하였다. 원래 HNR은 장애음성 분야에서 장애도를 측정하는데 활용되는 개념으로 잡음을 추정하는 방법에 따라 다양한 방법이 있다[9]. 본 논문에서는 잡음을 harmonic 정점 사이의 저점 값의 합으로 추정하였다. HNR은 다음과 같이 표현된다.

$$HNR(t) = 20 \cdot \log_{10}\left(\frac{\frac{1}{N_p}\sum_{i=0}^{N_p-1} |X_R(t,k_{p,i})|}{\frac{1}{N_v}\sum_{j=0}^{N_v-1} |X_R(t,k_{v,j})|} \right) \quad (14)$$

$X_R(t,k)$는 t번째 프레임의 입력신호($x(t)$)의 LP residual 신호의 주파수 응답이다. HNR 값이 낮을수록 주파수 영역에서 주기성이 낮기 때문에 비음성구간일 확률이 높고, 높을수록 주기성이 높아서 유성음 구간일 확률이 높다.

3.5 Harmonic-weighted 에너지

위 과정을 통해 얻은 HNR을 기반으로 harmonic-weight를 계산하기 위해서 sigmoid function을 이용하였다. 본 논문에서 제안한 harmonic-weighted 에너지는 다음과 같이 프레임 에너지와 harmonic-weight의 곱으로 나타낼 수 있다.

$$W_h(t) = f[HNR[t]] = \frac{1}{1+e^{-HNR(t)}} \quad (15)$$

$$E_{HW}(t) = W_h(t) \cdot E(t) \quad (16)$$

수식 (15)와 (16)은 각각 harmonic weight 와 harmonic-weighted 에너지를 나타낸다.

4. 실험 및 결과

제안한 특징 추출 방법의 성능을 검증하기 위해 깨끗한 음성으로 PBW (Phonetically Balanced Word) 452 단어 4세트에 babble, car, restaurant, subway, train 등의 5가지 AURORA 잡음 [10]을 0, 5, 10, 15, 20 dB의 5가지 SNR에 맞춰 인위적으로 합성하였다 (총 45200 = 452 x 4 x 5 x 5 개의 잡음 샘플). False alarm rate를 측정하기 위해서 음성구간 이전과 이후에 1초씩 순수 잡음 구간을 추가하였다. 단구간 신호처리를 위해 한 프레임의 크기는 20 ms 단위로 50% 씩 중첩 (overlap) 하였다. 주파수 분석을 위한 FFT 크기는 512로 설정하였고, 합성된 음성샘플은 8 kHz 로 표본화 되었고, 16 bit의 해상도를 가진다. 선형 예측 부호화를 위한 차수는 16, 하모닉 정점의 개수, N_p 는 7로 설정 하였다.

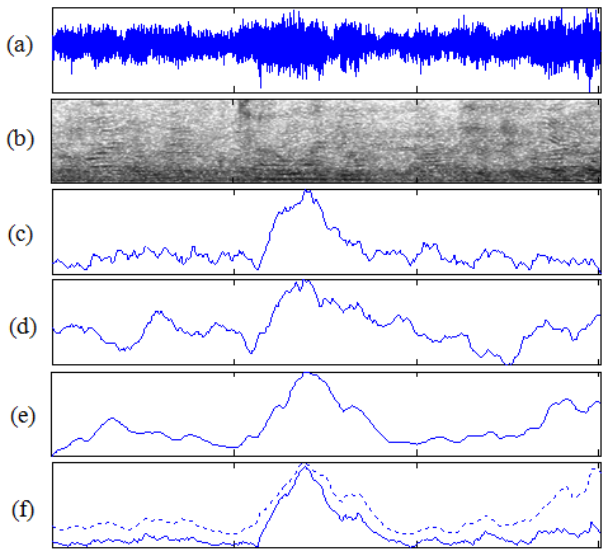

그림 2. 특징 개형 비교. (a) 음성파형, (b) 스펙트로그램, (c) HNR, (d) SE, (e) MDF, (f) STE (점선), 제안한 특징추출 방법 (실선)
(SE의 경우, 다른 특징들과 쉬운 비교를 위해 위아래를 뒤집음, 수평축의 한 칸은 1초에 해당함)

Figure 2. Comparison of feature contours. (a) Noisy waveform (0 dB, /chung-wa-dae/, babble noise), (b) Spectrogram, (c) HNR, (d) SE, (e) MDF, (f) STE (dotted), proposed (solid)
(For easy comparison with other features, the SE contour is reversed.
One second for a tick on the horizontal axis)

<그림 2>는 제안한 harmonic-weighted 에너지를 다른 feature contour와 비교한 것이다. <그림 2>의 (e)를 참조하면, 제안한 알고리즘이 순수 잡음구간에서는 (시작 1초, 마지막 1초 구간) feature 값이 작아진 반면에 음성구간은 비교적 유지되고 있다.

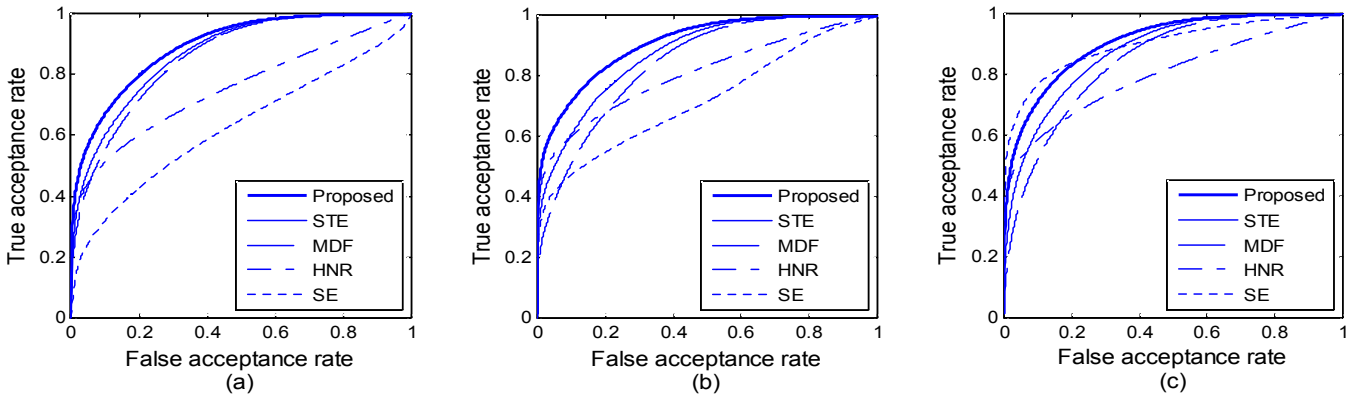

그림 3. 잡음 환경 별 여러 가지 특징추출 방법의 ROC 커브 비교. (a) babble, (b) car, (c) subway.
('Proposed' 는 제안한 방법, STE, HNR (수식 (14)), SE[2], MDF[3])
Figure 3. Comparison among ROC curves of the various feature extraction methods. (a) babble, (b) car, (c) subway.
(the proposed, STE, HNR in (14), SE[2], MDF[3])

제안한 특징 추출 방법을 보다 객관적으로 평가하기 위해서 ROC 커브를 측정하였다. 각 피쳐의 ROC 커브는 피쳐 도메인에서 잡음구간, 음성구간의 피쳐 분포에서 문턱치(threshold) 값을 옮겨가면서 True acceptance rate(TAR)와 False acceptance rate(FAR)을 측정한다. 음성구간의 피쳐 값의 평균이 잡음구간의 피쳐 값의 평균보다 크다고 가정했을 때, TAR은 음성구간의 피쳐 값이 특정 문턱치 보다 큰 경우, FAR은 잡음구간의 피쳐 값이 특정 문턱치 보다 큰 경우라고 할 수 있다. 따라서, ROC 커브의 한 점은 한 문턱치 값에 해당하는 TAR과 FAR의 좌표값이라 할 수 있고, 문턱치는 잡음구간의 피쳐 최소값부터 음성구간의 피쳐 최대값 까지 적당한 크기로 증가시킨다. <그림 3>은 잡음이 각각 babble, car, subway일 때의 ROC 커브를 그린 것이다. 커브가 좌상향 될수록 1에 가까운 값을 가지게 되므로 좋은 성능을 의미한다. ROC 커브의 아래 면적을 측정한 the area under an ROC curve (AUROC)는 전체 면적을 1로 봤을 때, 큰 값을 가질수록 좋은 성능을 나타낸다. <표 1> 은 다섯 가지 잡음에 대한 각 특징의 AUROC 값을 정리한 것이다. 제안한 특징, STE, HNR, MDF, SE가 평균적으로 0.9006, 0.8745, 0.7779, 0.8518, 0.6870을 나타내었다. <그림 3> 과 <표 1>를 보면 제안한 특징 추출 방법이 가장 좋은 성능을 보였다. 보다 성능에 신빙성을 더하기 위해 EER을 측정하였다. EER은 피쳐 도메인에서 음성을 잡음으로 잘못 검출하는 비율 (False reject rate, FRR)과 잡음을 음성으로 잘못 검출하는 비율 (False acceptance rate, FAR)이 같을 때의 에러 값으로 그 값이 낮을수록 변별력 있는 특징이라 할 수 있다. <표 2>는 다섯 가지 잡음환경에서 EER을 측정한 것이다. 평균적으로 제안한 특징, STE, HNR, MDF, SE 순으로 평균 19.32, 22.12, 29.86, 24.18, 36.52 의 EER을 보였다. ROC 커브와 EER 측정 결과를 보면, 다양한 잡음환경 및 전체 SNR에서 제안한 harmonic-weighted 에너지의 성능이 가장 높았다는 것을 알 수 있다. SE의 성능이 눈에 띄게 낮은 이유는 nonstationary 잡음에 harmonicity가 큰 경우가 포함되어 있을 뿐만 아니라, 특히 엔트로피가 잡음의 종류에 민감한 특성이 있다[3].

표 1. 다양한 잡음환경에서 특징추출 방법의 AUROC
Table 1. AUROC of feature extraction methods in various noisy environments

Noise	STE	HNR	MDF	SE	Proposed
Babble	0.8716	0.7528	0.8577	0.6255	0.8936
Car	0.8743	0.8078	0.8410	0.7187	0.9084
Restaurant	0.8773	0.7302	0.8667	0.6891	0.8920
Subway	0.8781	0.8000	0.8458	0.9004	0.9076
Train	0.8714	0.7989	0.8477	0.5011	0.9016
Average	0.8745	0.7779	0.8518	0.6870	0.9006

표 2. 다양한 잡음환경에서 특징추출 방법의 EER 결과 (%)
Table 2. EER results of feature extraction methods in various noisy environments(%)

Noise	STE	HNR	MDF	SE	Proposed
Babble	22.5	32.1	23.7	40.9	20.2
Car	22.4	27.6	25.5	36.1	18.7
Restaurant	21.4	33.8	22.6	36.7	19.9
Subway	21.6	28.2	24.5	17.6	18.4
Train	22.7	27.6	24.6	51.3	19.4
Average	22.12	29.86	24.18	36.52	19.32

5. 결론

본 논문은 비정상성 잡음이 존재하는 환경에서 음성구간검출의 성능을 향상시킬 수 있는 특징추출 방법에 대한 연구를 수행하였다. 기존의 특징 추출 방법인 에너지와 영교차율을 이용한 방법, spectral entropy를 이용한 방법, mean delta function을 이용한 방법 등과 ROC 커브, EER 결과를 비교해 본 결과 제안한 harmonic-weighted 에너지의 성능이 가장 높았다. 향후 연구 계획으로는 제안한 특징추출 방법을 EPD 결정룰 (decision rule)과 결합하여 EPD를 수행하고, 음성인식률을 측정하여 제안한 특징 추출 방법의 성능을 검증할 계획이다.

감사의 글

본 연구는 지식경제부의 산업원천기술개발과제의 일환으로 수행하였음. [10035252, 모바일 플랫폼 기반 대화모델 적용 자연어 음성인터페이스 기술 개발]

참고문헌

[1] Rabiner, L.R. (1975). An algorithm for determining the endpoints of isolated utterances. *The Bell System Technical Journal*, Vol. 54, No. 2, 297-315.

[2] Zoltan, T. (2005). Robust voice activity detection based on the entropy of noise-suppressed spectrum. *Interspeech*, 245-248.

[3] Ouzounov, A. (2004). A robust feature for speech detection. *Cybernetics and information technologies*, Vol. 4, No. 2, 3-14.

[4] Kondoz, A.M. (1994). *Digital speech: coding for low bit rate communication system*. UK: John Wiley & Sons.

[5] Rabiner, L.R. (1978). *Digital processing of speech signals*. USA: Prentice-Hall.

[6] Jeong, S. (2001). Speech quality and recognition rate improvement in car noise environments. *Electronics Letters*, Vol. 37, No. 12, 801-802.

[7] ETSI Std. (2005). Speech processing, transmission and quality aspects (STQ); distributed speech recognition; extended advanced front-end feature extraction algorithm; compression algorithm; back-end speech reconstruction algorithm. *ES 202 212* V1.1.2.

[8] Brandstein, M. (2001). *Microphone arrays: signal processing techniques and applications*. Berlin: Springer.

[9] Qi, Y. (1997). Temporal and spectral estimations of harmonics-to-noise ratio in human voice signals. *Journal of Acoustical Society of America*. Vol. 102, No. 1, 537-543.

[10] Hirsch, H. (2000). The aurora experimental framework for the performance evaluation of speech recognition systems under noisy conditions. *ISCA ITRW ASR2000*.

• **홍정표 (Hong, Jungpyo)**
한국과학기술원 전기 및 전자공학과
대전광역시 유성구 대학로 291
Tel: 042-350-5474 Fax: 042-350-7619
Email: hansin@kaist.ac.kr
관심분야: 음성신호처리
현재: 한국과학기술원 전기 및 전자공학과 박사과정

• **박상준 (Park, Sangjun)**
한국과학기술원 전기 및 전자공학과
대전광역시 유성구 대학로 291
Tel: 042-350-8074 Fax: 042-350-7619
Email: psj@kaist.ac.kr
관심분야: 음성신호처리
현재: 한국과학기술원 전기 및 전자공학과 박사과정

• **정상배 (Jeong, Sangbae)** 교신저자
경상대학교 공과대학(공학연구원)
경남 진주시 가좌동 900번지
Tel: 055-772-1727 Fax: 055-772-1729
Email: jeongsb@gnu.ac.kr
관심분야: 음성신호처리
현재: 경상대학교 전자공학과 조교수

• **한민수 (Hahn, Minsoo)**
한국과학기술원 전기및전자공학과
대전광역시 유성구 문지동 103-6
Tel: 042-350-6206
Email: mshahn@ee.kaist.ac.kr
관심분야: 음성신호처리
현재: 한국과학기술원 전기 및 전자공학과 교수